

A Survey on Gesture Recognition Technique for Hearing and Speech Disable Person

Anjali Singh¹, Balram Timande²

¹Digital Electronics
Department of Electronics and Telecommunication Engg.
RCET, Bhilai
anjali.singh15692@gmail.com

²Prof. Balram Timande
Associate Professor
Department of Electronics and Telecommunication Engg
RCET, Bhilai
balramtimande@gmail.com

Abstract: *Unable to speak is considered to be true disability. People having these disabilities uses different modes to communicate with others, there are many methods are available for their communication, one such common method of communication in sign language. Sign language allows people to communicate with human body language; each word has a set of human actions representing a particular expression. The main aim of this paper is to convert the human sign language to Voice with human gesture understanding and motion capture. A vision based Hand gesture Recognition system that recognizes hand gesture in the air, especially for physically challenged people, and provides recognized character or number as text and corresponding sound.*

Keywords: Sign language, Hand gesture, web cam, gesture recognition, image processing

1. Introduction

Communication is the activity of exchanging the information between two or more people in order to convey or get the meaningful message through a shared system of signs and semiotic rules. The basic steps of communication are the forming of intent of communication and message composition,

Message encoding then transmission of signal then signal reception and then decoding, and finally the decoded message received by the recipient.

Communication is an internal process and established between three main subject Categories: human beings, living organisms in general and communication-enabled devices.

Communication which come under category of living beings either human being or animal usually occurs through visual, auditory. Human communication is unique for its extensive use of language. But all kind of people or persons are not able to communicate like normal people and it is main problem in today's world. The devices which are invented for them are cannot be afforded by all kinds of persons that's why it is difficult for them to communicate with other peoples.

Many Deaf people cannot speak a natural language, technically they known as: "mute". The word "dumb" has a very old meaning that means "mute". When a person is unable to make speech or any type of sounds correctly or fluently, or has problem with their voices so that problem is known as **speech disorder**. Difficulty in pronouncing sounds, or voicing disorders and stammering are examples of speech disorders.

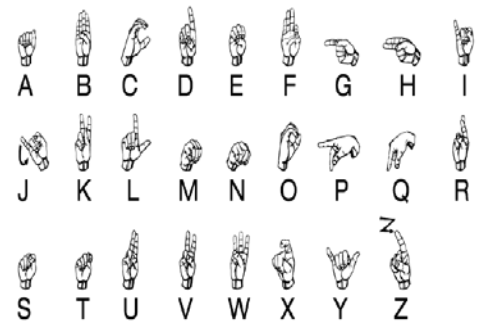


Figure.1 Indian Sign language Hand gesture

A gesture is a form of that communication which is either non-verbal communication or non-vocal communication in which instead of speech, visible body action communicate to display a particular message. Gesture can include the motion of hand, face, or other parts of the body. Gestures can be different from physical non-verbal communication that does not convey the specific messages, such as purely expressive displays, or displays of joint attention. Talking with the help of gestures allow individuals to communicate a variety of feelings and thoughts, through the action of their hand and face, rather than to speak.

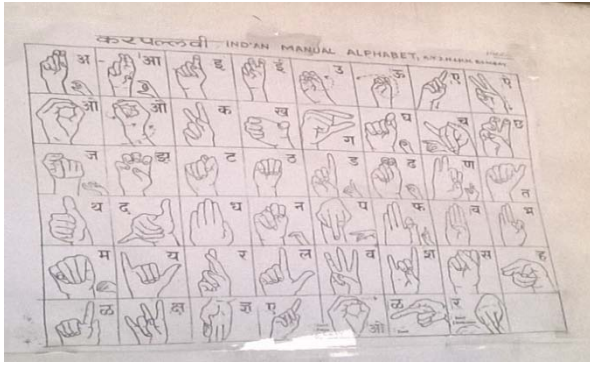


Figure. 2 Hand Gestures of HINDI Sign Language

Gesture recognition can be known as an interface network with computers using gestures of some parts of human body, typically hand movements are considered. Gesture recognition can be an important development for robots that work closely with humans. Gesture recognition is especially valuable in application involving interaction human/robot for several reasons.

2. SURF Algorithm

Earlier time a slow algorithm SIFT is used for key point detection and discription i.e., SIFT algorithm and needed more speed so SURF: Speeded Up Robust Feature[1] is introduced which is an advance version of SIFT. In SIFT, for finding scale space low approximated Laplacian of Gaussian with Difference of Gaussian. As comparing on SURF, it goes a little further and approximates LoG with Box Filter. Figure [3] shows a demonstration of such an approximation. Convolution with box filter can be easily calculated with the help of integral images is the big advantage of SURF algorithm. And it can be done in simultaneously for different scales. The SURF algorithm depends on determinant of Hessian matrix for both location and scale.

