A NEW SPARSE IMAGE REPRESENTATION ALGORITHM APPLIED TO FACIAL EXPRESSION RECOGNITION

Ioan Buciu and Ioannis Pitas
Department of Informatics
Aristotle University of Thessaloniki, GR-541 24 Thessaloniki, Greece
Phone: +30-231-099-6361
Fax: +30-231-099-8453
E-mail: nelu,pitas@zeus.csd.auth.gr
Web: http://poseidon.csd.auth.gr

Abstract. In this paper we present a novel algorithm for learning facial expressions in a supervised manner. This algorithm is derived from the local non-negative matrix factorization (LNMF) algorithm, which is an extension of non-negative matrix factorization (NMF) method. We call this newly proposed algorithm Discriminant Non-negative Matrix Factorization (DNMF). Given an image database, all these three algorithms decompose the database into basis images and their corresponding coefficients. This decomposition is computed differently for each method. The decomposition results are applied on facial images for the recognition of the six basic facial expressions. We found that our algorithm shows superior performance by achieving a higher recognition rate, when compared to NMF and LNMF.

INTRODUCTION

In recent years it is argued, from a visual neuroscience viewpoint, that the architecture of the visual cortex suggests a hierarchical organization, in which neurons become selective to progressively more complex aspects of image structure. The type of image encoding in the human visual system is related to the number of neurons that are active (respond) to a certain information represented by a specific sensory stimulus caused by the image. We refer to a *local* code when only a single individual specific cell is activated. We have a *dense* code when a large cell population with overlapping sensory input is activated and contributes to image representation. In between local and dense codes, we have the *sparse* codes, where only a fraction of a large neuronal population is active. It is a compromise between dense and local codes, combining their advantages and trying to eliminate their drawbacks.

Atick and Redlich [1] support and argue for compact, dense decorrelated codes for image representation. They have demonstrated that receptive fields of retinal ganglion cells can be viewed as local "whitening" filters that remove second-order correlations between image pixels. Bandpass, multi-scale and oriented receptive fields of V1 neurons may also be considered as filters that remove second-order correlation, such as Principal Component Analysis (PCA) does.

Numerous evidence about sparse image coding were brought by other researchers. They argue for a sparse representation that leads to an "efficient coding" in the visual cortex [14]. Since spatial receptive fields of simple cells (including V1 neurons) have been reasonably well described physiologically as being localized, oriented and bandpass, Olshausen and Field [20] bring evidence that an efficient image coding can be produced by considering an approach where the image is described by a small number of descriptors. This descriptors can be found by applying principles such as entropy minimization [2], which is equivalent to minimizing the mutual information in a such a way that the higher-order correlation between images is removed. Biederman came up with the theory of recognition-by-components (RBC) [5]. Empirical tests support his idea that the complex objects are segmented in components called 'geons' that are further used by humans for image understanding. There are also studies showing that Gabor elementary functions, which are a sparse image representation, are suitable for modeling simple cells in visual cortex [18].

Form engineering viewpoint, both dense and sparse (local) codes have been used by computer scientists in the attempt to analyze and represent the human face for face recognition, identification or facial expressions analysis. Representations of human face based on principal components give us a dense code and the post-processed images have holistic ("ghost" - like) appearance. PCA has been successfully applied to recognize facial identity in [9], [4] and [24], and facial expressions [10], [21] and [7]. Other researchers state that local or parts-based human face representation performs better than holistic representations. Bartlett et al. [3] used Independent Component Analysis (ICA) to represent faces for recognition. They employed an ICA configuration in such a way that the derived image features represent local features and they found that ICA outperforms PCA. ICA looks for components that are as independent as possible and produces such features whose properties are related to the ones of V1 receptive fields, e.g. orientation selectivity, bandpass nature and scaling ability.

The facial expression recognition methods can be classified in two categories: appearance-based methods and geometric feature-based methods. One of the most successful techniques that belongs to the first category is the one that implies a convolution of each image with Gabor filters, whose responses, extracted from the face images at fiducial points, form vectors that are further used for classification. Regarding the geometric feature-based methods, the positions of a set of fiducial points in a face form a feature vector that represents facial 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 [26], [23]. Several holistic and local representation methods have been studied and applied to classify facial actions by Donato et. al [12]. They have shown that the extraction of local features from the entire face space by convolving each image with a set of Gabor filters having different frequencies and orientations can outperform other methods that invoke the holistic representation of the face, when it comes to classify facial actions. They achieved the best recognition results by using ICA and Gabor filters. Regarding Gabor filters, they have been applied successfully not only to classify facial actions but to face recognition as well [25]. A survey on automatic facial expression analysis can be found in [13].

Within the local image representation framework, another two methods have been proposed recently for learning object parts. Lee and Seung [16] proposed an unsupervised learning technique, the so called Non-negative Matrix Factorization (NMF) which allows objects to be reassembled using purely additive combinations of the learned parts. Li et al. [19] have extended this technique by imposing additional constraints and developing a variant of NMF, named Local Non-negative Matrix Factorization (LNMF). Both methods have been applied for face representation and recognition. Li et al. found that, while NMF representation yields low recognition accuracy (actually lower than the one that can be obtained by using the PCA method), LNMF leads to a better classification performance. Chen et al. [8] successfully applied LNMF for face detection. LNMF has also been found to give higher facial expression recognition rate than NMF, when applied to recognize facial expressions [6].

In this paper, we further extend LNMF technique in order to enhance its performance regarding the recognition of the six basic expressions. All the previously mentioned representation methods (except FLD) are unsupervised. On the contrary, we propose here a novel supervised technique called Discriminant Non-negative Matrix Factorization (DNMF) that takes into account facial expression class information, which is not used in NMF and LNMF methods. Our technique is proven to perform better than the latter two methods by achieving a higher facial expression recognition rate.

NMF, LNMF, AND DNMF

Non-negative matrix factorization (NMF) has been proposed by Lee and Seung as a method that decomposes a given $m \times n$ non-negative matrix $\mathbf{X}$ into non-negative factors $\mathbf{Z}$ and $\mathbf{H}$ such as $\mathbf{X} \approx \mathbf{ZH}$, where $\mathbf{Z}$ and $\mathbf{H}$ are matrices of size $m \times p$ and $p \times n$, respectively [16]. Suppose that $i = 1, \ldots, m$, $j = 1, \ldots, n$ and $k = 1, \ldots, p$. Then, each element x_{ij} of the matrix $\mathbf{X}$ can be written as $x_{ij} \approx \sum_k z_{ik} h_{kj}$. The quality of approximation depends on the cost function used. Two cost functions were proposed by Lee and Seung in [17]: the Euclidean distance between $\mathbf{X}$ and $\mathbf{ZH}$ and KL divergence. In this

case, KL has the following expression:

$$D_{NMF}(\mathbf{X} \,\|\, \mathbf{ZH}) \triangleq \sum_{i,j} \left(x_{ij} \ln \frac{x_{ij}}{\sum_k z_{ik} h_{kj}} + \sum_k z_{ik} h_{kj} - x_{ij} \right), \tag{1}$$

This expression can be minimized by applying multiplicative update rules subject to $\mathbf{Z}, \mathbf{H} \geq 0$. The positivity constraints arise in many real image processing applications. For example, the pixels in a grayscale image have non-negative intensities. In the NMF approach, its proposers find appropriate to impose non-negative constraints, partly motivated by the biological aspect that the firing rates of neurons are non-negative. Since both matrices $\mathbf{Z}$ and $\mathbf{H}$ are unknown, we need an algorithm which is able to find these matrices by minimizing the divergence (1). By using an auxiliary function and the Expectation Maximization (EM) algorithm [11], the following update rule for computing h_{kj} is found to minimize the KL divergence at each iteration t [17]:

$$h_{kj}^{t} = h_{kj}^{t-1} \frac{\sum_i z_{ki} \frac{x_{ij}}{\sum_k z_{ik} h_{kj}^{t-1}}}{\sum_i z_{ik}}. \tag{2}$$

By reversing the roles of $\mathbf{Z}$ and $\mathbf{H}$ in (2), a similar update rule for each element z_{ik} of $\mathbf{Z}$ is obtained:

$$z_{ik}^{t} = z_{ik}^{t-1} \frac{\sum_j \frac{x_{ij}}{\sum_k z_{ik}^{t-1} h_{kj}} h_{jk}}{\sum_j h_{kj}}. \tag{3}$$

Both updating rules are applied alternatively in an EM manner and they guarantee a nonincreasing behavior of the KL divergence.

