Extraction of Head and Hand Gesture Features for Recognition of Sign Language

Paulraj M P[1] Sazali Yaacob[1] Hazry Desa[1] Hema C.R. [1]
Wan Mohd Ridzuan Wan Ab Majid[1]
[1]*School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), Perlis, Malaysia-02600.*
Email (wgengg@gmail.com)

Abstract

Sign language is the primary communication method that impaired hearing people used in their daily life. Sign language recognition has gained a lot of attention recently by researchers in computer vision. Sign language recognition systems in general require the knowledge of the hand's position, shape, motion, orientation and facial expression. In this paper we present a simple method for converting sign language into voice signals using features obtained from head and hand gestures which can be used by hearing impaired person to communicate with an ordinary person. A simple feature extraction method based on the area of the objects in a binary image and Discrete Cosine Transform (DCT) is proposed for extracting the features from the video sign language. A simple neural network models is developed for the recognition of gestures using the features computed from the video stream. An audio system is installed to play the particular word corresponding to the gestures. Experimental results demonstrate that the recognition rate of the proposed neural network models is about 91%.

1. INTRODUCTION

Sign language is the fundamental communication modality among the hearing impaired. Normal people generally enquire a translator to communicate with the hearing impaired. The hearing impaired people have developed their own culture and methods for communicating among themselves as well as with ordinary person by using sign gestures. Sign gestures are a non-verbal visual language, different from the spoken language, but serving the same function. Recently research works on sign gestures have gained a lot of attention by many researchers in computer vision, pattern recognition and natural language processing. Sign gesture system in general requires the knowledge of hand's position, shape, motion, orientation and facial expressions.

M.K. Bhuyan et al. [1] have developed a framework for hand gesture recognition, they propose a gesture recognition system based on object-based video abstraction technique. Their experimental result shows that their recognition system can be used reliably in recognizing some signs of native Indian sign language.

Thad Starner et al.[2] have presented an extensile system which uses a color camera to track hands in real time and interprets American Sign Language (ASL) using Hidden Markov Models (HMMs) with a vocabulary of 40 signs. Signs are modeled with four-states-HMM and they have achieved recognition accuracies between 75% and 99%.

Liang and Ouhyoung [3] used the HMM approach for recognition of continuous Taiwanese Sign Language with a vocabulary of 250 signs using Hidden Markov Models (HMMs). Eng-Jon Ong and Bowden [4] have developed an unsupervised approach to train a robust detector the presence of human hands within an image and classified the hand shape. In their work, a database of a hand gesture images were created and clustered into sets of similar looking hands using k-means clustering algorithm.

Chan-Su Lee et al. have developed real-time recognition system of Korean Sign Language based on elementary components. Basic component classifiers using Fuzzy Min-Max Neural network and fuzzy logic are used to understand the meaning of a gesture. These systems have recognized 31 Korean manual alphabet and 131 Korean signs in real-time with recognition rate 96.17% for Korean manual alphabets and 94.3% for Korean sign words.

Wilson and Bobick [6] have proposed a state-based technique for the representation and recognition of gesture. Noor Saliza Mohd Salleh et al. [7] have presented techniques and algorithms for hand detection

978-1-4244-2315-6/08/$25.00 ©2008 IEEE.

and gesture recognition process using hand shape variations and motion information as the input to the Hidden Markov Model based recognition system. Rini Akmelia et al. [8] have developed a real-time Malaysian sign language translation using colour segmentation and achieved a recognition rate of 90%.
Eun-Jung Holden et al.[9] have implemented an Australian sign language recognition system, which tracks multiple targets objects (the face and hands) throughout an image sequence by using the relative geometrical positioning and shapes of the target objects to extract features for recognition of sign phrases from a single colour camera. Hidden Markov Models (HMMs) are used to recognize Auslan phrases. Experiment were conducted using 163 test sign phrases and have achieved over 97% recognition on a sentence level and 99% success rate a word level. Nariman Habili et al. [10] have developed a hand and face segmentation methodology using color and motion cues for the content-based representation of sign language video sequences. George Awad et al. [11] have proposed a unified system for segmentation and tracking of face and hands for sign language recognition.
In this paper an intelligent system for converting human sign language into voice using the head and hand gestures is developed. A simple gesture extraction algorithm for extracting the features from the images of a video stream is also proposed. By applying a simple DCT on the variation of the shape and size of the head and hand gesture images, features are extracted. 32 different gesture signs are considered in this work. A simple neural network model is developed for the recognition of gestures using the features computed from the video stream. The recognized signs are connected to the specific audio signals using MATLAB software for the communication between the ordinary and deaf people.

