Sign Language Interpretation using Linear Discriminant Analysis and Local Binary Patterns

Mahmood Jasim
Department of Computer Science & Engineering
University of Dhaka
Dhaka - 1000, Bangladesh
Email: mahmood.jasim.0@gmail.com

Md. Hasanuzzaman
Department of Computer Science & Engineering
University of Dhaka
Dhaka - 1000, Bangladesh
Email: hzamancsdu@yahoo.com

***Abstract*—This paper presents a computer vision-based hand sign gesture recognition system for sign language interpretation. Haar-like feature-based cascaded classifier is used for hand area detection. Hand gestures portraying sign language are recognized using Linear Discriminant Analysis and Local Binary Pattern based feature extractors separately. The sign gestures are classified using Nearest Neighbor algorithm. For testing the system the Chinese and Bangladeshi Numeral Gesture datasets are prepared containing sign gestures describing the numerals of 0 to 9 for the respective languages. The mean accuracy of LDA based sign language interpretation on the Chinese numeral gesture dataset is 92.417% and on the Bangladeshi numeral gesture dataset is 88.55%. The mean accuracy of LBP based sign language interpretation on the Chinese numeral gesture dataset is 87.13% and on the Bangladeshi numeral gesture dataset is 85.10%.**

I. Introduction

Computer vision based automatic sign language interpretation have featured prominently in research for the last decade. This paper introduces a comprehensive system of recognizing hand sign gestures using computer vision based approach for sign language interpretation. The motivation behind using simple camera based image capture, rather than expensive devices like Microsoft Kinect or Sony's Wii-Mote is simplicity, availability and accessibility of the system to countries where these devices are not available. The challenges of sign language interpretation are segmentation of hand area from the image sequence, extracting important features for interpretation, understanding the pattern associated with each sign, etc. Different approaches have been taken to solve these challenges. For static hand gesture based signs, Principal Component Analysis (PCA) has been extensively used by many researchers. Huang et. al. used PCA to recognize sign gestures[1]. Lu et. al. used modifications of PCA to recognize sign gestures[2]. Other systems for sign language interpretation, Sterner et. al.'s American Sign Language interpretation[3], Liang and Ouhyoung's Taiwanese Sign Language interpretation[4], Ong's British Sign Language interpretation[5] etc. are notable. In this paper, Haar-like feature based cascaded classifier is used to segment the hand area[6]. Hand sign gesture features are extracted using Linear Discriminant Analysis (LDA) and Local Binary Patterns (LBP). Sign gestures classification is done using Nearest Neighbor algorithm (NN).

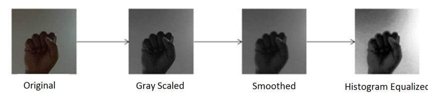

Fig. 1. Hand sub-image pre-processing steps.

This paper is organized as follows. The next section describes the proposed LDA and LBP based sign language interpretation system. Section III presents the Chinese numeral gesture dataset, Bangladeshi numeral gesture dataset and experimental results of the proposed system with appropriate discussion. The conclusion is given in Section IV.

II. Linear Discriminant Analysis and Local Binary Pattern based Sign Gesture Interpretation

The images containing the signs are captured using a fixed monocular camera. From the images, the hand areas are segmented using Haar-like feature based cascaded classifiers. The segmented area is then pre-processed by gray scaling, resizing the image to 100×100, Gaussian smoothing and Histogram Equalization. Fig 1 presents the pre-processing steps.

A. Linear Discriminant Analysis

The Linear Discriminant Analysis (LDA) performs a class specific dimension reduction.[7] It finds the combination that best separates different classes. To find the class separation, LDA maximizes both between class and within class scatters instead of maximizing the overall scatter. As a result same class members cluster together and different class members stay far apart from each other in the lower dimensions. Let, χ be a vector with samples from c classes.

Let, χ be a vector with samples from c classes.

$$\chi = \{\chi_1, \chi_2, ..., \chi_c\}$$

$$\chi_i = \{x_1, x_2, ..., x_c\}$$

The between class and within class scatters, S_B and S_W are calculated as follows.

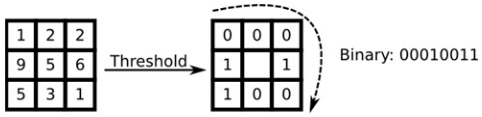

Fig. 2. Example of LBP code generation

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

$$S_W = \sum_{i=1}^{c} \sum_{x_j \in \chi_i} (x_j - \mu_i)(x_j - \mu_i)^T$$

Here, μ and μ_i are the mean of vector data and mean of the class i, where $i = 1, ..., c$.

$$\mu_i = \frac{1}{|\chi_i|} \sum_{x_j \in \chi_i} x_j$$

LDA finds a projection, W that maximizes the class separation criterion.

$$W = argmax_W \frac{|W^T S_B W|}{|W^T S_W W|}$$

The rank of S_W is at most $(N - c)$, where N is the number of samples and c is the number of classes. Almost always the number of samples is less than the dimension of the image data in pixels. Principal Component Analysis (PCA) is performed on the image data and projected on a $(N - c)$ dimensional space. LDA is performed on this reduced data. The transformation matrix, W projecting the sample in to $(c - 1)$ dimensional space is,

$$W = W_{fld}^T W_{pca}^T$$

Here,

$$W_{pca} = argmax_W |W^T S_T W|$$

$$W_{fld} = argmax_W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|}$$

B. Local Binary Pattern (LBP)

Local Binary Patterns (LBP) was proposed by Ojala et. al.[8] It performs local operations on the neighborhood of an image pixel. The neighborhood of a pixel is the pixels adjacent to that particular pixel. In LBP an 8 bit binary code is for a 3×3 pixel neighborhood of image I is,

$$b_j = \begin{cases} 1, & \text{if } (x_i, y_i) > (x_0, y_0) \\ 0, & \text{otherwise} \end{cases}$$

Here $-1 \leq i \leq 1$ and $j = 0...7$. In clockwise order, the neighborhood pixel values are thresholded against the center pixel to generate the 8 bit code. If the value is greater, the code is 1 and otherwise the code is 0. The process of generating LBP codes is shown in Fig 2.

An extension of the basic LBP operator with Uniform Patterns is used in the proposed system. An LBP pattern is called uniform if it possesses at most two transitions from 0 to 1 or 1 to 0. This system uses an LBP operator of 2 pixel radius with uniform patterns and 8 sample points, $LBP_{8,2}^{U2}$. The operator is presented in Fig 3.

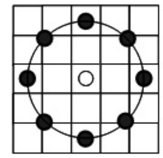

Fig. 3. LBP operator used in this system

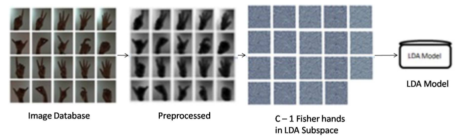

Fig. 4. Example of gesture model generation using LDA features.

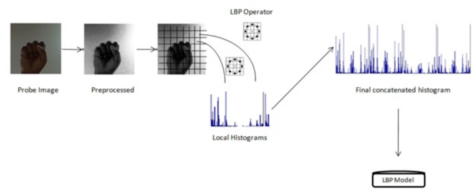

Fig. 5. Example of gesture model generation using LBP features.

This operator is used to produce a histogram of LBP codes. The image is divided into 8×8 regions and local histograms are generated from each region. The histograms are concatenated together to create the final LBP histogram. This histogram is used as the feature in recognizing the sign gestures.

C. Sign Gesture Model Training

For training and testing of the system, Chinese numeral gesture dataset and Bangladeshi numeral gesture dataset is used. The datasets will be described in Section III. Each dataset contains a total 2000 images of ten numeral signs from ten different people of respective language.
The images are loaded and pre-processed first. After the pre-processing, feature extraction is performed for respective models. For an LDA model, the images are projected onto the LDA space and for the LBP model; LBP histograms are generated and saved in the model. Both processes are shown in Fig 4 and Fig 5 respectively.

D. Sign Gesture Classification

Nearest Neighbor algorithm (NN) is used for sign gesture classification for both systems. The signs based on LDA features are projected on the LDA model. The Euclidean norm is used as the dissimilarity measurement to find the closest match. For LBP based classification, Chi-square difference is used as the dissimilarity measure. Ahonen showed the efficiency of Chi-square method as the dissimilarity measure for LBP histogram features[9].

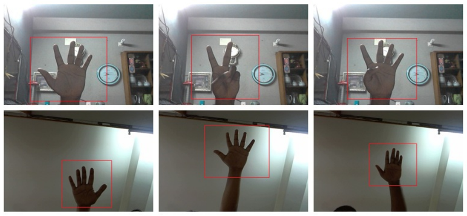

Fig. 6. Example of sign language recognition in different backgrounds.

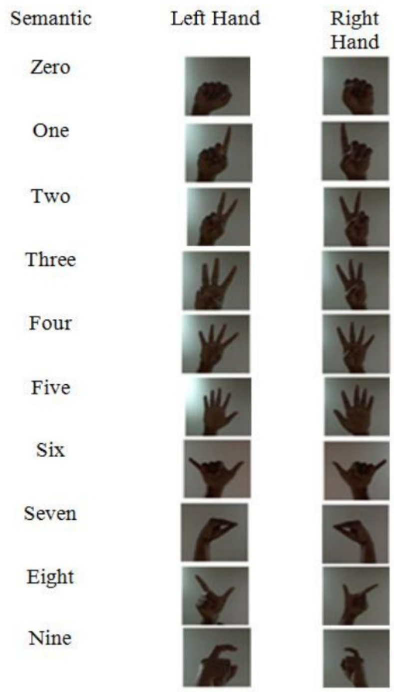

Fig. 7. An example set of Chinese numeral gesture dataset.

III. Experimental Result and Discussion

The system is tested in a moderate system containing Intel Core - i5 2400 processor with four physical cores, 8 GB of RAM and 500 GB of secondary storage. The images are captured using Logitech 310 web camera. All of the following experiments were conducted in both controlled and cluttered backgrounds. Fig 6 shows example of sign language recognition in different backgrounds. It shows the presence of static objects in the background do not reduce system performance.

A. Chinese Numeral Gesture Dataset

The Chinese numeral gesture dataset is a collection of ten Chinese numeral symbol gestures taken from ten different volunteers in different backgrounds. Each gesture is captured from both Left hand and Right hand of each volunteer. Each gesture was taken ten times. There are a total of 2000 gesture images from both the Right and Left hands. An example set of Chinese numeral gestures is shown in Fig 7.

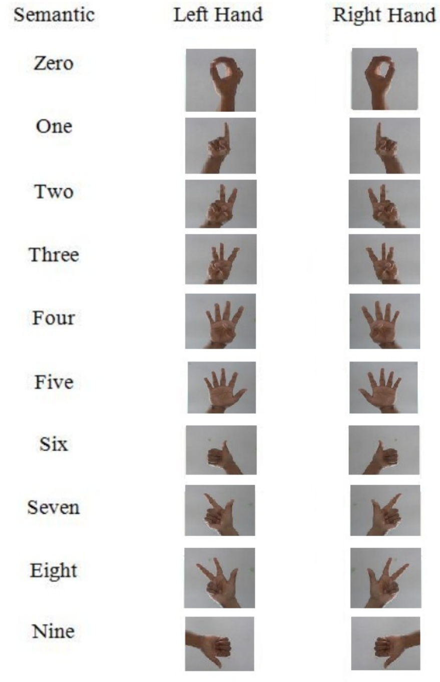

Fig. 8. An example set of Bangladeshi numeral gesture dataset.

B. Bangladeshi Numeral Gesture Dataset

The Bangladeshi numeral gesture dataset is a collection of ten Chinese numeral symbol gestures taken from ten different volunteers in different backgrounds. Each gesture is captured from both Left hand and Right hand of each volunteer. Each gesture was taken ten times. There are a total of 2000 gesture images from both the Right and Left hands. An example set of Bangladeshi numeral gestures is shown in Fig 8.

