Real Time Sign Language Interpreter

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

Arun C S
Department of Electronics and Communication
Sree Buddha College of Engineering, Pattoor
Kerala, India
csaruncs@gmail.com

***Abstract*— Sign Language is a medium for communication for many disabled people. The sign language recognition has many applications including gesture controlled activities like human computer interaction, gesture controlled home appliances and other electronic devices and many applications that uses gesture as the trigger input. The most important application is it provides a communication aid for deaf and dumb people. The system for sign language recognition for deaf and dumb people is implemented in ARM CORTEX A8 processor board using convex hull algorithm and template matching algorithm. Image is obtained using webcam. This hand sign image is converted to text so as to develop a communication between normal and deaf and dumb people. Open CV is the software tool that provides the support with image processing techniques. The system converts sign language to text for deaf and dumb people to communicate with normal people. Moreover the system is used to control devices like Robot, Car Audio Systems, home appliances etc.**

Keywords—Convex Hull; Sign Language; HCIs; ARM CORTEX; Convex points; Defect points;Template matching.

I. INTRODUCTION

The paper employs a novel method of recognizing sign language employed in an embedded hardware. The system is used for many applications including deaf and dumb people communication. The paper presents a new approach of recognizing hand sign that can be used by the deaf and dumb community to interact with the normal people. The system can also be used for many other applications including home automation, triggering audio systems in automobile etc. The system also includes an automatic speech recognition system. Normal people are completely unaware of the sign language that the deaf and dumb people use. Hence communication with them is not possible unless they took proper training to learn this sign language.

Sign language recognition can be of two types based on the gestures used for communication. The gestures used for communication can be static or dynamic depending on the movement of hand signs. Hence the sign language can be static hand poses or dynamic hand gestures. Static signs are those with no motion. This provides only the 2-Dimensional information regarding the hand sign image. Mostly static symbols are shown to recognize alphabets and numbers. Dynamic gestures are shown corresponding to words that are either continuous string of alphabets or a gesture itself for a word [1]. The sign language interpreter system developed is capable of recognizing the standard American Sign Language (ASL) symbols.

The system is developed using an algorithm which does not require training for the system to start working. The algorithm used is Convex Hull algorithm or Jarvis algorithm for recognition of numbers and template matching for recognition of alphabets. The hardware is implemented on Beagle Bone Board with ARM CORTEX A8 processor architecture. The system requires an operating system to support large processing. The operating system involved here is Ubuntu based on Linux kernel.

The image processing is done using the software Open source computer vision [2]. The software tool Open CV provides with sufficient image processing functions that helps the processing of hand sign images using three basic approaches including recognition, detection and reconstruction. These approaches help in identification and detection of hand signs and obtain the corresponding text that normal people can understand.

II. LITERATURE REIVIEW

Gesture recognition has been an area of interest for many years [3]. There are several methods employed for gesture recognition. These methods of gesture recognition can be categorized into 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 [4]. 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 [5]. The input image and 2-D appearance of the hand models are compared. 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. 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 [6]. 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 subsequent method to be performed is After segmentation subsequent method to be performed is feature extraction and classification which will be done exploitation totally different mapping techniques. One shape matching that rely on contour that make use of Fourier descriptors and wave descriptors. B- Spline is another form matching technique supported contour. Another shape matching technique is local features based mostly object matching that used SURF and SIFT algorithms [10]. 2 classical approaches for extracting the analytical visage and trimming the volume are Principal component analysis (PCA) [8] and Linear Projection technique (LPP) [9][12]. Hand sign recognition based on PCA and LPP to extract feature vectors once the hand shape samples are normalized in rotation and size. Another system uses finger and palm arrangements for hand sign recognition. It employs contour points and biconvex hull or Jarvis algorithmic program to condense the features for posture recognition [11]. one-dimensional and unbend signs is recognized by this method

In this paper an embedded system is developed for deaf and dumb people communication using a Beagle Bone Black with ARM Cortex A8 processor architecture that act as the processing unit. A webcam is used for acquiring the input images that is forwarded for preprocessing actions. The image features are extracted using the edge detection methods and is forwarded for classification and matching procedures. The corresponding text for each hand sign is shown on the display device.

III. System Architecture

The system consists of a camera system, beagle bone black board and a display device that can be connected to a network. Fig 1 illustrates the hardware architecture of the proposed system. A USB camera is used to capture the image for the proposed system. The optical resolution provided by the system is 1280×960. It supports the system with a video capture up to 1024×768 pixels. The camera system supports a frame rate of 30 fps and the focal length of 4.0mm. The hand signs are captured by the camera system. T he beagle bone board is used to implement the system.

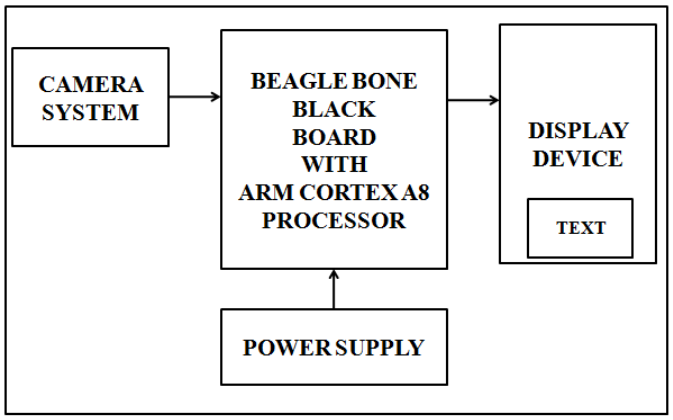

Fig. 1Hardware architecture

It consists of ARM CORTEX A8 processor. It is a 32 microprocessor with 1 GHZ processing speed. It supports high level operating systems. It consumes low power. The camera system is connected through USB connection. The camera system is used to acquire the images if sign language symbols. Display device attached shows the corresponding text of sign language symbols.

IV. Methodology

The recognition of hand sign and conversion to text can be accomplished in two stages. The numbers recognised using convex hull detection method and alphabets using template matching method. This section describes how numbers are identified using convex hull algorithm and alphabets using the template matching algorithm in real-time. The initial step in template matching is to create template images and store it in a database

A. Image Recognition Procedure

The image captured by the webcam undergoes various pre processing and processing stages to extract features and finally identification. Fig 2 shows the various processing stage of hand sign image. From the physical environment image is acquired using a camera module that can be directly connected to the controller board. Subsequent step is pre-processing that include skin colour extraction, reducing dimension using blurring, RGB to binary conversion, edge detection. After pre-processing 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 template matching algorithms from the database constituting template images created during the training phase.

In order to obtain the complete recognition system three different approaches are used. For pre processing process to obtain the binary image from the colour image canny edge detection algorithm is used. For feature extraction and obtaining the contour of the image and hence information about finger tips convex hull or Jarvis algorithm is used. For identification and recognition of finger alphabets template matching algorithm is utilized.

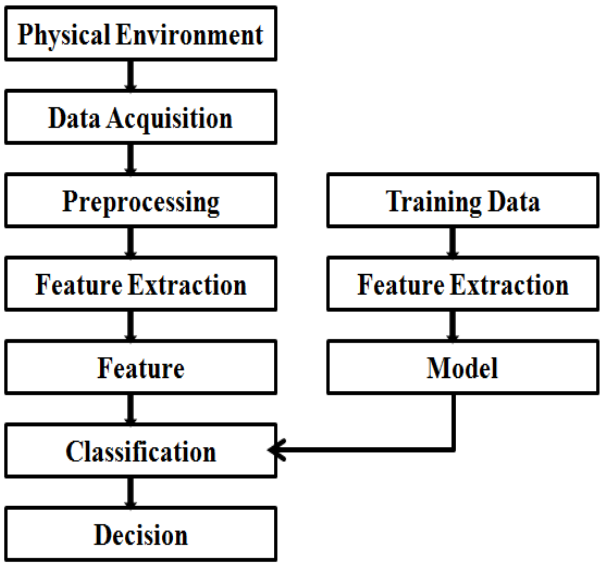

Fig 2 Hand Sign Recognition Process

In order to detect edges canny edge detection algorithm is employed. It is a multistage detection algorithm. In open CV all the steps in canny edge detection algorithm is concentrated to a single function cv2.Canny (img, 200, 100). Different steps are performed in this single step. First step is noise removal from the image. Edge detection is susceptible to various noise present in the image. 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. 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:

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

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

These processes are used to obtain the template images for creating the database during the training phase. The edge detection and thresholding to obtain binary images are also performed on the real time hand sign image captured before the recognition process.

