Vision Based Hand Gesture Recognition for Indian Sign Languages Using Local Binary Patterns with Support Vector Machine Classifier

K.Muthukumar, S. Poorani, S. Gobhinath

1Assistant Professor, Department of EEE, Sri Krishna College of Engineering and Technology
2Professor, Department of EEE, Karipagam University, Coimbatore.

Received 28 January 2017; Accepted 22 March 2017; Available online 28 April 2017

ABSTRACT
Sign Language is a language which uses visually transmitted sign patterns to convey meaning by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluently express one's thoughts / communicate with others and is commonly used by the physically impaired people who cannot speak or hear. Automatic Sign Language system requires fast and accurate techniques for identifying static signs or a sequence of produced signs to help interpret their correct meaning. Major components of a Sign Language are Hand Gesture. In this paper, a robust approach for recognition of bare-handed static sign language is presented, using a novel combination of features. These include Local Binary Patterns (LBP) histogram features based on color and depth information, and also geometric features of the hand. Linear binary Support Vector Machine (SVM) classifiers are used for recognition, coupled with template matching in the case of multiple matches. This research aim towards working on hand gesture recognition for sign language interpretation as a Human Computer Interaction application.

KEYWORDS: Indian Sign Language (ISL), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Local Binary Pattern (LBP).

INTRODUCTION

Sign language (SL) is a natural language used for communication by hearing and / or speech impaired persons. It is a language which uses manual communication and body language to convey meaning, as opposed to acoustically conveyed sound patterns. They share many similarities with spoken languages which is why linguists consider both to be natural languages, but there are also some significant differences between signed and spoken languages [1]. Though sign language is spread and used all over the world, it is not universal. Hundreds of sign languages are in use around the world and are at the core of local deaf cultures. Some sign languages have obtained some form of legal recognition, while others have no status at all. Regionally American Sign Language (ASL), French Sign Language (LSF), German Sign Language (GSL), Indian Sign Language (ISL), British Sign Language (BSL) etc. have been evolved.

Indian Sign Language is one of the first known sign language systems and is considered extremely important in the history of sign languages, but it is rarely used today. In linguistic terms, sign languages are as rich and complex as any spoken language, despite the common misconception that they are not “real languages.”
Professional linguists have studied many sign languages and found that they exhibit the fundamental properties that exist in all languages [2] [3]. The elements of a sign are Hand shape, Orientation (or Palm Orientation), Movement, and Facial Expression summarized in the acronym HOLME [4]. The main idea behind the proposed method is to exploit a novel combination of color, depth, and geometric information of a hand sign in order to increase recognition performance while most approaches only attempt to use a combination of two or less [6][9-15]. This enables recognition of a wider set of hand signs even though they appear to be very similar.

![Diagram of hand pose recognition system](image)

**Fig. 1:** Overview of the proposed hand pose recognition system.

**LITERATURE SURVEY:**
Developing a vision based human computer interaction system for sign language interpretation is a challenging problem for researchers. This chapter presents theoretical foundation and literature survey. It reviews the research based on sign language and the challenges involved. Few unaddressed issues and problems that the spoken and written language of a country is different from other country. Although the same language has been used by a number of countries, however, the syntax and semantics of a language is dependent on a country/region. For example, English is the official language of the UK, USA and many other nations. The usage of English differs at country level. Similarly, the sign language of a country is not similar than that other country.

The focus of this study is on the development of sign languages at international level [2]. To acquire data for SLI, earlier data gloves and accelerometers were used to specify hand. The location, orientation and velocity were measured using a tracker [14] or data gloves [15]. While these techniques gave accurate positions, but they had the disadvantage of high costs and restricted mobility, which changed the signs [6]. Hence vision based systems came into existence and have become popular. In case of vision input, a sequence of images are captured from a combination of cameras. A sequence of images is captured using either monocular [16], stereo [17] or orthogonal [18] cameras. Feris et al. [19] used external light sources to illuminate the scene and then a multi-view geometry to construct a depth image.

Xiaolong Zhu et al. [18] proposed advances in the methodology of Hybrid Classification architectures considering Face and Hand gesture recognition. They build this hybrid architecture using an ensemble of connectionist networks- radial basis functions (RBF) and inductive decision trees (DT), which combines the merits of ‘holistic’ template matching with those of ‘abstractive’ matching using discrete features and subject to both positive and negative learning. C. Huang et al. [20] have investigated affective body gesture analysis in video sequences beyond facial expressions. They proposed to fuse facial expressions and body gesture at the feature level using Canonical Correlation Analysis (CCA). Z. Ren et al. [15] have proposed an integration of face and hand gesture recognition. They have claimed that face recognition rate can be improved by hand gesture recognition. They have proposed a security elevator scenario. They have claimed that the integration of two search engines proposed by them is general and is not only for face and hand gesture recognition.

**HAND GESTURE RECOGNITION:**
A sign in a Sign Language (SL) as discussed earlier consists of three main parts: Manual features, non-manual features and finger spelling [1]. To interpret the meaning of a sign, all these parameters are to be analyzed simultaneously. Sign language poses a main challenge of being multichannel. Each channel in this
system is separately built and analyzed and the output of each channel is combined at the final stage to draw conclusion.

The research in Sign Language Interpretation (SLI) began with Hand Gesture Recognition (HGR). Hand gestures are extensively used in human non-verbal communication by hearing impaired and speech impaired people. Even normal people sometimes use sign language for communication. Though sign language is spread and used all over the world, it is not universal. Wherever hearing impaired community exists, sign languages develop. To make communication between the hearing impaired and normal people simple and efficient, it is necessary that this process be automated. Number of techniques have been developed for automatic HGR. The overall process of Hand Gesture Recognition (HGR) system block diagram is as shown in figure 2. There are three similar steps in HGR:

1. Hand acquisition which deals with hand extraction from the given static image and hand extraction and tracking from a video.
2. Feature extraction which basically deals with compressed representation of the data which will facilitate the recognition of the hand gesture.
3. Classification/ recognition of the hand gesture following some rule.

Fig. 2: Block Diagram for Process of Hand Gesture Recognition.

DATA SETS ACQUISITION:

In this research, two different data sets are used in ISL recognition system. The data sets are ISL digits (0-9) and single handed ISL alphabets (A-Z). For data set acquisition, dark background for uniformity and easy in manipulation of images for feature extraction and classification is initially chosen. A Sony digital camera, Cyber shot H70, is used for capturing the images. All images are captured with flash light in an intelligent auto mode. The common file format JPEG is used to capture the images as it is a common image standard now a day. Each original image is 4608×3456 pixel and required approximately 5.5 MB storage space. To create an efficient data set with a reasonable size, the images are cropped to 200×300 RGB pixels and barely 25 KB memory space is required per image. The data set is collected from 100 signers. Out of these signers, 69 are male and 31 are female with average age group of 27. The average height of a signer is about 66 inches. The data set contains isolated ISL numerical signs (0-9). Five images per ISL digit sign is captured from each signer. Therefore, a total of 5,000 images are available in the data set. The sample images of the data set are shown in figure 3.

Fig. 3: The ISL Digit Signs Data set
In this data set, a total of 2600 images cropped to 200×300 RGB pixel sizes are available. The images are collected from four males and six females. The backgrounds of sign images are dark, as only hand orientations are required for the feature extraction process. The images are stored in JPEG format because it can be easily exported and manipulated in different hardware and software environments. Each pre-processed ISL sign image required nearly 25 KB of storage space with 72 dpi. The size of the images is 200×300 pixels. The skin colors of these images are neither very dark complexion nor very white complexion. This is due to the proposed application can be useful in India only. The colors corresponding to human skins are mainly used in capturing the sign images. A sample data set is shown in figure 4.

![Fig. 4: The ISL Single Handed Alphabet Signs Data Sets](image)

**HAND GESTURE SEGMENTATION USING LINEAR DISCRIMINANT ANALYSIS AND LOCAL BINARY PATTERN:**

Segmentation is used to detect hand from background [5]. The experimentation in this work is carried out using two datasets representing hand gestures performed with one hand for alphabets A to Z using Indian Sign Language. The images of this dataset before and after preprocessing stage are shown in fig. 5.

![Fig. 5: (a) Original images of alphabet 'A', (b) Images after RGB to Gray conversion and resizing.](image)

**A. Linear Discriminant Analysis (LDA):**

The Linear Discriminant Analysis (LDA) performs a class specific dimension reduction. [7] It finds the combination that best separates different classes. To find the class separation, LDA maximizes both between class and within class scatters instead of maximizing the overall scatter. As a result, same class members cluster together and different class members stay far apart from each other in the lower dimensions. Let, \( X \) be a vector with samples from \( c \) classes.

Let, \( X \) be a vector with samples from \( c \) classes.

\( X = \{x_1, x_2, ..., x_c\} \)

\( X_l = \{x_1, x_2, ..., x_c\} \)

The between class and within class scatters, \( S_B \) and \( S_W \) are calculated as follows.
Mean of vector data and mean of the class i, where i = 1; :::; c.

LDA finds a projection, W that maximizes the class separation criterion.

\[ W = \underset{W}{\text{argmax}} |W^T S_W W| \]

The rank of SW is at most (N c), where N is the number of samples and c is the number of classes. Almost always the number of samples is less than the dimension of the image data in pixels. Principal Component Analysis (PCA) is performed on the image data and projected on a (N to c) dimensional space. LDA is performed on this reduced data. The transformation matrix, W projecting the sample in to (c 1) dimensional space is,

\[ W = W_{\text{lda}}^T W_{\text{pca}}^T \]

Here,

\[ W_{\text{pca}} = \underset{W}{\text{argmax}} |W^T S_W W| \]
\[ W_{\text{lda}} = \underset{W}{\text{argmax}} |W^T S_B W| \]

B. Local Binary Pattern (LBP):

Local Binary Patterns (LBP) was proposed by Ong et. al.[11] It performs local operations on the neighborhood of an image pixel. The neighborhood of a pixel is the pixels adjacent to that particular pixel. In LBP an 8 bit binary code is for a 3 X 3 pixel neighborhood of image I is,

\[ b_j = \begin{cases} 1, & \text{if } (x_i, y_i) > (x_0, y_0) \\ 0, & \text{otherwise} \end{cases} \]

Local Binary Pattern (LBPs) have proven to be very effective for image representation and have been applied in various analysis. The LBPs are tolerant against monotonic illumination changes and are able to detect various texture primitives like spot, line end, edge, corner etc. The most popular and efficient version of LBP i.e. Block LBP (figure 8) with uniform / no uniform patterns is used as the first technique for extracting hand features [6].
FEATURE EXTRACTION USING SUPPORT VECTOR MACHINE CLASSIFIER:

The feature extraction [6] approaches in image processing, extracts valuable information present in an image. This deals with conversion of a high dimensional data space into lower dimensional data space. The lower dimensional data extracted from images should contain precise information which is the representative of the actual image. The image can be reconstructed from the lower dimensional data space. The lower dimensional data is required as input to any classification technique as it is not feasible to process higher dimensional data with speed and accuracy. The inputs to an automatic sign language recognition system are either static signs (images) or dynamic signs (video frames) [2]. In order to classify input signs in an automatic sign language recognition system, extraction of valuable features from signs is required. All the algorithms that are used for facial feature extraction are used for Hand feature extraction as well.

Classification is a technique part of machine learning. The technique is used to classify each item in a data set into one of a predefined set of groups. Classification methods use mathematical models including linear programming, decision trees, statistics and neural networks for pattern classification. In classification, a software module is created that could learn the art of classifying the data items into different groups. With initial experimentation using multiclass SVM and decision trees, a huge number of misfits have been identified in the process of classification. Hence these classifiers are not further used for final experimentation towards recognition.

During SVM classification, if more than one sign returns a positive match for a test image pair, the template matching process is executed. Firstly, the test image pair is checked with all the signs which returned a positive match if it falls within the range of height to width ratios of that sign defined by rmn and rmax. If the range of ratios of a sign does not fall into that of the test image pair, the sign will not be considered as a positive match in the subsequent template matching steps. The cosine distance d cosine is then calculated between the feature vector φ of the test image pair and the average feature vector φ avg of each sign that returned a positive match.

An edge template similarity metric sedge is also calculated according to (3). Here a bitwise AND operation is performed between the edge template of the test image pair Xtest and the edge template of each sign that returned a positive match Xsign. The sum of the number of white pixels in the resulting image is considered to be edge. Although the image pairs are of different sizes, the resizing of the edge template into a standard size allows a direct bitwise AND operation to be performed.

\[ s_{edge} = \sum_{i,j} X_{test}(i,j) \cap X_{sign}(i,j) \]

The total similarity metric stot is then defined according to (4). Here \( \alpha = 0.001 \) and \( \beta = 1.2 \) were chosen as it produced optimum results. The sign for which the similarity metric stot returns a maximum will be considered as the final output sign.

\[ s_{tot} = \alpha s_{edge} + \beta (1 - d_{cosine}) \]
Although 26 classes are present in ISL single handed alphabet, the system is able to predict single handed characters with more than 95% accuracy. This is possible with LBP and SVM feature extraction technique. In figure 9, a sample output is shown for single handed ISL sign 'B'. The input sign image is processed through the system and a prediction is shown in the right-hand side of the output screen. The sign interpreted as single handed 'B' which is the correct prediction.

A. Performance of Sign Language Interpretation System:

For sign language interpretation, N-fold cross validation method was used with the N = 5. For a single hand (left or right) each fold is consisting of 200 images. The system is trained using 800 images from four of the five folds and tested against the remaining fold of 200 images. For both hands, each fold has 400 images. The system is trained using 1600 images from four of five folds and tested against the remaining fold of 400 images. The system was tested under three criteria. Sign gestures performed by left, right and both hands. The accuracy of all three criteria is measured using the following condition, where NC is the number correctly classifier sign gestures and N is the number of all test sign gestures.

\[ \text{Accuracy(\%)} = \frac{NC}{N} \]

An overall accuracy of 92.14% was obtained with a relatively small training dataset. The results are superior to those achieved in [1] and [2]. It could be seen that the system coped well with variation of individual signs caused by different users as well as the similarity that exist among different signs.

Conclusion:

In this work, a vision-based automatic sign language recognition was presented which is able to recognize sentences in Indian Sign Language. Several features and different methods to combine them were investigated and experimentally evaluated. Tracking algorithms with applications to hand and head tracking were presented and experiments were carried out to determine the parameters of these algorithms. An emphasis was put on appearance-based features that use the images itself to represent signs. Other systems for automatic sign language recognition usually require a segmentation of input images to calculate features for the segmented image parts. The algorithm was designed to function in real-time without requiring excessive computational power. The results indicate that it is possible to train the system to recognize more static Indian Sign Language hand signs while maintaining high accuracy. It is also desirable to build on the framework to recognize dynamic sign language. Future depth sensor technology with higher depth and color resolution and more accurate skeletal tracking has the potential to improve the results of the proposed algorithm even further. The results presented in this work show that the usage of appearance-based features yields a promising recognition performance.

REFERENCES