Real-Time Recognition of Indian Sign Language

Muthu Mariappan H
Department of Computer Science and Engineering
National Engineering College
Kovilpatti, Tamil Nadu, India
0000-0001-7801-3512

Dr Gomathi V
Department of Computer Science and Engineering
National Engineering College
Kovilpatti, Tamil Nadu, India
0000-0003-3639-485X

***Abstract* – The real-time sign language recognition system is developed for recognising the gestures of Indian Sign Language (ISL). Generally, sign languages consist of hand gestures and facial expressions. For recognising the signs, the Regions of Interest (ROI) are identified and tracked using the skin segmentation feature of OpenCV. The training and prediction of hand gestures are performed by applying fuzzy c-means clustering machine learning algorithm. The gesture recognition has many applications such as gesture controlled robots and automated homes, game control, Human-Computer Interaction (HCI) and sign language interpretation. The proposed system is used to recognize the real-time signs. Hence it is very much useful for hearing and speech impaired people to communicate with normal people.**

Keywords – ISL, Sign language recognition, HCI, Fuzzy c-means clustering

I. INTRODUCTION

World Health Organization's (WHO) survey states that above 6% of the world's population is suffering from hearing impairment. In March 2018, the number of people with this disability is 466 million, and it is expected to be 900 million by 2050. Also, the 2011 census of India states that 7 million Indians are suffering from hearing and speech impairment. They do not think these impairments as disabilities; it is another way of a different life. However, their circle is very much limited. They should not be part of the deaf world alone, which seems cloistered sometimes. Text messaging, writing, using visual media and finger spelling are a few methods used to establish communication between normal and hearing and speech impaired people. However, they prefer sign language only because they can express their emotions and feelings through signs only. So conversing in their regional sign language brings more comfort for the people to share their ideas and thoughts among their near and dears.

Sign languages are a visual representation of thoughts through hand gestures, facial expressions, and body movements. Sign Languages also have several variants, such as American Sign Language (ASL), Argentinean Sign Language (LSA), British Sign Language (BSL) and ISL. The hearing and speech impaired people prefer the sign language, which is mostly used in their region. Moreover, in India, there is no universal sign language. Though there exist many sign languages, the normal people do not know about sign languages. Hence communicating with deaf and dumb people becomes more complex.

Recognition of sign language can be done in two ways, either glove based recognition or vision based recognition. In glove based technique a network of sensors is used to capture the movements of the fingers. Facial expressions cannot be recognized in this method and also, wearing a glove is always uncomfortable for the users. This method cannot be implemented massively since data gloves are very much expensive. So, the proposed system uses the non-invasive vision based recognition method. The vision-based recognition can be achieved in two ways. They are Static recognition or Dynamic recognition. In static recognition system, the input may be an image of hand pose. It provides an only 2D representation of the gesture, and this can be used to recognize only alphabets and numbers. For recognition of continuous sign language, the dynamic gesture recognition system is used. Here the real-time videos are given as inputs to the system, a sequence of hand movements form the gesture of the word/sentence. Information Technology with its modern methodologies such as artificial intelligence and cloud computing has an impressive role in enhancing intercommunication among people with vocal disabilities and normal people.

978-1-5386-9471-8/19/$31.00 ©2019 IEEE

II. RELATED WORKS

Hand gesture recognition is the key area of research for the past two decades. Researchers have done lots of research and tried variety of techniques for gesture recognition. Geethu Nath and Arun C.S. [1] developed a system using ARM CORTEX A8 processor for recognising the ASL symbols. The system uses Jarvis algorithm to recognize the numbers and template matching algorithm to recognize the alphabets. Kumud Tripathi [2] designed a system for recognising continuous ISL gestures using Principal Component Analysis (PCA) with various distance classifiers. From the own data set, the features from the keyframes are extracted using Orientation Histogram and given as input to the system. Manasa Srinivasa H.S. and Suresha H.S. [3] used the codebook algorithm for background subtraction and generated binary images from the given image frames. The binary image is used to calculate convex hull and convexity defects and depending upon the calculation of defect points the fingers which are unfolded are counted.

Joyeeta Singha and Karen Das [4] described a novel approach to recognize alphabets of ISL. An eigenvector-based technique is used for feature extraction, and eigenvalue weighted Euclidean distance technique is used for Classification of 24 different ISL alphabets. Archana S. Ghotkar and Gajanan K. Kharate [5] explored rule-based and dynamic time warping (DTW) based method to recognize ISL words. Their experiments proved that the performance of DTW is very much higher for continuous word recognition. M.K. Bhuyan [6] segments frame into video object planes (VOPs), to obtain a semantically meaningful hand position. Key VOPs and temporal information are tracked to form a complete gesture sequence. The test results concluded that by using keyframes, a gesture could be uniquely represented as a finite state machine with keyframes and corresponding frame duration as states. M.M.Gharasuie and H.Seyedarabi [7] proposed a system that recognizes the real-time hand gestures of numbers from 0 to 9 using Hidden Markov Models (HMMs).In this proposed system, preprocessing and tracking steps took place in hand trajectory extraction phase and feature extraction is taking place in the classification phases. The system is capable of providing 93% recognition rate.

Kairong Wang [8] presented a Codebook (CB) modelling method and spatial moments for recognizing dynamic hand gestures. Background subtraction with skin colour detection is used for hand region segmentation. Palm centre is identified by spatial moments of hand contour, and fingers are tracked by a curvature-based algorithm. Francke H [9] developed a real-time hand gesture recognition system by using active learning and bootstrap training techniques. The usage of colour-based hand tracking and a multi-gesture classification tree increases the robustness of the system. This innovative use provides the system with 86% accuracy, better than the similar systems. Hari Prabhat Gupta [10] uses accelerometer and gyroscope sensors for recognizing continuous hand gestures recognition. The starting and ending points of meaningful gesture segments are developed using an automatic gesture spotting algorithm. This gesture code is compared with the gesture database using DTW algorithm to recognize the corresponding gesture. Noor Tubaiz [11] suggested sequential data classification using Modified k-Nearest Neighbor (MKNN) approach. The hand motions are sensed using data gloves. The raw data are augmented using window-based statistical features, which are computed from the previous raw feature vectors and future raw feature vectors. Based on the existing systems the proposed system has been developed with novel techniques to recognize words of Indian Sign Language (ISL).

III. METHODOLOGY

The proposed system has a camera unit for capturing the gestures of the hearing and speech impaired people. The real-time sign language recognition system was designed as a portable unit for more convenience of the users. The raw videos taken in a dynamic background is given as an input to the system. The image frames are resized to maintain the equality among all the videos. OpenCV (Open Source Library for Computer Vision) is used for feature extraction and video classification.

A. Preprocessing

During preprocessing, high-intensity noises from the video frames are eliminated. The first and foremost step in preprocessing is smoothing or

978-1-5386-9471-8/19/$31.00 ©2019 IEEE

blurring, and the most popular goal of blurring is to reduce noise. The blurred image is obtained by performing a convolution operation with a low-pass box filter. A 3x3 normalised box filter can be represented as: $K = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

Coloured object extraction can be achieved more easily in HSV colourspaces. Hence the images are converted from BGR colorspaces to HSV colorspaces with the range of H varies from 0 to 179, the range of S varies from 0 to 255, and the range of V varies from 0 to 255. At the end of the preprocessing, binary images are obtained where the white coloured area is the skin region, and the black coloured represents the rest.

B. Morphological Transformations

Morphological transformations are operated on the binary images based on the shape of the image. It requires the original image and the structuring element or kernel as inputs. Erosion and Dilation are the two basic morphological operators. Erosion removes all the noises near the edges, based on the kernel size. Thus erosion can be very much useful in removing small noises from the foreground. The erosion is followed by dilation. It increases the foreground object or the white coloured region in the output image, because the object may shrink while eroding.

Let E be a Euclidean space and A be a binary image in E and B be the structuring element. Then,
Erosion of A by B is,

$$A \ominus B = \bigcap_{b \in B} A_{-b}$$

Here A_{-b} denotes the translation of A by –b

Dilation of A by B is,

$$A \oplus B = \bigcup_{b \in B} A_b$$

A_b is the translation of A by b.

C. Background Noise Removal

The two morphological operations are repeated until a clear foreground object is extracted. While performing the morphological operations, the selection kernel depends upon the needs of the system, and it may be created manually using the OpenCV module. Morphological operations along with median blurring achieve high efficiency in noise removal. In the proposed system a 5x5 elliptical kernel as shown below is used.

$$MORPH\_ELLIPSE = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The median blurring technique is very much efficient in removing the salt and pepper noises in the image. In median filtering, the mid value is updated as the median of all neighbouring pixels. After applying morphological operations and median blurring, a simple threshold function is used to obtain the final image after preprocessing.

D. Finding Contours

In the proposed system contours are used for detecting the object. Contour is a curve that joins all the points in the edges, having same colour or intensity or a contour refers to the outline or silhouette of an object.

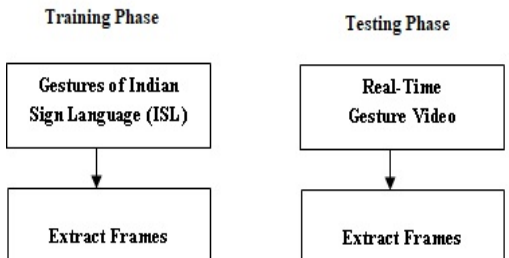
Contours work very well on binary images. Hence threshold or canny edge detection is applied before finding contours. Contours are a list of the entire contour in the image. Each contour is an array of (x, y) coordinates of boundary points of the object. Area of all the contours is calculated and using them the top three contours are selected. Those three contours represent Face, Left and Right hand, which contributes to the gestures.

E. Feature Extraction

Feature selection and extraction are very crucial steps in an image processing application. The most relevant features should be identified and extracted for the correct functionality.

Criteria for feature selection/extraction:

- ✓ Either improve or maintain the accuracy of the classifier
- ✓ Simplify the complexity of the classifier



978-1-5386-9471-8/19/$31.00 ©2019 IEEE

Fig.1. Flow Diagrams of Training and Testing

These two criteria must be satisfied while doing feature selection and extraction. The top three contours obtained earlier completely covers the Regions of Interest (Face, Left hand and Right hand). The required features are extracted from these regions as vector features for each frame in a video.

F. Fuzzy Clustering

Clustering is known as the process of grouping of similar data items together, while the items in the other clusters are as dissimilar as possible. In fuzzy clustering, the data items may belong to more than one cluster. Among several fuzzy clustering algorithms, fuzzy c-means clustering (FCM) algorithm is used most widely, and this can be used for both supervised learning and unsupervised learning, depending upon the needs.

The FCM, partitions, n data elements $X=\{x_1, x_2, ...x_n\}$ into c clusters $C = \{c_1, c_2, ...c_c\}$ based on the criteria used.

The partition matrix,

$$W = w_{i,j} \in [0,1], i = 1,...n, j = 1,...c$$

Where w_{ij} is the membership of the x_i in c_j .

The FCM minimises the following objective function,

$$\arg\min_{C} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| x_i - c_j \right\|^2$$

Where m > 0, and the fuzzy partitioning is carried out by an iterative optimisation of the above function with the update of w_{ij} .

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}}$$

Here k is an iteration step. During training, the extracted features are given to the c-means algorithm and it partitions the input data items into a specified number of clusters. During testing, it matches the test file with the existing clusters and returns the id of the cluster centre which has the highest degree of membership.

IV. EXPERIMENTAL ANALYSIS

The data samples are collected for 80 words and 50 sentences of everyday usage terms of ISL. The videos are recorded from ten volunteers of our collaborator school, using a digital camera.

Table.1: Distribution of Dataset

Sign Type	Number. of Signs	Number of signers	Total Samples
Word	80	10	800
Sentence	50	10	500

The data collection camp is planned for two sessions, where the samples for 40 words and 25

978-1-5386-9471-8/19/$31.00 ©2019 IEEE

sentences are recorded in each session. At the earlier stage, the system was developed to recognize 40 words. Eight samples of each sign were used for training, and two samples were used for testing.

Table.2: Samples from Dataset

Word	Book, Note, Pen, Face, Father, Mother, Brother, Soap, Temple, School
Sentences	What is your name? How are You? Where are you going?

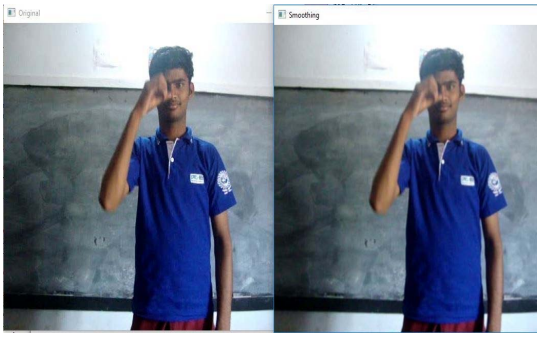


Fig.2. Comparison of the original frame (left) and smoothened frame.

In Fig. 2, the right side image is obtained by applying smoothing or blurring technique to reduce the noise in the image.

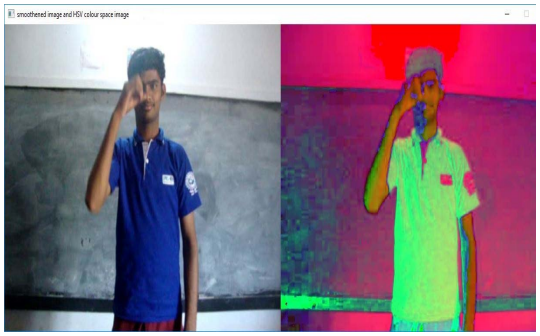


Fig.3. HSV colorspaces conversion

BGR image is converted into HSV colorspaces since coloured object tracking can be done effectively in HSV colorspaces and the result is displayed in Fig. 3.

The morphological operations are performed on the HSV image to remove the noises present in the foreground. The morphological transformation gives the binary images as shown in Fig. 4. Here the white region represents the skin area and rests are in black.

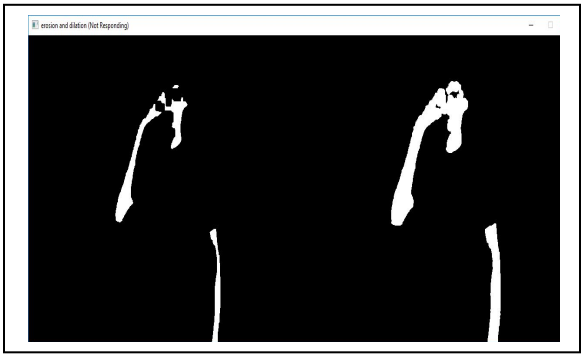


Fig.4. Erosion (left) and Dilation (right)

The contours are used to track the foreground object (skin area). All the contours in the current frame are identified; among them, the top three represents our ROI. Fig. 5, Depicts the contours covering the face and both hands.

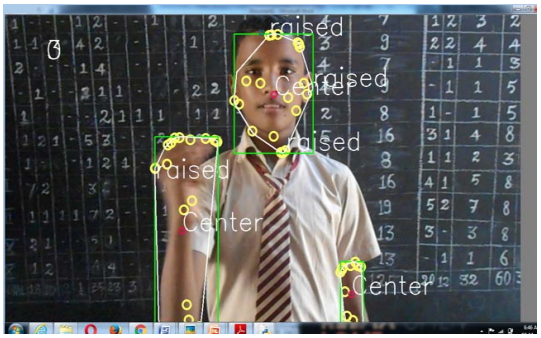


Fig.5.Segmentation of ROIs with contours

The features such as number of points in the convex hull, number of defect points and distance from the centre to each finger are extracted from the Regions of Interest, through these three contours, and the orientation between the contours is also kept track.

The FCM algorithm is used to group similar data items. FCM assigns membership to data points, corresponding to the cluster centres. More the data point is near to the cluster, the higher the value of the membership. The sum of membership of data points one. Membership and cluster centres are updated iteratively. Finally, the algorithm returns cluster centres and membership of each data points. By

978-1-5386-9471-8/19/$31.00 ©2019 IEEE

using these, the Fuzzy c-means prediction algorithm classifies the new data items. The cluster with the highest membership for the corresponding data points is chosen as gesture id. The identifications of the gestures are made by using this gesture id.

V. RESULTS

This FCM based real-time sign language recognition system, for recognising the words of Indian Sign Language has produced 75 % accuracy in gesture labelling and this is somewhat higher than the similar systems. Also, the developed system is much better than other systems, since it is capable of recognising 40 words of ISL in real-time while the similar systems have the capability to recognize static gestures only. The FCM is more efficient and reliable than the other clustering algorithms in many applications by its performance.

VI. CONCLUSION

The system for recognizing real-time Indian Sign Language (ISL) portrays an impressive role in enhancing casual communication among people with hearing disabilities and normal persons. Though FCM is efficient, it requires more computation time than the others. Also, for high dimensionality datasets, most of the traditional algorithms suffer. Hence it is planned to extend the system by combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to capture the spatial and temporal features. In future work, more words will be added to the system.

VII. ACKNOWLEDGEMENT

Sincere thanks to EPICS in IEEE for providing the initial funding to develop this assistive product. The research team appreciates and heartily thanks the high school volunteers for their contribution to the dataset.

REFERENCES

[1]. Geethu G Nath and Arun C S, "Real Time Sign Language Interpreter," 2017 International Conference on Electrical, Instrumentation, and Communication Engineering (ICEICE2017).

[2]. Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 523 – 531.

[3]. Manasa Srinivasa H S and Suresha H S, "Implementation of Real Time Hand Gesture Recognition," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 5, May 2015.

[4]. Joyeeta Singha and Karen Das, "Automatic Indian Sign Language Recognition for Continuous Video Sequence," ADBU Journal of Engineering Technology 2015 Volume 2 Issue 1.

[5]. Archana S. Ghotkar and Gajanan K. Kharate, "Dynamic Hand Gesture Recognition and Novel Sentence Interpretation Algorithm for Indian Sign Language Using Microsoft Kinect Sensor," Journal of Pattern Recognition Research 1 (2015) 24-38.

[6]. M.K. Bhuyan, "FSM-based recognition of dynamic hand gestures via gesture summarization using key video object planes," World Academy of Science, Engineering and Technology Vol: 6 2012-08-23.

[7]. M.M.Gharasuie and H.Seyedarabi, "Real-time Dynamic Hand Gesture Recognition using Hidden Markov Models," 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP).

[8]. Kairong Wang, Bingjia Xiao, Jinyao Xia, and Dan Li, "A Dynamic Hand Gesture Recognition Algorithm Using Codebook Model and Spatial Moments," 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics.

[9]. Francke H., Ruiz-del-Solar J. and Verschae R., "Real-Time Hand Gesture Detection and Recognition Using Boosted Classifiers and Active Learning," Advances in Image and Video Technology. PSIVT 2007. Lecture Notes in Computer Science, vol 4872. Springer, Berlin, Heidelberg.

[10]. Hari Prabhat Gupta, Haresh S Chudgar, Siddhartha Mukherjee, Tanima Dutta, and Kulwant Sharma, "A Continuous Hand Gestures Recognition Technique for Human-Machine Interaction using Accelerometer and Gyroscope sensors," IEEE Sensors Journal (Volume: 16, Issue: 16, Aug.15, 2016)Page(s): 6425 – 6432.

[11]. Noor Tubaiz, Tamer Shanableh, and Khaled Assaleh, "Glove-Based Continuous Arabic Sign Language Recognition in User-Dependent Mode," IEEE Transactions on Human-Machine Systems, Vol. 45, NO. 4, August 2015.

978-1-5386-9471-8/19/$31.00 ©2019 IEEE