Proceeding of the IEEE
International Conference on Robotics and Biomimetics (ROBIO)
Shenzhen, China, December 2013

A Novel Hand Gesture Recognition Method using Principal Directional Features

Mahmood Jasim, Tao Zhang and Md. Hasanuzzaman

Abstract— **This paper presents a novel hand gesture recognition method based on Principal Directional Features (PDF). The image sequence is captured using a fixed mounted monocular camera to recognize dynamic gestures. Haar-like feature based cascaded classifier is used for hand area segmentation. Text based Principal Directional Features are extracted from the segmented images. Longest Common Subsequence algorithm is used to recognize the gestures from text based PDF. The Directional Gesture dataset is prepared containing complex dynamic gestures to test this system and achieved 94% accuracy in recognizing dynamic hand gestures.**

I. INTRODUCTION

Computer vision based automatic hand gesture recognition methods have been a prominent research topic for the last few decades. This paper introduces a new method of recognizing hand gestures using computer vision based approach. The challenges of hand gesture recognition include segmentation of hand area from the image sequence, capturing the motion associated with the gestures and the interpretation of the motion to recognize the gestures. The goal of this paper is to cope with all these challenges and present an efficient way of recognizing hand gestures from image sequences. Different approaches have been taken to solve these challenges. To segment the hand area from the images, color gloves or skin color cue based methods have been applied with limited success by Keskin [1] and Manresa [2]. A more recent approach is using Haar-like feature based cascaded classifiers by Chen [3] to detect the hand. To extract hand features for static pose recognition, Principal Component Analysis (PCA) has been extensively used by many researchers. Huang [4], and Lu [5], used PCA or modifications of PCA to recognize hand poses from the image stream. For dynamic hand gesture recognition, Finite State Machine (FSM) or Hidden Markov Model (HMM) based methods are popular choices for the researchers. HMM has many variations and were used by Huang [5] and Bowden [6] for dynamic hand gesture recognition. Other methods including Linguistic based Framework by Derpanis [7], Local Orientation Histogram Feature Description Model by Zhou [8], Motion Divergence Fields by Shen and Wu [9] etc. are notable. In this paper, Haar-like feature based cascaded classifier is used to segment the hand area, proposed text based Principal Directional Features (PDF) are used as directional features and Longest Common Subsequence is used to classify the hand gestures from the PDF.

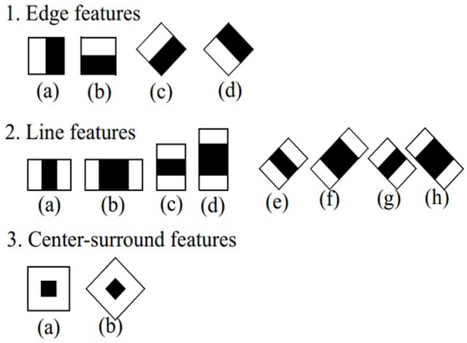


Fig. 1. Example sets of Haar-like features

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 III presents the proposed Principal Directional Feature based hand gesture recognition method. Section IV presents the proposed Directional Gesture dataset and experimental results. Finally, this paper is concluded in Section V.

II. HAND AREA SEGMENTATION USING HAAR-LIKE FEATURES

Haar-like feature based hand area segmentation is a statistical approach that concentrates on certain areas of an image, not individual pixel values. The concept of "Integral Image" is used to compute a set of Haar-like features introduced by Viola and Jones [10]. It achieves true scale invariance which eliminates the need for a multi-scale image pyramid for different scales of object image. This algorithm selects features based on AdaBoost [11] learning. The Viola Jones method performs 15 times faster without sacrificing the accuracy compared to other object detection methods. Fig.1 presents example sets of Haar-like features.

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

This work was not supported by any organization

M. Jasim is with the Department of Computer Science & Engineering, University of Dhaka, Dhaka - 1000, Bangladesh mahmood.jasim.0@gmail.com

T. Zhang is with the Department of Automation, School of Information Science & Technology, Tsinghua University, Beijing 100084, China. taozhang@mail.tsinghua.edu.cn

M. Hasanuzzaman is with the Department of Computer Science & Engineering, University of Dhaka, Dhaka - 1000, Bangladesh. hzamancsdu@yahoo.com

978-1-4799-2744-9/13/$31.00 ©2013 IEEE

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 the 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 the better accuracy.

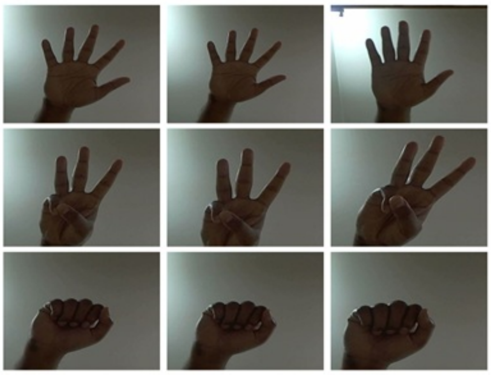


Fig. 2. Examples of positive images for Haar-like feature based cascaded classifier training

In this paper, we propose to train a common Haar-like feature based cascaded classifier to segment the hand position from the image.

Fig. 3. Examples of negative images for Haar-like feature based cascaded classifier training

In this system, a Haar-like feature based cascaded classifier is generated for the proposed method using 3000 positive images and 2000 negative images. Examples of positive and negative images are shown in Fig. 2 and Fig. 3 respectively. The object size is set to 20 × 20 and OpenCV [12] library is used to train the classifier.

III. 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 II. The Principal Directional Features are extracted from the segmented images and classification is performed by Longest Common Subsequence for text based PDF.

A. 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 using equation 1.

$$C_i(x, y) = \left(\frac{H_i(Width)}{2}, \frac{H_i(Height)}{2} \right) \tag{1}$$

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 presented in equation 2.

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

Equations 3 and 4 calculates the linear displacements between two consecutive centroids, $C(x, y)$ and $\acute{C}(\acute{x}, \acute{y})$ respective to x and y axis.

$$d(x) = (C(x) - C'(x') \tag{3}$$

$$d(y) = (C(y) - C'(y') \tag{4}$$

It is also possible to calculate the angular displacement between two subsequent centroids using the inverse tangent equation, shown in equation 5.

$$\theta = tan^{-1}(\frac{y}{x}) \times \frac{\pi}{180} \tag{5}$$

Let us consider Fig. 4, where 15 images of a hand gesture is presented. From these 20 frames, 20 centroids are calculated. From the subsequent centroids, the linear displacements of corresponding x and y axis are calculated using the equations 3 and 4. In case of angular displacement, equation 5 is used. Using the linear displacements, the Directional Features are extracted using the encoding condition as shown in equation 6.

$$D_i = \begin{cases} N, & \text{if } |d_i(x)| < |d_i(y)| \text{ and } d_i(y) < 0 \\ E, & \text{if } |d_i(x)| > |d_i(y)| \text{ and } d_i(x) > 0 \\ S, & \text{if } |d_i(x)| < |d_i(y)| \text{ and } d_i(y) > 0 \\ W, & \text{if } |d_i(x)| > |d_i(y)| \text{ and } d_i(x) < 0 \end{cases} \tag{6}$$

In case of the angular displacements, the Directional Features can be extracted using the encoding condition as shown in equation 7.

