Journal of Communications Technology, Electronics and Computer Science, Issue 8, 2016
ISSN 2457-905X

MonVoix - An Android Application for hearing impaired people

AishwaryaDanoji[1], AishwaryaDhage[1], Rachana Kamat[1], PriyaPuranik[1], Prof Sharmila Sengupta[2]
BE Students, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai-400074, India[1]
Professor, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Mumbai-400074, India[2]
aishwarya.danoji@ves.ac.in, aishwarya.dhage@ves.ac.in, rachana.kamat@ves.ac.in, priya.puranik@ves.ac.in, sharmila.sengupta@ves.ac.in

Abstract — **Human communication is the most valuable foundation in developing a cooperative environment for sharing information and knowledge by interactive sessions. Normal individual vocalizes his views through shared intentions like facial expressions and hand gestures. People with acoustical disabilities are obligated to rely on interpreters for day-to-day conversations. Interpretation of various Sign Languages is important as it will shorten the social drift and act as an agent of communal integration. This paper proposes an Android Application for a coherent interpretation of Sign Languages. MonVoix - a French remark for my voice, would act as a boon for the deaf and the mute by completely eliminating the requisite of a human interpreter. This approach utilizes a user's Smartphone camera to capture a series of hand gestures and convert the image file to the corresponding message and audio using image processing techniques and database emulation for identification of the image.**

Keywords — Sign Language, hand gestures, Android Application, Otsu Binarization, Image processing, FFT, Canny edge detection, histogram equalization, dilation.

I. Introduction

Since the beginning of human communication, sign language has been an essential aspect of communication. Ancient human used to communicate by simple basic gestures to express ideas. Sign language has gradually evolved with time. There are approximately 300 sign languages with new language emerging at a rapid rate.

Various technologies have been deployed to aid the acoustically disabled like the Video Relay Service, W-Talker, TV-Listener Infrared Radio Systems, Door Signalers and Telephone to Telephone Typewriters. The proposed system will help the deaf to lead an interpreter-independent life and bestow a voice to the mute. An Android phone camera is used for capturing the gesture; OpenCV Python is used for processing the sign language to text conversion and SQLite Manager for storing the image database.

Many researchers have contributed to the advent of sign language across the globe. Juan Pablo Bonet presented a dictionary that included the first sign language alphabet. Charles Michel De L'Eppe, celebrated as developed a dictionary for French sign language that included symbolic gestures that conveyed concepts as opposed to just letters. Thomas Hopkins Gallaudet and Laurent Clerc together promoted the learning of American Sign Language. The work of these renowned researches provided the much needed base for understanding the semantics behind the gestures. Modern day developers consequently embedded these dictionary contents to create applications.

II. LITERATURE SURVEY

Gestures are nonverbal method of communication. There are different types of gestures. Symbolic gestures which have a single meaning in each culture example sign language. Deictic gestures direct the listener's attention to specific events or objects in the environment example pointing gesture. Iconic gesture represents meaningful objects or actions. Pantomimic gestures that depict objects or actions, with or without accompanying speech example mimic gestures.

Gesture applications can be broadly classified into multidirectional and a symbolic language. Some of the applications include 3D design where objects can be rotated or translated to obtain their 360 degree view, the real time ROBOGEST system [1] constructed at University of California; San Diego presents a natural way of controlling an outdoor autonomous vehicle by use of a language of hand gestures, Virtual reality [2] and sign language.

Hand gestures are probably the most frequently used type of gestures. Various applications of hand gesture recognition are Human computer interaction, Visualization, computer gaming etc. Sign language is also one of the applications of gesture recognition.

There are various algorithms used for hand gesture recognition such as Hidden Markov Models which uses multi-colored gloves and has an accuracy of 91.7% [3]. Traditionally around 1977, gloves for hand gesture identification with light based sensors and flexible tubes were used. This method comprised of having a light source on one end and photo cell on the other. Proximity sensors were used to determine the bent of fingers based on the amount of light hitting the photocell. Additionally there were inertial sensors to deduce the amount of twist in the hand movement and flex sensors that formed a voltage divider network .This network helped in deducing the bending of fingers on the account of the change in resistance and the stretching of the elastomer. Nowadays the gloves use an adjustable palm strip controller and 15 discrete sensors which include accelerometer for linear acceleration, gyroscope for angular velocity and magnetometer for orientation.

