International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 07 | July 2019 www.irjet.net p-ISSN: 2395-0072

Gesture Recognition for Indian Sign Language using HOG and SVM

[1]Manjushree K, [2]Divyashree B A

[1]Student, Department of Computer Science and Engineering, BNMIT, Bengaluru, India
[2]Professor, Department of Computer Science and Engineering, BNMIT, Bengaluru, India

---

ABSTRACT: *Gesture is a distinct form of sign language which involves movement of body such as hands or face to express the meaning. Hand gesture has gained immense importance now days, reason being the problems faced during the communication between mute people and normal people. The motivation behind is to develop a successful human machine interfaces system to recognize Indian Sign Language (ISL) under ambiguous conditions. The proposed work mainly aims to classify the single handed sign language by using histogram of oriented gradients feature matching and support vector machine classifier. The Natural Language Processing (NLP) toolkit is used in order to evaluate the spelling of sign words. The english sign alphabet (excluding J and Z) dataset which has a complex background are trained using three different classifiers to predict the better classifier for testing. The overall accuracy achieved by histogram of oriented gradient features along with the SVM classifier was 97.1%, the recognised sign will be presented in form of text and speech.*

Keywords: Hand gesture recognition, static and dynamic images, Image pre-processing techniques, SVM, NLP and OpenCV.

I. INTRODUCTION

Hand gestures are generally used by the mute people to communicate in their daily life. Hand sign are the primary component of sign language. Impaired people often faces problem to convey their feeling to the normal people, therefore the human-computer interfaces system has to be implemented to help the impaired people in the society. The signs that are used by the impaired people can be visualized by the normal people to understand the signs that are used. Sign language has their own grammar and lexicon [6]. Hand gesture are quite similar to the signs where signs indicates the letters such as "C", "M" and "1" etc, but gesture indicates the specific symbols like married, food, women and drink etc,. Every gesture will have specific meaning to communicate and understand. They are two types of sign i) alternate sign: sign language that has a specific context which is of standard signs and ii) primary sign: sign that are communicate between group of people who belongs to same locality. Different part of the world makes use of different sign language like ASL (American Sign Language), CSL (Chinese Sign Language) and ISL (Indian Sign Language) etc, [1]. According to the survey in India 7.5% of people are dumb and 5.8% people are both dumb and deaf. They are many differences like grammar and structural compared to other nations. There are basically two approaches for decoding the hand signs, I) Sensor Based: It is difficult to process because it involves the huge hardware components and sensor based hand gloves which results in troubleshoot during pre-processing. II) Image Based: It is one of the easiest implementation and widely used method for recognizing the sign or gestures [8]. In sign recognition system, section II represents the Literature Review, section III presents the proposed approach and section IV shows the Experimental Results followed by conclusion and future work in section V. Standard single handed sign for an Indian english character from A to Z are shown in figure 1.

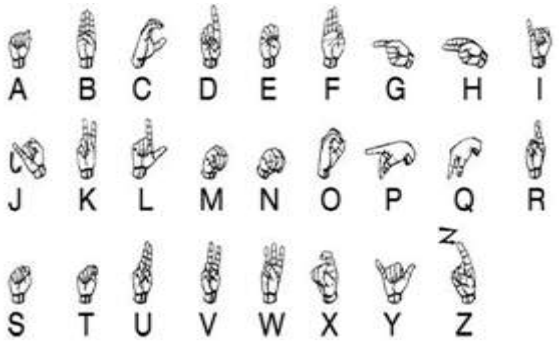

Figure 1: Single handed ISL English alphabets

II. LITERATURE REVIEW

Literature review explains the perspectives of the previous work carried out for Indian and other sign language recognition. Divya Deora and Nikesh Bajaj [1] presented segmentation to partition the images into different or multiple pixel and RGB image is detected to separate the images based on the colors to calculate the threshold frequency. Binarization technique concatenates single image which is red in color due to its high intensity. Finger-tip finding algorithm is applied to extract the four features such as thinning using distance formula, determining the two dimensional pixels, finding the edges and eliminating the angle between two fingers points. A clustering algorithm is used to calculate the image with the values obtained in finger-tip finding algorithm. Principal Component Analysis (PCA) tool is applied to calculate the eigen vector of the matrix of each frame and to recognize the sign correctly.

The authors describes the different steps and techniques used when performing the pre-processing in three important phase [2]. First phase is pre-processing the image with removing the background of an image, reducing the noise, luminous and gray scale conversion to quotation the required features from image which are further used for

classification. Second phase is classification phase which involves haar cascade algorithm to precisely classify the extracted features. Training and Testing are the two phase in classification. Training stage consists of 1000 images and 50 images are considered to test in separate folders which are trained with different size, orientations and color etc, stored in the database. Testing stage will make use of classifier to distinguish between the trained datasets and different signs to be tested and match the sign exactly. Third phase are speech synthesis phase used for converting the text into speech format or vice-versa. The speech is converted using OpenCV tool with a built-in function termed as system.speech.synthesis which gives the speech as the output to the recognised sign.

The number of gesture is reduced by classification the gestures into single and double hands and also solve the issues associated with it in each subcategory [3]. Morphological operation involves the process of dilation and closing operation for single and double handed gestures. Filtered binary images are used to extract the four geometric features: solidity, longest diameter and shortest diameter, conic section and bounding box ratio, the above obtained values are used to classify the single or double handed gestures. The distribution of orientation and magnitude of the images are the HOG features. Classification is done by applying K- Nearest Neighbouring (KNN) algorithm on the geometric feature extracted along with the HOG features. The comparison is also done with Support Vector Machine (SVM) to check for the better accuracy.

The human-computer interaction system is implemented on the mobile to help impaired people to communicate with society without ant barriers. Different techniques are used to build the automatic application which translates the sign into speech [4]. The image is processed using Hue, Saturation and Value (HSV) algorithm for skin detection. Two detection methods are i) Blob Detection: detect the largest blob by calculation the distance between centroid and other end point of the image ii) Contour Detection: the grayscale image is accepted and if it is greater than 200, holes is filled with white pixels. Eigen object recognizer is employed to extract the values for the pre-processed images and calculates the value of eigen image, distance threshold and average images to recognize the identical signs. To analysis the sign and convert into text and speech, the Principal Component Analysis (PCA) tool is used.

Hand glove are used to detect the gesture along with the sensor in-built, in order to recognise the movement of the hand. The recognised hand gesture are then implemented on android phones has an application to communicate [5]. The movements are tracked by the glove, they are three types i) Finger Bends using Flex Sensor: it measure the pressure of the finger that is bend to depict signs ii) Angular Movement using Gyroscope: determines the angle between each finger when there is movement in space and audit the rate of changes occur along each axis iii) Orientation using Accelerometer: the direction and magnitude is calculated. The movements of hands are stored as the features to recognise the sigs. Arduino is used for writing and uploading the code ATMEG32. Android application connected to bluetooth will recognize the gesture based on the angles detected by the movements of the flex sensors and display the output on the mobile screen.

A set of features obtain under complex background is efficiently used in Conditional Random Field (CRF) recognition system [6]. The author describes two types of motions: i) Global Motion: the motion which represents the gesture of hands along with the fingers and ii) Local Motion: the motion which represents only the fingers of the hands. The pre-process of image includes skin color segmentation method that is used to convert RGB color into HSV color space and the frame differencing method is used for subtracting the background objects. Contour matching algorithm, calculates the values based on the each edge pixels. The features are extracted on the probability of segmentation and naming the data one after the other by using Conditional Random Field distribution formulas. Finally the system recognizes the gesture with single hand or double hand based on the features extracted.

III. PROPOSED METHODOLOGY

A block diagram of system architecture represents the proposed approach as shown in Figure 2. The static datasets is downloaded from the www.kaggle.com, real-time images and videos are captured using web camera which consist around 200 images for each ISL with complex background and 10 dynamic signs videos of english alphabets. The 70% of dataset is trained using different machine learning algorithm like SVM, KNN and Naive Bayes in-order to predict which algorithm fit better and 30% dataset is used to test the model. The sign language recognition system works into two stages, the first stage is preprocessing the image are extracting the hand shape, orientation and other features from the image using image preprocessing techniques and the second stage is classification, where classifier produces high accuracy in recognise the sign. Lastly the output of recognised sign is in the form of text and speech.

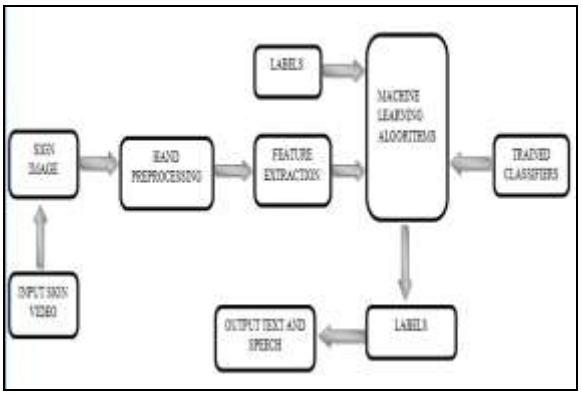

Figure 2: System Architecture of proposed system

A. Image and Video Acquisition

The first step in sign recognition is collecting the datasets of ISL english alphabets and videos. They are two types of datasets: I) Standard datasets which can be downloaded using www.kaggle.com or any government websites. II) Real-time datasets is capturing the hand sign images using web camera or high resolution mobile camera. Even web camera of laptop can be used but does not provide high clarity of images. The captured images are more than 100 RGB images per alphabets with static, different rotations and complex background. The 10 dynamic videos are captured, where each videos present the signs of ISL english words consisting of 5 to 6 letters such as "HELP", "PLACE", "WATER", "DRINK" and "WALK" etc,.

B. Image Pre-processing and Binarization

Every RGB images has to be resized and converted into frame vector (X) and name it corresponding to the sign language alphabets denoting (Y) to define a label for each RGB image. Pre-processing phase involves extracting frames from the video stream as frame per seconds (fps) based on the length of the videos. The frame rate corresponds to the number of images projected per second. The movement of hands are essential to process segmentation. The various image processing techniques are shown in Figure 3.

- Colour Space Enhancement: RGB image is converted into Hue, Saturation and Value (HSV) images to separate image luminance (brightness) from color information.
- Binarization: The HSV image is then converted into black and white image. The binary image is composed by calculating a simple threshold value of the pixels in the grayscale Image. Threshold value lies in-between 0 to 255 to produce a binary image.
- Image Denoising Techniques: Denoising involves manipulation or reducing the noise from the image to produce visually high quality image. Gaussian function is used to reduce the noise and contract the edges. Median blur operation is used to replace all the pixels in the kernel area by the median value and primarily processes the edges to remove noise.
- Morphological Operations: It involves erosion dilation operation based on the hand shape. The work of the dilate operator is to increase the pixel of the object area and uses accentuate features and work of erode operator is to away the boundaries of foreground object and used to diminish the features of an image.

C. Canny Edge Detection

A wide range of edges in image are detected by applying an edge detection operator in multi-stage algorithm. Gaussian function is used to remove the noise and smoothen the binary image. The edge intensity and direction of images is calculated by gradient of each pixel which results in change of intensity compared to original image. Non- maximum algorithm is used to find the gradient matrix intensity with maximum value (i.e, 255) in edge detection and finally double thresholding technique is applied to make the weaker matrix as a stronger matrix to achieve perfect edge detection.

D. Histogram of Oriented Gradients features

The Hog algorithm is used to count the phenomenon of gradient orientation in the localized portions of an image and store it has vector number for corresponding sign. When histogram of oriented gradient features used with SVM classifier it gives better accuracy compared to other classifier. The first step in Hog feature extraction is pre-processing where the input sign image is patched in many locations, then the horizontal and vertical gradients along with the direction and magnitude is calculated. The histogram of gradient in 8x8 cells is processed to obtain the nine bin histogram vector. Finally the normalization function is enforced on the vector obtained and calculates the Hog feature vector for the given sign image.

E. SVM Classifier

Support vector machine is a supervised learning method which deals with the understanding and inferring from the function by labelling the training set of data. After performing hand pre-processing, edge detection and feature extraction the output values are classified into different classes. The algorithm first creates hyper planes to separates the data into two classes, any number of hyper planes can be constructed to classify the data. The approximate hyperplane can be represented as the one which offers largest separation or margin between any two classes. The hyperplane for which the margin is maximum is the optimal hyperplane. The hyperplane is measured in terms of distance present from the adjacent data point on where it is maximized on each side.

F. Reverse Sign Recognition

The reverse process is essential in sign language recognition system due to dual mode of communication. In reverse recognition sign text (English alphabets) is given as the input, the text maps on to the labels which consist of X and Y values and matches the sign image to parade the RGB image along with speech. Example when user provide "WALK" text as input, the RGB image of W, A, L and K sign image will be displayed sequentially to user with speech. In OpenCV python Pyttsx3 module is invoked for the conversion of text to speech. The pyttsx3 is an offline cross platform library function which supports multiple text –to-speech (TTS) engine.

G. Natural Language Processing (NLP)

In sign language recognition the NLP module is used in-order to check the spelling of the words. When user communicates faster some real-time sign may not be able to detect by the computer, formerly it may display wrong text to avoid miss-spell the NLP module named "pypspellchecker" is used. A spellchecker is a software tool that identifies and corrects any spellings mistakes in a text or images. Example user provides a real-time sign video has "HEPL" as input, the spellchecker corrects the sign word and exhibits output has "HELP" each character of word in sequence with text as well as speech.

IV. EXPERIMENTAL RESULTS

The proposed system is implemented on the OpenCV Python software. The expected result of the system is successfully displayed with text and Speech. Speech has an audio output is not able to exhibit in the result section. Three different algorithms were used to train the dataset and check the accuracy to predict the classifier to test the data as shown in Figure 4. The SVM algorithm achieved good performance compared to other algorithm in trained phase consequently the test datasets used SVM classifier as testing phase. The four parameters are used to calculate the accuracy of trained dataset, i) True Positives (TP): where actual class and predicted class are yes. ii) True Negatives (TN): where the actual value and predicted value are both no. iii) False Positives (FP): where the actual class are no and predicted class are yes. iv) False Negatives (FN): where actual class are yes and predicted class are no. Therefore the formula for accuracy is,

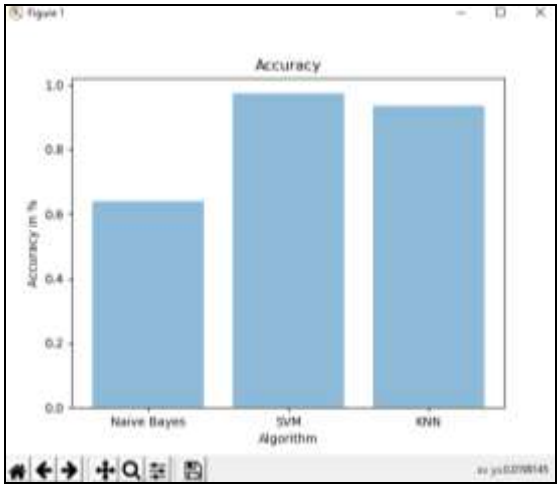

Figure 4: Accuracy comparison of trained dataset

$$\text{Accuracy} = \frac{TP+TN}{FP+FN+TN}$$

The system of vision based gesture recognition for Indian sign language english alphabets goes through image processing techniques to get a processed quality image for further image analysis such as edge detection, feature extraction and classification as shown in following figures.

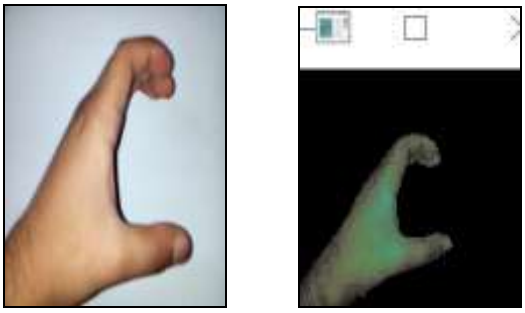

Figure 5: Color Enhancement of Sign "C"

The figure 5 depicts the conversion of RGB image into HSV image to separate image luminance from color information.

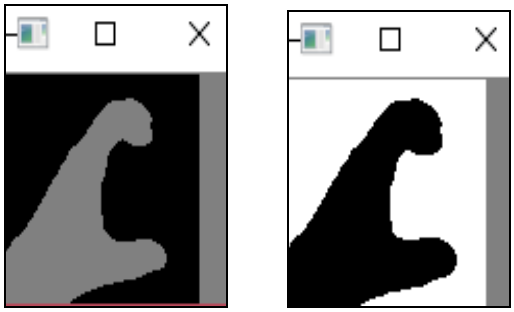

Figure 6: Binary Conversion of Sign "C"

The Figure 6 shows skin masking which deals with the recognition of *skin* colored along with the orientation and size of the sign image. The transformation of RGB values (24 bit) into grayscale value (8 bit), reduces the complexity for the pixels and displays the binary image with the values of 0 (black) for all pixels of the hand region and 1 (white) for background region. Conversion of grayscale image into binary image is done by thresholding technique where the threshold value is stated in between 0 to 255.

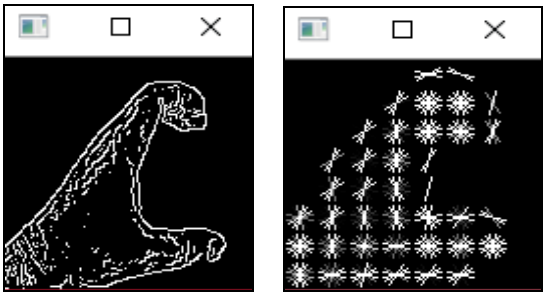

Figure 7: Edge detection and HOG features of Sign "C"

The Figure 7 presents the canny edge detection to detect the edges of sign and the features like orientation and rotation is processed by histogram of oriented gradients.

Figure 8: Outputs the recognised sign "C" along with text and speech.

The Figure 8 shows the Indian sign alphabet "C" is recognised when the image "C" is fed to be classified using SVM. The classified image gets displayed along with labels, text and speech. The reverse of the sign recognition system outputs when the user provides ISL english letter such as "PLACE" the image of given text will be converted into sign image and speech in the sequence of text stated by the user.

Figure 9: Recognition of sign words using NLP

Figure 9 depicts the correction of sign words using NLP modules, the input is given in form of sign videos and output displays sign images, text and speech for both before NLP and after usage of NLP module. Experimental results of single handed Indian sign language recognition clearly shows the success of human-computer interfaces system in above Figure (4 to 9) having the average accuracy rate of 97.1%.

V. CONCLUSION AND FUTURE WORK

The implementation was proposed for recognizing single handed ISL to remove the communication barrier between mute and normal people in the society. The results were calculated using different set of image preprocessing techniques and machine learning algorithms. The complete result was achieved by HOG features along with the SVM classifier was 97.1% of high accuracy. The system is not only focused on the sign to words but also on the speech and text using OpenCV python modules. The benefit of the proposed approach is that the training set up is done before the real time usage, to reduce the processing power and increase the efficiency in testing time.

The future work of human-computer interface system can be more versatile working on Android application and to build a language translator system which involves all the sign language dictionaries of different nation to help mute people to communicate easily across the world.

REFERENCES

[1] Divya Deora and Nikesh Bajaj, "Indian Sign Language Recognition", International Conference on Emerging Technology Trends in Electronics, communication and Networking, 2012.

[2] Kanchan Dabre and Surekha Dholay, "Machine Learning Model for Sign Language Interpretation using Webcam Images", International Conference on Circuits, System, communication and Information Technology Applications, 2014.

[3] Akanksha Singh, Saloni Arora, Pushkar Shukla and Ankush Mittal, "Indian Sign LanguageGesture Classification as Single or Double Handed Gestures", International Conference on Image Information Processing, 2015.

[4] Sunny Patel, Ujjayan Dhar, SurajGangwani, Rohit Lad and Pallavi Ahire, "Hand-Gesture Recognition for Automated Speech Generation", IEEE Interanational Conference On Recent Trends In Electronics Information Communication Technology, May, 2016.

[5] S. Yarisha Heera, Madhuri K Murthy, Sravanti and Sanket Salvi, "Talking Hands- An ISL to Speech Translating Gloves", International Conference on Innovative Mechanisms for Industry Applications, 2017.

[6] Ananya Choudhury, Anjan kumar and Kandarpa Kumar, "A Conditional Random Field based ISL Recognition System under Complex Background", International Conference on Communication Systems and Network Technologies, 2014.

[8] Umang Patel and Aarti G Ambekar, "Moment Based Sign Language Recognition for Indian Languages", International Conference on Computing, Communication, Control and Automation, 2017.

[9]Purva A Nanivadekar and Dr. Vaishali Kulkarni, "Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation", International Conference on Circuits, System, Communication and Information Technology Application, 2014.