Sign Language Recognition System

Er. Aditi Kalsh[1], Dr. N.S. Garewal[2]
1 Assistant Professor, Department of Electronics & Communication, GCET, PTU
2 Associate Professor, Department of Electronics & Communication, GNE, PTU

ABSTRACT:

Communication is the process of exchanging information, views and expressions between two or more persons, in both verbal and non-verbal manner. Hand gestures are the non verbal method of communication used along with verbal communication. A more organized form of hand gesture communication is known as sign language. In this language each alphabet of the English vocabulary is assigned a sign. The physically disabled person like the deaf and the dumb uses this language to communicate with each other. The idea of this project is to design a system that can understand the sign language accurately so that the less fortunate people may communicate with the outside world without the need of an interpreter. By keeping in mind the fact that in normal cases every human being has the same hand shape with four fingers and one thumb, this project aims at designing a real time system for the recognition of some meaningful shapes made using hands.

KEYWORDS: *Edge Detection, Human Computer Interface (HCI), Image Processing, kNN Search, Peak Detection, Sign Language Recognition System (SLRs),*

I. INTRODUCTION

A gesture may be defined as a movement, usually of hand or face that expresses an idea, sentiment or emotion e.g. rising of eyebrows, shrugging of shoulders are some of the gestures we use in our day to day life. Sign language is a more organized and defined way of communication in which every word or alphabet is assigned some gesture. In American Sign Language (ASL) each alphabet of English vocabulary, A-Z, is assigned a unique gesture. Sign language is mostly used by the deaf, dumb or people with any other kind of disabilities. With the rapid advancements in technology, the use of computers in our daily life has increased manifolds. Our aim is to design a Human Computer Interface (HCI) system that can understand the sign language accurately so that the signing people may communicate with the non signing people without the need of an interpreter. It can be used to generate speech or text. Unfortunately, there has not been any system with these capabilities so far. A huge population in India alone is of the deaf and dumb. It is our social responsibility to make this community more independent in life so that they can also be a part of this growing technology world. In this work a sample sign language [1] has been used for the purpose of testing. This is shown in figure 1.

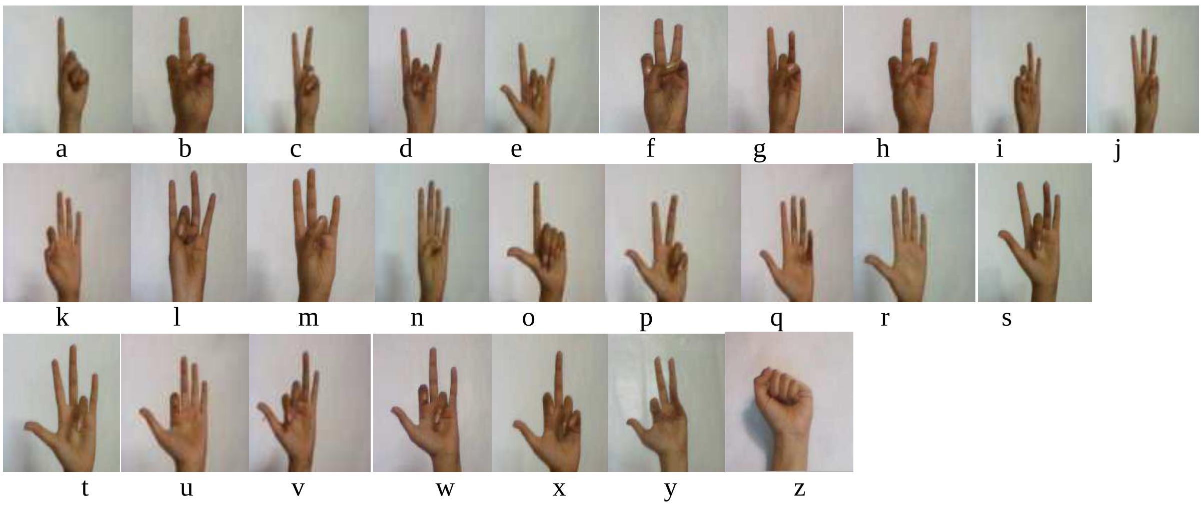


Figure.1. Sample sign language

No one form of sign language is universal as it varies from region to region and country to country and a single gesture can carry a different meaning in a different part of the world. Various available sign languages are American Sign Language (ASL), British Sign Language (BSL), Turkish Sign Language (TSL), Indian Sign Language (ISL) and many more. There are a total of 26 alphabets in the English vocabulary. Each alphabet may be assigned a unique gesture. In our project, the image of the hand is captured using a simple web camera. The acquired image is then processed and some features are extracted. These features are then used as input to a classification algorithm for recognition. The recognized gesture may then be used to generate speech or text. Few attempts have been made in the past to recognize the gestures made using hands but with limitations of recognition rate and time. This project aims at designing a fully functional system with significant improvement from the past works.

II. LITERATURE REVIEW

The importance of sign language may be understood from the fact that early humans used to communicate by using sign language even before the advent of any vocal language. Since then it has been adopted as an integral part of our day to day communication. We make use of hand gestures, knowingly or unknowingly, in our day to day communication. Now, sign languages are being used extensively as international sign use for the deaf and the dumb, in the world of sports by the umpires or referees, for religious practices, on road traffic boards and also at work places. Gestures are one of the first forms of communication that a child learns to express whether it is the need for food, warmth and comfort. It increases the impact of spoken language and helps in expressing thoughts and feelings effectively. Christopher Lee and Yangsheng Xu [2] developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and able to update the model of each gesture in the system in online mode, with a rate of 10Hz. Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove. The most successful commercially available glove is by far the VPL Data Glove as shown in figure 2.

Figure. 2. VPL data glove

It was developed by Zimmerman during the 1970's. It is based upon patented optical fiber sensors along the back of the fingers. Star-ner and Pentland developed a glove-environment system capable of recognizing 40 signs from the American Sign Language (ASL) with a rate of 5Hz. Hyeon-Kyu Lee and Jin H. Kim [3] presented work on real-time hand-gesture recognition using HMM (Hidden Markov Model). P. Subha Rajam and Dr. G. Balakrishnan [4] proposed a system for the recognition of south Indian Sign Language. The system assigned a binary 1 to each finger detected. The accuracy of this system was 98.12%. Olena Lomakina [5] studied various approaches for the development of a gesture recognition system. Etsuko Ueda and Yoshio Matsumoto [6] proposed a hand-pose estimation technique that can be used for vision-based human interfaces. Claudia Nölker and Helge Ritter [7] presented a hand gesture recognition modal based on recognition of finger tips, in their approach they find full identification of all finger joint angles and based on that a 3D modal of hand is prepared and using neural network. In 2011, Meenakshi Panwar [8] proposed a shape based approach for hand gesture recognition with several steps including smudges elimination orientation detection, thumb detection, finger counts etc. Visually Impaired people can make use of hand gestures for writing text on electronic document like MS Office, notepad etc. The recognition rate was improved up to 94% from 92%.

III. SIGN LANGUAGE RECOGNITION SYSTEM

The sign language recognition done using cameras may be regarded as vision based analysis system [9]. The idea may be implemented using a simple web camera and a computer system. The web camera captures the gesture image with a resolution of 320x240 pixels. The captured image is then processed for recognition purpose. The idea may be represented by the block diagram as shown in figure 1.

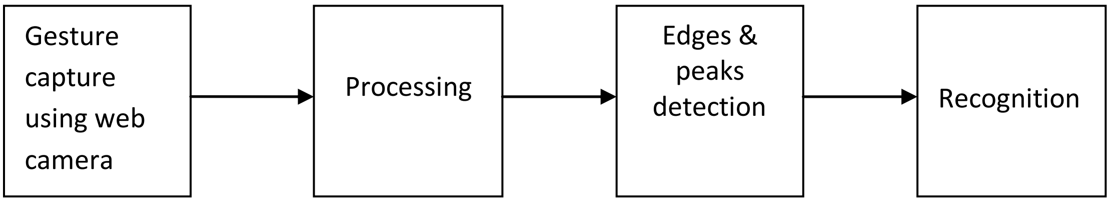


Figure.3. System Block Diagram

3.1 Gesture capture using web camera

The first step towards image processing is to acquire the image. The acquired image that is stored in the system windows needs to be connected to the software automatically. This is done by creating an object. With the help of high speed processors available in computers today, it is possible to trigger the camera and capture the images in real time. The image is stored in the buffer of the object.As has been already discussed, the image is acquired using a simple web camera. Image acquisition devices typically support multiple video formats. When we create a video input object, we can specify the video format that you want the device to use. If the video format as an argument is not specified, the video input function uses the default format. Use the imaqhwinfo [10] function to determine which video formats a particular device supports and find out which format is the default. As an alternative, we can specify the name of a device configuration file, also known as a camera file or digitizer configuration format (DCF) file.

