Face Gestures Detection using Improved Viola-Jones Algorithm

Ahmed I. Taloba
Department of Information System
Assiut University
Assiut, Egypt

Yasser A. Dawood
Department of Computer Science
Assiut University
Assiut, Egypt

ABSTRACT

Many studies indicate that words and sentences represent only 7% of our ways of communicating with our world, while the bulk goes to our movements and gestures that are translated into the minds of the recipients unconsciously of them. Face detection is an important topic in computer vision. Viola-Jones is a powerful approach of identifying objects, especially facial recognition. Viola-Jones is subjected to a number of parameters that affect its performance. The measuring factor has the most impact on facial recognition and provides a balance between speed and accuracy in face detection. In this paper, the improved Algorithm of Viola Jones has been used by modifying the measurement factor using a genetic algorithm to determine face gestures.

In this paper, the KNN classifier, MFCC (Mel Frequency Cepstral Coefficient), and the Extract feature are used in evaluating the detection of face gestures. First, we take the picture as input and use signal processing techniques to convert it to time domain references. For detecting the front face gestures accurately and efficiently, MFCC will be used with LBP (Local Binary Pattern) and DOG (Difference of Gaussians) to extract the feature. Based on the result of feature extracting, the KNN classifier classifies the activity of our input signal. Experimental results show that the proposed approach is superior to other ways of recognition.

Keywords
Gestures, Face Detection, Face Recognition, Viola-Jones, Scale Factor, Genetic Algorithm.

1. INTRODUCTION
The detection of face gestures is part of the feelings and is considered one of the important topics in these days. The traditional response in the various facial images under conditions is not controlled such elements as the security situation and lighting and different changes to the person in recognizing the face and different frequencies of sound in order to identify emotions for any database of face detection system and its indication of gesture[1][2].

The frequency, test, and results of the sound mileage are calculated by creating the database for the face and storing these features in the database. This database is then, used to evaluate facial gestures using different algorithms [3].

The expression of gesture and the recognition of a person's affective state are abilities indispensable for natural human interaction and social integration. The study of gesture has attracted interest of researchers from very diverse areas ranging from psychology to the applied sciences. Face detection and emotion features are currently very active area of research in the computer vision field. Different kinds of face detection application are currently used such as image database management system, monitoring and surveillance analysis, biomedical image that help the interfaces and drivers alertness systems [4][5].

Recently, many commercial systems for face recognition have become one of the most active applications of pattern recognition, image analysis and understanding. Non-verbal gestures give us more vital details about the person as the facial recognition plays a vital rule in human computer interaction. The verification system identifies a person's identity by comparing the captured image templates stored in the system and recognizes a person by checking the entire template database for a match [6].

Viola Jones is one of the most important algorithms in face recognition subject at the moment. There are a number of parameters that play a great role in the process of face recognition using Viola Jones. One of these parameters is the Scale Factor. In [7] the optimal value of the scale coefficient has been selected using a genetic algorithm where Genetic Algorithms are one of the most important algorithms currently used in the processes of improvement.

In this paper, the improved algorithm of Viola Jones has been used in face gestures detection. To detect foreshorten face gestures accurately and efficiently, we will use MFCC along with LBP and DOG for feature extraction. Based on the results of feature extraction, the KNN classifier best classifies the activity of our input signal.

The rest of this paper is organized as follows: the improved algorithms Viola-Jones is discussed in section 2. Local Binary Pattern (LBP) is discussed in section 3. Difference of Gaussians (DOG) is discussed in section 4. Mel frequency cepstral (MFCC) is discussed in section 5. K-nearest neighbors classifier (KNN) is discussed in section 6. Methodology is discussed in section 7. Experimental result is presented in section 8 while in section 9, we conclude our work.

