

A Review on Indian Sign Language Recognition

Anuja V. Nair

M. Tech Signal Processing
Department of Electronics and Communication
Engineering
SCT College of Engineering,
Thiruvananthapuram, Kerala, India

Bindu V.

Department of Electronics and Communication
Engineering
SCT College of Engineering,
Thiruvananthapuram, Kerala, India

ABSTRACT

Automatic Sign Language Recognition is an extensive research area in the field of human computer interaction. Such recognition systems are meant to replace sign language interpreters. With the development of image processing and artificial intelligence techniques, many techniques have been recently developed in this area. Most of the signs in Indian Sign Language (ISL) are double handed and hence it is more complex compared to single handed American Sign Language (ASL). So, most of the researchers use ASL signs for creating their database. Recently, researchers from India have started working on ISL to develop automatic Indian sign language recognition systems. Mainly three steps are involved in sign language recognition-preprocessing, feature extraction and classification. The important classification methods used for recognition are Artificial Neural Networks (ANN), Support Vector Machine (SVM), Hidden Markov Models (HMM) etc.

General Terms

Pattern Recognition, Human computer interaction, Artificial intelligence

Keywords

Sign language recognition, Indian Sign Language, ANN, SVM, HMM

1. INTRODUCTION

Sign language is the medium of communication for the hearing impaired people. It uses gestures instead of sound to convey meaning. It combines hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns for conveying messages. It is not universal and just like spoken language, it has regional dialects. American Sign Language (ASL), British Sign Language (BSL), Indian Sign Language (ISL) etc. are some of the common sign languages in the world. In India, over 2 million people are deaf. They find it difficult to communicate with the normal people as the hearing or normal people are unfamiliar of sign language. There arises the need for sign language interpreters who can translate sign language to spoken language and vice versa. But, the availability of such interpreters is limited and expensive. This resulted in the development of automatic sign language recognition systems which could automatically translate the signs into corresponding text or voice without the help of sign language interpreters. Such systems can help in the development of deaf community through human computer interaction.

Sign language recognition is an important research area since there are a lot of challenges in developing an automatic recognition system. Most of the researchers in this area concentrate on the recognition of American Sign Language (ASL) since most of the signs in ASL are single handed and thus, complexity is less. Another attractive feature is that ASL already has a standard database that is available for use. When compared with ASL, Indian Sign Language (ISL) relies on both hands and thus, an ISL recognition system is more complex. The research works carried out by the researchers in the recognition of ISL is very less. Recently, more researchers have started doing research in this area. A project is currently going on at IIT, Guwahati, India which is aimed to develop an automatic ISL education and recognition platform for hearing impaired students of India.

2. SIGN LANGUAGE RECOGNITION- DIFFERENT APPROACHES

Sign language recognition is an important application of gesture recognition. There are mainly two different approaches in sign language recognition - Glove based approach and vision based approach. The first category requires signers to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation during processing. The drawback of this approach is that the signer has to wear the sensor hardware along with the glove during the operation of the system. Vision based approach uses image processing algorithms to detect and track hand signs as well as facial expressions of the signer. This approach is easier to the signer since there is no need to wear any extra hardware. However, there are accuracy problems related to image processing algorithms and these problems are yet to be modified.

There are again two different approaches in vision based sign language recognition: 3D model based and appearance based [1]. 3D model based methods make use of 3D information of key elements of the body parts. Using this information, several important parameters, like palm position, joint angles etc., can be obtained. This approach uses volumetric or skeletal models, or a combination of the two. In computer animation industry and for computer vision purposes, volumetric approach is better suited. This approach is very computational intensive and also, systems for live analysis are still to be developed.

Appearance-based systems use images or videos as inputs. They directly interpret from these videos/images. They don't use a spatial representation of the body. The parameters are derived directly from the images or videos using a template database. Some templates are the deformable 2D templates of

the human parts of the body, particularly hands. Deformable templates are sets of points on the outline of an object, used as interpolation nodes for the object's outline approximation. One of the simplest interpolation functions is linear. It performs an average shape from point sets, point variability parameters and external deformations. These template-based models are mostly used for hand-tracking, but could also be used for simple gesture classification. A second approach in sign language gesture detection using appearance-based models uses image sequences as the gesture templates. Either the images themselves, or certain features derived from these images can be used as the parameters.

3. OVERVIEW OF SIGN LANGUAGE RECOGNITION SYSTEM

A simple block diagram of a sign language recognition system is shown in Figure 1. The entire recognition process can be divided into two phases- training and testing. In the training phase, the classifier has to be trained using the training dataset. The database can be either created by the researcher himself or an available database can be used. An external webcam, digital camera or inbuilt webcam in the laptops can be used to capture the training images. Most of the sign language recognition systems classify signs according to hand gestures only or in other words, facial expressions are excluded. The important steps involved in training phase are creation of database, preprocessing, feature extraction and training the classifier. The testing phase includes video/image acquisition (input can be videos or images), preprocessing, feature extraction and classification.

3.1 Preprocessing

A preprocessing step is carried out on the training images to extract the region of interest (ROI). The ROI can be hands if only hand gestures are considered or both face and hands if the facial gestures are also included. Usually the preprocessing step consists of filtering, image enhancement, image resizing, segmentation and morphological filtering. Filtering and image enhancement can be any one of the commonly used methods. For segmentation, the algorithm that better suits the input video/images has to be selected. Otsu's thresholding [2], Background subtraction [3], skin color based segmentation [4] and motion based segmentation [5] are the commonly used segmentation techniques. During testing phase, the test images or videos are also preprocessed to extract the region of interest.

3.2 Feature Extraction

Feature extraction is one of the most crucial steps of sign language recognition since the inputs to the classifier are the feature vectors obtained from this step. The techniques used for feature extraction should find shapes reliably and robustly irrespective of changes in illumination levels, position, orientation and size of the object in a video/image. Objects in an image are represented as collection of pixels. For object recognition we need to describe the properties of these groups of pixels. The features can be obtained in different ways: wavelet decomposition [6], Haar wavelets, Haar-like features [7], texture features [8], orientation histogram[9], scale invariant feature transform[10], Fourier descriptors etc. In some cases the ROI pixels will be used as the feature vector after a dimensionality reduction using Principal Component Analysis (PCA) [11]. The feature vector thus obtained using any one of the feature extraction methods is used for training the classifier.

3.3 Classification

A classifier is needed in sign language recognition to classify the input signs into different classes. The feature vector obtained from the training database is used to train the classifier during the training phase. When a test input is given, the trained classifier identifies the class corresponding to the sign and displays the text or plays the sound. The test inputs can be images or videos. Most commonly used classifiers are Hidden Markov Models (HMM), Artificial Neural Networks (ANN), Multiclass Support Vector Machines (SVM), Fuzzy systems, K Nearest Neighbor (KNN) etc. The performance of the classifier is measured in terms of recognition rate.

3.3.1 Hidden Markov Models

A Hidden Markov model [20] is a collection of finite states connected by transitions. Each state is characterized by two sets of probabilities: a transition probability and either a discrete output probability distribution or continuous output probability density function which, given the state, defines the condition probability of emitting each output symbol from a finite alphabet or a continuous random vector. The HMM approach to gesture recognition is motivated by the successful application of Hidden Markov modeling techniques to speech recognition problems. HMM is a doubly stochastic model and is appropriate for coping with the stochastic properties in gesture recognition. Instead of using geometric features, gestures are converted into sequential symbols. HMMs are employed to represent the gestures, and their parameters are learned from the training data. Based on the most likely performance criterion, the gestures can be recognized by evaluating the trained HMMs.

3.3.2 Artificial Neural Networks

An artificial neural network [16] involves a network of simple processing elements (artificial neurons) which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations. There are several neural networking algorithms which can be used for gesture recognition. The different networks are feed forward networks, Elman neural networks, Self-organizing networks etc. There are several backpropagation algorithms available for training the neural networks.

3.3.3 Support Vector Machine

The SVM [15] is a popular pattern recognition technique with supervised learning. Since it divides the feature space for each class, the SVM can handle unknown data well, although it is not suited to grouping sample data. It is originally developed by Vapnik and colleagues at bell laboratories. It was actually developed for solving binary decision problems. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output. Thus, it can be called as a non-probabilistic binary linear classifier. For multi-class problems, such problems are decomposed into several two-class problems that can be addressed directly using several SVMs. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly

mapping their inputs into high-dimensional feature spaces. The different approaches for solving multi-class problems with SVM are one against all, one against one, decision directed acyclic graphic approach etc

4. ISL RECOGNITION

Figure 2 shows the signs in ISL corresponding to the English alphabets A, B, C, D and E. ISL recognition is comparatively new in the field of sign language recognition. A practical framework for ISL gesture based human robot interaction had been proposed in [12]. ISL video had been used to compose several hand gestures to train the HOAP-2 robot in real time. Orientation histogram feature was extracted for real time classification using Euclidean distance metric. ISL video had been captured by selecting several dynamic gestures (i.e. sequence of frames) in real time using webcam. The classifier was trained using twenty ISL gestures. The work was carried out in WEBOTS platform using real time JAVA based software (developed by them). The classified ISL gestures with average accuracy of 90% were entirely mapped with specific gesture based applications on humanoid robot.

In [15], the authors took their dataset for dual-hand gesture from the 26 alphabets of ISL. The dataset consisted a total of 2340 ($30 \times 26 \times 3$) images, 3 signers were requested to sign all

26 alphabets 30 times. Features were extracted from the images using Histogram of Orientation Gradient (HOG) and Histogram of Edge Frequency (HOEF). Finally, gestures were classified using Support Vector Machine (SVM). They had proposed a novel HOEF feature, using which they achieved a recognition rate of 98.1% of dual-handed gesture, and they have claimed in their paper that HOEF is better than HOG which is widely used by many for recognition.

A system had been proposed [14] by P.V.V. Kishore and P. Rajesh Kumar to automatically recognize gestures of ISL from a video stream of the signer. Their system converts words and sentences of ISL into voice and text in English. To accomplish the task, they had used powerful image processing techniques such as frame differencing based tracking, edge detection, wavelet transform and image fusion techniques to segment shapes in the videos. Then Elliptical Fourier descriptors were used for shape feature extraction and PCA for feature set optimization and reduction. Database of extracted features were compared with input video of the signer using a trained fuzzy inference system. The proposed system converts gestures into text and voice message with 91% accuracy. Around 80 gestures from 10 different signers were used. The entire system was developed in a user friendly environment by creating a Graphical User Interface (GUI) in MATLAB.

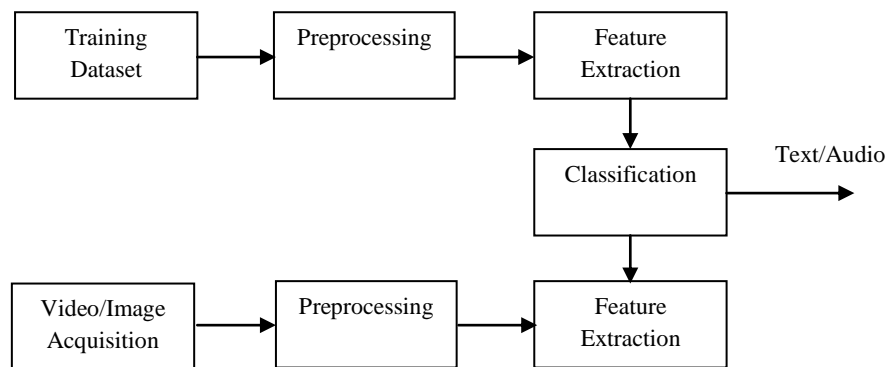


Figure 1. Simple Block Diagram of Sign Language Recognition System

P.V.V.Kishore and P.Rajesh Kumar [16] again proposed a real time approach to recognize gestures of ISL. The input video to the sign language recognition system was made independent of the environment in which signer was present. Active contours were used to segment and track the non-rigid hands and head of the signer. The energy minimization of active contours was accomplished by using object color, texture, boundary edge map and prior shape information. A feature matrix was designed from segmented and tracked hand and head portions. This feature matrix dimensions were minimized by temporal pooling creating a row vector for each gesture video. Pattern classification of gestures was achieved by implementing fuzzy inference system. The proposed system could translate video signs into text and voice commands. Their data base had 351 gestures with each gesture repeated 10 times by 10 different users. A recognition rate of 96% for gestures in all background environments was achieved.

The same authors proposed a complete skeleton of isolated Video Based Indian Sign Language Recognition System (INSLR) [17] that integrates various image processing techniques and computational intelligence techniques in order

to deal with sentence recognition.. A wavelet based video segmentation technique was proposed which detects shapes of various hand signs and head movement in video based setup. Shape features of hand gestures were extracted using elliptical Fourier descriptions which to the highest degree reduces the feature vectors for an image. PCA was used to minimize the feature vector again for a particular gesture video and the features were not affected by scaling or rotation of gestures within a video. Features generated using these techniques made the feature vector unique for a particular gesture. Recognition of gestures from the extracted features was done using a Sugeno type fuzzy inference system which used linear output membership functions. Finally the INSLR system employed an audio system to play the recognized gestures along with text output. The system was tested using a data set of 80 words and sentences by 10 different signers. Their system had a recognition rate of 96%.

The same authors summarize various algorithms used to design a sign language recognition system [18]. They designed a real time sign language recognition system that could recognize gestures of ISL from videos under different complex backgrounds. They have done a lot of works in the

field of ISL recognition. They had used fuzzy classification and Artificial Neural network classification. Segmenting and tracking of non-rigid hands and head of the signer in sign language videos was achieved by using active contour models. Active contour energy minimization was done using signers hand and head skin color, texture, boundary and shape information. Classification of signs was done by an artificial

neural network using error back propagation algorithm. Each sign in the video was converted into a voice and text command. The system has been implemented successfully for 351 signs of Indian Sign Language under different possible video environments. The recognition rates were calculated for different video environments.

Table 1. Comparison of Various Indian Sign Language recognition Systems

Method	Input	Segmentation	Feature Vector	Classification	Recognition Rate	Platform
Anup Nandy, Soumik Mondal	Real time video	N/A	Orientation Histogram	Euclidean distance	90%	Real time JAVA based software.
Himanshu Lilha, Devashish Shivmurthy	Images	N/A	Histogram of Edge Frequency (HOEF).	Support Vector Machine	98.1%	N/A
P.V.V Kishore, P. Rajesh Kumar	Video	Canny fused Wavelet based	Elliptical Fourier descriptors	Fuzzy	91%	Matlab
P.V.V.Kishore, P.Rajesh Kumar	Real time	Active contours	Texture features	Fuzzy	96%	Matlab 7.0.1
P.V.V.Kishore,P. Rajesh Kumar	Video	Wavelet based	Elliptical Fourier descriptors	Fuzzy	96%	Matlab 7.0.1
P.V.V. Kishore, P.Rajesh Kumar	Real-time	Active contours	Texture features	ANN- error back propagation algorithm	93%	Matlab
Geetha M., Manjusha U C	Images	Boundary tracing	A novel method based on B-spline approximation	Support vector machine	Numbers- 93.2% Alphabets- 91.83%	OpenCV

Geetha M and Manjusha U C proposed [19] a novel vision-based recognition of Indian Sign Language Alphabets and Numerals using B-Spline Approximation. Gestures of ISL alphabets were complex since it involves the gestures of both the hands together. Their algorithm approximates the boundary extracted from the Region of Interest, to a B-Spline curve by taking the Maximum Curvature Points (MCPs) as the Control points. Then the B-Spline curve was subjected to iterations for smoothening resulting in the extraction of Key Maximum Curvature points (KMCPs), which are the key contributors of the gesture shape. Hence a translation & scale invariant feature vector was obtained from the spatial locations of the KMCPs in the 8 Octant Regions of the 2D Space. SVM was used for classification.

5. CONCLUSION

Recent research works have focused mainly on the recognition of static signs of ISL from images or video sequences that have been recorded under controlled conditions. Special hardware like data gloves, colored gloves or markers is used in some systems. Certain systems use only bare hands but only one background is considered. Most of the systems are signer-dependent and also, the signer needs to wear full sleeve dark colored jackets. Facial features are not included in majority of the systems. Development of signer independent systems which could recognize signs from both facial and hand gestures is a major challenge in the area of sign language recognition. The researchers should focus on the development of a well suited segmentation scheme which is capable of extracting the hand and face region from videos/images having any background. More emphasis should also be given for extracting such features which could completely distinguish each sign irrespective of hand size, distance from the source, color features and lighting conditions.

6. REFERENCES

- [1] Noor Adnan Ibraheem, RafiqulZaman Khan, 2012, Survey on Various Gesture Recognition Technologies and Techniques, *International Journal of Computer Applications*, Volume 50, No. 7
- [2] Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung, 2001, A Fast Algorithm for Multilevel Thresholding, *Journal of Information Science and Engineering* 17, pp. 713-727
- [3] Dr. Alan M McIvor, Background subtraction techniques, *Image and Vision Computing Newz Zealand 2000 (IVCNZ00)*
- [4] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai, Skin Segmentation Using Color and Edge Information, *Proceedings on International Symposium on Signal Processing and its Applications*, 1-4 July 2003, Paris, France
- [5] Jorge Badenas, Josee Miguel Sanchiz, Filiberto Pla, 2001, Motion-based Segmentation and Region Tracking in Image Sequences, *Pattern recognition* 34, pp. 661-670
- [6] Sanjay Kumar and Dinesh K. Kumar, 2005, Visual Hand Gestures classification Using Wavelet Transform and Moment Based Features, *International Journal of Wavelets, Multiresolution and Information Processing*, Volume 3, Issue 1

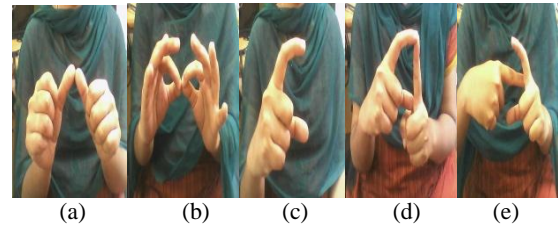


Figure 2. ISL Signs for (a) A, (b) B, (c) C, (d) D and (e) E

- [7] Qing Chen, Nicolas D. Georganas, Emil M. Petriu, Real-Time Vision-based Hand Gesture Recognition Using Haar-Like Features, *Instrumentation and Measurement Technology Conference-IMTC 2007*, Parsaw, Poland, May 1-3
- [8] J. Rekha, J. Bhattacharya, and S. Majumder, Shape, Texture and Local Movement Hand Gesture Features for Indian Sign language Recognition, *3rd International Conference on Trendz in Information Sciences and Computing*, 8-9 Dec, 2011
- [9] Hyung-Ji Lee, Jae-Ho Chung, Hand Gesture Recognition Using orientation histogram, *TENCON 99, Proceedings of the IEEE Region 10 Conference*, 1999
- [10] Chieh-Chih Wang and Ko-Chih Wang, Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction, *Recent Progress in Robotics, LNCIS* 370, pp. 317-329, 2008
- [11] Deng-Yuan Huang, Wu-Chih Hu, Sung-Hsiang Chang, Vision-Based hand Gesture Recognition Using PCA+Gabor Filters and SVM, *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009
- [12] Anup Nandy, Soumik Mondal, Jay Shankar Prasad, Pavan Chakraborty and G.C.Nandi, 2010 Recognizing & Interpreting Indian Sign Language Gesture for Human Robot Interaction, *International Conf. on Computer & Communication Technology*, 2010 |ICCT'10|, pp. 712-717
- [13] Qing-song Zhu, Yao-qin Xie, Lei Wang (2010) Video Object Segmentation by Fusion of Spatio-Temporal Information Based on Gaussian Mixture Model, *Bulletin of advanced technology research*, vol. 5, No. 10, pp 38-43.
- [14] P.V.V Kishore, P. Rajesh Kumar, E. Kiran Kumar & S.R.C.Kishore, 2011, Video Audio Interface for Recognizing Gestures of Indian Sign Language, *International Journal of Image Processing (IJIP)*, Volume 5, Issue 4, 2011 pp. 479-503
- [15] Himanshu Lilha, Devashish Shivmurthy, 2011, "Evaluation of Features for Automated Transcription of Dual- Handed Sign Language Alphabets", *International Conference on Image Information Processing (ICIIP)*, 3-5 Nov. 2011
- [16] P.V.V.Kishore, P.Rajesh Kumar, 2012, A Model For Real Time Sign Language Recognition System, *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 2, Issue 6, June 2012 pp. 30-35

- [17] P. V. V. Kishore and P. Rajesh Kumar, 2012, Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic, IACSIT International Journal of Engineering and Technology, Vol. 4, No. 5, October 2012 pp. 537-542.
- [18] P.V.V.Kishore, P.Rajesh Kumar,2012, Segment, Track, Extract, Recognize and Convert Sign Language Videos to Voice/Text, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, pp. 35-47
- [19] Geetha M., Manjusha U. C. 2013,A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, International Journal of Computer Science and Engineering (IJCSE)
- [20] Tie Yang, Yangsheng Xu, 1994, Hidden Markov Model for Gesture recognition
- [21] Aseema Sultana, T. Rajapushpa, Vision Based Gesture Recognition for Alphabetical Hand gestures Using the SVM Classifier, International Journal of Computer Science and Engineering Technology, Volume 3, No. 7, 2012