It has been shown that, if the matrix $\mathbf{X}$ contains images from an image database one in each matrix column, then the method decomposes them into basis images (columns of $\mathbf{Z}$) and the corresponding coefficients (or hidden components) (rows of $\mathbf{H}$) [16]. The resulting basis images contain parts of the original images, parts that are learned thorough the iterative process in the attempt of approximating $\mathbf{X}$ by the product $\mathbf{ZH}$. In this context, m represents the number of pixels in the image, n is the total number of images and p is the number of the subspaces in which basis images lay.

Local non-negative matrix factorization (LNMF) has been developed by Li et al [19]. This technique is a version of NMF which imposes more constraints on the cost function that are related to spatial localization. Therefore, the localization of the learned image features is improved. If we use the notations $[\mathbf{u}_{ij}] = \mathbf{U} = \mathbf{Z}^T\mathbf{Z}$ and $[\mathbf{v}_{ij}] = \mathbf{V} = \mathbf{H}\mathbf{H}^T$, the new function has to be minimized subject to three additional issues: 1) $min \sum_j \mathbf{u}_{jj}$, 2)$min \sum_{j \neq k} \mathbf{u}_{jk}$ and 3)$max \sum_j \mathbf{v}_{jj}$.

Therefore, the new cost function takes the form of the following divergence:

$$\begin{aligned} D_{LNMF}(\mathbf{X}||\mathbf{ZH}) \triangleq\ & D_{NMF}(\mathbf{X}||\mathbf{ZH}) \\ & + \alpha \sum_{ij} u_{ij} - \beta \sum_i v_{ii}, \end{aligned} \tag{4}$$

where α, $\beta > 0$ are constants. A solution for the minimization of relation (4) can be found in [19]. Accordingly, if we use the following update rules for image basis and coefficients:

$$h_{kj}^t = \sqrt{h_{kj}^{t-1} \sum_i z_{ki} \frac{x_{ij}}{\sum_k z_{ik} h_{kj}^{t-1}}}. \tag{5}$$

$$z_{ik}^t = \frac{z_{ik}^{t-1} \sum_j \frac{x_{ij}}{\sum_k z_{ik}^{t-1} h_{kj}} h_{jk}}{\sum_j h_{kj}}. \tag{6}$$

$$z_{ik}^t = \frac{z_{ik}^t}{\sum_i z_{ik}^t}, \quad \text{for all } k \tag{7}$$

the KL divergence is nonincreasing.

Let us suppose now that we have c distinctive image classes $\{\mathrm{Q}_1, \ldots, \mathrm{Q}_c\}$. Each image from the database corresponding to one column of matrix $\mathbf{X}$, belongs to one of these classes. We denote the arithmetic mean of each class Q_l by $\mu_l = \frac{1}{n_l} \sum_{r=1}^{n_l} h_r$ and the global arithmetic mean by $\mu = \frac{1}{n} \sum_{j=1}^{n} h_j$, where n_l is the cardinality of class Q_l, n is the total number of images and $l = 1, \ldots, c$. Both NMF and LNMF consider the database as a whole and treat each image in the same way. There is no class information integrated into the cost function. Here, we extend the cost function given by the LNMF technique by proposing a class-dependent approach called *Discriminant Nonnegative Matrix Factorization* (DNMF). The decomposition coefficients encode the image representation in the same way for each image. Therefore, by modifying the expression for the coefficients in a such a way that the basis images incorporate class characteristics, we obtain a class-dependent image representation. We preserve the same constraints on basis as for LNMF and we only introduce two more constraints on the coefficients:

1. $\mathbf{S}_w = \sum_l \sum_{h_r \in \mathrm{Q}_l} (h_r - \mu_l)(h_r - \mu_l)^T \longrightarrow \min$. $\mathbf{S}_w$ represents the within-class scatter matrix and defines the scatter of the class samples around their mean. The dispersion of samples that belong to the same class around their corresponding mean should be as small as possible.

2. $\mathbf{S}_b = \sum_l (\mu_l - \mu)(\mu_l - \mu)^T \longrightarrow \max$. $\mathbf{S}_b$ denotes the between-class scatter matrix and defines the scatter of the class mean around the global mean μ. Each cluster formed by the samples that belong to the same class must be as far as possible from the other clusters. Therefore, $\mathbf{S}_b$ should be as large as possible.

