FACIAL EXPRESSION ANALYSIS UNDER PARTIAL OCCLUSION

I. Buciu   I. Kotsia   and   I. Pitas

Department of Informatics, Aristotle University of Thessaloniki
GR-54124, Thessaloniki, Box 451, Greece, {nelu,ekotsia,pitas}@zeus.csd.auth.gr

ABSTRACT

Six basic facial expressions are investigated when the human face is partially occluded, i.e. when the eyes and eyebrows or the mouth regions are occluded. Such occlusions occur when a person wears glasses (e.g. in VR application) or a mouth mask (e.g. in medical application). More specifically, we are interested in finding the part of the face that contains sufficient information in order to correctly classify these six expressions. Two facial image databases are employed in our experiments. Each image from the database is convolved with a set of Gabor filters having various orientations and frequencies. The new feature vectors are classified by using a maximum correlation classifier and the cosine similarity measure approaches. We find that, overall, the facial expression recognition method provides robustness against partial occlusion, the classification accuracy only decreasing from 89.7 % (no occlusion) to 84 % (eyes region occlusion) and 83.5 % (mouth region occlusion) for the first database and from 94.5 % (no occlusion) to 91.5 % (eyes region occlusion) and 87.2 % (mouth region occlusion) for the second database, respectively.

1. INTRODUCTION

The non verbal communication systems, such as the facial expression mechanism, have captured an increased interest not only from a psychological perspective, 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. It is difficult to quantify emotions since there is no pure emotion. Rather, a particular emotion is a combination of several facial expressions that can be coded according to, for instance, the Facial Action Coding System (FACS) by a set of parameters called action units (AUs) which define measurements of appearance changes in the face [1].

Plenty of 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) [2] or a larger set of facial actions (AUs) [3]. 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 may happen that the face is occluded by a scarf or sunglasses, causing the classifier accuracy to decrease. Such occlusions occur when a person wears glasses (e.g. in VR application) or a mouth mask (e.g. in medical application). 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 [5]. 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 local interpretation of the face model into a general classification score.

The purpose of the work presented in this paper is to simulate this scenario and to perform experiments 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. The first database we used for our experiments contains 213 images of Japanese female facial expression (JAFFE) [6]. 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 [7] that contains single or combined action units. Facial action (action units) have been converted into emotions according to

This work has been conducted in conjunction with the "SIMILAR" European Network of Excellence on Multimodal Interfaces of the IST Programme of the European Union (www.similar.cc).

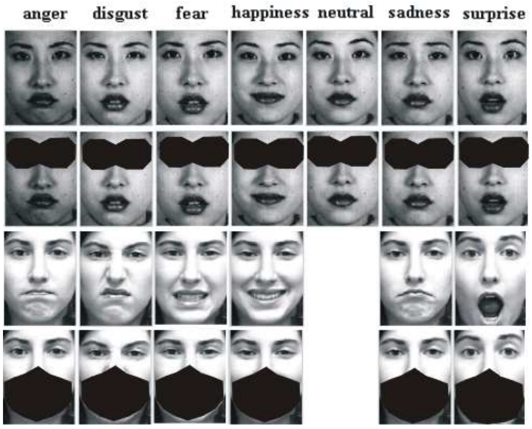

Fig. 1. An example of one expresser from JAFFE database posing 7 facial expressions having occluded eyes and one from Cohn-Kanade database posing 6 facial expressions with occluded mouth, respectively.

[8]. Thirteen persons have been chosen to create the second database. Each person expresses six basic emotions and each emotion has 3 intensities. Therefore, the total number of images in the second database is 234. We superimposed a black rectangle around the eyes and mouth regions to occlude them partially. Then each image is cropped and downsamped in a such way that the final image size is 80 $\times$ 60 pixels. Figure 1 presents one expresser from JAFFE database posing 7 facial expressions having occluded eyes and one from Cohn-Kanade database posing 6 facial expressions with occluded mouth, respectively.

