A Phoneme Based Sign Language Recognition System Using Skin Color Segmentation

Paulraj M P, Sazali Yaacob, Mohd Shuhanaz bin Zanar Azalan, Rajkumar Palaniappan

School of Mechatronic Engineering,
University of Malaysia Perlis, Ulu Pauh Permanent Campus, 02600 Arau, Perlis, Malaysia
prkmect@gmail.com

Abstract: **A sign language is a language which, instead of acoustically conveyed sound patterns, uses visually transmitted sign patterns. Sign languages are commonly developed for deaf communities, which can include interpreters, friends and families of deaf people as well as people who are deaf or hard of hearing themselves. Developing a sign language recognition system will help the hearing impaired to communicate more fluently with the normal people. This paper presents a simple sign language recognition system that has been developed using skin color segmentation and Artificial Neural Network. The moment invariants features extracted from the right and left hand gesture images are used to develop a network model. The system has been implemented and tested for its validity. Experimental results show that the average recognition rate is 92.85%.**

Keywords**: Sign language recognition, hand gesture, Moment invariants.**

I.INTRODUCTION:

Normal people can communicate their thoughts and ideas to others through speech. The only means of communication method for the hearing impaired community is the use of sign language. The hearing impaired community has developed their own culture and methods to communicate among themselves and with ordinary person by using sign gestures. Instead of conveying their thoughts and ideas acoustically they convey it by means of sign patterns. Sign gestures are a non-verbal visual language, different from the spoken language, but serving the same function. It is often very difficult for the hearing impaired community to communicate their ideas and creativity to the normal humans. Development of sign to voice system will be more useful for the hearing impaired to communicate with the normal people more fluently.

II.RELATED WORKS ON GESTURE RECOGNITION

Attempts to recognize sign language automatically began to appear in the literature in the 90s. Many researchers are trying to develop automatic sign language recognition system in various sign languages. Various works have been carried out previously on various sign language recognition techniques. The research on Gesture recognition system can be classified into two [1] types, first is the use of electromechanical devices. This type of system affects the signer's natural signing ability. The second category is classified into two types, one is the use of colored gloves and the other is not using any devices which might affect the signer's natural signing ability.

Eng-Jon Ong and Bowden [2] presented a novel, unsupervised approach to train an efficient and robust detector which is capable of not only detecting the presence of human hands within the image but classifying the different hand shapes. Their approach detects the location of the hands using a boosted cascade of classifiers to detect shape alone in grey scale image. Symeonidis [3] used orientation histograms to recognize static hand gestures, specifically, a subset of American Sign Language (ASL).The system was implemented with a perceptron network.

Foong, Low& Wibowo [4] developed a prototype sign to voice system (S2V) using a feed forward neural network to detect a two-sequence sign. The two-sequence sign language or hand gestures were tested with an average recognition rate of 78.6%.

Segan [5] used an edge based technique to extract image parameters from simple silhouettes developing a system capable of recognizing ten distinct poses in real time using motion orientation vector as features. Vafadar and Behrad [6] developed a hidden markov model based sign language recognition system. A classification accuracy of 90% was reported. Maung [7] developed a hand gesture recognition system to recognize real time gesture in unstrained environments. A pattern recognition system with a transform that converts an image into a feature vector was compared with the feature vectors of a training set of gestures. Experiments show that their system can achieve 90%

recognition rate. The proposed system in this paper is designed to visually recognize 9 phonemes in English.

III.SYSTEM DESIGN:

A video image acquisition process is subjected to many environmental concerns such as the position of the camera, number of camera's used, lighting sensitivity and background condition. The camera is fixed at the center of a cross arm. The cross arm is fixed over a slider of a vertical stand. This is to place the camera at any desired height from the ground level. The illumination level is controlled by using a special lighting arrangement fitted on the cross arm. A normal illumination level of 18lux is maintained for video sequence capturing. The camera is placed at a distance of 1 meter from the ground level and 1.5 meter away from the subject. The experimental set up is shown in Fig. 1.

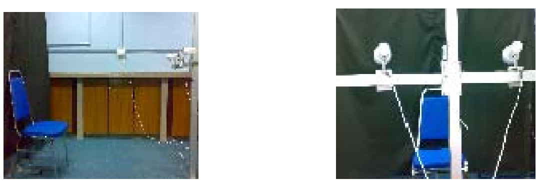

Fig. 1 Experimental set up

IV.METHODOLOGY:

The system is designed to recognize the gestures of English phonemes. The proposed system is very simple and the subject is not required to wear any gloves or any electromechanical device.

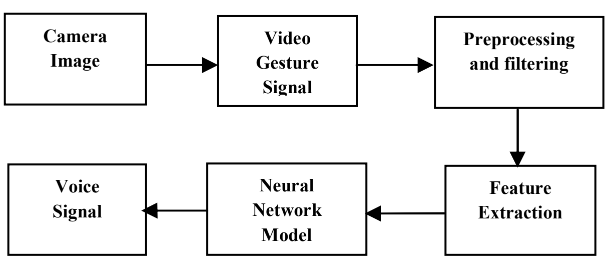

Fig. 2 Block diagram of the proposed system

The proposed system has three processing stages namely preprocessing, feature extraction and gesture classification. Fig. 2 shows the block diagram of the proposed system. Video of 9 different gesture signs are captured. Gesture signs are recorded with a resolution of 640X480 pixels and 30 frames per second (fps).40 frames are required to record a gesture completely. Fig. 3 shows the original image frame. Only frames 9 to 26 were taken for processing from the obtained 40 frames to avoid unwanted data.

Fig. 3 Original Image frame

During the preprocessing stage the skin color detection and region segmentation is carried out. The skin color detection techniques are classified broadly as cbcr color space, RGB Color space, HS (Hue ,Saturation) color space, Normalized RGB & HSV(Hue ,Saturation Value). The RGB color space that defines skin region rule and gives the boundary of skin cluster is used in this work. Based on the values of the RGB (Red, Green and Blue) in the image frame the skin color is detected. The skin region algorithm [8] is as follows.

Step1: Acquire RGB image frame
Step2: Separating R, G and B components.
Step3: if R>95 and G>40 and B>20, do Step3 else it is not skin
Step3: if Max{R, G, B}-Min{R, G, B}>15, do Step4 else it is not skin
Step4: if |R-G|>15 and R>G and R>B it is skin else it is not skin

The Skin color detection algorithm is applied to each of the gesture frame images and segmented. The skin color detected in a particular image frame using the above mentioned algorithm is shown in Fig.4. After the skin color detection the segmentation of left and right hand is carried out as shown Fig. 5. From the segmented right and left hand the moment invariant [9] is obtained. In the feature extraction stage the moment invariant is calculated from the blob alone in the set of image frames. The blob is the high intensity value 1 and the background is 0 in the binary image. This moment invariants value of the right and left hand from the set of frames is taken as the feature vector for the classification purpose. In the gesture classification stage, a simple neural network model is developed for the recognition of gestures signs using the features computed from the video stream.

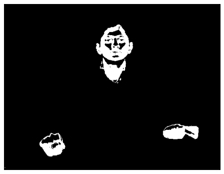

Fig. 4 Skin Color Segmented Image frame

(a) (b)

Fig. 5 (a) Segmented Right Hand, (b) Segmented Left Hand

The system comprises of gesture for nine phonemes in English. The subjects were instructed to make nine different gestures. Three subjects where used in the experimental study. Each subject performs the nine gestures representing the nine different phonemes. Ten such trials were performed by each subject for each gesture. Thus a database containing of 270 samples was created. While creating the gestures for the various phonemes, it is also considered that there is no overlap between the two hands to avoid occlusions caused due to over lapping of two hands in normal signs.

V.MOMENT INVARIANTS:

Moment invariants algorithm has been known as one of the most effective methods to extract descriptive feature for object recognition applications [10]. Essentially, the algorithm derives a number of self-characteristic properties from a binary image of an object. This moment invariants are taken as feature from the gesture video streams. The moment invariants properties are invariant to rotation, scale and translation. The 2-D moment of order (p+q) of a digital image $f(x, y)$ of size M × N is defined as

$$m_{pq} = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} x^p y^q f(x, y) \tag{1}$$

where p=0, 1, 2… and q=0, 1, 2… are integers. The corresponding central moment of order (p+q) is defined as

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x-\bar{x})^p (y-\bar{y})^q f(x, y) \tag{2}$$

for p=0,1,2,… and q=0,1,2,…

where $\bar{x} = \frac{m_{10}}{m_{00}}$ $\quad \bar{y} = \frac{m_{01}}{m_{00}}$ (3)

The normalized central moments, denoted η_{pq} ,are defined as

$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$, where $\gamma = \frac{p+q}{2} + 1$ for p+q=2, 3…

Moment invariants derived from the second order moments.

$$\phi_1 = \eta_{20} + \eta_{02} \tag{4}$$

Second order Moment invariants corresponding to the gesture sign shown in Figure 5(a) (with resizing to half size, translated and rotation of 45º) are listed in Table 1.

TABLE I

Second Order Moment invariants for the Figure 5(a).

Invariant	**original**	**Resized**	**Translated**	**Rotated 45º**
Ø1	0.2271	0.2271	0.2271	0.2271

VI.GESTURE CLASSIFICATION:

Artificial Neural Network (ANN) provides alternative form of computing that attempts to mimic the functionality of the brain [11].The neural network architecture has three layers consisting of an input layer, one hidden layer and an output layer. To classify the different gestures a simple neural network models using error back propagation is developed. The network model has 34 input neurons which represents the moment invariant feature vector. There are 20 hidden neurons and 4 output neurons. Initial weights for the neural network are normalized between 0 and 1 and randomized. The feature dataset contains 270 feature vectors. The network is trained with 162 samples and tested with the remaining 108 samples.

