Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 53(67), Fascicola 2, 2008

Facial Expression Recognition under Noisy Environment Using Gabor Filters

Ioan Buciu[1], I. Nafornita[2], I. Pitas[3]

Abstract – Facial expression recognition is a major task concerning human-computer interaction issue. Plenty of techniques were proposed to recognize an expression either in still images or image sequences. However, most of them were applied for images recorded under controlled recording conditions. This paper aims at describing Gabor filters' application to extract facial features required to classify facial expression when the images are disturbed by various noise levels. The experiments indicate a satisfactory performance for Gabor filters when compared to another state-of-the-art method named principal component analysis (PCA).

Keywords: facial expression recognition, noisy environment, Gabor wavelets

I. INTRODUCTION

Computer-based facial expression recognition is a difficult task due to several factors such as face pose, scale, or illumination variations. Moreover, reliable facial expression recognition is a challenging task as there is no pure emotion. Rather, any specific emotion is a combination of several facial expressions. Generally, two sorts of approaches exist in this regard: appearance-based and geometric feature-based methods. The fiducial points are either manually selected [1] or automatically [2]. The face images are convolved then with Gabor filters and the output locally extracted from the face image at the corresponding fiducial points are used to build a new feature vector. Gabor filters can also be applied to the entire image [3] while the features are formed of the whole pixel space. As far as the geometric approaches are concerned, the respective points taken as features represent the face geometry through their spatial coordinates. It was proven when both approach types are combined the recognition performance is improved compared to a single approach. The facial expression recognition can be performed either on still images [1] or image sequences where temporal information is considered [4]. Guo and Dyer [5] addressed facial expression classification, when a small number of training samples was only available. A new linear programming-based technique was developed for both feature extraction and classification and a pairwise framework for feature selection was designed instead of using all classes simultaneously. Gabor filters were used to extract facial features and large margin classifiers such as support vector machines (SVMs) and AdaBoost were employed to recognize facial expressions. Their approach named "feature selection via linear programming" (FSLP) is able to automatically determine the number of selected features for each pair of classes in contrast to AdaBoost, which heuristically determines the number of features. A survey on automatic facial expression recognition techniques can be found on [6], [7], and [8].

All state-of-the-art methods were typically applied under controlled environmental conditions. A few works have been dedicated to the analysis of recognition performance for unfavorable environment such as occlusion [9], for example. Yet, no work was reported for facial expression recognition when the face images suffered noise distortion. This paper describes facial expression recognition task under noisy face images for two different facial expression databases. The features are extracted using global Gabor features which are further classified with the help of two distance measure-based classifiers. The approach is also compared with PCA. Experiments indicate superior performance for Gabor features compared to the one extracted by PCA, for several noise levels.

II. GABOR FEATURES

Gabor filters (known also as wavelets or functions) are based on physiological studies of simple cells in the human visual cortex. The cells are selectively tuned to orientation as well as spatial frequency, and their response can be accurately enough approximated by 2D Gabor filters [10]. Thus, the increased

[1] Electronics Department, Faculty of Electrical Engineering and Information Technology, University of Oradea, Universitatii ,1, 410087, Oradea, Romania, e-mail: ibuciu@uoradea.ro

[2]Electronics and Communications Faculty, „Politehnica" University of Timisoara, Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail: ioan.nafornita@etc.upt.ro

