A Survey of Facial Expression Recognition Methods

Shail Kumari Shah

Department of CSE, Rajasthan College of Engineering for Women, Jaipur, India

Abstract: - Facial expression recognition (anger, sad, happy, disgust, surprise, fear expressions) is application of advanced object detection, pattern recognition and classification task. Facial expression is one of the most powerful and natural means for human beings to show their emotions. It has found its applications in human-computer interaction (HCI), robotics, border security systems, forensics, machine vision, video conferencing, user profiling for customer satisfaction, physiological research etc. Although humans can detect facial expressions with less effort and delay but it is still a challenge for the machine to fast and effectively detect facial expressions. Therefore algorithms should be developed to thought machines to understand facial gestures. This paper focuses on a review of different techniques for facial expression recognition.

Keywords: - *Gabor filter, Log-Gabor filter, Local Binary Pattern, PCA, LDA, SVM, K-NN*

I.   INTRODUCTION

Facial expression is one of the most Facial expression is one of the most powerful and natural means for human beings to show their emotions. It is the position or movement of the muscles beneath the skin of the face. This movements show the emotional state of an individual. The research study by Mehrabian [11] has indicated that 7% of the communication information is transferred by linguistic language, 38% by paralanguage, and 55% by facial expressions in human face-to-face communication. This, therefore, shows that facial expressions provide a large amount of information in human communication.

Recognition of facial expression is often a hard task. FACS (Facial Action Coding Systems) describes the changes in facial expression that human can detect by observing changes in facial muscles. Each component of facial movement is called an Action Units. It was published by Ekman and Friesen in 1978 [1]. FACS describes 44 Action Units.

Facial expression recognition is an interesting and challenging area. Its application is found in many areas like human-computer interaction (HCI), robotics(AIBO robots), border security systems, forensics, machine vision, video conferencing, user profiling for customer satisfaction, physiological research etc.

 Facial expression analysis consists of two different approaches and each approach has two different methodologies. When whole of the frontal face is use and processes in order to end up with the classifications of 6 universal facial expression prototypes: disgust, fear, joy, surprise, sadness and anger gives the outlines the first approach. Instead of using the face images as a whole, dividing them into some sub-sections for further processing forms up the main idea of the second approach for facial expression recognition. Geometric Based Parameterization is an old way which consists of tracking and processing the motions of some spots on image sequences. Facial motion parameters and the tracked spatial positioning & shapes of some special points on face, are used as feature vectors for the geometric based method. Rather than tracking spatial points and using positioning and movement parameters that vary within time, colour (pixel) information of related regions of face are processed in Appearance Based Parameterizations.

In this paper we provide a critical review of the most recent development in facial expression recognition. This paper is organized as follows: In Section II pre-processing task is explained. Section III provides a detailed review of some facial expression recognition methodologies. Finally, conclusions are in Section IV.

II.   IMAGE PRE-PROCESSING

Noise removal, normalisation against the variation of pixel position or brightness, segmentation, location or tracking of the face or its part comes under pre-processing phase. Rotation of the head and changes in illumination are one of the most important factors affecting the performance of facial expression recognition. Illumination affect can be removed by using Gabor filter. Turning of head can be eliminated by choosing central points of the eyes manually. Then the images are turned up to a point that the X parameter gets the same dimensions as the central points of the eye [12]. Then the face is cropped in rectangular according to face model explained in [8]. A variety of face detection techniques exists like Viola-Jones Method, Exhaustive search, Branch and Bound. However robust detection of face is still a difficult task.

III. METHODOLOGY

A. Feature Extraction

Feature extraction is one of the Feature extraction is one of the most imporatnt part of facial expression recognition. It is the process of extracting and isolating imporant desired feature from the face. Feature extraction converts pixel data into a higher-level representation - of shape, motion, colours, texture, and spatial configuration of the face or its components [fer-brief tutorial]. Deriving an effective facial representation from original images is an important step for successful facial expression recognition. If inadequate features are used, even the best classifier fails to achieve accurate recognition. In this paper we will discuss Gabor filter, Log-Gabor filter, Local Binary Pattern for feature extraction.

1) Gabor filter: For feature extraction Gabor filters are used. Gabor filters are applied to image to extract features aligned at certain angle and frequencies. It has advantage of having optimal localisation in both frequency and spatial domain. Certain orientations and frequencies are selected and used to differentiate between different facial expressions in images. A Gabor filter can be represented by the following equation:

$$g(x, y) = s(x, y)w(x, y)$$

Where *s(x, y)* is a complex sinusoidal known as the carrier, and *w(x, y)* is a 2-D Gaussian-shaped function known as the envelope [10].
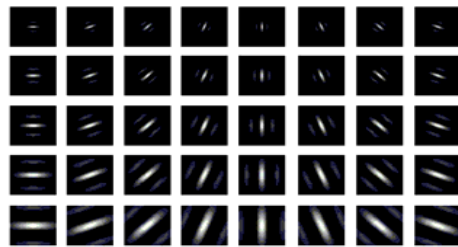

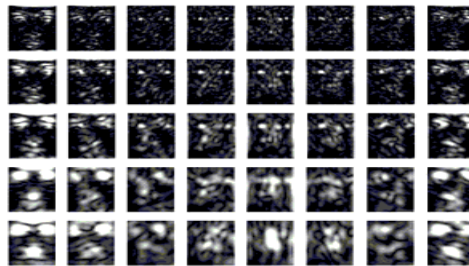Fig.1 The real part of the Gabor Filters

Fig.2 Sample image

Fig.3 The magnitudes of the Gabor feature representation of the sample face image

There are various forms to define this function, one normalized 2-D form being:

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)$$

Where

$$x' = x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$

In this equation *(x, y)* is the pixel position in spatial domain, λ represents the wavelength (reciprocal of frequency) in pixels of sinusoidal plane wave, θ represents the orientation of the Gabor filter, ψ is the phase offset, σ is the spatial width of the Gaussian envelope and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function. In most of the cases Gabor filter bank with 5 frequencies and 8 orientations is used to extract the Gabor features for face representation.

2) Log-Gabor Filter: Gabor filters are a traditional choice for obtaining localized frequency information. They offer the best simultaneous localization of spatial and frequency information. However they have two main limitations. The maximum bandwidth of a Gabor filter is limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization. An alternative to the Gabor function is the Log-Gabor function proposed by Field [1987]. Log-Gabor filters can be constructed with arbitrary bandwidth and the bandwidth can be optimized to produce a filter with minimal spatial extent. Gabor filters are not optimal to achieve broad spectral information with the maximum spatial localization. Furthermore, the Gabor filters are band pass filters, which may suffer from lost of the low and the high-frequency information. To achieve the broad spectral information and to overcome the bandwidth limitation of the traditional Gabor filter, Field proposed Log-Gabor filter. 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. One cannot construct Gabor functions of arbitrarily wide bandwidth and still maintain a reasonably small DC component in the even-symmetric filter. Log-Gabor functions, by definition, always have no DC component.

$$G(f) = exp\left\{\frac{-\left[log(\frac{f}{f0})\right]^2}{2\left[log(\frac{\sigma}{f0})\right]^2}\right\}$$

Therefore Log-Gabor filter can achieve better performance than Gabor filter.

3) Local Binary pattern: The original LBP operator was introduced by Ojala et al. [7] and was proved a powerful means of texture description. LBP operator takes the signs of the pixel differences between a pixel and its neighboring pixels to a binary code called LBP codes. Then the histogram of the binary code of image block is used for further analysis.

The first experiment of the LBP operator worked with eight neighbors i.e. 3×3 neighborhood of each pixel with the center value as a threshold. Where the center pixel's value is greater than the neighbor's value, we write "1". Otherwise, write "0". This gives 8 bit binary number. This binary number is then converted into decimal one. Based on the operator, each pixel of an image is labeled with an LBP code. Each code is called pattern. In this way each pixel is numbered.
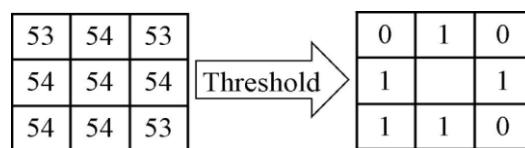
Pattern: 11101010
LBP code: 234
Fig. 4 An example of basic LBP operator

The limitation of the basic LBP operator is its small 3 ×3 neighborhood which cannot capture dominant features with large scale structures. Hence the operator later was extended to use neighborhood of different sizes. Using circular neighborhoods and bilinear interpolating the pixel values allows any radius and number of pixels in the neighborhood. In extended LBP operator , notation (P,R) is used to denotes neighborhood of P equally spaced sampling points on circle of radius R. LBP is robust against illumination variations because it only depends on magnitude relation between center pixel and its surrounding pixels.

B. Feature Reduction

Feature reduction or feature compression is a process of reducing excessive dimensionality of extracted feature from feature extraction step. Linear methods project the high dimensional data onto a lower dimensional space. There are two classical methods called PCA (Principle Component Analysis) and LDA (Linear Discriminent Analysis).

1) Principle Component Analysis (PCA): Principle Component Analysis also known as Karhunen-Loeve method is one of the most popular method for dimension reduction. It is also known as Eigenface method. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data come to lie on the first coordinate (called the first principle component), the second greatest variance on the second coordinate and so on. [13].

Consider a set of N sample images, $\{X_1, X_2, ...X_N\}$ represented by *t*-dimensional Gabor feature vector. The PCA [5] [4] can be used to find a linear transformation mapping the original *t*-dimensional feature space into a *g*-dimensional feature subspace, where normally $g << t$, the new feature vector is conducted with following equation:

$$\gamma_i = W_p X_i \quad i = 1...N$$

Where W_p is the linear transformations matrix, i is the number of sample images.
The columns of Wp are the *g* eigenvectors associated with the *g* largest eigenvalues of the scatter matrix S_T, which is defined as

$$S_T = \sum_{i=1}^{N} (X_i - \mu)(X_i - \mu)^T$$

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$$

2) Linear Discriminant Analysis (LDA): LDA is a powerful method for face recognition. It linearly transforms the original data space into low-dimensional feature space where data is well separated. The main difference between PCA and LDA is that in PCA the shape and location of the original data changes after transform while LDA doesn't change the location but tries to provide more class separability by drawing a decision region between the classes.

- Let suppose there are *C* classes.
- Let μ_i be the mean vector of class *i, i =1,2,3....,C*
- Let M_i be the number of samples within class *i, i=1,2,...C*
- Let $M = \sum_{i=0}^{C} M_i$ be the total number of samples.

In LDA, within class and between class scatter are used to formulate criteria for class separability. Within class scatter is the expected covariance of each of the classes. The scatter measures are computed using the equation:

$$S_{w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (Y_j - \mu_i)(Y_j - \mu_i)^T}$$

Between classes scatter matrix can be calculated using equation:

$$S_b = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T$$

Where μ is the mean of the entire dataset.

$$\mu = \frac{1}{C} \sum_{i=1}^{C} \mu_i$$

LDA computes transformation that maximizes the between class scatter while minimize the within class scatter.

Maximize=det (S_b)/det (S_w).

The solution obtained by maximizing this criterion defines the axes of the transformed space. The linear transformation is given by a matrix whose columns are the eigenvector of $S_w^{-1} S_b$.

IV.   CLASSIFICATION

Different expressions are categorised by a classifier. The two main types of classes used in facial expression recognition are action units (AU_s) [3], and the prototypic facial expressions defined by Ekman [2]. The 6 prototypic expressions are happiness, sadness, anger, surprise, fear, and disgust. An AU is one of the 46

atomic elements of visible facial movements or its associated deformation. These AU$_s$ are described in Facial Action Coding System (FACS).

1) Support Vector Machine (SVM): SVM are based on the concept of decision planes that define decision boundaries [6]. In this approach optimal classification of a separable two class problem is achieved by maximizing the width of the empty area (margin) between two classes. The margin width is defined as the distance between the discrimination hyper surface in n-dimensional feature space and the closest training pattern. The output of the SVM system is a label that classifies the grid under examination to one of the two basic (neutral, happy) facial expressions. One of the disadvantages of SVM is that it cannot be applied when the feature vectors defining samples have missing entries. However advantage of SVM over other classifiers is that SVM can achieve better generalization performance.

2) K-Nearest Neighbours (K-NN): K-NN is one of the simplest machine learning algorithms used for classification. The K-Nearest Neighbours (K-NN) method is a classical classification algorithm where the input feature vector is classified based on the class represented by the majority of the K nearest feature vectors obtained during the training process. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the Hamming distance [14].

V.   CONCLUSIONS

This paper has attempted to review a significant number of papers to cover the recent development in the field of facial expression recognition. Generally speaking, facial expression recognition is a difficult task. Present study reveals that for better facial expression recognition new algorithms can be develop like using Log Gabor filter (feature extractor) , PCA+LDA (feature reduction) and ANN may yield better result in terms of performance and accuracy.

REFERENCES

[1]   P. Ekman and W.V. Friesen, *Facial Action Coding System*, Palo Alto: Consulting Psychologist Press, 1978.
[2]   P. Ekman, Emotion *in the Human Face*, Cambridge University Press, 1982.
[3]   G. Donato, M.S. Bartlett, J.C. Hager, P.Ekman, T.J.Sejnowski, "*Classifying Facial Action*", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 21, No.10, pp.974-989,1999.
[4]   R. O. Duda, P. E. Hart, D. G. Stork, "*Pattern Classification.*" Wiley, New York (2001).
[5]   A. M. Martinez, A. C. Kak, "*PCA versus LDA*", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, 2001, pp.228-233.
[6]   C.W. Hsu and C. J. Lin, "*A comparison of Methods for multiclass Support Vector Machines*", IEEE Transactions on Neural Networks, vol. 13, no. 2, pp.415-425, 2002.
[7]   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, ISSN 971–987, 2002.
[8]   F.Y. Shih, C. Chaung, "*Automatic extraction of head and face boundaries and facial feature*", Information Sciences, Vol. 158, 2004, pp.117-130.
[9]   Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang, "A *New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA*", International Journal of Information Technology, Vol. 11, No. 11 ,2005.
[10]   Nectarios Rose", *Facial Expression Classification using Gabor and Log-Gabor Filters*", IEEE Computer Society, 0-7695-2503-2/06, 2006.[
[11]   A. Mehrabian, Nonverbal Communication. London, U.K.: Aldine, 2007.
[12]   Saeid Fazli, Reza Afrouzian, Hadi Seyedarabi, " *High Performance facial Expression Recognition using Gabor Filter and Probabilistic Neural Network*", 978-1-4244-4738-1/09, 2009 IEEE.
[13]   http://en.wikipedia.org/wiki/Principle_components_analysis.
[14]   http://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm.