ICA AND GABOR REPRESENTATION FOR FACIAL EXPRESSION RECOGNITION

I. Buciu C. Kotropoulos and I. Pitas

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

ABSTRACT

Two hybrid systems for classifying seven categories of human facial expression are proposed. The £rst system combines independent component analysis (ICA) and support vector machines (SVMs). The original face image database is decomposed into linear combinations of several basis images, where the corresponding coef£cients of these combinations are fed up into SVMs instead of an original feature vector comprised of grayscale image pixel values. The classi£cation accuracy of this system is compared against that of baseline techniques that combine ICA with either two-class cosine similarity classi£ers or two-class maximum correlation classi£ers, when we classify facial expressions into these seven classes. We found that, ICA decomposition combined with SVMs outperforms the aforementioned baseline classi£ers. The second system proposed operates in two steps: £rst, a set of Gabor wavelets (GWs) is applied to the original face image database and, second, the new features obtained are classi£ed by using either SVMs or cosine similarity classi£ers or maximum correlation classi£er. The best facial expression recognition rate is achieved when Gabor wavelets are combined with SVMs.

1. INTRODUCTION

From the perspective of psychology and anthropology, the evolution of human brain is re¤ected by the human language that represents a major milestone in the human evolution. The language alone does not seem to be suf£cient for successful social (human-to-human) interaction. Therefore, the evolution of a nonverbal signaling system, such as the facial expression mechanism has captured an increased attention in psychology and anthropology for a better understanding of the social context [1]. Unlike the human-to-human interaction that takes into consideration the facial expression, human computer interaction systems that use facial expression analysis have been introduced only recently [2], [3], [4] and [5]. Reliable facial expression modeling and, particularly, human emotions recognition, are challenging tasks since there is no pure emotion. A particular emotion rather is a complex combination of several facial expressions. Moreover, the emotions can vary in intensity, which makes emotion recognition even more dif£cult.

Basically, there are two types of approaches to cope with **facial expression recognition**: appearance-based methods and geometric feature-based methods. For the £rst method the £ducial points of the face are selected either manually [6] or automatically [7]. The face images are convolved with Gabor £lters and the responses extracted from the face images at £ducial points form vectors that are further used for classi£cation. Alternatively, the Gabor £lters can be applied to the entire face image instead to speci£c face regions. Regarding the geometric feature-based methods, the positions of a set of £ducial points in a face form a feature vector that represents the face geometry. Although the appearance-based methods (especially Gabor wavelets) seem to yield a reasonable recognition rate, the highest recognition rate is obtained, when these two main approaches are combined [6], [8]. Several other techniques for the recognition of 6 single upper action units (AUs) and 6 lower face AUs are described in [4]. These works suggest that the GW-based method can achieve high recognition accuracy for facial expression classi£cation. Similar results are obtained by using ICA.

The paper is organized as follows. After a brief image database description in Section 2, the independent component analysis is described in Section 3. In Section 4, we present the Gabor image representation. ICA/SVMs and GWs/SVMs system, training and testing procedures and the experiments conducted are described in Section 5. Conclusions are drawn in Section 6.

2. DATA SET DESCRIPTION

The database we used in our experiments contains N = 213 images of Japanese female facial expression (JAFFE) [9]. Ten expressers posed 3 or 4 examples for each of the 7 basic facial expressions (anger, disgust, fear, happiness, neutral, sadness, surprise) yielding a total of 213 images of facial expressions.

Each original image has been manually cropped and aligned with respect to the upper left corner. The cropped image size is 160 $\times$ 120. Each cropped image has been downsampled by a factor of 2 yielding an image of 80 $\times$ 60 pixels.

3. INDEPENDENT COMPONENT ANALYSIS

The goal of independent component analysis (ICA) is to decompose a set of observations into a basis whose components are statistically independent or, at least, are as independent as possible. Prior to the application of ICA on the 80 $\times$ 60 face images, each image has been scanned rowwise in order to form a column feature vector of size M = 4800. All the feature vectors have been collected in a matrix $\mathbf{X}$ whose columns contain the images. Hence, the matrix $\mathbf{X}$ is $M \times N$. Let us now suppose that each image in the database (columns of $\mathbf{X}$) represents a linear combination of some underlying basis images. In the matrix form we can write $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ are the basis images associated with a set of independent coef£cients vector (source) of $\mathbf{S}$ [10]. All we want

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

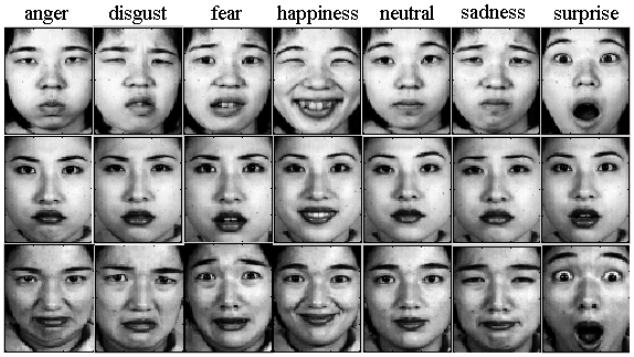

Fig. 1. Examples by three different expresser poses 7 facial expressions (emotions). The original images have been downsampled by a factor of 2, cropped and aligned with respect to the upper left corner.

to do is to estimate $\mathbf{A}$ by $\mathbf{D}^{-1}$, where the unmixing matrix $\mathbf{D}$ is the learned ICA weight matrix, such that $\mathbf{C}_{train} = \mathbf{DX}$ and $\mathbf{C}_{train} \approx \mathbf{S}$. Therefore, each column of $\mathbf{C}_{train}$ consists of the independent coef£cients, $\mathbf{c}_{train}$, for the linear combination of basis images in $\mathbf{A}$ that comprised each face image $\mathbf{x}$. Since ICA attempts to make $\mathbf{C}_{train}$ as independent as possible, $\mathbf{C}_{train}$ is called a factorial code for face images [10]. Note that, under this con£guration the pixels are independent across the same image. That is, the coef£cients $\mathbf{c}$ found in the columns of $\mathbf{C}_{train}$ are independent and not the basis images.

In order to have a controlled reduction of the number of independent components, we choose m linear combinations of the original images, namely the £rst m PCA coef£cients of the images. ICA was performed on the matrix $\mathbf{R}_m^T$ of dimension $m \times N$ whose columns are the PCA coef£cients of the training images. Let $\mathbf{P}_m^T$ be the modal matrix of dimension $m \times M$ where rows are the m principal eigenvectors. Matrix $\mathbf{R}_m^T$ is then given by $\mathbf{R}_m^T = \mathbf{P}_m^T\mathbf{X}$. Hence:

$$\mathbf{C}_{train} = \mathbf{DR}_m^T. \tag{1}$$

Subsequently, a *whitening* process is applied to $\mathbf{R}_m^T$ to normalize the data. If the row means are substracted from $\mathbf{R}_m^T$ and the resulting matrix is passed thorough a zero-phase whitening £lter which is twice the inverse square root, the whitening transformation is written as $\mathbf{W} = 2(\frac{1}{N}\mathbf{R}_m^T\mathbf{R}_m)^{-\frac{1}{2}}$. Therefore, the zero-mean input matrix can be decomposed as the product of the unmixing matrix and the whitening matrix $\mathbf{D}_w = \mathbf{DW}$. Accordingly, (1) is rewritten as $\mathbf{C}_{train} = \mathbf{D}_w\mathbf{R}_m^T$. The unmixing matrix $\mathbf{D}_w$ must be learned by ICA during training. An iterative process for updating $\mathbf{D}_w$ yields the independent coef£cients. Different approaches exist for this purpose. In our experiments we used the maximum entropy method. Let $\mathbf{c}_{train,i}$ be the i-th column vector of $\mathbf{C}_{train}$, $\mathbf{c}_{train,i} = (c_{i1}, c_{i2}, \ldots, c_{ij}, \ldots, c_{im})^T$, and $g(\xi) = 1/(1+e^{-\xi})$ be a nonlinearity applied component wise to the elements of $\mathbf{c}_{train,i}$ to yield the vector $\mathbf{z}_i = g(\mathbf{c}_{train,i})$. An updating equation for $\mathbf{D}_w$ based on $\mathbf{c}_{train,i}$ at each iteration k, is given by [11]:

$$\mathbf{D}_w(k+1) = \mathbf{D}_w(k) + \eta[\mathbf{I} + (\mathbf{1} - 2\mathbf{z}_i(k))\mathbf{c}_{train,i}^T(k)]\mathbf{D}_w(k) \tag{2}$$

where η is the learning rate, $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a $m \times 1$ of ones. We used $k = 1000$ and η was set to 10^{-6}. Obviously, $\mathbf{c}_{train,i}(k) = \mathbf{D}_w(k)\mathbf{r}_{m,i}$ where $\mathbf{r}_{m,i}$ is the i-th column of $\mathbf{R}_m^T$ and $\mathbf{z}_i(k) = g(c_{train,i}(k))$. Once we have £nished training and obtained $\mathbf{C}_{train}$, a test image $\mathbf{x}_{test}$ can be represented as $\mathbf{c}_{test} = \mathbf{D}_w\mathbf{P}_m^T\mathbf{x}_{test}$, where $\mathbf{x}_{test}$ has zero mean.

4. GABOR WAVELETS

2D Gabor wavelets have been widely used for computer vision applications and modeling biological vision, since recent studies have shown that Gabor elementary functions are suitable for modeling simple cells in visual cortex [12]. A 2D Gabor wavelet transform is de£ned as the convolution of the image $\mathcal{I}(\mathbf{z})$:

$$\mathcal{J}_{\mathbf{k}}(\mathbf{z}) = \int\int \mathcal{I}(\mathbf{z}')\psi_{\mathbf{k}}(\mathbf{z}-\mathbf{z}')d\mathbf{z}' \tag{3}$$

with a family of Gabor £lters [7]:

$$\psi_{\mathbf{k}}(\mathbf{z}) = \frac{\mathbf{k}^T\mathbf{k}}{\sigma^2}\exp(-\frac{\mathbf{k}^T\mathbf{k}}{2\sigma^2}\mathbf{z}^T\mathbf{z})\Big(\exp(i\mathbf{k}^T\mathbf{z}) - \exp(-\frac{\sigma^2}{2})\Big), \tag{4}$$

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

$$\mathbf{k} = \begin{pmatrix} k_\nu \cos\varphi_\mu & k_\nu \sin\varphi_\mu \end{pmatrix}^T \tag{5}$$

with

$$k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \;\; \varphi_\mu = \mu\frac{\pi}{8}, \;\; \nu = 0,1,2,3,4, \;\; \mu = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}. \tag{6}$$

The parameters ν and μ de£ne the frequency and orientation of the £lter. 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.

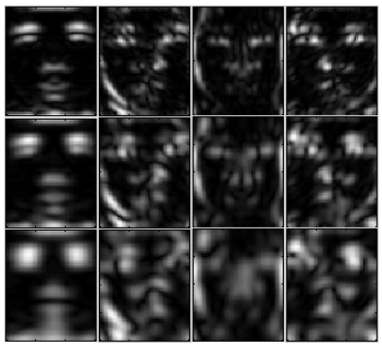

Fig. 2. Magnitude of Gabor representation of the £rst image from Figure 1 convolved with 12 Gabor £lters for ν = 0,1,2 and μ = 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$.

Each 80 × 60 image has been convolved with 12 Gabor £lters corresponding to high frequency range and orientation, downsampled by 3 resulting an image of 20 × 15 pixels and scanned row by row to form a vector 1 × 300 for each Gabor £lter. The 12 outputs have been concatenated to form a new longer feature vector of dimension 1 × 3600. Hence the £nal matrix $\mathbf{X}$ is 3600 × N, where the feature vectors have been stored in the columns of $\mathbf{X}$. The same procedure has been applied for the low frequency range.

We took only the magnitude of Gabor representation, because it varies slowly with the position, while the phase is very sensitive to it. The magnitudes for one sample image are shown in Figure 2.

5. TRAINING AND TESTING

The the 7 basic facial expressions (i.e. *anger, disgust, fear happiness, neutral, sadness, surprise*) form the set $\mathcal{L} = \bigcup_{j=1}^{7} \mathcal{L}_j$, where $\mathcal{L}_j$ is a class that corresponds to the facial images having one particular expression. Let us enumerate the expressions by the number $j = 1, 2, \ldots, 7$. We used "one-vs-all" multiclass strategy, i.e., we wanted to classify a test image in class $\mathcal{L}_j$ against images that belong to its complement $\mathcal{L}_j^C = \mathcal{L} - \mathcal{L}_j$. Let $l_j = 1$ and $l_j^C = -1$ the labels for each two-class problem. In a such classi£cation problem, we construct a classi£er where the output, predicted value of the classi£er $p_i \in \{-1, +1\}$. The recognition rate is de£ned as $RR = \#\{p(\mathbf{c}_{test}) = l(\mathbf{c}_{test})\}$, where $l(\mathbf{c}_{test})$ is the ground truth for $\mathbf{c}_{test}$. Once $\mathbf{C}_{train}$ is trained in the case of ICA and $\mathbf{X}$ containing the Gabor representation is formed, three classi£ers are employed to classify a test image:

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 $\mathbf{c}_j$ (in the case of ICA) be a column vector of $\mathbf{C}_{train}$ that corresponds to the nearest class $\mathcal{L}_j$. Let also $\mathbf{c}_j^C$ be the nearest $\mathcal{L}_j^C$ class neighbor column vector for a test coef£cient vector $\mathbf{c}_{test}$. We compute the quantities:

$$d_j = \frac{\mathbf{c}_{test}^T \mathbf{c}_j}{\|\mathbf{c}_{test}\| \|\mathbf{c}_j\|} \quad \text{and} \quad d_j^C = \frac{\mathbf{c}_{test}^T \mathbf{c}_j^C}{\|\mathbf{c}_{test}\| \|\mathbf{c}_j^C\|}, \tag{7}$$

where d_j and d_j^C are the cosines of the angle between a test feature vector and the nearest training one. We assign $\mathbf{c}_{test}$ to $\mathcal{L}_j$, if $d_j > d_j^C$. Otherwise $\mathbf{c}_{test} \in \mathcal{L}_j^C$.

2. *Maximum correlation classi£er* (MCC). The second classi£er is the minimum Euclidean distance classi£er. The Euclidean distance from $\mathbf{c}_{test}$ to $\mathbf{c}_j$ is expressed as

$$\begin{aligned} \|\mathbf{c}_{test} - \mathbf{c}_j\|^2 &= -2[\mathbf{c}_j^T \mathbf{c}_{test} - \frac{1}{2}\|\mathbf{c}_j\|^2] + \mathbf{c}_{test}^T \mathbf{c}_{test} \\ &= -2h_j(\mathbf{c}_{test}) + \mathbf{c}_{test}^T \mathbf{c}_{test}, \end{aligned} \tag{8}$$

where $h_j(\mathbf{c}_{test})$ is a linear discriminant function of $\mathbf{c}_{test}$. A test image is classi£ed by this classi£er by computing two linear discriminant functions $h_j(\mathbf{c}_{test})$ and $h_j^C(\mathbf{c}_{test})$ and assigning $\mathbf{c}_{test}$ to the class corresponding to the maximum discriminant function value.

3. *Support vector machines* (SVMs). If N_T denotes the number of training images, then the class membership for a new test vector $\mathbf{c}_{test}$ is given by the sign of the following decision function [13]:

$$f(\mathbf{c}_{test}) = \sum_{i=1}^{N_T} \alpha_i \; l_i \; K(\mathbf{c}_{test}, \mathbf{c}_i) + b \tag{9}$$

where $K(\mathbf{z}_1, \mathbf{z}_2)$ is a kernel function that de£nes the dot product between $\Phi(\mathbf{z}_1)$ and $\Phi(\mathbf{z}_2)$ in an Hilbert space $\mathcal{H}$, and α_i are nonnegative Lagrange multipliers associated with the quadratic optimization problem:

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + E\textstyle\sum_{i=1}^{N_T} \xi_i \\ &\text{subject to} && l_i(\mathbf{w}^T\Phi(\mathbf{c}_i) + b) \geq 1 - \xi_i, i = 1, \ldots, N_T. \end{aligned} \tag{10}$$

In (10), $\mathbf{w}$ and b are the parameters of the optimal hyperplane in $\mathcal{H}$ that attempts to separate the classes. That is, $\mathbf{w}$ is the normal vector to the hyperplane, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin, with $\|\mathbf{w}\|$ denoting the Euclidian norm of vector $\mathbf{w}$. E is a parameter which penalizes the errors and ξ_i are positive slack variables. Frequently used kernel functions are the polynomial kernel, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + n)^q$ and the Exponential Radial Basis Function (ERBF) kernel, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\gamma|\mathbf{x}_i - \mathbf{x}_j|\}$. We used $q = 1$ (equivalent to a linear classi£er), $q = 2, 3, 4$ and $\gamma = 0.005$ in our experiments.

Since the database is limited, the generalization error (recognition rate) is measured over identity using a leave-one-out strategy which makes maximal use of the available data for training. The results were averaged over the subjects and classes. Along with the systems described in the paper we carried out experiments with SVMs when the input is represented by the original image data for comparison purposes. The best and worst recognition rates (BRR, WRR) with respect to expressions, together with the corresponding index j to the facial expression, and the average recognition rates (ARR) over the expressers and expressions for different methods are ranked with respect to ARR from top to bottom in Table 1.

Table 1. Best recognition rate (BRR), worst recognition rate (WRR) with respect to expressions , the corresponding classes and average recognition rate (ARR) (%) over the expressers and expressions for several methods employed.

Methods	BRR	j	WRR	j	ARR
GWs (hfr) + SVMs (q = 2)	94.70	2	87.18	6	90.34
GWs (hf) + SVMs (q = 3)	94.72	2	87.77	6	90.08
GWs (hf) + SVMs (q = 1)	94.73	2	87.27	6	90.00
GWs (hf) + ERBF SVMs	92.49	2	85.86	5	89.53
GWs (lf) + ERBF SVMs	92.92	4	86.30	5	89.52
GWs (lf) + SVMs (q = 3)	93.79	2	84.52	5	89.44
GWs (hf) + SVMs (q = 2)	95.16	2	83.09	5	89.07
ERBF SVMs	92.47	4	85.84	5	89.07
GWs (lf) + SVMs (q = 1)	95.18	2	84.43	5	88.80
GWs (hf) + SVMs (q = 4)	90.63	4	84.79	5	88.34
SVMs (q = 2)	93.94	4	82.94	5	88.32
SVMs (q = 1)	90.52	2	83.46	5	87.86
SVMs (q = 4)	91.99	7	79.13	1	86.30
ICA + ERBF SVMs	86.39	2	84.54	3	85.58
GWs (hf) + CSM	90.13	2	79.15	5	84.07
GWs (lf) + CSM	88.64	2	77.41	6	83.70
GWs (hf) + MCC	88.74	2	78.27	5	83.54
GWs (lf) + MCC	88.19	2	78.10	5	83.37
ICA + SVMs (q = 4)	85.41	5	72.64	7	79.91
ICA + CSM	81.22	4	72.25	5	76.43
ICA + MCC	84.13	5	69.09	3	75.61
ICA + SVMs (q = 2)	80.71	5	68.57	7	75.75
Method described in [5]					75.00
ICA + SVMs (q = 3)	76.27	5	59.09	2	63.86

