Automatic Indian Sign Language Recognition for Continuous Video Sequence

Joyeeta Singha[1], Karen Das[2]

[1]National Institute of Technology, Silchar,
Silchar-788010, Assam, INDIA
joyeeta_singha90[at]yahoo.com
[2]School of Technology, Assam Don Bosco University
Airport Road, Azara, Guwahati - 781017, Assam, INDIA
karen.das[at]dbuniversity.ac.in

Abstract— Sign Language Recognition has become the active area of research nowadays. This paper describes a novel approach towards a system to recognize the different alphabets of Indian Sign Language in video sequence automatically. The proposed system comprises of four major modules: Data Acquisition, Pre-processing, Feature Extraction and Classification. Pre-processing stage involves Skin Filtering and histogram matching after which Eigen vector based Feature Extraction and Eigen value weighted Euclidean distance based Classification Technique was used. 24 different alphabets were considered in this paper where 96% recognition rate was obtained.

***Keywords**: Eigen value, Eigen vector, Euclidean Distance (ED),Human Computer Interaction, Indian Sign Language (ISL), Skin Filtering.*

Cite as: Joyeeta Singh, Karen Das "Automatic Indian Sign Language Recognition for Continuous Video Sequence", ADBU J.Engg.Tech., 2(1)(2015) 0021105(5pp)

1. Introduction

Sign Languages are natural languages which use various forms of expression for communication in daily life. For deaf and mute people, we can say that it is the only means of communication for them to interact with the other people. There has been widespread research in this field. Many researchers have already worked in this field like Japanese Sign Language, Korean Sign Language, American Sign Language, British Sign Language and many more. But till date, very few works have been done in Indian Sign Language recognition. Recognition of sign language is important not only from the engineering point of view but also for its impact on human society.

Unfortunately, many studies have shown that deaf and mute people are illiterate and become isolated because the normal people never try to learn these signs. Moreover, finding a qualified interpreter was always not possible. Thus automatic sign language recognition was developed where the computer was programmed to translate the sign language into some text format proving it to be reliable solution to the above problem. It also provided a replacement of speech for deaf and mute people.

Earlier researches [1-4] mainly focused on the recognition of static images of different hand gestures. In our previous works [1][3]we have used Karhunen-Loeve Transform [1] for recognition of different hand gestures, where the success rate obtained was 96% but had a limitation that only single hand gestures could be recognized. In [2] change in abduction angle and change in inflexion angle were considered as features. Recognition rate obtained in average was 93% but a limitation was it was suitable for single hand recognition. Artificial Neural Network was used to recognize the Ethiopian Sign language [3] where recognition rate obtained was 98.5% but was limited to single hand gestures. In [4][5] problem faced with two hand gestures was solved. Eigen value weighted Euclidean distance was used in [4] for classification of different static alphabets of Indian Sign Language and success rate obtained was 97%. In [5] gesture comprising of both hand could be recognized using Hidden Markov Model but an external accelerometer architecture was used. They attained a success rate of 94%. Nowadays, researchers are more interested in recognition of gestures in video or real time. [6] used Neural Network based features and Gaussian Hidden Markov Model to recognize the gestures in video sequences. Hidden Markov Model was used in [7] for which accuracy rate obtained was 99% but limitation was colored gloves were used in this paper. Paulraj [8] used moment invariants features and Artificial Neural Network for recognition of different gestures with an accuracy rate of 92.85%. In [9][10]system was designed to recognize Taiwanese Sign Language. In [9] Hidden Markov Model was used in real time and accuracy rate obtained was 84%. Limitations of this paper were use of data gloves and recognition of single hand gestures. Both static and dynamic hand gestures were being recognized in [10] with the use of Support vector Machines and Hidden Markov Model but this system has to be operated using color gloves.

Thus a system has been proposed to recognize different alphabets of Indian Sign Language for continuous video sequences. Our system has tried to remove the above said limitations of other related works and has attained a success rate of 96%.The experiment was carried out with bare hands and the result shows a vast improvement over other approaches. The work in [4] has been extended for video sequence in this paper.

Different alphabets of Indian Sign Language are shown in Fig.1 which involves the use of either single hand or both hands.

