

Contour Potential Energy Based Method for Indian Sign Language Recognition

Adithya V, Vinod P R, Usha Gopalakrishnan

Abstract— Sign Language is the only way of communication for the people with hearing problems. Sign Language is composed of a number of gestures formed by various handshapes, body movements or facial expressions. Sign language is known and used by deaf people and people living close to them. Most hearing people do not understand sign language. This causes the isolation of deaf people in the society. In order to facilitate the communication between the deaf people and the hearing people an automatic system for sign language interpretation is essential. Such a system will definitely support the integration of deaf community with the hearing community. There are various sign languages across the world. The visual language used by the deaf community in India is called Indian sign language. This paper proposes a vision based method for the automatic recognition of fingerspelling in Indian sign language. The signs considered for recognition include the letters of the English alphabet and the numerals from 0-9. The method makes use of the images of the bare hands and is inexpensive, therefore facilitating the deaf-dumb people to use it.

Index Terms— Indian Sign Language, Hand Segmentation, Contour Potential Energy, Fourier Transform, Central Moments, Classification, Artificial Neural Network.

1 INTRODUCTION

Sign Language is the natural way of communication for the deaf-dumb people. Deaf people use hand gestures, body movements and facial expressions to communicate with others. Sign language is considered as a core distinctive feature of the deaf community. Sign language recognition is the process of decoding various gestures and converting them to textual or voice form. Sign language recognition provides a way for the interaction of deaf people with the hearing people. Usually deaf people depend on sign language interpreters for decoding the gestures. But this system is very expensive and may not be available throughout the life period of a deaf person. To overcome these difficulties, an automatic sign language recognition system is necessary. Such a system can reduce the gap in communication between deaf people and normal people in the society.

Indian Sign Language (ISL) is the visual language used by the deaf community of India. ISL consists of word level signs as well as fingerspelling. Fingerspelling is used for letter by letter signing. It is used to code a word for which the sign does not exist or to emphasize a particular idea or thought. This paper proposes a method for automatically recognizing the fingerspelling in Indian sign language [1]. The signs considered for recognition include the letters of the English alphabet and the numerals from 0-9. In the proposed method, the hand shape is considered as the feature for differentiating different gestures. Shape features are

extracted using contour based potential energy and the classification is done using artificial neural network. The representations of ISL alphabet and numerals are given in Fig 1.

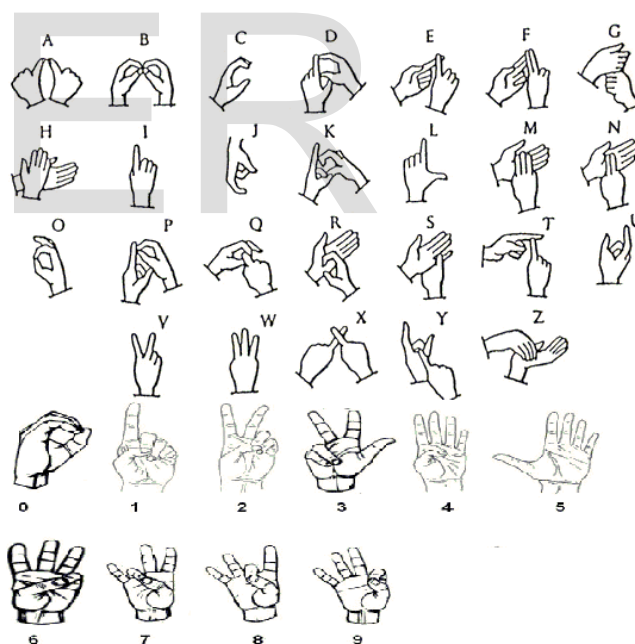


Fig1. ISL Alphabet and Numerals

This paper is organized in five sections: section II discusses about the previous research works done in the area of sign language recognition. The proposed method is presented in detail in section III. The experimental results are given in section IV. Finally section V summarizes and concludes the work.

- Adithya V is currently pursuing M.Tech degree in Computer Science with specialization in Image Processing at College of Engineering, Chengannur, Kerala, India.
- Vinod P R is currently working as Assistant Professor at College of Engineering, Chengannur, Kerala, India.
- Usha Gopalakrishnan is currently working as Assistant Professor at Musaliar College of Engineering and Technology, Pathanamthitta, Kerala, India.

