Facial Expression Recognition- Review

Umesh Balkrishna Chavan,
Department of Information Technology,
Walchand College Of Engineering, Sangli,Maharashta,,India

Dr.Dinesh B Kulkarni,
Department of Iinformation Technology,
Walchand College Of Engineering,Sangli,Maharashta,India

Abstract- expression recognition (happy, sad, disgust, surprise, angry, fear expressions) is application of advanced object detection, pattern recognition and classification task. Facial expression recognition techniques detecting emotion of people' using their facial expressions. This has found applications in technical fields such as Human-computer-Interaction (HCI) and security monitoring. It generally requires fast processing and decision making. Therefore, it is imperative to develop innovative recognition methods that can detect facial expressions effectively and efficiently. Although humans recognize facial expressions virtually without efforts or delay, reliable expression recognition by machine remains a challenge as of today. To automate recognition of emotional state, machine must be taught to understand facial gestures. This paper focuses on a review of different techniques for face recognition, face detection and emotion recognition are presented.
Keywords – HCI , GPU, CUDA,PCA,ICI,ANN,FFT,MLP,RBF.

I. INTRODUCTION

The facial expression recognition is to detect human emotion based on expression. Facial expression recognition follows the research framework of pattern recognition. This is composed of three steps: detection of face, feature (facial) extraction and expression classification. The amount of research carried out in each of these categories is quite sizable and noteworthy. These three categories are concerned with the central background pertaining to the issue of facial emotion recognition. Apart from them, another core area is development of appropriate facial database for such studies.

The rest of the paper is organized as follows. Related work (literature survey) is presented in section II. Methodology is explained in section III. Experimental results are presented in section IV. Concluding remarks are given in section V.

II RELATED WORK

TABLE 1: FACE DETECTION ALGORITHMS

Method	Note	Comment
Exhaustive search	Evaluate all possible subsets of features	Optimal, but too complex,
Branch and bound	Use branch and Bound Algorithm	Can be optimal. Complexity $O(2^n)$
Viola-Jones Method.	The first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola	it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection

TABLE II FACE RECOGNITION METHODS

Method	Note
DCT-mod 2 by Sanderson and Paliwal.	Applies DCT to sub regions of facial image to extract a number of DCT coefficients.
Support Vector[14] Machines	Non-linear data-processing tool, maps training samples to another high dimensional space

TABLE III FACIAL EXPRESSION RECOGNITION METHODS

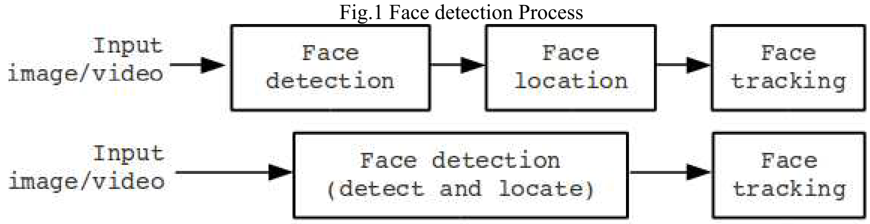
Methodology	Recognition Rate (%)
Context Method	79.20
Gabor Method	74.10
40 Gabor method	73.80
LGBP	77.67
RGB tensor	80.11

III METHODOLOGY

A. Face detection Algorithm

Face detection is a computer technique that determines the locations and sizes of human faces in arbitrary (digital) images. It detects facial features and ignores anything else, such as buildings, trees and other bodies. Some systems detect and locate faces at the same time, others first perform a detection routine and then, try to locate the face. Some tracking algorithm is needed

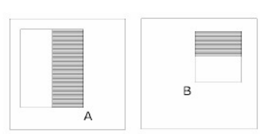
Fig.1 Face detection Process

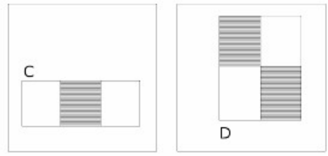


Early face detection algorithms focused on the detection of frontal human faces, whereas newer algorithms attempt to solve the more general and difficult problem of multi-view face detection. These algorithms take into account variations in the image or video by factors such as face appearance, lighting, and pose. Face detection is a two class problem where one has to decide if there is a face or not in a picture. This approach can be seen as a simplified face recognition problem. Face recognition has to classify a given face, and there could be many classes as candidates. Consequently, many face detection methods are very similar to face recognition algorithms. Or put another way, techniques used in face detection are often used in face recognition. In biometric, face detection is used, often as a part of (or together with) a facial recognition system. It is also used in video surveillance, human computer interface and image database management. For autofocus purpose in some recent digital cameras use face detection.

