Available online at www.elixirpublishers.com (Elixir International Journal)

Advanced Engineering Informatics

Elixir Adv. Engg. Info. 33 (2011) 2308-2311

ISSN: 2229-712X

A static sign language recognition using discrete cosine transform and Ann Back propagation

P.V.V.Kishore[1] and P.Rajesh Kumar[2]

[1]Department of ECE, Vignan's Institute of Information Technology, Duvvada, Visakhapatnam, Andhra Pradesh, India-530046.

[2]Department of ECE, Andhra University College of Engineering, Visakhapatnam, Andhra Pradesh, India-530001.

ARTICLE INFO

Article history:

Received: 28 February 2011;

Received in revised form:

23 March 2011;

Accepted: 30 March 2011;

Keywords

Sign Language Recognition,

Neural Networks,

DCT,

Sobel and

Canny Edge Detector.

ABSTRACT

The paper presents a system developed for recognizing gestures of Indian sign language from images of gestures. The proposed system is based on Discrete Cosine Transform (DCT) and neural networks used for gesture pattern recognition. Unlike the systems proposed by other researchers such as using a radio frequency or colored gloves to achieve the recognition our system does not impose any such constraints. Features are extracted from the images using DCT which greatly reduces the size of the feature vector. Neural networks error back propagation algorithm is used to recognize gestures of alphabets of English language. The system was implemented with 130 sample images of gestures of alphanumeric characters with a maximum of 5 images per gesture. Experimental results show that the neural network is able to recognize gestures with an accuracy of 99.52%.

Introduction

Visual gesture language is employed by a large percentage of deaf populations to communicate among themselves. It has an extensive vocabulary, its own grammatical patterns, and its own rules of usage. For a common person to communicate with a deaf person, a translator is considered necessary to decode sign language into usual spoken language and vice versa.

The key element of many sign languages is to recognize hand shapes, signing area, hand movement, and head movement. When two people are observed in conversation of sign language, their bodies bend, facial expressions change constantly, and their hands seem to be constantly in motion. Some signs require no hand movement.

These signs are called stationery signs. Some static signs are expressed with one hand and other with two hands. A gesture is a form of non-verbal communication made with a part of the body, used instead of or in combination with verbal communication (Wikipedia).

Sign language recognition aspires to convert sign language into text or speech is an efficient and exact way. This can be accomplished by using digital image processing techniques and neural networks. In this paper are tried to recognize gesture of letters of English alphabets for Indian sign language. Recently sign language gestures recognition has gained a lot of interest by researchers in the areas of computational intelligence, neural networks and image pattern recognition.

Related work

Mohammed Mohandas et. Al [1] has proposed an image based systems for Arabic sign language recognition using hidden Markov Model for recognition; they have used a Gaussian skin color model to detect the signal face. The proposed system achieved a recognition accuracy of 98% for a date set of Arabic sign gestures. J.Han, G.Awad et.al.[2] proposed a video based framework for segmenting and tracking skin objects which consists of two parts: a skin color model and a skin object tracking system.Syed Atif Meohadi et. Al [3] examines the possibility of recognizing sign language gestures using sensor gloves. The paper describes the uses of a sensor gloves to capture ASL and uses NN to recognize these gestures. M.Shimada. et.al[4] has proposed an algorithm for figure spelling recognition of static and motion images. The recognition rate is 85.2v.

M.K.Bhuyar et.al [5] developed a framework for hand gesture recognition system based on object based video abstraction technique.

The experiment results show that the developed system can recognize signs of Indian sign language. Rini Akmelia et.al [6] proposed a real time Malaysian sign language translation using color segmentation technique which has a recognition rate of 90%.

Nariman Habili et.al [7] proposed a hand and face segmentation technique using color and motion cues for the content based representation of sign language video sequences.

In this research, a system is developed for to recognize all static gestures of Indian sign language.

The experiments are conducted on gestures of English alphabets by using computerized image processing techniques and Neural Networks. The developed system converts gesture signs of Indian sign language into text.

The key difficulty in such a system is to make the system sign independent, since different signs have different hand shapes, sizes, head sizes, body sizes etc. Our system uses the powerful discrete cosine transform to extract features from the images along with other image processing techniques.

The extracted features from the image form an input to the neural network for a parallel distributed processing. The recognized signs are successfully converted into text messages using MATLAB software.

Tele:

E-mail addresses: pvvkishore@gmail.com,rajeshauce@gmail.com

Our approach to system design

The system design has four phases: pre-processing, segmentation, feature extraction and pattern recognition as shown in the figure 1.

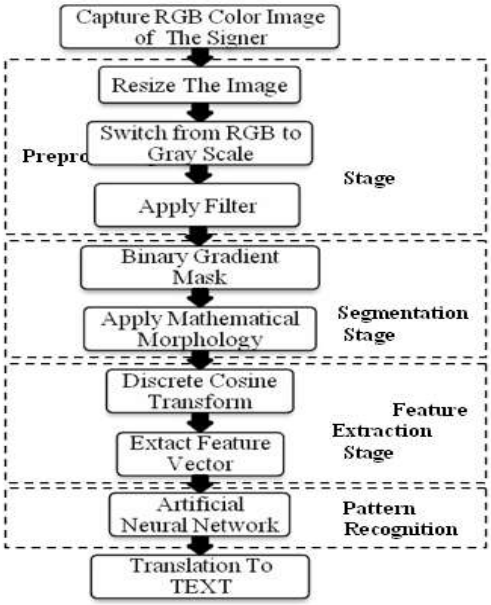

Figure 1: Block Diagram Representation of the System

Images are acquired using a two mega pixel cell phone camera that captures RGB color images of size 1024×1024 pixels. Higher resolution causes longer delays in the image acquisition process and higher execution times. The images of gestures are acquired under normal lighting conditions to simulate real time environments.

The illumination level on the person is kept between 12-15 lux using an electric white bulb of power 15 watts during the entire image acquisition process.

The only constraint in this system is the background of the image and clothes of the person in the image should be of dark color. For this we have chosen a dark blue color background and the person in the image is also wearing a dark blue color full sleeves shirt.

system architecture

The pre-processing stage consists of image resize, RGB (Red, Blue and Green) to gray scale conversion and filtering.

The acquired image is resized to 256×256 by using nearest neighbourhood interpolation method.

Pre-processing Stage

To make the system more robust we bring the idea of capturing the images under different lightning conditions.

Two different color spaces are used that is RGB and HSV (Hue, Saturation and Value) which gives color according to human perception to demonstrate the problem with lightening conditions. Binary images consisting two levels are created from both RGB and HSV images as shown in figure 2.

It is shown from figure 2, converting an image from RGB to gray scale and then to binary results in spiky and unambiguous details for the image in our project.

It is observable that the RGB color to HSV color space conversion then to binary image produced image that is deficient in many features of the sign.

Further the gray scale image is filtered using smoothing or averaging filter to remove noise present in the image added during image acquisition process.

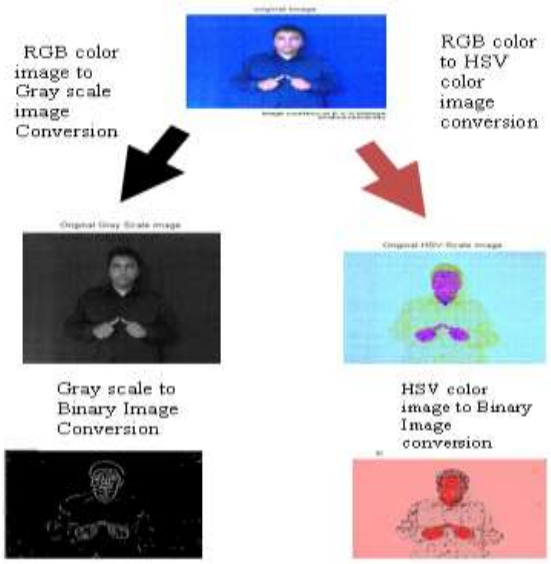

Figure 2: Variation in image conversion techniques for converting a RGB color image to Binary image.

Segmentation Stage

The objective of image segmentation is to subdivide the image into constituent head and hand objects. This is accomplished using edge detection methods and mathematical morphology techniques such as dilation and erosion. A good quality edge detector should mark as many real edges as possible. We used sobel edge detector since it provided a finest edge detection solution. Sobel edge detector is a superior edge detector compared to canny edge detector as shown in the figure 3.

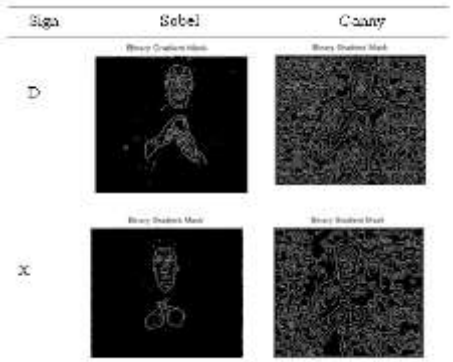

Figure 3: Choosing suitable Edge Detection Technique

As shown in the figure 3 sobel edge detector is suitable for our task of recognition signs from the image as it marks numerous details to make a good feature vector. While the canny edge detector marks large number of false edges and the sign information is lost.

Most of the times the sobel edge detector marks extra details than required which can be eliminated by using appropriate threshold .After testing with diverse thresholds from 0.4 to 0.9 we determined to use 0.9 after considering the results shown in figure 4.

The linear gaps will disappear if the sobel image is dilated using horizontal and vertical structural elements. Unwanted gaps are filled with holes and borders are cleared. In order to make the segmented object look natural we smoothen the object by eroding the image with a diamond structuring element. This segmented image is used to extract features to design a feature vector for the neural network.

Sobel edge With out Threshold | Sobel edge with 0.5 Threshold | Sobel edge with 0.9 Threshold

Figure 4: Sobel edge detector with and without threshold

Feature Extraction Stage

In feature extraction we generally search for invariance properties so that the extraction process does not change with the specified conditions. The technique should find shape consistently whatever may be the value of the parameter that controls shape. For this work we choose not to vary the background of the image and only changes in shapes happen in the hands orientation and head movement. Here we extract head and hand shapes which vary according to the sign gesture.

Our algorithm is based on studying each images areas and extracting shape features to identify the sign. For this process we have chosen very popular image transform DCT which is a real transform that transforms a set of real data points from the image onto a real spectrum. DCT coefficients are calculated for the binary image as shown in the figure 5. From the figure 5 we were able to clearly spot that most of the energy is concentrated in the top left corner.

Since most of the energy is concentrated in the top left corner after many trials we have decided to consider the first horizontal row of the DCT matrix. For our paper we have considered 26 images of English alphabets. The size of the feature vector is 26×2. The feature vector is transformed to binary matrix of same size which is given as input to the neural network.

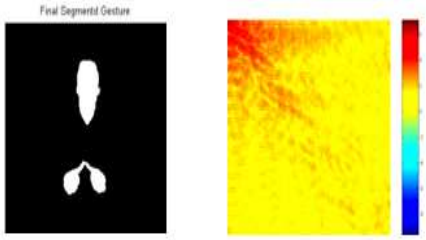

Figure 5: DCT of a Gesture Sign 'A'

Pattern Recognition Stage

We have employed an artificial neural network to accomplish the task of recognizing and classifying gesture signs. The neural network has 130 neurons in the input and output layers along with 256 neurons in its hidden layer. This particular neural network object can take 130 input images and classify 130 output images successfully. The proposed system is based on supervised learning which employs error back propagation learning algorithm. This algorithm is a iterative gradient based algorithm proposed to minimize an error between actual output vector and desired output vector. The size of our target matrix is 130×2.

The neural network object created is feed forward back propagation network. The weights and bias for each neuron are initialized randomly and network is ready for training. The training process requires a set of examples of proper network behaviour, network inputs and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network performance function which in case of feed forward networks is mean square error.

The training function used is gradient descent with momentum and adaptive learning rate which updates the weights and bias of neurons during training. The number of epochs used for training is 15000. The system was tested with 104 images previously unseen by the network in the testing phase. The network is tested more than once during testing phase.

Results and analysis

This section presents the results of the experiments performed using a personal computer to classify gestures of sign language. The training set of data consists of 26 gray scale images of size 256x256. Also 5 samples of each sign were taken under different conditions and small changes in orientation with a total of 130 images of signs of 26 English alphabets. Of the 5 signs for each alphabet 3 are used for training and remaining 2 are used for testing along with images used in the training phase. We trained the system for three samples for each sign with a total number of 78 images. The mean squared error tolerance is fixed at 0.0001 for training the samples.

The learning rate and momentum factor are chosen as 0.25 and 0.9. The hidden and output neurons are activated using hyperbolic tangential sigmoid transfer function. A total of 78 samples are used for training the network and 130 samples are used for testing the network. The mean square error versus epoch graph is shown in figure 6.

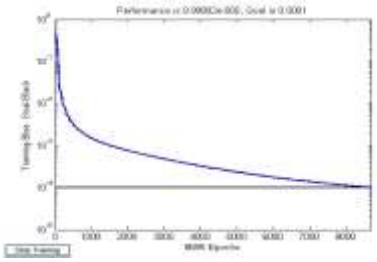

Figure 6: Mean Square versus Epoch graph

The tested results for both training and testing the neural network are shown in table 1.

The Proposed system was able to reach a recognition rate of 99.52%.

Conclusion and future work

In this paper a straightforward system for sign language recognition system was developed. The segmentation algorithm involving edge detection and mathematical morphology is used. The features are extracted using DCT and the gesture patterns are recognized using feed forward neural network. In future this work can be used in real time environments to help hearing impaired people.

References

[1] Syed Atif Meohadi and Mohamed Deriche,Image based Arabic sign language recognition, IEEE, 2005.

[2] J. Han, G. Awad, A. Sutherland, Automatic skin segmentation and tracking in sign language recognition IET Computer. Vision. 2009, Vol. 3, Iss. 1, pp. 24–35

[3] Syed AtifMehdi, Yasir Niaz Khan, sign language recognition using sensor gloves, Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 5.

[4] M. Shimada, S. Iwasaki , and T. Asakura, Finger Spelling Recognition using Neural Network with Pattern.

[5] Recognition Model, SICE Annual Conference in Fukui, August 4-6.2003.
[6] M.K Bhuyan, D. Ghoah and P.K. Bora, 'A Framework for Hand Gesture Recognition with Applications to Sign Language', Annual India Conference 2006 , pp. 1-6.
[7] 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 2006 ,pp. 1-6.
[8] Nariman Habili, Cheng Chew Lim and Alireza Moini , '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 2004 , Vol. 14, No. 8, , pp.1086 – 1097
[9] S.N. Sivanandam, M. Paulraj An Introduction to Neural Networks, 2003, Vikhas Publications Company Ltd.India.

TABLE I

RESULTS OF TRAINING AND TESTING OF NEURAL NETWORK

Number of input neurons : 130 Number of hidden neurons: 256 Number of output neurons: 26 Activation Function: Hyperbolic Tangential sigmoid Learning rate: 0.25 Momentum Factor: 0.9 Error Tolerance: 0.0001 Number of training samples used: 78 Number of testing samples: 130			
Data	Total Number of Samples	Correctly Recognized samples	Performance Rate of the Neural Network (%)
Training	78	78	100%
Testing	130	129	99.21%
Total	208	207	99.52%