Machine Learning based sign and voice communication system

Mouna shree K[#1], Dr.Venugeetha Y[#2]

[#1] 8th Sem Student, Dept of CSE,
[1]mounaraj33@gmail.com

[#2]Associate Professor, Dept of CSE, Global Academy of Technology, Bengaluru
[2] venugeetha@gat.ac.in

Abstract—"Deaf/Dumb" and blind people face problems in communicating with others with difficulties in dealing with the communication technology. The goal of this paper is to design a desktop human computer interface application that is used to facilitate communication between normal, "deaf/dumb" and blind people. Optical character recognition systems have been effectively developed for the recognition of printed characters.Optical character recognition is an awesome computer vision technique with various applications ranging from saving real time scripts digitally and deriving context based intelligence using natural language processing from the texts. Human Skin detection deals with the recognition of skin-colored pixels and regions in a given image. Skin color is often used in human skin detection because it is invariant to orientation and size and is fast to process.

Keywords— Machine printed characters,Deaf-blind people, Deaf-blind communication devices,Skin Detection.

Introduction

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. 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. Each system is restricted by several limitations based on either the direction of inputs and outputs; or the methodology they are used i.e. (hardware/software, vision based system or hybrid system)[1]. The Arabic sign dictionary system is a vision basedsoftware- 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 asMicrosoft Word.[2]

An 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.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. System analyses video and voice streams on real time with high speed network connection and high performance computing capabilities.

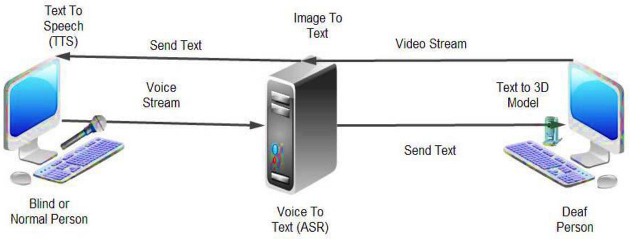


Figure 1: Block diagram of Sign and Voice communication system.

As mentioned earlier, the overall process of an Average Template technique involves using a number of samples initially to account for variability in the speech signal for each word. The overall definitive reference pattern/template is composed from a set of templates rather than one single template. This also serves as to improve the quality of the reference template/pattern by diminishing the effect of a low quality sample out of the overall sample cluster. The most crucial advantage is that the overall procedure is computationally efficient.[3].

Optical character recognition is an awesome computer vision technique with various applications ranging from saving real time scripts digitally and deriving context based intelligence using natural language processing from the texts. One such application is the recognition of machine printed characters. This low cost Computer vision based technique can detect machine part features and serial numbers and creates one to one map between them for identification and quality control. The serial numbers are printed on metal parts. This required us to solve it in two stages. A preprocessing stage which extracts the region of text and segments out the characters and OCR stage which identifies the characters based on a pretrained model built using machine learning techniques.[4].

Two Skin detection is the process of finding skin-coloredpixels and regions in an image or a video. This process is typically used as a preprocessing step to find regions that potentially have human faces and limbs in image. Skin image recognition is used in a wide range of image processing applications like face recognition, skin dis-ease detection, gesture tracking and human-computer interaction. One simple method is to check if each skin pixel falls into a defined color range or values in some coordinates of a color space. There are many skin color spaces like RGB, HSV, YCbCr, YIQ, YUV, etc. that are used for skin color segmentation. We have proposed a new threshold based on the combination of RGB, HSV and YCbCr values. The following factors should be considered for determining the threshold range:
1) Effect of illumination depending on the surroundings.
2) Individual characteristics such as age, sex and body parts.
3) Varying skin tone with respect to different races.
4) Other factors such as background colors, shadows and motion blur.

The skin detection is influenced by the parameters like Brightness, Contrast, Transparency, Illumination, and Saturation. The detection is normally optimized by taking into consideration combinations of the mentioned parameters in their ideal ranges.[5].

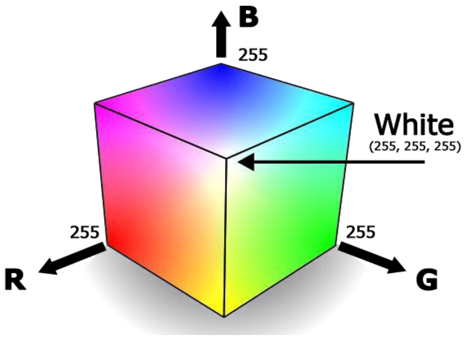


Fig. 2 Color Model

Speech to Video Conversion

The Direction from blind to "deaf/dumb" (voice to images) as shown in fig.3 The voice from blind is converted into its corresponding text using Speech-to-Text (STT) API. Then, the natural language is audio stream from normal/blind user. The microphone is to in be at the tip of the tongue all the recording time this allows the ambient noise to be minimized.

