Recognition of Facial Expressions in Presence of Partial Occlusion

Ioan Buciu, Irene Kotsia, Ioannis Pitas

Computer Vision and Image Processing Group
AIIA Laboratory, Department of Informatics
Aristotle University, 54124 Thessaloniki, Greece
E-mail: {nelu,ekotsia,pitas}@zeus.csd.auth.gr
Web: http://poseidon.csd.auth.gr

Abstract. Nowadays, both computer vision researchers and psychology experts show an increased interest for human facial expression analysis. Despite the huge amount of research that has been dedicated to this area, almost all of them concern data recorded in controlled laboratory conditions, which does not always reflect the real world environment in which the human face is partially occluded. Six basic facial expressions are investigated in that case, i.e. when the eyes and eyebrows or the mouth regions are left out. We are interested in finding which part of the face comprised sufficient information with respect to the entire face, in order to correctly classify these six expressions. Each image from the two databases used is convolved with a set of Gabor filters having various orientations and frequencies. The new feature vectors are classified using a maximum correlation classifier and the cosine similarity measure approaches. Overall, the method provides robustness against partial occlusion.

1 Introduction

The non verbal communication systems, such as the facial expression mechanism have seized an increased interest not only from a psychological perspective point of view but also from the computer vision researchers who try to develop a complex human-computer interface that is capable of automatically recognizing and classifying the human expressions or emotions. Biometrics is another area where facial expressions seem to have an impact. A person is identified by determining the authenticity of a specific physiological or behavioral characteristic possessed by a user under different various face expressions. This task encounters many difficulties, as it has to overcome the difficulties that appear when the human face is under different environmental pose conditions (different illumination conditions, emotion intensity variation, etc). Furthermore, the quantification of emotion is something quite ambiguous, as there is no clear definition of pure emotion. More precisely, a particular emotion is a combination of several facial expressions that can be coded according to, for instance, the Facial Action Coding System (FACS), which is built by a set of parameters called action units (AUs), that are used to define measurements of appearance changes in the face [3].

Much work has been done on facial expression recognition. Approaches to automatic facial expression analysis attempt to recognize either a small set of prototypic emotional facial expressions (anger, disgust, fear happiness, sadness, surprise) [10] or a larger set of facial actions (AUs) [2]. A survey on automatic facial expression analysis can be found in [4].

Although promising results have been reported on facial expression analysis, the experiments have been conducted in controlled laboratory conditions which do not always reflect the real-world conditions. For example, it is a quite common case that the face is occluded by a scarf or sunglasses, causing the classifier accuracy to decrease. Despite the importance of building an automatic facial expression classifier capable to cope with occluded faces, there is no much research in this regard. Recognition of facial expressions in the presence of occlusion is investigated in the paper [1]. The approach is based on localized representation of facial expression feature, and on fusion of classifier outputs. Facial points are automatically tracked over an image sequence, and used to represent a face model. The classification of facial expressions is then performed by using decision level fusion that combines a local interpretation of the face model and a general classification score.

The purpose of the work presented in this paper is to simulate this scenario and to perform experiments in order to determine the part of the face that contains the most discriminative information for facial expression recognition.

2 Data Description

The experiments have been performed using two databases. For the first database, called Japanese female facial expression (JAFFE) [8], ten expressers posed 3 or 4 examples of each of the 6 basic facial expressions (anger, disgust, fear, happiness, sadness, surprise) plus neutral pose, for a total of 213 images of facial expressions. A second database have been derived from Cohn-Kanade AU-coded facial expression database [6] that contains single or combined action units. Facial action (action units) has been converted into emotions according to [11]. A number of thirteen subjects have been chosen to form the second database. Each subject expresses six basic emotions and each emotion has 3 intensities. Therefore, the total number of samples that forms the second database is 234.

Each original image $\mathbf{x}$ has been aligned with respect to the upper left corner. We superimposed a black rectangle around the eyes and mouth regions to partially occlude the face, and cropped each occluded image and downsamped it afterwards, in a such way that its final size was 80×60 pixels. Figure 1 presents one expresser from JAFFE database and one from Cohn-Kanade database, respectively, posing 7 facial expressions (emotions) and the corresponding images that were formed for the eyes occlusion case.

3 System Description

The block diagram of the method proposed in the paper is described in the Figure 2.

3.1 Feature Extraction

We used 2D Gabor wavelets for feature extraction since the Gabor wavelet-based method can achieve high sensitivity for facial expression classification and give the

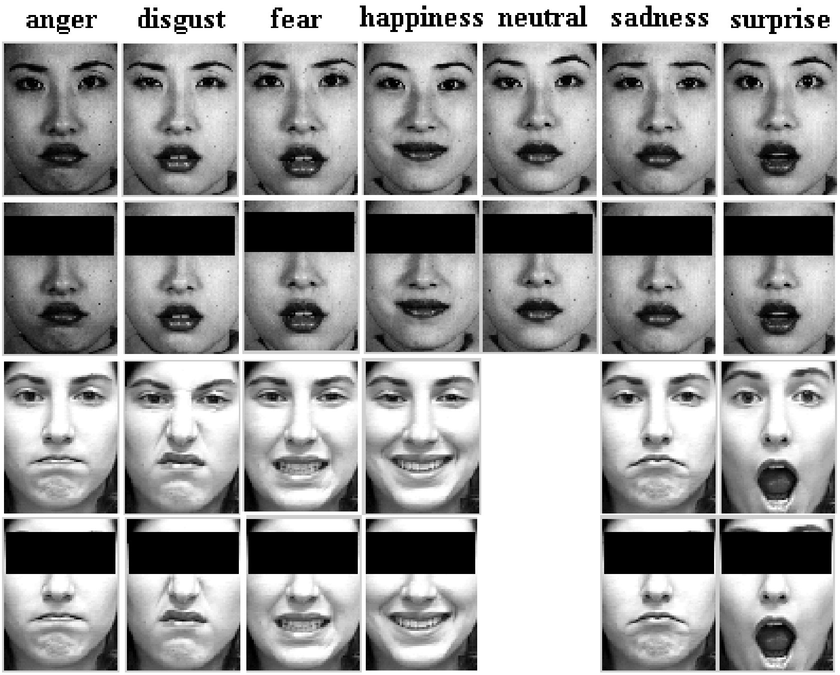

Fig. 1. An example of one expresser from JAFFE database and one from Cohn-Kanade database, respectively, posing 7 facial expressions (emotions) and the corresponding eyes occluded region.

best reported results [2, 8, 9, 12]. Moreover, recent studies have shown that Gabor elementary functions are suitable for modeling simple cells in visual cortex [7]. We applied Gabor filters (GF) to the entire face area instead of using specific regions, avoiding to manually select the interested regions in order to extract facial features.

A 2D Gabor wavelet transform is defined as the convolution of the image $\mathcal{I}(\mathbf{z})$:

$$\mathcal{J}_{\mathbf{k}}(\mathbf{z}) = \int\int \mathcal{I}(\mathbf{z})\psi_{\mathbf{k}}(\mathbf{z}-\mathbf{z}')\mathbf{dz} \tag{1}$$

with a family of Gabor filters [13]:

$$\psi_{\mathbf{k}}(\mathbf{z}) = \frac{k^2}{\sigma^2}\exp(-\frac{k^2}{2\sigma^2}x^2)\Big(\exp(i\mathbf{kx}) - \exp(-\frac{\sigma^2}{2})\Big), \tag{2}$$

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

$$\mathbf{k} = \begin{pmatrix} k_\nu \cos\varphi_\mu \\ k_\nu \sin\varphi_\mu \end{pmatrix},$$

with

$$k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \;\; \varphi_\mu = \mu\frac{\pi}{8}. \tag{3}$$

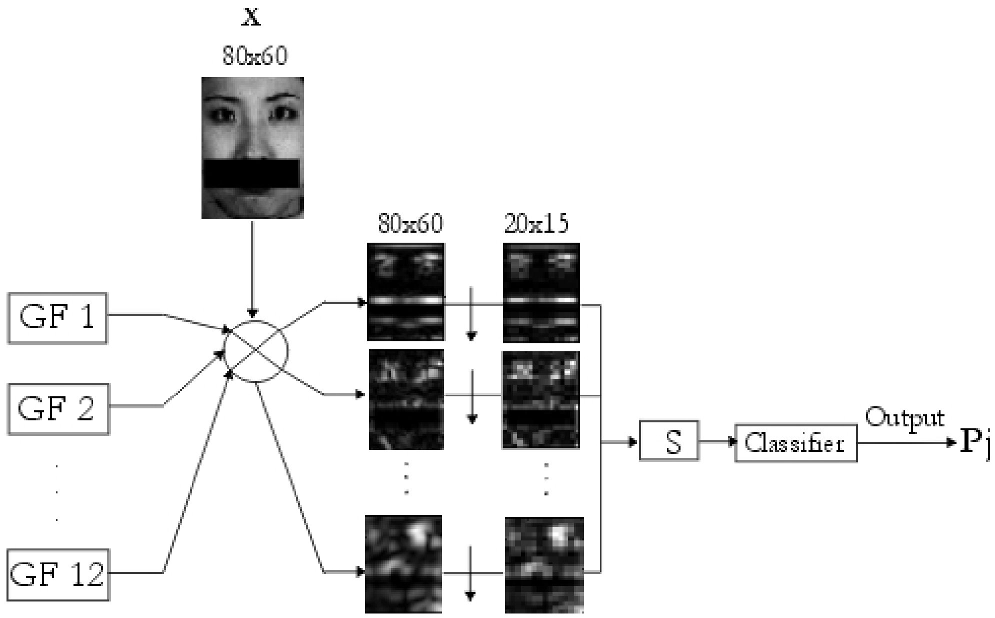

Fig. 2. Diagram block of the proposed system

The parameters ν and μ define the frequency and orientation of the filter. In our implementation, we used four orientations 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ and two frequency ranges: high frequencies (hfr) with ν = 0,1,2 and low frequencies (lfr) with ν = 2,3,4.

In the beginning, a new feature vector is formed by convolution of the 80 $\times$ 60 image **x** with 12 Gabor filters corresponding to high frequency range and orientation, and then downsampled to an image of 20 $\times$ 15 pixels and scanned row by row in order to form a vector of dimension 1 $\times$ 300 for each Gabor filter. The magnitude of Gabor representation, was the only result used, as it varies slowly with the position, while the phases are quite sensitive. The 12 outputs have been concatenated to form a new longer feature vector **s** of dimension 1 $\times$ 3600. The same procedure has been followed for the low frequency range.

3.2 Classification Procedure

The 6 basic facial expressions plus the neutral pose, form 7 classes. Let us denote the classes by $\mathcal{L}_j, j = 1,2,\ldots,7$. The label $\mathcal{L}_j$ is denoted by l_j. Hence, $\mathcal{L} = \{$*anger, disgust, fear happiness, neutral, sadness, surprise*$\}$. In the classical classification problem, we construct a classifier where the output (predicted value) of the classifier for a test sample $\mathbf{s}_{test}$, is p_j. The classifier accuracy is defined as $\#\{l(\mathbf{s}_{test}) = p(\mathbf{s}_{test})\}$. Once we have formed the 7 classes of the new feature vectors (or prototype samples) two kinds of classifiers are employed in order to classify a new test sample:

1. *Cosine similarity measure* (CSM). This approach is based on the nearest neighbor rule and uses as criterion the angle between a test sample and a prototype one. We compute the quantity:

$$CSM = argmin_{j=1,\dots,7}\{d_j\} \tag{4}$$

where

$$d_j = \frac{\mathbf{s}_{test}\mathbf{s}_j^T}{||\mathbf{s}_{test}||||\mathbf{s}_j||} \tag{5}$$

and d_j are the cosines of the angle between a test sample $\mathbf{s}_{test}$ and the prototype one $\mathbf{s}_j$.

2. *Maximum correlation classifier* (MCC). The second classifier is a minimum Euclidean distance classifier. The Euclidean distance from $\mathbf{s}_{test}$ to $\mathbf{s}_j$ is expressed as

$$\begin{aligned} ||\mathbf{s}_{test} - \mathbf{s}_j||^2 &= -2[\mathbf{s}_j\mathbf{s}_{test}^T - \frac{1}{2}||\mathbf{s}_j||^2] + \mathbf{s}_{test}\mathbf{s}_{test}^T \\ &= -2h_j(\mathbf{s}_{test}) + \mathbf{s}_{test}\mathbf{s}_{test}^T, \end{aligned} \tag{6}$$

where $h_j(\mathbf{s}_{test})$ is a linear discriminant function of $\mathbf{s}_{test}$. A test image is classified by this classifier by computing seven linear discriminant functions and choosing:

$$MCC = argmax_{j=1,\dots,7}\{h_j(\mathbf{s}_{test})\} \tag{7}$$

4 Experimental Results and Discussion

Since the database is limited, the classifier accuracy is measured using the leave-one-out strategy which makes the most of the available data used and produces averaged results as well. Table 1 shows the experimental results with no occlusion, as well as for mouth and eyes region occluded for both classifiers and Gabor filters with low and high frequency range. The highest accuracy concerning the experiments of the occluded face regions has been achieved using Gabor filters with low frequency range. However, the same conclusion is derived when the experiments have been conducted with no occlusion images. More precisely, for the JAFFE database, the accuracy is almost the same, regardless of the occluded region, while for the Cohn-Kanade derived database it

Database	Classifier	No occlusion		Mouth occluded		Eyes occluded	
		lfr	hfr	lfr	hfr	lfr	hfr
JAFFE	CSM	88.8	81.7	83.5	80	83	77.2
	MCC	89.7	82.6	83.5	80.8	84	78.2
Cohn-Kanade	CSM	94.5	90.6	86.4	83.3	91.5	88.5
	MCC	93.6	90.2	87.2	83	92.3	88.9

Table 1. Classifier accuracy in percentage (%) with no occlusion, mouth and eyes region occluded for both classifiers and Gabor filters with low and high frequency range.

turns out that the eyes region is not as important in terms of class separability, as the mouth region, since in the latter case the classifier accuracy is decreased more when the eyes are occluded.

In order to verify this result and have some knowledge about the inter-class and intra-class dispersion, we measured the class separability by computing the within-class scatter matrix S_W, the between-class scatter matrix S_B and the mixture scatter matrix $S_M = S_W + S_B$ corresponding to each class [5]. The criterion used for the class separability measurement was $J = trace(S_M)/trace(S_W)$. This number is large when S_B is dispersed or the scatter of S_W is small.

In the following matrices, the results have the following coding:

EA occluding Eyes Area
MA occluding Mouth Area
HFRG applying High Frequency Range Gabor filtering
LFRG applying Low Frequency Range Gabor filtering
EAHFRG occluding Eyes Area and applying High Frequency Range Gabor filtering
MAHFRG occluding Mouth Area and applying High Frequency Range Gabor filtering
EALFRG occluding Eyes Area and applying Low Frequency Range Gabor filtering
MALFRG occluding Mouth Area and applying Low Frequency Range Gabor filtering

Jaffe database experiments	J
EA	1.0767
MA	1.0937
HFRG	1.1470
LFRG	1.1578
EAHFRG	1.0889
MAHFRG	1.1274
EALFG	1.1083
MALFRG	1.1368

Table 2. Class separability measurement for Jaffe database.

From the above matrices, we can see that the results produced are consistent with the ones shown in Table 1. More precisely, the recognition accuracy achieved when low frequency range Gabor filters are used, is greater than that when applying high frequency range Gabor filtering. For example, in Jaffe database, J is equal to 1.1274, for the results produced using high frequency range Gabor filters when occluding the mouth area, and is less than the equivalent J computed for the low frequencies, 1.1368. The results derived from the experiments for the Cohn-Kanade database are in agreement with the assumptions made above (where J is equal to 1.0408 and 1.0428 for the case of mouth area occluding and applying high and low frequency range Gabor filtering, respectively).

Cohn-Kanade database experiments	J
EA	1.0409
MA	1.0256
HFRG	1.0655
LFRG	1.0779
EAHFRG	1.0590
MAHFRG	1.0408
EALFRG	1.0674
MALFRG	1.0428

Table 3. Class separability measurement for Cohn-Kanade database.

5 Conclusion

Facial expression recognition in the presence of mouth and eyes occlusion has been investigated to determine the part of the face that contains the most discriminative information for emotion classification task. The feature vectors that have been extracted from the original images by convolving them with a set of Gabor filters are classified by using maximum correlation classifier and cosine similarity measure.

We compared the classification results data with the equivalent data when no occlusion is observed to find that, overall, the system is robust against partial occlusion of a face. The accuracy achieved when partial occlusion is observed, is almost the same for the different cases of occlusion (eyes area or mouth area occlusion), while it turns out to be greater if low frequency range Gabor filtering is used.

Aknowledgement

This work was supported by the European Union Research Training Network "Multimodal Human-Computer Interaction (HPRN-CT-2000-00111).

References

1. F. Bourel, C. C. Chibelushi, and A. A. Low. Recognition of facial expressions in the presence of occlusion. *Proceedings BMVC Conference (Poster Session 1)*, 2001.
2. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.
3. P. Ekman and W.V Friesen. Manual for the facial action coding system. *Consulting Psychologists Press*, 1977.
4. B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 1(30):259–275, 2003.
5. K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, 1990.

6. T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *Proceedings IEEE International Conference on Face and Gesture Recognition*, pages 46–53, March 2000.
7. T. Lee. Image representation using 2d gabor wavelets. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996.
8. M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. *Proceedings 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, April 1998.
9. M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, December 1999.
10. M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, December 2000.
11. M. Pantic and L. J. M. Rothkrantz. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, (18):881–905, March 2000.
12. Y.-Li Tian, T. Kanade, and J. Cohn. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proceedings 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 229–234, May 2002.
13. L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 19(7):775–779, December 1997.