The best recognition rate (95.18 %) for a novel expresser is obtained when the feature vectors are obtained by Gabor representation with low frequency range and the classi£cation is done

with linear ($q = 1$) SVMs. On the opposite side, the polynomial SVM ($q = 3$) combined with ICA decomposition gives the worst recognition rate with respect to expressions. It turns out from Table 1 that the Gabor representation provides more discriminability power than the original feature vector (image pixel values) and ICA decomposition. Overall, the best average recognition rate over the expressers and expressions was achieved by high frequency Gabor representation with a polynomial SVM of degree 2. Although ICA combined with SVMs outperforms MCC and CMS, the performance of this system is inferior to GWs/SVMs facial expression recognition system. However, both of them outperformed the method of Lyons and all [5]. The most dif£cult expression to be recognized was the "neutral", while the easiest was "disgust". In order to verify this result and gain 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 [14]. 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. The results are shown in Table 2. The results presented in Table 2 are consistent with the results shown in Table 1 since the smallest J are J_5 and J_6 that correspond to the worst recognition rates for j equals 5 and 6 ("neutral" and "sadness"), whilst J_2 corresponding to the index 2 ("disgust") has a large value. "Fear" has been reported in [6] as being the emotion that is most dif£cult to be recognized. Indeed, the corresponding J_3 has a small value supporting this claim. As one can see, the Gabor representation produces a better class separation than the original feature vector or ICA decomposition.

Table 2. Class separability measurement when data are represented by the: 1. original image grayscale pixel values; 2. the GW(hfr) representation; 3. the GW(lfr) representation 4. ICA decomposition.

Case	J_1	J_2	J_3	J_4	J_5	J_6	J_7
1	1.025	1.017	1.016	1.018	1.013	1.011	1.032
2	1.032	1.034	1.016	1.031	1.015	1.013	1.067
3	1.035	1.027	1.016	1.034	1.015	1.015	1.051
4	1.004	1.003	1.014	1.013	1.003	1.005	1.008

6. CONCLUSIONS

The experiments carried out on several methods for facial expression recognition demonstrate that the GWs/SVMs - based system yields the best average recognition rate.

Aknowledgement

The authors thank Michael J. Lyons from ATR Human Information Processing Research Labs for providing them the JAFFE facial expression database.

7. REFERENCES

[1] K. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *Yearbook of Physical Anthropology*, , no. 44, pp. 3–24, 2001.

[2] J.-J. J. Lien, J. F. Cohn, T. Kanade, and C. C. Li, "Subtly different facial expression recognition and expression intensity estimation," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, pp. 853–859, June 1998.

[3] 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.

[4] 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.

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

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

[7] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, December 1997.

[8] Y.-Li Tian, T. Kanade, and J. Cohn, "Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," in *Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 229–234, May 2002.

[9] 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.

[10] M. S. Bartlett, H. M. Lades, and T. K. Sejnowsky, "Independent component representations for face recognition," in *Proc. SPIE Conf. Human Vision and Electronic Imaging III*, vol. 3299, pp. 528–539, 1998.

[11] A. J. Bell and T. J. Sejnowski, "An information - maximization approach to blind separation and blind deconvolution," *Neural Computing*, vol. 6, no. 10, pp. 1129–1159, 1995.

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

[13] V. N. Vapnik, *Statistical Learning Theory*, J. Wiley, N.Y., 1998.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.