We modify the divergence by adding these two more constraints. The new cost function is expressed as:

$$\begin{aligned} D_{DNMF}(\mathbf{X}||\mathbf{ZH}) \quad \triangleq \quad & D_{LNMF}(\mathbf{X}||\mathbf{ZH}) + \gamma \sum_l \sum_{h_r \in \mathrm{Q}_l} (h_r - \mu_l)(h_r - \mu_l)^T - \\ & -\delta \sum_l (\mu_l - \mu)(\mu_l - \mu)^T, \end{aligned} \tag{8}$$

where γ and δ are constants. Following the same EM approach used by NMF and LNMF techniques, it can be proven that the following update expression for each element h_{kj} of the coefficient matrix $\mathbf{H}$ is obtained:

$$h^t_{kj(l)} = \frac{2\mu_l - 1 + \sqrt{(1-2\mu_l)^2 + 8\xi h^{t-1}_{kj(l)} \sum_i z_{ki} \frac{x_{ij}}{\sum_k z_{ik} h^{t-1}_{kj(l)}}}}{4\xi} \tag{9}$$

$$h^t_{kj} = [h^t_{kj(l_1)} \,|\, h^t_{kj(l_2)} \,|\, \ldots \,|\, h^t_{kj(l_c)}] \tag{10}$$

where "|" denotes concatenation and $\xi = \gamma - \beta$. The expression for updating the image basis remains unchanged from LNMF.

The method proposed here is a supervised method that preserves the sparseness characteristic of basis images through (6), while enhancing the class separability by the minimization of $\mathbf{S}_w$ and the maximization of $\mathbf{S}_b$ through (9).

FACIAL EXPRESSION RECOGNITION EXPERIMENT

We have tested our method along with NMF and LNMF approaches for recognizing the six basic facial expressions namely, anger, disgust, fear, happiness, sadness and surprise from face images. The facial images used come from Cohn-Kanade AU-coded facial expression database [15]. The facial action (action units) that are described in the image annotations have been converted into emotion class labels according to [22]. Thirteen persons have been chosen to create the image database that has been used in our experiments. Each person expresses six basic emotions and each emotion has 3 intensities. Therefore, the total number of images in the database is n = 234. Each original image was cropped to a central face image containing the main facial fiducial points (as eyebrows, eyes, nose and chin) The uniform background was eliminated. The cropped face images have been aligned with respect to their upper left corner. The cropped face images of size 80 $\times$ 60 pixels were downsampled to 40 $\times$ 30 pixels. The face image pixels were stored into a $m = 1200$ - dimensional vector for each image. These vectors form the columns of matrix $\mathbf{X}$.

In the classical facial expression classification context, the original data are split in two disjoint parts, the training and test data sets. To form the training set, 164 face images were randomly chosen from the Cohn-Kanade derived database, while the remaining images were used for testing, forming the test face image set. Out of the training images we formed the basis images corresponding to NMF, LNMF, DNMF by executing the algorithms described in this paper. The first 10 basis images learned by NMF, DNMF and LNMF for the facial expression recognition experiment are depicted in Figure 1. It can be noticed by visual inspection that the basis images retrieved by DNMF are not as sparse as those extracted by LNMF but are more sparse than the basis images found by NMF. The training procedure was applied for various

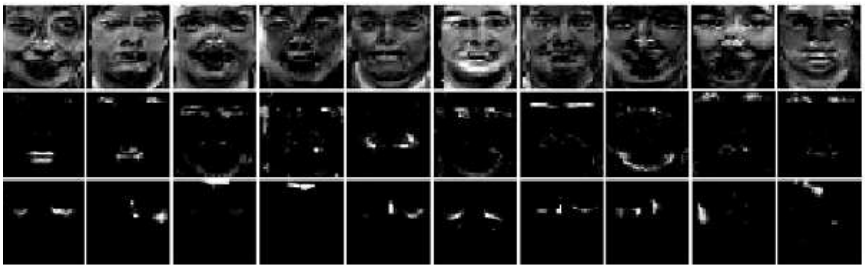

Figure 1: A set of 10 basis images learned by NMF (top), DNMF (middle) and LNMF (bottom).

numbers of basis images. The image data are then projected into the image basis in an approach similar to the one used in PCA, yielding a new feature vector $\mathbf{F} = \mathbf{Z}^T(\mathbf{X} - \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a matrix whose columns represent the average face $\psi = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$. In the test phase, for each test face image $\mathbf{x}_{test}$, a new test feature vector $\mathbf{f}_{test}$ is then formed as $\mathbf{f}_{test} = \mathbf{Z}^T(\mathbf{x}_{test} - \psi)$.

If we construct a classifier whose class label output for a test sample $\mathbf{f}_{test}$ is $\widetilde{l}$ then, the classifier accuracy is defined as the percentage of the correctly classified test images when $\{\widetilde{l}(\mathbf{f}_{test}) = l(\mathbf{f}_{test})\}$, where $l(\mathbf{f}_{test})$ is the correct class label. Once we have formed 6 classes of new feature vectors (or prototype samples), a nearest neighbor classifier is employed to classify the new test sample, by using the *Cosine Similarity Measure* (CSM). This approach is based on the nearest neighbor rule and uses as similarity the angle between a test feature vector and a prototype one. We choose $\widetilde{l} = \text{argmin}_{l=1,\ldots,c}\{d_l\}$, where $d_l = \frac{\mathbf{f}_{test}\mathbf{f}_l^T}{\|\mathbf{f}_{test}\|\|\mathbf{f}_l\|}$ and d_l is the cosine of the angle between a test feature vector $\mathbf{f}_{test}$ and the prototype one $\mathbf{f}_l$.

PERFORMANCE EVALUATION AND DISCUSSIONS

We have tested the algorithms for several number of basis images (subspaces). The results are shown in Figure 2. Unfortunately, the accuracy does not increase monotonically for none of the methods with the number of basis images. A maximum classification accuracy of 82.85 % is obtained for 81 and 100 basis images in the case of DNMF, 81.42 % is yielded by LNMF corresponding to 25 basis and 77.14 % for NMF and 36 basis. However, mean and standard deviation for accuracy averaged over the number of basis images are (%): 79.04 and 3.19 for DNMF, 76.34 and 5.53 for LNMF and 71.58 and 4.42 for NMF, respectively. Hence, a maximum average classification accuracy and a its minimum deviation around the mean is obtained by DNMF, indicating that DNMF has a better overall classification rate.

A very careful attention must be paid to the choice of the parameter ξ in (9). Due to the fact that the cost function defined by DNMF is formed by several terms that are simultaneously optimized (minimized or maximized),

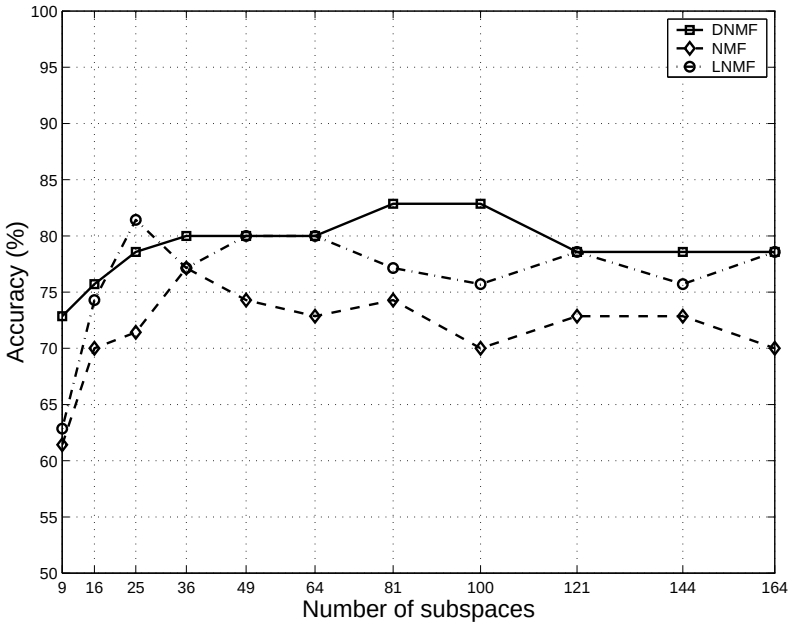

Figure 2: Accuracy achieved for DNMF, NMF and LNMF methods versus number of basis images (subspaces).

its global optimization may suffer. Although the cost function is globally minimized, each term has its own rate of convergence. The parameter ξ governs the convergence speed for minimizing $\mathbf{S}_w$ while maximizing $\mathbf{S}_b$. However, it also interferes with the expression that minimizes the approximation $\mathbf{X} \approx \mathbf{ZH}$, i.e., the term $D_{NMF}(\mathbf{X} \,\|\, \mathbf{ZH})$. An overly small value of ξ will speed up the decrease of $\mathbf{S_w}$, the increase of $\mathbf{S_b}$ and the minimization of $D_{NMF}(\mathbf{X} \,\|\, \mathbf{ZH})$). However, the algorithm may stop too early and the number of iterations might not be sufficient to reach a local minimum for $D_{DNMF}(\mathbf{X} \,\|\, \mathbf{ZH})$ and to learn corresponding sparse basis images. On the other hand, the algorithm may converge very slowly if an overly large value of ξ is chosen. Experimentally, we have chosen a value of $\xi = 0.5$ in our experiments that gave us a good trade-off between sparseness and convergence speed.

CONCLUSION

In this paper we have presented a new image representation approach that has been applied to facial expression recognition. We found that it produces a higher recognition accuracy than NMF or LNMF approaches. In the light of the sparse image coding theory, the neural interpretation of this model is that a simple cell in V1 area performs sparse coding on the visual input, hav-

ing its receptive fields closely related to the sparse coding basis images and firing rates proportional to the representation coefficients. DNMF presents a sparse structure of the basis images The basis image sparse "active" patterns (just a few pixel patterns have non-zero value) are selected by the representation coefficients that convey class information. The proposed approach is a supervised learning algorithm that keeps the original non-negative constraints on basis and coefficients borrowed from the original NMF approach, enhances the sparseness of basis images (with respect to NMF) by adding the constraints taken from LNMF approach and improves the classification accuracy by following a class discriminant approach. As far as basis image sparseness is concerned, DNMF is a good trade-off between local image representation produced by LNMF and the holistic image representation produced by NMF.

AKNOWLEDGEMENT

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

REFERENCES

[1] J. J. Atick and A. N. Redlich, "What does the retina know about the natural scene ?" **Neural Computation**, vol. 4, pp. 196–210, 1992.

[2] H. Barlow, "Unsupervised learning," **Neural Computation**, vol. 1, no. 3, pp. 295–311, 1989.

[3] M. S. Bartlett, H. M. Ladesand and T. J. Sejnowsky, "Independent component representations for face recognition," **Proc. SPIE 3299**, pp. 528–539, 1998.

[4] 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.

[5] I. Biederman, "Recognition-by-components: a theory of human image understanding," **Psychological Review**, vol. 94, no. 2, pp. 115–147, 1987.

[6] I. Buciu and I. Pitas, "Application of non-negative and local non-negative matrix factorization to facial expression recognition," **Int. Conf. on Pattern Recognition**, 2004.

[7] A. J. Calder, A. M. Burton, P. Miller, A. W. Young and S. Akamatsu, "A principal component analysis of facial expressions," **Vision Research**, vol. 41, pp. 1179–1208, 2001.

[8] X. Chen, L. Gu, S. Z. Li and H.-J. Zhang, "Learning representative local features for face detection," **Int. Conf. Computer Vision and Pattern Recognition**, pp. 1126–1131, 2001.

[9] G. Cottrell and M. Fleming, "Face recognition using unsupervised feature extraction," **Proc. of Int. Neural Network Conference**, pp. 322–325, 1990.

[10] G. Cottrell and J. Metcalfe, "Face, gender and emotion recognition using

holons," **Addvances in Neural Information Procesing Systems**, vol. 3, pp. 7564–571, 1991.

[11] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," **Royal Stat. Soc.**, vol. 39, pp. 1–38, 2001.

[12] 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.

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

[14] D. Field, "What is the goal of sensory coding ?" **Neural Computation**, vol. 6, no. 4, pp. 559–601, 1994.

[15] 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, March 2000.

[16] D. D. Lee and H. S. Seung, "Learning the parts of the objects by non-negative matrix factorization," **Nature**, vol. 401, pp. 788–791, 1999.

[17] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," **Advances Neural Information Processing Systems**, pp. 556–562, 2001.

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

[19] S. Z. Li, X. W. Hou and H. J. Zhang, "Learning spatially localized, parts-based representation," **Int. Conf. Computer Vision and Pattern Recognition**, pp. 207–212, 2001.

[20] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," **Network Computation in Neural Systems**, vol. 7, no. 2, pp. 333–339, 1996.

[21] C. Padgett and G. Cottrell, "Representing face images for emotion classification," **Addvances in Neural Information Procesing Systems**, vol. 9, pp. 894–900, 1997.

[22] 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, March 2000.

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

[24] M. Turk and A. Pentland, "Eigenfaces for recognition," **Cognitive Neuroscience**, vol. 3, no. 1, pp. 71–86, 1991.

[25] 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.

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