International Journal of Image and Graphics

A Real - Time Computer Vision - Based Static and Dynamic Hand Gesture Recognition System

Mahmood Jasim*

Department of Computer Science & Engineering, University of Dhaka, 1, Lalbagh
Dhaka, Dhaka - 1000, Bangladesh
mahmood.jasim.0@gmail.com

Tao Zhang†

Department of Automation, School of Information Science & Engineering, Tsinghua University
Beijing 100084
China
taozhang@mail.tsinghua.edu.cn

Md. Hasanuzzaman‡

Department of Computer Science & Engineering, University of Dhaka, 1, Lalbagh
Dhaka, Dhaka - 1000, Bangladesh
hzamancsdu@yahoo.com

This paper presents a novel method for computer vision-based static and dynamic hand gesture recognition. Haar-like feature-based cascaded classifier is used for hand area segmentation. Static hand gestures are recognized using Linear Discriminant Analysis and Local Binary Pattern based feature extraction methods. Static hand gestures are classified using Nearest Neighbor algorithm. Dynamic hand gestures are recognized using the novel text-based Principal Directional Features (PDF), which are generated from the segmented image sequences. Longest Common Subsequence algorithm is used to classify the dynamic gestures. For testing, the Chinese numeral gesture dataset containing static hand poses and directional gesture dataset containing complex dynamic gestures are prepared. The mean accuracy of LDA based static hand gesture recognition on the Chinese numeral gesture dataset is 92.42%. The mean accuracy of LBP based static hand gesture recognition on the Chinese numeral gesture dataset is 87.23%. The mean accuracy of the novel dynamic hand gesture recognition method using Principal Directional Features (PDF) on directional gesture dataset is 94%.

Keywords: Hand gesture recognition; Linear Discriminant Analysis; Local Binary Pattern; Principal Directional Features; Chinese numeral gesture; Directional gesture; Nearest Neighbor.

*Department of Computer Science & Engineering, University of Dhaka, 1, Lalbagh, Dhaka, Dhaka - 1000, Bangladesh
†Department of Automation, School of Information Science & Technology, Tsinghua University, Beijing 100084, China
‡Department of Computer Science & Engineering, University of Dhaka, 1, Lalbagh, Dhaka, Dhaka - 1000, Bangladesh

1. Introduction

Computer vision based automatic hand gesture recognition systems have featured prominently in research for the last decade. This paper introduces a novel system of recognizing hand gestures using computer vision based approach. 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 hand gesture recognition include segmentation of hand area from the image sequence, extracting important features for static gesture classification, capturing the motion associated with the gestures and the interpretation of the motion to recognize the dynamic gestures. Different approaches have been taken to solve these challenges. To segment the hand area from the images, color gloves based methods have been applied with limited success by Keskin et. al.[1] Skin color based methods were used to segment the hand area by Manresa et. al.[2] A more recent approach is using Haar-like feature based cascaded classifiers by Chen et. al.[3] For static hand gesture recognition, Principal Component Analysis (PCA) has been extensively used by many researchers. Huang et. al. used PCA to recognize static hand gestures.[4] Lu et. al. used modifications of PCA to recognize static hand gestures.[5] Other methods including Linguistic based Framework by Derpanis et. al.,[6] Local Orientation Histogram Feature Description Model by Zhou et. al.,[7] etc. are notable. For dynamic hand gesture recognition, Finite State Machine (FSM) or Hidden Markov Model (HMM) based techniques are popular. HMM was used by Huang et. al.[4] Variation of HMM was used by Bowden et. al. for dynamic hand gesture recognition.[8] In this paper, Haar-like feature based cascaded classifier is used to segment the hand area. Static hand gesture features are extracted by Linear Discriminant Analysis (LDA) and Local Binary Patterns (LBP). Static hand gestures are classified using Nearest Neighbor algorithm (NN). For dynamic hand gesture recognition, the novel text based Principal Directional Features (PDF) are used to extract directional features and Longest Common Subsequence (LCS) is used to classify the hand gestures from the PDF. The use of LCS is a novel technique to utilize text string matching to classify gestures from text based features.

This paper is organized as follows. The next section describes the Haar-like feature based cascaded classifiers, used in detecting the hand area from the images. Section 3 presents the proposed LDA and LBP based static hand gesture recognition system. Section 4 presents the novel Principal Directional Feature (PDF) based dynamic hand gesture recognition system. Section 5 presents the Chinese numeral gesture dataset, directional gesture dataset and experimental results of the proposed system with appropriate discussion. The conclusion is given in Section 6.

2. Hand Area Segmentation using Haar- like Features

Haar-like feature based hand area segmentation is a statistical approach towards segmenting the hand area. It concentrates on certain areas of an image, not indi-

vidual pixel values. The concept of "Integral Image" is used to compute a set of Haar-like features introduced by Viola and Jones.[9] It achieves true scale invariance and eliminates the need for a multi-scale image pyramid for different scales of object image. This algorithm selects features based on AdaBoost learning by Friedman et. Al.[10] The Viola Jones method performs 15 times faster without sacrificing the accuracy compared to other object detection methods.
Haar-like features calculate the difference between light and dark regions within a kernel. Every Haar-like feature is consisting of two to three light and dark rectangles. The rectangles are interconnected with each other. The value of Haar-like feature is the difference between the sums of pixels values of the dark and light rectangles. The accuracy of a single Haar-like feature is not sufficient. A better result is achieved by using a series of weak classifiers. The AdaBoost learning algorithm is used to increase the accuracy of these weak classifiers. Initially, AdaBoost trains a weak classifier using a single Haar-like feature, which achieves the best performance for all the training samples. In the next iteration, the misclassified samples in the first iteration are weighted up. Finally a cascade of linear combination of the selected weak classifiers which is a strong classifier is achieved. This strong classifier is capable of achieving better accuracy.

Fig. 1. Examples of positive images for Haar-like feature based cascaded classifier training.

Chen proposed to train individual classifiers to recognize the respective hand pose.[3] This paper proposes to use a common classifier to segment the hand position from the image.

In this system, a Haar-like feature based cascaded classifier is generated for the proposed system using 3000 positive images from simple and plain background and 2000 negative images containing landscape, human faces and other complex non-hand objects. Examples of positive and negative images are shown in Fig 1 and Fig 2 respectively. The object size was set to 20×20 and OpenCV library by Bradsky

Fig. 2. Examples of negative images for Haar-like feature based cascaded classifier training.

was used to train the classifier.[11] The extended Haar-like feature set containing 14 feature prototypes proposed by Lienhart is used to generate the Haar-like feature based cascaded classifier.[12] Fig 3 shows Lienhart's extended Haar-like feature set.

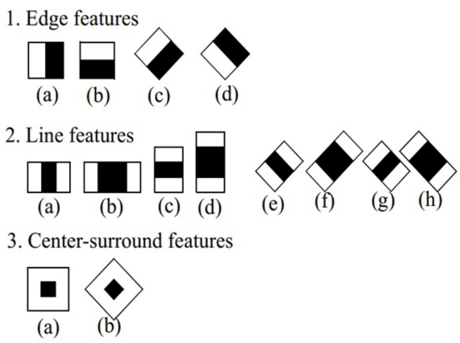

Fig. 3. The extended set of Haar-like features.

3. Linear Discriminant Analysis and Local Binary Pattern based Static Hand Gesture Recognition

The images containing the gesture are captured using a fixed monocular camera. From the images, the hand areas are segmented using Haar-like feature based cascaded classifier, described in Section 2. The segmented area is then pre-processed by gray scaling, resizing the image to 100×100, Gaussian smoothing and Histogram Equalization. Fig 4 presents the pre-processing steps.

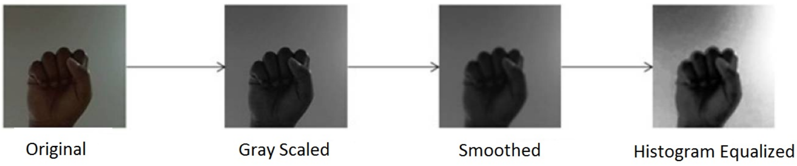

Fig. 4. Hand sub-image pre-processing steps.

3.1. *Linear Discriminant Analysis*

The Linear Discriminant Analysis (LDA) performs a class specific dimension reduction.[13] 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.

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

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

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,

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

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

3.2. *Local Binary Pattern (LBP)*

Local Binary Patterns (LBP) was proposed by Ojala et. al.[14] 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 5.

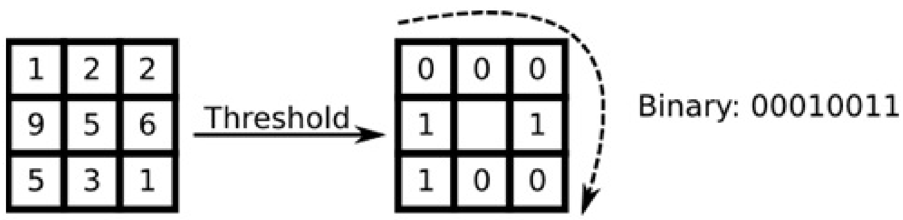

Fig. 5. Example of LBP code generation

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 6.

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 static hand gestures.

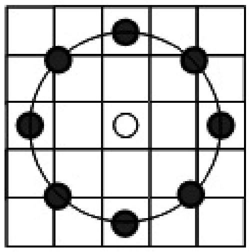

Fig. 6. LBP operator used in this system.

3.3. *Static Hand Gesture Model Training*

For training and testing of the system, Chinese numeral gesture dataset is used. The dataset will be described in Section 5. The dataset contains a total 2000 images of ten Chinese Numeral symbols from ten different people.
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 7 and Fig 8 respectively.

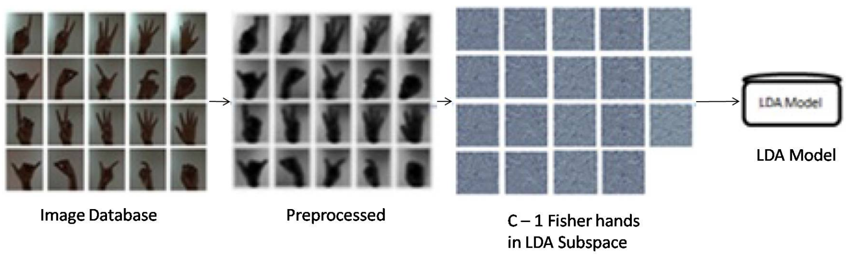

Fig. 7. Example of gesture model generation using LDA features.

3.4. *Static Hand Gesture Classification*

Nearest Neighbor algorithm (NN) is used for static hand gesture classification for both methods. The gestures 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.[15]

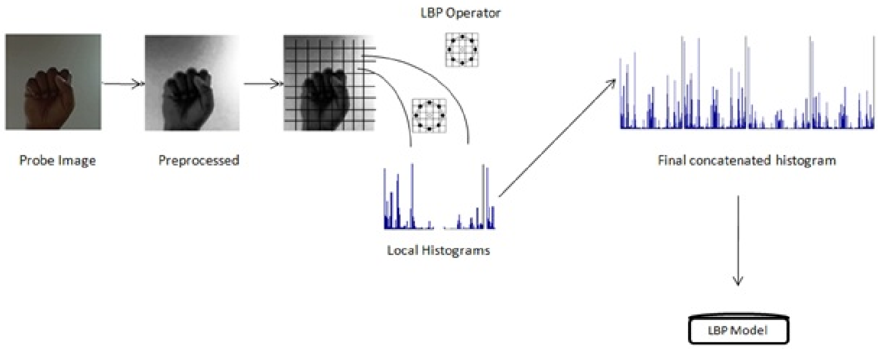

Fig. 8. Example of gesture model generation using LBP features.

4. Principal Directional Feature - Based Hand Gesture Recognition

The image sequences are captured using a fixed monocular camera. From the image stream, the hand areas are segmented using Haar-like feature based cascaded classifier, described in Section 2. The Principal Directional Features are extracted from the segmented images and classification is performed by Longest Common Subsequence for text based PDF.

4.1. *Principal Directional Feature Extraction*

After the hand areas, $H_i(width \times height)$, have been segmented from the image sequence, the centroids, $C_i(x, y)$ are calculated.

$$C_i(x, y) = \left(\frac{H_i(Width)}{2}, \frac{H_i(Height)}{2} \right)$$

From these centroids, the Directional Features are calculated. The Directional Features are calculated from the displacement measurement between subsequent centroids. The general displacement equation between two subsequent centroids, $C(x, y)$ and $\acute{C}(\acute{x}, \acute{y})$ is,

$$d = \sqrt{(C(x) - C'(x'))^2 + (C(y) - C'(y'))^2}$$

The linear displacements between two consecutive centroids, $C(x, y)$ and $\acute{C}(\acute{x}, \acute{y})$ in X and Y respectively are,

$$d(x) = (C(x) - C'(x')$$

$$d(y) = (C(y) - C'(y')$$

Let us consider Fig 9, where 15 images of a hand gesture is presented. From these 15 frames, 15 centroids are calculated. From the subsequent centroids, the linear displacements of corresponding x and y axis are calculated. Using the linear displacements, the Directional Features are extracted using the following encoding condition.

$$D_i = \begin{cases} N, & \text{if } |d_i(x)| < |d_i(y)| \text{ and } d_i(y) < 0 \\ E, & \text{if } |d_i(x)| > |d_i(y)| \text{ and } d_i(x) > 0 \\ S, & \text{if } |d_i(x)| < |d_i(y)| \text{ and } d_i(y) > 0 \\ W, & \text{if } |d_i(x)| > |d_i(y)| \text{ and } d_i(x) < 0 \end{cases}$$

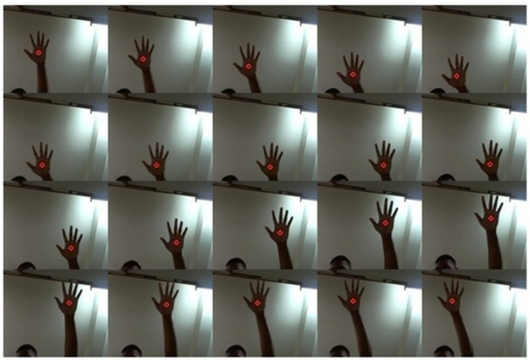

Fig. 9. A sequence of 20 images portraying a dynamic hand gesture.

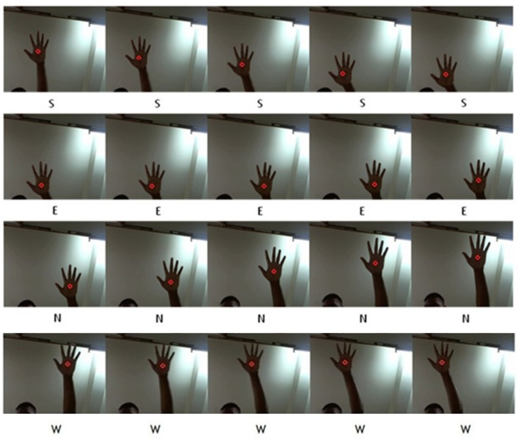

Fig. 10. Directional Feature encoding on the image sequence corresponding to Fig 9.

Table 1. Processing of Longest Common Subsequence

	E	S	N	E
S	0	1	1	1
E	1	1	1	2
S	1	2	2	2
W	1	2	2	2

From the displacement of two consecutive centroids, four direction is detected (N = North, W = West, S = South and E = East). Both linear and angular displacements provide similar encoding. By applying the coding scheme on the image sequence of Fig. 9; the Directional Feature, DF is generated as shown in Fig 10. The generated Directional Feature(DF) is $DF = SSSSSEEEEENNNNN$. The Directional Feature carries redundant information. All the directional coding in this gesture has a sequence of five continuing occurrences. This may not occur in a practical scenario. A threshold is applied on the feature sequence to reduce the Directional Features into Principal Directional Features (PDF). Any code (N, W, W or E) continuing at least five or more occurrences is considered to be a Principal Directional Feature (PDF).

Based on this consideration, the DF is reduced to Principal Directional Feature (PDF) $PDF = SEN$. This text based feature vectors are used in recognizing the dynamic hand gestures from image sequence.

4.2. *Principal Directional Feature Model Training*

The PDF is text based in nature. Hence, a text based model is trained with the help of the proposed directional dataset described in Section 5. The dataset is consist of complex directional gestures. After extracting features from the complex gesture, the PDF is stored in a text based feature model.

4.3. *Dynamic Gesture Classification based on Principal Directional Features*

In this system, to classify the dynamic hand gestures from the text based Principal Directional Features, a robust, efficient and accurate text matching algorithm, the Longest Common Subsequence (LCS) is used.

The Principal Directional Feature sequence is matched against the pre-stored text model to classify the dynamic gestures. Longest Common Subsequence (LCS) finds the longest sequence of characters present in two text strings where the characters may or may not reside in contiguous blocks. The length of the LCS from two Principal Directional Features, $K[1...i] = SESW$ and $L[1...j] = ESNE$ can be found by considering the following two cases.

Table 2. Dynamic Hand Gesture Classification Algorithm

```
Initialize MAX_LENGTH to 0
Initialize GESTURE to NULL
Load Model MN containing N gestures
Compute PDF from probe gesture
WHILE i < N
   LCS (PDF, MN)
   IF LCS_LENGTH > MAX_LENGTH
      MAX_LENGTH = LCS_LENGTH
      GESTURE = LCS_GESTURE
WHILE END
RETURN GESTURE
```

Case 1: If $K[i] \neq L[j]$, then one of $K[i]$ or$L[j]$ is discarded.

$$LCS[i,j] = MAX(LCS[i-1,j], LCS[i,j-1])$$

Case 2: If $K[i] = L[j]$, then $K[i]$ and $L[j]$ are matched.

$$LCS[i,j] = 1 + LCS[i-1,j-1]$$

From these cases, it is evident that filling up a matrix over all possible values of i and j is sufficient to find the sequence length. Let us consider Table 1; where K is along the leftmost column and L is along the topmost row. By filling out this matrix row by row, the length of the overall Longest Common Subsequence is found between K and L in the lower right corner. It takes a constant amount or $O(1)$ time to fill up each cell of the matrix. It takes $O(mn)$ time to complete the whole matrix, where m and n is the length of K and L. Here K and L may or may not be of equal size.

To find the sequence, backtracking is used from the lower right corner to all the way up. There are two move criteria. If the cell above and on the left has the exact same value as the current cell, move to that cell. If both of them have the same value, move to any one of the cells. The second criterion is, if these cells have strictly lesser values than the current cell, then move diagonally to the top-left cell and output the corresponding character of the cell that was just left behind. For K and L the length is 2 and the sequence is SE.

The output sequence will be in reverse order. A reverse string algorithm is applied to find the original Longest Common Subsequence string. The Principal Directional Features from a test gesture sequence is matched against all the features from the gesture model using Longest Common Subsequence for classification. The proposed algorithm for dynamic hand gesture classification is presented in Table 2.

5. 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 11 shows example of hand gesture recognition in different backgrounds. It shows the presence of static objects in the background do not reduce system performance.

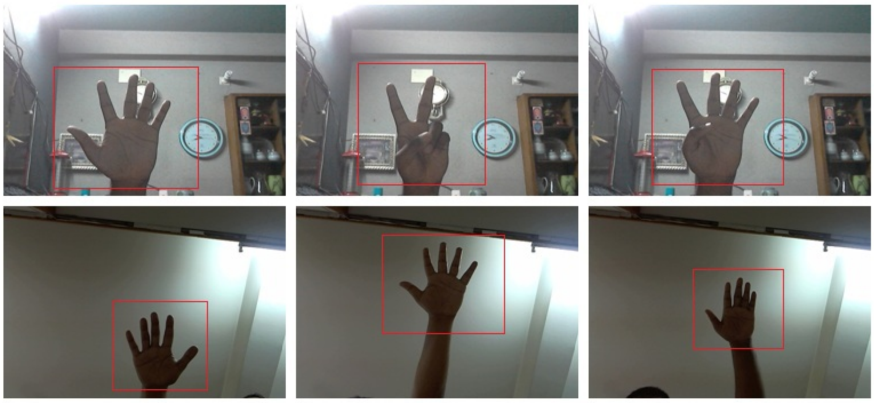

Fig. 11. Example of hand gesture recognition in different backgrounds.

5.1. *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 gestures is shown in Fig 12.

5.2. *Directional Gesture Dataset*

The directional gesture dataset is a collection of complex directional gestures. The dataset contains ten complex different directional. Each directional gesture is consisting of complex gestures. An example set of gestures from directional gesture database are presented in Fig 13.

5.3. *Performance of Hand Area Detection*

To test the Haar-like feature based cascaded classifier, 2000 test images are

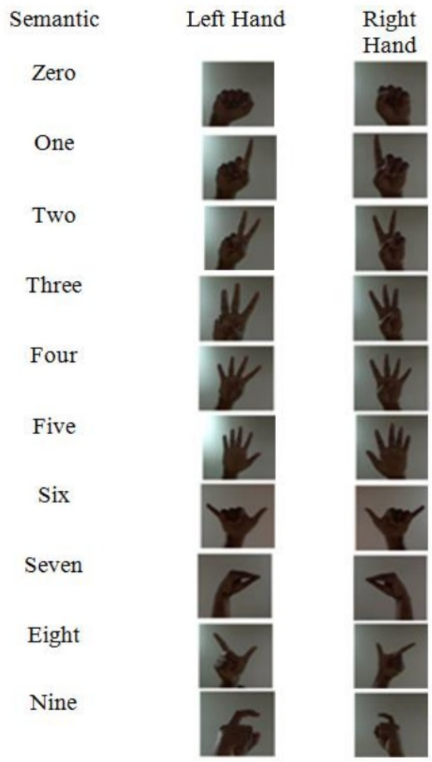

Fig. 12. An example set of Chinese numeral gesture dataset.

Table 3. Comparison Between Hand Detection Methods

Condition	Skin-color Cue (%)	Haar-like Features (%)
Offline controlled	89.40	98.30
Offline cluttered	82.90	95.60
Online controlled	82.20	96.70
Online cluttered	77.50	93.80

captured for the online and offline mode each. The images are taken from different backgrounds. The result comparison between the Haar-like feature based cascaded classifier and traditional skin color cue based method for hand detection is shown in Table 3, where the Haar-like feature based method surpasses the traditional method by far.

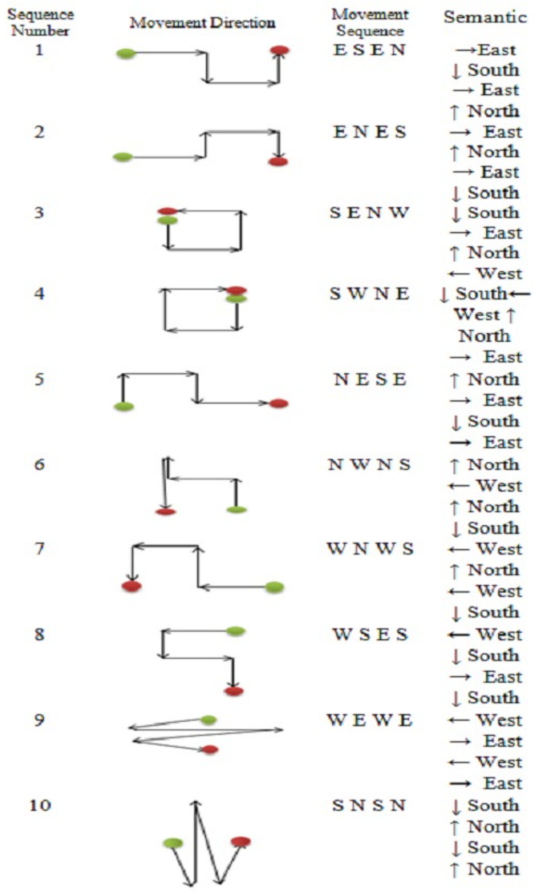

Fig. 13. An example set of gestures from directional gesture dataset.

5.4. *Performance of Static Hand Gesture Recognition System*

For static hand gesture recognition, 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. Static 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 gestures and N is the number of all test gestures.

Table 4. Accuracy comparison amongst static hand gesture recognition systems for gestures performed by left hand

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 5. Accuracy comparison amongst static hand gesture recognition systems for gestures performed by right hand

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 6. Accuracy comparison amongst static hand gesture recognition systems for gestures performed by both hands

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

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

Tables 4, 5, and 6 present the accuracy comparison amongst the PCA based system and the proposed LDA and LBP based systems respectively using N fold cross validation.

The mean accuracy of LDA based static hand gesture recognition on the Chinese numeral gesture dataset is 90.35% for gestures performed by left hand, 94.15% for gestures performed by right hand and 92.75% for gestures performed by both hands.

The mean accuracy of LBP feature based static hand gesture recognition on the Chinese numeral gesture dataset is 85.85% for the gestures performed by left hand, 89.25% for the gestures performed by right hand and 86.60% for gestures performed by both hands.

Table 7. Computation cost comparison amongst static hand gesture recognition systems for gestures performed by both hands

Methods	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

The efficiency is measured based on the time it takes for each method 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, preprocess, 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, LBP and PCA based static hand gesture recognition systems.

5.5. *Performance of Dynamic Hand Gesture Recognition System*

Ten sets of gestures, prepared from ten different persons ($10 \times 10 = 100$ gestures) are used in testing the novel Principal Directional Feature based dynamic hand gesture recognition system. The hand gesture recognition accuracy was measured on the basis of the following condition, where N_C is the number correctly classifier gestures and N is the number of all test gestures.

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

Table 8 presents the accuracy measure of each dynamic gesture from the Directional Gesture dataset. The mean accuracy is 94%.

In the testing system configuration, it takes an average of 2.537 milliseconds to capture a frame and another 2.332 milliseconds to detect the hand position for computing the directional feature from the image sequence. The average computation cost of processing a single dynamic gesture consisting of minimum 20 frames is 97.423 milliseconds.

6. Conclusion

The paper presents a novel method for computer vision-based static and dynamic hand gesture recognition. The mean accuracy of LDA based static hand gesture recognition on the Chinese numeral gesture dataset is 92.42%. The mean accuracy of LBP based static hand gesture recognition on the Chinese numeral gesture dataset is 87.23%. It takes 2.7004 microseconds to complete a single matching using LDA features and 45.105 microseconds to complete a single matching using LBP

Table 8. Dynamic hand gesture recognition accuracy

Sequence Number	Gesture Sequence (%)	Accuracy (%)
1	E S E N	90
2	E N E S	100
3	S E N W	100
4	S W N E	100
5	N E S E	90
6	N W N S	90
7	W N W S	90
8	W S E S	80
9	W E W E	100
10	S N S N	100
Mean		94%

features. The mean accuracy of the novel dynamic hand gesture recognition method on directional gesture dataset is 94%. The average computation cost of processing a single dynamic gesture consisting of minimum 20 frames is 97.423 milliseconds. Due to high performance and simplicity of the system, more complex hand gesture recognition is possible if the system is trained accordingly. The outcome of the system can be utilized for future human-computer or human-robot natural interaction.

7. References

References

1. C. Keskin, A. Erkan, L. Akarun, "Real Time Hand Tracking and 3D Gesture Recognition for Interactive Interfaces Using HMM", *In Proceedings of International Conference of Artificial Neural Networks* (2003).
2. C. Manresa, J. Varona, R. Mas, F. J. Perales, "Hand Tracking and Gesture Recognition for Human-Computer Interaction", *Electronic Letters on Computer Vision and Image Analysis* (1964), p. 96.
3. Q. Chen, N. D. Georganas, E. M. Petriu, "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features", *In Instrumentation and Measurement Technology Conference Proceeding*, (2007), p. 149.
4. C. Huang, S. Jeng, "A Model-based Hand Gesture Recognition System. Machine Vision and Applications", *Springer-Verlag 2001*, (2001), p. 243.
5. W. L. Lu, J. J. Little, "Simultaneous Tracking and Action Recognition Using the PCA-HOG Descriptor", *IEEE Computer Society*, (2006), p. 6.
6. K. G. Derpanis, R. P. Wildes, J. K. Tsotsos, "Hand Gesture Recognition Within a Linguistic-based Framework", *ECCV, Springer, 2004*, (2004), p. 282.
7. H. Zhou, D. J. Lin, T. S. Huang, "Static Hand Gesture Recognition Based on Local Orientation Histogram Feature Distribution Model", *CVPRW, 2004* (2004), p. 161.
8. R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady, "A Linguistic Feature Vector for the Visual Interpretation of Sign Language", *Springer-Verleg 2004*, (2004), p. 391-401.

9. P. Viola, M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", *IEEE CVRP, 2001* (2001), p. 511.
10. J. Friedman, T. Hastie, R Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting", *Annals of Statistics*, (1998), p. 2000.
11. G. Bradski, "The OpenCV Library", *Dr. Dobbs Journal of Software Tools*, (2000).
12. R. Lienhart, J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", *IEEE ICIP 2002* (2002), p. 900.
13. 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.
14. 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.
15. T. Ahonen, A. Hadid, "Face recognition with Local Binary Patterns", *ECCV 2004*, (2004), p. 469.

Mahmood Jasim received his B.Sc. and M.Sc. degree in Computer Science & Engineering from the Department of Computer Science & Engineering, University of Dhaka, Bangladesh in 2011 and 2013, respectively. His current research interests include Human-Computer Interaction, Image Processing, Computer Vision and Artificial Intelligence.

Dr. Tao Zhang received his B.S., M.S. and Ph.D.degrees in Electrical Engineering from Tsinghua University of China, in 1993, 1995 and 1999, respectively. In 2002, he received his second Ph.D. degree in Electrical Engineering in the Department of Advanced System Control Engineering, Graduate School of Science and Engineering, Saga University, Japan. He was a postdoctoral researcher in the National Institute of Informatics of Japan.
Since 2012, he has been serving as a Professor in the Department of Automation, Tsinghua University, Beijing, China. His current research interests include Nonlinear System Control Theory and Application, Intelligent Control of Robot, Micro-satellite Engineering, Fault Diagnosis and Reliability Analysis and Signal

Processing. He is a member of IEEE, IEIEC and IFAC.

Dr. Md. Hasanuzzaman received Ph.D.degree from the Department of Informatics, National Institute of Informatics (NII), The Graduate University for Advanced Studies, Tokyo, Japan. He graduated (with Honors) in 1993 from the Department of Applied Physics and Electronics, University of Dhaka, Bangladesh. He completed Masters of Science (M.Sc.) in Computer Science in 1994 from the University of Dhaka. He joined as a Lecturer int the Department of Computer Science & Engineering, University of Dhaka, Bangladesh in 2000. Since March 2013, he has been serving as a professor in the Department of Computer Science & Engineering, University of Dhaka, Bangladesh. His current research interests include Human-Computer Interaction, Image Processing, Computer Vision and Artificial Intelligence.