S.Adhan and C.Pintavirooj proposed geometric invariance based algorithm for gesture recognition [4].This image acquisition system takes an input image using a web camera and converts the RGB image into gray scale. The pre-processing techniques involve Binarization using Otsu thresholding and Canny edge detection for boundary extraction.The boundary is sharpened by using B spline curvature contouring. Features like the internal angle (α_1, α_2, α_3 are internal angles of its triangle), normalization area (A_{i1}) and an adjacent area ratio (A_{i2}) are extracted from the triangle i.e

$$F_i=\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, A_{i1}, A_{i2}. \quad (1)$$

These points are compared with the reference template and the minimum error is calculated to match the equivalent hand gesture.

Thinning algorithm is proposed by S. N. Omkar and M. Monisha[5]. This system initially separated the RGB and foreground components to calculate the Eucledian distance. The pre-thinning procedure involved threshold based on the Eucledian distance, erosion and dilation of the image. The thinning process translated the structuring element to every pixel position possible within the image and if a match is found with the underlying image the pixel value is changed to zero. The termination of extraction points is calculated by checking the value of surrounding eight pixels. Mathematically, thinning is represented as

$$\text{thin}(A, B)= A- \text{hit and miss}(A-B) \quad (2)$$

Other methods include vision based approach, feature extraction approach], multiclass classifier integrated with several binary classifier such as Support Vector Machine Algorithm (SVM) having accuracy of 93% [6]. Wavelet Packet Decomposition & principle curvature based region detection having accuracy of 93.1% [7]. Machine Learning Model for Sign Language Interpretation using Webcam Images [8] involves background subtraction, Blob analysis, noise reduction, brightness normalization, contrast adjustment and scaling for pre-processing the image. The database includes 500 positive cases, 500 negative cases and 50 test images from different locations with different people. The Haar Cascade classifier compares the pre-processed image to the test images and produces an output with an accuracy of 92.68%.

Dynamic & Static gesture recognition [9] proposes detecting gesture by video object plane (VOP) for abstraction. Redundant frames are eliminated and the displacement of VOP centroids is taken into consideration. For global displacement, Hausdroff distance based shape similarity is used. Local displacement understands the shape and pose through FSM matching. Image Segmentation using Eigen value function is followed by Fourier transformations. These transformations are used to extract the feature vectors which are then classified by K- Nearest Neighbor (KNN) algorithm to determine Arabic Sign Language (ArSL).

III. PROPOSED SYSTEM

Sign Language for beginners [10] is an American and British Sign Language Learning application that provides a prolonged sectional list of sign language with clear images describing the respective action. These sections are segregated according to alphabets, numbers, colors and commonly used words along with the facts and history of the gesture. ASL-American Sign Language [11] has an image library presenting alphabets, numbers, greetings and videos demonstrating the gestures. To make the learning session more interactive the developers have incorporated quizzes and picture matching exercises.

Sign Language ASL kids [12] provides a platform for mute kids between the age group of 1-12 years to learn the hand signs for understanding finger spellings. The application has a sound button as an add-on to assist in better interpretation. An Android application offered by Magoosh as available on Play Store as ASL Flashcards [13] uses smart algorithms to track the user's development in learning the sign language and provides image flashcards with the corresponding word. Brazilian developers-Technologias Assistivas have created an application named ProDeaf translator [14] that converts the phrases in Portuguese to Brazilian Sign Language through animation. Also, if the input is an English phrase, then it translates the same into American Sign Language.

On exploring the extensive list of Android applications available on Google Play Store, it can be concluded that majorly they provide the sign language images as output for the input text. The unique feature of MonVoix that makes it stand out of this pool of applications is that it converts the dynamically captured real-time gesture to text. Also, it converts the converted text to speech for prompt outcome and coherent understanding. In addition, all the applications work primarily in interpreting American, British, French and Brazilian Sign Language. MonVoix proposes to efficiently work on interpreting various segments of Indian Sign Language such as alphabets, numbers, greetings, colors, emotions, days and other everyday words.

The application MonVoix focuses on minimizing the number of errors whiles capturing the hand gestures so as to improve robustness. It is a real time vision based system that provides flexibility to user has higher portability and eliminates the dependency on sensors by using only mobile camera thus minimizing the number of requirements.