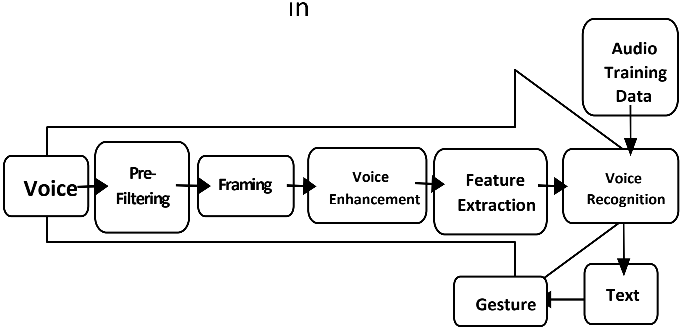


Fig.3 Direction from blind to "deaf/dumb"

- *Audio/Utterance Recording:* Starting points can be found by comparing ambient audio levels (acoustic energy in some cases) with the sample just recorded. Endpointdetection is harder because speakers tend to leave "artifacts" including breathing/sighing, teeth chatters, and echoes.
- *Pre-Filtering:*- The "Bank-of-Filters" method is used. This method utilizes a series of audio filters to prepare the sample. The Linear Predictive Codingmethod uses a prediction function to calculate differences (errors). Different forms of spectral analysis are also used.
- *Framing/Windowing:* - This is involved separating the sample data into specific sizes. This is often done as previously mentioned. This step also involves preparing the sample boundaries for analysis (removing edge clicks, etc.)
- *Comparison and Matching*: - The final preparation for each window before comparison and matching is done such as time alignment and normalization. Combinations of many techniques are used for comparison and matching involving comparing the current window with known samples. These combinations are Hidden Markov Models, Dynamic time wrapping, frequency analysis, differential analysis, linear algebra

techniques/shortcuts, spectral distortion, and time distortion methods. All these methods are used to generate a probability and accuracy match.

B. *Voice Enhancement*

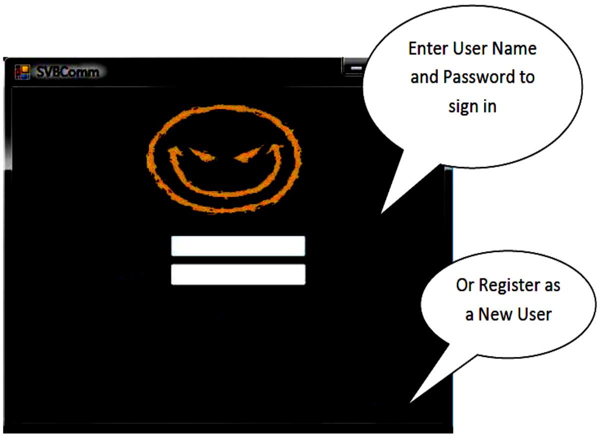
This subcomponent is responsible for enhancing audio to be correctly featuring extracted by normalizing and removing any noise that code results through the recording process. This is done using an image averaging filter such as Cepstral Mean Subtraction (CMS). This filter removes mean value from cepstral parameters to reduce convolution noise, in the cepstral domain by assuming that there is enough of a signal that the mean is not significantly influenced by the speech component of the signal.

C. *Feature Extraction and Dynamic Time Wrapping*

It is important to resolve the quality of the designed reference templates because the accuracy of the DTW-based speech recognition systems is greatly relies on [6]. The system selects reference templates by selecting one example template then tests its recognition rate. If the recognition rate is rising, this reference is kept; or another template should be selected. Vector quantization (VQ) is used to improve the recognition performance. VQ prepares reliable templates for the feature vectors used are the Mel Frequency Cepstral Coefficients (MFCC).[3]

System Implementation

System builds a graphical user interfaces with a web-based and desktop applications as shows in Fig. 1. System is implemented using different tools such as Microsoft Visual Studio 2005 , Matlab , Aforge.net libraries, XNA 0.2(Skeletal Animation Programming) and StarUML in order to maximize the productivity and quality with its automatically generating numerous results. This enables the system to solve technical computing problems faster and interactive exploration and design.

