This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.

Spotting and Recognition of Hand Gesture for Indian Sign Language Recognition System with Skin Segmentation and SVM

S. Reshna[1] and M. Jayaraju[2]
[1]ECE, T K M College of Engineering, Kollam, Kerala
[2]M E S Institute of Technology and Management, Kollam, Kerala
Email: [1]reshna.s@tkmce.ac.in [2]jayarajum@gmail.com

***Abstract*—The hearing impaired people use sign language for communication. It relates letters, words, and sentences of spoken language to gestures, helping them to communicate among themselves. An automation system that can convert sign language to spoken language can help the hearing impaired community to intermingle with normal people. It will aid them to strengthen their abilities and make them aware that they can do better for the growth and development of humanity. In this paper we have tried to develop an automation system to recognize sign language in complex backgrounds. Video of the signer is taken and part of the hand showing the sign is segmented with respect to skin colour. Features that can identify the sign are extracted from the hand image and classified to recognize the sign. The classification is done using support vector machine.**

***Index Terms*—Sign Language (ISL), Sign Language Recognition (SLR) System, Skin Segmentation, Support Vector Machine (SVM), YCbCr Color Space.**

I. Introduction

Hearing impaired becomes isolated and lonely as normal people fail to interact with them due to their ignorance of sign language. This will affect their social and working life. A translator is desirable when a person wants to converse with hearing impaired, but there is a lack of availability of such experienced and educated interpreter. An automation system that can translate sign language into spoken language can aid the hearing impaired in interacting with normal people. It can help them to enhance their abilities and make them capable of doing better things for the growth and development of mankind. Recognition of a sign language is significant from the technical point of view and also for its impact on the society.

Sign languages are natural languages used by deaf and dump which uses various means of expression for communication in everyday life. It relates letters, words, and sentences of a spoken language to gestures facilitating hearing impaired people to communicate among themselves. Every region has their sign language. It is used by the deaf people, parents of the deaf children, children of deaf parents and deaf educators.

The native language commonly practiced by the deaf community of India is the Indian Sign Language. It allows a deaf to convey thoughts and ideas using hands, arms, and face.

ISL shares the grammatical features like the use of space and simultaneity, hand shape, hand location, hand orientation, movement and non-manual features like body position, movement of head and facial expression with other sign languages.

India is diversified in language, culture, and religious belief. So no standard sign language is adopted in India. Various dialects of ISL with etymological variation are found in various parts of the country. Still, the language construction is similar for most of the gestures. Phonological work on ISL started in 1970's. In 1977, with support from the National Science Foundation (USA) Vasishta, Woodward, and Wilson visited India and collected signs from different centres in India for linguistic analyses. Subsequent efforts by Vasishta et al [1] stemmed in lexicons of ISL regional varieties and articles.

The organization of the paper is as follows: Section II discusses the method with various techniques available in the literature. Section III describes the proposed sign language recognition system with the algorithm for skin segmentation and support vector machine. Section IV focuses on the experimental outcomes and Section V is discussion and conclusion.

II. Literature Survey

Study of Indian Sign Language, its dialects and variations and efforts taken in standardization of ISL are done in [2]. This paper proposed a methodology to distinguish static single hand gestures from a set of Indian sign language. The system they proposed is for signer-independent and utilizes a single web camera for acquiring sign.

J. Bhattacharya J. Rekha, and S. Majumder Scientists of Surface Robotics Laboratory, Central Mechanical Engineering Research Institute (CMERI-CSIR) Durgapur proposes an automatic gesture recognition approach for ISL recognition [3]. Skin color segmentation using YCbCr skin color model to segment the hand region, feature extraction by Principle Curvature Based Region (PCBR) detector and Multi class nonlinear support vector machines (SVM) recognition is proposed and a rate of 91.3% is achieved. Another approach to recognize sign language gestures in a real time environment using SURF & Hu Moment Invariant methods [4], K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) are are developed by the same team. In addition, using Hidden Markov Model (HMM) a lexicon based method for finger spelled word recognition is developed. They also proposes another system

978-1-5090-4442-9/17/$31.00 ©2017 IEEE

for hand tracking and isolating face from hands [5] using ICondensation algorithm and BRIEF feature descriptors.

Research work on ISL recognition system is taking place at the Department of Electronics and Communication Engineering, Indian Institute of Technology Guwahati and Indian Institute of Information Technology, Allahabad. An algorithm for automatic recognition of continuous Indian Sign Language against various backgrounds using DWT for feature extraction and HMM for recognition is proposed in [6]. Experiments are performed on numerous backgrounds like colored one, one with multiple objects etc. and this is found to have vary less time complexity and space complexity. A system that can handle both static and dynamic hand gestures in vision based platform is proposed in [7]. The palm and finger movements are represented as 'shape change' of the hand. The global hand movement is represented as motion trajectory in 2D space.

A spatio-temporal feature-extraction techniques with applications to online and offline recognitions of isolated Arabic Sign Language gestures is proposed in [8]. A hand gesture recognition system to recognize continuous gesture before stationary background using HMM with recognizing rate is above 90% is developed to recognize 20 different gestures in [9]. A real-time methodology for the spotting, representation, and recognition of hand gestures from a video stream for human–computer interaction is proposed in [10]. Hand gesture spotting and tracking is done with skin color analysis and spatiotemporal motion detection with Motion-Feature-Warping Algorithm. The recognition rate was over 90% with a library of 12 gestures.

Segmentation of the Face and Hands in the video sequences are done by Color and Motion Cues [11] by Nariman Habili et al. Skin Detection Mask (SDM) is formed by classifying image pixels as skin or non skin based on the Mahalanobis distance and Change Detection Mask (CDM) is developed by conducting the F test to focus moving objects in the vedio sequence. Finally SDM and CDM are fused together to create a face and hand segmentation mask (FHSM).

III. Proposed Work

Indian Sign Language gestures are formed by incorporating hand shapes, hands movement, hand and head orientation and location and facial expressions. The aim of sign language recognition system is to reproduce verbal or text related to the given sign. Research on SLR involves computer vision, pattern recognition, natural language processing and psychology.

The acquisition of images using a video camera is the first step in this process. We can use a handycam or webcam or even mobile camera for acquiring the sign gesture. To determine the sign, the human body position, configuration (angles and rotations), and movement (velocities) need to be sensed. In the proposed system, we are using the camera of our laptop.

The video taken is divided into frames and each frame have to be processed alone and the hand frame is treated as a posture and segmented. The video of ISL gestures are composed of various moving entities besides the hand and head of the signer. The signer's hand and head are the foreground image and rest of the objects are the background image in the video sequence. The unwanted objects in the background are removed by background subtraction algorithm and skin segmentation is done to segment hands and head of signers. Most of the existing skin segmentation techniques comprise the classification of individual image pixels on the basis of pixel color into skin and non skin categories because the human skin has consistent colors with different textures and are distinct from other colors.

Hand and head tracking are done to find the position of hand and head in video frame. A feature matrix for each sign is formed from shape features extracted from segmentation and location information from tracking and is stored in a database. It can be used as templates at training stage or as inputs to pattern classifiers. After modeling and analysis of the input hand image, gesture classification algorithm is used to recognize the sign and generate sound messages corresponding to the recognized sign.

Statistical tools and soft computing techniques are used for the recognition of each sign. Commonly used Statistical Tools are (i) Hidden Markov Model (HMM) (ii) Principal Component Analysis (PCA) (iii) Learning Vector Quantization (LVQ) (iv) Finite State Machine (FSM) (v) Support Vector Machine (SVM) (vi) K means Clustering. The Soft Computing Tools used are (i) Artificial Neural network (ANN) (ii) Fuzzy Means clustering (FMC) (iii) Genetic Algorithms (GA). In our system, classification is done by using SVM.

The challenge that must be tackled by any SLR system is to track the signer in the video with a variation of background clutter and lighting conditions. Britta and Karl-Friedrich [12] reported some problems associated with sign language:

i. The frame for Sign recognition may contain occlusions.
ii. The position of signer in front of camera may vary so that the camera can't acquire some extremes of the sign.
iii. Camera has lack of depth information.
iv. While performing the sign, speed and position of performing the sign may vary from person to person or with same person. This creates variation of sign image in time and space.
v. The frames between two signs should not be considered as sign and is to be removed (Co-articulation problem).

A. Skin Segmentation

In the proposed system, segmentation is done by skin detection using YCbCr color space. This method is based on the threshold value of the components Y, Cb and Cr.