A. Features:

1. Convert gestures to text
2. Convert text to audio

3. Display gesture to be used for a given text.
4. Voice Recognition: This allows a hearing person to respond with voice to the person signing. It then converts the audio to text.
5. Database having List of various sign languages used in the system.

B. System Architecture

MonVoix uses Android Studio for the User Interface, OpenCV Python for image processing and SQLite Manager for storing training images and their identification parameters. The component diagram is presented in Fig. 1.

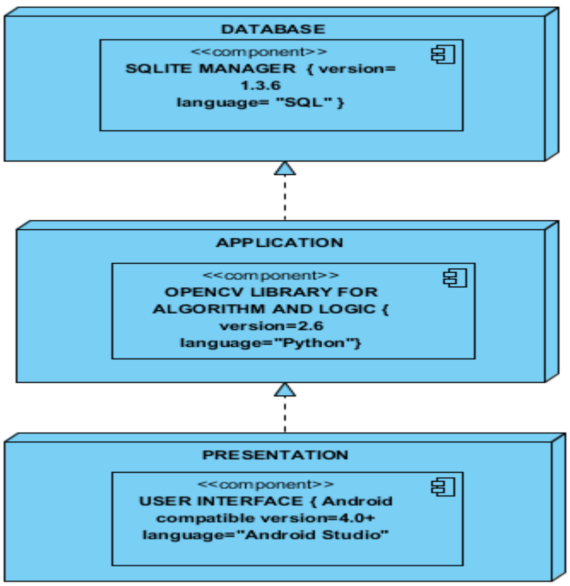


Fig.1. Component diagram of MonVoix Application.

IV. IMPLEMENTATION

The application comprises of four modules- Number recognition, alphabet recognition, text to audio and audio to text conversion.

A. Number Recognition:

We first capture the gesture for numbers using the smartphone camera. The captured image is then processed as follows:

1) Convert the captured RGB image to Grayscale to get the Region of Interest (ROI) using cv2.COLOR_BGR2GRAY in OpenCV.

2) Perform Gaussian blurring using cv2.GaussianBlur () for smooth transition between different colors and to reduce the background noise.

3) Threshold the blurred image with cv2.threshold () for image segmentation which helps to highlight ROI by using OTSU Binarization technique that calculates the threshold value by binarizing the image.

4) Draw contours using cv2.drawContours () and find the convexity defects to determine the angle of deviation from the convex hull. By determining the deepest points of deviation on the contour we can deduce the number of extended fingers.

5) If the calculated angle is less than or equal to 90 degree then increment the defect count and continue to scan the contour to determine other angles.

6) Based on the number of defect counts, display the corresponding recognized number.

B. Alphabet recognition

Fig. 2 presents the flowchart for alphabet recognition.

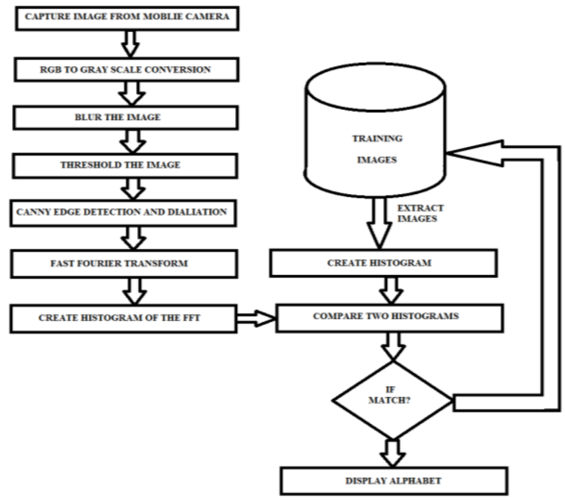


Fig. 2. Flowchart for alphabet recognition.

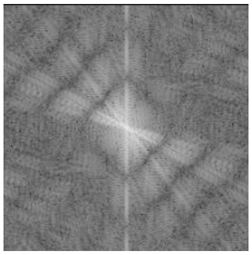
The steps involved in alphabet recognition involve processing of two sets of images training and testing datasets: Image processing on training images is done as follows:

1. Capture the image using a smartphone camera. Process it by altering the RGB image to Grayscale to get the Region of Interest (ROI).
2. Execute Gaussian blurring for smooth transition between different color .Threshold the blurred image for image segmentation which helps to highlight ROI by using OTSU Binarization technique that calculates the threshold value by binarizing the image.
3. An edge detection algorithm is applied on the thresholded image to highlight the edges and remove noise from the image. Dilation is then performed to smooth the broken edges as sown in Fig. 3 and 4.

Fig. 3. Alphabet C.

Fig. 4. Alphabet W.

4. Perform Fast Fourier Transform (FFT) of the dilated image to convert the image in spatial domain to frequency domain. By taking FFT, all the frequencies present in the image according to their positions can be seen. Fig. 5 and 6 present the FFT analize of Fig. 3 and 4.



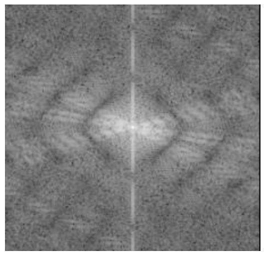


Fig. 5. FFT of Fig 3 Fig. 6. FFT of Fig 4

5. The histogram is then taken of the transformed image. Histogram processing gives the numbers of pixels present for the frequencies present in the image. The image processing operations performed are presented in Fig. 7.

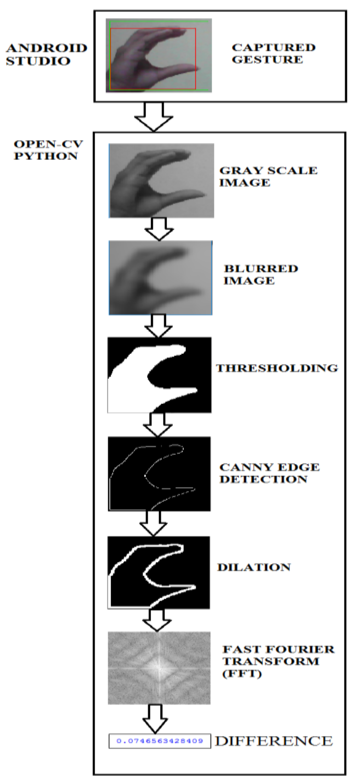


Fig 7. Real-time capturing of gesture and stepwise image processing.

COMPARING THE TRAINING AND TESTING IMAGES

FFT output images are stored as training images in the database. These images are retrieved from the database. Histogram processing is performed on the retrieved images. The histogram of the captured image is compared with the histograms of the training images in the database. The histogram with the least difference is the histogram of the desired output. The alphabet in the alphabet column in the database corresponding to the least difference histogram of the FFT is the output. Fig. 8 presents the result after comparing the difference of testing image C with training database images.

rowid	title	image	alphabet
1	A	BLOB (Size: 63392)	A
2	B	BLOB (Size: 63160)	B
3	C	BLOB (Size: 62140)	C
4	W	BLOB (Size: 61990)	W

Fig. 8. Snapshot of a section of table in database.

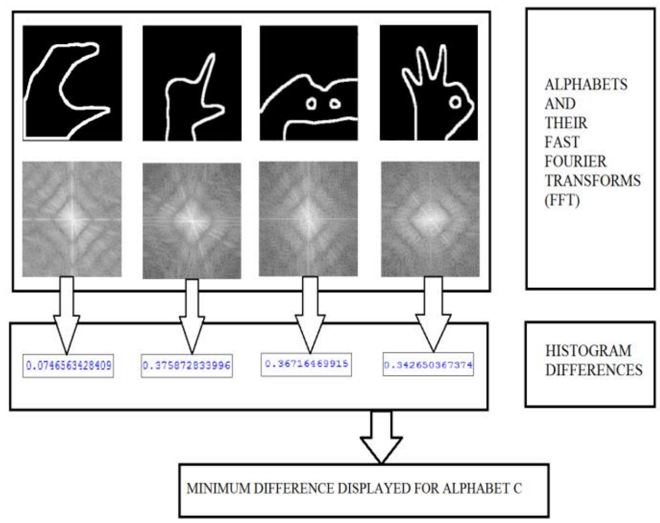


Fig. 9. Result after comparing the difference of testing image C with training database images.

C. *Text to Audio*

Android allows conversion of text to speech and this is done using the class TextToSpeech. To use this class one needs to initialize an object of this class and also specify initListener. To set a specific language we need to call setLanguage() method. Fig. 10 ad 11 present the graphical user interface.