The learning rate, momentum factor, training and testing tolerance are shown in Table 2. The performance of the network model is determined using different sets of initial weights. The mean squared error versus epoch graph is shown in Fig. 6. From Table 2 it is observed that the best accuracy obtained for phoneme gesture recognition is 92.85%

TABLE II

NETWORK LEARNING PARAMETERS.

Number of Input neurons:34 Activation Function: Binary sigmoid Training Tolerance: 0.01 Number of samples used for training: 162	Number of hidden neurons:20 Learning Rate: 0.5 Testing Tolerance: 0.3 No. samples used for testing: 108			Number of output neurons:4 Momentum Factor: 0.9 Total Samples:270		
	No of Epoch:			Classification Rate:		
Trail	Min Epoch for Training	Max Epoch for Training	Mean Epoch for Training	Min Classification Rate (%)	Max Classification Rate (%)	Mean Classification Rate (%)
1	7641	8544	8012	83.57	90	88.33
2	7121	10096	8659	85.71	90.71	89.25
3	7229	8916	8340	82.13	92.85	90.15
4	7998	10443	8523	84.26	91.43	88.72
5	7346	8440	8273	85	88.57	87.45
Average	7467	9287	8361	84.13	90.71	88.78

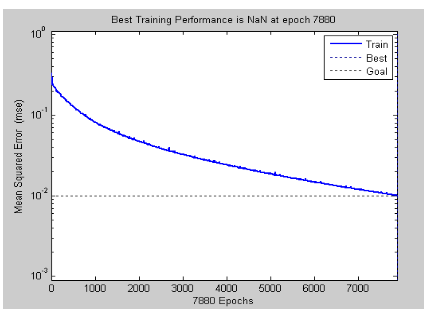

Fig. 6.The Mean Squared Error (mse) versus Epochs Graph

VII.CONCLUSION:

A simple sign language recognition system was developed by using the skin color segmentation and neural network model. The proposed system shows maximum classification accuracy of 92.85%. In future this work will be extended to all the phonemes in English, to improve the results obtained and to develop a Sign to voice system which helps the hearing impaired people to communicate with normal people more fluently.

ACKNOWLEDGMENT

The authors would like to acknowledge the support and encouragement given by the Vice Chancellor of University of Malaysia Perlis (UniMAP), Brig.Jeneral Dato' Prof.Dr.Kamaruddin Hussin. This work is financially assisted by the Fundamental Research Grant Scheme (FRGS):9003-00188 by the Ministry of Higher Education, Malaysia.

REFERENCES:

[1] Qutaishat Munib , Moussa Habeeb, Bayan Takruri, Hiba Abed Al-Malik, "American sign language (ASL) recognition based on Hough transform and neural networks", Expert Systems with Applications 32 (2007) 24–37

[2] Eng-Jon Ong, Richard Bowden, "A Boosted Classifier Tree for Hand Shape Detection," pp.889, Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'04), 2004

[3] Klimis Symeonidis (2000), "Hand Gesture Recognition Using Neural Networks", Thesis, School of Electronic and Electrical Engineering, Centre for Vision, Speech and Signal Processing, Surrey University.

[4] Oi Mean Foong, Tan Jung Low, and Satrio Wibowo, " Hand Gesture Recognition: Sign to Voice System S2V"Proceedings Of World Academy Of Science, Engineering And Technology Volume 32 AUGUST 2008 ISSN 2070-3740

[5] Jakub Segan, "Controlling computer with gloveless gesture", in Virtual Reality System'93, 1993, pp. 2-6.

[6] Maryam Vafadar and Alireza Behrad, "Human Hand Gesture Recognition Using Motion Orientation Histogram for Interaction of Handicapped Persons with Computer" Volume 5099/2009, Springer Berlin / Heidelberg.

[7] Tin Hninn Hninn Maung, "Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks", World Academy of Science, Engineering and Technology 50 2009.

[8] Yona Falinie bte Abdul Gaus, Farrah Wong and Kenneth teo, "Malaysian Sign Language Recognition Using Neural Network". Proceeding of 2009 Conference on Research and development (SCOReD 2009).

[9] William K.Pratt, "*Digital Image Processing-PIKS Scientific Inside 4th ed",* A Wiley-Interscience publication.

[10] Zhongliang, Q., Wenjun, W. "Automatic Ship Classification by Superstructure Moment Invariants and Two-stage Classifier". ICCS/ISITA '92. Communications on the Move, (1992)544-547.

[11] S.N Sivanandam and Paulraj M, "*An Introduction to Artificial Neural Networks"*, Vikhas Publication, India, 2003.