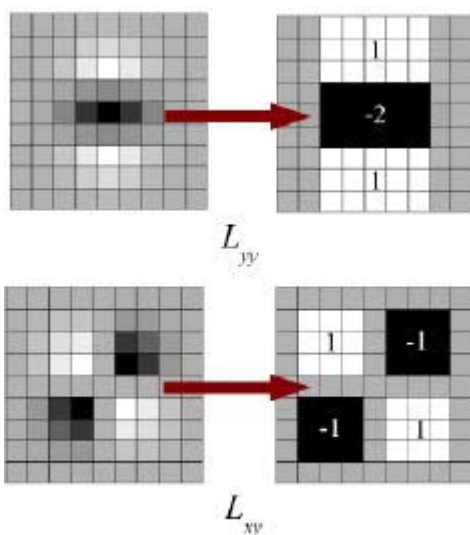


Figure 3:

SURF uses wavelet responses in parallel and perpendicular direction for orientation assignment of a neighborhood of size 6s. Satisfactory Gaussian weights are also applied to the algorithm. By calculating the

addition of total responses within an orientation window of angle 60 degrees the dominant orientation is estimated. By using integral images very easily at any scale wavelet response can be found and it is very interesting for many applications, no requirement of rotation invariance, so there is no need of finding this orientation also, which speeds up the process. Functionality like Upright-SURF or U-SURF is provided by SURF. It improves speed and is robust upto $\pm 15^\circ$. OpenCV[8] supports both, depending upon the flag, upright. Orientation or prenotation is calculated if it is 0 or it is not calculated if it is 1 and then it will be faster.

3. OSTU's Method

Some time when the image is store or recognized some noise also get inside the image which can distort the image and can create an error, so to it is very necessary to eliminate the noise, Ostu's method [2] is used to eliminate the noise. It based on a very simple idea i.e., find the threshold that minimizes the weighted within-class variance. Then that will turns out to be the same as maximizing the between-class variance. Operates directly on the gray level histogram [e.g. 256 numbers], so it's fast once it is computed. Histogram (and the image) which is used that should be bimodal. Spatial coherence or any other notion of object structure is not used. Stationary statistics are assumed but can be modified to be locally adaptive. Uniform illumination is also assumed, so the bimodal brightness behavior arises from object when the appearance of object gets differences only.

4. Skin Color Segmentation

When we start with a color image, the first stage is to transform it to a skin-likeness [3] image. That involves transforming of every pixel from RGB representation to chroma representation and based on the equation given in the previously it determine the likelihood value. The skin-likeness image will be a gray-scale image whose gray values represent the likelihood of the pixel belonging to skin. All skin regions (like the face, the hands and the arms) were shown brighter than the non-skin region.

It is very necessary to note that the detected regions may not necessarily correspond to skin. It is necessary to consummate that the detected region has the same color as that part of skin which is recognized. It is very important that this process can point out regions that do not have the same color of the skin and such part would not need to be considered in the face finding process.

The skin regions are brighter than the other parts of the images, through a threshold process the skin regions can be segmented from the rest of the image. It is not possible to be found a fixed threshold value to process different images of different people with different skin. Since people with different skins have different likeness, a process is required to achieve the optimal threshold value for each run which is known as adaptive threshold process.

The adaptive threshold is based on the observation that stepping the threshold value down may intuitively excise the segmented region. Anyway, the growth in segmented region will gradually decrease (as percentage of skin regions is 100%), but will increase sharply when the threshold value is considerably too small that other regions which are non-skin get included. The value of threshold at which the minimum increase in region size is observed while, when step down the threshold value will be the original threshold. In this program, the threshold value is decreased from 0.65 to 0.05 in steps of 0.1. The minimum increase occurs when the threshold value was changed from 0.4 to 0.3, then the original threshold will be taken as 0.4.

5. Hidden Markov Model (HMM)

A hidden Markov model (HMM) is used for statistical model [1] in which the system being designed by assuming to be a Markov process with unseen (hidden) states. A HMM can be presented as the simplest form of dynamic Bayesian network. L. E. Baum and coworkers are the developers of the mathematics behind HMM. In speech recognition and just a while ago in handwriting recognition Hidden Markov Model has been successfully used.

In simple Markov models (like a Markov chain), the possibilities of state transition are the only parameters because the state is directly seen by the observer. But in a hidden Markov model, there is some different the state is not directly seen, but output, dependent on the state, is seen. Over each possible output tokens all state has a probability distribution. Therefore some information about the sequence of states can be taken by sequence generated by HMM. The word 'hidden' refers to the state disposition, through which the model passes, not to the content of the model. Even if these content are known exactly the model is still referred to as a 'hidden' Markov model.

The main application of Hidden Markov models [1] are temporary pattern recognition such as voice, writing with hand, hand gesture recognition, tagging in part-of-speech, musical score following, and bioinformatics.

A hidden Markov model can be taken as an observation of a mixture model where the unseen variables (or latent variables), which regulates the mixture of the component, which are going to be selected for each observation, and then they are related through a Markov process instead of independent of each other. Nowadays, hidden Markov models have been paired with Markov models and triplet Markov models by which it can allow the consideration of more complex data structures and the modeling of non stationary data.

6. Real Time Video Analysis

OpenCV stands for Open Source Computer Vision [8]. It is a full library of programming functions which are mainly aimed at real-time computer vision technology. The original developer of Open CV was Intel research center which is situated in Nizhny

Novgorod. The library is cross-platform and free for use under the open-source BSD license.