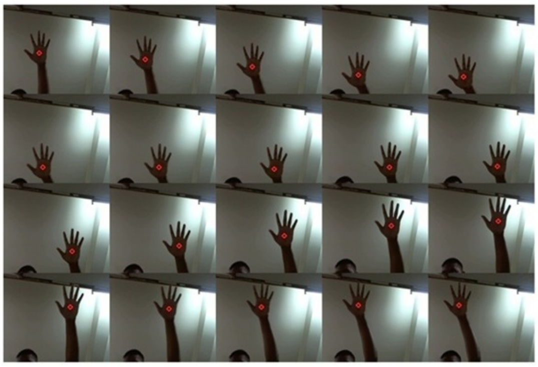


Fig. 4. A sequence of 20 images portraying a dynamic hand gesture

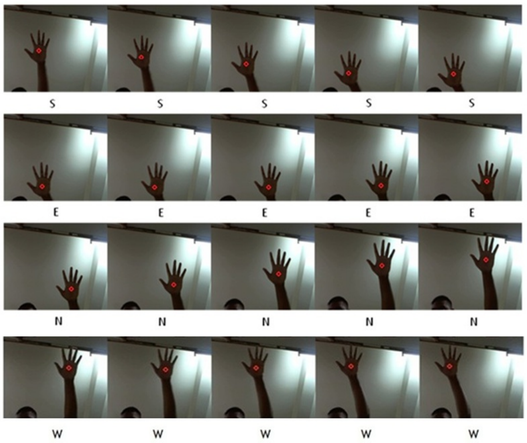


Fig. 5. Directional Feature encoding on the image sequence corresponding to Fig. 4.

$$D_i = \begin{cases} N, & \text{if } \theta \geq 45 \text{ and } \theta < 135 \\ W, & \text{if } \theta \geq 135 \text{ and } \theta < 225 \\ S, & \text{if } \theta \geq 225 \text{ and } \theta < 315 \\ E, & \text{if } \theta \geq 315 \text{ and } \theta < 45 \end{cases} \tag{7}$$

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 any coding scheme on the image sequence of Fig. 4; the Directional Feature, DF is generated as shown in Fig. 5. The generated Directional Feature is shown in Fig. 6.

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) and is stored as a feature.

DF = SSSSSEEEEENNNNNWWWWW

Fig. 6. The Directional Feature generated in Fig. 5.

PDF = SENW

Fig. 7. The Principal Directional Feature (PDF) generated from Directional Feature of Fig. 6.

Based on this consideration, the DF from Fig. 6 is reduced to Principal Directional Feature (PDF) of Fig. 7. This text based feature vectors are used in recognizing the dynamic hand gestures from image sequence.

B. 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 IV. The dataset is consist of complex directional gestures. As an example, the S E N W gesture sequence in the Directional Gesture dataset is a complex gesture of MOVE SOUTH and PROCEED, then TURN EAST and PROCEED, then TURN NORTH and PROCEED and finally TURN WEST and PROCEED. After extracting features from the complex gesture, the PDF is stored in a text based feature model.

C. 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 is used. The proposed algorithm for dynamic hand gesture classification is presented in Table I.

The Principal Directional Feature sequence is matched against the pre-stored text model to classify the dynamic gestures. Longest Common Subsequence 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 Longest Common Subsequence from two Principal Directional Features, $S[1...i] = SESW$ and $T[1...j] = ESNE$ can be found by considering two cases shown in equation 8 and equation 9.

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

$$LCS[i,j] = MAX(LCS[i-1,j], LCS[i,j-1]) \tag{8}$$

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

$$LCS[i,j] = 1 + LCS[i-1,j-1] \tag{9}$$

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 II; where S is along the leftmost column and T is along the topmost row. By filling out this matrix row by row, the length of the overall Longest Common Subsequence is found between S and T in the lower right corner. It takes a constant amount or $O(1)$ time

TABLE I

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
```

TABLE II

PROCESSING THE 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

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 S and T. Here S and T 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 S and T 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.

IV. EXPERIMENTS

The system is tested in a computer system with Intel Core - i5 2400 processor with four physical cores, 8 GB of RAM and 500 GB of secondary storage. The images are captured using Logitec 310 web camera. For system coding, Microsoft Visual Studios 2012 and OpenCV is used.

A. Directional Gesture Dataset

The Directional Gesture dataset is a collection of complex directional gestures. The dataset contains ten complex different directional gestures. Each directional gesture is consisting of complex gestures. The gestures are not allowed to have repeated patterns. For example, a MOVE SOUTH gesture is not followed by another MOVE SOUTH gesture. Table III presents the Directional Gesture Dataset.

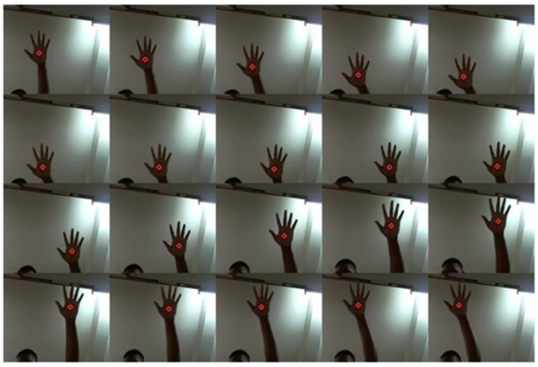


Fig. 8. An example of a directional gesture in 20 frames (S E N W) from the Directional Gesture dataset

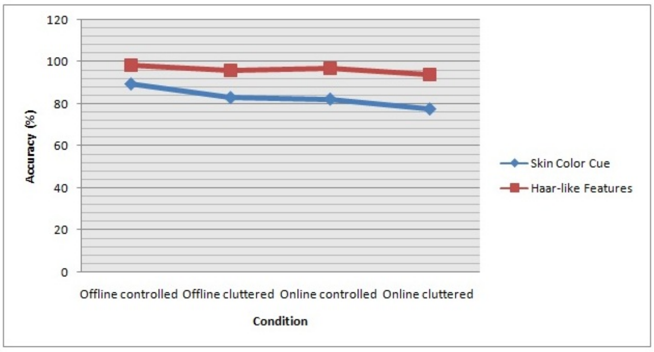


Fig. 9. Comparison between hand detection methods

Each gesture is consist of minimum 20 continuous image frames. Each of the ten different gestures is gathered from ten different persons. There are a total of 100 gestures $(10 \times 10 = 100)$ present in the dataset taken in various lighting, environment and from different people. An example of complex gesture is shown in Fig. 8.

B. Experimental Result of Hand Detection Methods for hand area segmentation

To test the Haar-like feature based cascaded classifier, 2000 test images were captured in the online mode and 2000 images were captured in the offline mode. The images were taken from controlled and cluttered backgrounds to test the classifier. The result comparison between the Haar-like feature based cascaded classifier and traditional skin color cue based method for hand detection is shown in Fig. 9 and Table IV, where the Haar-like feature based method outperforms the skin color cue based system by far.

C. Experimental Results of Dynamic Hand Gesture Recognition

Ten sets of gestures, prepared from ten different persons $(10 \times 10 = 100)$ are used in testing the proposed Principal Directional Feature based dynamic hand gesture recognition method. The hand gesture recognition accuracy is measured on the basis of equation 10.

$$Accuracy(\%) = \frac{N_C}{N} \tag{10}$$

TABLE III
DIRECTIONAL GESTURE DATASET

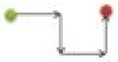
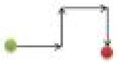
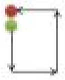
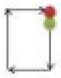
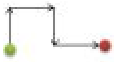
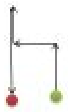
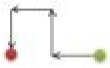
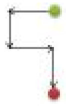
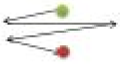
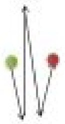
Gesture Number	Movement Directions	Movement Sequences	Semantic
1		E S E N	East South East North
2		E N E S	East North East South
3		S E N W	South East North West
4		S W N E	South West North East
5		N E S E	North East South East
6		N W N S	North West North South
7		W N W S	West North West South
8		W S E S	West South East South
9		W E W E	West East West East
10		S N S N	South North South North

TABLE IV
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

TABLE V
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%

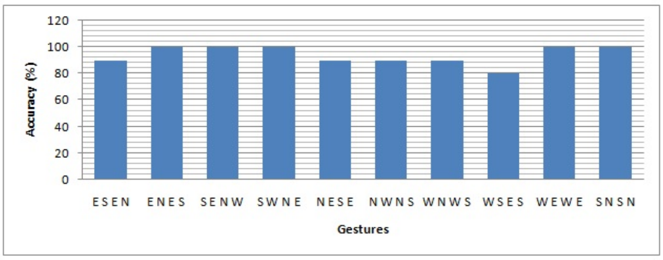


Fig. 10. Dynamic hand gesture recognition accuracy

Here, N is the number of all test gestures and NC is the number of correctly classified gestures. Fig. 10 and Table V presents the classification accuracy measure of each dynamic gesture from the Directional Gesture dataset. The mean accuracy is 94%.

V. CONCLUSION

This paper proposed a novel dynamic hand gesture recognition method based on Principal Directional Features, (PDF). The system utilizes the temporal positional displacement of the hand centroids from a sequence of images to extract most prominent direction of motion and encodes them in to PDF. Hand positions from the images are detected using Haar-like feature based cascaded classifier. To test the performance of the method, the Directional Gesture dataset consisting of complex gestures is prepared. Longest Common Subsequence is used for text based Principal Directional Feature based gesture classification. The system shows 94% mean accuracy on the Directional Gesture dataset of 100 complex gestures.

A drawback of this system is the dependency on the hand detection accuracy. In future, a larger hand image dataset can be used to train a robust classifier for hand detection. An experimental deployment with application extension towards robot movement control by gestures can also be considered.

REFERENCES

[1] C. Keskin, E. 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", 2005.

[3] Q. Chen, N. D. Georganas, E. M. Petriu, “Real-time vision-based hand gesture recognition using Haar-like features”, *IEEE In Instrumentation and Measurement Technology Conference Proceedings, 2007*.

[4] C. Huang, S. Jeng, “A model-based hand gesture recognition system”, Machine Visions and Applications, Springer-Verlag 2001, ch 12.pp. 243-258.

[5] W. L. Lu, J. J. Little, “Simultaneous tracking and action recognition using the pca-hog descriptor”, In proceedings on *3rd Canadian Conference on Computer and Robot Vision, 2006*.

[6] R. Bowden, D. T. K. Windridge, A. Zisserman, M. Brady, “A linguistic feature vector for the visual interpretation of sign language”, In *European Conference on Computer Vision*, Vol. 1, Springer-Verlag, 2004, pp. 391-401.

[7] K. G. Derpanis, R. P. Wildes, J. K. Tsotsos, “Hand gesture recognition withing a linguistic-based framework”, In proceedings *ECCV*, Springer, 2004, pp. 282-296.

[8] H. Zhou, D. J. Lin, T. S. Huang, “Static hand gesture recognition based on local orientation histogram feature distribution model”, In proceedings of *The Conference on Computer Vision and Pattern Recognition Workshop, 2004, CVPRW ’04*, Vol. 10.P. 161.

[9] X Shen, G. Hua, Y. Wu, L. Williams, “Motion divergence fields for dynamic hand gesture recognition”, In *IEEE International Conference on Automatic Face and Gesture Recognition Workshop, 2011*, FG 2011, pp. 492-499.

[10] P. Viola, M. Jones, “Rapid object detection using a boosted cascade of simple features”, *IEEE CVPR ’01*, 2001, pp. 511-518.

[11] J. Friedman, T. Hastie, R Tibshirani, “Additive logistic regression: a statistical view of boosting”, Annals of Statistics, 1998, Vol, 28, p. 2000.

[12] G. Bradski, “The OpenCV library”, Dr. Dobbs Journal of Software Tools, 2000.