ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
Research Article
www.ijptonline.com

MULTIMODAL FEATURES AND HYBRID CLASSIFICATION FOR INDIAN SIGN LANGUAGE RECOGNITION

M. Suresh Anand[1], Dr. N. Mohan Kumar[2], A. Kumaresan[3]
[1]Research Scholar, Anna University, Chennai, TamilNadu, India
[1]Assistant Professor, Department of Computer Science and Engineering,
Sri Sai Ram Engineering College, Anna University, Chennai, TamilNadu, India.
[2]Professor, Department of Electronics and Communication Engineering,
SKP Engineering College, Anna University, Tirvannamalai, TamilNadu, India.
[3]Assistant Professor, Department of Computer science and Engineering,
SKP Engineering College, Anna University, Tirvannamalai, TamilNadu, India.
Email: suresh.anandm@rediffmail.com

Received on: 03-02-2017 *Accepted on: 12-03-2017*

Abstract

In this paper, an efficient approach for Indian Sign Language Recognition (ISLR) system is presented. A methodology to extract features from two different modalities; hand gesture images and Non-Audible Murmur (NAM) speech signal is discussed. Discrete Wavelet Transform (DWT) is used for extracting image based features and Mel Frequency Cepstral Coefficients (MFCC) is used for signal based features. These two features are independently analyzed and also fused together for the recognition. Two classification approaches; K Nearest Neighbor (KNN) and Hidden Markov model (HMM) are independently applied on the extracted features for the recognition and further improvement on recognition accuracy is made by hybrid classification with multimodal features. The proposed ISLR system has the ability to recognize the Indian sign language with a high accuracy of 88.06%.

Keywords: Sign language recognition, hand gesture, NAM microphone, KNN and HMM.

1. Introduction

The two basic forms of communication between people are speaking and hearing. However, there are some problems occurred to the physically challenged people like deaf and dump people. To ease the communication between them, an efficient approach for ISLR is proposed. Recently, many algorithms are designed to recognize sign language. Some of them are outlined in this section. An approach for ISLR by the use of single and double handed gesture images with camera is discussed in **[14]**. It uses histogram of gradients and geometric descriptors as features with

KNN and support vector machine as classifiers. The different backgrounds are considered for the analysis. Discrete cosine transform based hand gesture recognition for ISLR is discussed in [1]. In the preprocessing stage, the acquired images are enhanced and then converted into gray scale format. Region and pixel based segmentation approaches are used to segment the skin region for feature extraction. Finally, recognition is achieved using neural network classifier. The detection of edges and peaks in the hand gesture images are used for ISLR system in [3]. Only six alphabets are tested with KNN search algorithm. The edges are detected by canny operator. An ISLR system using wavelet transform is described in [2]. It uses video sequences for the recognition. The noises in the frames are removed by Gaussian filter and segmentation is achieved by canny operator, DWT with thresholding. Elliptical Fourier descriptors and fuzzy based system are used for feature extraction and classification respectively.

HMM based hand gesture recognition is discussed in [4]. Manual features by tracking, non tracking and hand shape with non manual features are extracted and these are used to train the HMM classifier. Wavelet based ISLR system is discussed in [9]. Wavelet transform based features with binary features are extracted from the segmented hand region and KNN classifier is used for the recognition of only 13 English alphabets. All the distance measures in KNN are analyzed.

Non-Audible Murmur (NAM) microphone was developed by Nakajima **[5]**. The speech signals obtained from NAM are used for effective sign language recognition with HMM classifier. Wavelet transform based voice conversion from NAM speech signal is discussed in [6]. The speech signal is preprocessed for de-nosing, echo cancellation and also framing and windowing. The preprocessed signals are decomposed using DWT to obtain features. Based on the standard deviation, the corresponding accurate speech signal is obtained.

HMM distances and principal component analysis based Japanese language recognition using NAM signals is described in [7]. The experiments are conducted using the stethoscope and also microphones regarding the recognition of NAM speech. In [8], a probabilistic model is designed to convert the NAM signal to speech signal. Acoustic features of NAM are extracted and these features are modeled by a Gaussian Mixture Model (GMM) for the recognition. NAM speech signal is recognized by HMM modal in [10]. Self designed microphone is used to acquire NAM speech signal. A transition movement model is designed for the recognition of sign language in [11]. It uses a temporal clustering algorithm to cluster mass transition movements and then iterative segmentation method is used to segment transition parts. HMM based American sign language recognition is described in [12] using videos. The language is recognized by tracking hands, angle of axis of least inertia and eccentricity of bounding ellipse in each video frame. Speaker adaptive training based NAM speech signal recognition is discussed in [13]. It uses factorized transforms for extracting acoustic characteristics and analyzed on both NAM speech signal and normal speech. Acoustic features based NAM to speech and NAM to whisper conversion is implemented in [15]. GMM classifier is trained with spectral envelope at each speech frames.

In this paper, an efficient ISLR system is presented which uses hand gesture images and NAM signal with two classifiers; KNN and HMM. The rest of the paper is organized as follows: Section 2 discusses the methods and materials used in the proposed ISLR system and the results obtained by the system are discussed in section 3. Finally, conclusion is given in section 4.

2. Methods and Materials

The overall flow of the proposed ISLR system is shown in Figure 1. The proposed system extracts features from two different modalities; hand gestures and NAM signals. These two features are analyzed independently using KNN and HMM classifiers and both features and classifier are hybridized for effective ISLR system.

2.1 Features from Hand Gesture Images

DWT based energy features and hand area features are extracted from the hand gesture images [9]. It consists of two sequential processes; hand region segmentation and feature extraction. Before applying these processes, the acquired images are transformed into gray scale format in order to ease the computation process.

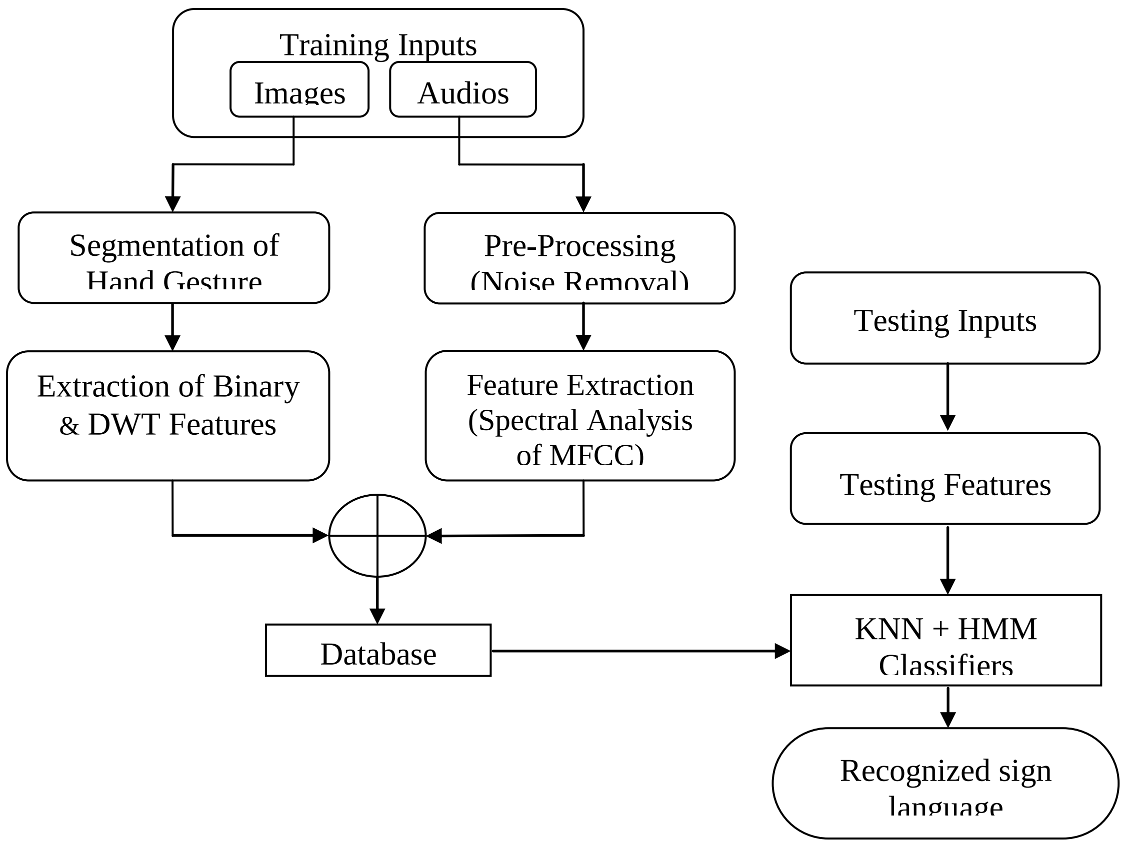

Fig.1. Overall flow of the proposed ISLR system.

An effective background and foreground segmentation based on Otsu thresholding [16] is employed. It is based on the intensity value of the images. To use this thresholding approach, the images are acquired with clear background and foreground information as shown in Figure 2 (a). The corresponding gray scale image and thresholded images are shown in Figure 2 (b) and (c).

(a) (b) (c)

Figure 2 (a) Hand gesture image (b) RGB to Gray conversion (c) Otsu segmentation

After segmentation, only hand region is cropped and processed using mathematical morphological operations such as connected component analysis, flood fill operation. The detailed segmentation process can be obtained from [9]. The area of hand region is computed from the result Otsu segmentation by counting the number of white pixels. In order to compute the DWT energy features, the segmented region is superimposed with the original hand image and only the hand gesture image is obtained. Figure 3 shows the extracted hand gesture image.

Figure 3 Hand gesture image of Figure 2(a)

As DWT, a multiresolution analysis, the segmented hand gesture image can be viewed in different resolution level. In each level, energy is computed form the obtained wavelet sub-bands by the following eqn.

$$E_k = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\mathrm{wavlet\_sub\,band}_k(i,j)\right| \tag{1}$$

where M, N is the width and height of the k^{th} sub-band and $|wavelet\_subbnad_k(i,j)|$ is the pixel value of the k^{th} sub-band at location *(i,j)*. The obtained energy features and the hand area are combined and used as features for the given hand gesture image.

2.2 Features from NAM speech signal

To acquire NAM speech signal, stethoscope based NAM microphone using the concept of Nakajima [5] is designed. It captures an inaudible murmur. From this signal, features are extracted for the recognition. Among the various features used in speech recognition, MFCC is one of the features which have proved highly successful in the recognition. Hence, in the proposed NAM speech signal based ISLR system, MFCC feature is extracted. The following steps are applied to get the MFCCs [17].

1. The given NAM speech signal is divided into number of frames by the use of Hamming window with a size of 30 milliseconds and a shift of 10 milliseconds.
2. Apply Fourier transform to each frame to obtain frequency spectrum
3. From the spectrum, calculate the power according to Mel filter bank.
4. At each Mel scale, take logarithm of each power
5. Compute discrete cosine transform of all energies
6. Take the amplitudes which gives MFCCs

The obtained MFCCs are used as spectral features for the corresponding NAM signal. It includes various information such as bandwidth, shape of vocal tract and spectral energy. These features are combined with the hand gesture features to improve the accuracy of the proposed ISLR system.

2.3 Classification

In this stage, two different classifiers; KNN and HMM are used for the recognition. In KNN, the hypotheses for the recognition are based on nearest neighbor using the distance measure such as Euclidean, city block, cosine, and correlation. HMM is applied in many pattern recognition approaches such as handwriting, speech, and hand gesture recognition. The first step for the classification is to divide the dataset for each class into two sets. Then train the HMM per class using one set. The remaining set is used for testing which is based on the likelihood of each model. They are related by the use of Markov process rather than independent of each other.

3. Results and Discussions

The validation of the proposed system is carried out using hand gesture images and NAM signal. The hand gestures are acquired using C170 Logitech digital camera. In order to differentiate the hand gesture from the background, all the persons are asked to wear a black jacket. 100 hand gesture images per alphabet is captured from different individuals. Similarly NAM speech waveforms are obtained from the same individuals. The proposed hand area with wavelet features are extracted from the hand gestures and MFCC features are extracted from the NAM speech signals. Two different classifiers; KNN and HMM are used for the classification. The proposed ISLR system recognizes one of the English alphabets. Hence, it is a 26-class problem and in order to train the classifier efficiently, half of the samples in each English alphabet are used. The evaluation is carried out using the remaining samples. The selection of training and testing samples are based on random selection and it is repeated for 10 times. Then the average accuracy is computed for the analysis. It is defined by

$$\text{Recognition Accuracy}(\%) = \frac{\#\text{correctly recognized sign language}}{\#\text{total test inputs}} x100 \qquad (1)$$

Initially, the performance of the proposed ISLR system is analyzed by using KNN classifier. The proposed features for both hand gesture and NAM speech signals are tested independently. Then the features are fused by the fusion approach and their performances are analyzed. Table 1 shows the accuracy of the proposed ISLR system using KNN classifier.

Table 1. Accuracy of the proposed ISLR system using KNN classifier.

Alphabets	**NAM signal**	**Hand Gesture Images**	**Fusion**	**Alphabets**	**NAM signal**	**Hand Gesture Images**	**Fusion**
A	64	65	78	N	60	71	71
B	58	73.80	69	O	57	75.60	75.60
C	55	65	66	P	56	72.60	63
D	56	72	74	Q	60	70.60	68.40
E	58.20	67	79.20	R	57.80	72.60	72.60
F	57	68	72.60	S	55.20	72	76.20
G	63	70	70	T	65	74.60	78.60
H	56.20	73	84.60	U	58	62	68
I	62.60	72.20	67	V	59.60	72	76
J	58	75	71	W	57	70	65
K	59.80	66	74	X	59	71.60	69.60
L	66	68	68	Y	56.40	73	71
M	58	70	80	Z	58	70	78
				Average	58.88	70.48	72.55

It is observed from Table 1 that the proposed ISLR system provides only 72.55% accuracy while using KNN classifier with the fusion approach and their individual counterpart gives 58.88% (NAM speech signal) and 70.48% (Hand gestures). The same features are given to the HMM classifier and their results are given in table 2.

Table 2. Accuracy of the proposed ISLR system using HMM classifier.

Alphabets	NAM signal	Hand Gesture Images	Fusion	Alphabets	NAM signal	Hand Gesture Images	Fusion
A	64	75	78	N	63	78	75
B	60	73.80	75	O	59.20	86	87
C	58	87	89	P	58	78	79
D	61	82	83	Q	65	82	84
E	60.20	78	79.20	R	59.80	79.40	81.20
F	59.80	80.60	82.40	S	63.60	78	79
G	66	74.40	76.60	T	68	79	81.20
H	59.20	78	84.60	U	62	85	86
I	66	86	88	V	66	83	88
J	72	75	77	W	59	79.60	82
K	65.20	76	77	X	63	86	89.80
L	67	82	85	Y	59	81.80	84
M	61	77	80	Z	62	88.20	88.60
				Average	62.58	80.34	82.29

It is observed from Table 2 that there is approximately 10% increase in average accuracy by the HMM classifier in comparison with KNN classifier using the fusion approach. Also, the accuracies of the proposed ISLR system increase while using NAM speech signal and hand gesture images individually. In order to further analyze the proposed system, both the classifiers are hybridized and their performances are evaluated. Table 3 shows the accuracy of the proposed ISLR system using hybrid classification of KNN and HMM classifier.

Table 3. Accuracy of the proposed ISLR system using hybrid classifier.

Alphabets	NAM signal	Hand gesture Images	Fusion	Alphabets	NAM signal	Hand gesture Images	Fusion
A	64	88	92	N	66.60	86	88
B	65	87	90	O	62.80	87	96
C	64.40	87	89	P	66.40	92	94
D	66	82	95	Q	65	82	84
E	68.60	81	84	R	68.80	86.80	88
F	63	80.60	82.40	S	64.60	93	97
G	60.20	80	82	T	68	79	81.20
H	67.80	93	94	U	62.40	85	86

I	64.60	86	88	V	64	83	88
J	63.80	80	82	W	63.60	79.60	82
K	67.60	85.60	87	X	65	86	89.80
L	62.60	82	85	Y	69.60	81.80	84
M	64.60	90.20	92.60	Z	62	88.20	88.60
				Average	65.04	85.07	88.06

It is observed from table 3 that the proposed ISLR system using hybrid classifier gives a maximum average recognition accuracy of 88.06%. Also, the accuracy of each alphabet is over 82% by hybrid the classifier as well as the features obtained from both NAM and hand gesture images.

4. Conclusion

In this study, hybrid classification approach for ISLR system is presented by the use of KNN and HMM. It consists of two difference feature extraction approaches; signal processing approach using NAM speech signal and image processing approaches for hand gesture images. MFCC features from the NAM signal and wavelet based features along with hand area from the hand gesture images are extracted. The obtained features are given to hybrid classifier for recognizing all English alphabets. Results show that the proposed ISLR system provides 88.06% average accuracy using the hybrid features with hybrid classification.

7. Reference

1. Tewari, D. and Srivastava, S.K. (2012) “A Visual Recognition of Static Hand Gestures in Indian Sign Language Based on Kohonen Self-Organizing Map Algorithm”. International Journal of Engineering and Advanced Technology (IJEAT), 2, 165-170.
2. Kishore, P.V.V. and Kumar, P.R. (2012) “A Video Based Indian Sign Language Recognition System (INSLR) using Wavelet Transform and Fuzzy Logic”. IACSIT International Journal of Engineering and Technology, 4, 537-542.
3. Cooper, H., Holt, B. and Bowden, R (2011) “Sign Language Recognition”. In: Moeslund, T.B., et al., Eds., Visual Analysis of Humans, Springer, London, 539-562.
4. Kalsh, A. And Garewal, N.S. (2013) “Sign Language Recognition System”. International Journal of Computational Engineering Research, 3, 15-21.
5. Punikos Herucleous, Yoshituka Nakajima, Akinobu Lee, Hiroshi Suruwuturi, Kiyohiro Shikuno, “Accurate Hidden Markov Models for Non- Audible Murmur (NAM) Recognition Based on Iterative Supervised Adaptation”, 2003 IEEE.

6. K. Kalaiselvi, M. S. Vishnupriya, “Non - Audible Murmur (NAM) Voice Conversion by Wavelet Transform”, International Journal of Science and Research (IJSR), Volume 3 Issue 3, March 2014.
7. Panikos Heracleous, Viet-Anh Tran, Takayuki Nagai, And Kiyohiro Shikano, Fellow, IEEE, ”Analysis and Recognition of NAM Speech using Hmm Distances and Visual Information” in IEEE transactions on Audio, Speech, And Language Processing, vol. 18, no. 6, august 2010.
8. Daisuke Miyamoto, Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano, ”Acoustic Compensation Methods for Body Transmitted Speech Conversion”, 2009 IEEE.
9. Anand, M.S., Kumar, N.M. and Kumaresan, A. (2016) “An Efficient Framework for Indian Sign Language Recognition using Wavelet Transform”. Circuits and Systems, 7, 1874-1883.
10. Mathavan Suresh Anand, Nagarajan Mohan Kumar and Angappan Kumaresan, (2016),”Recognition of Nam Speech – Indian English Alphabets Using Self-Designed Nam Microphone with Hmm-Viterbi” in I J C T A, pp. 159-166.
11. Fang, Gaolin, Wen Gao, and Debin Zhao. "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models." IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 37.1 (2007): 1-9.
12. Starner, Thad, and Alex Pentland. "Real-Time American Sign Language Recognition from Video using Hidden Markov Models." Motion-Based Recognition. Springer Netherlands, 1997. 227-243.
13. Denis Babani, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano,” Acoustic Model Training for Nonaudible Murmur Recognition using Transformed Normal Speech Data”, 2011 IEEE.
14. Singh, Akanksha, et al. "Indian Sign Language Gesture Classification as Single or Double Handed Gestures." 2015 Third International Conference on Image Information Processing (ICIIP). IEEE, 2015.
15. Toda, Tomoki, Mikihiro Nakagiri, and Kiyohiro Shikano. "Statistical Voice Conversion Techniques for Body-Conducted Unvoiced Speech Enhancement." IEEE Transactions on Audio, Speech, and Language Processing 20.9 (2012): 2505-2517.
16. Otsu, N. (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9, 62-69.
17. Koolagudi, S.G., Rastogi, D. and Rao, K.S., 2012. Identification of language using mel-frequency cepstral coefficients (MFCC). Procedia Engineering, 38, pp.3391-3398.