SIGN LANGUAGE INTO VOICE SIGNAL CONVERSION USING HEAD AND HAND GESTURES

Conference Paper · February 2008

CITATIONS
2

READS
830

1 author:

Hazry Desa
Universiti Malaysia Perlis
169 PUBLICATIONS **1,209** CITATIONS

SEE PROFILE

Proceedings of the ISCO 2008 on 1-2, February 2008

Coimbatore - 32; India

SIGN LANGUAGE INTO VOICE SIGNAL CONVERSION USING HEAD AND HAND GESTURES

PAULRAJ M P SAZALI YAACOB HAZRY DESA HEMA C.R. M. HARIHARAN
WAN MOHD RIDZUAN WAN AB MAJID
School of Mechatronics Engneering,
Universiti Malaysia Perlis,
02600 Jejawi,Jalan Kangar-Arau,Perlis
MALAYSIA
wanmohdridzuan@gmail.com

Abstract: - The use of gestures as means to convey information is an important part of human communication. Human gesture is a form of visual communication among people. Visual –based automatic gesture recognition has recently acquired much attention and many research works have been carried out to develop intelligent and natural interfaces between users and computer systems based on body movements. The detection and understanding of human gestures in videos is high value for many applications such as surveillance, collaborative environments, training and entertainment, medical support system, sign language, recognition and human computer interaction. In this paper, a simple feature extraction method based on centroid and Discrete Cosine Transform (DCT) is proposed for extracting the features from the video of sign language. 10 different video of sign language are considered in this work. Two neural network models are developed for the recognition of 10 sign languages. The simulation result shows that neural network provides comparable recognition rate of 80%.

Key-words: sign language recognition; head and hand gestures, centroid, neural network

1 Introduction

During recent years human-machine-interfaces have experienced a growing interest. Systems for the analysis of body motion have been developed carefully as a first step for sign language interaction between user and the computer. A special case of body motion is sign language. The sign language is the fundamental communication method between people who suffer from hearing defects. Sign language consists of static and dynamic gestures. In order for an ordinary person to communicate with deaf people, a translator is usually needed to translate sign language into natural language and vice versa. The detection and understanding of human sign language in videos is of high value for many applications, such as human computer interaction, surveillance, collaborative environments, training and entertainment, and medical support.

There are many earliest reported work on sign language recognition have been develop by other researcher. Starner and Pentland have developed a glove-environment system capable of recognizing a subset of the American Sign Language (ASL)[1,2]. Liang and Ouhyoung [3] used the HMM approach for recognition of continuous Taiwanese Sign Language with a vocabulary of 250 signs.Hidden Markov Models (HMMs) are widely used for modeling temporal structures. Yang and Ahuja used Time-Delay Neural Networks (TDNN) to classify motion patterns of ASL [4]. Eng-Jon Ong and Bowden [5] have presented a novel, unsupervised approach to training an efficient and robust detector which applicable of not only detecting the presence of human hands within and image but classifying the hand shape. The Korean Manual Alphabet (KMA) by Jung-Bae Kim, Kwang-Hyun Park and Z.Zenn Bien [6], present a vision-based recognition system of Korean manual alphabet which is a subset of Korean Sign Language. Wilson and Bobick [7] resented a state-based technique for the representation and recognition of gesture.

Noor Saliza Mohd Salleh et al.[8] have presented a research progress and findings on techniques and algorithms for hand detection as it will be used as an input for gesture recognition process. Rini Akmelia et al.[9] have develop real-time Malaysian sign language translation using colour segmentation and neural network where it achieved the recognition rate of over 90%. Eun-Jung Holden et al. [10] presented Australian sign language recognition which tracks multiple target objects (the face and hands) throughout an image sequence and extracts features for recognition of sign phrases.

This paper proposes an intelligent system for converting human sign language into voice obtained from the head and hand gestures. Using camera, system receives sign language video from the deaf people in the form of video streams in RGB (red-green-blue) colour with a screen bit depth of 24-bits and a resolution of 320 x 240 pixels. From each frames of images, head and two hand regions are segmented and then converted into binary image. This method

have been use by Qutaishat Munib et al.[11] in their recent research on American sign language recognition based on Hough transform and neural network. The centroid of the binary images of each image frames are calculated and then Discrete Cosine Transform (DCT) is applied on centroids of all image frames. Artificial Neural Network (ANN) provides alternative form of computing that attempts to mimic the functionality of the brain [12]. A simple neural network model is developed for sign recognition from the features computed from the video stream. An audio system is installed to play the particular word for the communication between the ordinary people and deaf people.

2 Experimental Setup

Video of sign language is captured by using Web Digital Camera in the form of still images and video streams in RGB (red-green-blue) colour with a screen bit depth of 24-bits and a resolution of 320 X 240 pixels. The resolution is chosen to satisfy the execution time constraint. Higher resolution causes considerable delay in the execution of the acquisition process and longer processing time. The wearing of long sleeve jacket while doing the sign is included in this research work. These long sleeve jackets will hide the arms of signer which shown only signer hand.

In this research work, video/image acquisition process is subjected to many environmental concerns such as the position of camera, lighting sensitivity and background condition. The camera is placed to focus on an area that can capture the movement of signer hands. Thus, the proposed position of the camera will be about 1 meter from floor and 1 meter from the signer. In this way, the movement of the hand can be detecting by camera. Sufficient lighting is required to ensure that head and hand is clear enough to be seen and analyzed.

Fig. 1. Experimental setup

3 System Design and Implementation

Our system is designed to visually recognize all sign language of Malaysian Sign Language (MSL) such as simple sign like left. The users/signers are not required to wear any gloves or to use any devices to interact with the size, body size, operation habit and so on, which bring about more difficulties in recognition. The users/signers also have to wear dark colour long sleeve shirt or jacket. However, different signers vary their hand shape size, body size, operation habit and so on, which bring about more difficulties in recognition.

The system have three phases: video/image processing, the feature extraction phase and the classification. Videos were captured from camera and save as in avi file. The video/image processing applied an video/image processing technique which involves using algorithms to detect and isolate various desired portions of the digitized sign. During this phase, single movie frame is converted into indexed image. Average filter is applied for this image for removing the unwanted image noise. From each frames of images, head and two hand regions are segmented and then converted into binary image. The goal of segmentation and binary are to mark the points in an image (sign image) which show the signer head and hands. For the feature extraction phase, the centriod of the binary images of each image frames are calculated and then DCT model is applied on centriods of all image frames.

In the classification stage, a simple neural network model is developed for sign recognition from the features computed from the video stream. An audio system is installed to play the particular word.

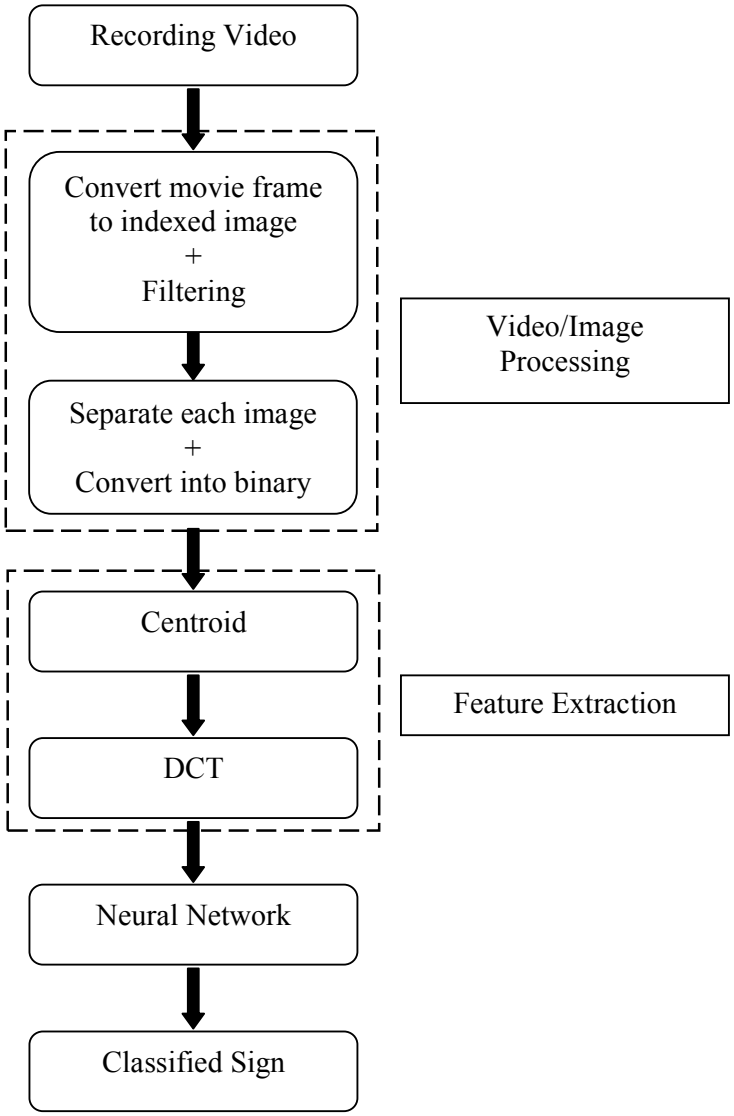

Fig. 2. System Block Diagram

Fig. 3. Video of sign

3.1 Video/Image processing phase

3.1.1 Convert movie frame to indexed image and filtering

Video of sign that have recorded will be saved in .avi file with resolution of 320x240 pixels and 40 frames per second (fps). The "frame2im" function will be used for getting image from movie frame. This function will converting the single movie frame into the indexed image and associated colormap. "fspecial" filter is applied for each image to remove the unwanted image noise.

Fig. 4. Grayscale image frame

3.1.2 Separated each image and convert into binary

From each frames of images, head and two hand regions are segmented and then converted into binary image. Binary images are image whose pixels have only two possible intensity values. They are normally displayed as black and white. Numerically, the two values are often 0 for black, and either 1 or 255 for white. Binary images are often produced by thresholding a grayscale or color, in order to separate an object in the image from the background. The color of the object (usually white) is referred to as the foreground color. The rest (usually black) is referred to as background color. However, depending on the image which is to be thresholded, this polarity might be inverted, in which case the object is displayed with 0 and the background is with a non-zero value.

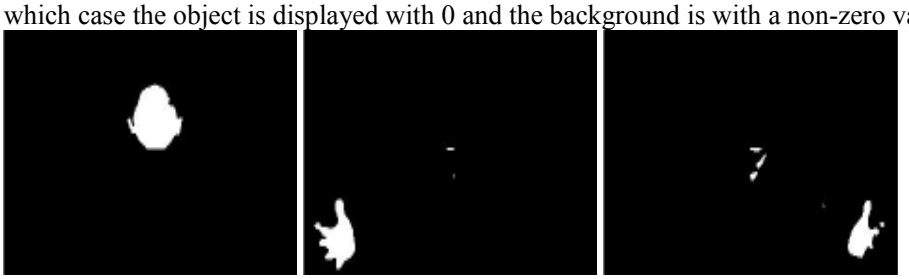

Fig. 5. Separated image frame and binary image

3.2 Feature Extraction phase

In feature extraction phase, centroid function will be used in calculate the centroids of each image frame. Centroid is the function that takes image as an argument (suitably should contain only one object whose centroid is to be obtained) and return the x and y coordinates of its centroid. From all centroids of each image frame, the Discrete Cosine Transform (DCT) is applied. Figure 6 and figure 7 show the centroid of different sign language for the right and left hand.