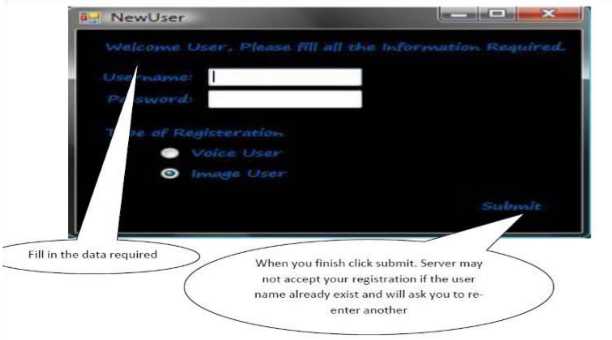


Enter User Name and Password to sign in

Or Register as a New User

Fig. 1 System Graphical User Interface

The user has two singing options as shown in Fig. 2. The first is to register as new users with specifying wither the user is a deaf or blind/normal. The second is to sign in if he is already a registered user. If the user is a new blind user, the voice identification process will start. The user will record as many words as needed for user contrast well,



Welcome User, Please fill all the Information Required.

Username:

Password:

Type of Registeration

○ Voice User

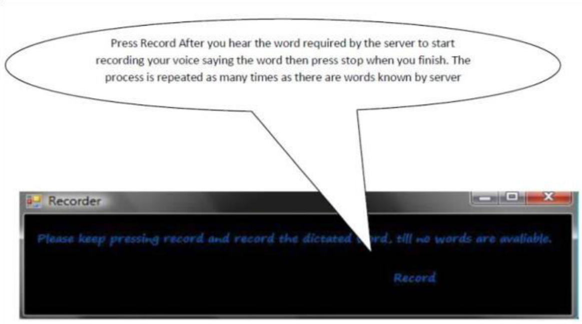
● Image User

Fill in the data required

When you finish click submit. Server may not accept your registration if the user name already exist and will ask you to re-enter another

Fig. 4.System signing options

If the user is a new deaf user, the image identification process will start. The user will supply image to the system representing the dictated word to be sent to the server to train the neural networks. The user may be asked to test his voice for the words that will be dedicated by the system. The user will press record after hearing the sentence "Start Recording your Voice". The user will start speaking then he press stop after finishing as shown in Fig. 3All hypertext links and section bookmarks will be removed from papers during the processing of papers for publication. If you need to refer to an Internet email address or URL in your paper, you must type out the address or URL fully in Regular font.

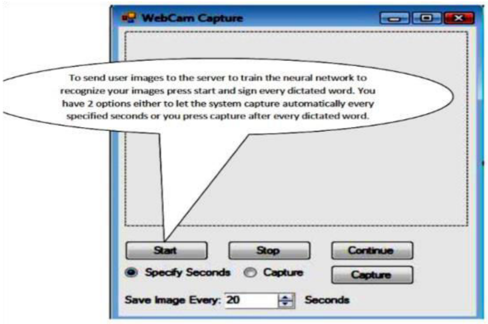


Press Record After you hear the word required by the server to start recording your voice saying the word then press stop when you finish. The process is repeated as many times as there are words known by server

Please keep pressing record and record the dictated word, till no words are avaliable.

Fig. 3.User Voice Testing Phase

The process is repeated as many times as there are words are needed to be recognized by the system. The user could supply image to the system representing the dictated word. The user presses start to send images to the system in order to train the neural networks for the recognition process. The user has two options either to let the system capture automatically every specified seconds or to press himself capture after every dictated word as shown in Fig. 4.



To send user images to the server to train the neural network to recognize your images press start and sign every dictated word. You have 2 options either to let the system capture automatically every specified seconds or you press capture after every dictated word.

● Specify Seconds ○ Capture

Save Image Every: 20 Seconds

Fig. 4. System Image User Option

After finishing registration, the user now is able to sign in with the new information easily. Server may not accept the registration if the user name already exists. The user will be asked to re-enter another log in information as shown in Fig. 5.