B. Recognition of Number Using Convex Hull Method

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. The convex hull is performed to the threshold image to obtain the convex and defect points. The convex hull includes a set of continuous points in the Euclidean space that is connected to contours. On joining all convex points, a pentagon-shaped skull is drawn around the contour. Convex hull works as an envelope around the hand contour. Minimum points are used to form the hull and to incorporate contour points inside or on the hull while not disturbing the convexity property. This causes the formation of defects in the convex hull with respect to the contour drawn on hand. The count number of fingers in the hand sign shown is found by making use of the defect points in the hand posture

C. Recognizing Alphabets Using Template Matching

Template matching is a technique used in image processing to find small parts of an image that matches with a template provided in the database. It includes source image and template image. The source image (I) is the image in which we expect to find a match to the template image. The template image (T) is the image that will be compared to the source image. Cv2.matchTemplate is a function used in Open CV to calculate the matches. The template images are fed to the system by creating a database. The template matching algorithm aims at identifying the maximal matching area. The template image slides over the input image, the comparison starts with the template image and patch of the input image under the template image.

D. Database of Template Images

During the training phase the the raw image of various hand signs of alphabets are collected using webcam. This is pre-processed to obtain the features for template matching. The images are blurred and threshold to obtain the binary image which is the required template image. This pre recorded template images are stored in the database which is used for the matching purpose for real time detection of hand signs. The template images are shown in Fig 3. These images are obtained by following the steps explained in section III A. When a hand sign is shown in real time the input image after pre-processing is slided over each template image to obtain the corresponding match. After a particular match is obtained the corresponding text is identified and displayed.

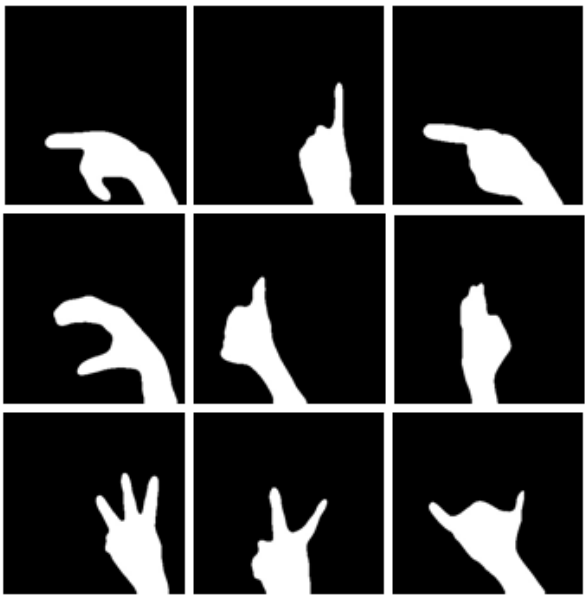

Fig. 3 Template Images Stored in Database

V. IMPLEMENTATION AND RESULTS

The system has been implemented using beaglebone board with Linux operating system and ARM CORTEX A8 processor. The python programming is used to implement the system. The camera system is connected to the controller through USB. The hardware implementation is shown in fig.4.