2. IMPROVED VIOLA-JONES ALGORITHMS
In this section, we discuss our method of improving the performance Viola-Jones face detection algorithm. Viola-Jones algorithm has been widely used in face detection [8] due to it rescales the detection instead of the input image and run the detector many times through the image each time. This method saves processing time because it requires the same number of calculate whatever the used size [9]. Viola-Jones algorithm enables a fast and accurate detection of the integral image for feature computation, AdaBoost for selection and cascade classifiers for efficient computation resource allocation [10].Viola-Jones algorithm is governed by a number of parameters that affect its face detection performance: suppression, search mode, scaling factor and scaling mode. Among these parameters, the scaling factor has

the most impact on the process of face detection because it introduces a balance between speed and accuracy. However, the value of scaling factor is selected to be fixed which may not be global in different datasets. To improve this drawback, we propose the use of genetic algorithm of Viola-Jones algorithm to obtain dynamic scaling factor and have global solution. In the proposed genetic algorithm, we use crossover operation to determine the coefficient of the scale by the fitness value. The process is described using flowchart in Fig.1.
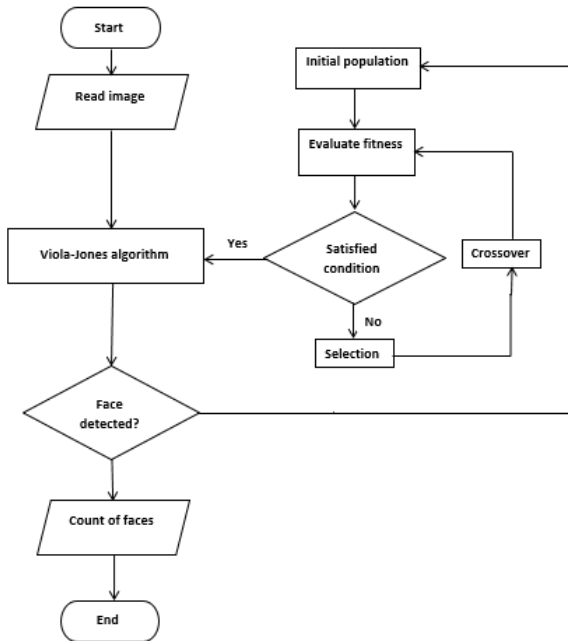

Fig. 1. The improved Viola-Jones algorithm based on the genetic algorithm.

The image is feed to the Viola-Jones algorithm and at the same time Viola-Jones demean for the value of fitness value for the scaling factor. This value is used to identify the face in the image in all Viola-Jones algorithm stages. If the process is succeeds, the algorithm will stop. Otherwise, a new value for scaling factor is generated by the genetic algorithm and the process is repeated until the suitable value is obtained.

3. LOCAL BINARY PATTERN (LBP)

In various applications, Local Binary Pattern (LBP) features have performed many tasks such as texture classification, segmentation, image retrieval and surface inspection [11-13].

LBP specifies the 3 x 3 region per pixel with the center pixel value and looks at the result as a binary number. The original LBP driver names the pixels of the image to look at facial shape information then, divided the facial images into small, non-overlapping R R0, R1, ... RM

Each face image is regarded as a composition of micro-patterns which is effectively detected by the LBP operator. The LBP histograms extracted from each sub-region are then, connected into a single, spatially enhanced feature histogram which can be defined as:

$$H_{i,j} = \sum_{x,y} I(f_1(x,y) = i) \ I \ ((x,y)\epsilon R_j) \qquad (1)$$

Where i = 0, ... , L-1, j = 0, ..., M-1. The extracted feature histogram describes the local texture and global shape of face images.

4. DIFFERENCE OF GAUSSIANS (DOG)

DOG is an algorithm which improves the features that subtracts an unclear version of an image from less visible one. Indistinct images are resulted after simplifying images to gray scale and combining the original images with Gaussian kernels. High -frequency spatial information is prevented by the blurring of the image using a Gaussian nucleus [14-16].

The subtraction of one image other than others preserves the spatial information between the frequency ranges stored in unclear images. Accordingly, the Gaussians are a filter which excludes the spatial frequencies and maintains them.

The difference from Gaussians is used as an optimization feature to increase the visibility of the edges and other details in a digital image. The Gaussians distinguish the high-frequency details that include random noise making it suitable for high-noise image processing.

The main problem in application our algorithm is the inherent reduction in the total picture contrast resulting from the process. There is a variety of alternative edge sharpening filters that work by improving the details of high frequencies.

The random noise with high spatial frequency can be undesirable fact because many of these sharpening filters tend to enhance noise.

5. MEL FREQUENCY CEPSTRAL (MFCC)

MFCC's are very useful features in audio processing. In the presence of noise, MFCC values not robust. Accordingly, there has been an increased effort to find new features that are more noise robust than MFCCs. Such as features as spectro-temporal modulation features are more robust to noise, but they are computationally expensive. Some researchers propose modification to the basic MFCC algorithms to improve robustness.Skowronski and Harris studied the effects of wider filter bandwidth on noise robustness. To modify MFCC, they used the known relationship between center frequency and critical bandwidth. The features of MFCC approximate the frequency decomposition along the basilar membrane by a short-time Fourier Transform. DCT (discrete cosine transform) is used to de-correlate the features. Compression is expressed as a log function and the auditory critical bands are modeled using triangular filters [17].

6. K-NEAREST NEIGHBORS CLASSIFIER (KNN)

KNN (k Nearest Neighbors) classifier is one of the most commonly used methods. It has been applied on a variety of cases. Among several methods for pattern recognition, KNN classifier works as follows. First, for each one of the training set elements a classification of it is performed based on various neighborhoods. The k value that maximizes the DC of each classification is found thus, for each training set there corresponds a particular k value which is considered the best available. Then, for each unknown element, the nearest neighbor is found and its k value is assumed (based on the "optimum" k array). After that, the KNN classifier is applied on that test element using that k value [18].

The KNN algorithm is given below

1. Choose k number of samples from the training set to generate initial population (p1).

2. Calculate the distance between training samples in each chromosome and testing samples as fitness value.

3. Choose the chromosome with the highest fitness value and store it as global maximum (Gmax).For i = 1 to L do

i. Perform reproduction

ii. Apply the crossover operator.

iii. Perform mutation and get the new population.

iv. iv. Calculate the local maximum (Lmax).

v. If Gmax < Lmax then a. Gmax = Lmax;

Output: the chromosome which obtains Gmax has the optimum K-neighbors and the corresponding labels are the classification results.

7. METHODOLOGY

The main aim of this work is to implement an efficient method to detect the face and some non-verbal gestures of a person. This work is divided into two parts. The first part focuses on storing the features of the face and the features of the voice of humans. The second part deals with an evaluation of the face and some non-verbal gestures of a person using the features database. We use enhancement of Viola Jones face detection algorithm to create the database for the face detection. The numbers of input images are collected for the creation of database then, the database for the features of voice with the different non-verbal gestures of the person is created. Fig 2 shows the overall implementation of the proposed system.
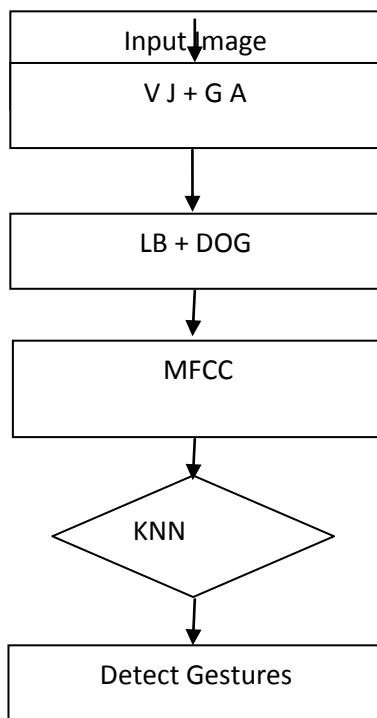

Fig 2. Use of the improved Viola-Jones algorithm for detecting gestures

8. EXPERIMENTAL RESUULT

The following figure shows the overall experimental results of the proposed system. We considered Fig. 3 (a) as an input image and applied signal processing techniques to convert it into time domain signals. Then, conversion signals are decomposed and considered one decomposition image as shown in Fig. 3 (b). To foreshorten face gestures accurately and efficiently, we will use the MFCC along with LBP (Local Binary Pattern) and DOG for feature extraction, from this we will get normalized signals as shown in Fig. 3 (c). On the basis of the results of feature extraction, the KNN classifier classifies the activity of our input signal. We obtain output according to the normalized signals as shown in Fig. 3 (d).
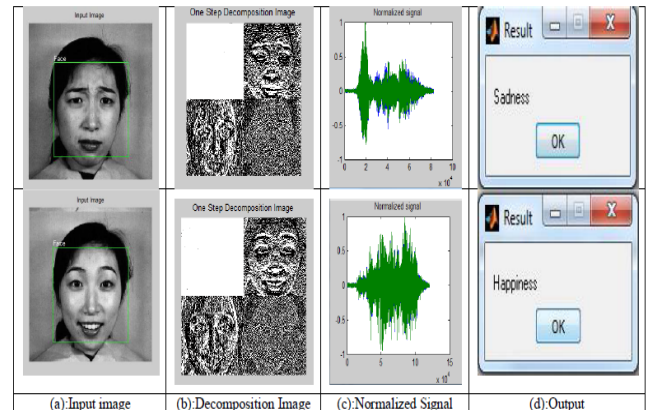

Fig 3: Results of the proposed Work

9. CONCULSION

The improved algorithm of Viola Jones has been used by modifying the measurement factor using a genetic algorithm to determine face gestures. To evaluate the detection of face gestures, the KNN classifier, the MFCC (Mel Frequency Cepstral Coefficient), Extract feature are used. First, we take the picture as input and use signal processing techniques to convert it to time domain references. To detect the front face gestures accurately and efficiently, MFCC has been used with LBP (Local Binary Pattern) and DOG (Difference of Gaussians) to extract the feature. Based on the results of feature extraction, the KNN classifier best classifies the activity of our input signal. Experimental results show that the proposed approach is superior to other ways of face recognition.

10. REFERENCES

[1] Sushmita Mitra and Tinku Acharya, " Gesture Recognition: A Survey" IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) Volume: 37 , Issue: 3 , May 2007).

[2] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas S. Hunag, " Human Computing and Machine Understanding of Human Behavior: A Survey", 2007.

[3] Anshun Raghuwanshi and Preeti D Swami, "An automated classroom attendance system using video based face recognition" 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)

[4] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico and Marc Schr¨oder, "Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing" ,IEEE Vinciarelli, Alessandro, et al.

"Bridging the gap between social animal and unsocial machine: A survey of social signal processing." IEEE Transactions on Affective Computing, 2012.

[5] Albert Mehrabian, " Nonverbal Communication ", 2017.

[6] Meghana B M, K Mahantesh, Dept. Electronics and Communication Engineering, S.J.B. Institute of Technology, Bangalore, India, "Classification of Human Emotions for Face and Voice Using KNN Classifier", May(2016).

[7] Ahmed I. Taloba, Adel A. Sewisy, and Yasser A. Dawood, " Accuracy Enhancement Scaling Factor of Viola-Jones Using Genetic Algorithms", ICENCO 2018.

[8] A. S. Dhavalikar and R. Kulkarni, "Face detection and facial expression recognition system," in Electronics and Communication Systems (ICECS), 2014 International Conference on, 2014, pp. 1-7: IEEE.

[9] S. Tikoo and N. Malik, "Detection of Face using Viola Jones and Recognition using Back Propagation Neural Network," arXiv preprint arXiv:1701.08257, 2017.

[10] M. El Rhazi, A. Zarghili, and A. Majda, "Comparative Study of Harris and Active Contour using Viola-Jones Algorithm for Facial Landmarks Detection," Transactions on Machine Learning and Artificial Intelligence, vol. 5, no. 4, 2017.

[11] Zhenhua Guo, Lei Zhang, and David Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification", Submitted to IEEE Transactions on Image Processing, 2010.

[12] Baochang Zhang, Yongsheng GAO, Sanqiang Zhao, and Jianzhuang Liu, " Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 2, FEBRUARY 2010.

[13] Bo Yang and Songcan Chen, " A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image ", Elsevier 2013.

[14] D. Giordano, R. Leonardi, F. Maiorana, G. Scarciofalo, and C. Spampinato, " Epiphysis and Metaphysis Extraction and Classification by Adaptive Thresholding and DoG Filtering for Automated Skeletal Bone Age Analysis", 2007.

[15] Juan F. Ramirez-Villegas, Eric Lam-Espinosa, and David F. Ramirez-Moreno, " Microcalcification Detection in Mammograms Using Difference of Gaussians Filters and a Hybrid Feedforward-Kohonen Neural Network", 2009.

[16] Eric C. Larson and Damon M. Chandler, " Most apparent distortion: full-reference image quality assessment and the role of strategy", Journal of Electronic Imaging 19(1), 011006 (Jan–Mar 2010).

[17] Cristina Elena Munoz Mulas, "Speech Signal Feature Extraction Model for Speaker's Gender and Age Identification System", 2014.

[18] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang and Yihong Gong, "Locality-constrained Linear Coding for image classification", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.