



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Real-Time Vision Based Sign Language Recognition System

Akanksha Medhi, Mitali Doshi, Prajкта Pawar, Apurva Kesarkar, Prof. Rupali Dalvi

B. E Student, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Karvenagar
Pune, India

Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Karvenagar
Pune, India

ABSTRACT: This paper presents a novel system to aid in communicating with those having vocal and hearing disabilities. It discusses an improved method for sign language recognition and conversion of speech to signs. The algorithm devised is capable of extracting signs from video sequences under minimally cluttered and dynamic background using skin color segmentation. It distinguishes between static and dynamic gestures and extracts the appropriate feature vector. These are classified using Support Vector Machines. Speech recognition is built upon standard module Sphinx. Experimental results show satisfactory segmentation of signs under diverse backgrounds and relatively high accuracy in gesture and speech recognition.

KEYWORDS: Computer and information processing, Feature extraction, Gesture recognition, Image processing, Sign Language Recognition, Speech recognition

I. INTRODUCTION

a. MOTIVATION

Sign language is the basic means of communication for those with hearing and vocal disabilities. Those disadvantaged have difficulty in their day to day lives. We aim to develop a system that would ease this difficulty in communication. Sign language consists of making shapes or movements with your hands with respect to the head or other body parts along with certain facial cues. A recognition system would thus have to identify specifically the head and hand orientation or movements, facial expression and even body pose. We propose the design for a basic yet extensible system that is able to recognize static and dynamic gestures of American Sign Language, specifically the letters a-z (where j and z are dynamic with hand or fingertip movement while the rest are static). American Sign Language was chosen since it is utilized by a majority of those disabled. The system is able to perform in dynamic and minimally cluttered background with satisfactory result as it relies on skin color segmentation [3] to separate the gestures, i.e. focus on the hands and face. Since the lexicon under consideration (letters a-z) does not involve facial cues, we eliminate the face using Viola-Jones face detection followed by subtraction of the detected region. We classify the gesture as static or dynamic by measuring the distance moved by the hand in subsequent frames. For static gestures, we use Zernike moments, a well-known shape descriptor in image processing as shown in [4]. For dynamic gestures we extract a curve feature vector which shows high accuracy in uniquely identifying paths [7]. These feature vectors are then classified using pre-trained SVM classifiers. This helps map the gesture to a particular alphabet. For speech recognition, Sphinx module is used which maps the spoken alphabet to text with high accuracy. This text is then mapped to a picture if it is a static gesture or a video if it is a dynamic gesture. The paper is organized as follows - Section II describes the past research and advances made in the same fields. In Section III we give a brief overview of our system. Section IV details gesture recognition system step-by-step. Section V briefly discusses the speech recognition system. Section VI shows the experimental results. In Section VII we compare our method with other common and prevalent systems. In the last section, Section VIII we give the conclusion of our method.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

b. BACKGROUND

Hand gesture recognition is a rapidly growing field in various applications, such as sign language, computer vision, natural language processing, biomedical, biometrics, pattern recognition, and much more [15]. Hand Gesture Recognition can be seen as a way for computers to understand the intentions of the human being, creating an interface between machines and humans. With the advent of technologies such as smartphones, Hand Gesture Recognition is expected to add new ways of interacting with mobile devices. Thus, gestures captured by the camera of the mobile can be analyzed to place orders without having to touch the screen. In our paper, we chose to make our system on an Android smartphone. The choice of the Android operating system is not arbitrary but for several technical and non-technical reasons. Among the non-technical reasons is that Android is the world's most popular operating system for mobile phones and the majority of smartphone users use the android system [1]. And for technical reasons, we have chosen to work with Android platform because it allows integrating libraries for image processing. In our case, we have chosen to integrate a free graphics library specializes in real time image processing named Open CV (Open Computer Vision) library [2] to perform image processing. And then we have chosen to integrate the SVM classifier (Support Vector Machine) which can model real-world problems such as image classification, hand-writing recognition, text, bioinformatic and bio sequence analysis. Our system, unlike systems realized before, is designed to perform all necessary operations in order to recognize hand gestures, by the smartphone and without the need to use another device. Some applications that use gestures as an Human-Machine Interface were first imposed in game consoles (Microsoft Kinect and PlayStation Eye for Sony) and could now be used in other areas such as smart television and smartphone

II. RELATED WORK

There are two categories for vision-based hand gesture recognition. The 3-D hand model based method and appearance based method. The 3-D hand model based method works by comparing the input frames and the 2-D appearance projected by the 3-D hand model. However a huge database is required to deal with all the possible projections of the 3-D hand model making it less practical. In contrast, appearance-based techniques extract the image features and model these as visual characteristics of a hand posture and compare them with those features extracted from the live video feed of a user performing a gesture. They are real-time in nature because 2-D image features are used. The appearance-based techniques can again be categorized as hand posture detection (static) and gesture detection (dynamic). Nasser H.D et.al [1] considers this approach where the key features extracted are SIFT (Scale Invariant Feature Transform) key-points. They further constructed a grammar from a sequence of hand postures for detecting dynamic gestures. Emil M.P. et.al [2] proposes the same but uses Haar-like features for describing the image and AdaBoost for classification. It discusses the advantages like being fast in computation and disadvantages such as the need to use a large number of features, which makes the training stage of the system difficult since the AdaBoost classifier is used. The identification and extraction of hands from a cluttered and dynamic background poses another problem. One of the most popular methods is Skin color detection. In [3] P.K.Bora et.al introduced HSV - Hue, Saturation, Value color space and it states that the skin color, irrespective of gender or race falls in particular ranges for H and S. Utilizing this information we can extract skin colored objects from any background. In [4] Macheal V et.al had carried out the evaluation of various shape descriptors including Zernike moments and displays its superiority. Zernike moments have become increasingly adapted to image processing schemes as their higher order moments can be calculated independent of lower order moments. Also reconstruction and rotation invariance make them an attractive feature for shape description. Athira et.al in their project report [10] shows the viability of Zernike Moments by building a system to recognize only the static gestures of ASL against a uniform background. It also features a speech engine to convert speech to gesture. In [5] a basis for usage of Hidden Markov Models (HMM) is established by drawing an analogous relationship between speech recognition and gesture recognition. HMMs can be used to model time series data, and here the movement of the hand along the co-ordinate axis is tracked and each direction is taken as a state. This paper makes use of a lexicon of 40 gestures and achieves an accuracy of 95 percent. It also states its disadvantage that as the lexicon grows the need to describe the hand configuration along with hand trajectory also will grow making the designing of HMM more complex and time-consuming. We needed a method to describe dynamic gesture in a simpler way. In [6] an object

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

tracking algorithm, Hausdorff object tracker is used to extract frames from a live video feed that will best encompass the translation an object has gone through.

III. SYSTEM OVERVIEW

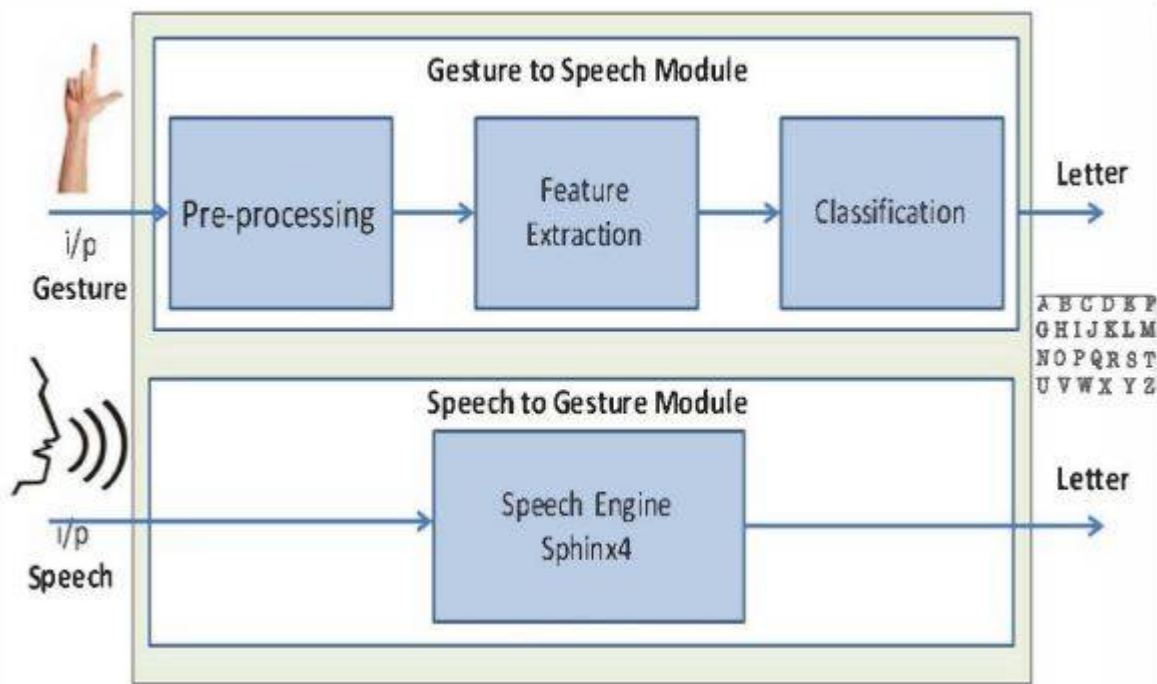


Fig 1: High Level View of System

On a high level the system comprises of two independent modules as shown in the above diagram. The two modules being the 'Gesture to Speech Module' and 'Speech to Gesture Module'. The Gesture to Speech Module takes a gesture as an input in the form of an image/video and produces the speech corresponding to the sign as the output, while the second module does the vice-versa where it takes speech from the user as the input and produces the corresponding image/video as the output. Here the image input/output is meant for a static gesture while a video in the same case is used for a dynamic gesture.

IV. GESTURE RECOGNITION

The input can be a static or dynamic gesture, so to make it as general as possible a video recording of two seconds is passed as the input at the rate of 6fps. The flow diagram of the gesture to speech module is shown below (Fig 2). The gesture is extracted and depending on its type (static/dynamic) certain features are extracted. These are then classified using pre-trained SVM classifiers. The details of image extraction and classification is explained in the below section.

A. SKIN - COLOR SAMPLING

It is theorized that in the Hue, Saturation, Value color space the skin color, irrespective of gender or race lies in-between certain values of Hue and Saturation[21][22]. However if this global range is used for thresholding it would give us some noise and/or false positives.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

B. IMAGE PRE-PROCESSING AND HAND SEGMENTATION

Since skin color segmentation is used to detect the hand we have to make sure that there is no other prominent skin colored regions in the frame. The face is often more prominent than the hand. Hence the first step in image preprocessing is the detection and removal of face using Viola and Jones algorithm.

C. POST PROCESSING AFTER HAND SEGMENTATION

After the image preprocessing, we get a binary image of the hand (Fig 6) which we invert (Fig 7). To be able to recognize dynamic gestures we need to detect and track the finger-tips and COG to know whether they are moving.

D. THRESHOLD COMPARISON

Now that we have detected the CoG and fingertips we can identify if the gesture is static or dynamic by tracking the distance they move. In this paper the entire class of ASL gestures has been divided into three. First there is the basic classification of Static gestures and Dynamic gesture. Dynamic gestures are then further subdivided into dynamic gestures with significant hand movement and dynamic gestures with significant fingertip movement. Significant movement of the hand means substantial movement of the CoG of the hand while in the other class CoG almost remains stationary but the fingertip moves substantially.

E. FEATURE EXTRACTION

For static gestures, we use Zernike moments to identify the orientation of the hand. From experimental results and [10] we were able to deduce Zernike moments up to the seventh order and hence a seven feature vector is sufficient to identify a static gesture. For dynamic gesture involving the tracking of CoG or a particular fingertip we characterize the path traced by a curve feature vector of 5 elements. To do this we first identify Key points or representative points on the path. As demonstrated below a circle would require more key points to characterize the path whereas a square would only require four. Next we extract the features which are

1) Trajectory Length/Location feature:

The total length of the trajectory can be calculated by computing the distance between the key-points and summing them up. It is translation invariant as even if there is a shift the location where the gesture is drawn the length of the gesture will not change. Location feature is the distance between the center of gravity of the trajectory and the key-points. Since there are a number of key points the average of the sum is considered. It is also translation invariant as distance between center of gravity and key points would not vary even if there is a shift in location of the gesture. Thus the ratio, (Trajectory length/Location feature) would be a scale and translation invariant ratio.

2) Number of significant changes in hand orientation:

The orientation change (i.e. the angle formed) along the set of three key-points is considered, if it is greater than a predefined threshold say, an angular displacement ≥ 45 it is counted as one significant change, else ignored. The total count of such changes is accounted here.

3) Average velocity: It is the average velocity over the whole trajectory.

4) Standard deviation of the speed: Variation of the speed over the whole trajectory length indicates the smoothness of the trajectory.

5) Number of minima velocity points: Number of points that have a speed below some threshold $T-v(\min)$. This indicates turns in the trajectory and hence the smoothness of the trajectory.

F. CLASSIFICATION

Sanjay in his master's thesis [16] and Neelam et.al in their review [17] show that SVM classifiers are a good choice for classification of hand gestures, with some variation for different methods adopted for feature extraction. Classifiers are trained and a classification model is built using Java Machine Learning library- JA V AML [19]. Two classification models, one for Zernike feature vectors and another for trajectory feature vectors, are built which successfully classified the gesture to corresponding alphabet.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

V. SPEECH RECOGNITION

Speech Recognition (SR) is the conversion of spoken words into text, also known as "Automatic Speech Recognition" (ASR). There are two types of SR Systems, Speaker independent[12] and Speaker-dependent systems, where the latter has a training phase which uses a person's specific voice for the recognition purpose. A number of Speech applications are available in the market like Dragon NS [13]. For experimental purposes of this paper the Speech recognition module was setup in two ways, first using a speech engine (Sphinx 4), where a restricted set of grammar for alphabets from A-Z was used, just like a command recognition system. The SPHINX Speech Recognition Engine of CMU [14] was utilized. A SPHINX engine just like any other Speech Engine requires two files as input an Acoustic Model and a Language model or a Grammar File.

This general model of SPHINX is depicted in below diagram (Fig 9). Its working is based on Hidden Markov Models (HMM).

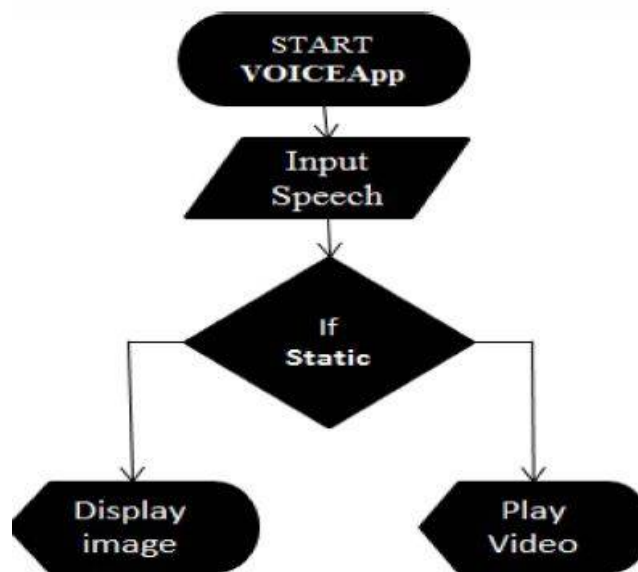


Fig 12: Flow

The SPEAK button (Fig. 1 1) can be used to give a speech input to the system and the output is either an image or a video corresponding to the gesture depending on whether it is static or dynamic. The accuracy shows huge variation based on the type of usage and the amount of background noise.

VI. CONCLUSION

The system is novel approach to ease the difficulty in communicating with those having speech and vocal disabilities. Since it follows an image-based approach it can be launched as an application in any minimal system and hence has near zero-cost. There are further areas of improvement such as increasing the system performance under robust and unfavorable environment (lot of clutter, poor lighting). We also need to expand the current feature set to be able recognize more gesture (like those involving two hands or facial cues). We also need to deal with co-articulation.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

REFERENCES

- [1] Nasser H. Dardas and Nicolas D. Georganas. Real-time handGesturedetection and recognition using bag-of-features and support vector machine techniques. IEEE
- [2] Emil M. Petriu Qing Chen, Nicolas D. Georganas. Real-time visionbasedhand gesture recognition using haar-like features, 2007.
- [3] P.K. Bora M.K. Bhuyan and D. Ghosh. Trajectory guided recognition of hand gestures having only global motions. International Science Index, 2008.
- [4] Michael Vorobyov. Shape class i_cation using zernike moments, 2011.
- [5] Thad Eugene Starner. Visual recognition of american sign language using hidden markov models. Master's thesis, Massachusetts Institute of Technology, Cambridge MA, 1995.
- [6] P.K. Bora M.K. Bhuyan and D. Ghosh. Trajectory guided recognition of hand gestures having only global motions. International Science Index, 2008.
- [7] Emil M. Petriu Qing Chen, Nicolas D. Georganas. Feature extraction from 2d gesture trajectory in dynamic hand gesture recognition, 2006. [8] Manisha M. Ingle Neelam K. Gilorkar. A review on feature extraction for indian and american sign language, 2014.
- [9] Chih-Chung Chang and Chih-Jen Lin. A review on feature extraction for indian and american sign language, 2013.
- [10] ATHIRA P K ALEENA K RAJ and DEEPA I K. Sign language conversion software for people with hearing and vocal disabilities, 2013.
- [12] Fifth Generation Computer Corporation-"Speaker Independent Connected Speech Recognition" [Online]-Available: Fifthgen.com
- [13] NUANCE. Dragon Speech Recognition Software [Online]. Availabe: <http://www.dragonsys.com>
- [14] Kai-Fu Lee, Hsiao-Wuen Hon, and Raj Reddy, An Overview of the SPHINX Speech Recognition System. IEEE Transactions on Acoustics, Speech and Signal Processing.
- [15] Developer-Android - "SpeechRecognizer" [Online] - Available: developer.android.com
- [16] SANJA Y MEENA. A Study on Hand Gesture Recognition Technique, 2011.