2. SYSTEM DESIGN

RGB video images of head and hand gestures are captured using a USB web camera. The captured images have a resolution of 320 X 240 pixels. Higher resolution causes considerable delay in the execution of the acquisition process and longer processing time. A video image acquisition process is subjected to many environmental concerns such as the position of camera, lighting sensitivity and background condition.
The camera is fixed at the center of a cross arm. The camera cross arm is fixed over a slider of a vertical stand so that the camera can be placed at any desired height from the ground level. The cross arm has two adjustable sliding plates on either side of the camera. On each sliding plates, an electric white lamp bulb (9watt) is fixed on each sliding plates. The illumination level on the subject is controlled by adjusting the position of the sliding plates fitted with the electric bulb. A normal illumination level of 15-18 lux is maintained for image capturing. While capturing the gesture images, the camera is placed in front of the subject at a distance of 1 meter from the ground level and 1.5 meter from the subject. The experimental setup is shown in Figure 1.

Figure 1 Experimental setup

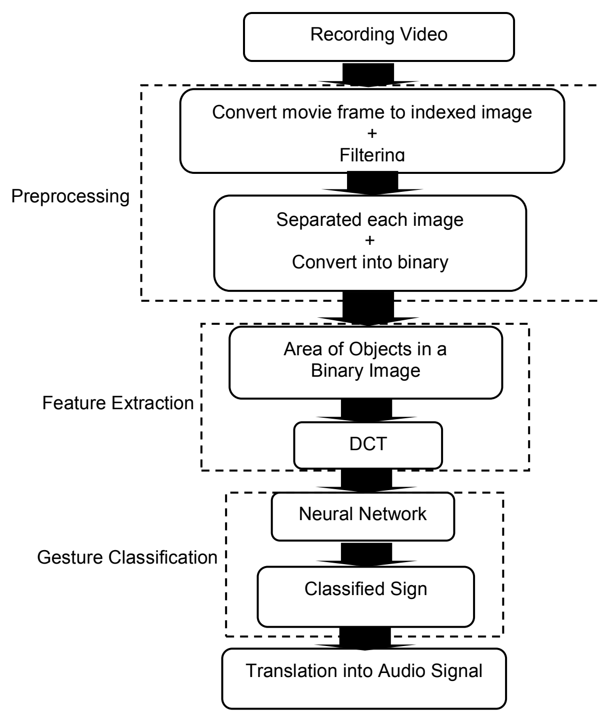


Figure 2 System block diagram

3. METHODOLOGY

The system is designed to visually recognize the gestures of Malaysian Sign Language (MSL). The proposed system is very simple and the subject is not required to wear any gloves. The only restriction for the system is that the subject must wear a dark color long sleeve shirt or jacket.

The proposed system has three processing stages namely preprocessing, feature extraction and gesture classification. Figure 2 shows the block diagram of the proposed method. Videos of 32 different gesture sign are captured. Gesture signs are recorded for one second with a resolution of 320x240 pixels and 42 frames per second (fps). Figure 3 shows a sequence of video stream images that represents a typical sign gesture for the word ‘Floor’. From the recorded 42 frames, the first and last eight frames are removed; only the middle 26 frames are considered for processing. During the preprocessing, the movie frames are converted into indexed image format. Average filter is applied on these images and the unwanted noises are removed.

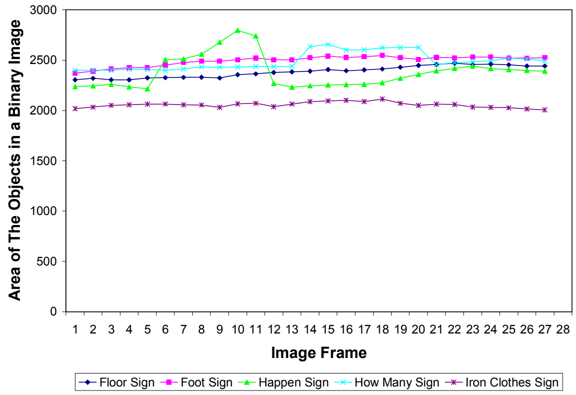
Figure 3 Discrete video sequences to represent a Floor Sign gesture.

Each image frame is segmented into three regions namely, the head, the left hand and the right hand regions. Then these segmented images are converted into binary images. Figure 4 shows the three segmented binary images. In the feature extraction stage, for each frame the area of the segmented binary images are calculated.

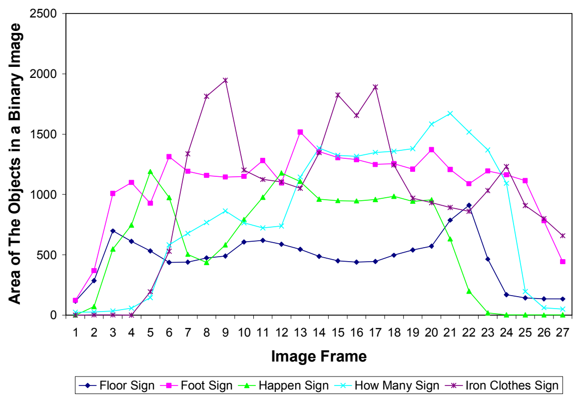
Figure 4 Separated image frame and binary image

Hence each frame has three segmented areas, namely the head area, the left hand area and the right hand area. The variation of the segmented head, left hand and right hand area for all the 26 frames of five different gestures namely ‘floor sign’, ‘foot sign’, ‘happen sign’, ‘how many sign’ and ‘iron clothes sign’ are shown in Figure 5(a), 5(b) and 5(c) respectively. From Figures 5(a), 5(b) and 5(c), it can be observed that each gesture has sweeped different segmented areas. Now treating each segmented area as a discrete event, the DCT is applied and the first 15 DCT coefficients corresponding to each segmented areas are considered as features. The combination of DCT coefficients from the three segmented image areas are used as a feature vector for gesture classification. The variation of DCT coefficient of the different gesture is show in Figure 6(a), 6(b) and 6(c).

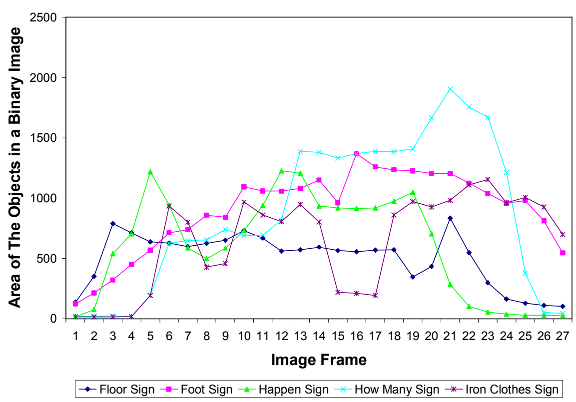
In the gesture classification stage, a simple neural network model is developed for the recognition of gestures signs using the features computed from the video stream.



(a) Segmented Object Area Variation (Head Portion)

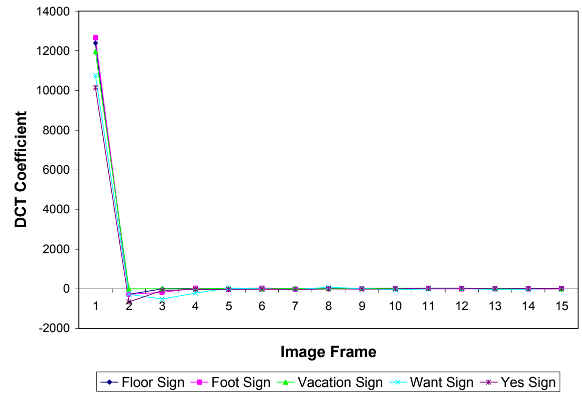


(b) Segmented Object Area Variation (Right Hand Portion)

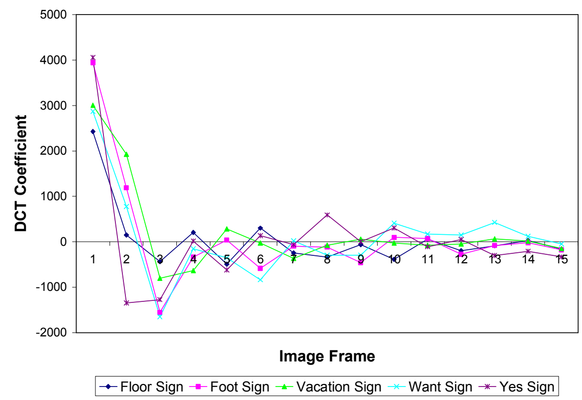


(c) Segmented Object Area Variation (Left Hand Portion)

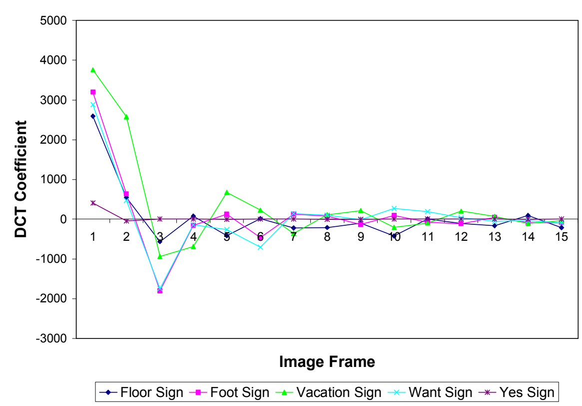
Figure 5 Segmented Object Area Variations



(a) DCT Coefficient for Head Portion



(b) DCT Coefficient for Right Hand Portion



(c) DCT Coefficient for Left Hand Portion

Figure 6 DCT Coefficient

Gestures Classification

Artificial Neural Network (ANN) provides alternative form of computing that attempts to mimic the functionality of the brain [12]. A simple neural network models is developed for sign recognition using the features computed from the video stream. The Neural Network architecture has three layers consisting of an input layer, one hidden layer and an output layer. Table 1 summarizes the different gestures used in the analysis.

Table 1 Sign Language Used

Used Sign
Floor, foot, game, get along, happen, how many, hungry, important, iron clothe, language, play, swimming, take up, ten, tend to, tennis, tired, toilet, toy, truck, vacation, visit, walk, wall, want, wash clothes, what up, when, which, wow, yes, your turn

To classify the different gestures a simple neural network model using error back propagation is developed. The neural network model has 45 input neurons and 6 output neurons. It also has one hidden layer with 32 hidden neurons. The initial weights for the neural network are normalized between 0 and 1 and randomized. The performance of the network model is determined using different sets of initial weights.

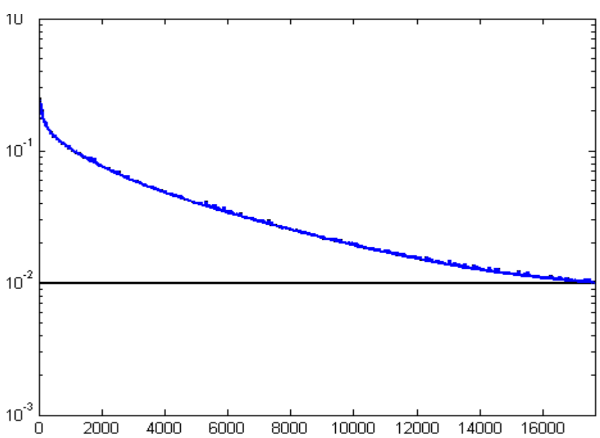


Figure 7 The Cumulative Error versus Epoch Graph

The network is trained with five trial sets of weights. Each trial set consists of 50 weight samples and five such trials sets are used in this experiment. The mean squared error tolerance is fixed as 0.01 for training the network and the network is tested with a testing tolerance of 0.2.

Table 2 Neural Network Training Results

Number of input neurons: 45
Number of Hidden Neurons : 32
Number of output neuron: 6
Activation Function: Binary sigmoid
Learning Rate: 0.25 Momentum Factor: 0.85
Training Tolerance: 0.01
Testing Tolerance: 0.2
Number of samples used for training: 144
No. samples used for testing: 480

Trial No.	Number of Min Epoch for Training	Number of Max Epoch for Training	Number of Mean Epoch for Training	STD Deviation
1	9809	19083	13829	2012.096
2	9809	17645	13628	1945.557
3	10055	17870	13879	1908.999
4	9877	19360	14454	2260.337
5	9809	19202	13865	2225.643
Average	9871.8	18632	13931	2070.526

Table 3 Neural Network Classification Rate

Number of input neurons: 45
Number of Hidden Neurons : 32
Number of output neuron: 6
Activation Function: Binary sigmoid
Learning Rate: 0.25 Momentum Factor:0.85
Training Tolerance: 0.01
Testing Tolerance: 0.2
Number of samples used for training: 144
No. samples used for testing: 480

Trial No.	Min Classification Rate (%)	Max Classification Rate (%)	Mean Classification Rate (%)
1	88.19	95.83	91.75
2	88.19	95.83	91.63
3	88.19	95.14	92.22
4	89.58	95.83	92.86
5	88.19	95.83	91.89
Average	88.47	95.69	92.07

The network is trained using conventional back propagation procedure with momentum and adaptive learning rate. The learning rate and momentum factor are chosen as 0.25 and 0.85 respectively. The hidden and output neurons are activated by binary sigmoidal activation function. 144 samples are used for training the network and the network is tested with 480 samples. The cumulative error versus epoch graph is shown in Figure 7. The network is trained with five trial sets of weight. The training results for the networks namely the minimum epoch, maximum epoch, mean epoch and the STD deviation are shown in Table 2. Table 3 shows the performance of the network classification, namely, the minimum, the maximum and the mean classification rate. From Table 3, it can be observed that the minimum and maximum classification rates are 88.47 and 95.69 respectively.

4. CONCLUSION

In this paper a simple gesture recognition system for Malaysian Sign Language is developed. A simple feature extraction technique based on DCT is proposed. A simple neural network model to classify the gestures is also developed. The developed system has a classification rate of 92.07%. In future work, this work can be extended to help primarily for deaf to solve their communication problems.

ACKNOWLEDGMENT

The authors expressed sincere thanks to Brig. Jen. Prof. Dato' Dr. Kamaruddin Hussin, Vice Chancellor, Universiti Malaysia Perlis for his constant encouragement. This research is funded by the MOHE, FRGS funding [9003-00128] of the Malaysian Government.

REFERENCES

[1] M.K Bhuyan, D. Ghoah and P.K. Bora (2006), 'A Framework for Hand Gesture Recognition with Applications to Sign Language', *Annual India Conference*, pp. 1-6.

[2] T. Starner and A. Pentland (1995), 'Real-time American Sign Language recognition from video using hidden markov models', *International Symposium on Computer Vision,* pp. 265-270.

[3] R. Liang and M. Ouhyoung (1998), 'Real-time continuous gesture recognition system for sign language', *Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition,* pp. 558-567.

[4] Eng-Jon Ong and Richard Bowden (2004), 'A Boosted Classifier Tree for Hand Shape Detection', *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 889-894.

[5] Chan-Su Lee, Zeungnam Bien, Gyu-Tae Park, Won Jang, Jong-Sung Kim and Sung-Kwon Kim (1997), 'Real-time recognition system of Korean sign language based on elementary components', *Sixth IEEE International Conference on Fuzzy Systems*, Vol 24, pp. 1463-1468.

[6] Wilson, A. and A. Bobick (1995), 'Configuration States for the Representation and Recognition of Gesture'. *International Workshop on Automatic-Face and Gesture-Recognition,* Zurich, Switzerland, pp. 129-134.

[7] Noor Saliza Mohd Salleh, Jamilin Jais, Lucyantie Mazalan, Roslan Ismail, Salman Yussof, Azhana Ahmad, Adzly Anuar and Dzulkifli Mohamad (2006), 'Sign Language to Voice Recognition: Hand Detection Techniques for Vision-Based Approach', *Fourth International Conference on Multimedia and Information and Communication in Education*, pp. 967.

[8] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang (2007), 'Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network', *IEEE on Instrumentation and Measurement Technology Conference Proceeding,Warsaw, Poland*, pp. 1-6.

[9] Eun-Jung Holden, Gareth Lee and Robyn Owens (2005), 'Australian sign language recognition', *Machine Vision and Applications,* Vol.16, No. 5, pp. 312-320.

[10] Nariman Habili, Cheng Chew Lim and Alireza Moini (2004), 'Segmentation Of The Face And Hands In Sign Language Video Sequences Using Color And Motion Cues', *IEEE Transactions on Circuits and Systems For Video Technology*, Vol. 14, No. 8, , pp.1086 – 1097

[11] George Awad, Junwei Han and Alistair Sutherland (2006), 'A Unified System for Segmentation and Tracking of Face and Hands in Sign Language Recognition', *18th International Conference on Pattern Recognition*, Vol. 1, pp. 239 – 242.

[12] S.N. Sivanandam, M. Paulraj (2003), *An Introduction to Neural Networks*, Vikhas Publications Company Ltd. India.