Vision Based Hand Gesture Recognition Using Eccentric Approach for Human Computer Interaction

Vishal Bhame
Student, Pune Institute of Computer Technology,
Pune, India
vishalbhame09@gmail.com

R.Sreemathy
Prof., Pune Institute of Computer Technology,
Pune, India
rsreemathy@pict.edu

Hrushikesh Dhumal
Software Developer, Motion Vista, Excell Technology Pvt. Ltd.,
Pune, India
hrushikesh.dhumal@gmail.com

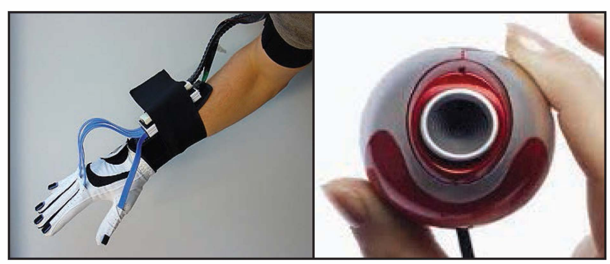
***Abstract*— There has been growing interest in the development of new approaches and technologies for bridging the human–computer barrier. Hand gesture recognition is considered as an interaction technique having potential to communicate with machines. Human computer interaction (HCI) was never an easy task and lots of approaches are available to build such systems. Hand gesture recognition (HGR) using wearable data glove provides a solution to build a HCI system, but lags in terms of its computational time and poor interface. Pattern matching is one more solution which uses vision based techniques and provides strong interface to build HCI systems. But again, it requires complex algorithms which takes lots of computational time and hence limits its use in real time HCI applications. In this paper, we presented an eccentric approach for hand gesture recognition which is simple, fast and user independent and can be used to develop real time HCI applications. Based on proposed algorithm we built a system for Indian Sign Language recognition which converts Indian Sign numbers into text. The algorithm first captures the image of single handed gesture of speech/hearing impaired person using simple webcam and then using our proposed algorithm it classifies the gesture into its appropriate class. It uses simple logical conditions for gesture classification which make its use in real time HCI applications.**

Keywords—Features Extraction; Hand Gesture Recognition (HGR); Human computer Interaction (HCI); Pattern Matching.

I. INTRODUCTION

In present days, the interaction with digital computing devices is advanced in such way that for human, it has become a difficult to live without it. For this reason, Human Computer Interaction has become an active research field since last few years. HCI has several applications mainly classified into three categories as ‘Machine Control’, ‘Sign Language Recognition’ and ‘Vision Based Gaming Navigations’. Controlling machines includes vision based robot control [1], Music player control, TV channel/volume control [2], mouse cursor, power point presentation control etc. The field of sign language recognition includes the development of systems that can facilitate the communication between speech/hearing impaired and vocal people which can be used to integrate the speech and hearing impaired people in our society. The remaining category includes the development of vision based gaming consoles which can be used to play PC games without using conventional hardware like keyboard, joysticks etc. For each said application of HCI, it mainly involves the interaction of human hand or fingers with machine. Hence Hand Gesture Recognition (HGR) is a branch of Human Computer Interaction which deals with interpretation of signs into meaningful information which can be used to develop HCI applications.

The term, Hand Gesture Recognition is collectively referred as the process of tracking human gestures and then interpreting them in the form of meaningful commands. Technologies available for achieving this task are the use of either contact based devices or vision based devices. Figure 1 shows the examples of contact and vision based devices. Contact based devices are usually based on technologies like data glove [3], accelerometers, multi-touch screen etc which uses several detectors based on temperature, propagation, reflection, speed, time etc. Fig 1 (a), shows the Cyberglove II data gloves containing optical fibers as sensing units. Use of sensors and wires in data gloves restricts the movements of user hand and hence limits its use in natural means of human computer interface. Vision based devices are alternatives to this which uses one or more simple image capturing devices like simple web camera as shown in figure 1 (b).



a) Cyberglove II b) Simple web camera

Fig. 1. Contact and Vision based devices used for hand gesture recognition

The main advantage of this method is the degree of freedom for hand movement which is very essential in natural means of communication with machines. In this method, the image of hand gesture is captured using capturing devices and then using various hand detection and tracking algorithm the hand region is separated. This separated hand region is then passed through various recognition algorithms for gesture recognition and classification. Due to the ease of implementation and accuracy in recognition, the vision based

978-1-4799-3080-7/14/$31.00 ©2014 IEEE

hand recognition method is widely accepted by the researchers from worldwide.