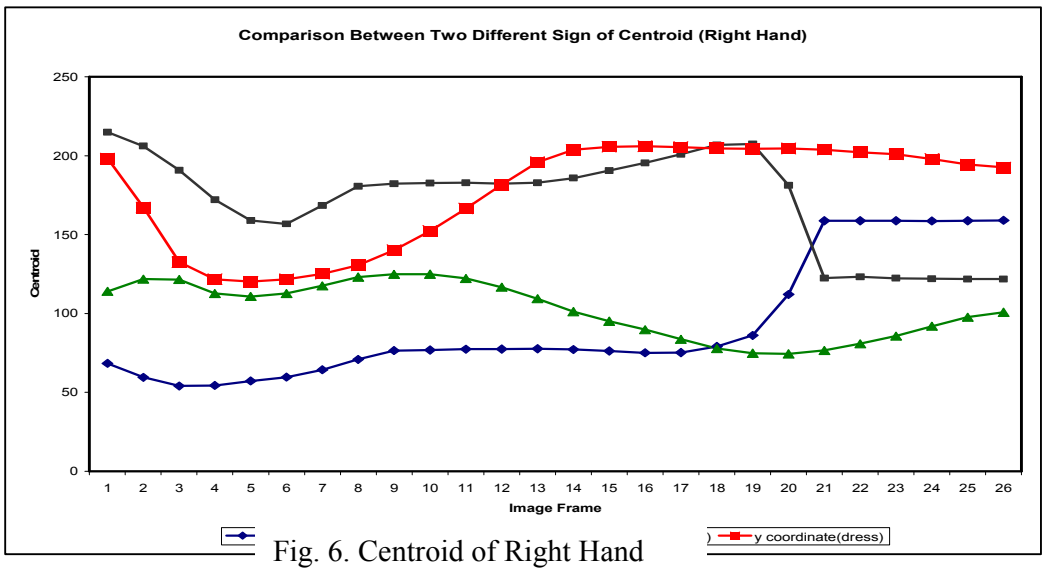

Fig. 6. Centroid of Right Hand

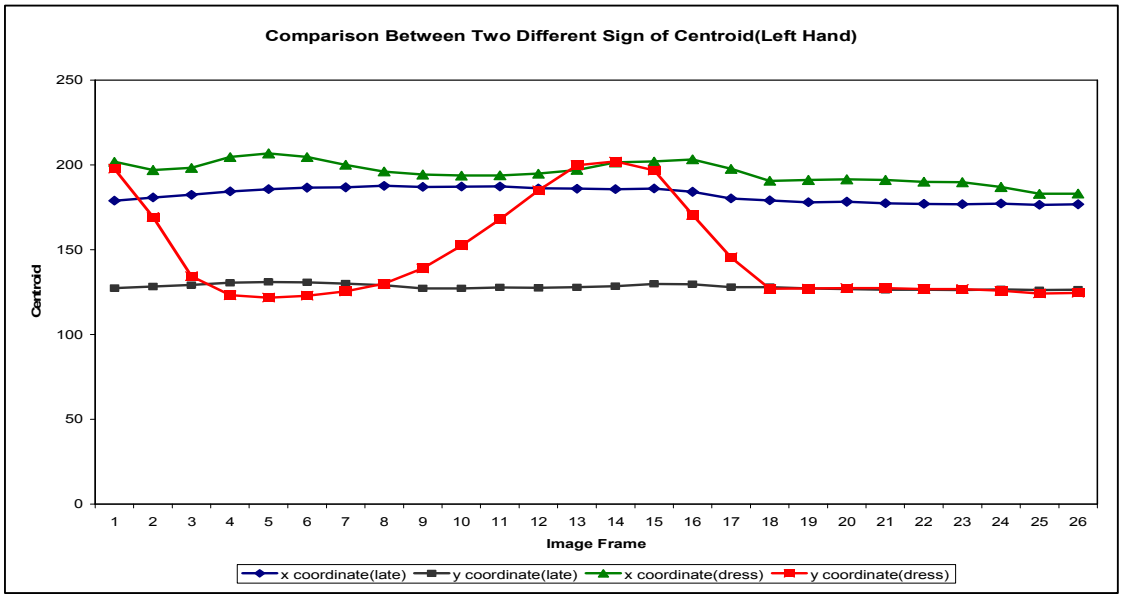

Fig. 7. Centroid of Left Hand

3.3 Neural Network phase

Artificial Neural Network (ANN) provides alternative form of computing that attempts to mimic the functionality of the brain [12]. Two simple neural network models are developed for sign recognition from the features computed from the video stream. The Neural Network architecture has three layers consists of input, hidden and the output layer. 60 input neurons represents of Discrete Cosine Transform (DCT) coefficient from 10 different sign are fed to both network as input data. The first neural network model is used for classifying the following sign, dress, fat, feel, fence and fine sign. Similarly the second neural network model is used for classifying the following sign, have, heart, heavy, here and late sign. The hidden layer consists of 16 neurons which are determined by trial and error and the output consists of 5 neurons. The initial weights for the neural network are randomized between 0 and 1 and normalized. A trial weight set consist of 50 sets of randomized weight samples are considered. The mean squared error tolerance is fixed as 0.01.The learning rate and momentum factor are chosen as 0.1 and 0.9 respectively. Both networks are trained by the conventional back propagation procedure with momentum and adaptive learning rate. The hidden and output neurons are activated by the binary sigmoidal activation function. 60 samples are used for training the neural network and tested with 75 samples. The neural network training results are tabulated in TABLE I.

TABLE I
NEURAL NETWORK TRAINING RESULTS

Number of input neurons:60, Number of Hidden Neurons:16, Number of output neuron: 5
Activation Function: Binary sigmoidal function
Learning Rate: 0.1 Momentum Factor:0.9
Training Tolerance: 0.01 qh=1.0 qo=1.0
Testing Tolerence:0.5
Number of samples used for training: 60
No. samples used for testing: 75

Trial No.	Mean Classification Rate (%)		Number of Mean Epoch for Training	
	NN Model -I	NN Model -II	NN Model -I	NN Model -II
1	87	89	459	405
2	87	89	445	402
3	86	89	460	412
4	87	89	457	421
5	86	88	465	436
Mean	87	89	457	415

4 Conclusion

In this paper, a simple feature extraction algorithm has been presented using centroid and Discrete Cosine Transform (DCT) technique. Two neural network models have been developed for sign recognition from features computed from the video stream. From the experimental results, it was observed that the recognition rate of the proposed two neural network models is 87 and 89 respectively. The simulation result shows that the recognition rate of the system is over 80%. In future work, this work can be extended to helps primarily for deaf, deafened and hard of hearing people to solve their communication problems and gives them the same whole life experience to which their hearing colleagues have access.

Reference:

[1] T. Starner and A. Pentland, Real-time American Sign Language recognition from video using hidden markov models, International Symposium on Computer Vision, 1995, pp. 265-270

[2] T. Starner, J. Weaver and A. Pentland, Real-time american sign language recognition using desk and wearable computer based video, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 12, 1998, pp. 1371 - 1375.

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

[4] M. H. Yang, N. Ahuja and M. Tabb, Extraction of 2D motion trajectories and its application to hand gesture recognition, IEEE Trans on.Pattern Analysis and Machine Intelligence, Vol. 24, No. 8, 2002, pp. 1061-1074.

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

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

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

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

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

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

[11]. Qutaishat Munib, Moussa Habeeb, Bayan Takruri and Hiba Abed Al-Malik., American sign language (ASL) recognition based on Hough transform and neural networks, Expert Systems with Application, Vol.32, 2007, pp.24-37.

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