1) Viola-Jones Algorithm: The features employed by the detection framework universally involve the sums of image pixels within rectangular areas. This is based on Haar basis functions. The figure 2 illustrates the four different types of features used. The value of any given feature is the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles. Rectangular features of this sort are rather primitive as compared to alternatives such as steerable filters. These are sensitive to vertical and horizontal features, their feedback is considerably coarser. With the use of an image representation (integral image), rectangular features can be evaluated in constant time, which gives them a considerable speed advantage over their more sophisticated relatives. As each rectangular area in a feature is always adjacent to at least one other rectangle, it follows that any two-rectangle feature can be computed in six array references, any three-rectangle feature in eight, and any four rectangle feature in just nine.

Figure 2: feature type used by Viola-**Jones** method.





2) Exhaustive Search Algorithm: Human face detection is a key technology in face information processing, the speed of which is very important during real-time face detection for input images or input video sequences. This algorithm presents a novel face window searching algorithm based on evolutionary agent when detecting faces in gray-scale images. It can quickly AND the candidate face windows through the evolutionary computation of distributed agents, which represent different kind of windows. Experimental results prove that the evolutionary agent-based searching algorithm can increase the detection speed by 5-7 times compared with the traditional exhaustive searching method used in some general algorithms. The traditional face window searching starts from left-top corner of the image to classify all possible sub windows marks the windows which meet classification characteristics. The exhaustive searching method can't identify the face like windows quickly because generally there are so many non-face regions in the input image. It is not only computationally expensive but also has high false detection rate. As we all know, human can find the face regions quickly when browsing an image. If there are several faces in the image, people will focus on them simultaneously rather than search the face windows from the left-top corner to the right-bottom corner of the image, and it is very easy for human to know how many faces exist in the image by counting them.

3) Branch and Bound algorithm: Recently, researchers have proposed many face recognition methods with the aim of improving the accuracy rate of face recognition. However, few face recognition methods only focus on computational cost. For reduction in the computational cost of face recognition, an effective face recognition method using Haar wavelet features and a branch and bound method is used. In this method extracts features of the Haar wavelet from a normalized face image, and recognizes the face by classifiers learned with the Ada-Boost M1 algorithm. To accelerate the recognition process we select features according to the accuracy of classification and apply a branch and bound method to the recognition tree into which the classifiers of an individual in the face database are merged.

B. Feature Extraction/Reduction Methods

The feature extraction phase represents a key component of any pattern recognition system. Feature extraction involves detecting and isolating various desired features of patterns. It is the operation of extracting features for identifying interpreting meaningful information from the data. In facial expression recognition this is an essential pre-processing step as in pattern recognition. Following methods are used in feature extraction.

1) Fisher's Linear Discriminate: (FLD): This method projects high dimensional data onto a line and performs classification in this space. If there are two classes, the projection maximizes the distance between the means and minimizes the variance within each class. It can reduce the number of variables in the input by projecting data onto a possibly uncorrelated and low dimensional space. It reduces the number of features in the input to a manageable level. Some variables with information which are not related to facial expression can be excluded during the projection. This helps neural network to prevent from learning unwanted details in input. This improves the network classifiers performance and generalization. It is best for classification task which improves the performance and feature reduction technique.FLD considerers maximizing the following objective.

$J(w) = W^T S_B W / W^T S_W W$ Where S_B is the "distance between class scatter matrix" and S_W is the "within class scatter matrix". Then we formulate two matrixes which contain specific information of the data.

2) Principal Component Analysis (PCA): PCA is a linear transformation technique that transforms the data to a new coordinate system such that the greatest variance by projection of the data comes to lie on the first coordinate(called the principal component), the second greatest variance on the second coordinate and so on[14].PCA can be used for dimensionality reduction in a data set while retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring high-order ones. Such low-order components often contain the most important aspects of the data.

