iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** V **Month of publication:** May 2019

DOI: https://doi.org/10.22214/ijraset.2019.5414

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Sign Voice Bi-Directional Communication System (SVBi) for Blind, Deaf/Dumb and Normal People

Sharanya R[1], Sneha Bhatt U[2], Suma H R[3], Tejashwini G[4], Jyothi R[5]
[1, 2, 3, 4, 5]*Department of Computer Science and Engineering, Global Academy of Technology,Bangalore-560098*

Abstract: Blind and deaf/dumb people face problems in communicating with others facing the difficulties in dealing with the communication technology. The goal of this paper is to design a communication system that is used to facilitate communication between blind, deaf/dumb and normal people using machine learning techniques. SVBiCommunication system helps blind person to hear voice through speaker, saying the word gestured by the deaf/dumb while the deaf will receive text that can be seen on LCD, representing the word said by the blind. SVBiCommunication works in two main directions, the first direction is processing the input from image to speech and also to text. The animated word gestures that are taken from the image are mapped with predefined language knowledge base into text. Then, the relevant audio is generated. The second direction is processing from speech to text. The voice from blind is converted into its corresponding text using ARM-Voice Recognition application.

Keywords: Deaf-blind people; Deaf-blind communication devices; Machine learning techniques; text-to-speech conversion; speech-to-text conversion

I. INTRODUCTION

Deaf/dumb and blind people usually face problems on normal communication with other people in society. It has been observed that, they sometimes find it difficult to interact with normal people with their gestures. Because people with hearing problems cannot speak like normal people, they have to depend on some kind of visual communication in most cases. To overcome these problems, we have proposed a system which enables the communication from deaf/dumb to blind and also vice-versa.

According to the World Health Organization (WHO) [1] , there are around 466 million people worldwide have disabling hearing loss, and more than 28 million of these are Americans, 13 million people within Egypt across all age groups. The estimated number of people visually impaired in the world is 285 million, 39 million blind and 246 million having low vision. Egypt has approximately 1 million blind people and 3 million visually impaired. It is necessary to find basic means of communication among hard - of-hearing or deaf people, blind and normal people. It is important to give the deaf and blind people an individualized and appropriate communication. system that supports different communication techniques, strategies and modes.

These systems should reflect their assessed needs and respects their choice with facilitating their life by integrating them into the society. Those technology- based solutions facilitate face-to-face longer-distance communication needs. Systems act as an interpreter that performs the bidirectional translation of sign language and spoken language between vocal and hearing-impaired people. Few systems were developed to perform the bidirectional translation. The Arabic sign dictionary system [1] is a vision based software- system working from text to signals direction. This system generates static signals as typing font for static signs (letters, numbers), can be integrated with different software as Microsoft Word. The Glove-Based Systems for example is the portable glove system [2] is a hardware system working from signs to text direction. In this system speaker wear electronic hardware gloves to measure hands and face motion level up to six degrees of freedom.

Those systems are very efficient on its measurements and can exactly describe each moving object. But these kinds of systems suffer from high cost and the speaker has to be in a static place. Moreover, they suffer from hard and long time to adapt with the hardware used from the user's perspective.

Tessa [3] system translates English UK accent sentences clerks into British Sign Language (BSL) signs, by using a 3D virtual human. Transactions time is longer that is worse than transactions handled without the system. This system limits the sentences that the system accepts to a set of pre-defined ones that caused these negative results. The system described in [4] supports medical conversation between a hearing physician and a deaf patient by showing the written transcription of physician's spoken sentences together with medical images (e.g., diet plans) on a tabletop display. The system uses forms of communication without signs do not integrate sign language support. iCommunicator [5] makes effective two-way communication possible for persons who are deaf, hard-of-hearing or experience unique communication challenges. The iCommunicator translates in real-time:

Speech to Text then to Sign-Language and Video to text and then to Computer Generated Voice. Limitations of this system are the difficulty to adapt with the hardware as contains many hardware pieces with different purposes.

SVBiComm system aims to develop a vision-based system for translating sign language (sequence of images) into text and then to speech as one direction and translating voice spoken words to text. SVBiComm works in two directions: - from video to speech: - deaf to normal\blind, and from speech to video: - normal\blind to Deaf as shown in Figure 1.1.

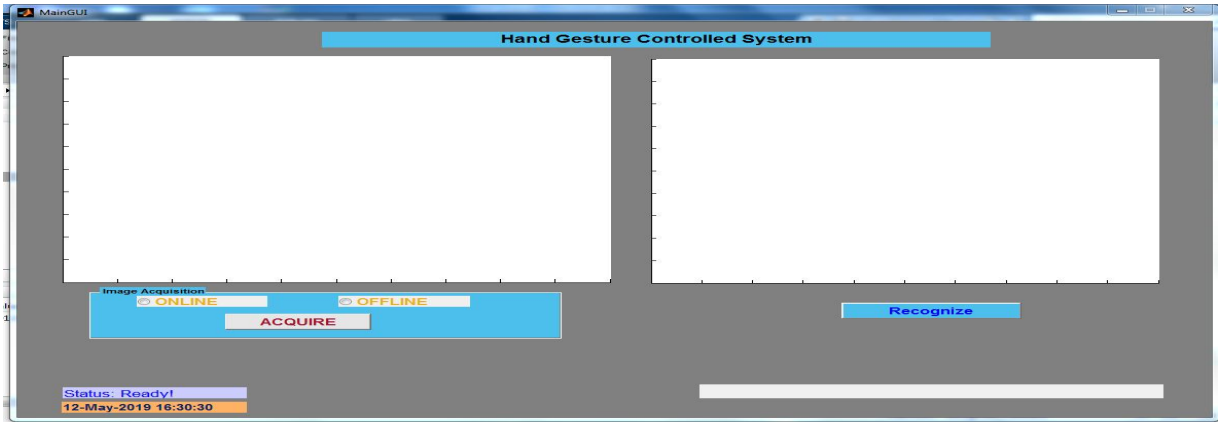


Figure 1.1: GUI design for communication from deaf/dumb to blind using MATLAB.

SVBiComm system with its video and voice capturing devices uses the object oriented programming language to implement image processing algorithm. The image could be captured from any distance with non/black background without any skin colored objects. Speech should be in a quite space with minimum noise. The proposed vision-based hand tracking system does not require any special markers or gloves that can operate with low-cost cameras. SVBiComm system provides user by: static sign translation; isolated words animation and translation; continuous sentence translation and playing by voice; and finally a "deaf/dumb"-keyboard. SVBiComm system analyses video and voice streams on real time with high speed network connection and high performance computing capabilities.

II. METHODOLOGY

The framework of SVBiComm System is shown in Figure 2.1. The system is mainly consists of two modules, first module is the translation of the sign language to text and then to speech as one direction. The second module is the translation of the voice to text represented by a 3D model.

1) *Generation Of The Database:* Here our system takes the hand movements through the web camera. In this proposed method, 26 combinations of Indian characters are developed by the use of right Hand saved in training database. This is implemented using a well-trained Neural Network (training & recognition) / Machine Learning Component: - A typical Back Propagation ANN learning algorithm is used to train the network i.e. learn the training data. The black nodes (on the extreme left) are the initial inputs. Training such a network involves two phases. In the first phase, the inputs are propagated forward to compute the outputs for each output node. Then, the error for each output node is calculated by subtraction of each output from its desired output. Then each output errors is passed in reverse and the weights are fixed. The setup phases is continued until the sum of square of output errors reaches an acceptable value. Network Training; a set of trained samples in all patterns is responsible for classifying the training data into a set of classes and mapping each pattern to its owned unique class.
2) *Image Pre-Processing And Segmentation:* The pre-processing takes place on these recorded input gestures. Then the segmentation Hands are performed to separate object and background.
3) *Feature Extraction:* The segmented hand image is represented with certain features. The characteristics are used for gesture recognition with the template matching algorithm that gives Optimized results.
4) *Sign Recognition:* The given character gesture is recognized with the skin colour recognition and the template Matching from the record.
5) *Sign To Text And Speech Conversion:* The recognized sign is then mapped into text and further converted into speech with various libraries.

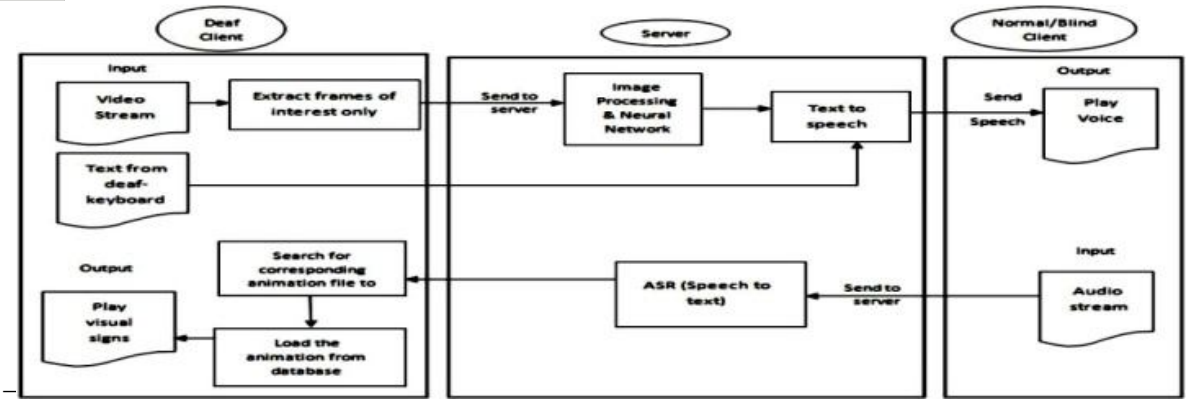


Figure 2.1: SVBiCommunication System Framework

The first direction: The input could be a text which is directly sends to the server using a special purpose deaf- keyboard. Otherwise the video stream is captured by the "deaf/dumb" user's camera. Frames of interest are extracted and passed down to server to be processed. The server then applies some image processing techniques on the passed frames. In order to get the black and white image, the skin detection mechanism is applied on the image. The white parts represent the skin detected parts. The next step is passing down the image to a median filter to remove any possible noise on the image. The image is then portioned to pieces.

Afterwards, the pieces are passed to a scale down function to make them all of one size. Then merge the pieces again into one image in one specific order starting from top left. The last step is to feed the image into a well-trained neural network to recognize the image and return the corresponding text. Finally, the produced text is converted into speech using TTS tool, then this speech is played on the normal/blind client's machine. The second direction: - The system records audio stream from the normal/blind client's microphone and send it to the server to be recognize. The speech is filtered before recognition then feature extraction function is applied. The speech is recognized by using

Dynamic Time Wrapping (DTW), to return the corresponding text. The text is then passed down to the 3D graphical model to animate the text as visual signals.

III. FIRST MODULE - IMAGE TO SPEECH CONVERSION

The first module is the translation of the sign language to text and then to speech as one direction as shown in Figure 3.1. The animated word gestures are mapped with language knowledge base into voice using frame formation. The captured input video stream is fed to the frame extractor subcomponent to extract the frames of interest. Then, the relevant audio is generated using Text-to-Speech (TTS) API.

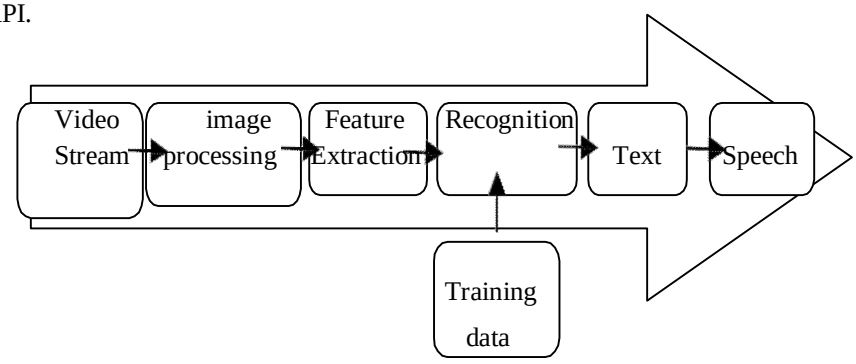


Figure 3.1: SVBiComm System Framework Direction from "deaf/dumb" to blind Page Layout

This module will implement many steps in order to convert the sign language into text then to speech as follow.

1) Image Pre-processing

a) Noise Removal

b) Skin Detection

c) Skin Classifier and Binary Image Processing:

2) Feature Extraction
3) Image Recognition
4) Text to Speech Conversion

A. Input

The input could be a video stream which is captured by the deaf user's camera. Frames of interest are extracted and passed to server to be processed. Or the input could be directly sent to the server using a special purpose deaf-keyboard.

B. Image Preprocessing

This phase composed of three steps noise removal, skin detection and binary image processing.

1) Noise Removal using Median Filter: Median filtering is used to maintain edges and remove noise. Each pixel is convoluted with its nearby neighbors to decide whether it represents its surroundings or not [2]. The pixel's median is calculated by replacing each entry with the median of the surrounding neighborhood pixel entries. If the neighborhood pixels are an even number then the algorithm calculates the average of the double pixels values in the middle.

2) Skin Detection: Skin detection aims to find regions that have human faces and hands in images and to reject as much "non-skin" parts. This is done by transforming a specific pixel into the corresponding colored space [4]. Then a skin classifier is used to tag each pixel with a skin label or a non-skin label. The human skin color fall in and clustered at a small area in the color space [2]. There are many color spaces that are used in skin detection for digital images. RGB (Red, Green, and Blue) color space is the most commonly used due to its simplicity but it has some limitations. Another space is the YCbCr (Luminance, Chrominance) orthogonal color space which is that most appropriate for encoded video media. The image is then segmented based on skin color. Then image is transformed from RGB into any of YCbCr space by straight forward linear transformation resulting in an invariant to human race skin detectors [3]. All these color spaces separate the illumination channel (Y) from two orthogonal chrominance channels (CbCr). The YC_BC_R space segments the image into a luminosity component and color components then remove the influence of luminosity. The Y component gives all information about the brightness C_B (blue) and C_R (red). Those components are independent of the luminosity. C_B and C_R values are used as shown in Figure 3.2. The faces were differentiated using the "ground truth data".

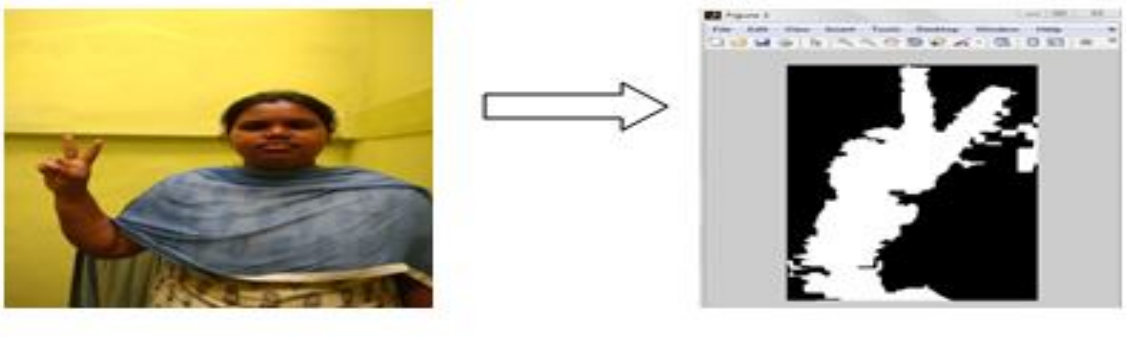


Figure 3.2: The face and hands Skin Detection algorithm

3) Skin Classifier and Binary Image Processing: Is used to indicate the boundary of the skin color class in a feature space by identifying whether a pixel is part of the skin or not. All pixels in the image plane are classified into object and background pixels. Thresholding the values of C_B and C_R allow separation of figure/ground of an image that consists of object with quite high intensity, and a background of relatively low intensity. A binary image function can then be built by making all pixels above the threshold are `1' and below the threshold are `0'. A black and white mask is applied on all the face in addition to some body parts background to reduce the effect of the background and to remove holes within faces. The mask is derived from binary morphological operations [4]. Separate face image is corroded with a 9x9 pixel face-shaped kernel to eliminate the small background objects. Then this damaged image is stretched with a 10x10 pixel face-shaped kernel to refill gaps within the faces. To supplement the dilation, a hole/filling algorithm is used to fill any remaining holes inside faces. Extraction of objects of interest (hands and face) depends on the size and position (location), to detect left and right hands as shown in Figure. 3.3

Result after skin classification detection and skin classification are shown below.

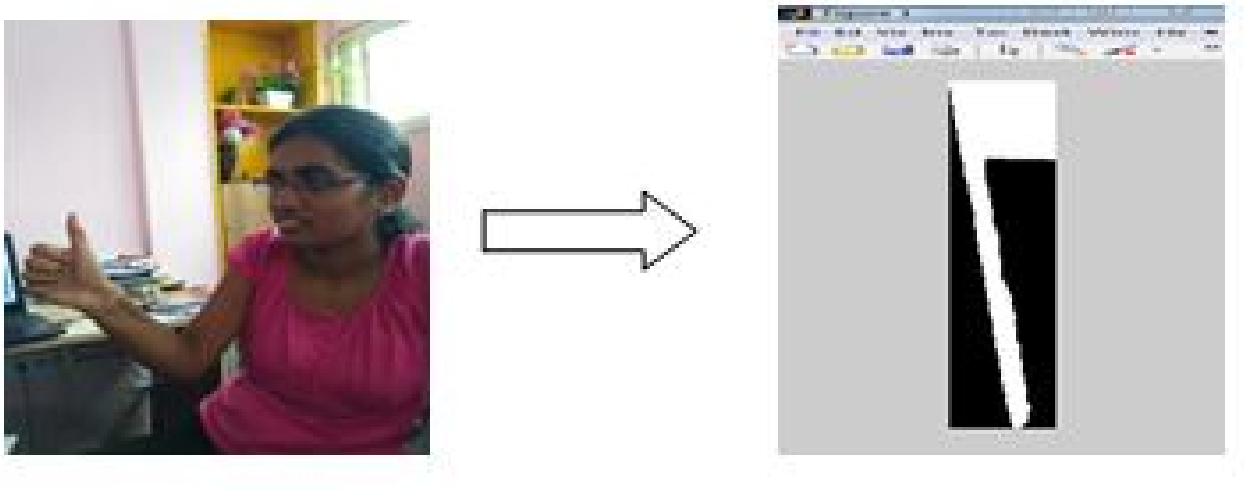


Figure 3.3: Result after skin detection

C. Image Recognition

The image is needed to be prepared in order to avoid miss matching and different motion during the recognition in the neural network. So, the objects is recombined together according to the location i.e. depending on the position of each object using Left Top technique where the object with the smallest left pixel location is the one with the highest priority and if 2 objects have the same left pixel the top pixel determines which one to precede. The implemented system generates a method to map each blobs combination to its corresponding class. Three different classes are used to classify the recombined blobs oneObject, twoObject and threeObject in order to help the Neural Network training and learning and hence generalization

Figure 3.4: Result after Skin Classifier

D. Data Training

This is implemented using a well-trained Neural Network (training & recognition) / Machine Learning Component: - A typical Back Propagation ANN learning algorithm is used to train the network i.e. learn the training data. The black nodes (on the extreme left) are the initial inputs. Training such a network involves two phases. In the first phase, the inputs are propagated forward to compute the outputs for each output node [2]. Then, the error for each output node is calculated by subtraction of each output from its desired output. Then each output errors is passed in reverse and the weights are fixed. These two phases is continued until the sum of square of output errors reaches an acceptable value. Network Training; a set of trained samples in all patterns is responsible for classifying the training data into a set of classes and mapping each pattern to its owned unique class.

E. Text to Speech Conversion

Speech synthesis techniques are used to convert input text to computer generated voice. It is used by the blind to listen to written material. Text-to-speech is very helpful in this system as it is used to convert the text generated from processing the deaf images into voice hearable by the blind/normal user.

IV. SECOND MODULE-SPEECH TO VIDEO CONVERSION

The second module is the second direction from blind to "deaf/dumb" (voice to images) as shown in figure 4.1. The voice from blind is converted into its corresponding text using ARM-Voice Recognition application.

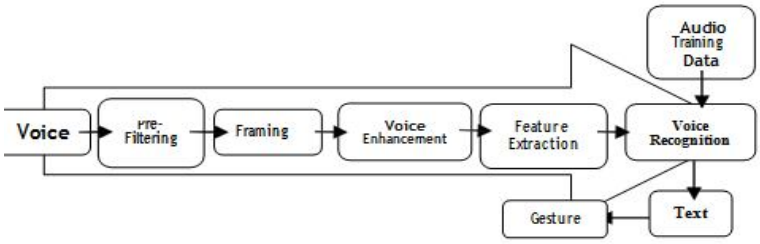


Figure 4.1: Second SVBiCommunication System Direction from Blind to deaf/dumb

V. SYSTEM IMPLEMENTATION

SVBiComm system builds a graphical user interfaces with a web-based and desktop applications. SVBiComm System is implemented using different tools such as Microsoft Visual Studio 2005 , MATLAB , Aforge.net libraries, XNA 0.2(Skeletal Animation Programming) and Star UML in order to maximize the productivity and quality with its automatically generating numerous results. Initially the gestures are taken from deaf/dumb through MATLAB, then the unwanted noise is removed. Once the noise is removed, then the feature extraction is done for the given image, then final recognition of the image is done. Once the image is recognized, then the gestures are converted into text. Then that text is converted into speech and can be listened through the speaker, and the text can be seen on LCD to the deaf.

In another way communication, there is a ARM-VOICE RECOGNITION application, where we can record the voices and speech could be recognized by deaf/dumb through LCD.

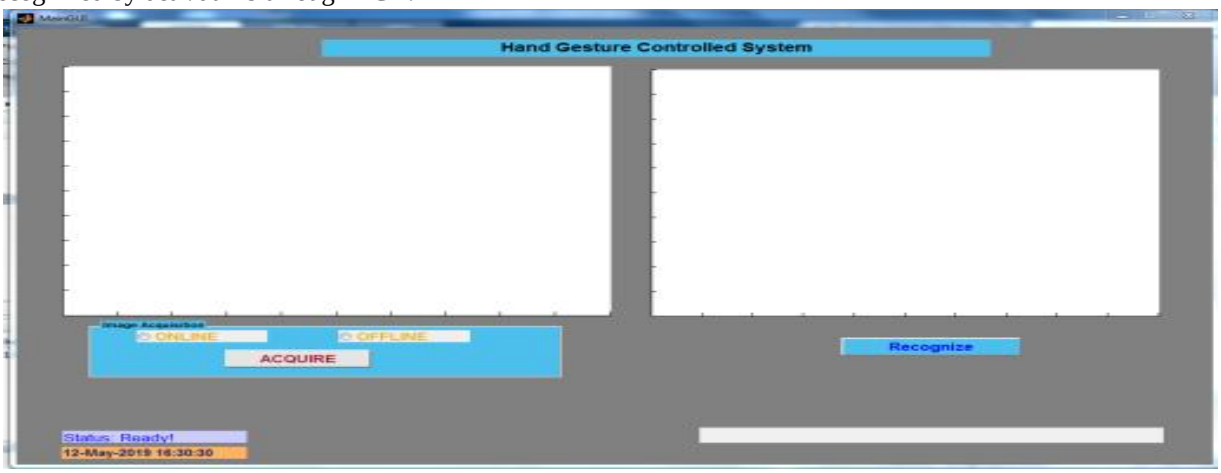


Figure 5.1: SVBi Communication Graphical User Interface

VI. HARDWARE IMPLEMENTATION

Hardware module is sub-divided into following modules:

A. PIC16F877A
B. Liquid Crystal Display(LCD)
C. GP Transmitter and GP Receiver
D. Step-Down Transformer
E. APR Voice Play
F. Speaker.
G. Serial Transmitter.
H. Bluetooth module.

It is divided into 3 modules:

1) *Module Name:* Power Supply Design

a) Functionality: A power supply unit is required to provide the appropriate voltage supply. The 230V AC supply is converted into 5V AC supply through the transformer. The output of the transformer has the same frequency as in the input AC power. This AC power is converted into DC power through diodes.

b) Input: 230 volts of AC from Transformer.

c) Output: 12 volts of AC which is transmitted to GP Receiver.

2) *Module Name:* Embedded Microcontroller working and LCD part

a) *Functionality:* Transferring the character type values obtained from MATLAB to

b) *Input:* Through UART protocol GP Receiver transmits the values obtained from MATLAB to GP Receiver.

c) *Output:* Character type values, which is sent to the APR Voice module to record the voice and also sent to LCD to display the values.

3) Module Name: Output at LCD and Speaker

a) *Functionality:* The values are displayed on the LCD for Deaf/dumb and voice recorded using APR Voice Play will be sent to Speaker.

b) *Input:* Character type values obtained from MATLAB.

c) *Output:* Values on LCD can be seen by Deaf/Dumb and Voice can be heard by the blind through speaker.

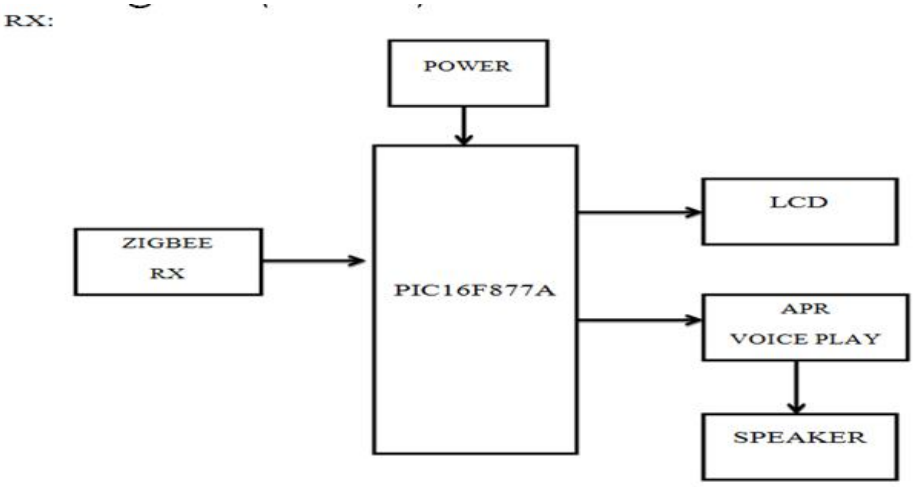


Figure 6.1: Block Diagram of Receiver with PIC micro-controller.

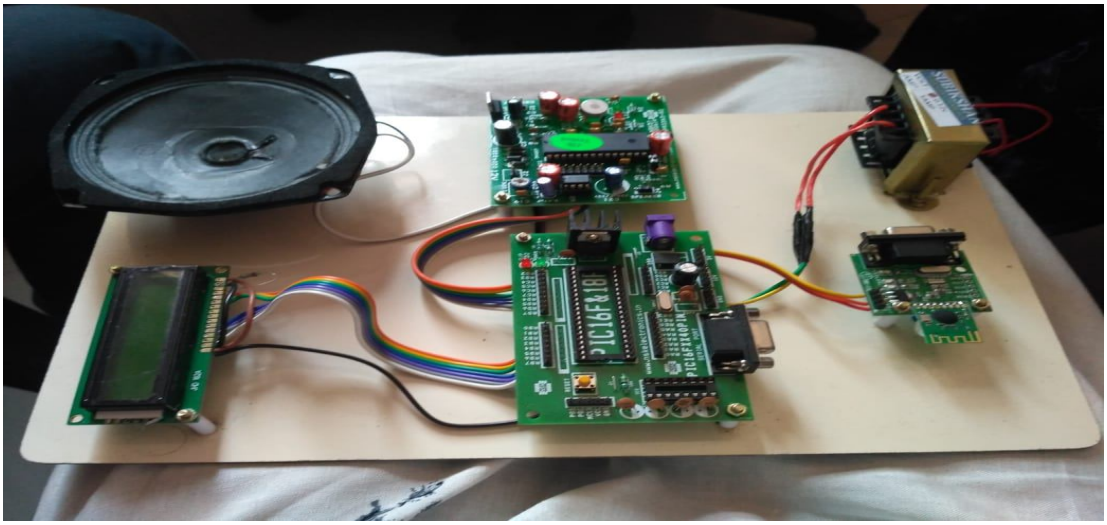


Figure 6.2: Overall hardware setup with PIC 16F877A micro-controller

VII.RESULTS

The 'deaf/dumb' user send images to the server for processing. The server converts them into their equivalent voice to be sent to the user. The received text will be animated by the 3D graphical model. The user receives an automatically played voice from the server. The received text also displayed as a text on the LCD. The below snapshots explains how graphical user interface designed for communication from deaf/dumb to deaf and also to Blind. For this communication, MATLAB is a tool used to have the communication, where the gestures are sent by the user. Once the images are received, frames of interest are extracted, and the noise are removed if any, then the feature extraction is done from the image. At the end, the images are recognized. Then the gestures are then sent to the serial transmitter for further processing. Step-down transformer is used to convert the 230volts AC to 12volts AC. Then the diode which is a DC-converter, converts the 12AC to12 DC. Then 7805 regulator converts the 12volts DC to 5volts DC. Finally, the gestures are converted into the speech whichcan be listened through the speaker and also can be seen on the LCD. This is the communication from deaf/dumb to blind.

Another way of communication is blind to deaf/dumb. For this, ARM-voice recognition application is used to record the voice, so that communication happens between the blind to deaf/dumb. Whatever the voice is recorded, it gets displayed on the LCD.

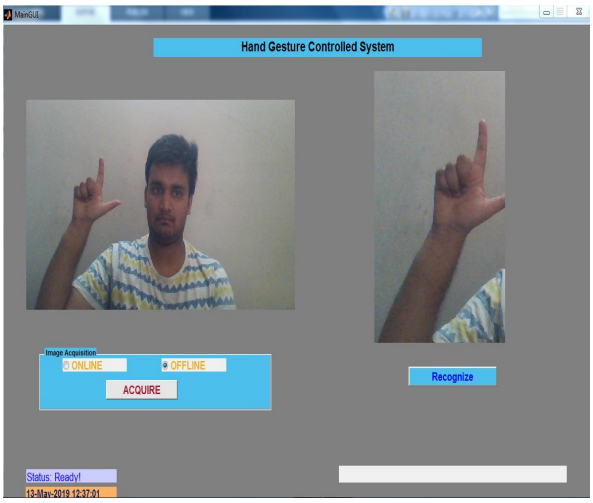


Figure 7.1: Feature extraction

The unwanted noise are cropped from the image using dilation process. The required gestures are segmented using k-means and fuzzy-c means.

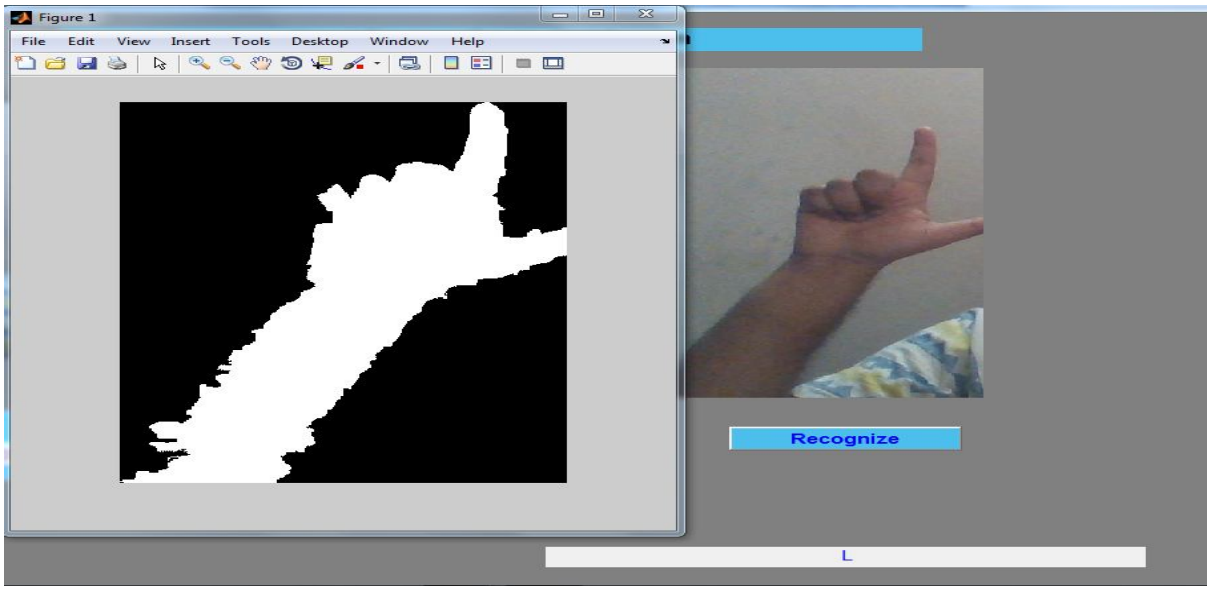


Figure 7.2: Recognizing the images using skin detection technique.

The segmented gesture is recognized using skin detection technique.

A. Software Requirements

B. Embedded C

1) Embedded C is one of the most popular and most commonly used Programming Languages in the development of Embedded Systems.

2) Embedded C is perhaps the most popular languages among Embedded Programmers for programming Embedded Systems.

3) There are many popular programming languages like Assembly, BASIC, C++ etc. that are often used for developing Embedded Systems but Embedded C remains popular due to its efficiency, less development time and portability.

VIII. CONCLUSION

Blind/Normal to Deaf/Hard-of-Hearing Chat system (SVBiComm) is implemented to translate sign language gestures into the corresponding computer generated/human speech gestures using machine learning techniques. The system is divided into two basic parts. The first part is the static sign recognition that is based on posture recognition such as alphabet recognition, finger spelling recognition or words that require no motion of hands or face. The second part is the voice recognition through an application ARM-VOICE recognition is either isolated or continuous, isolated where word by word is supplied to the server to recognize, continuous on the other hand where a complete sentence is supplied to the server to recognize. The proposed system puts no restrictions on the environment illumination or background simplicity. The speech is needed to be in quite space. SVBiComm with its vision-based hand tracking system does not require special markers or gloves in tracking the hands and faces. SVBiComm can operate on a commodity PC with low- cost cameras. The proposed system faces some difficulties and restrictions; the system needs the training data for both images and voice to be noiseless to train the recognition systems well. The voice capturing phase needs as much as possible the surrounding space to be silent to have the most minimum noise Also, there are many sign language limitations such as : Persons make up to use famous shapes or marks to represent persons, places, streets, names and so on, and the same sign used in different places with different meanings so it is needed to develop a generic tool to try to accept any group specific signs.

REFERENCES

[1] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott. Tessa, a system to aid communication with deaf people. In Assets '02: Proc. 5th int'l ACM conf. Assistive technologies, pages 205–212, New York, NY, USA, 2002. ACM Press

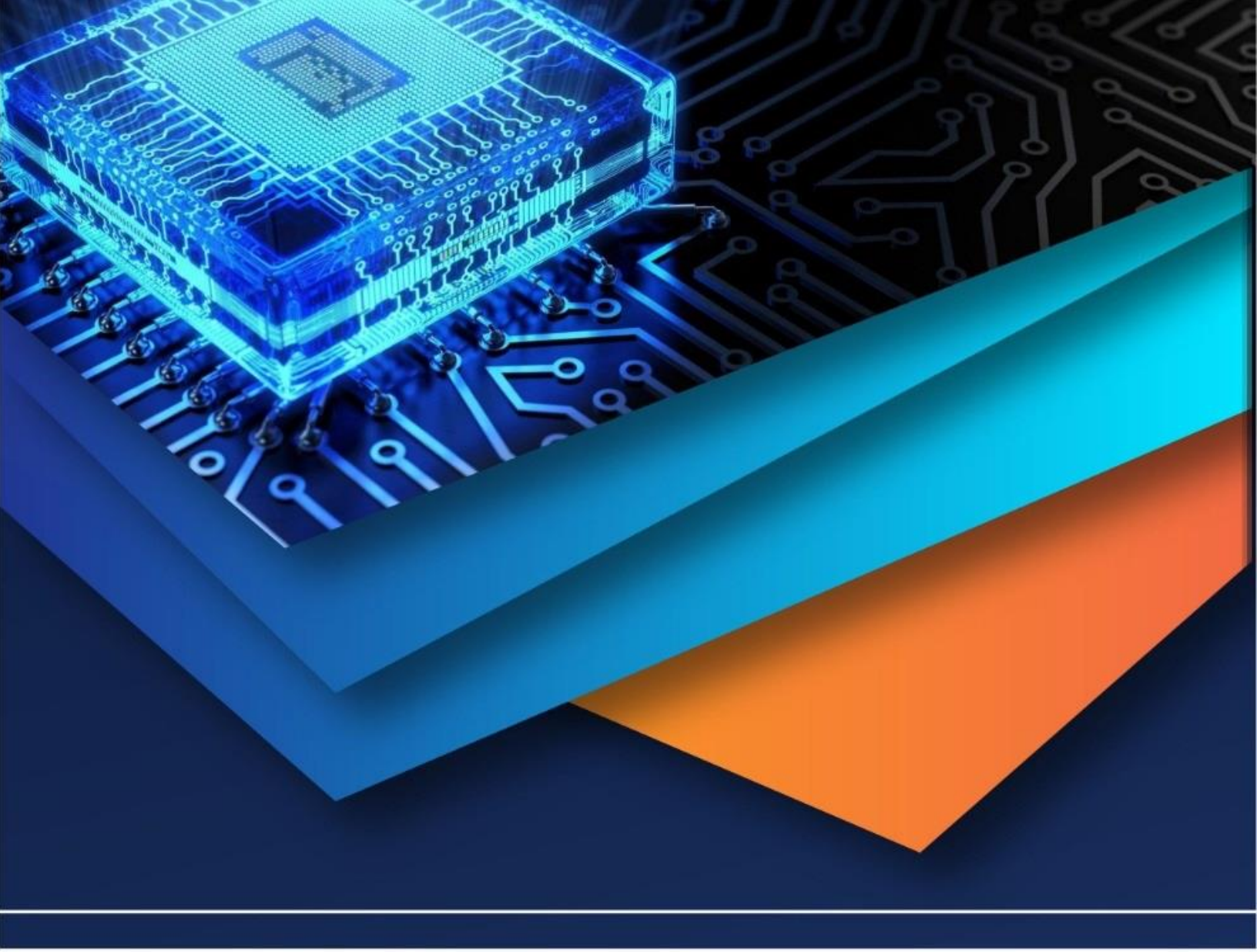
[2] Almohimeed, Abdulaziz, Wald, M. and Damper, R.I. "Arabic Text to Arabic Sign Language Translation System for the Deaf and Hearing- Impaired Community " EMNLP 2011: The Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), United Kingdom., 2011. pp. 101-109.

[3] Sousa, L., Rodrigues, J.M.F., Monteiro, J., Cardoso, P.J.S., Lam, R."GyGSLA: a portable glove system for learning sign language alphabet", Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9739, 2016. pp. 159–170. Springer, Cham doi:10.1007/978-3-

[4] Aniket Patil, Mrinai Dhanvijay,' Blob Detection Technique Using Image Processing For Identification Of Machine Printed Characters', novateur publications international journal of innovations in engineering research and technology [IJIERT] issn: 2394-3696 volume 2, issue - 10, oct.- 2015

[5] Elgammal, Ahmed, Crystal Muang, and Dunxu Hu. "Skin detection-a short tutorial," Encyclopedia of Biometrics, pp.1-10, 2009.

[6] Aniket Patil, Mrinai Dhanvijay,' Blob Detection Technique Using Image Processing For Identification Of Machine Printed Characters', novateur publications international journal of innovations in engineering research and technology [IJIERT] issn: 2394-3696 volume 2, issue - 10, oct.- 2015



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY