A REGION BASED METHODOLOGY FOR FACIAL EXPRESSION RECOGNITION

Anastasios C. Koutlas
Dept. of Medical Physics, Medical School, University of Ioannina, Ioannina, Greece

Dimitrios I. Fotiadis
Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science University of Ioannina, Ioannina, Greece

Keywords: Facial expression recognition, Gabor filters, filter bank, artificial neural networks, Japanese Female Facial Expression Database (JAFFE).

Abstract: Facial expression recognition is an active research field which accommodates the need of interaction between humans and machines in a broad field of subjects. This work investigates the performance of a multi-scale and multi-orientation Gabor Filter Bank constructed in such a way to avoid redundant information. A region based approach is employed using different neighbourhood size at the locations of 34 fiducial points. Furthermore, a reduced set of 19 fiducial points is used to model the face geometry. The use of Principal Component Analysis (PCA) is evaluated. The proposed methodology is evaluated for the classification of the 6 basic emotions proposed by Ekman considering neutral expression as the seventh emotion.

1 INTRODUCTION

Facial expression recognition is an active research field that spawns across different subjects such as Human Computer Interaction (HCI), Smart Environments and medical applications. Recognizing facial expressions is a difficult task and therefore several limitations exist such as limitation due to lighting conditions, facial occlusions or facial hair.

In 1971 Ekman et.al determined 6 basic emotions; anger, fear, surprise, happiness, disgust and sadness (Ekman and Friesen, 1971). The neutral face expression is usually considered as the seventh basic emotion. Basic emotions are universal and exist in different human ethnicities and cultures. Even though the term emotion is used for categorization, emotions do not rely solely on visual information (Fasel and Luettin, 2003).

The task of Facial Expression Recognition can be divided into three main steps which are face recognition so that the face in an image is known for further processing, facial feature extraction which is the method used to represent the facial expressions and finally classification which is the step that classifies the features extracted in the appropriate expressions.

In general there are two approaches to represent the face and consequently the facial features. The first, often referred to as holistic approach, treats the face as a whole. Essa (Essa and Petland, 1997) treated the face holistically using optical flow and measured deformations based on the face anatomy. Donato (Donato et. al. 1999) has used several methods for facial expression recognition. Fisher linear discriminates (FLD) were used to project the images in a space that provided the maximal separability between classes and Independent Component Analysis (ICA) to preserve higher order information.

Instead of using the whole face, one can isolate and use the prominent features of a face, such as eyes, eyebrows, mouth, etc. Using fiducial points to model the position of the prominent features one can symbolize the face geometry in a local manner. The number of fiducial points used varies and mainly depends on the desired representation, as it is reported that different positions hold different

C. Koutlas A. and I. Fotiadis D. (2008).
A REGION BASED METHODOLOGY FOR FACIAL EXPRESSION RECOGNITION.
In *Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing*, pages 218-223
DOI: 10.5220/0001056602180223

information regarding the expressions (Lyons et. al., 1999). The way that these fiducial points are identified in an image can either be automatic (Gu et. al., 2005) or manual (Lyons et. al. 1999), (Guo and Dyer, 2005), (Zhang et. al. 1998).

It has been shown that simple cells in the primary visual cortex can be modeled by Gabor functions (Dougman, 1980), (Dougman, 1985). This solid physiological connection between Gabor functions and human vision has yielded several approaches to facial expression recognition (Lyons et. al. 1999), (Gu et. al., 2005), (Guo and Dyer, 2005), (Zhang et. al. 1998), (Liu and Wang, 2006), (Lyons and Akamatsu, 1998). Zhang (Zhang et. al., 1998) compared the Gabor function coefficients with the coordinate positions of the fiducial points and concluded that the first represent the face better than the latter. Donato (Donato et. al., 1999) reported that Gabor functions performed better than any other method used in both analytic and holistic approaches.

In this work we present a methodology for the classification of human emotions which is based on Gabor coefficients of the fiducial points. The methodology is based on Gabor coefficients which are extracted from a region around the fiducial points. It is noted in the literature that the feature vector is formed using single pixel values at the locations of the fiducial points. The proposed approach forms the feature vector from a region around each fiducial points gathering more information and avoiding in such a way artifacts which might exist close to the fiducial point. Furthermore, an alternate set of fiducial points is presented using just 19 landmark positions. We also attempted to reduce the number of fiducial points and to make the approach more efficient using PCA. The methodology is evaluated using the Japanese Female Facial Expression (JAFFE) database (Lyons and Akamatsu, 1998) in two cases: (a) using its full annotation and (b) excluding fear.

2 MATERIALS AND METHODS

The proposed methodology includes three stages (a) construction of the Gabor Filter Bank, (b) extraction of the Feature vector and (c) classification (Fig. 1).

2.1 Gabor Function

A two dimensional Gabor function $g(x,y)$ is the product of a 2-D Gaussian-shaped function referred as the envelop function and a complex exponential

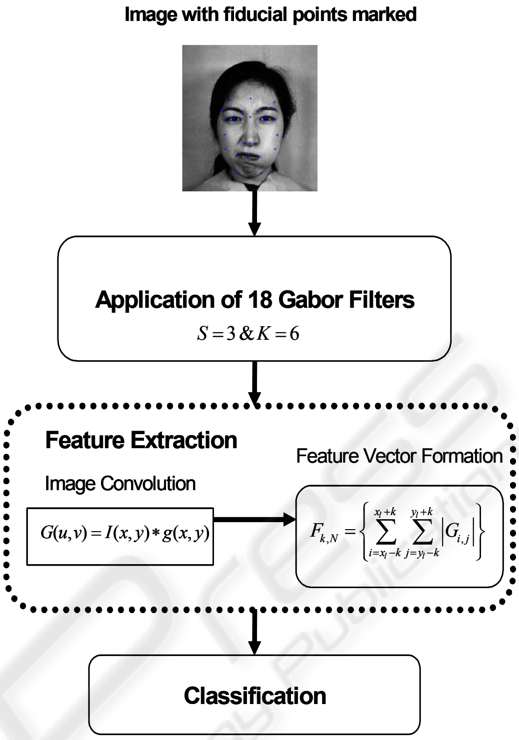


Figure 1: Flow chart of the proposed method.

(sinusoidal) known as the carrier and can be written as (Dougman, 1980), (Dougman, 1985), (Manjunath and Ma, 1996):

$$g(x,y)=\left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jW\right], \tag{1}$$

where x, y are the image coordinates, σ_x, σ_y are the variances in the x, y coordinates respectively and W is the frequency of the sine wave.

Its Fourier Transform $G(u,v)$ can be written as:

$$G(u,v)=\exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}, \tag{2}$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$.

2.2 Gabor Filter Bank

A Gabor filter bank can be defined as a series of Gabor filters at various scales and orientations. The application of each filter on an image produces for each pixel a response. The above representation (Eq. (1)) combines the even and odd Gabor functions as are defined in (Dougman, 1980).

If $g(x,y)$ is the mother function, we can derive the Filter bank functions using a series of rotations and dilations on the mother function:

$$g'(x,y) = g(x',y'), \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (3)$$

where $\theta = n\pi / K$, K is the total number of orientations and $n = 0, 1, \ldots, K-1$.

Manjunathan showed that Gabor filters form a nonorthogonal basis and that redundant information is included in the images produced by the filter (Manjunath and Ma, 1996), (Guo and Dyer, 2005). This leads to the following equations for the filter parameters a, σ_u and σ_v:

$$a = \left(\frac{U_h}{U_l}\right)^{\frac{1}{S-1}}, \quad W = a^m U_l, \quad (4)$$

$$\sigma_u = \frac{(a-1)W}{(a+1)\sqrt{2\ln 2}}, \quad (5)$$

$$\sigma_v = \tan\left(\pi/2K\right)\sqrt{\frac{W^2}{2\ln 2} - \sigma_u}, \quad (6)$$

where a is the scaling factor, S is the number of scales, $m = 0, 1, \ldots, S-1$, and U_h and U_l are the high and low frequencies of interest.

In this work we have chosen $U_h = \sqrt{2}/4$, $U_l = \sqrt{2}/16$ with three scales ($S = 3$) and six orientations ($K = 6$) differing each by $\pi/6$. Thus 18 complex Gabor filters were defined in total which will be used to extract the feature vector for each image. In Figure 2 the real part of the resulting filters is displayed.

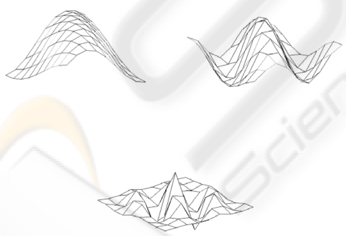


Figure 2: The real part of the Gabor filter when $\theta = 2\pi/6$ at all scales used.

2.3 Gabor Features

For any given image $I(x,y)$ its Gabor decomposition at any given scale and orientation can be obtained by convolving the image with the particular Gabor filter.

$$G(u,v) = I(x,y) * g(x,y) \quad (7)$$

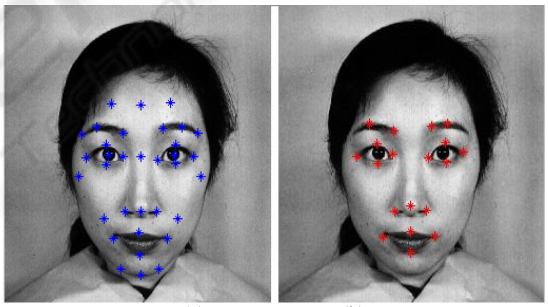
The magnitude of the resulting complex image is given:

$$|G| = \sqrt{Re(G)^2 + Im(G)^2} \quad (8)$$

All features derive from $|G|$ and the feature vector $F_{k,N}$ is formed according to the following formula:

$$F_{k,N} = \left\{ \sum_{i=x_l-k}^{x_l+k} \sum_{j=y_l-k}^{y_l+k} \left| G_{i,j} \right| \right\}, \quad l = 0,1,\ldots,N, \quad k = 0,1,\ldots,5, \quad (9)$$

where N is the number of the fiducial points, equalled to 19 and 34 respectively here. k is the number of neighbouring pixels used to form the regions. The feature vector can be portrayed as a square 1-norm of the matrix when $k \neq 0$, which corresponds to the intensity values of the mask around each fiducial point.



(a) (b)

Figure 3: Typical Positions of fiducial points (a) 34 points (b) 19 points.

2.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) are well known classifiers and can be used in multi-class problems. In the presented work we employed feed forward back propagation ANNs. The architecture of the ANNs consists of three layers. The first layer (input layer) consist of T input nodes where T is the dimension of the feature vector ($F_{k,N} \in R^T$). The second layer (hidden layer) consists of $T+C/2$ neurons, where C is the number of the classes. The sigmoid function is used as activation function for these hidden neurons. Finally the third layer (output layer) consists of C neurons. The activation function of the output neurons is the linear function. In order to train the ANNs the mean square error function is used and the number of epochs are 500.

2.5 Principal Component Analysis

In several cases T is quite large (for example when N in Eq. (9) is set to 34, the resulting feature vector has a dimension of 612). PCA is applied to reduce the input number features so that the retained features account for 95% of the total variance (sum of variances).

2.6 Dataset

The JAFFE (Lyons and Akamatsu, 1998) database was used for the evaluation of the proposed method. It features ten different Japanese women posing 3 or 4 examples for each basic emotion containing a total of 213 images. Including in the annotation of the dataset, neutral position is considered as a seventh basic emotion.

An alternate dataset derives from JAFFE database containing 181 images when fear is excluded. This can be justified in (Zhang et. al., 1998). Hereafter the two different datasets would be addressed as JAFFE-7 and JAFFE-6 with the latter excluding fear.

3 RESULTS

Several different sets of experiments were contacted with respect to:

i. The annotation used for classifications (i.e. either JAFFE-6 or JAFFE-7 datasets)
ii. The number of fiducial points used (N in Eq. (9) is equal to 19 or 34)
iii. The neighborhood size used to construct the feature vector (Single Pixel, 3x3, 5x5, 7x7, 9x9, 11x11)
iv. The employment or not of PCA for dimensionality reduction

The combination of the aforementioned sets leads to 48 different feature sets. For the evaluation the ten fold stratified cross validation method was used.

In the tables that will be presented below the abbreviations used correspond to the emotions, (SU for surprise, DI for disgust, FE for fear, HA for happy, NE for neutral, SA for sadness and finally AN for anger).

3.1 JAFFE-7

In this series of experiments the full annotation of the JAFFE dataset was used along with both facial representations (34 and 19 fiducial points). Table 1 displays the accuracy of each approach; the best performance was obtained when a neighborhood 11x11 of pixels was used with 34 fiducial points representing the face. When 19 fiducial points were used the accuracy declined only by 0.9% at max.

Table 1: Performance using the JAFFE-7 Dataset.

Region	34 Points	34 PCA	19 Points	19 PCA
Single Pixel	72.8%	53.5%	63.4%	47.4%
3x3	81.7%	74.6%	73.2%	60.1%
5x5	84.0%	79.3%	78.4%	71.4%
7x7	85.0%	78.9%	82.2%	73.7%
9x9	87.3%	82.6%	84.0%	80.8%
11x11	87.8%	83.6%	86.9%	82.6%

Table 2 displays the confusion matrix for the best performing approach. It can be seen that the poorest performance was obtained for the emotions of disgust and fear where the first was classified often as anger and the latter as sadness. Following the reasoning of Zhang (Zhang et. al., 1998) a second series of experiments were conducted.

Table 2: Confusion matrix for 34 fiducial points and 11x11 region.

	SU	DI	FE	HA	NE	SA	AN
SU	30	0	0	0	0	0	0
DI	0	24	0	0	0	1	4
FE	1	1	23	2	1	3	1
HA	0	0	0	27	3	1	0
NE	0	0	0	0	29	1	0
SA	1	0	1	1	1	27	0
AN	0	2	1	0	0	0	27

3.2 JAFFE-6

In this series of experiments fear was excluded from the classification process. The accuracy for each approach is shown in Table 3. The best performance was still obtained when using 34 fiducial points with accuracy 92.3%. Still the alternate dataset with 19 fiducial points provided similar results with accuracy 90.1%.

Table3: Performance using the JAFFE-6 Dataset.

Region	34 Points	34 PCA	19 Points	19 PCA
Single Pixel	75.7%	60.2%	65.2%	53.0%
3x3	85.6%	79.0%	76.8%	68.5%
5x5	87.3%	81.2%	81.8%	72.9%
7x7	89.5%	82.9%	84.0%	79.0%
9x9	91.7%	85.1%	85.6%	85.1%
11x11	92.3%	87.3%	90.1%	86.2%

In Table 4 and Table 5 the confusion matrices of these best performing experiments are presented. Disgust still is confused with anger in both cases. This yields that both these sets of fiducial points are not adequate enough to separate correctly these two emotions.

Table 4: Confusion matrix for 34 fiducial points and 11x11 region excluding fear.

	SU	DI	HA	NE	SA	AN
SU	29	0	0	1	0	0
DI	0	24	0	0	2	3
HA	0	0	31	0	0	0
NE	0	0	0	30	0	0
SA	3	0	1	1	26	0
AN	0	3	0	0	0	27

Table 5: Confusion matrix for 19 fiducial points and 11x11 region excluding fear.

	SU	DI	HA	NE	SA	AN
SU	29	0	0	1	0	0
DI	0	24	0	0	2	3
HA	0	0	31	0	0	0
NE	0	0	0	30	0	0
SA	3	0	1	1	26	0
AN	0	3	0	0	0	27

4 DISCUSSION

A facial expression recognition method, using a Gabor Filter Bank was presented. All redundant information in the construction of the filter bank was avoided by specially designing the filters. Two different facial representations were used using 19 and 34 fiducial points, respectively. Furthermore, the employment of a region based approach was investigated to avoid misclassification due to artefacts.

The manual feature reduction performed with the alternate dataset has reduced the feature vector by a factor of 0.4. The use of PCA, produced competitive results and has decreased the dimension of the feature vector by a factor of 0.9. In this work the fiducial points in the image were marked manually. This approach is possible to introduce errors, for example choosing a different point of interest instead of the one indented. By using regions the possibility of such errors taking place was minimized. The classifier performed weakly when tried to classify disgust and anger. Larkin (Larkin et. al., 2002) reported that males also made errors when decoding facial expressions of disgust, confusing it with anger. Facial expression recognition is a multi-class problem. Zhang (Zhang et. al. 1998), using a slightly different ANN, have reported accuracy ~90% when dealing with JAFFE-7 and 92.2% when using JAFFE-6. Guo (Guo and Dyer, 2005) had used JAFFE-7 and compared the performance of different classifiers. When the same feature vector was used (dimension equaled to 612) they reported accuracy 63.3% for the Simplified Bayes, 91.4% when using linear Support Vector Machines and 92.3% when using non linear (Gaussian Radial Basis Function Kernel) Support Vector Machines. Both of these approaches use a pixel-based feature extraction approach; in our case we employed a region-based feature extraction process, which permits some flexibility in the selection of the fiducial points and the affect of artifacts is minimized.

Further improvement of the presented method consists primarily of making the method automated. This is mainly related to the identification of the fiducial points that currently are manually marked. Furthermore, the use of a three-dimensional filter bank will be investigated by using time as a third constant and applied in a new, preferably video based, dataset.

ACKNOWLEDGEMENTS

This work was partly funded by the European Union and the General Secretariat for Research and Technology of the Hellenic Ministry of Development (PENED 2003 03OD139).

REFERENCES

Ekman, P, Friesen, WV, 1971, "Constants Across Cultures in the Face and Emotion", *J. Pers. Psycho.,* vol. 17, no. 2, pp. 124-129.

Fasel, B, Luettin, J, 2003, "Automatic Facial Expression Analysis: a survey", *Pattern Recognition,* vol. 36, no. 1, pp. 259-275.

Essa, I, Pentland, 1997, "Coding, Analysis, Interpretation, Recognition of Facial Expressions", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 757-763.

Donato, G, Bartlett, MS, Hager, JC, Ekman, P, Sejnowski, TJ, 1999, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 10, pp. 974-989.

Lyons, MJ, Budynek, J, Akamatsu, S, 1999, "Automatic Classification of Single Facial Images", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 12, pp. 1357-1362.

Gu, H, Zhang, Y, Ji, Q, 2005, "Task Oriented Facial Behaviour Recognition with Selective Sensing", *Computer Vision and Image Understanding,* vol. 100, no. 1-2, pp. 385-415.

Guo, G, Dyer, CR, 2005, "Learning From Examples in the Small Sample Case: Face Expression Recognition", *IEEE Trans. System, Man and Cybernetics-Part B: Cybernetics,* vol. 35, no. 3, pp. 477-488.

Zhang, Z, Lyons, M, Schuster, M, Akamatsu, S, 1998, "Comparison Between Geometry Based and Gabor Wavelet Based Facial Expression Recognition Using Multi Layer Perceptron", In *Proc. 3rd Int. Conf. Automatic Face and Gesture Recognition,* pp. 454-459.

Dougman, J, 1980, "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles", *Vision Research,* vol. 20, pp. 846-856.

Dougman, J, 1985, "Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Fields", *J. Opt. Soc. Am. A.,* vol. 2, no. 7, pp. 1160-1169.

Liu, W, Wang, Z, 2006, " Facial Expression Recognition Based on Fusion of Multiple Gabor Features", In Proc. 18th Int. Conf. on Pattern Recogntion,vol. 3, pp. 536-539.

Lyon, M, Akamatsu, S, 1998, "Coding Facial Expressions with Gabor Wavelets", In *Proc. 3rd Int. Conf. Automatic Face and Gesture Recognition,* pp. 200-205.

Manjunath, BS, MA, WY, 1996, "Texture Features for Browsing and Retrieval of Image Data", *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 837-842.

Larkin, KT, Martin, RR, McClain, SE, 2002, "Cynical Hostility and the Accuracy of Decoding Facial Expressions of Emotions", *J. Behavioural Medicine,* vol. 25, no. 3, pp. 285-292.