3. SYSTEM DESCRIPTION

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

3.1. Feature extraction

Gabor wavelet-based method can achieve high sensitivity for facial expression classification and give the best reported results [3], [6], [9]. We applied Gabor filters (GF) to the entire face instead to specific regions avoiding the manual selection of the interesting regions to extract facial features. A 2D Gabor wavelet transform is defined as the convolution of the image $\mathcal{I}(\mathbf{z})$:

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

with a family of Gabor filters [6]:

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

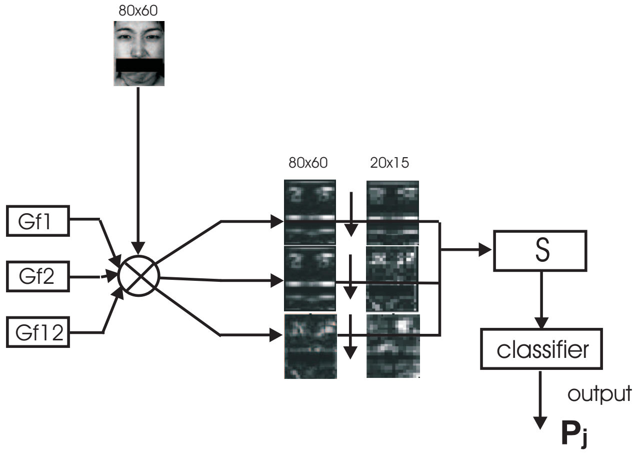

Fig. 2. Diagram block of the proposed system

where $\mathbf{z} = (x, y)$ and $\mathbf{k}$ is the characteristic wave vector, $\mathbf{k} = (k_\nu \cos\varphi_\mu \quad k_\nu \sin\varphi_\mu)^T$ with $k_\nu = 2^{-\frac{\nu+2}{2}}\pi$, and $\varphi_\mu = \mu\frac{\pi}{8}$. 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.

A feature vector is formed by convolving the 80 $\times$ 60 image $\mathbf{x}$ with 12 Gabor filters corresponding to high frequency range and orientation, downsampled to an image of 20 $\times$ 15 pixels and scanned row by row to form a vector of dimension 1 $\times$ 300 for each Gabor filter. We only took the magnitude of Gabor representation, because it varies slowly with the position, while the phases are very sensitive to it. The 12 outputs have been concatenated to form a new longer feature vector $\mathbf{s}$ of dimension 1 $\times$ 3600. The same steps have been performed 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, $\mathcal{L}$ = {*an, di, fe, ha, ne, sa, su)*}. 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 $\#\{p(\mathbf{s}_{test}) = l(\mathbf{s}_{test})\}$. Once we have formed 7 classes of new feature vectors (or prototype samples) two types of classifiers are employed to classify a new test sample, namely, *cosine similarity measure* (CSM) and *maximum correlation classifier* (MCC) [10].

4. EXPERIMENTAL RESULTS AND DISCUSSION

Since the database is of limited size, the classifier accuracy is measured using a leave-one-out strategy which makes maximal use of the available data and the results are averaged. Along with experiments including partial occlusion of the face we performed tests with no occlusion of any facial region to serve as baseline. Table 1 shows the experimental results when face images are not occluded to serve as a baseline. We should notice that the results obtained here differs from those of [6] due to the different feature extraction and classification procedure involved. In [6] the feature vector is formed by taking the information from 36 fiducial points manually located on the face plane, while in our work the whole image contributes to representation of the corresponding feature vector. Table 2 shows the experimental results with mouth and eyes occluded for both classifiers and Gabor filters with low and high frequency range.

Table 1. Classifier accuracy in percentage (%) with non occluded facial images for Jaffe and C-K (Cohn Kanade) databases.

Database	Classifier	No occlusion	
		lfr	hfr
JAFFE	CSM	88.8	81.7
	MCC	89.7	82.6
C-K	CSM	94.5	90.6
	MCC	93.6	90.2

Table 2. Classifier accuracy in percentage (%) with no occlusion, mouth and eyes region occluded for both classifiers and Gabor filters with low and high frequency range for Jaffe and C-K (Cohn Kanade) databases.

Database	Classifier	Mouth occluded		Eyes occluded	
		lfr	hfr	lfr	hfr
JAFFE	CSM	83.5	80	83	77.2
	MCC	83.5	80.8	84	78.2
C-K	CSM	86.4	83.3	91.5	88.5
	MCC	87.2	83	92.3	88.9

The highest accuracy for occluded face regions experiments has been achieved by using Gabor filters with low frequency range. The same conclusion holds when the experiments have been conducted with non occluded facial images. Overall, for JAFFE database, the accuracy is almost the same regardless of the occluded region. For the Cohn-Kanade database, it turns out that the eyes region is not as important in terms of expression discrimination as the mouth region since, in the latter case, the classifier accuracy decreases more than when the eyes are occluded.

Table 3. Confusion matrix when the eyes region is occluded for MCC and low frequency range for JAFFE database

	an	di	fe	ha	ne	sa	su
an	26	3	0	0	0	1	0
di	1	25	2	0	0	1	0
fe	0	2	28	1	0	0	1
ha	0	0	1	26	2	2	0
ne	0	0	0	1	24	5	0
sa	1	1	2	1	4	22	0
su	0	0	0	1	1	1	27

Due to the fact that the results presented in the Table 2 are average over all facial expressions and do not provide us any information with respect to a particular expression, the confusion matrix has been computed to handle with this problem.

Table 4. Confusion matrix when the mouth region is occluded for MCC and low frequency range for JAFFE database

	an	di	fe	ha	ne	sa	su
an	28	2	0	0	0	0	0
di	0	27	0	1	0	1	0
fe	0	0	25	2	0	3	2
ha	0	0	0	22	6	2	1
ne	0	0	1	2	24	3	0
sa	0	2	1	2	2	24	0
su	0	0	1	1	1	0	27

Table 5. Confusion matrix when the eyes region is occluded for MCC and low frequency range for Cohn-Kanade database

	an	di	fe	ha	sa	su
an	35	1	1	1	1	0
di	1	38	0	0	0	0
fe	1	0	37	1	0	0
ha	1	1	2	35	0	0
sa	3	0	0	0	36	0
su	0	3	1	0	0	35

Confusion matrices for JAFFE database are presented in the Tables 3 and 4 when MCC classifier is involved for low frequency range. As can be seen from the Tables 3 and 4, for JAFFE database, 'sadness' has been misclassified 5 and 3 times as 'neutral' when the eyes and mouth region is occluded, respectively. This result shows that the eyes region plays a major role when 'sadness' is involved. The eyes and mouth region have approximately the same discriminative information for the 'disgust' expression, as we have 3 and 2 misclassifications as being 'anger' when eyes and mouth region is occluded, respectively. 'Neutral' has been misclassified 6 times as 'happiness' in the absence of mouth compared with only 2 misclassifications when eyes are occluded, since the most discriminative features for this expression are mainly conveyed by the mouth, as we have expected. The mouth region also provides important discriminative information when it comes to 'fear' and 'neutral' that has been wrongly classified as 'sadness' (3 times as 'fear' and 3 times as 'neutral').

Table 6. Confusion matrix when the mouth region is occluded for MCC and low frequency range for Cohn-Kanade database

	an	di	fe	ha	sa	su
an	32	1	4	1	1	0
di	1	37	0	1	0	0
fe	1	0	34	2	2	0
ha	1	0	3	34	1	0
sa	2	9	4	0	33	0
su	0	1	2	0	2	34

The confusion matrices for Cohn-Kanade database are presented in the Tables 5 and 6. 'Anger' is misinterpreted 3 times as 'sadness' in the absence of eyes, while the most problematic expressions when the mouth was occluded have been 'anger' and 'sadness' being misclassified 4 times as 'fear'.

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 most discriminative information for facial expression classification task. The feature vectors that have been extracted from the original images by convolution with a set of Gabor filters are classified by using the maximum correlation classifier and cosine similarity measure. We compared the classification results when the eyes and mouth regions are occluded with no occlusion data and we found that, overall, the system is robust against partial facial occlusion. Moreover, each expression has been analyzed by performing the confusion matrix. The results shows that 'anger', and 'fear' are emotions that mouth region has more discriminative power than the eyes region, while the opposite happens for 'happiness'. An ambiguous expression is the 'neutral' one that has been confused between 'fear' and 'sadness' when mouth is occluded.

6. REFERENCES

[1] P. Ekman and W.V Friesen, "Manual for the facial action coding system," *Consulting Psychologists Press*, 1977.

[2] 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, 2000.

[3] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, October 1999.

[4] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 1, no. 30, pp. 259–275, 2003.

[5] F. Bourel, C. C. Chibelushi, and A. A. Low, "Recognition of facial expressions in the presence of occlusion," in *Proc. of the Twelfth British Machine Vision Conference*, vol. 1, pp. 213–222, 2001.

[6] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. Third IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 200–205, April 1998.

[7] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 46–53, March 2000.

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

[9] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, December 1999.

[10] I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor representations for facial expression recognition," in *Proc. IEEE Int. Conf. On Image Processing*, pp. 1054–1057, 2003.