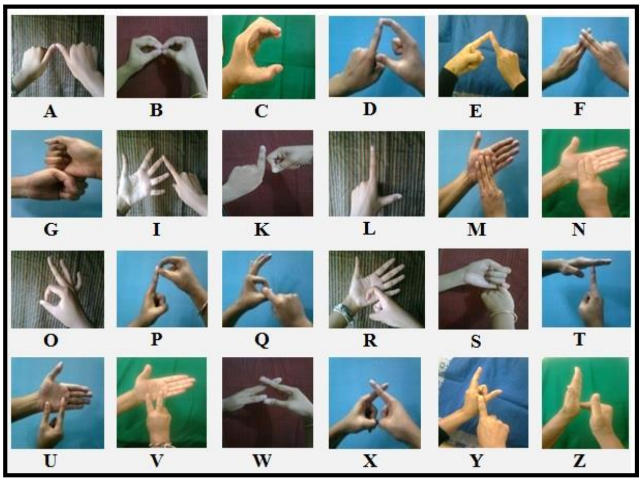

Figure 1: Indian Sign Language

2. Proposed Methodology

Fig.2 describes our proposed system which consists of four major phases- Data Acquisition, Pre-processing, Feature Extraction and Classification. Pre-processing phase includes Skin Filtering and histogram matching. Eigen vector were extracted as features and Eigen value weighted Euclidean distance based Classification was used which is briefly discussed in [4].

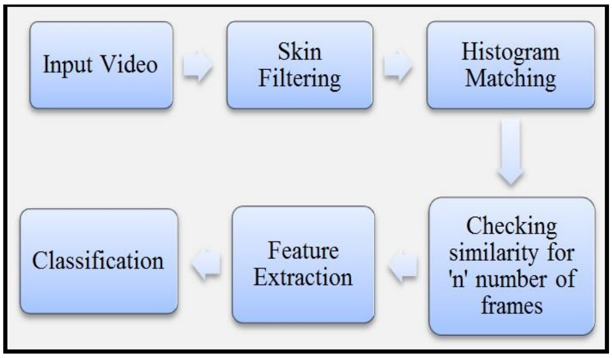

Figure 2: System Overview

2.1 Data Acquisition

The video is taken using webcam where different alphabets of the Indian Sign Languages were considered. Some of the video frames of a continuous video sequence are shown in Fig.3.We have captured video for each alphabet from 10 different people but for simplicity only a single video is shown in this paper.

2.2 Skin Filtering

After the video was captured by the webcam, skin filtering was performed in order to extract out the region of interest from the background. In this step, the skin colored pixels was separated from the non-skin colored pixels so that the hand regions can be extracted out from the surroundings.

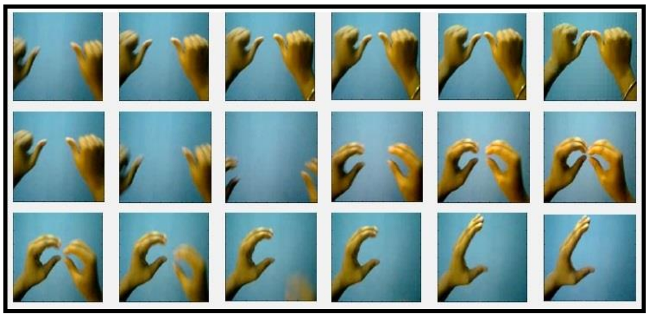

Figure 3: Some of the video frames acquired

Fig.4 describes the steps used in Skin Filtering as given in [4]. Briefly skin filtering includes steps like conversion of RGB image which in our case in the video frames to HSV color space and then filtering; smoothing and finally binary image was obtained. But in order to differentiate hands from other skin colored objects from the background Biggest BLOB was found out. Hence, the hand was extracted from the background which was processed in further steps.

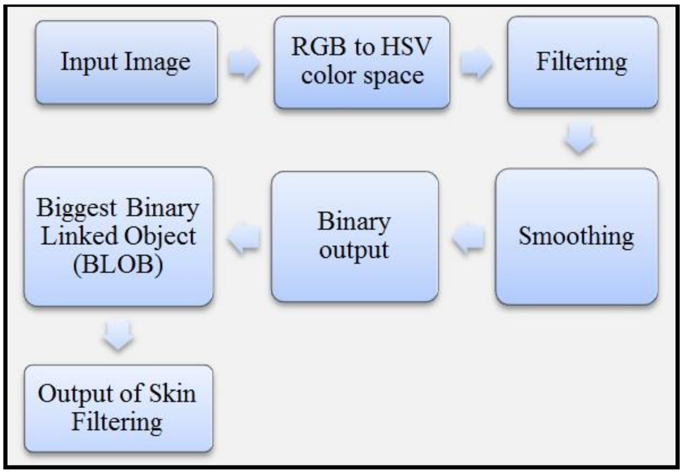

Figure 4: Skin Filtering

Some of the results obtained after skin filtering step after the video was started are given in Fig.5.

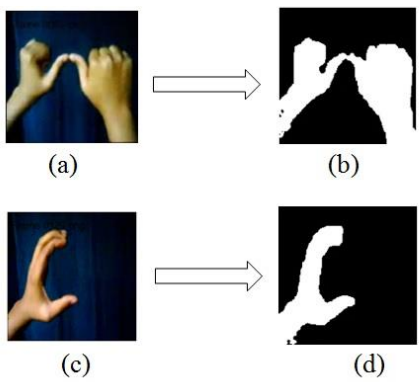

Figure 5: Skin filtering: (a) and (c) shows the input video frames and (b) and (d) shows the respective output after skin filtering

2.3 Histogram Matching

After the video frames are skin filtered, the next step is to find out the histogram for each frame and compare the histogram of consecutive frames in order to find out the similarity between them. If the histogram difference is below a threshold (th) value then we continue with the comparison between the next frames until 17 number of frames is reached. Thus we can say that we are processing a sign and then we proceed for the feature extraction stage. But if the histogram difference is above the threshold value then it means the two frames are not similar and we reject the previous frames and start with next frame. Fig. 6 shows the summary of the Histogram Matching step.

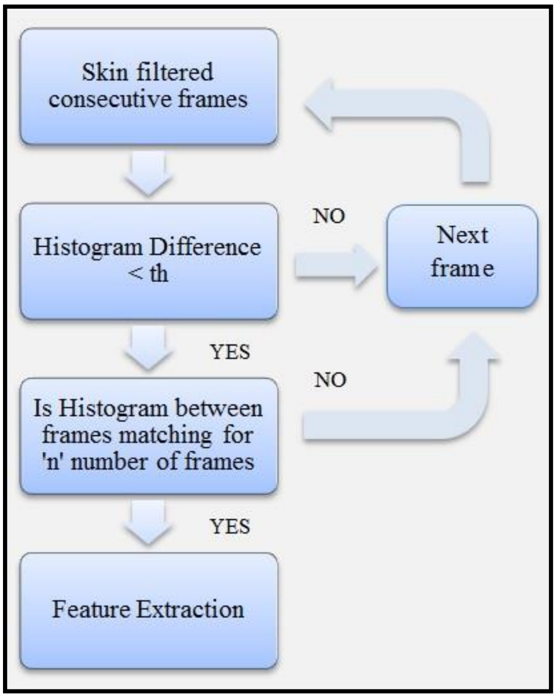

Figure 6: Summary of Histogram Matching step

2.4 Feature Extraction

Eigen values and Eigen vectors of the present frame are found out and matched with the Eigen values and Eigen vectors of the database images. They are found out mathematically as:

- Let the present frame be Y which is acting as the test image. The image has been resized to 70 by 70.
- Mean and covariance of the vector Y is found out as described in [4][11].
- Then finally the Eigen value and Eigen vector are being found out from the above computed covariance. The Eigen values are arranged in decreasing order.

In our proposed system, five principle Eigen values and five Eigen vectors were considered out of seventy resulting in reduction of dimension of the matrix. Thus data compression was achieved with only a little loss of information.

2.5 Classification

After the features were extracted, the next phase includes recognition of the last frame acquired by us during the video sequence. This was done using Classification based on Eigen value weighted Euclidean distance as given in [4]. Steps used in this classification were:

- Euclidean Distance was found out for every Eigen vector using the formula:

$$Euclidean\ Distance = \sqrt{\sum_{n=1}^{5}(EV1(n) - EV2(n))^{2}} \quad (1)$$

  where EV1 represents the Eigen vector of the current frame acquired and EV2 represents the Eigen vector of the images present in database.

- Difference between the Eigen value of the current video frame and the Eigen value of the trained images were found out. This difference was then multiplied with the Euclidean distance obtained.
- The sum of the results obtained for each image was added and minimum of all was found out which is the recognized symbol obtained.
- Finally, the result was displayed in the form of alphabet to which the current frame matches.

3. Experimental Results

Our proposed system was able to recognize 24 different alphabets of Indian Sign Language for continuous video sequences with an accuracy rate of 96%. For our system we have used MATLAB version 7.6 (R2008a) as software and Intel® Pentium® CPU B950 @ 2.10GHz processor machine, Windows 7 Home basic (64 bit), 4GB RAM and a webcam of resolution 320x240.Each video frame was tested with 240 different images i.e. 10 samples for each alphabets of Indian Sign Language.

Table I describes one of the video frame and its results obtained using Eigen value weighted Euclidean distance based classification technique for few images. Similar procedure is carried out for other video frames.

Again this proposed system was compared with other similar approaches and difficulties faced by the other approaches were discussed in Table II. Our proposed system was able to remove the difficulties faced by these approaches with a high accuracy rate of 96%.

4. Conclusion and Future work

Advantages like good accuracy, hand gesture recognition using single and both hands, performing with bare hands and video processing were achieved when compared with the previous works. Input video was skin filtered; each current video frame was matched with its consecutive frames. If matched for 17 video frames, it was forwarded to feature extraction and classification phase where calculation of Eigen value, Eigen vector, Euclidean distance, Eigen value weighted Euclidean Distance and sum was made. We have extended our work from still image recognition to video sequence recognition in this paper. We wish to extend our work further in real time with better accuracy. And moreover we have dealt with only alphabets of Indian Sign Language. We will try to extend it towards recognition of words and sentences.

TABLE I. EIGEN VALUE WEIGHTED EUCLIDEAN DISTANCE BASED CLASSIFICATION

Current videoframe	**Database image**	**Eigen value weighted ED(1st Eigen vector)**	**Eigen value weighted ED (2nd Eigen vector)**	**Eigen value weighted ED (3rd Eigen vector)**	**Eigen value weighted ED (4th Eigen vector**	**Eigen value weighted ED (5th Eigen vector)**	**Sum**	**Recognized symbol**
	A	0.0826	1.1543	0.3821	1.4290	0.0975	**3.1155**	
	B	2.1895	4.6304	1.7922	1.6347	0.0347	10.2815	
	C	0.6578	5.0470	0.2782	0.4448	0.0200	6.4478	
	D	8.7941	4.6638	0.4544	0.5875	0.8638	15.3636	
	E	0.3944	5.1792	1.0782	0.3894	0.0230	7.0642	
	F	1.4021	2.5786	0.8992	0.2134	0.0752	5.1685	
	G	0.5953	5.3613	2.5118	0.4610	0.4517	9.3811	
	I	4.3281	1.6915	0.2347	1.2731	1.4047	8.9321	**“A”**
	K	4.9311	2.1584	0.4284	0.3202	0.7027	8.5408	
	L	2.6093	0.9297	0.6307	0.4778	0.3848	5.0323	
	M	0.8821	3.6198	0.9947	0.3509	0.0757	5.9232	
	N	0.8390	3.1534	0.2733	0.3164	0.0331	4.6152	
	O	1.1857	0.2525	4.2474	1.2707	0.5076	7.4639	
	P	1.8514	4.1573	0.6155	0.4069	0.4883	7.5194	
	Q	1.4926	0.1010	0.0369	1.0951	0.8603	3.5859	
	R	4.7437	2.2137	0.4400	0.1312	0.3470	7.8756	
	S	5.4566	1.5585	2.1420	0.3504	0.2013	9.7088	
	T	1.1238	1.5760	1.8215	0.1843	0.0432	4.7488	
	U	1.1073	1.0829	0.3421	1.0141	0.4446	3.9910	
	V	0.0214	1.4439	1.2304	1.5067	0.1617	4.3641	
	W	2.6469	2.3819	1.0448	3.7922	1.2076	11.0734	
	X	9.0963	4.3775	0.5477	0.5366	0.2994	14.8575	
	Y	1.7486	2.5087	0.3461	0.6502	0.6318	5.8854	
	Z	3.3249	5.9975	0.0796	0.1369	0.0078	9.5467	

TABLE II. COMPARITIVE STUDY BETWEEN OUR WORK AND OTHER APPROACHES

Name of the technique used	Success Rate	Remarks
Karhunen-Loeve Transform [1]	98%	Recognition of only single hand gestures. Few hand gestures considered. Some gestures could not be recognized. Worked on static images.
Eigen value weighted Euclidean distance [4]	97%	Recognition of both single and two hand gestures was made possible. Worked on static images.
Hidden Markov Model [7]	99%	Worked on video sequence. Use of colored gloves-a limitation. Recognition of American Sign Language with single hand gestures.
HMM [9]	84%	Though worked on real time but very Low accuracy rate Use of Data Gloves Recognition of single hand gestures
Our work	**96%**	Worked on video sequences. Recognition of both single and two hand gestures. High accuracy rate in video processing. Use of bare hands.

Acknowledgement

We are sincerely thankful to Dr. Sunandan Baruah, Head of the Department, Department of Electronics and Communication Engineering, DBCET, Assam Don Bosco University for his kind help and encouragement for the proposed work.

References

[1] J. Singha and K. Das, “Hand Gesture Recognition Based on Karhunen-Loeve Transform”, Mobile and Embedded Technology International Conference (MECON), January 17-18, 2013, India, pp. 365-371.

[2] M.K. Bhuyan, M.K. Kar, and D.R. Neog, “Hand Pose Identification from Monocular Image for Sign Language Recognition”, IEEE International Conference on Signal and Image Processing Applications, 2011, pp. 378-383.

[3] A.S. Ghotkar, R. Khatal, S. Khupase, S. Asati, and M. Hadap, “Hand Gesture Recognition for Indian Sign Language”, International Conference on Computer Communication and Informatics (ICCCI), Jan. 10-12, 2012, Coimbatore, India.

[4] J. Singha and K. Das, “Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique”, International Journal of Advanced Computer Science and Applications, Vol. 4, No. 2, March2013, pp. 188-195.

[5] H. Brashear, T.Starner, P.Lukowicz, and H. Junker, “Using Multiple Sensors for Mobile Sign Language Recognition”, Seventh IEEE International Symposium on Wearable Computers, 2003.

[6] Y.L. Gweth, C. Plahl, and H. Ney, “Enhanced Continuous Sign Language Recognition using PCA and NeuralNetwork Features”, IEEE, 2012, pp. 55-60.

[7] T. Starner and A. Pentland, “Real-Time American Sign Language Recognition from VideoUsing Hidden Markov Models”, IEEE, 1995, pp. 265-270.

[8] M.P. Paulraj, S. Yaacob, Mohd S.Z Azalan, and R.Palaniappan, “A Phoneme Based Sign Language RecognitionSystem Using Skin Color Segmentation”,6th International Colloquium on Signal Processing & Its Applications (CSPA), 2010, pp. 86-90.

[9] R.H. Liang and M.Ouhyoung, “A Real-time Continuous Gesture Recognition System for Sign Language”, IEEE International Conference on Automatic Face and Gesture Recognition, 1998, Japan, pp.558-567.

[10] B.L. Tsai and C.L. Huang, “A Vision-Based Taiwanese Sign Language Recognition System”, IEEE International Conference on Pattern Recognition, 2010, pp. 3683-3686.

[11] J. Singha and K. Das, “Recognition of Indian Sign Language in Live Video”, International Journal of Computer Applications, Vol. 70, No. 19, May 2013, pp. 17-22.

Author Profile

Joyeeta Singha received her M.Tech degree in 2013 from Assam Don Bosco University, Guwahati and B.E degree from G.I.M.T Guwahati in 2011. She is currently pursuing Ph.D in the Department of Electronics and Communication Engineering at National Institute of Technology, Silchar, India. Her research interests include Image and video processing, Hand gesture recognition.

Karen Das received his M.Tech degree from Tezpur University, Tezpur and is currently working as an Assistant Professor in School of Technology, Assam Don Bosco University,Guwahati, Assam, India. He is pursuing his Ph.D in the department of Electronics and Communication Engineering at School of Technology, Guwahati, Assam, India. His research interest includes Machine Learning, Image and video processing, Nano Electronics.