C. Performance of Sign Language Interpretation System

For sign language interpretation, N-fold cross validation method was used with the N = 5. For a single hand (left or right) each fold is consisting of 200 images. The system is trained using 800 images from four of the five folds and tested against the remaining fold of 200 images. For both hands each fold has 400 images. The system is trained using 1600 images from four of five folds and tested against the remaining fold of 400 images.

The system was tested under three criteria. Sign gestures performed by left, right and both hands. The accuracy of all three criteria is measured using the following condition, where N_C is the number correctly classifier sign gestures and N is the number of all test sign gestures.

$$Accuracy(\%) = \frac{N_C}{N}$$

Tables 1 and 2 present the accuracy comparison of the proposed LDA and LBP based systems respectively on Chinese numeral gesture dataset and Bangladeshi numeral gesture dataset using N fold cross validation based on the sign gestures performed by left hand with the existing Principal

TABLE I
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR CHINESE NUMERAL GESTURES PERFORMED BY LEFT HAND BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	82.50	89.50	85.25
Fold - 2	83.50	91.50	86.50
Fold - 3	83.25	90.00	86.00
Fold - 4	82.50	91.50	86.50
Fold - 5	82.00	89.25	85.00

TABLE II
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR BANGLADESHI NUMERAL GESTURES PERFORMED BY LEFT HAND BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	80.50	86.50	83.25
Fold - 2	81.25	88.25	84.50
Fold - 3	81.50	86.25	84.50
Fold - 4	80.50	87.50	84.00
Fold - 5	82.00	88.25	83.00

Component Analysis (PCA) based system. Tables 3 and 4 present the accuracy comparison of the proposed LDA and LBP based systems respectively on Chinese numeral gesture dataset and Bangladeshi numeral gesture dataset using N fold cross validation based on the sign gestures performed by right hand with the existing Principal Component Analysis (PCA) based system. Tables 5 and 6 present the accuracy comparison of the proposed LDA and LBP based systems respectively on Chinese numeral gesture dataset and Bangladeshi numeral gesture dataset using N fold cross validation based on the sign gestures performed by both hands with the existing Principal Component Analysis (PCA) based system.

The mean accuracy of LDA based sign language interpretation on the Chinese numeral gesture dataset is 90.35% for the sign gestures performed by left hand, 94.15% for the sign gestures performed by right hand and 92.75% for sign gestures

TABLE III
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR CHINESE NUMERAL GESTURES PERFORMED BY RIGHT HAND BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	87.25	94.00	90.00
Fold - 2	87.50	94.50	89.50
Fold - 3	88.00	94.50	89.50
Fold - 4	88.50	93.50	88.25
Fold - 5	87.25	94.25	89.00

TABLE IV
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR BANGLADESHI NUMERAL GESTURES PERFORMED BY RIGHT HAND BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	82.50	90.50	86.25
Fold - 2	83.50	90.50	87.25
Fold - 3	83.00	89.25	87.50
Fold - 4	82.25	88.50	86.00
Fold - 5	83.25	89.00	87.25

TABLE V
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR CHINESE NUMERAL GESTURES PERFORMED BY BOTH HANDS BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	84.50	92.50	87.00
Fold - 2	85.00	93.50	86.50
Fold - 3	85.25	93.00	86.50
Fold - 4	84.50	92.50	87.00
Fold - 5	84.00	92.25	86.00

TABLE VI
COMPARISON OF SIGN LANGUAGE INTERPRETATION SYSTEMS FOR BANGLADESHI NUMERAL GESTURES PERFORMED BY BOTH HANDS BASED ON ACCURACY

Fold Number	PCA (%)	LDA (%)	LBP (%)
Fold - 1	81.50	88.25	85.00
Fold - 2	82.25	88.50	84.50
Fold - 3	82.25	89.25	84.00
Fold - 4	81.00	88.50	85.50
Fold - 5	82.25	89.25	84.00

performed by both hands. The mean accuracy of LBP based sign language interpretation on the Chinese numeral gesture dataset is 85.85% for the sign gestures performed by left hand, 89.25% for the sign gestures performed by right hand and 86.60% for the sign gestures performed by both hands. The mean accuracy of LDA based sign language interpretation on the Bangla numeral gesture dataset is 87.35% for the sign gestures performed by left hand, 89.55% for the sign gestures performed by right hand and 88.75% for sign gestures performed by both hands. The mean accuracy of LBP based sign language interpretation on the Bangla numeral gesture dataset is 83.85% for the sign gestures performed by left hand, 86.85% for the sign gestures performed by right hand and 84.60% for the sign gestures performed by both hands. It is clear from the above tables that both systems performs better than the existing PCA based system, separately.

The efficiency is measured based on the time it takes for each systems to complete a single matching against the train set. It is important to note that the training time is not considered for the efficiency measurement. Only the time required to perform a single test against the train model is considered. Each fold has a total of 400 test images for both hands. The time to load an image, pre-process, extract features and recognize the gesture using a single test image against the training set is considered for performance measurement. Table 7 presents the computation cost comparison to perform a single matching amongst LDA and LBP based sign language

TABLE VII
COMPUTATION COST COMPARISON AMONGST SIGN LANGUAGE INTERPRETATION SYSTEMS FOR SIGN GESTURES PERFORMED BY BOTH HANDS

Systems	Left Hand (μs)	Right Hand (μs)	Both Hands (μs)
PCA	52.719	52.756	50.350
LDA	2.838	2.944	2.700
LBP	79.919	80.644	45.105

interpretation systems with the existing PCA based system.

IV. Conclusion

The paper presents a system for computer vision-based sign language interpretation. Linear Discriminant Analysis and Local Binary Patterns are used in this sign language interpretation system separately. LDA and LBP based systems have featured heavily in face recognition systems. In this paper, these systems are introduced in the hand sign gesture recognition genre. The results prove the effectiveness of the system. The mean accuracy of LDA based sign language interpretation system on the Chinese numeral gesture dataset is 92.417% and on the Bangladeshi numeral gesture dataset is 88.55%. The mean accuracy of LBP based sign language interpretation system on the Chinese numeral gesture dataset is 87.13% and on the Bangladeshi numeral gesture dataset is 85.10%. Both systems provides better performance as compared to the existing PCA based sign language interpretation system. Due to high performance and simplicity of the system, more complex sign language interpretation is possible if the system is trained accordingly. The future scope of the system is utilization in human-computer or human-robot natural interaction, controlling robot navigation and interaction with intelligent agents.

References

[1] C. Huang, S. Jeng, "A Model-based Hand Gesture Recognition System. Machine Vision and Applications", *Springer-Verlag 2001*, (2001), p. 243.

[2] W. L. Lu, J. J. Little, "Simultaneous Tracking and Action Recognition Using the PCA-HOG Descriptor", *IEEE Computer Society*, (2006), p. 6.

[3] T. Starner, A. Pentland, J Weaver," Real-time American Sign Language Recognition using Desk and Wearable Computer Based Video", *IEEE Transaction on Pattern Analysis and Machine Intelligence.*, 1998, p. 1371-1375.

[4] R. H. Liang and M. Ouhyoung, " A Real-Time Continuous Gesture Recognition System for Sign Language", *In Proceedings of 3rd International Conference on Automatic Face and Gesture Recognition*, 1998, p. 558-565.

[5] E. Ong, H. Cooper, N. Pugeault, R. Bowden, " Sign Language Recognition using Sequential Pattern Trees", *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.

[6] P. Viola, M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", *IEEE CVRP, 2001* (2001), p. 511.

[7] P. N. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1997), p. 711.

[8] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Pattern", *IEEE Transaction on pattern analysis and machine intelligence*, (2002), p. 971.

[9] T. Ahonen, A. Hadid, "Face recognition with Local Binary Patterns", *ECCV 2004*, (2004), p. 469.