Fig. 10. Start Screen. Fig. 11. Text to hand gesture.

D. Audio to text

Android comes with an inbuilt feature speech to text through which we can provide speech input to your app. the speech input will be streamed to a server, on the server voice will be converted to text and the final text will be sent back to our app.

The heart of Speech to text Android API is class android. speech. RecognizerIntent. Basically, we trigger intent (android. speech. RecognizerIntent) which shows dialog box to recognize speech input. Speech to text conversion outcome is returned back to the Activity. The below figures present the user interface.

Fig. 12. Captured hand gesture.

Fig. 13. Gesture to text output.

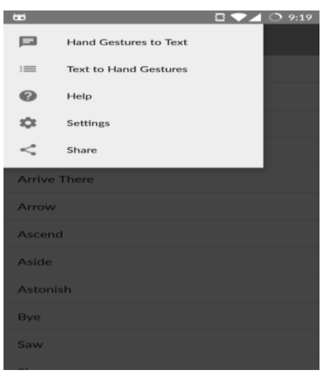


Fig. 14. Menu bar options

Fig. 15 Help Menu.

V. Conclusion And Future Work

We have observed that a hearing impaired person has to rely completely on a human interpreter for an active communication and create a drift from leading a normal life in the society. Thus, the proposed android application will not only make them independent and helps in reducing the social barriers but also boost their moral. Personalizing the gestures according to the user for swift performance will further improve the efficiency of the application. This pocket application can find their application not only for SLR, but also be exploited in other domains like mobile gaming and virtual reality where in-built movements can augment the visual outcomes.

VI. References

[1] ArunKatkere, Edward Hunter, Don Kuramura, Jennifer Schlenzig, Saied Moezzi, and Ramesh Jain,"Robogest: Telepresence using hand gestures",Technical report, University of California, San Diego, Visual Computing Laboratory, Technical Report No. VCL-94-104, December 1994.

[2] Hank Grant, Chuen-Ki Lai, "simulation modeling with artificial reality technology (smart): an integration of virtual reality simulation modeling" , Proceedings of the Winter Simulation Conference, 1998.

[3] S. Liwicki and M. Everingham. Automatic recognition of finger spelled words in british sign language. In IEEE Workshop on CVPR for Human Communicative Behavior Analysis, 2009.

[4] SuchinAdhan and ChuchartPintavirooj "Alphabetic Hand Sign Interpretation using Geometric Invariance" The Biomedical Engineering International Conference (BMEiCON-2014)

[5] S. N. Omkar and M. Monisha. "Sign Language Recognition using Thinning algorithm".ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING, AUGUST 2011, VOLUME: 02, ISSUE: 01.

[6] J. Jones. (1991, May 10). Networks (2nd ed.) [Online].Available:http://www.atm.com R. Kurdyumov, P. Ho, J. Ng. (2011, December16) Sign Language Classification Using WebcamImagesOnline]. Available:http://cs229.stanford.edu/.../KurdyumovHoNgSignLanguageClassificationUsi...

[7] J.Rekha, J. Bhattacharya, S. Majumder, "Shape, texture and local movement hand gesture features for Indian Sign Language recognition ", 3rd International Conference on Trendz in Information Sciences and Computing (TISe), pp. 30 - 35, 2011.

[8] KanchanDabre, SurekhaDholay. "Machine Learning Model for Sign Language Interpretation using Webcam Images". 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)

[9] M.K.Bhuyan and P.K.Bora, "A Frame Work of Hand Gesture Recognition with Applications to Sign Language ", Annual India Conference, IEEE, pp.I-6.

[10] Sign language for beginners (2016, May 23). Retrieved from https://play.google.com/store/apps/details?id=com.sign.language.learn1234

[11] ASL-American Sign Language (2016,September 21). Retrieved from https://play.google.com/store/apps/details?id=tenmb.asl.americansignlanguagepro

[12] Sign Language ASL kids (2016, February 18). Retrieved from https://play.google.com/store/apps/details?id=com.basvanderwilk.aslkids

[13] ASL Flashcards (2016, July 9). Retrieved from https://play.google.com/store/apps/details?id=com.magoosh.gre.flashcards.asl

[14] ProDeaf Translator (2016, August 9). Retrieved from https://play.google.com/store/apps/details?id=com.Proativa.ProDeafMove