The generic skin model is implemented by defining skin color range in a color spaces like RGB, HSV, YCbCr etc. We can represent the skin pixels in the RGB space [13] as follows:

(i) under uniform day light

$$\begin{aligned}&(R > 95)\ \&\ (G > 40)\ \&\ (B > 20)\ \&\\&(\max\{R,G,B\} - \min\{R,G,B\} > 15)\ \&\\&(|R-G| > 15)\ \&\ (R > G)\ \&\ (R > B)\ \&\ (G > B)\end{aligned} \tag{1a}$$

(ii) under the flashlight

$$(R > 220)\ \&\ (G > 210)\ \&\ (B > 170)\ \&$$
$$(|R - G| > 15)\ \&\ (R > G)\ \&\ (R > B)\ \&\ (G > B). \quad (1b)$$

In RGB color space the luminance (intensity) and chrominance (mixing of color) components are not decoupled and have non uniform characteristics. So it is not preferred for hand gesture recognition using color based detection and color analysis. Also, the feature space is for RGB is 3-D and that for chrominance is 2-D. In the YCbCr color space, Y is the luminance component and is scaled from 16 to 235. Cb and Cr, the chrominance components represents color differences B-Y and R-Y respectively and have a range of 16 to 240. The conversion from RGB to YCbCr can be done by the formula,

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

$$Y = 0.99R + 0.587G + 0.114B \quad (3a)$$
$$Cr = R - Y\ \ Cb = B - Y. \quad (3b)$$

The transformation and efficient separation of color and intensity information in YCbCr color space is easy as compared to other color models. This color space is effective and efficient for the recognition of image pixels in color images. This approach distinguish color and intensity information even under uneven illumination conditions. So YCbCr color space can be used for the complex images with different illumination.

Many algorithms have been proposed for skin color pixel classification like multilayer perceptron (ANN), piecewise linear classifiers, Gaussian classifiers and Bayesian classifier with the histogram technique. The decision boundaries range from simple shapes like rectangle and ellipse to more complex parametric and nonparametric forms. A color pixel x is considered as a skin pixel if

$$\frac{p(x/skin)}{p(x/non\ skin)} \geq T$$

where $p(x/skin)$ and $p(x/non\ skin)$ are the respective conditional probability density functions of skin and nonskin colors and T is threshold [14]. The value of T depends on the apriori probabilities of skin and nonskin values. The conditional probability density functions can be calculated using histogram.

The threshold value of the individual component of color pixels is the basis of skin detection using YCbCr color space. As RGB color space depends on non-uniform and device characteristics, it is not used commonly for color based analysis [15]. The algorithm for the detection of human skin color using YCbCr Color Space can be explained as follows.

1) Input image is the image obtained after preprocessing the output from video camera.

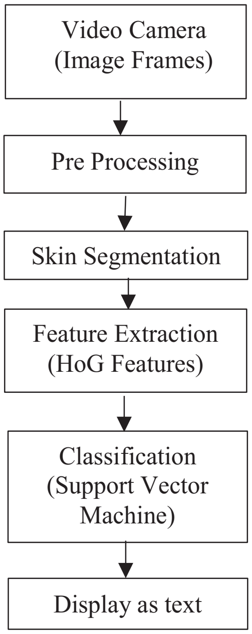

Fig. 1. Block diagram of the proposed system.

2) The above image in RGB color space is converted into YCbCr color space using (Eq. 2).
3) Histogram is computed for all three components and the threshold value are determined. The threshold for chrominance components are $150 < Cr < 200$ and $100 < Cb < 150$.
4) Spatial filtering is done for skin pixels in the test image.
5) Threshold is applied to the filtered image.
6) Smoothening is done on the threshold image.
7) The output image comprises of only skin pixels of the hand image.

B. Support Vector Machine (SVM)

The support vector machine (SVM) deals with binary classification problems (two class problems) in which the data are separated by a hyper plane defined by a number of support vectors. Support vectors are a subset of training data used to define the boundary between the two classes. Each instance in the training set contains one target value and features. The goal of SVM is to produce a model which predicts target value of data instance in the testing set which consists of only features (see Fig. 1).

SVM is a supervised machine learning method that produces input-output mapping functions from a set of training data [16]. The mapping function can be either a classification function or a regression function. For classification, nonlinear kernel functions are used to convert input data to a high-dimensional feature space where the input data is more separable with respect to the original input space. Fig. 2 shows the input plane and the hyperplane that separates the data into two spaces. Maximum-margin hyperplanes are then created. This model depends on only a subset of the training data near

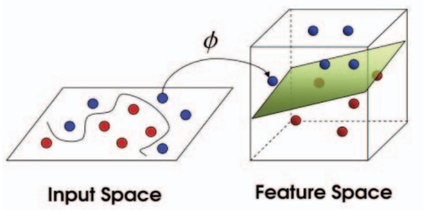

Fig. 2. Input space and the hyperplane that separates the data into two spaces.

the class boundaries. Similarly, the model produced by support vector regression ignores any training data that is sufficiently close to the model prediction. SVMs are also said to belong to "kernel methods".

Consider the case of classifying the set of training vectors belonging to two separate classes $\{x_1, x_2, \ldots\}$ which are vectors in R^A. Consider a decision function of the following form:

$$y\left(x\right) = W^T \varphi\left(x\right) + b$$

x_i is a class label, $t_i \in -1, +1$. a decision function is to be constructed such that, $y(x_i) > 0$ for all i such that $t_i = +1$, and $y(x_i) < 0$ for all i such that $t_i = -1$ that is,

$$t_i y\left(x_i\right) > 0, \quad \forall i.$$

SVM multi-class problem is used to decompose an M-class problem into a series of two-class problems. Let the jth decision function, with the maximum margin that separating the class i from the remaining classes, be

$$Y_j\left(x\right) = W_j^T \varphi\left(x\right) = b_i.$$

Here, w_i is n dimensional vector, $\varphi(x)$ is mapping function which maps x into n dimensional space. If the problem is linearly separable, the training data belonging to class k satisfy

$$y_k(x) = 0$$

and data belonging to other classes must satisfy

$$y_k(x) \leq 1.$$

In the case of inseparable problem unbounded support vectors satisfy the condition

$$(x) = 1$$

and bounded support vectors belonging to class k satisfy the condition

$$y_k(x) \leq 1$$

and other data belonging to other classes satisfy the condition

$$y_k(x) \geq -1.$$

This support vector classifier recognize the hand gestures based on the trained point features. In this paper, classification is done by using SVM technique with $C = 2.67$, $gamma = 5.383$.

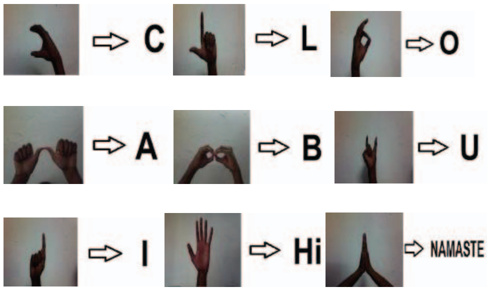

Fig. 3. Different set of training images.

IV. Experimental Results

Database is created by taking the training images for sign language recognition system. We have taken the 11 symbols in the ISL system. We have processed 500 training images for each symbol. The training images are resized (200 × 200) to a new dimension and renamed. These are automatically stored in an array. The images in the array are passed through a Sobel filter for detecting the edges. Sobel filter gives two types of edges, vertical and horizontal edges (i.e. in x and y direction). This Cartesian coordinates are converted to polar coordinates. Now the HoG features of the images are obtained through numpy library of python. Hence we have the samples of the training images. Classification is done by using SVM technique with $C = 2.67$, $gamma = 5.383$. Now we have different classes for different set of training images as shown in Fig. 3.

For the testing of the system real time video captured through the camera of the laptop. The video acquired through the camera is converted to frames simultaneously. The RGB frames obtained were converted to YCbCr images. Threshold was applied to these images for converting the image to a feature matrix using skin color detection. Bitwise AND operation is done between features of training and test images and a new image is obtained, which was resized. The comparison is done with help of cv2 library functions in python. The corresponding sign is predicted by comparing with the help of database stored. The predicted gesture is converted to corresponding text and displayed. Results are shown in Fig. 4.

Similarly 11 gestures are trained to detect. Training has been done in different background and postures. In total 200 training images were fed for each gesture. Video captured was then converted into frames and then these frames were compared with the trained images. If the 2 images were similar, result stored for each gesture is displayed. Gestures for A, B, C, I, L, Namaste, O, U, V, W, Hi, was trained to detect.

V. Conclusion

This work was successfully completed in Python background. We developed the system to work in complex backgrounds by segmenting the face and hand with respect to skin colour. Here we have considered the static gestures only but dynamic gestures can also be incorporated by comparing videos instead of image comparison. Similar

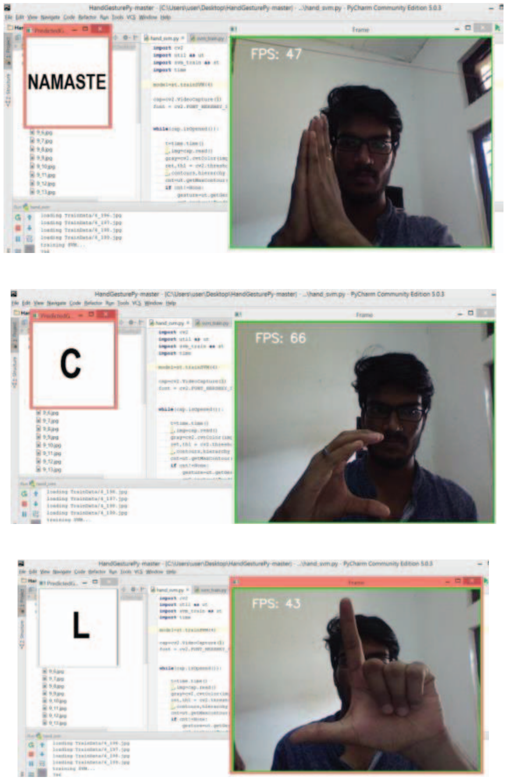

Fig. 4. Gesture for showing C, L and Namaste.

gestures were difficult to extract. By using better methods of feature extraction this can be done more accurately. Major disadvantage of this system is the lighting conditions in different situations. This can be avoided by using flash lights and more powerful camera. But it will increase the cost of the system.

India is diversified in language, culture, and religion. In India there is no standard sign language. Various dialects of ISL with verbal variation are found in various parts of the India. Even in the small state of Kerala, there are many versions of Sign language. We are trying to develop a sign language recognition system for recognizing the sign of our region.

References

[1] Ulrike Zeshan, Madan M. Vasishta, and Meher Sethna, "Developmental articles-implementation of indian sign language in educational settings," *Asia Pacific Disability Rehabilitation Journal*, 16, vol. 16, no. 1, 2005.

[2] Pravin R. Futane and Dr. Rajiv V. Dharaskar, "HASTA MUDRA - an interpretation of indian sign hand gestures," in *3rd International Conference on Electronics Computer Technology (ICECT)*, 8–10, vol. 2, Apr. 2011, pp. 377–380.

[3] J. Rekha, J. Bhattacharya, and S. Majumder, "Shape, texture and local movement hand gesture features for Indian sign language recognition," in *3rd International Conference on Trendz in Information Sciences & Computing (TISC2011)*, 8–9, Dec. 2011, pp. 30–36.

[4] J. Rekha, J. Bhattacharya, and S. Majumder, "Hand gesture recognition for sign language: a new hybrid approach," in *International Conference on Image Processing, Computer Vision and Pattern Recognition, IPCV'11*, Jan. 2011, pp. 80–86.

[5] J. Rekha, J. Bhattacharya, and S. Majumder, "Improved hand tracking and isolation from face by ICondensation multi clue algorithm for continuous indian sign language recognition," in *International Conference on Advanced Computing, Networking and Security ADCONS'11*, pp. 106–116, Berlin, Heidelberg: Springer-Verlag ©2012.

[6] Kumud Tripathi, Neha Baranwal, and G. C. Nandi, "Continuous dynamic indian sign language gesture recognition with invariant backgrounds," *Advanced computing, networking and security*, vol. 7135 of the series *Lecture notes in computer science*, pp. 106–116, 2015.

[7] M. K. Bhuyan, D. Ghosh, and P. K. Bora, "Hand motion tracking and trajectory matching for dynamic hand gesture recognition," *Journal of Experimental and Theoretical Artificial Intelligence,Taylor and Francis*, vol. 18, no. 4, pp. 435–447, 2006.

[8] Tamer Shanableh, Khaled Assaleh, and M. Al-Rousan, "Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 3, pp. 641–650, June 2007.

[9] Feng-Sheng Chen, Chih-Ming Fu, and Chung-Lin Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image and Vision Computing*, vol. 21, pp. 745–758, 2003.

[10] Yuanxin Zhu and Guangyou Xu, and David J. Kriegman, "A real-time approach to the spotting, representation, and recognition of hand gestures for human–computer interaction," *Computer Vision and Image Understanding*, vol. 85, pp. 189–208, 2002.

[11] Nariman Habili, Cheng Chew Lim, and Alireza Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 8, pp. 1086–1097, 2004.

[12] B. Bauer and Karl-Friedrich, "Towards an automatic sign language recognition system using subunits," *LNAI 2298, GW-2001*, Springer, pp. 34–47.

[13] Junwei Han, George M. Award, Alistair Sutherland, and Hai Wu "Automatic skin segmentation for gesture recognition combining region and support vector machine active learning," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06).*

[14] S. L. Phung, A. Bouzerdoum, & D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148–154, Jan. 2005. Copyright IEEE 2005.

[15] Khamar Basha Shaika, P. Ganesan, V. Kalist, B. S. Sathish, and J. Merlin Mary Jenithab, "Comparative study of skin color detection and segmentation in HSV and YCbCr color space," in *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015) Procedia Computer Science*, vol. 57, 2015, pp. 41–48.

[16] Andrew Ng, CS229 Lecture notes, Part V - Support Vector Machines.