[3] Dept. of Informatics, Aristotle University of Thessaloniki, GR – 541-24, Box 451, Thessaloniki, Greece, e-mail: pitas@aiia.csd.auth.gr

popularity of this approach is biologically well justified. A 2D Gabor wavelet transform is defined as the convolution of the image $I(\mathbf{z}')$:

$$J_k(\mathbf{z}) = \int\int I(\mathbf{z}')\psi_k\left(\mathbf{z}-\mathbf{z}'\right)d\mathbf{z}' \qquad (1)$$

with a family of Gabor filters [2]:

$$\psi_k(\mathbf{z}) = \frac{\mathbf{k}^T\mathbf{k}}{\sigma^2} e^{\left(-\frac{\mathbf{k}^T\mathbf{k}}{2\sigma^2}\mathbf{z}^T\mathbf{z}\right)}\left(e^{i\mathbf{k}^T\mathbf{z}} - e^{-\sigma^2/2}\right) \qquad (2)$$

where $\mathbf{z} = (x,y)$ and $\mathbf{k}$ is the characteristic vector:

$$\mathbf{k} = \left(k_\nu \cos\varphi_\mu \quad k_\nu \sin\varphi_\mu\right)^T \qquad (3)$$

with $$k_\nu = 2^{-\frac{\nu+2}{2}\pi}, \quad \varphi_\mu = \mu\frac{\pi}{8} \qquad (4)$$

$$\nu = 0,1,2,3,4, \quad \mu = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$$

The parameters ν and μ define the frequency and orientation of the filter. Four orientations $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$ are used in our experiments. Two frequency ranges, i.e., high frequencies (hfr) with $\nu = 0,1,2$ and low frequencies (lfr) with $\nu = 2,3,4$ are considered.

III. DATA SETS DESCRIPTION

Two facial expression sets were used for experiments. The first data set of facial images used for the facial expression recognition task comes from the Cohn-Kanade AU-coded facial expression database [11]. The database was originally created for the representation of Action Units (AU) appearing in the FACS coding system and not for explicit facial expression recognition. The facial actions (action units) that are described in the image annotations have been converted into facial expression class labels according to [12]. Only thirteen subjects were picked up to form the data set as they displayed all six facial basic expressions, namely anger, disgust, fear, happiness, sadness and surprise. Each subject from Cohn-Kanade (C-K) database forms an expression over time starting from a neutral pose and ending with a very intense expression, thus having several video frames with different expression intensities. However, the number of these intermediate video frames is not the same for the various posers. We have selected three poses with low (close to neutral), medium and high (close to the maximum) facial expression intensity and used them to form the database utilized in our experiments. The total number of frontal images used was 234. The registration of each original frontal image x was performed by mouse clicking on the eyes, thus retrieving the eyes coordinates, followed by an image shift step for centering the eyes. Furthermore, the images were rotated to align the face horizontally according to the eyes. In the next step, the face region was cropped in order to remove the image borders, while keeping the main facial features (as eyebrows, eyes, nose and chin). Each 80 × 60 image was convolved with 12 Gabor filters, corresponding to the low frequency range and the four orientations aforementioned. Each resulting image was further downsampled by a factor of 3 to an image of 20 × 15 pixels, which was scanned row-wise to form a final feature vector of dimension 300 for each Gabor filter output. The 12 outputs have been concatenated to form a new longer feature vector of dimension 3600. Each feature vector has been next stored into the matrix $\mathbf{X}_{Gabor}$ of 3600 x n.

The second database contains 213 images of Japanese female facial expressions (JAFFE) [13]. Ten subjects produced three or four examples of each of the six basic facial expressions plus a neutral pose, thus producing a total of 213 images of facial expressions. Image registration was performed in the same way as for the C–K database, and Gabor features extraction was accomplished similarly.

IV. EXPERIMENTAL RESULTS

Prior running the experiments, the whole dataset (expressed now by the matrix X) is split into two disjoint sets called training set $X^{training}$ and test set X^{test}, one for each database. To form the training set, 164 and 150 face images were randomly chosen from the C–K derived and the JAFFE database, respectively, while the remaining 70 and 63 images were used for testing, thus forming the test image set. Both the training and the test set contain all facial expressions.

Our goal is to analyze the recognition performance corresponding to Gabor features compared to the features discovered by PCA for various face image degradation levels imposed by noise. As no real noisy facial expression image database exists to date, we have simulated this environment by adding white Gaussian noise only to the test set images. To quantify the noise amount we use the Signal-to-Noise Ratio (SNR) expressed in decibels (dB) as:

$$SNR = 10\cdot\log_{10}\frac{\sigma_x^2}{\sigma_N^2} \qquad (5)$$

where σ_x and σ_N is the signal and noise variance, respectively. We have employed five gradually noise levels expressed by the SNR for the following values SNR = {20, 10, 5, 2, 1}. The noisy face images from the test set are depicted in Figure 1 for the five SNR values. As more noise quantity is added the facial expression gets harder to be recognized. See, for instance, the fourth sample which corresponds to a "fear" expression which can be easily confused with a "smile" expression for SNR = 1 dB.

Gabor feature vectors corresponding to the original training images and the noisy test images are extracted using the procedure mentioned in the previous Section, leading to the Gabor feature vectors $\mathbf{F}_{Gabor}^{training}$. For the PCA method, eigenimages [14] $\mathbf{W}$ are computed and the training feature vectors are formed by projecting the original training images into the eigenimage subspace, i.e. $\mathbf{F}_{PCA}^{training} = \mathbf{W}^T\mathbf{X}$. Analogously, a noisy test image $\mathbf{x}_{test}$ is projected into the same eigenimage subspace yielding the test PCA feature vector, $\mathbf{f}_{PCA}^{test} = \mathbf{W}^T\mathbf{x}^{test}$.

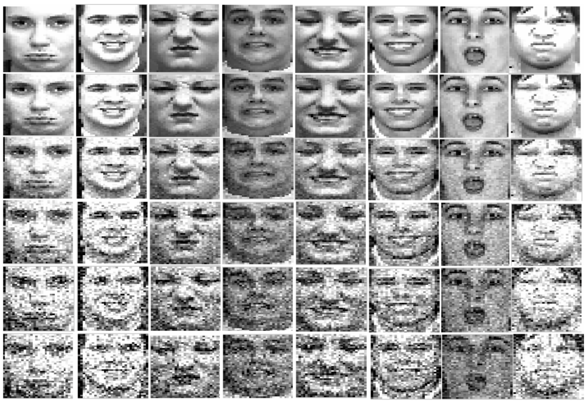
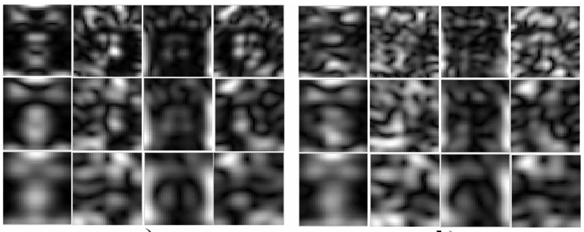


Figure 1 Each row depicts eight samples from the C-K test set for different noise levels expressed through SNR. First row illustrates noise-free face images, following, from top to bottom, images having different SNR values varying as SNR = {20, 10, 5, 2, and 1}.

We stress again the noise was added only for the test images.



a) b)

Figure 2 Magnitude of Gabor representation of the 50th image from the test convolved with 12 Gabor filters for $\nu = 0,1,2$ $\mu = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$. Here a) shows noise-free Gabor features while b) depicts noisy Gabor features corresponding to SNR = 1dB.

In Figure 2 the magnitude of the Gabor features (corresponding to hfr) is shown for one original test image and the Gabor features for the same face image disturbed by noise having SNR =1 dB.

Once the features are extracted, two distance measure based classifiers are employed in order to assign the test facial expression to a hopefully correct facial expression class. The recognition rate is defined as $RR = \#\{p(\mathbf{c}_{test}) = l(\mathbf{c}_{test})\}$ where $l(\mathbf{c}_{test})$ is the ground truth for $\mathbf{c}_{test}$, and $p(\mathbf{c}_{test})$ is the predicted value of the classifier. Two classifiers are used to classify the Gabor and PCA feature vectors:

1. Cosine similarity measure (CSM). This approach is based on the nearest neighbor rule and uses as similarity the angle between a test vector and a training one. Let $\mathbf{x}_j^{training}$ (in the case of Gabor) be a column vector of $\mathbf{X}_{Gabor}^{training}$ that corresponds to the nearest class L_j. Let also $\mathbf{x}_j^C$ be the nearest L_j^C class neighbor column vector for a test coefficient vector $\mathbf{x}_{Gabor}^{test}$. We compute the quantities:

$$d_j = \frac{\mathbf{x}_{Gabor}^{T(test)}\mathbf{x}_j}{\left\|\mathbf{x}_{Gabor}^{T(test)}\right\| \cdot \left\|\mathbf{x}_j\right\|} \text{ and } d_j^C = \frac{\mathbf{x}_{Gabor}^{T(test)}\mathbf{x}_j^{C(training)}}{\left\|\mathbf{x}_{Gabor}^{T(test)}\right\| \cdot \left\|\mathbf{x}_j^{C(training)}\right\|}$$

where d_j and d_j^C are the cosines of the angles between a test feature vector and the nearest training one. We assign $\mathbf{x}_{test}^{Gabor}$ to L_j if $d_j > d_j^C$. Otherwise, $\mathbf{x}_{Gabor}^{test} \in L_j^C$.

2. *Maximum correlation classifier* (MCC). The second classifier is the minimum Euclidean distance classifier. The Euclidean distance from $\mathbf{x}_{Gabor}^{test}$ to $\mathbf{x}_j^{training}$ is defined as:

$$\begin{aligned}\left\|\mathbf{x}_{Gabor}^{test} - \mathbf{x}_j^{training}\right\|^2 &= -2[\mathbf{x}_j^{T(training)}\mathbf{x}_{Gabor}^{test} - \\ &-\frac{1}{2}\left\|\mathbf{x}_j^{training}\right\|^2] + \mathbf{x}^{T(test)}\mathbf{x}_{Gabor}^{test} = \\ &= -2h_j(\mathbf{x}_{Gabor}^{test}) + \mathbf{x}_{Gabor}^{T(test)}\mathbf{x}_{Gabor}^{test}\end{aligned}$$

where $h_j(\mathbf{x}_{Gabor}^{test})$ represents a linear discriminant function of $\mathbf{x}_{Gabor}^{test}$. A test image is classified by this classifier computing two linear functions $h_j(\mathbf{x}_{Gabor}^{test})$ and $h_j^C(\mathbf{x}_{Gabor}^{test})$ and assigning the test image to the class associated to the maximum disciminant function value.

The facial expression recognition results for C-K and JAFFE databases are presented in Table 1 and 2 for both low and high frequency ranges. Several observations can be drawn accordingly. Firstly, the Gabor features based facial expression recognition is superior to the PCA based recognition regardless of the noise level. For the C-K facial expression database, low frequencies lead to higher recognition performance. On contrary, for the JAFFE images, a slight improvement in recognition rate is obtained with high frequency when the noise level is low. However, as the noise level increases, low frequencies based Gabor features outperform, in terms of facial expression recognition, the features extracted by high frequency Gabor filters as noted from Table2. As far as the classifiers are concerned, CSM is preferred as it conducts to higher recognition performance for both Gabor and PCA features and both facial expression databases.

Table 1 Recognition results (%) corresponding to the C-K database for the two classifiers (CSM and MCC). "hfr" and "lfr" stands for high and low frequency range (see text for details). The highest recognition rate is drawn in bold.

Environment	Classifier	Feature extraction method		
		Gabor		PCA
		hfr	lfr	
SNR = 20 dB	CSM	80	**84.29**	74.28
	MCC	80	82.86	72.85
SNR = 10 dB	CSM	80	**82.86**	72.85
	MCC	80	**82.86**	71.55
SNR = 5 dB	CSM	80	**81.43**	72.85
	MCC	80	**81.43**	71.55
SNR = 2 dB	CSM	78.57	**81.43**	72.85
	MCC	74.29	72.86	68.55
SNR = 1 dB	CSM	77.14	**77.14**	71.43
	MCC	74.29	74.29	68.57

Table 2 Recognition results corresponding to the JAFFE database and for the two classifiers (CSM and MCC). "hfr" and "lfr" stands for hig and low frequency range (see text for details). The highest recognition rate is drawn in bold.

Environment	Classifier	Feature extraction method		
		Gabor		PCA
		hfr	lfr	
SNR = 20 dB	CSM	**77.78**	76.19	63.49
	MCC	**77.78**	76.19	60.31
SNR = 10 dB	CSM	**76.79**	76.19	63.49
	MCC	73.02	76.19	60.31
SNR = 5 dB	CSM	74.60	**75.37**	58.07
	MCC	72.78	**75.37**	58.50
SNR = 2 dB	CSM	68.25	**73.02**	59.60
	MCC	69.84	69.84	59.60
SNR = 1 dB	CSM	69.84	**73.02**	59.90
	MCC	69.84	**73.02**	58.32

III. CONCLUSIONS

The papers dealt with the facial expression recognition for noise face images. Global Gabor as well as PCA based features are extracted followed by a classification procedure using two standards distance based classifiers. The experiments were run for five different noise levels and they revealed the superiority of the Gabor features over the PCA features for each noise environment condition.

REFERENCES

[1] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometric-based and Gabor-wavelets-based facial expression recognition using multi-layer parceptron", in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 454 – 459, 1998

[2] L. Wiskott, J.- M. Fellous, N. Kruger, and C. von der Malsburg, "*Face recognition by elastic bunch graph matching*", IEEE. Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775 – 779, 1999.

[3] I. Buciu, C. Kotropoulos and I. Pitas, "*ICA and Gabor representations for facial expression recognition*" Proc. 2003 IEEE International Conference on Image Processing. pp. 1054 – 1057, 2003.

[4]. G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "*Dynamics of facial expression extracted automatically from video*", Image and Video Computing, vol. 24, no. 6, pp. 615-625, 2006.

[5] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: face expression recognition," IEEE Tran. On Systems, Man, and Cybernetics, Part II, vol. 35, no. 3, pp. 477 – 488, 2005.

[6] M. Pantic and L .J. M. Rothkrantz, "*Automatic analysis of facial expressions: The state-of-the-art*," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424--1445, Dec., 2000.

[7] M. Pantic and L .J. M. Rothkrantz, "*Facial action recognition for facial expression analysis from static face images*," IEEE Trans. Systems, Man, and Cybernetics, Part B, vol. 34, no. 3, pp. 1449--1461, June, 2004.

[8] B. Fasel and J. Luettin, "*Automatic Facial Expression Analysis: A survey*," Pattern Recognition, vol. 1, no. 30, pp. 259--275, 2003.

[9] I. Kotsia, I. Buciu, and I. Pitas, "*An analysis of facial expression recognition under partial facial image occlusion*", Journal of Image and Visual Computing, vol. 26, Issue 7, pp. 1052-1067, July, 2008.

[10] T. Lee, "*Image representation using 2d Gabor wavelets*", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 10, pp. 959--971, 1996.

[11] T. Kanade, J. Cohn and Y. Tian, "*Comprehensive database for facial expression analysis*," in Proc. IEEE Inter. Conf. on Face and Gesture Recognition, pp. 46--53, 2000.

[12] M. Pantic and L. J. M. Rothkrantz, "*Expert system for automatic analysis of facial expressions*," Image and Vision Computing, vol. 18, no. 11, pp. 881--905, 2000.

[13] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "*Coding facial expressions with gabor wavelets*," in Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit., vol. 8, pp. 200–205, 1998.

[14] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "*Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.