2 LITERATURE REVIEW

Sign language recognition methods are mainly classified into two broad categories: device based methods and vision based methods. Device based approach needs special hardware devices to extract the physical features of the hand sign such as dimension, angle, motion and colour. In comparison, vision based methods use image processing algorithms to detect, track and interpret hand signs. This approach has the advantage that the user does not have to wear clumsy devices.

The earliest reported work on sign language recognition was done by Starner and Pentland [2]. They developed a glove based system to recognize American Sign Language. This method uses a colour camera to track hands from video and the interpretation is done using hidden markov model. Deng-Yuan Huang¹ et.al. proposed a hand gesture recognition system in which the features are extracted using Gabor filters [3]. Rini Akmelia et.al. presented a method [4] for the recognition of Malaysian Sign Language recognition. A real time Arabic sign language recognition system proposed by Nadia R. Albelwi and Yasser M. Alginahi is presented in [5]. This method utilizes a Haar like method to track hand in the video frames and the bounded hand region becomes the area of interest. The resultant images are transformed into the spectral domain using Fourier transformation to form the feature vectors and the classification is done using KNN method.

Stephan Liwicki and Mark Everingham presented a method for recognizing the British sign language [6], in which the hand shapes are described by a joint histogram over quantized gradient orientation and position and the classification was done using HMM. Azadeh Kiani Sarkalehl et.al. proposed a system for Persian sign language recognition in [7]. This method uses discrete wavelet transform to extract features from the gesture images and a multilayered perceptron neural network for the classification of gestures. Al-Jarrah and Halawani [8] presented a method for automatic translation of gestures of the manual alphabet in the Arabic sign language. This system utilizes the feature values that comprise of some length measures which indicate the positions of the fingertips. Classification is done using a subtractive clustering algorithm and fuzzy inference system. An approach for the recognition of static alphabetic signs of Spanish sign language is addressed in [9]. Rokade et.al. [10] used thinning method for the recognition of American Sign Language numbers.

A video based Indian sign language recognition system was developed by P. V. Kishore and Rajesh Kumar using wavelet transform and fuzzy logic [11]. They used wavelet based video segmentation technique to detect shapes of various hand signs. Shape features of hand gestures are extracted using elliptical Fourier descriptions and PCA is used to reduce the dimensionality of the feature space. Recognition of gestures from the extracted features is done using a fuzzy inference system. J. Rekha et.al. used a feature descriptor which is a combination of shape, texture and local movements of hand features for the recognition of Indian sign language alphabet. The shape, texture and finger features of each hand are extracted using Principle Curvature Based Region (PCBR) detector. Wavelet Packet

Decomposition (WPD-2) and complexity defects algorithms respectively for hand posture recognition process. To classify each hand posture, multi class non linear support vector machine (SVM) was used [12].

Geetha M. and Manjusha U. C. proposed a method [13] for the recognition of Indian sign language alphabet and numerals using B-Spline approximation. Their algorithm approximates the boundary extracted from the region of interest, to a B-Spline curve by taking the maximum curvature Points as the control points. Then the B-Spline curve is subjected to iterations for smoothing resulting in the extraction of key maximum curvature points, which are the key contributors of the gesture shape. Support vector machine (SVM) is used as the classification tool in this method.

3 PROPOSED METHOD

In this paper, we propose a method that recognizes the fingerspelling in Indian sign language. The proposed method has three major steps.

1. Hand Segmentation
2. Feature Extraction
3. Gesture Classification

3.1 Hand Segmentation

Image segmentation is the process of extracting the object of interest from an image. Efficient and accurate segmentation of hand region has a key role in sign language recognition. There exist many different algorithms for segmenting the hand region from the captured image. Among them the skin colour based segmentation is considered to be the best due to its invariance to geometrical transformations and low computational complexity.