Following are the areas in which OpenCv software is included:

- 2D and 3D feature toolkits
- Ego motion estimation
- Facial recognition system
- Recognition of Gesture
- HCI (Human Computer Interaction)

OpenCV can be written in many languages such as C++ and C++ is in its primary interfaces, but it still retains a not comprising though extensive older C interface. There are combining in Python, Java and MATLAB. The API for these interfaces can be found in the online documentation. Collaboration in other languages such as C#, Perl Ch, and Ruby have been developed to encourage adoption by many people. The C++ interface is now used to run all developments and algorithm of Open CV

7. Lots Notation

It gives a new and improved method in which with the help of a representation technique (LOTS Notation) [5] which will map the nomadic data which are parallel to each word and animate it dynamically we have improved the overall accuracy and a speed of animation. This application consists of three main parts : first to enter words it requires an interface which allows the user to enter the words, a language processing system that converts into English text to Indian Sign Language format and a virtual expression that acts as an conveyor which convey the message at the interface. A novel method based dynamically animated gestures are devised by us in order to map the nomadic data for the parallel word. The word after translation into ISL will be checked in the stored database where the word lies the notation format for each word or not. This notation called as LOTS notation [5].

8. Conclusion

From the above literature survey it is seen that various method to convert gesture into speech. But till now the work has been done for American Sign Language (ASL), British Sign Language (BSL), and for Indian Sign language (ISL). Image extraction and segmentation is done to extract the features of image and then captured image is then converted into text and then it is translated into audio. HMM algorithm, OpenCV software [8], Ostu's method, Speech playback., Microsoft Kinect Sensor, BRIDGE technology [4], LOTS Notation [5], following are the device and methods are used to translate text to audio. This makes possible for hearing and speech disable person to communicate with other people with much less difficulties. Up to 90% of accuracy has been achieved but still there is much to develop.

So the problem which is identified is the “language” in which the project is done is English which is not understandable by all level of persons and of all countries and secondly the accuracy of conversion is to be increased for better communication and third is the delay time during the conversion of gesture into audio or speech is more that have to be reduced.

Refrences

- [1]. Hongwu Yang, Xiaochun An , Dong Oei, Yiong Liu . 2014. IEEE, Towards realizing gesture-to-speech conversion with a HMM- based bilingual speech synthesis system.
- [2]. D.K. Vishwakarma and Rajiv kapoor , 2012. Simple and intelligent system to recognize the expression of speech-disabled person. IEEE , International Conference on intelligent human computer interaction.
- [3]. Prof. R.RItkarkar and Dr. Anil V Nandi , 2013. Hand Gesture to speech conversion using Matlab, IEEE.
- [4]. Prof. Prashant G Ahire, Kshitija B, Tilekar, Tejaswini A, Jawake, Pramod B Warale. ,2015, IEEE, International conference on computing communication control and automation.
- [5]. Ruvinash J Raghavan, Kiran A Prasad, Rahul Muraleedharan, Geetha.M. 2013, Animation system for Indian sign language communication using LOTS notation.
- [6]. M.Sethuraman , Abhinav Rajan, K.P Sai Saran, 2015, Beta Real Time Interface for deaf and Global Entity, IEEE ,2015 International conference on smart technology

Author Profile

Anjali Singh[1], I am the student of M.Tech, branch Digital Electronics in RCET, Bhilai, I have completed my B.E from Ashoka Institute of Technology and Management,Rajnandgaon,Chhattisgarh in 2013. I live in Rajnandgaon.



and management for computing, communication, controls, energy, and materials.

- [7]. Rajaganapathy.S, Aravind.b, Keerthana.B, Sivagami.M , 2015, Conversation of Sign language to Speech with Human Gestures, ELSEVIER .
- [8]. Itisha Gupta, Vinti Parmar, Shubham Tayal, February 2014, Hand Gesture Recognition for Hindi Vowels Using OpenCV, International Journal of Application or Innovation in Engineering & Management (IJAIEM).
- [9]. Neha Poddar, Shrushti Rao, Shruti Sawant,Vrushali Somavanshi, Prof. Sumita Chandak, February 2015, Study of Sign Language Translation using Gesture Recognition. International Journal of Advanced Research in Computer and Communication Engineering
- [10]. Shreyashi Narayan Sawant , M.S. kumbhar, 2014, Real Time Sign Language Recognition using PCA ,IEEE International Conference on Advanced Communication Control and Computing Technologies(ICACCCT),

Websites

- 1 .<https://www.en.wikipedia.org/wiki/Gesture>.
2. https://en.wikipedia.org/wiki/List_of_gestures.
3. https://riweb-bafaru_caroline_2005_1.pdf
4. http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_surfer_intro/py_surf_intro.html.
5. <http://www-cs-students.stanford.edu/~robles/ee368/skinsegment.html>
6. https://en.wikipedia.org/wiki/Hidden_Markov_model
7. <https://en.wikipedia.org/wiki/Opencl>