VOL. 9, NO. 8, AUGUST 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

INDIAN SIGN LANGUAGE RECOGNITION USING NEURAL NETWORKS AND KNN CLASSIFIERS

Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo
Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University, Greater Noida, India
E-Mail: ashoksahoo2000@yahoo.com

ABSTRACT

Sign language recognition is helpful in communication between signing people and non-signing people. Various research projects are in progress on different sign language recognition systems worldwide. The research is limited to a particular country as there are country wide variations available. The idea of this project is to design a system that can interpret the Indian sign language in the domain of numerals accurately so that the less fortunate people will be able to communicate with the outside world without need of an interpreter in public places like railway stations, banks, etc. The research presented here describes a system for automatic recognition of Indian sign language of numeric signs which are in the form of isolated images, in which only a regular camera was used to acquire the signs. To use the project in real environment, first we created a numeric sign database containing 5000 signs, 500 images per numeral sign. Direct pixel value and hierarchical centroid techniques are used to extract desired features from sign images. After extracting features from images, neural network and kNN classification techniques were used to classify the signs. The result of these experiments is achieved up to 97.10% accuracy.

Keywords: Indian Sign Language, Feature Extraction, kNN classifier, Neural Networks.

INTRODUCTION

Sign language (SL) is a visual-gestural language used by deaf and hard-hearing people. They use three dimensional spaces and the hand movements (and other parts of the body) to convey meanings. It has its own vocabulary and syntax which is entirely different from spoken and written languages. A gesture may be defined as a movement, usually of hand or face that expresses an idea, sentiment or emotion e.g. rising of eyebrows, shrugging of shoulders is some of the gestures we use in our day to day life. Sign language is a more organized and defined way of communication in which every word or alphabet is assigned to a particular gesture. With the rapid advancements in technology, the use of computers in our daily life has increased manifolds. Our aim is to design a Human Computer Interface (HCI) system that can understand the sign language accurately so that the signing people may communicate with the non signing people without the need of an interpreter [1]. Around 5% of world community in all parts of the world is using sign language as a medium of communication [2]. Regionally different languages have been evolved as ASL (American Sign Language) in America, GSL (German Sign Language) in Germany, BSL (British Sign Language) in the UK or ISL (Indian Sign Language) in India [1]. There are mainly two different motivations for developing sign language recognition model. The first aspect is the development of an assistive system for the deaf or hard hearing people. For example development of a natural input device for creating sign language documents would make such documents more readable for deaf people. Moreover hearing people have difficulties in learning sign language and likewise the majority of those people who were born deaf or who became deaf early in life, have only a limited vocabulary of accordant spoken language of the community in which they live. Hence a system of translating sign language to spoken language would be of great help for deaf as well as for hearing people. A second aspect is that sign language recognition serves as a good basis for the development of gestural human-machine interfaces [3].

Sign gesture can be divided into two types: static and dynamic. Static gestures have fixed position of hand whereas dynamic gestures have movement of hands and body parts.

Sign Language Recognition is the machine recognition of gestures. Gesture recognition can be done in either way, Device based approach or Vision based approach. The later one is commonly used in pattern recognition. There is no common way of recognition of sign language gestures, so a recognition system is to be formalized. Till now; there is no translator or machine available to help the unblessed community in public places in India. We are in the process of developing a recognition system to help the unblessed community.

RELATED WORK

Balakrishnan, G., P. S. Rajam, *et al*., [4] proposed a method of recognizing a 32 set of combinations, 10 for each up and down position of fingers to get corresponding Tamil letters. The method was used up/down position of fingers into decimal numbers which is further categorized to recognize the Tamil alphabets. A set of static data in the form of images of sizes 640×480 pixels were captured. Palm image extraction was used to convert RGB to Greyscale images. The experimental result was 96.87% recognition rate for static method and 98.75% accuracy rate reported for dynamic method.

Ghotkar, A. S., *et al*., [5] in their work described four modules for recognition of sign languages. First method used was the hand tracking system in which CAM-SHIFT method was employed. Second was the hand

segmentation in which HSV color model and neural network. Third one was used for feature extraction in which Generic Fourier Descriptor (GFD) method was used. Fourth was the gesture recognition in which Genetic algorithm. The authors did not describe the database and no result was reported used in their work.

Futane, P. R., *et al*., [6] in their work proposed two approaches, one for device based and the second is vision based for Indian sign language recognition. In the device based approach they used an electronic glove for sign input. In the vision based approach two processes involved, first one was used for sign capturing and second one was used for sign analysis. They used three modules for capturing and analysis of sign symbols. First one was used for Gestures Acquisition, pre-processing and creation of gestures database. Second module was on Gestures classification and Key features extraction and analysis. Third one they used for Building Gestures recognition engine. Neither any database nor any result was discussed in this paper.

Deora, D. and N. Bajaj [7] developed an Indian Sign Recognition System for twenty five English alphabets (double handed signs) and for nine numeral signs. The signers used for data acquisition were required to wear blue and red gloves. Segmentation and finger tip algorithm for feature extraction and PCA for classification of signs were employed in this research. The overall recognition rate reported was 94%.

Rekha, J., *et al*., [8] proposed an approach to recognize ISL double handed static and dynamic alphabet signs. 23 static ISL alphabet signs from 40 signers were collected as training samples and 22 videos were used as testing samples. The shape features were extracted by the method of Principle Curvature Based Region Detector, texture features of hand were extracted by Wavelet Packet Decomposition and features from fingers were extracted by complexity defects algorithms. Multi class non-linear SVM, KNN and DTW were used as sign classification. The recognition rate achieved were 91.3% for static signs and 86.3% for dynamic signs.

Kar, M. K., D. R. Neog, M. K. Bhuyan, *et al*., [9] in their work proposed a novel approach for hand pose recognition for analyzing the textures and key geometrical features of the hand. For hand segmentation Bayesian classifier method was used. For feature extraction, the Gaussian distribution method and for proximity measurement distance measure was used. Tracking of fingers was computational expensive so they proposed abduction and adduction movement of fingers and use Homogeneous Texture Descriptor (HTD) for inflexible variation of fingers. The overall recognition rate was 93.38% for all eight gestural patterns.

After studied related works carried by various authors we found many limitations in these works. The first and foremost drawback is the database. No standard database is available for research work on ISL. The second is that in all the databases were developed in some laboratory. The recognition system we planned is for use in public places where different background (noises) may be available in acquiring sign images. Some authors used various methods for classification and feature extraction which are not clear and they were unable to justify the usage. In these research papers the authors developed some laboratory based systems which is either high cost or needs more processing power. The systems are not portable for use in mobile or hand held devices.

THE PROPOSED SYSTEM

The paper is organized in the following order.

a) Database collection/Image acquisition
b) Image pre-processing
c) Feature extraction
d) Classification
e) Results and discussion
f) Conclusion
g) Future work

DATABASE COLLECTION

Dark background for acquiring the images is selected. The selection of this background is due to uniformity in background and its pixel values in capturing features, also it is helpful in deleting background in order to extracting important features. A Sony digital camera was used for image acquisition. The camera was capable of capturing images up to 16.1 Mega Pixel with 10X Optical Zoom. The common file format JPEG was used to capture the images. There are 5000 images for database. Each image is 4608*3456 pixel size and a 5.5 MB size approximately. In order to create an efficient database with reasonable size, the images were cropped to 200×300 RGB pixels and each image requires 25 KB memory. The database was collected from 100 users. Out of these 100 users, 69 were male and 31 were female. The database was of numerical signs for 0, 1, 2...9 contains 10 signs (Figure-1). Each signer contributed five samples of each individual sign. So a total 5000 signs were collected. Then database was divided into two groups, one set was used for training and other was used for testing.

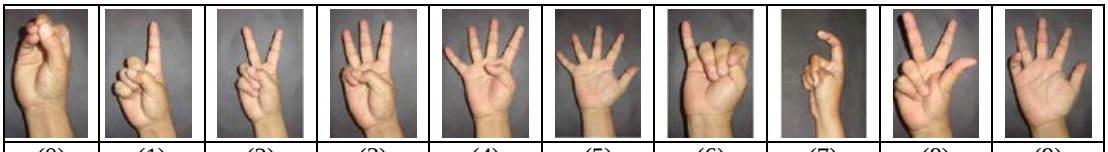

Figure-1. ISL static signs for isolated numerals.

IMAGE PRE-PROCESSING

After collecting the database from users, the images were pre-processed. Firstly the RGB images were converted to gray scale image by rgb2gray function available in Matlab environment. It converts the true color image RGB to the gray scale intensity image. The function converts RGB images to gray scale by eliminating the hue and saturation information while retaining the luminance [12]. We used first derivative Sobel edge detector method because it computes gradient by using discrete difference between rows and columns of 3×3 neighbors. The Sobel method finds edges using the Sobel approximation to the derivative. Where the gradient of image is maximum, Sobel returns edge points. Sobel is the best in amongst because it provides good edges, and it performs reasonably well in the presence of noise.

FEATURE EXTRACTION

Feature extraction [1] is a form of dimensionality reduction. Input images are too large for processing, so to process these images in time we reduce the dimension of the input image by feature extraction. Transforming input data into feature is called feature extraction. Feature extraction is chosen in such a way that image information must be retained. Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. It is often decomposed into feature construction and feature selection.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a *classification* algorithm which *over fits* the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Feature extraction techniques used in this project are direct pixel value and hierarchical centroid.

In direct pixel value feature extraction method, original image (200×300) was resized to 20×30 pixels and then the image matrix was converted into one dimensional array containing exactly 600 elements.

Hierarchical centroid [10], which uses the centroid method of finding the centroid of the image, through the centroid image is partitioned into two different zones, left and right zones. Iteratively this method was performed at most seven times then we get 124 features out the image. The end result of the extraction task is set of features, commonly called a feature vector and feature vector constitutes a representation of the image.

CLASSIFICATION

Extracted features are needed as input for classification. Classification techniques are helpful to recognize the gestures. There are a number of classification techniques available [1]. Classification is identifying inputs to a set of class on basis of training data set. The scope of image classification is to identify and convey meaning. Image classification is the most important part of digital image analysis. Classification techniques generally aim to provide a reasonable answer for all possible inputs and to perform "most likely" matching of the inputs, taking into account their statistical variation. In our work, kNN and Neural Network Pattern recognition tools ware used in recognizing the numeral gestures of ISL.

K-Nearest Neighbor (kNN)

K-nearest neighbor (kNN) classifier classifies objects on the basis of feature space. kNN uses supervised learning algorithm. Nearest neighbor algorithm is most popular classification technique proposed by Fix and Hodges. kNN classify method classifies each row of the data in sample into one of the groups in training using the nearest-neighbor method [11]. SAMPLE and TRAINING must have same number of columns. Group is a grouping variable for training and unique values define groups. Each element defines the group that the corresponding row of training belongs. Group can be a numeric vector, a string array, or a cell array of strings in the Matlab environment. Training and group have same number of rows. Class indicates group of each row of sample that it has been assigned to, and is of the same type as group. The default behavior of this method is to use majority rule i.e., a sample point is assigned to the class from which the b of the K nearest neighbors are from. When classify more than two groups and 'k' is even number then there is a tie so to break tie random is used to break tie in nearest neighbor. The default is majority rule for nearest tie breaker.

Neural network pattern recognition tool (NPR Tool)

NPR Tool leads you through solving a pattern-recognition classification problem using a two-layer feed-forward patternnet network with sigmoid output neurons. Twenty five neurons were used on the hidden layer. The network has ten output neurons because there are ten target values associated with each vector. Pattern recognition networks are feed forward networks that can be trained to classify inputs according to target classes. The target data for pattern recognition networks should consist of vectors of all zero values except for a 1 in element i, where i is the class they are to represent.
In this method target matrix is set to as follows:

Class [0-9]	**Target Matrix**
0	[1 0 0 0 0 0 0 0 0 0]
1	[0 1 0 0 0 0 0 0 0 0]
2	[0 0 1 0 0 0 0 0 0 0]
3	[0 0 0 1 0 0 0 0 0 0]
4	[0 0 0 0 1 0 0 0 0 0]
5	[0 0 0 0 0 1 0 0 0 0]
6	[0 0 0 0 0 0 1 0 0 0]
7	[0 0 0 0 0 0 0 1 0 0]
8	[0 0 0 0 0 0 0 0 1 0]
9	[0 0 0 0 0 0 0 0 0 1]

For the purpose of training the input data is divided into three sets: Training, Validation and Testing.

Training data was used for adjusting the weight and biases. Validation was used to decide when to stop the training process, to avoid over fitting which is a situation where the network memorizes the training data, rather than training the low that governs them. Testing data was used to measure the performance of the trained network. It is important that this data do not participate in the training process.

In this method training, 70% of the original data, 15% were used to validate that the network is generalizing and to stop training before over fitting and for testing, 15% were used as a completely independent test of network generalization.

The training algorithm 'trainscg' was used. This is a network training function that updates weight and bias values according to the scaled conjugate gradient method. The training parameters for TrainSCG are:

epochs	**Maximum number of epochs to train**
Show	Epochs between showing progress
Goal	Performance goal
Time	Maximum time to train in seconds
min_grad	Minimum performance gradient
max_fail	Maximum validation failures
Sigma	Determines change in the weight for second derivative approximation.
lambda	Regulates the indefiniteness of the hessian.

RESULTS AND DISCUSSIONS

The data set divided into two groups, one used for training and other for testing. The training set consists of 70% of the aggregate data and remaining 30% are used as testing. We also perform experiments on same (30% or 70%) dataset which is training as well as testing for kNN classifier. The results on these experiments have a 100% accuracy rate. This means, if the user who is supposed to use this project has already contributed to our dataset earlier, the system will guarantee 100% recognition rate. The results from other experiments are shown as follows.

Feature extractor	**Classifier**	
	kNN	**NPRTOOL**
Direct pixel value	78.63%	97.10%
Hierarchical centroid	70.96%	94.30%

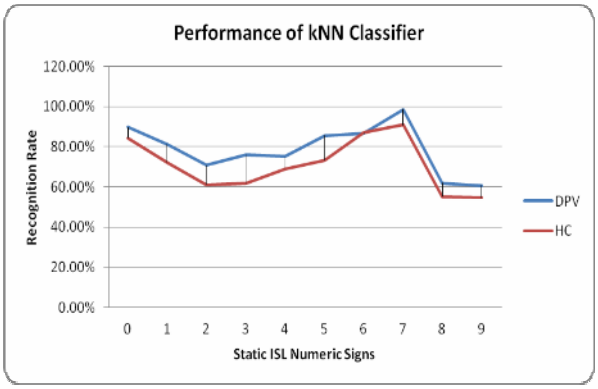

Figure-2. Performance of kNN classifier against Direct Pixel Value (DPV) and Hierarchical Centroid (HC) feature extraction methods.

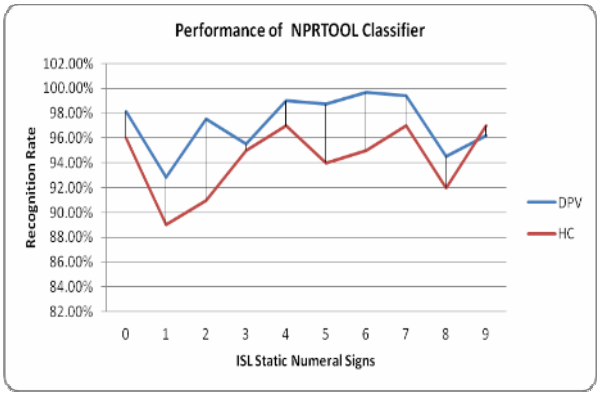

Figure-3. Performance of NPRTOOL classifier against Direct Pixel Value (DPV) and Hierarchical Centroid (HC) feature extraction methods.

From the Figures (2 and 3) we can inference that the direct pixel value feature extraction performs better than that of hierarchical centroid method. The signs '9' and '5' are confused against each other as both have very similar shapes. Also signs '8' and '9' are confused against each other and also '4' and '5' are confused against each other.

CONCLUSIONS

The gesture recognition system is capable of recognizing only numerical ISL static signs with 97.10% accuracy. The experimental result shows that system can

be used as a "working system" for Indian Sign Language numerical recognition.

FUTURE WORK

The system can be useful for static ISL numeral signs only. The ISL recognizer system cannot be considered as a complete system, as for complete recognition of sign language, we have to include ISL alphabets, words and sentences. These signs can be included in future. Also other feature extraction algorithms like Wavelet transform, Invariant moments, Shapelets descriptors and other existing methods can be included in conducting experiments for improvement in the results. Other classifiers like multi class Support Vector Machine (SVM), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) or a combination of these classifiers can be included in conducting experiments to improve the recognition rate.

REFERENCES

[1] Sahoo, A. K., G. S. Mishra and K. K. Ravulakollu. 2014. Sign Language Recognition: State Of The Art. ARPN Journal of Engineering and Applied Sciences. 9(2): 116-134.

[2] Zeshan U., M. M. Vasishta and M. Sethna. 2005. Implementation of Indian Sign Language in Educational Settings. Asia Pacific Disability Rehabiliation Journal. (1): 16-40.

[3] Tanibata, N, N. Shimada and Y. Shirai. 2002. Extraction of hand features for recognition of sign language words. In: Proceedings of International Conference on Vision Interface. pp. 391-398.

[4] Balakrishna, G. and P. S. Rajam. 2012. Recognition of Tamil Sign Language Alphabet using image processing to aid Deaf-Dumb People. International Conference on Communication Technology and System Design. 30: 861-868.

[5] Ghotkar, A. S., S. Khatal, S. Khupase, S. Asati and M. Hadap. 2012. Hand Gesture Recognition for Indian Sign Language. International Conference on Computer Communication and Informatics (ICCCI-2012), Coimbatore. pp. 10-12.

[6] Futane, P. R. and R. V. Dharaskar. 2011. Hasta Mudra: An Interpretation of Indian sign hand Gestures. International Conference on Electronics Computer Technology (ICECT). : 377-380.

[7] Deora, D. and N. Bajaj. 2012. Indian sign language recognition. 1st International Conference on Emerging Technology Trends in Electronics Communication and Networking (ET2ECN). : 1-5.

[8] Rekha J., J. Bhattacharya and S. Majumder. 2011. Shape, texture and local movement hand gesture features for Indian Sign Language recognition. 3rd International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35.

[9] Kar, M. K., D. R. Neog and M. K. Bhuyan. 2011. Hand Pose Identification from Monocular Image for Sign Language Recognition. IEEE International Conference on Signal and Image Processing Applications.

[10] Vamvakas, G., B. Gatos and J. Stavros. Perantonis. 2010. Handwritten character recognition through two-stage foreground sub-sampling. Pattern Recognition. 43(8): 2807-2816.

[11] Mitchell, T. M. 1997. Machine learning. McGraw-Hill Science/ Engineering/Math.

[12] http://www.mathswork.in/help/images/ref/rgb2gray.html (Accessed on 20.04.2014).