

Real-Time Sign Language Recognition: A Review

Ms.R.S.Gaikwad¹.Mrs.P.P.Belagal²

¹Assistant Professor, Electronics and Telecommunication dept.SBGI Miraj.(MH)India

²Associate Professor, Electronics dept. JJM COE Jaysingpur.(MH)India

Abstract

Real time sign language recognition is multidisciplinary research area involving computer vision, image segmentation, pattern recognition and neural language processing. Sign language recognition is a comprehensive problem because of the complexity of the shapes of hand. Sign language recognition requires the knowledge of hands position, shape, motion and orientation. A functional sign language recognition system can be used to generate speech or text making the hearing impaired person more independent. The most complicated part of sign language recognition system is to recognize the simplest hand gestures that must be detected in the image. In this paper, a review of different sign language recognition techniques is provided and a comparison of the results of different methods is done.

1. Introduction

Many researches began in late 90s to deal with recognition of sign language automatically in various languages. Since ages communication has served as a medium to build relationships, know people, understand technology and allow rapid growth and development on a global basis. Normal people can communicate their thoughts and ideas to others through speech. One important means of communication method for the hearing impaired community is the use of sign language. 5,00,000 to 2,00,000 people use Sign Language as their major daily communication tool. The population of deaf people in India is in the millions - some estimates are as high as 60 million of the total population. Ninety percent of deaf children are born to hearing parents who may not know sign language or have low levels of proficiency with sign language. Unlike hearing children of English speaking parents or deaf children of signing parents, these children often lack the access to language at home which is necessary for developing linguistic skills.

2. Gesture recognition by utilizing bio-mechanical characteristics

Gesture recognition by utilizing bio-mechanical characteristics, proposed in [1] is inspired by the observation that all forms of hand signs include finger-joint movements from a starting posture to a final posture. The concept of 'range of motion' from the Bio- Mechanical literature at each joint is used to abstract the movement. Range of motion (ROM) is a quantity which defines the joint movement by measuring the angle from the starting position of an axis to its position at the end of its full range of the movement. For example, if the position of a joint axis changes from 20° to 50° with respect to a fixed axis, the range of motion for this joint is 30°. The range of motion per joint is obtained by using the sensor values acquired by the sensory device.

The range of motion of each section of the hand participating in a sign, relative to the nonparticipating sections, is a user-independent characteristic of that sign. This characteristic provides a unique signature for each sign across different users. Given a sensory device with n sensors, each ASL static sign can be represented with a set of n values. Let us call this set $S_i = (s_1, s_2, \dots, s_n)$ where i is an ASL sign and S is the set of sensor values. One issue in gesture recognition is that different users making the same sign (gesture) would generate different S_i 's, i.e., S_i is not unique for different users. The objective is to transform S_i to NR_i where NR_i is unique across different users. Suppose S_0 and S_i represent the sets of the initial and final postures for a specific sign respectively.

The range of motion tuple R_i is as follows: $R_i = S_i - S_0$. Subsequently, we find the maximum and minimum values within R_i and represent them with $M(R)$ and $m(R)$ respectively. Each value in R_i is normalized and the result of this normalized R consists of values between 0 and 1 with NR . Finally, we discretize the values of NR with a given discretization parameter $k (> 1)$. For example, if $k=2$, we replace each value of NR with 0 if its value is less than 0.5. Since NR represents the characteristic of

movement of the sensors making a particular sign, it provides an abstraction for that sign. We call this abstraction the signature of the sign and observe that while the signature is unique for each sign, it is identical among different users making the same sign. That is, if different users wear the sensory device and make a specific ASL sign, while the raw data generated by the sensors are completely different, the calculated NR are almost identical across all of them. This also implies that by abstracting the sign with its signature, we eliminate the effect of inevitable noise produced by the sensors during the data collection process. The uniqueness of this signature provides us with the very important property of user independency. In order to recognize an unknown static sign made by a user, we require comparing its signature with the signatures of some known samples. Consequently, the first step is collecting the data for each static sign once and calculating its corresponding NR. We call this process 'registration' and save all the registered signs in our registration database. While we have this database, we compare the signature of each unknown gesture with all registered signatures and the unknown gesture is labeled with the signature with the least distance (according to a distance metric, e.g. Euclidian distance). For this experiment, a glove model is used which has three flexion sensors per finger, four abduction sensors, a palm-arch sensor, and sensors to measure flexion and abduction. A picture of this glove, indicating the location of each sensor is displayed in Figure 1. CyberGlove is a glove that provides up to 22 joint-angle measurements. It uses proprietary resistive bend-sensing technology to transform hand and finger motions into real-time digital joint-angle data.

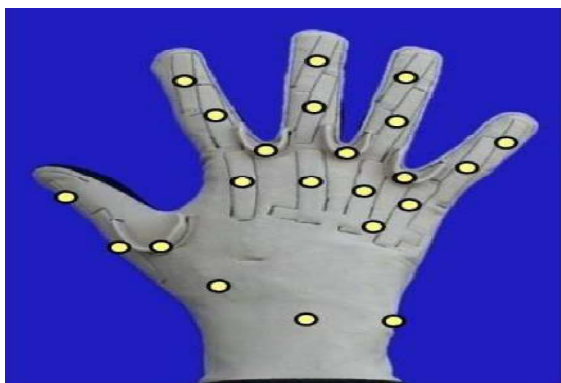


Figure1. Cyber Glove and the location of its sensors

3. SLR based on skin color [2, 3].

This is an intelligent & simple system for converting sign language into voice signal by tracking head & hand gestures. This system proposes a simple gesture extraction algorithm for extracting features from the images of a video stream. This system is very simple & subject is not required to wear any glove. But the subject must wear a dark color, long sleeve shirt. Then the gesture signs are recorded. Each image frame is segmented into three regions- head, left hand and right hand. Then these segmented images are converted into binary images. In feature extraction stage, for each frame the area of objects in segmented binary image is calculated. So, each frame has 3 segmented areas- head area, left hand area and right hand area. There is a different segmented area for each gesture type. Each segmented area is treated as a discrete event & DCT is applied to it. First 15 DCT coefficients are considered as features. They correspond to each segmented area. Combination of DCT coefficients from 3 segmented image areas are used as feature vector for Neural Network (NN). Here, a simple NN model is developed for sign recognition. The features computed from video stream are given as an input to this NN. To classify the gesture, a NN model uses error back propagation. So we get the classified signs at the end of gesture classification phase. Finally, an audio system is used to play words corresponding to gesture. The different steps involved are shown in figure 2.

4. Neural network approach

Neural networks are based on the parallel architecture of human brains. It can be defined as a multiprocessor system network with high degree of interconnection and adaptive interaction between elements. A neural network image processor can free imaging applications from various constraints in terms of video acquisition, lighting and hardware settings. This degree of freedom is possible in application to surveillance networks because a neural network allows you to build a network by learning examples. The more examples are learned, the more expert the network and sometimes it is quite easy to automate the learning. Sign language recognition using neural networks [4] is based on the learning of the gestures using a database set of signs. A combinational neural network model [8] is developed for feature extraction. A three layer network called back propagation is used to build the CNN. The network layer consists of 3 stages: stage 1, stage 2 and stage 3. Each stage acts as a back propagation neural network layer that takes the elements of

feature vector as input and outputs the class of object. Each stage receives their input elements from the feature extraction layer.

In the neural network approach, a video sequence of the signer is obtained using a camera. The image frames are then filtered using a median filter or mean average filter and after background subtraction [5] are given to the neural network for feature extraction.

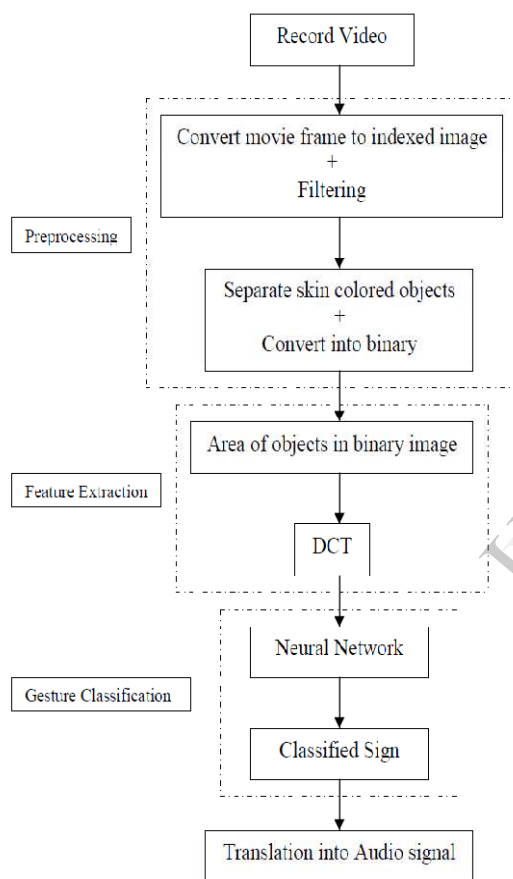


Figure2. Flowchart for SLR based skin color

5. SLR using hidden markov models

Hidden Markov Models are a type of doubly stochastic models [7]. In developing the system where the observations are characterized as a finite set of symbols, discrete HMM is used. A typical discrete HMM can be specified by N distinct states, M distinct observation symbols per state, and the probability matrices of state transitions, observation symbols, and the initial state of the HMM process. A discrete HMM can be presented as $\lambda = (A, B, \Pi)$,

where, $A = \{a_{ij}\}$ is a matrix of the state transition probability distribution, and a_{ij} specifies the probability that state S_i changes to state S_j ($1 \leq i, j \leq N$); $B = \{b_j(k)\}$ is the observation symbol probability distribution in state S_j , and it represents the probability that the system will output an observable symbol O_k at the state S_j ($1 \leq j \leq N; 1 \leq k \leq M$); Π is a vector representing the probability that each state is the initial state of the HMM process. There are three fundamental tasks in the HMM design: (1) Given an observation sequence, compute the probability with which those observations can be generated by a given HMM model; (2) Determine the most likely sequence of internal states in a given model which will give rise to a given observation sequence; (3) Adjust the model parameters of an HMM to optimize the probability distribution matrices for a given set of observations. The details of these tasks are described in reference [7].

Multi-dimensional HMM

A multi-dimensional HMM for better recognition rates compared with one dimensional HMM is used in [6]. In this model, each dimension of the HMM state corresponds to the data from each sensor channel. The multiple data streams from the sensory glove and 3-D motion tracker are the inputs to the HMM process, and this raw data corresponds to the sequence of observations in the HMM. There are 21 channels of raw data, so we define one 21-dimensional state for each ASL alphabet in the HMM model. The probability matrices of the HMM (A, B, Π) are specified by clustering Gaussian distribution. A 5-Bakis-state HMM is used in the system (Figure 3), i.e. one state in an HMM can reach the same state or one of next two states. For instance, 5 consecutive

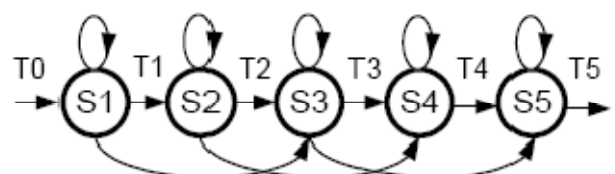


Figure 3. - Bakis -state HMM.

readings obtained from time instant T_1 to T_5 are fed into the 5-state HMM process. The first reading at T_1 is recognized as S_1 . The second reading data is then recognized as either the same state S_1 (the sensory data does not change at time T_1 and T_2), or a new state S_2 (the sensory data at T_2 is different from the one at T_1). In this case of alphabet recognition, these successive readings are typically classified as the same state if the user intends to gesture a certain ASL alphabet. Due to the large amount of words spelled

by alphabets, we assume equal probabilities for the transition of internal states and the initial state of an HMM process. By clustering the Gaussian distribution, we define multiple B matrices for the dimensions of each HMM state.

Since a discrete HMM is used, it is necessary to represent a gesture as a sequence of discrete symbols. We must preprocess the raw gesture data, which in this case are the values of the 15 bending and abduction angles, 3 position coordinates and 3 orientation angles, which are obtained from the Cyber glove and the Flock of Birds tracker [6]. The retrieval rate of the raw data is 40Hz and the system segments the data stream based on a velocity trigger. The segmentation procedure starts when the hand is stationary, i.e. the speed of the hand of the user is below a preset threshold, and ends when the speed gets above a threshold.

6. Comparison of results

a. Gesture recognition by utilizing bio-mechanical characteristics

Gesture Recognition based on Biomechanical Characteristic' provides higher accuracy, it addresses detecting the similar hand gestures without having them labeled, a problem that most traditional classification methods fail to address.

b. SLR Based on Skin Color

This gesture recognition method based on skin color is an intelligent & simple system for converting sign language into voice signal using head & hand gesture. This system is very simple & subject is not required to wear any glove. But the subject must wear a dark color, long sleeve shirt. This system has minimum and maximum classification rates as 88.47 and 95.69 respectively.

c. Neural Network Approach

The system is designed to recognize simple gestures or signs. The design is very simple and does not require any kind of gloves to be worn. Also the system is applicable to different backgrounds. This sign language recognition approach requires a computer with at least 1GHz processor and at least 256 MB of free RAM. The training set consists of all alphabets A to Z (26 patterns).

d. SLR using Hidden Markov Models

The sign language recognition system using HMM can perform online training and real-time recognition of ASL alphabets and basic hand shapes. The evaluation results show that the proposed method allows fast training and online learning of new

gestures and reliable recognition of the trained gestures afterwards.

7. Conclusion

There are various methods for sign language recognition. Four such methods are discussed and compared. The skin color based method is simple & does not require the signer to wear any glove. The accuracy of this method is also good. The gesture recognition by utilizing bio-mechanical characteristics requires cyber gloves to be worn by the signer. The accuracy is less here. The neural network approach provides greater speed in recognizing the sign language. But, a large amount of labeled examples are required to train the network for accurate recognition. Each of these four methods has its own pros and cons. So, depending on the type of application, appropriate method should be chosen.

8. References

- [1] Farid Parvini, Dennis McLeod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, Shahram Ghandeharizadeh Computer Science Department University of Southern California Los Angeles, California 90089-0781 [fparvini,mcleod,cshahabi,navai,bzali,shahram]@usc.edu
"An Approach to Glove-Based Gesture Recognition"
- [2] Paulraj M P, Sazali Yaacob, Hazry Desa, Hema C.R., "Extraction of Head & Hand Gesture Feature for Recognition of sign language", International Conference on Electronic Design, Penang, Malaysia, December 1-3, 2008.
- [3] Paulraj M P, Sazali Yaacob, Mohd Shuhanaz bin Zanar Azalan, Rajkumar Palaniappan, "A Phoneme based sign language recognition system using skin color segmentation", Signal Processing and Its Applications (CSPA), pp: 1 – 5, 2010.
- [4] Priyanka Mekala, Ying Gao , Jeffrey Fan , Asad Davari "Real-time Sign Language Recognition based on Neural Network Architecture" 978-1-4244-9592-4/11/\$26.00 ©2011 IEEE
- [5] Jong Bae Kim,Hye Sun Park,Min Ho Park,Massimo Piccardi,'Background subtraction techniques: a review',Systems, Man and Cybernetics,vol.4,2004,IEEE International Conference,pp:3099-3104.
- [6] HONGGANG WANG , MING C. LEUAND CEMIL OZ , "American Sign Language Recognition Using Multi-dimensional Hidden Markov Models" JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 22, 1109-1123 (2006)
- [7] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, Vol. 77, 1989, pp. 257-286.
- [8] P. Mekala, S. Erdogan, Jeffrey Fan, "Automatic object recognition using combinational neural networks in surveillance networks", IEEE 3rd International Conference on Computer and Electrical Engineering (ICCEE'10), Chengdu, China, Vol. 8, pp. 387-391, November 16-18, 2010.