Vision based hand gesture recognition is classified into two categories as static and dynamic gestures. Static hand gestures are the orientation and position of hand in the space for fixed amount of time without any movement whereas Dynamic hand gestures includes gestures with movement of hand like waving of hand or movements of fingers. Ghosh presented a static hand gesture recognition algorithm using k-mean based radial basis function neural network [4] which uses preprocessing techniques for hand detection and tracking. A localized contour sequence (LCS) based feature is then used for gesture classification. The proposed algorithm is applicable only for static images containing plain background. Vieriu [5] presented fast yet simple solution for addressing problems in Static HGR. Author used Discrete Hidden Markov Models (DHMMs) that use features extracted from the hand contours. Adithya [6] proposed a method based on artificial neural network for finger spelling recognition based on which an Indian Sign Language recognition system is developed.

Most of the methods proposed in literature use pattern matching techniques which compares extracted input features with database features and hence takes high computational time for comparisons. In this paper we proposed an eccentric approach for hand gesture recognition which does not use a conventional way of extracting input features and comparing it with all database features instead it uses variable distance feature and a simple logic for counting active fingers involved in gestures which makes it to recognize fast and hence can be used in real time HCI application. Using our proposed algorithm, we have recognized Indian Sign Language (ISL) numbers from zero to nine as shown in figure 2.

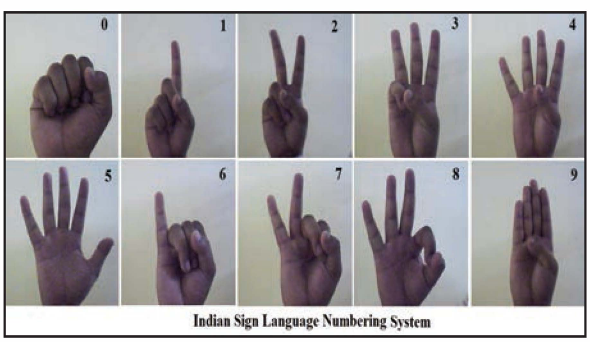


Fig. 2 . Typical Indian sign language numbering system

II. PROPOSED HAND GESTURE RECOGNITION SYSTEM

Figure 3 explains the flow of hand gesture recognition system. Any gesture recognition system mainly composed of image acquisition, image segmentation followed by morphological operations and feature extraction for gesture classification. In this case, the image of hand region with a resolution of 360 x 280 is first captured by using iball C12 Webcam. Hand region is then cropped from captured image by using skin color segmentation. Using 'sobel' edge detection algorithm, edge is calculated from which Centroid and Euclidean features are extracted. Then using our eccentric approach active fingers involved in gesture is calculated which further used for ISL number classification from zero to nine. The System Architecture for HGR consists of six stages as:

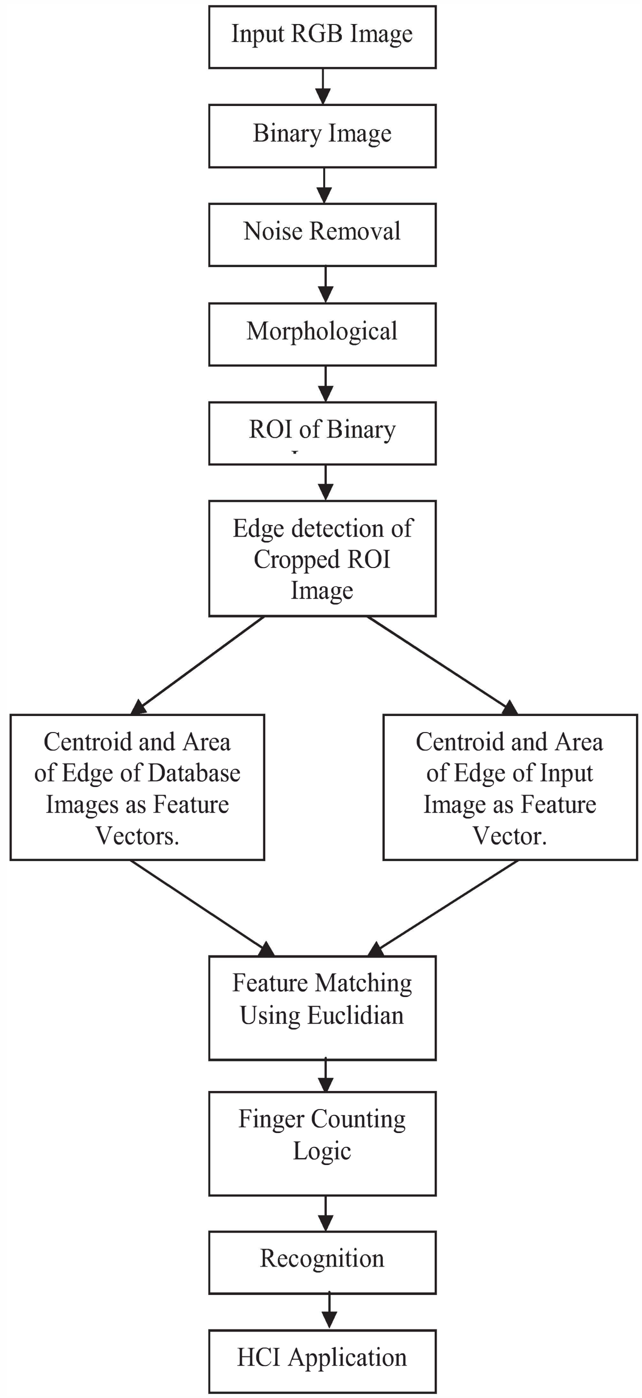


Fig. 3 . Flow of hand gesture recognition system

A. *Image Capturing*

In this proposed method, a 12 megapixel iBall C12 webcam is used for capturing the hand image. Image is captured in RGB color format and then resized to 160 x 120 in order to reduce the computational time required while processing. In this paper, the proposed algorithm is explained with the example of recognizing the gesture of ISL ‘one’, ‘two’ and ‘nine’. Figure 4 shows the captured image of said gestures in RGB colorspace.

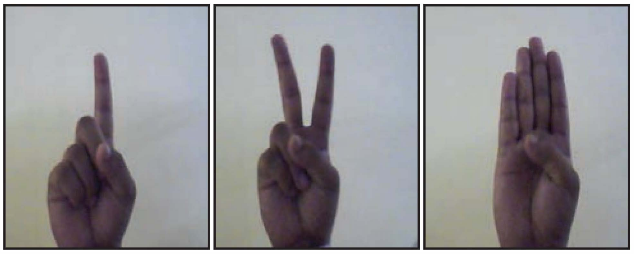


Fig. 4. Captured input images of ISL gesture one, two and nine.

B. Image Preprocessing

This stage consists of hand segmentation followed by morphological operations. Dawod in [7] proposed an adaptive skin color model for hand segmentation by mapping YCbCr color space into YCbCr color plane. Dhruva in [8] explored the various possible ways of segmentation using different color spaces and models. Kook-Yeol Yoo [9] proposed robust hand segmentation and tracking method based on HSV histogram. Most of the methods proposed in literature can segment and track the hand region in simple as well as complex background. For simplicity, we used simple segmentation technique based on computing maximum and minimum skin probabilities of input RGB image as suggested by author in [10]. Figure 5 shows the result images obtained after segmentation. Morphological operations like erosion and dilation are used to remove noise and preserve edges.

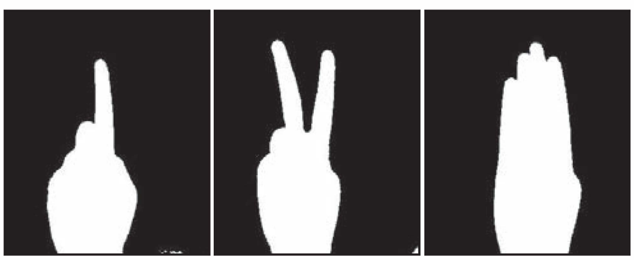
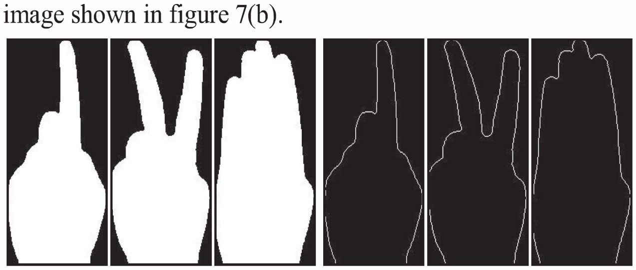


Fig. 6. Result after segmentation of input images.

C. Region of Interest Extraction

First 8-componant connectivity of segmented binary image is calculated using bwlabel and then for cropping hand region regionprop is applied. Figure 7 shows the extracted region of interest. ‘Sobel’ edge detection algorithm is applied on ROI image shown in figure 7(b).



a) Binary image *b) Edge image*

Fig. 7. Extracted region of interest containing only hand region

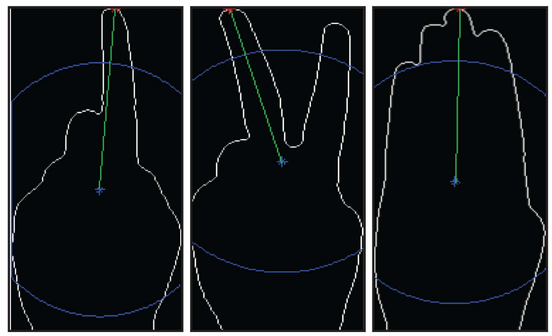
‘Sobel’ algorithm finds edges using those points where the gradient of input binary image is maximum.

D. Finger Counting Logic

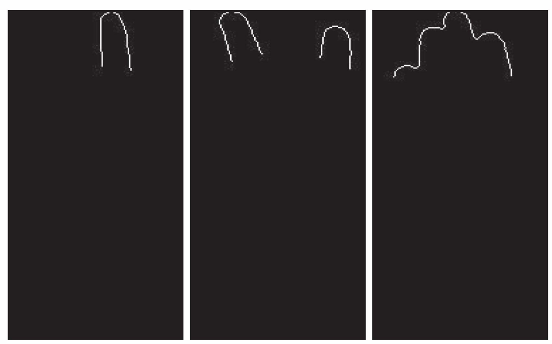
In this section, active fingers involved in input gestures are calculated using variable Centroid distance feature. First, Centroid of the ROI image C (x, y) is calculated using equation:

$$x = \frac{\sum_{i=0}^{i=k} x_i}{k}, \quad y = \frac{\sum_{i=0}^{i=k} y_i}{k} \tag{1}$$

Where X_i and Y_i are x and y coordinates of the i^{th} pixel. Then maximum Euclidian distance ‘Dmax’ between two points, Centroid and the point located on counter of edge image is calculated. In figure 8, blue marker denotes the Centroid point



a. Circle enclosing palm region and intersecting all fingers



b. Gesture recognized as count 1, 2, and 1 respectively.

Fig. 8. Logic for finger counting

whereas red marker denotes the farthest point from Centroid on contour of edge image. Line joining these two points gives maximum distance Dmax shown in green color.

After calculating Dmax, a circle is drawn with radius R = 0.72 of Dmax centered at Centroid point. The figure 0.72 is selected after several experiments at which circle intersect all the fingers. The value of radius R varies with Dmax hence this method is invariant to size, distance from webcam and orientation of user hand. The interior region of circle is masked as non skin region (Pixel value=0) along with last 10 rows of the image as shown in figure 8 (b). Using bwlable and regionprop the number of objects from masked image are counted which gives the number of active fingers involved in gesture. Using above steps a gesture of one, two, three, four and five can be recognized directly. To recognize remaining gestures, simple logical conditions are used as explained under Gesture Classification.

E. Gesture Classification

In this section, from the value of 'finger count' the classification of input gesture is carried out. From observation, it is clear that gesture from 1 to 5 recognized correctly and only task remained is to recognize remaining gestures. Gesture of zero, one, six and nine are recognized as 'count 1'. If input gesture recognized as 'count 1' then its width to height ratio w/h, Number of pixels in Quadrant I say pixcount1 and Quadrant II say pixcount2 are calculated. If the w/h ratio is greater than 0.94 the gesture is classified as 'Zero' else if pixcount1 is greater than pixcount2 the gesture is classified as 'Six'. If difference between pixcount2 and pixcount2 is less than 10, the gesture is classified as 'Nine'. Figure 9 demonstrates the logic used to differentiate between 'one' and 'six'. Same logic is used to classify the gesture of 'two' and 'seven', 'three' and 'eight' when recognized count is two and three respectively.

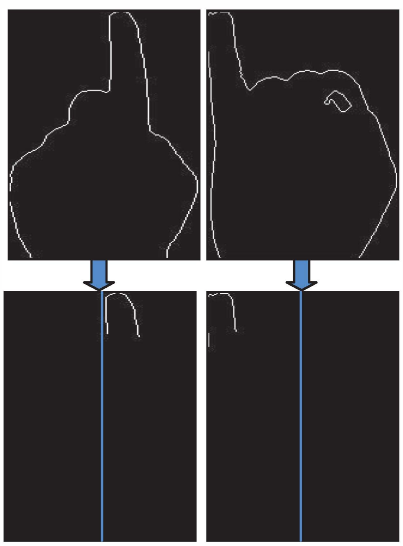


Fig. 9. Logical conditions used for gesture classification

Figure 9 shows the classification of 'count 1' into ISL 'One' and 'Six'. The input image of gesture one and six (Figure 9) are recognized as 'count 1' after passing through proposed algorithm. Then the output image is vertically divided into two quadrants shown by blue lines. The number of active pixels in each quadrant is calculated based on which the gesture is classified as one or six.

F. Display Recognition Result

The recognized hand gestures are displayed in text format. Graphical user interface designed for the same is shown in figure 10.

III. EXPERIMENTAL RESULTS

In the experiments, total 300 samples of hand gestures are collected from 30 different individuals having different hand shapes and sizes and in different lightning conditions. Hence each ISL gesture is tested for 30 samples and recognition result is recorded as given in table 1. Recognition accuracy is calculated using following equation.

$$\text{Recognition accuracy} = \frac{\text{Number of gestures correctly recognised}}{\text{Total number of input gestures}}$$

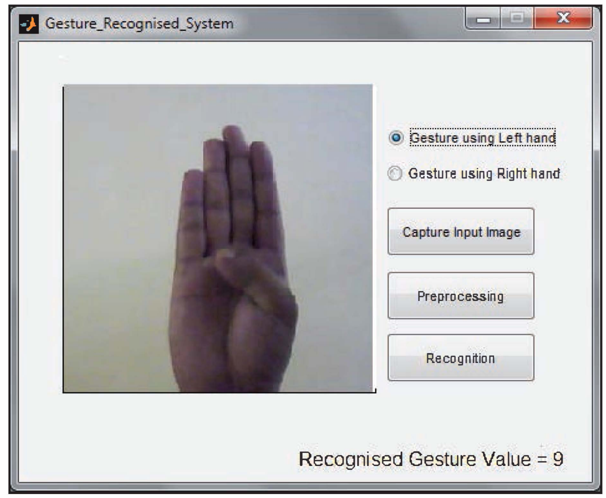


Fig. 10. GUI of Gesture Recognition System

Table 1. The experimental result of recognizing ISL gestures.

Input ISL Gesture	Correctly Recognized Out of 30	Recognition Accuracy in %
Zero	26	86.66
One	30	100
Two	30	100
Three	30	100
Four	30	100
Five	29	96.66
Six	30	100
Seven	28	93.33
Eight	27	90.00
Nine	25	83.33

As the value of radius R varies with Dmax, this method is invariant to size, distance from webcam and orientation of user hand. Also unlike pattern matching techniques, this method does not require any feature comparisons; hence it takes very less computational time and can process up to 6 frames per second which makes its use in real time HCI applications.

IV. CONCLUSIONS

We have developed a method to recognize the ISL input gestures by using variable Centroid distance features. Using this approach, we recognized total ten gestures from Indian sign language. In the experiments, we assume stationary background in order to get good segmentation results. The system is completely autonomous and easy to use as the users do not have to wear any gloves or accelerometer thus making it more feasible. The above procedure aims at providing cost efficient and much simpler approach and can be made even faster if written in C/C++/Open CV.

REFERENCES

[1] Hasanuzzaman M., Ampornaramveth V., Tao Zhang, Bhuiyan, M.A. Shirai and Y. Ueno, "Real-time Vision-based Gesture Recognition for Human Robot Interaction," *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on* , vol., no., pp.413,418, 22-26 Aug. 2004.

[2] Shimada A., Yamashita T. and Taniguchi R., "Hand gesture based TV control system — Towards both user- & machine-friendly gesture applications," *Frontiers of Computer Vision, (FCV), 2013 19th Korea-Japan Joint Workshop on* , vol., no., pp.121,126, Jan. 30 2013-Feb. 1 2013.

[3] Youngmo Han, "A low-cost visual motion data glove as an input device to interpret human hand gestures," *Consumer Electronics, IEEE Transactions on* , vol.56, no.2, pp.501,509, May 2010.

[4] Ghosh D.K. and Ari S., "A static hand gesture recognition algorithm using k-mean based radial basis function neural network," *Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on* , vol., no., pp.1,5, 13-16 Dec. 2011.

[5] Vieriu, R.L., Mironica I. and Goras B.T., "Background invariant static hand gesture recognition based on Hidden Markov Models," *Signals, Circuits and Systems (ISSCS), 2013 International Symposium on* , vol., no., pp.1,4, 11-12 July 2013.

[6] Adithya, V., Vinod P.R. and Gopalakrishnan U., "Artificial neural network based method for Indian sign language recognition," *Information & Communication Technologies (ICT), 2013 IEEE Conference on* , vol., no., pp.1080,1085, 11-12 April 2013.

[7] Dawod A.Y., Abdullah J. and Alam M.J., "Adaptive skin color model for hand segmentation," *Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on* , vol., no., pp.486,489, 5-8 Dec. 2010.

[8] Dhruva N., Rupanagudi S.R., Sachin S.K., Sthuthi B., Pavithra R. and Raghavendra, "Novel segmentation algorithm for hand gesture recognition," *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on* , vol., no., pp.383,388, 22-23 March 2013.

[9] Kook-Yeol Yoo, "Robust hand segmentation and tracking to illumination variation," *Consumer Electronics (ICCE), 2014 IEEE International Conference on* , vol., no., pp.286,287, 10-13 Jan. 2014.

[10] V. Bhame, R. Sreemathy, H. Dhumal, "Vision Based Calculator for Speech and Hearing Impaired using Hand Gesture Recognition," *International Journal of Engineering Research & Technology (IJERT),* Vol. 3, Issue 6, June – 2014.