International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 01 | Jan 2019 www.irjet.net p-ISSN: 2395-0072

Image Processing Techniques for Hand Gesture and Sign Recognition

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

[1]Professor, Department of Computer Science and Engineering, BNMIT
[2]M.Tech Student, Department of Computer Science and Engineering, BNMIT, Bangalore, 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 received a greater importance over the last few years because to remove the barrier of communication between mute people and normal people. It is an object consists of distinct features to extract and recognize the gestures or signs exactly, therefore gesture recognition presents a most difficult and challenging tasks in the fields of image processing, computer vision and image analysis. The images are subjected to image processing steps. The processing steps consist of image acquisition, pre-processing, segmentation, feature extraction and classifiers. In order to achieve a better accuracy the image processing and machine learning techniques and algorithms are used.

***Key Words*: Hand Gesture Recognition, Static Images, Image Pre-processing Methodology and OpenCV.**

1. INTRODUCTION

Sign language is the basic communication method for those who suffer from hearing impairment. The primary component of a sign language is hand gestures. Gesturing is an instinctive way of communication to present a specific meaning. Sign language is the visual manual modality to convey meaning which is quite similar to the hand gestures [7]. Language is expressed via the manual sign-stream in combination with non-manual elements. Sign languages are full-fledged natural languages with their own grammar and lexicon. In India over 21 millions of people are suffering some form of disability out of that about 7.5% people are dumb and 5.8% people are both dumb and deaf [9]. Gesture involves the hands, head or face to communicate and sign involves only the hands to communicate. Humans express the idea or meaning by using the sign or gesture in order to convey the message to the mute people. A computer can detect and reconstruct the message addressed by the human gestures and translate approximately. Regionally different sign languages are used in different part of the world such as ASL (American Sign Language), CSL (Chinese Sign Language) and ISL (Indian Sign Language) etc, [1]. They are two types of sign language recognition i) Alternate Sign Language: Hand signs developed by speakers for a conversation in a specific context used have the standard signs and ii) Primary Sign Language: it is used by the group of people to communicate with each other who belong to the same family or locality. The decomposition of the signs can be simplified into single handed or double handed signs [3]. Classifying single or double handed reduces the number of signs in each subcategory. Indian sign language is most commonly used language among the mute people in India. There are almost 615 languages used in India based on the region and culture as the result to analyze the sign recognition is difficult, due to the involvement of both hands along with fingers with results in overlapping the hands [5]. Sign language recognition (SLR) is an auxiliary tool or a translator for deaf and dumb people to communicate without any barriers in society. Various methods of pre-processing, segmentation, feature extraction and classification are experimented in order to recognize the gesture correctly and achieve the better accuracy.

2. MOTIVATION TO DEVELOPSIGN LANGUAGE RECOGNITION MODEL

A deaf person exists all over the world, when the communities of deaf people abide automatically the sign language is generated to communicate, therefore the sign language recognition model is essential. The first aspect is to develop the assistive system for the deaf person, create some document of gestures or words of signs that are in readable form and also the system is designed to translate the sign language into spoken language which would be of great help for deaf as well as for hearing people. A second aspect is that sign language recognition serves as a good basis for the development of gestural human-machine interfaces [1].

3. APPLICATIONS

Sign language is used in most of the hearing impaired schools to help the children's to learn sign language using human-system interfaces. Android phone consist more number of applications such as talking hands, spread the signs and hand talk translator etc, which helps to easily communicate with the society. There are basically two approaches for decoding the hand gestures [8]:

A. **Sensor Based:** It is difficult to process because it involves the huge hardware components and sensor based hand gloves. Pre-processing the sensor based hand gloves image results in troubleshoot.
B. **Image Based:** It is one of the easiest implementation and widely used method for recognising the sign or gestures.

4. STEPS FOR IMAGE PROCESSING

The Diagram 1 shows the steps involved in preprocessing and classifying the hand gesture and sign.

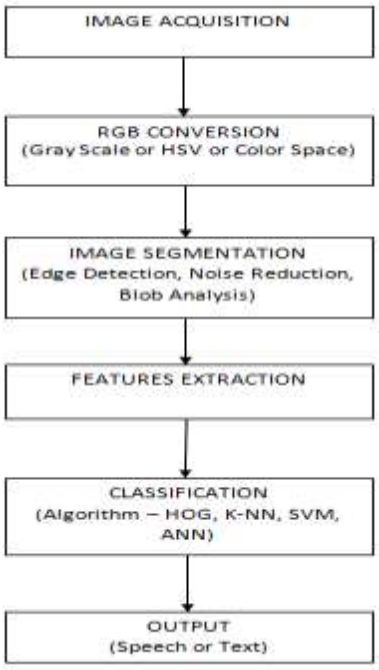

Diagram 1: Steps for Image Pre-Processing

Image Acquisition: The action of retrieving an image from source, usually a hardware-based source for processing (webcam or camera). It is the first step in the workflow sequence because without an image the next preprocessing steps are not possible.

RBG Conversion: A pixel color is a combination of three colors (RGB) which ranges from 0 to 255. Any digital image has to be converted from RGB values (24 bit) into grayscale value (8 bit) or L*A*B color space.

Image Segmentation: It is the process of partitioning a digital image into multiple segments. Segmentation is done to simply or changes the representation of an image into easier analyze of the image.

Feature Extraction: A feature is a piece of information which is extract from the relevant application for solving the computational task.

Classification: The task of extracting information classes from a multiband raster image. The resulting raster from image classification can be used to create thematic maps.

5. METHODS AND TECHNIQUES FOR HAND 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, finding perimeter pixels, finding corner points and eliminating the corner points. Finally the clustering algorithm is used to calculate the image with the values obtained in finger-tip finding algorithm. Principal Component Analysis (PCA) tool is used 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] [9]. First phase is pre-processing the image with background subtraction, blob analysis, noise reduction, brightness 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. The classification phase consists of two stages: Training stage and Testing stage. Training stage consists of 500 positive, 500 negative and 50 test image samples in each folder 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 software used is Emgu CV cross platform.NET wrapper to OpenCV tool and Voice synthesis can be accomplished by “system.speech” and Microsoft.NET framework provides library function named “system.speech.synthesis”.

A noval approach involves classifying gesture into single or double handed subcategories which simplifies the overall problems that are associated with the gesture recognition and reduces the number of gestures 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, Major to minor axis, eccentricity, bounding box ratio and equivalent diameter to minor axis ratio, the above obtained values are used to classify the single or double handed gestures. Histogram of Oriented Gradients (HOG) techniques are used to extract local orientations and intensity distribution for detecting the shapes of the object. Classification is done by applying K- Nearest Neighboring (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.

Sunny Patel narrates that the system runs on the mobile computing device which provides the technology for automated translation of the ISL into speech in the English language by providing the duplex communication between

the impaired people and the normal people [4]. The image is processed using Hue, Saturation and Value (HSV) algorithm for skin detection. Two detection methods are i) Blob Detection: used to calculate the distance between the centroid and identify the large blob. ii) Contour Detection: the grayscale image is accepted and if it's 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. Finally the Principal Component Analysis (PCA) function is used to analysis the sign and converts the text into speech.

S. Reshna [5] performs skin segmentation using YCbCr colour space and the threshold value of the components on each value of Y, Cr and Cb are applied on spatial filter to the image. A (SVM) Support Vector Machine classifier is used to separate the data into input space and hyper planes to recognize the hand gestures based on the trained point features.

Sensor incorporated on a glove detects the gestures and convert it to speech with the help of a blue-tooth module and an android phone [6]. The gloves will track three kinds of movements i) Finger Bends using Flex Sensor: calculates the resistance for bend in each fingers ii) Angular Movement using Gyroscope: the gyroscope calculates the angular movement in space by checking rate of change of angle along each axis iii) Orientation using Accelerometer: calculates the orientation of the hands in spaces by determining the axes readings. Arduino is used to provide an interface to write and upload the code on ATMEGA32. 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.

Conditional Random Field (CRF) based ISL recognition system is effective under complex background using a noval set of features [7]. 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 feature extraction is based on the probability for segmentation and labeling the sequential data by using Conditional Random Field distribution formulas. Finally the system recognizes the gesture with single hand or double hand based on the features extracted.

Motionlets matching with adaptive kernels describes the 3D motion captures the sign language in a video format [11]. The video is converted into frames and segmented into two phases: i) Skeleton and ii) Normal picture. Skeleton images are taken to identify the motion segmentation and non-motion segmentation. The four databases generated with motion joints segmented images uses Tracking Based Classification, further it is used to extract the finger shapes from images by using the formula of relative distance and joint relative angles. The three Adaptive kernel models are used to define the motion, the distance of joint motion and adaptive kernel matching technique is used recognize the query sign using ranked database.

A smart wearable hand device is used for American Sign Language recognition, which consists of 5 flex sensors, two pressure sensors and three-axis inertial motion sensor [12]. A wearable device was built using android pro mini controller to extract the finger gesture and to serve as input to the classifier, flexible filaments to enable the function of each finger joints and shapes etc, flex sensor used to recognize the movement of both the direction (upwards and downwards). An adafruit BNO055 device was used to abstract the sensor fusion and orientation data in quaternion based on the angles and vector format. Finally SVM classifiers are used to recognize the sign and display the text and audio in mobile application via the blue-tooth.

Key-frame Centered Clips (KCC) is a sign clip centered on a key-frame that is extracted by key-frame based approach in raw RGB sign data based on the natural characteristic of sign [13]. The three major folds are: i) Proposing KCC: to extract the hand features, locate key-frame and generate KCC's. ii) Multimodal Features: the obtained KCC features are fed into the neural networks: Conventional Neural Network (CNN) makes use of softmax layer with K output unit that is cascaded in CNN networks. The softmax is removed and fully connected layer is extracted as CNN features by a forward propagation algorithm. Depth Motion Maps (DMMs) is used to project the depth image onto these orthogonal cartesian planes to form 2D projection and Trajectory features extract the real-time position of skeleton joints. Later sub-work-to-KCC, KCC to words and words to sentence is identified to feed into the encoder- decoder LSTM network. Cross entropy loss is used to train the encoder-decoder network.

Dynamic Sign Language Recognition (DSLR) is used as a smart house application where the datasets are trained using K means++ cluster and tested using non-linear SVM [14]. Dynamic Sign Language Recognition consists of two modules: i) Image Processing Model and ii) Linear Formal Grammar. Different steps used in image processing model are IP modules which recognize the individual words of the sign language, Bags-of –features is used to model the hand gesture recognition from the video, 3D HOG descriptor to represent the features, K++ means to cluster the visual words. The classification was conducted based on non-linear Support Vector Machine and Linear Formal Grammar module that is used to analyze whether the grammar sentence is valid or not.

Implementation of real-time static hand gesture recognition methods deals with computer vision and machine learning

[15]. A computer vision focuses on acquiring the images with the support of image processing and extracts the essential data of the image by using HSV color space technique, it is the most common RGB color model representation for digital image and the angle from the central vertical axis is calculated using HSV formula. Gaussian blue technique is the process to smooth blur that resembles viewing the image through translucent screen. Contours technique is used to extract the features from the signs and finally the Artificial Neural Network (ANN) classifier is used to recognize the sign accurately. In speech translation, few words are trained and stored in database, the signs recognized are framed into words using grammar translation and the web-text are used to display speech or text.

The development of vision based static hand gesture recognition system uses web-camera in real-time application [16] [18]. The model is developed using three major folds: i) Pre-processing techniques: skin color segmentation and binarization which converts RGB color into gray-scale image. Morphological filtering and hand region detection process removes noise and well defined smooth closed hand gesture, the cropped images are resized to 128x128 pixel resolutions. The characteristic of the object and sign patterns information are extracted using Discrete Wavelet Transform (DWT) and Fisher-Ratio techniques, hence the SVM classifier exactly classifies the hand gesture. An author makes use of LOO-CV to train and test the datasets. The Leave-P-Subject-out Cross-Validation is used to evaluate the performance of the techniques with classification accuracy.

The signs 3D spatio-temporal information of each sign was interpreted using Joint Angular Displacement Maps (JADMs) which encode the sign as a color texture image [17] [19]. The method involves three steps: i) Compute JADMs from the 3D data which identify and calculate the 3D position of the image. The extraction of joint distance, angle and time evolution is done using T-Frame sign formula. ii) Encoding JADMs into RGB images, it is the standard mapping procedure for jet color map with each frame representing the color-coded X, Y and Z axis planes to improve the accuracy and iii) CNN Classifier Architecture is used to match the sign, the training and testing is done with help of the tensor flow and keras libraries that is implemented using python 3.6.

6. CHALLENGES FACED FOR HAND RECOGNITIONS

Images with certain angle of rotation will limit the number of samples.Irrelevant object might overlap with the hand for wrong object extraction. The usage of hand gloves also results in calculating the corner tip of gestures wrongly [1]. Reduce the performance recognition algorithm and system limitations restrict the applications. Ambient light affects the color detection threshold. The system gives better accuracy only when the images are static compared to the motion gestures [2]. Due to the lack of training datasets and improvising of the speech synthesizer cannot pronounce the sign to word input [3]. The system makes use of lighting condition in different situation which is difficult to pre-process the image. Ambient light affects the color detection threshold [5]. Few systems recognize the gestures based on the sensors placed on the hand which is difficult to carry all through the system [6]. The noise is high, if background color is same as the hand color [7] and the system fails to detect objects placed in front of webcam, so system may recognize any object rather than a specific object [8].

7. COMPARSION TABLE

The Comparison Table 1 brief out the methods, techniques and accuracy gained while recognizing the hand gesture in each paper. Image acquisition is the first and foremost step in identifying and collecting the images. Images are captured using laptop camera or to get better clarity the web cam is used. The captured images (datasets) consists of hand gestures and sign are Indian Sign Language alphabets (A-Z), numbers (0-9) and few gestures like “Hi”, “Bye”, “Namaste”, “hello”, “close”, “open”, “ drink” and “Study” etc,. The datasets can be in the form of static (images) or dynamic (video) format.

8. PROPOSED SYSTEM

Human sign language gestures are the combination of independent of hand and finger articulations. The Indian sign Language static images are fed as the input, the image is preprocessed using different image processing and machine learning algorithms and techniques. The extracted features are compared with the training datasets using a basic matching algorithm to match the signs accurately. Finally the sign image is converted into text and speech.

Table 1: Comparison Table for Hand Gesture Techniques

Year	Techniques	Accuracy
2012	Clustering Algorithm, and PCA Tool	94%
2014	Haar Cascade Classifier, Speech Synthesis Tool and Emgu CV Software	92.68%
2015	KNN Algorithm and SVM	88.46% (KNN) and 94.23% (SVM)
2016	Eigen Object Recognize and PCA Tool	High and Efficient accuracy obtained.

2017	Flex Sensor and Arduino Translator	94%
2014	Conditional Random Field Probability Techniques	90% (single hand) and 86% (double hand)
2018	Motion Segmentation Algorithm, Tracking based Classification and Adaptive kernel matching	75% (3D SLR in joint motionlets)
2017	SVM Classifier and Mobile Application	65.7% (only flex sensor) and 98.2% (both flex and pressure sensor)
2018	Proposing KCC , CNN, DMMs, Trajectory Feature	87%
2015	K means++ Cluster and Non-Linear SVM	97% (without grammar syntax) and 98% (with grammar syntax)
2017	HSV, Gaussian blue, Contour technique, ANN	90%
2018	Morphological filtering, DWT and F-ratio techniques, SVM	98.60%
2018	JADMs Formulas and CNN	Better accuracy was obtained

9. CONCLUSION

Now a day's the impaired people are finding difficulties to communicate with society. It is difficult to common people to understand the sign and gestures done by the mute people, therefore it is very much needed to design the human-system interfaces. The image processing techniques are used to extract the features from the images. The most commonly used techniques are segmentation, Binarization, background subtraction and noise reduction. The extracted features are applied on the machine learning techniques to recognize the hand sign accurately.

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. Reshna and M. Jayaraju, "Spotting and Recognition of Hand Gesture for ISL Recognition System with Skin Segmentation and SVM", Conference, 2017.

[6] 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.

[7] 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.

[10]Mohd. Baqir Khan, Kavya Mishra and Mohammed Abdul Qadeer, "Gesture Recognition Using Open-CV", 7th International Conference on Communication Systems and Network Technologies, 2017.

[11] P V V Kishore, D. Anil Kumar, A S C S Sastry and E Kiran Kumar, "Motionlets Matching with Adaptive Kernels for 3D Indian Sign Language", IEEE Sensors Journal, 2018.

[12] B G Lee and S M Lee, "Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion", IEEE Sensors Journal, 2017.

[13] Shiliang Huang, Chensi Mao, Jinxu Tao and Zhongfu Ye, "A Novel Chinese Sign Language Recognition Method Based on Keyframe-Centered Clips", IEEE Signal Processing Letters, March, 2018.

[14] Muhammad Rizwan Abid, Emil M Petriu and Ehsan Amjadian, "Dynamic Sign Language Recognition for Smart Home Interactive Application using Stochastic Linear Formal Grammar", IEEE Transactions on Instrumentation and Measurement, March, 2015.

[15] Rosalina, Lita Yusnita, Nur Hadisukmana, R B Wahyu, Rusdianto Roestam and Yuyu Wahyu, "Implementation of Real Time Static Hand Gesture Recognition using Artificial Neural Network", IEEE Journal, 2017.

[16] Jaya Prakash Sahoo, Samit Ari and Dipak Kumar Ghosh, "Hand Gesture Recognition using DWT and F-ratio Based Feature Descriptor", The Institution of Engineering and Technology Image Processing, 2018.

[17] E Kiran Kumar, P V V Kishore, A S C S Sastry, M Teja Kiran Kumar and Dr. Anil Kumar, "Training CNNs for 3D Sign Language Recognition with Color Texture Coded Joint Angular Displacement Maps", IEEE Signal Processing Letters, 2018.

[18] Neha Baranwal, Neha Singh and G C Nandi, "Indian Sign Language Gesture Recognition Using Discrete Wavelet Packet Transform", International Conference on Signal Propagation and Computer Technology, 2014.

[19] G Anantha Rao, K Syamala, P V V Kishore and A S C S Sastry, "Deep Convolutional Neural Networks for Sign Language Recognition", SPACES-2018.