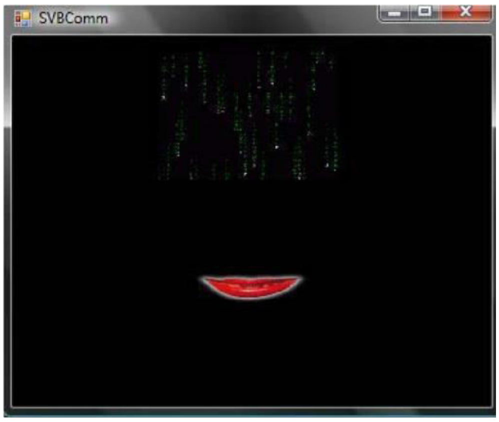
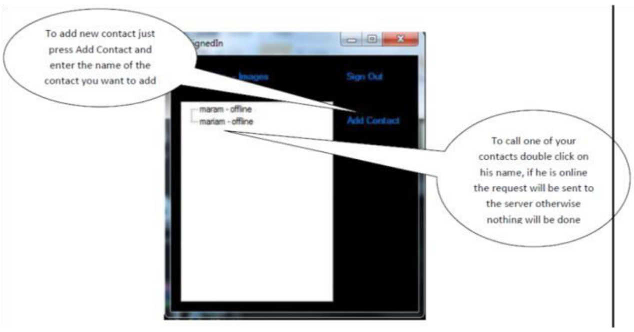


Fig. 5.User successfully Signing in

After successfully signing in, the server accepts the connection by processing the received information (username and password) to check whether the user exists on the application's database or not. Then, the system provides the user with a list with all the contacts he added/may add as longas he is an existed and registered user as shown in Fig. 6. By choosing any contact name, a message will be sent to the other client, in order to accept the call. If the contact is busy the request won't be sent to him, the user will receive a rejection message directly, if not a request is sent and the reply will be appeared on the original user screen as shown in Fig. 17 and Fig. 18.



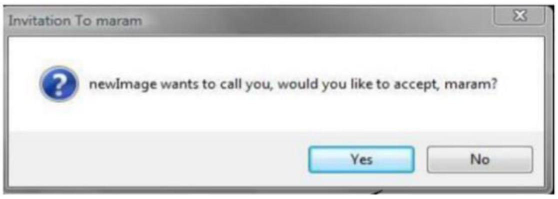
To add new contact just press Add Contact and enter the name of the contact you want to add

maram - offline
mariam - offline

To call one of your contacts double click on his name, if he is online the request will be sent to the server otherwise nothing will be done

Fig. 6. A contacts list provides by the system

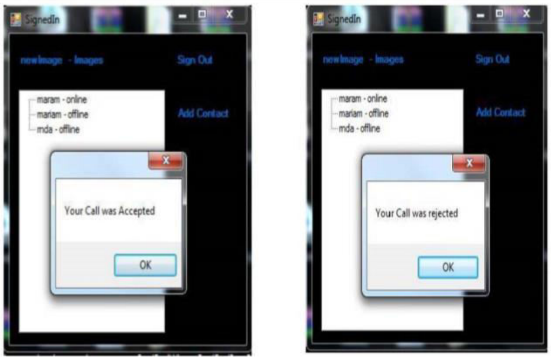
By choosing any contact name, a message will be sent to the other client, in order to accept the call. If the contact is busy the request won't be sent to him, the user will receive a rejection message directly, if not a request is sent and the reply will be appeared on the original user screen as shown in Fig. 7 and Fig. 8.



Invitation To maram

newImage wants to call you, would you like to accept, maram?

Fig. 7. call Acceptence Message



Your Call was Accepted

Your Call was rejected

Fig. 8. User Accepte/Reject n call

RESULTS

The 'deaf/dumb' user sends a video stream 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 text box.

IV. Conclusions

The version of this template In this paper, a Blind/Normal to Deaf/Hard-of-Hearing Chat system is implemented to translate sign language gestures into the corresponding computer generated/human speech gestures. Thus Blob detection technique is used for identification of machine printed characters. Edge detection of characters is also achieved with removal of unwanted blobs in an image. The reference template prepared by the mentioned average template technique is simply derived from a few examples of the words to be recognized rather than selecting just one example as a reference. It differs from the VQ technique in a sense that it requires fewer examples, and it doesn't incur any quantization errors. A threshold based algorithm which recognizes skin image using the RGB-HSV-YCbCr model.

References

[1] Mariam Moustafa Reda, Nada Gamal Mohammed ,Sign-Voice Bidirectional Communication System for Normal, "Deaf/Dumb"and Blind People based on MachineLearning.,2018.

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

[3] Mutcha Srinivasa Rao,"Pattern Normalization/Template Optimization in Order To Optimize Speech Recognition Process", International Journal of Scientific Research and Reviews, 2012, 1(2), pp.69

[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] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, J. Jatakia,"Human Skin Detection Using RGB, HSV and YCbCr Color Models", ICCASP/ICMMD-2016. Advances in Intelligent Systems Research. Vol. 137, Pp. 324-332, 2017

[6] Recognition & Verification System Using Back Propagation Neural Network"" Volume 2, No. 1, January, 2013,IJIEASR ISSN: 2319-4413