Figure. 4. Acquired image, gesture 'd'

Some image acquisition devices use these files to store device configuration information. The video input function can use this file to determine the video format and other configuration information. The imaqhwinfo function is used to determine if our device supports device configuration files. If the input is an RGB image, it can be of class uint8, uint16, single, or double. The output image, I, is of the same class as the input image. If the input is a colormap, the input and output colormaps are both of class double.
The acquired image is RGB image and needs to be processed before its features are extracted and recognition is made.

3.2 Processing

The captured image is a RGB image. This image is first converted into grayscale as some of the preprocessing operations can be applied on grayscale image only.

Figure. 5. Gray Scale Image, gesture‘d’

3.3 Edges & peaks detection

The edges are detected in the binary image. A number of edge detection techniques [10] [11] may be used in MATLAB. The Edge Detection block finds the edges in an input image by approximating the gradient magnitude of the image. The block convolves the input matrix with the Sobel, Prewitt, or Roberts kernel. The block outputs two gradient components of the image, which are the result of this convolution operation. Alternatively, the block can perform a thresholding operation on the gradient magnitudes and output a binary image, which is a matrix of Boolean values. If a pixel value is 1, it is an edge. For Canny, the Edge Detection block finds edges by looking for the local maxima of the gradient of the input image. It calculates the gradient using the derivative of the Gaussian filter. The Canny method uses two thresholds to detect strong and weak edges. It includes the weak edges in the output only if they are connected to strong edges. As a result, the method is more robust to noise, and more likely to detect true weak edges. In this project we have used canny edge detection.

The purpose of edge detection in general is to significantly reduce the amount of data in an image, while preserving the structural properties to be used for further image processing. Canny edge detection was developed by John F. Canny (JFC) in 1986. Even though it is quite old, it has become one of the standard edge detection methods and it is still used in research. The Canny edge detector works on gray scale image. In image processing finding edge is fundamental problem because edge defines the boundaries of different objects. Edge can be defined as sudden or strong change in the intercity or we can say sudden jump in intensity from one pixel to other pixel. By finding the edge in any image we are just reducing some amount of data but we are preserving the shape. The Canny edge detection algorithm is known as the optimal edge detector. Canny, improved the edge detection by following a list of criteria. The first is low error rate. Low error rate means edges occurring in images should not be missed and that there are NO responses to non-edges. The second criterion is that the edge points be well localized. In other words, the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum. A third criterion is to have only one response to a single edge. This was implemented because the first 2 were not substantial enough to completely eliminate the possibility of multiple responses to an edge . Based on these criteria, the canny edge detector first smoothes the image to eliminate and noise. It then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (non maximum suppression). The gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a non edge). If the magnitude is above the high threshold, it is made an edge. And if the magnitude is between the 2 thresholds, then it is set to zero The resulting image contains a number of discrete objects. The discontinuities are joined using k-

Nearest Neighbor search. The k-nearest neighbor (kNN) search helps to find the k closest points in X to a query point or set of points. The kNN search technique and kNN-based algorithms are widely used as benchmark learning rules—the relative simplicity of the kNN search technique makes it easy to compare the results from other classification techniques to kNN results. They have been used in various areas such as bioinformatics, image processing and data compression, document retrieval, computer vision, multimedia database, and marketing data analysis. You can use kNN search for other machine learning algorithms, such as kNN classification, local weighted regression, missing data imputation and interpolation, and density estimation.

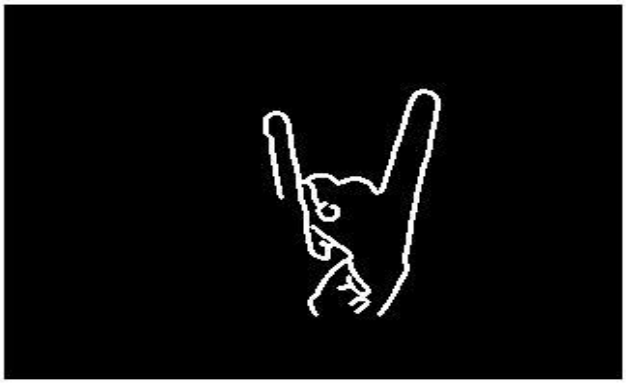


Figure. 6. After Canny's Edge Detection

Watershed Transform may be used in place of kNN search. Watershed transform computes a label matrix identifying the watershed regions of the input matrix A, which can have any dimension. The elements of L are integer values greater than or equal to 0. The elements labeled 0 do not belong to a unique watershed region. These are called watershed pixels. Once the edges are detected, our aim is to detect the finger tips. Wavelet family method is one of the techniques that may be used to detect the peaks. Wavelet analysis consists of decomposing a signal or an image into a hierarchical set of approximations and details. The levels in the hierarchy often correspond to those in a dyadic scale. From the signal analyst's point of view, wavelet analysis is a decomposition of the signal on a family of analyzing signals, which is usually an orthogonal function method. From an algorithmic point of view, wavelet analysis offers a harmonious compromise between decomposition and smoothing techniques. Unlike conventional techniques, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy will depend on the signal and experience. Often the level is chosen based on a desired low-pass cutoff frequency. The finger tips are detected by finding the 1s at the minimum rows. The width and height of the finger is predefined.

3.4 Recognition

Once the finger tips have been detected, our next aim is to match the gesture with the predefined gesture database. This is done using prediction tables. Fuzzy Rule set [10] may be used to make the classification after detecting the finger tips. The logical image is converted back to RGB. An RGB image, sometimes referred to as a true color image, is stored as an m-by-n-by-3 data array that defines red, green, and blue color components for each individual pixel. RGB images do not use a palette. The color of each pixel is determined by the combination of the red, green, and blue intensities stored in each color plane at the pixel's location. Graphics file formats store RGB images as 24-bit images, where the red, green, and blue components are 8 bits each. This yields a potential of 16 million colors. The precision with which a real-life image can be replicated has led to the nickname "truecolor image." An RGB array [12] can be of class double, uint8, or uint16. In an RGB array of class double, each color component is a value between 0 and 1. A pixel whose color components are (0,0,0) is displayed as black, and a pixel whose color components are (1,1,1) is displayed as white. The three color components for each pixel are stored along the third dimension of the data array. For example, the red, green, and blue color components of the pixel (10,5) are stored in RGB(10,5,1), RGB(10,5,2), and RGB(10,5,3), respectively. Wavelet family method is one of the techniques that may be used to detect the peaks. Wavelet analysis consists of decomposing a signal or an image into a hierarchical set of approximations and details. The levels in the hierarchy often correspond to those in a dyadic scale. From the signal analyst's point of view, wavelet analysis is a decomposition of the signal on a family of analyzing signals, which is usually an orthogonal function method. From an algorithmic point of view, wavelet analysis offers a harmonious compromise between decomposition and smoothing techniques. Unlike conventional techniques, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy will depend on the signal and experience. Often the level is chosen based on a desired low-pass cutoff frequency.

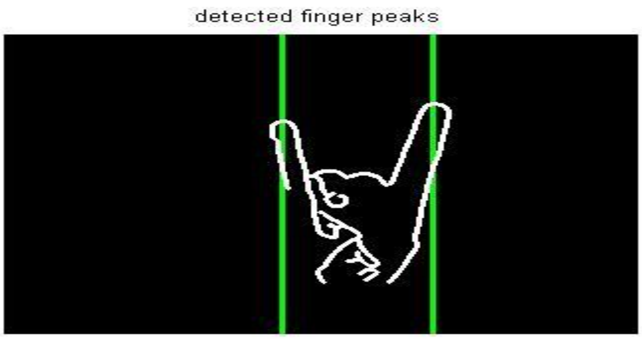


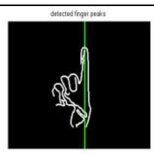
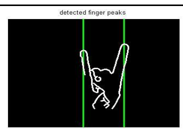
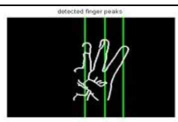
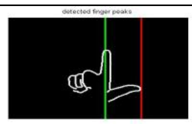
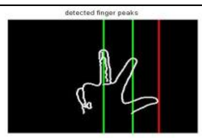
Figure. 7. Detected Finger Peaks

The finger tips are detected by finding the 1s at the minimum rows. The width and height of the finger is predefined. Once the gesture has been recognized, it may be used to generate speech or text.

IV. RESULT AND CONCLUSION

This project is designed to improve the recognition rate of the alphabet gestures, (A-Z), in previously done works []. Six alphabets are chosen for this purpose. These are alphabet A, alphabet D, alphabet J, alphabet O, alphabet P, and alphabet Q. Alphabet A has been chosen as it is recognized easily at 100% rate to show that this project not only improves the recognition rates of lacking alphabets, but also maintains the 100% recognition rate of other alphabets. The recognition time has also been improved significantly.It was observed that the recognition rate was fairly improved and recognition time was reduced significantly. This has been achieved by using knn search instead of contour method as is done before.

Table 1. Result

S.No.	Alphabet	Gesture	Recognized Gesture	Recognition Rate	Recognition Time (seconds)
	A			100%	1.52
	D			100%	1.50
	J			100%	1.61
	O			100%	1.50
	P			100%	1.57

This result is achieved only when the images are taken by following the given set of rules:
The gesture is made using right hand only.
The camera is at a distance of at least 1feetfrom the camera.
The background of the image is plain without any external objects.

The hand is in approximate centre.

This result was achieved by following simple steps without the need of any gloves or any specifically colored backgrounds. This work may be extended to recognizing all the characters of the standard keyboard by using two hand gestures. The recognized gesture may be used to generate speech as well as text to make the software more interactive.This is an initiative in making the less fortunate people more independent in their life. Much is needed to be done for their upliftment and the betterment of the society as a whole.

REFERENCES

[1] Panwar.M. Hand Gesture Recognition System based on Shape parameters. In Proc. International Conference, Feb 2012

[2] Christopher Lee and Yangsheng Xu. Online, interactive learning of gestures for human robot interfaces. Carnegie Mellon University, The Robotics Institute, Pittsburgh, Pennsylvania, USA, 1996.

[3] Hyeon-Kyu Lee and Jin H. Kim. An HMM-Based Threshold Model Approach for Gesture Recognition. IEEE transactions on pattern analysis and machine intelligence, Volume 21, October 1999

[4] P. Subha Rajam and Dr. G. Balakrishnan. Real Time Indian Sign Language Recognition System to aid Deaf-dumb People. ICCT, IEEE, 2011.

[5] Olena Lomakina. Development of Effective Gesture Recognition System. TCSET'2012, Lviv-Slavske, Ukraine, February 21-24, 2012.

[6] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai, Tsulasa Ogasawara. Hand Pose Estimation for Vision based Human Interface. In Proc. 10th IEEE International Workshop on Robot and Human Communication (Roman 2001) , pp. 473-478, 2001

[7] Claudia Nölker and Helge Ritter. Visual Recognition of Hand Postures. In Proc. International Gesture Workshop on Gesture-Based Communication in Human Computer Interaction, pp. 61-72, 1999

[8] Meenakshi Panwar and Pawan Singh Mehra. Hand Gesture Recognition for Human Computer Interaction. In Proc. IEEE International Conference on Image Information Processing, Waknaghat, India, November 2011

[9] Sanjay Meena. A Study of Hand Gesture Recognition Technique. Master Thesis, Department of Electronics and Communication Enginnering, National Institute of Technology, India, 2011

[10] http://www.mathworks.in/help/imaq/imaqhwinfo.html

[11] Wilson A.D, Bobick A.F. Learning visual behavior for gesture analysis. In Proc. IEEE Symposium on Computer Vision, 2011

[12] http://www.happinesspages.com/baby-sign-language-FAQ.html