



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

Neural Network Based Static Sign Gesture Recognition System

Parul Chaudhary¹, Hardeep Singh Ryait²

M.Tech Scholar, Department of ECE, Baba Banda Singh Bahadur Engineering College, Punjab, India¹

H.O.D, Department of ECE, Baba Banda Singh Bahadur Engineering College, Punjab, India²

ABSTRACT: Sign language is natural media of communication for the hearing and speech impaired all over the world. This paper presents vision based static sign gesture recognition system using neural network. This system enables deaf people to interact easily and efficiently with normal people. The system firstly convert images of static gestures of American Sign Language into Lab color space where L for lightness and (a, b) for the color-opponent dimensions, from which skin region i.e. hand is segmented using thresholding technique. The region of interest (hand) is cropped and converted into binary image for feature extraction. Then height, area, centroid, and distance of the centroid from the origin (top-left corner) of the image are used as features. Finally each set of feature vector is used to train a used to train a feed-forward back propagation network. Experimental results showed successful recognition of static sign gestures with an average recognition accuracy of 85 % on a typical set of test images.

KEYWORDS: American Sign Language, Binary image, Feed-forward back propagation network, Lab color space, Thresholding technique.

I. INTRODUCTION

COMMUNICATION is a fundamental requirement for survival and interaction provides a mean to communicate. Naturally, different communication ways are used for interaction such as language, eyes, body movement, facial expression, hand gesture and postures. A gesture is a form of non-verbal communication made with a part of the body and used instead of verbal communication (or in combination with it). A sign language is a language which uses gestures instead of sound to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. Sign language is a visual language and consists of 3 major components [2]: finger-spelling: used to spell words letter by letter, word level sign vocabulary: used for the majority of communication, non-manual features: facial expressions and tongue, mouth and body position. Sign language is one form of communication for the hearing and speech impaired.

Sign language recognition (SLR) is a multidisciplinary research area involving pattern recognition, computer vision, natural language processing and linguistics [1]. Moreover for Human Computer Interaction (HCI), as compared to the traditional interaction approaches such as keyboard, mouse, pen etc, and vision based hand interaction is more natural and efficient

Hand gestures fall into two categories, namely static and dynamic [3]. Some hand gestures also have both static and dynamic elements, as in sign languages [8]. Static hand gestures are characterized by the hand posture which are determined by a particular finger thumb-palm configuration and represented by a single image. Dynamic hand gestures are on the other hand characterized by the initial and final stroke motion of a moving gesture.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

II. RELATED WORK

Research showed various types of system and methods have been developed for sign language recognition. Acquired data in recognition system can be obtained either “Data-Glove based” or “Vision Based” approaches. The Data-Glove based methods use sensor devices for digitizing hand and finger motions into multi-parametric data [14]. These approaches can easily provide exact coordinates of palm and finger’s location and orientation, and hand configurations. However, the devices are quite expensive and bring much cumbersome experience to the users. In contrast, the Vision based methods require only a camera, thus considered easy, natural and less costly compared to glove based approach [7].

A vision based system was presented by Hienz et al [5] in which extracted feature vectors from video frame, recognize 262 different signs with an accuracy of 94%. Rule based classification was done on the images captured by single video camera by use of modular frame grabber system. Yin et.al [17] employed Restricted Coulomb Energy (RCE) neural network by taking L^*a*b color space as input and trained the output layer as skin layer class. Ranganath et.al [13] presented a hand gesture recognition system, they used image furrier descriptor as their prime feature and classified with the help of RBF. Ong et.al [12] detected hands with 99.8 percent accuracy in grey scale images with shape information alone; using a boosted cascade of classifiers Viola et.al [15] .Signers were constrained to wear long-sleeved dark clothing, in front of mostly dark backgrounds. Hemayed et.al [4] developed a edge based recognizer for Arabic sign language which uses Prewitt edge detector to extract the edges of the segmented hand gesture. Accuracy of 97% was achieved using Euclidean distance for classification. Murthy et.al [11] trained a supervised feed-forward neural network to count fingers and find direction in which user point. The vision based recognition system classified hand gestures into ten categories employing back propagation algorithm with an accuracy of 89% on a typical test set.

The paper presents a simple yet an efficient recognition system which will convert the static sign gestures of American Sign Language into text. Geometrical properties of the hand are transformed into features. Neural network is used for recognition and classification task. The rest of the paper will discuss different phases for the development of the system: image acquisition, image processing, feature extraction and classification in detail.

III. PROPOSED METHODOLOGY

The system is designed on the principle of pattern recognition. Pattern recognition is a process that takes raw data and makes an action based on the category of the pattern. Pattern recognition can be used for classification in which assign each input value is assigned to one of a given set of classes. The flow diagram of the system is shown in Fig. 1.

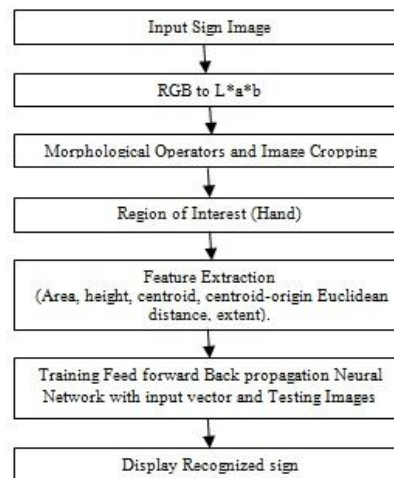


Fig. 1: Flow diagram for sign recognition system



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

A. Image Acquisition

In the first phase an image is taken from the webcam or from database. The system read input image from the database [6] which contain RGB images for ASL signs. The database contains samples of four signs performed by different users wearing long-sleeved clothing taken. The images are with uniform background of dark and light color under different lightening conditions. The image database consists of total of 160 images. 120 images: 40 images for each sign are used for training purpose while remaining 40 images: 10 images for each sign are used for training.

B. Image Processing

Processing is performed using three steps: color space conversion for skin area extraction, morphological operations to remove noise and image cropping for ease of feature extraction.

1) *Skin region detection*: The system used L^*a^*b color space for skin region detection using thresholding technique. L^*a^*b is Color space defined by the CIE (the International Commission on Illumination), based on one channel for Luminance (lightness) (L) and two color channels (a and b). Input RGB image is firstly converted to $L^*a^*b^*$ Color space to separate intensity information into a single plane of the image, and then calculates the local range in each layer. Skin color classification work well when chrominance component used for segmentation, therefore luminance component is discarded. Using thresholding values, second and third layer which represent the chroma component are converted into binary image. Two binary images are then multiplied to get resultant binary image which contain only hand region. Various morphological operations like opening, dilation were performed to remove the noise from segmented hand region.

2) *Image cropping*: Using connected component analysis the connected regions of the resultant image are labelled. Each connected component associates a bounding box with it which provides dimensions for the rectangular box for cropping hand region from input image.

C. Feature Extraction

The features are extracted on the shape based properties of the hand. Tanibata et.al [16] developed Japanese Sign Language using hand features like orientation, area, the flatness of the hand region. Mohammed et.al [9] proposed method of hand feature extraction using several geometrical dimensions like height, width, area etc. Using regionprops following image features are extracted and feature vector is formed which will act as input stage for the recognition and classification of the sign gesture.

1. *Area*: Calculated as the total number of white pixels (i.e., binary value '1').

2. *Centroid* = [Round ($\Sigma(x$ -values represent white pixels)/area), Round ($\Sigma(y$ -values represent white pixels)/area)].

$$X_c = \Sigma x_i / \text{Area}$$

Where x_i represent the X-coordinate of each boundary pixel of the image.

$$Y_c = \Sigma y_i / \text{Area}$$

Where y_i represent the Y-coordinate of each boundary pixel of the image.

The centroid of the image is (X_c, Y_c).

3. *Euclidian Distance of the Centroid from the Origin*. Given the two points (x_1, y_1) and (x_2, y_2), the Euclidian distance between these points is given by the formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The coordinates of the left-top most pixel of binary image is considered as the origin. If origin of image is (0, 0) then Euclidian distance between centroid and the origin is given by



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

$$d = \sqrt{X_c^2 + Y_c^2}$$

4. *Average Height of a Sign:* Height of a sign from binary image where 0's represent black pixels and 1's represent white pixels is calculated using pixel scanning method [10].

The average height of the binary image is calculated as

$$\text{Average Height} = N/M.$$

Where N is the total number of black pixels (0's) of the image and M is the total number of columns containing at least one black pixel (0).

Finally, the features collected from the above sections are combined to form a feature vector in the following order:

Feature vector, $V = [\text{area, x-centroid, y-centroid, centroid-distance, Average height}]$

Feature vectors of training images are stored in mat files of MATLAB and feature vector of input hand gesture image i.e. test image are calculated at run time.

D. Classification of Sign Gestures using Neural Network

The system uses feed-forward back propagation network for classification of sign gestures. Back propagation training algorithm is a supervised learning algorithm for multilayer feed forward neural network. Since it is a supervised learning algorithm, both input and target output vectors are provided for training the network. If there is an error, the Perceptron network will re-adjust the weights value until there is no error or minimized and then it will stop. Each pass through the input vectors is called epoch.

Input vector is the 1x5 feature vector so only five input neurons are used. The target vector is also defined corresponding to each hand gesture. The performance of the training is evaluated with MSE, correlation coefficient, i.e. regression (R) between the network outputs and corresponding target outputs and the characteristics of the training, validation, and testing errors. For the successful training, some of the conditions are set. These are the MSE set to 0.001, maximum validation failure set to 6 times, learning rate set to 0.05 and the maximum number of epochs set to 1000. Sim function is used to simulate the model. Finally output from the neural network is converted into text corresponding to each classified hand gesture. The training is stopped after 21 epochs since the validation error increased for more than six times as shown in Fig 2. The training, validation, and testing errors were in fairly good conditions with the characteristics set during training.

IV. EXPERIMENTAL RESULTS

For the implementation of the proposed system, image database is created for training and testing images. The image database consists of four static sign gestures of ASL in '.jpg format'.

The method is implemented using MATLAB R2012a. Skin region detection, image cropping resizing and feature extraction is performed using Image Processing Toolbox. Neural network toolbox is employed for classification of hand gestures. The MATLAB built in function (sim) simulates network. The behaviour of (sim) takes the network input, network object, then returning the network output. Trained neural network is tested with Test image database.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

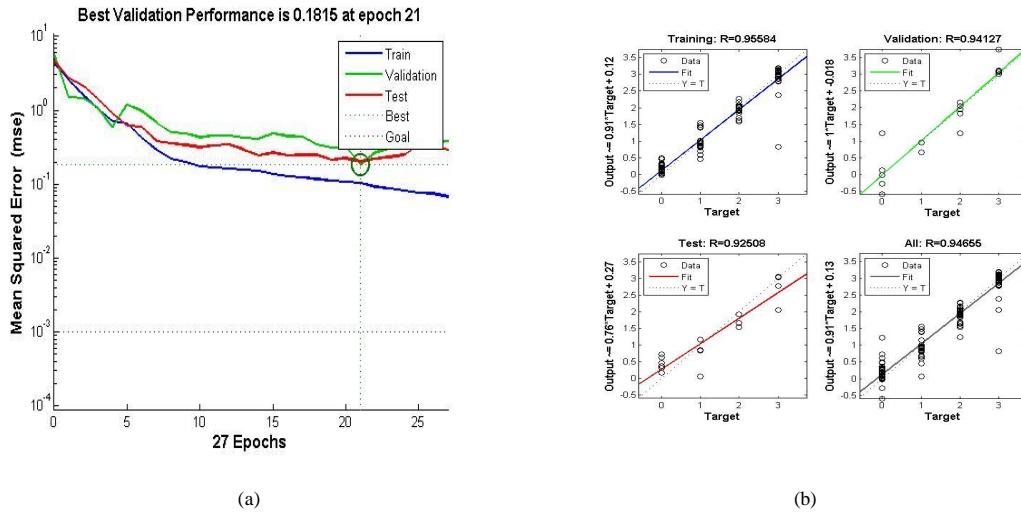
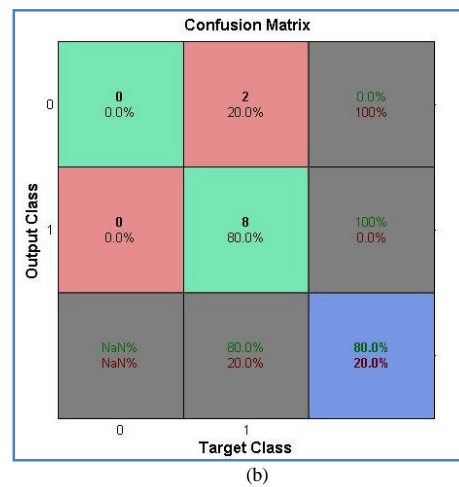
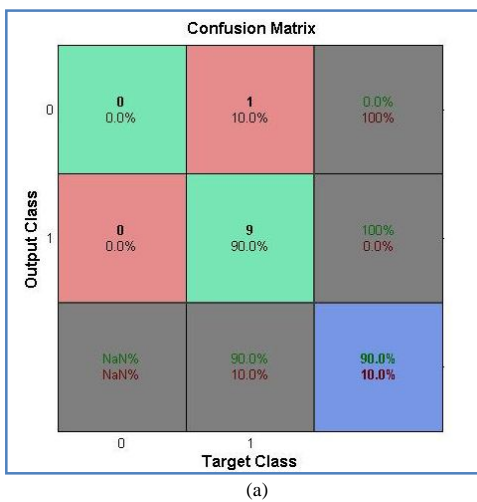


Fig.2 (a) Training Validation and Test errors during training with Levenberg-Marquardt algorithm (b) Neural network Regression plot

A. Confusion Matrix

Confusion matrix has been plotted to show the recognition accuracy for each hand gesture as shown in Fig.3. The green box show number of images correctly classified and blue box show overall recognition rate for test images dataset for particular hand gesture.

$$\text{Recognition Rate} = (\text{No of Recognized Letters} / \text{No of total samples of that Letters}) * 100\%.$$



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

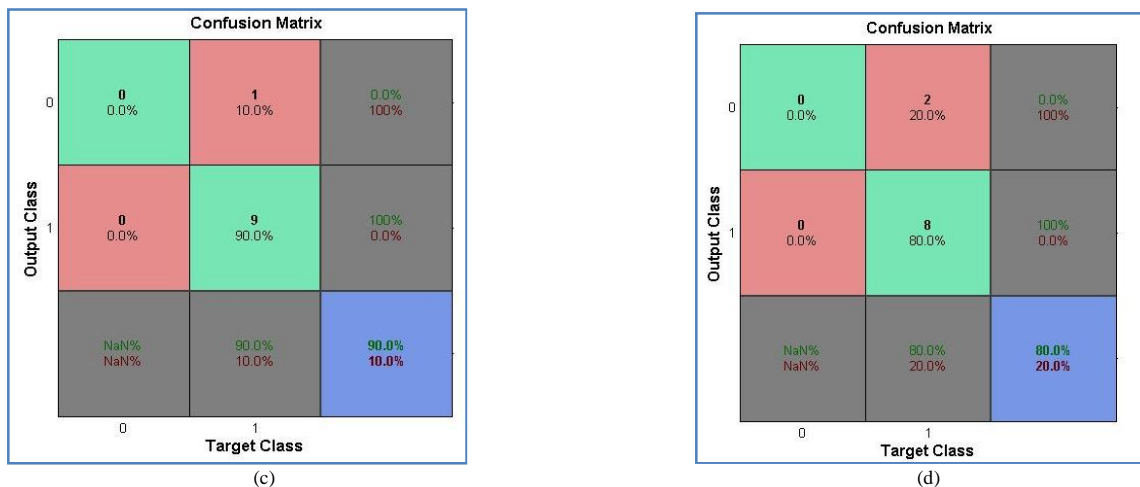


Fig.3 Confusion matrix (a) V (b) B (c) C (d) Point

B. GUI Simulating Results (sign to text).

For testing the images GUI (Graphical User Interface) has been created. The GUI consists of nine push buttons. GUI provides an easy way to the user, to interact with the system. The GUI shows different phases of the system: input image, hand region detection, feature extraction in which centroid of the hand is shown and finally the recognized character using neural network. An example of GUI is shown in Fig. 4 that shows different phases of system involved for sign 'V'

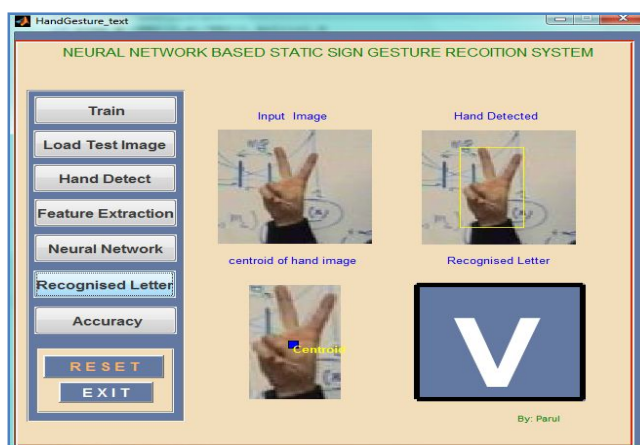


Fig.4 Example on GUI simulating sign 'V'

V. CONCLUSION

The system developed presents a simple yet an efficient method of gesture recognition using geometrical features based on the shape based properties of hand. Static hand gestures are recognized using neural network and converted into corresponding text each. The recognition rate of 85% is achieved for testing image. In future the system can be extended to recognize dynamic hand gestures in an unrestricted environment for real life applications.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

REFERENCES

1. Aran, O. (2008), 'Vision based sign language recognition: modeling and recognizing isolated signs with manual and non-manual components' (Doctoral dissertation, Bogaziçi University).
2. Bowden¹², R., Zisserman, A., Kadir, T., and Brady, M. (2003). 'Vision based interpretation of natural sign languages'.
3. Cutler. R, and Turk, M.(1998), 'View based interpretation of real time optical flow for gesture recognition'. IEEE International Conference on Automatic Face and Gesture Recognition.
4. Hemayed, E. E., and Hassanien, A. S. (2010, December), 'Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)'. In *Computer Engineering Conference (ICENCO), 2010 International* (pp. 121-127) IEEE.
5. Hienz, H., Grobel, K., and Offner, G.(1996) 'Real-Time hand-Arm Motion Analysis using a single video camera', *proc. International conference on Automatic Face and Gesture recognition*, pp. 323-327.
6. <https://sites.google.com/site/autosignlan/source/image-data-set>
7. Khan, R. Z., & Ibraheem, N. A. (2012), 'COMPARATIVE STUDY OF HAND GESTURE RECOGNITION SYSTEM'. *SIPM, FCST, ITCA, WSE, ACSIT, CS & IT*, 6, 203-213.
8. Mitra, S., and Acharya, T. (2007), 'Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*', *IEEE Transactions on*, 37(3), 311-324.
9. Mohammed, F., Mohammad, W., Kayes, and A.S.M., Poya, A. (2013), 'Geometric feature Extraction of Human Hand', *International Journal of Computer and Information Technology (SSN:2279-0764)*, Volume 02-Issue 04
10. Md. Aktaruzzaman, Md. Farukuzzaman Khan, and Professor Dr M. Ekin Uddin. (2009), 'Recognition of Offline Cursive Begali Handwritten Numerals using ANN' *Journal of the Peoples University of Bangladesh*, vol. 4, N0.1, July 2009, pp18-28, Bangladesh, ISSN 1812-4747
11. Murthy, G. R. S., and Jadon, R. S. (2010, February), 'Hand gesture recognition using neural networks'. In *Advance Computing Conference (IACC), 2010 IEEE 2nd International* (pp. 134-138). IEEE
12. Ong, E-J, Bowden, R., (2004), 'A Boosted Classifier Tree for Hand Shape Detection', *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 889-894.
13. Ranganath, S., Ng, C W (2002), 'Real-time gesture recognition system and application', *Image Vision Computer*, 20(13-14): 993-1007.
14. Sturman, D. J., & Zeltzer, D. (1994), 'A survey of glove-based input'. *Computer Graphics and Applications*, IEEE, 14(1), 30-39.
15. Viola, P, and Jones, M., (2002), 'Robust Real-Time Object Detection', *Proc. IEEE Workshop Statistical and Computational Theories of Vision*.
16. Tanibata, N., Shimada, N., and Shirai, Y. (2002), 'Extraction of Hand Features for recognition of Sign Language Words', *Proc. International Conference on Vision Interface*.
17. Yin, X. and Xie, M. (2001), 'Hand Gesture Segmentation, Recognition and Application' *Proc. Of 2001 IEEE International Symposium on Computational Intelligence in Robotics and Animation*.