Colour models represent a colour in a standard way. Many different colour spaces are employed in the literature, including RGB, YCbCr, and HSV etc. for skin detection application. RGB is the most commonly used colour space for storing and representing digital images. A pixel (x, y) belongs to skin if its (R, G, and B) components satisfy the following conditions:

$$R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and} \\ \text{Max}\{R, G, B\} - \text{Min}\{R, G, B\} > 15 \text{ and} \\ |R - G| > 15 \text{ and } R > G \text{ and } R > B.$$

YCbCr colour model separates RGB to luminance and chrominance components where Y is the luminance component and Cb, Cr are the chrominance components. RGB values can be transformed to YCbCr colour space using the following equations:

$$Y = 0.299R + 0.587G + 0.114B, \\ Cr = 128 + 0.5R - 0.418G - 0.081B, \\ Cb = 128 - 0.168R - 0.331G + 0.5B.$$

Hand region can be extracted by applying a thresholding operation on the pixel values of the YCbCr image. If the Y, Cb and Cr components of a pixel are within a predefined range of the skin colour, that pixel is set as white otherwise black. Thus a pixel is classified as belonging to skin if it satisfies the following relation:

$$75 < Cb < 135 \text{ and } 130 < Cr < 180 \text{ and } Y > 80.$$

HSV colour space also separates the RGB into luminance and chrominance components. Like YCbCr colour space it

eliminates the redundancy present in the RGB colour space. The transformation of RGB to HSV is invariant to high intensity at white lights, ambient light and Surface orientations relative to the light source. So HSV can be used as a very good choice for skin detection methods. A pixel in the HSV colour space is classified as skin pixel if it satisfies the following condition:

$$(s \geq .05 \ \& \ s \leq 1 \ \& \ (h \geq 0.7 \ \& \ h \leq 4 \ | \ h \geq 0 \ \& \ h \leq 0.6))$$

Although the skin colour based method gives good segmentation results the accuracy of segmentation may vary depending on the difference in human skin colour and the different lighting conditions. The combination of all the above colour spaces like RGB, YCbCr, and HSV gives a better skin detection performance. In the proposed method we have merged the results from all the three colour spaces to extract skin pixels and got better segmentation results.

The result of segmentation gives a binary image in which the pixels that define the hand region are in white colour and all others are in black colour. After skin detection all pixels may not be classified correctly due to noise or segmentation errors. Filtering and morphological operations are applied on the segmented image to decrease noise and segmentation errors if any.

3.2 Feature Extraction

Selection of good features has a significant role in any object recognition task. For robust recognition of gestures the hand sign in the captured image should be described and represented in a well defined manner. Shape is an important visual feature that characterizes each hand sign. So a shape based feature derived from the contour based potential energy is used in this work for differentiating various gestures.

3.2.1 Contour Potential Energy

The concept of potential energy is derived from physics. Potential energy of an object arises from the relative position of the object. Image potential energy is calculated from the positions of individual pixels in the image. It describes the fundamental feature of an image and is a new method for image feature processing and analysis. But the computation of the potential energy from the original image needs more time and storage space. Thus, to get better effect of using potential energy theory in image feature extraction the contour of the image is extracted first and calculates the potential energy from the image contour. The contour based potential energy is calculated as follows:

1. Extract the boundary of the segmented binary image.
2. Set the bottom and left border of the contour image as zero potential energy lines.
3. Find the sum of the pixel values multiplied by the pixel positions in the rows and columns of the contour image. This step returns two vectors.
4. Row vector R, where each element in this vector is the weighted sum of the pixel values in the corresponding row of the contour image.

5. Column vector C, where each element in this vector is the weighted sum of the pixel values in the corresponding column of the contour image.
6. The row vector and the column vector are the potential energy to the bottom and left border and is known as the two dimensional potential energy.

These two vectors are the one dimensional functions that characterize the hand shape in the input image. They are considered as the descriptors of the hand sign representing a particular gesture. These descriptors are subjected to further processing for making them robust and noise insensitive.

3.2.2 Fourier Descriptors

Fourier transform coefficients of the shape descriptors form the Fourier descriptors. In the proposed method, Fourier descriptors of the two potential energy vectors are calculated. The resultant Fourier transforms coefficients form the Fourier descriptors of the hand shape. These Fourier descriptors represent the feature vectors of the hand shape in the frequency domain. Fourier descriptors have strong discrimination power and less sensitivity to noise. Moreover, Fourier descriptors are information preserving and they can be normalized easily.

3.2.3 Feature Vector

The feature values are formed from the Fourier descriptors of the potential energy vectors by taking only the magnitude of the Fourier coefficients and ignoring the phase information. The feature values are normalized by dividing the magnitude values of all the Fourier coefficients by the magnitude value of the first coefficient which is called the DC component. Although the number of coefficients generated from the Fourier transform is usually large, a set of parameters describing the distribution of these coefficients are enough to capture the overall features of the shape. The feature vector for each gesture is composed of a set of six feature values which are the second, third and fourth central moments of the normalized Fourier coefficients of the row and column potential energy vectors. Thus a feature vector with six feature values that uniquely represent each hand sign is obtained.

3.2.4 Central Moments

Central moments are the moments about the mean. Central moments are a set of values by which the properties of a probability distribution can be usefully characterized. The higher order central moments are only related to the spread and shape of the probability distribution, rather than to its location. So they are preferred to ordinary moments for describing the probability distribution. For a real-valued random variable X, the kth moment about the mean $\mu_k = E[(X - \mu)^k]$ or k^{th} central moment is given by $\mu_k = E[(XE[X])^k]$, where E is the expectation operation. The first few central moments have intuitive interpretations: The "zeroth" central moment μ_0 is one. The first central moment μ_1 is zero. The second central moment μ_2 is called the variance, and is usually denoted as σ^2 , where σ represents the standard deviation of the distribution. The third central moment μ_3 and fourth central moment μ_4 are used to define the standardized

moments skewness and kurtosis respectively.

Variance is a measure of the dispersion of the data in a sample. It is a good descriptor of the probability distribution of a random variable. It describes the spread of the numbers from the mean value. In particular, variance is the second order moment of a distribution. Thus, it can be used as a parameter for distinguishing between probability distributions

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can have positive or negative values, or it can be undefined also. If the tail on the left side of the probability density function is longer than the right side, it indicates a negative skew. In this case, the bulk of the values possibly including the median lie to the right of the mean. A positive skew results when the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. When the values are evenly distributed on both sides of the mean, the skewness becomes zero.

Kurtosis is a measure of the "peakedness" of a probability distribution. It is a characteristic of the distribution of a real-valued random variable. Similar to the concept of skewness, kurtosis is also a descriptor of the shape of a probability distribution.

3.3 Gesture Classification

The feature vectors obtained from the feature extraction step are used as the input to the classifier that recognizes different signs. Artificial Neural Network (ANN) is used as the classification tool. The processing elements in an artificial neural network are artificial neurons which mimic the biological neurons. ANN is an adaptive system that learns to perform a function from an input-output map. Classification process involves two steps, training and recognition. The feature values extracted from the training image set are used for training the feed forward neural network. The system parameters are changed during training phase. In recognition phase the neural network parameters are fixed and the system is deployed to solve the problem.

The proposed method utilizes a feed forward neural network with back propagation learning mechanism for classifying the gestures. Feed forward neural network has a layered architecture with one input layer, one output layer and one or more hidden layers in which each layer is fully connected to the following layer. Back propagation is the most commonly used learning scenario for training the feed forward neural network. It works by the principle of "backward propagation of errors". Backpropagation is a supervised learning technique and the network is provided with the pairs of inputs and outputs that the network has to learn. The input patterns are given to the network through the neurons in the input layer and the output of the network is obtained through the neurons in the output layer. Then the backpropagation algorithm computes the difference between actual and expected results and this error value is propagated backwards. The backpropagation algorithm tries to minimize this error until the neural network learns the training data.

The number of input neurons is determined by length of the feature vector. The total number of hand signs

determines the number of neurons in the output layer. The number of neurons in the hidden layers is obtained by trial and error. The most compact network is chosen and presented. Neural network training was done using the matlab neural network toolbox.

4 EXPERIMENTAL RESULTS

The implementation results of the proposed sign language recognition method are given in this section.

4.1 Data Collection and Pre-processing

Images of the ISL alphabet and numerals are collected for conducting the experiment. There are no resources to download the dataset for ISL letters and numerals. So the dataset is created in our lab with suitable lighting and environmental setup. Image capturing is a random process. The resolution of various image capturing devices may not be the same. This results in different resolution of the captured images. For accurate comparison of features and to reduce the computational effort needed for processing, all the images should be scaled to a uniform size. Similarly the orientation of the objects in the captured images may not be the same always. In order to ensure the reliability and to enhance the robustness of gesture recognition, the images must be subjected to coordinate adjustments. For this, the object's major axis should be made parallel to the X-axis of the coordinate system.



Fig 2. Sample Dataset

4.2 Gesture Recognition

The proposed method has been implemented and tested using MATLABR2010a in a machine with Intel i3, 2.2GHz processor and 4GB RAM. A database containing 540 images with 15 images of each of the 36 signs are used for conducting the experiment. 10 images of each sign are used for training the system and 5 images of each sign are used for testing the system. A feed forward neural network is trained with the feature values from the training image set. The network is tested using the feature values extracted from the test images. The accuracy of the proposed method is estimated as the ratio of the number of signs recognized correctly to the total number of signs in the test dataset.



Fig 3. Hand Segmentation Results

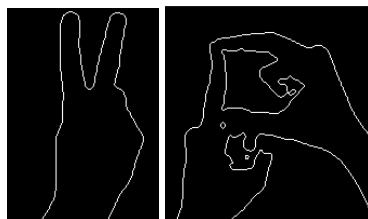


Fig 4. Binary Image Contour

We have got an average recognition rate of 92.22% from the experiment. This indicates a good recognition result when considering the large variety of signs in the dataset. The results are depicted in TABLE 1.

TABLE 1
PERFORMANCE OF THE PROPOSED METHOD

No. of Test Images	Correctly Classified	Wrongly Classified	Accuracy (%)
36 x 5 = 180	166	14	92.22

5 CONCLUSION

This paper presents a method for automatically recognizing the fingerspelling in Indian Sign Language. The signs are identified by the features extracted from the hand shapes. We used skin colour based segmentation for extracting the hand region from the captured image. Contour potential energy based feature is used to describe each hand sign. The features from the gesture images are used for training a feed forward neural network. The trained network can be used for the recognition of different hand signs. The method is implemented completely by utilizing digital image processing techniques so the user does not have to wear any special hardware device to get the features of the hand shape.

REFERENCES

[1] Adithya V, Vinod P R, Usha Gopalakrishnan "Artificial Neural Network Based Method for Indian Sign Language Recognition," 2013 IEEE Conference on Information and Communication Technologies.
[2] T.Starner and A. Pentland, "Real-time American sign language recognition from video using hidden markov models", Technical

Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.
[3] Deng-Yuan Huang1, Wu-Chih Hu2, Sung-Hsiang Chang , "Vision-based Hand Gesture Recognition Using PCA + Gabor Filters and SVM" ,2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.
[4] Rini Akmeiliawati, 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.
[5] Nadia R. Albelwi, Yasser M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition "International Conference on Communications and Information Technology 2012.
[6] R Stephan Liwicki, Mark Everingham, "Automatic Recognition of Fingerspelled Words in British Sign Language", School of Computing University of Leeds
[7] Azadeh Kiani Sarkalehl Fereshteh Poorahangaryan, Bahman Zan, Ali Karami, "A Neural Network Based System for Persian Sign Language Recognition" IEEE International Conference on Signal and Image Processing Applications 2009.
[8] Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems,"The Journal of Artificial Intelligence 133 (2001) 117-138.
[9] Incertis, J. Bermejo, and E. Casanova, "Hand Gesture Recognition for Deaf People Interfacing," The 18th International Conference on Pattern Recognition, 2006 IEEE.
[10] R. Rokade, D. Doye, and M. Kokare, "Hand Gesture Recognition by Thinning Method" International Conference on Digital Image Processing (2009).
[11] P. V. V. Kishore and P. Rajesh Kumar, "A 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
[12] J. Rekha, J. Bhattacharya, S. Majumder, "Improved Hand Tracking and Isolation from Face by ICondensation Multi Clue Algorithm for continuous Indian Sign Language Recognition", Surface Robotics Laboratory, Central Mechanical Engineering Research Institute, CSIR.
[13] Geetha M, Manjusha U. C, "A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation" International Journal on Computer Science and Engineering (IJCSE), March 2012.