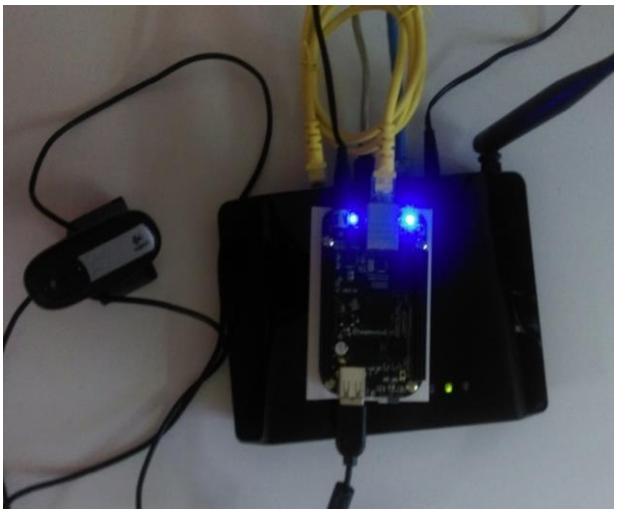

Fig. 4 Hardware implementation

The input to the system is shown in Fig 5.The symbols for numbers 2 and 3 and alphabets c and d are shown in Fig 5. The contour of the images is then obtained which is shown in Fig 6. The binary images are shown in Fig 7.this binary or threshold images are compared with the template images stored in database. After the match is performed corresponding text is obtained for each hand sign. This is shown in the output figure Fig 8.

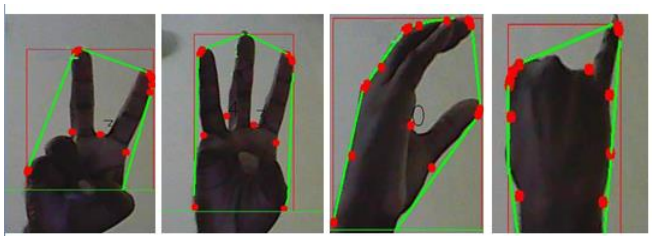

Fig 5 Input Hand Signs for Numbers 2,3 and Alphabets C, D

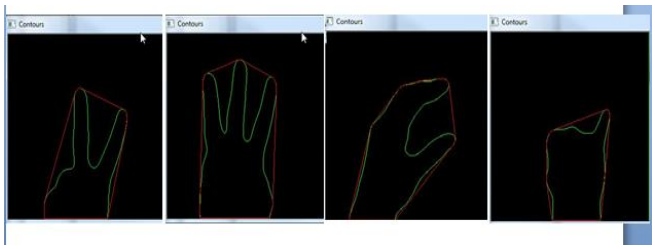

Fig 6 Contours of Signs for Numbers 2,3 and Alphabets C, D

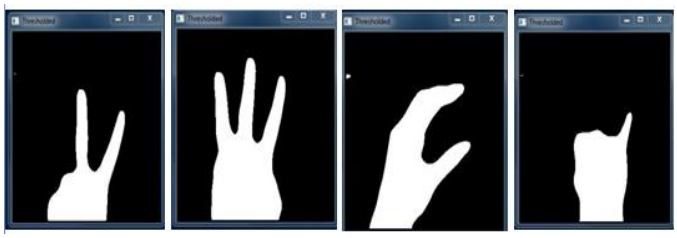

Fig 7 Threshold images of Numbers 2,3 and Alphabets C, D

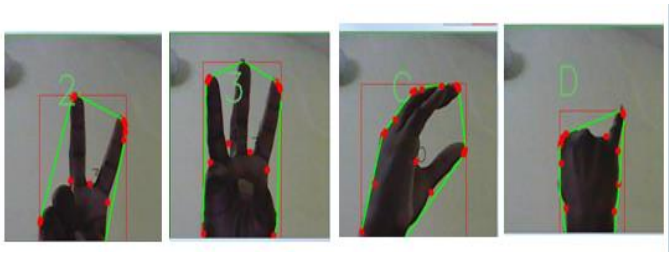

Fig 8 Output images with Text Display T of Numbers 2,3 and Alphabets C, D

A. Applications

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. The areas of application include Sign language for deaf and dumb, Controller less video gaming, Smart TV, Video surveillance, Human robot interaction, Biometrics

VI. CONCLUSION

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 without the assistance

of an interpreter in between. 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.

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

[3] 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.*

[4] 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*

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.