The PCA data reduction process follows for a data matrix A_i (i=1,...,6) of size M x K where each column of A represented an image feature vector of length M (Γk), K is the number of image representing each expression (here K=6) and i is the facial expression index, zero mean arrays $\hat{A}_i$ were calculated by subtracting the mean values for each row:

Γ^i_k corresponding column feature vector of image i of expression k.

$m_i = \sum \Gamma^i_k$ mean of training data.

$\square^{j}_{k} = \Gamma^{i}_{k} - m_i$, i=1,....,6, k=1,...,K which Normalized vector

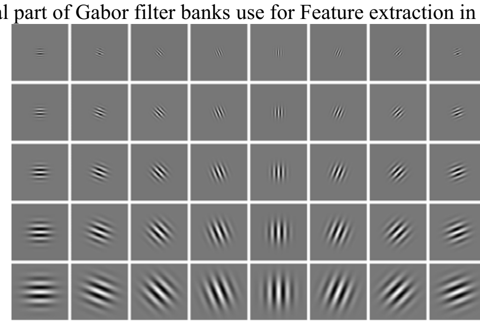
$\hat{A}^{k}_{i} = [\square^{j}_{k}\ \square^{j}_{k\ldots}\ \square^{j}_{k}]$ i=1,...,6

The feature arrays Ai were used to calculate the covariance arrays $L_i = \hat{A}^{T}_{I}\hat{A}_{i..}$

The eigenvectors arrays V_i of the L_i were determined and reduced in size from K x K to K x L (L<K) by removal vectors with small eigenvalues. The reduced-size eigenvectors $V^{reduced}_{i}$ formed the arrays U_i called the eigenfaces, and defines as: $V^{reduced}_{i} = U^{T}_{i}\hat{A}_i$

3) Gabor Filter: Field proposed Log-Gabor filters[9]. The response of the Log-Gabor filter is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high-frequency areas with desirable high pass characteristics. In this [10] a bank of 40 Log-Gabor filters is employed to extract the facial features.

Figure 3: Real part of Gabor filter banks use for Feature extraction in face recognition



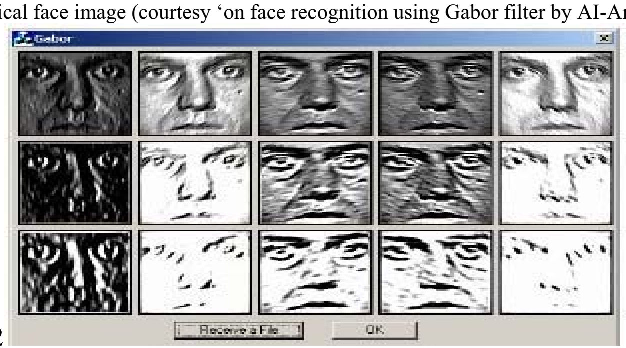
The polar form of 2-D Log-Gabor filters in frequency domain is given by

$$H(f,\theta) = \exp\left\{\frac{-[\ln(\frac{f}{f_o})]^2}{2[\ln(\frac{\sigma_f}{f_o})]^2}\right\}\exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\} ,B=2\sqrt{\frac{2}{\ln 2}} \times \ln\left(\frac{\sigma_f}{f_o}\right)$$

Where H (f,) is the frequency response function of the 2-D Log-Gabor filters, f and denote the frequency and the phase/angle of the filter, respectively, f_o is the filter's centre frequency and $_o$ the filter's direction. The constant σ_f defines the radial bandwidth B in octaves and the constant $\sigma_\square$ defines the angular bandwidth in radians .The ratio

is kept constant for varying fo, B is set to one octaves and the angular bandwidth is set to **Ω**= radians. This left only σ_f to be determined for a varying value of f_o. Six scales and four orientations are implemented to extract features from face images. This leads to 24 filter transfer functions representing different scales and orientations. The image filtering is performed in the frequency domain making the process faster compared with the special domain convolution. After the 2-D fast Fourier transform (FFT) into the frequency domain, the image arrays, X, are changed into the spectral vectors, X and multiplied by the 4 Log-Gabor transfer functions {H1, H2, …, H24}, producing 24spectral representations for each image [9]. The spectra are then transformed back to the spatial domain via the 2-D inverse FFT.

Fig. 4: Gabor filter response of a typical face image (courtesy 'on face recognition using Gabor filter by AI-Amin Bhutiyan and Chang Hung Liu)



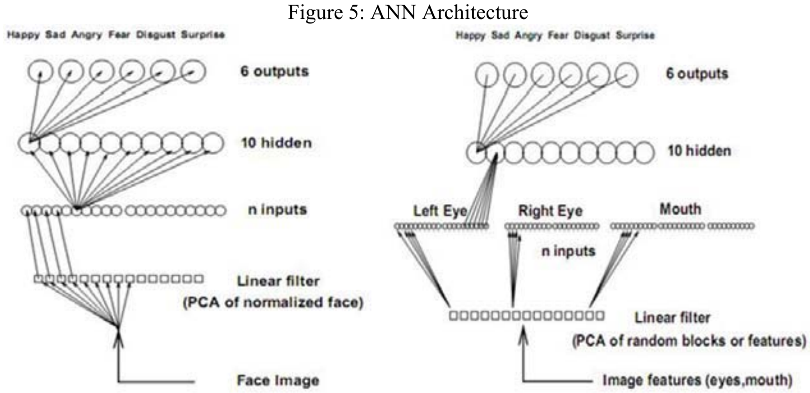
2

C. Face Expression (emotion) Classifications:

The research carried out by Ekman [4] on emotions and facial expressions is the main motivation behind the interest in the topic of emotion classification. First, face location determination and feature extraction was done from images, and then those extracted features were used as input to the classification system which in turn selects a predefined emotion category. Various methods have been introduced for the classification of emotions. All these approaches have targeted on classifying the six universal emotions. The non-universal emotion classification for emotions like wonder, amusement, greed and pity is yet to be taken into consideration. A good emotional classifier should be able to recognize emotions independent of gender, age, ethnic group, pose, lighting conditions, backgrounds, hair styles, glasses, beard and birth marks. In classifying emotions for video sequences, images corresponding to each frame have to be extracted, and the features extracted from the initial image have to be mapped between each frame.

1) Artificial Neural Networks (ANN): The most general method for answering this problem is through the use of a neural network which has obtained over 85% accuracy for emotion classification. Gargesha and Kuchi [5] has proposed a approach based on Multi Layer perceptrons and Radial Basis Function Networks (MLP and RBF networks) which classify the six basic types of emotions, here are classified into Sadness, Anger, Happiness, Fear, Surprise and Disgust. The description of neural network can be summarized as collection of units that are connected in some pattern to allow communication between the units. These entities (units) are referred as neurons or nodes. The output signals fed to other units along the connections which known as weight. Neural network has one/many hidden layer(s).

Figure 5: ANN Architecture



D. Facial recognition Methodology:

1) Gabor-Fisher Classifier (GFC) by Liu and Walsher.: The Gabor-Fisher Classifier is most robust to changes in illumination and facial expression. This method applies the Enhanced Fisher linear discriminate Model (EFM) to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images. The critical features of this method are -the derivation of an augmented Gabor feature vector, whose dimensionality is further, reduced using the EFM by considering both data compression and recognition (generalization) performance. The development of a Gabor–Fisher classifier for mulch-class problems extensive performance evaluation studies the feasibility of the new GFC method has been successfully tested on face recognition using 600 FERET frontal face images corresponding to 200 subjects, which were acquired under variable illumination and facial expressions. The novel GFC method achieves 100% accuracy on face recognition using only 62 features.

2) DCT-mod 2 by Sanderson and Paliwal.[7]:

DCT-mod 2 utilizes polynomial coefficients derived from 2D DCT coefficients obtained from horizontally & vertically neighbouring blocks via the use of various windows and diagonally neighbouring blocks. In DCT-mod2 feature extraction a given face image is analyzed on a block by block basis. Here the block is Np x Np and overlaps neighbouring blocks by 50%. Each block is divided in terms of 2D Discrete Cosine Transform (DCT) basis functions. A feature vector for each block is then obtained as:

$^{h}{}_{C0}\Delta^{\square}{}_{C0}\quad ^{h}{}_{C1}\Delta^{\square}{}_{C1}\ldots C_{M-1}]^{T}$

Where C_n represents the n-th DCT co-efficient. M is the number of retained DCT coefficients, $^{h}{}_{cn}$ and $^{v}{}_{cn}$ represent the horizontal and vertical delta coefficients respectively.

3) Self Organising Maps (SOMs):

With supervised learning technique there is a target output for each input pattern and the network learns to produce require output. In unsupervised learning the networks learn to form their own classifications of training data without external help. To do this we have to assume that class membership is classified broadly by the input patterns sharing common features and that the network will be able to identify those features across the range of input patterns. One particularly interesting class of unsupervised learning is based on competitive in which the neurons compete among themselves too be activated, with the result that only one is activated at one time. This activated neuron is called a winner. Such a competition can be induced or implemented by having lateral inhibition connections between the neurons. The result is that the neurons are forced to organise themselves. Such a network is called as Self Organising Map (SOM).

IV EXPERIMENTAL RESULTS

Methods of data analysis:

1) Performance Evaluation (Accuracy): To demonstrate the efficiency of the proposed method, extensive experiments will be conducted on the JAFFE/FERET [8] database. . Firstly, three samples of each facial expression per subject will be selected. The training images will select randomly and the rest will use for testing. Secondly, all the data in one subject will use as the test data and the remaining subjects will use as training samples. The experiments will repeat ten times, each time using a different subject as the test data. The confusion matrixes of the average recognition rate will present in Tables.

V. CONCLUSIONS AND FUTURE WORK

This work tries to address the problem of emotion recognition from an image based and a neural network based approach. Emotion recognition is still a difficult and a complex problem in computer science. Most researches have tried to address this problem in various ways. Emotions are expressed by humans in many different ways: facial expressions, pose, behavior, actions and speech. We chose facial expressions for the recognition of emotions in humans. This paper surveys some of the published work till date. With this literature survey which motivates us to introduce a parallel approach to implement over GPU to solve a facial expression recognition problem Our efforts will achieve better performance in terms of recognition rate and speed (more accurate and robust) than any existing methods. It even will have some capability in accurate emotion classifications when poses and lighting conditions vary. In future work, more detailed work and complex action units will be considered in our system. Our goal is to find a way which can acquire a better effect with getting higher recognition accuracy and speeding less recognition time.

V.ACKNOWLEGMENT

The authors would like to thank the anonymous reviewers of this paper for their valuable and constructive review comments which helped to improve the clarity and presentation of this paper.

VI .LIST OF FIGURES

VII .LIST OF TABLES

REFERENCES

[1] Proyecto Fin de Carrera "Face Recognition Algorithms", June 16, 2010 [2] Peter N. Belhumeur, " Ongoing Challenges in Face Recognition".
[2] Kyong, K. and Jung, K.," GPU Implementation of Neural Network. Pattern Recognition", vol. 37, no. 6, pp. 1311-1314. Pergamon, 2004.
[3] Ekman,"Facial action Coding System",COmsulting Psycologistic Press,CA 1998
[4] Gargesha and P Kunchi,EE 511-Artificial Neural Computing Systems, Spring 2002,Department Of Electrical Engineering,Arizon state Univer sity,Tempe,AZ 85287
[5] Yongsheng Gao, Member , IEEE , and Maylor K.H. Leung, Member , IEEE, "Face Recognition Using Line Edge Map",IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 6, june 2002
[6] Sanderson and K.K. Paliwal,"Fast feature extraction method for robust face verification C.",Elecronics Filters,5th Dec 2002.
[7] JAFFE database : h t t p : / / w w w . m i c . a t r . c o . j p / ~ m l yons/jaffe_temp.html
[8] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Amer. A, vol. 4, no. 12,pp. 2379–2394, Dec. 1987.
[9] Seyed Medi Lajevadi,Member IEEE,"Facial Expression Recognition in Perceptual Color Space" ,IEEE transaction on IMAGE PROCAESS- ING,Vol.21,No.8,Aug.2012.
[10] Juan Balarini,Nesmachnow,Rodriguez,"Facial recognition using Neural Networks over GPGPU",CLEI Elecronic Journal,Vol 15,No.3,Paper 6,Dec 2012.
[11] NVIDIA CUDA, "http://developer.download.nvidia.com/compute/Devzone/docs/html/C/doc/CUDA_C_Programmomg_Guide.pdf", Accessed June 2012.
[12] NVIDIA, "Nvidiacomputeunifieddevicearchitecture," http://www.nvidia.com/object/cuda.html, 2008.
[13] Zhiguo Niu ,Xuehong Que,"Facial expression recognition based on weighted principal component analysis and support vector machines ",Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference.