Application of non-negative and local non negative matrix factorization to facial expression recognition

Ioan Buciu and Ioannis Pitas
Dept. of Informatics
Aristotle University of Thessaloniki
GR-541 24, Thessaloniki, Box 451, Greece
{nelu,pitas}@zeus.csd.auth.gr

Abstract

In this paper two image representation approaches called non-negative matrix factorization (NMF) and local non-negative matrix factorization (LNMF) have been applied to two facial databases for recognizing six basic facial expressions. A principal component analysis (PCA) approach was performed as well for facial expression recognition for comparison purposes. We found that, for the first database, LNMF outperforms both PCA and NMF, while NMF produces the poorest recognition performance. Results are approximately the same for the second database, with slightly performance improvement on behalf of NMF.

1. Introduction

Is the perception of the whole image based on perception of its parts or it is viewed as an holistic process ? Despite the huge amount of psychological research done in this respect, there is no general consensus in answering these questions. Rather, the answer to the problem of how the visual cortex understands complex objects, and, in particular human faces, is a controversial one. Wachsmuth et. al [15] have drawn psychological and physiological evidence for parts-based objects representations in the brain. Biederman came up with the theory of recognition-by-components (RBC) [2]. Regarding the face recognition, which is a particular case of object recognition, Farah et al. [9] found that the features used to recognize faces are “holistic” in nature. As far as computer vision researchers are concerned, the same questions hold in attempting to create an automatic human face analyzer (facial identification/facial expression recognition). While some researchers found suitable to follow holistic approaches for facial expressions recognition [4], by employing approaches as principal components analysis (PCA) and Fisher linear discriminant (FLD), other researchers state that local or parts-based human face representation performs better than holistic representations. Several holistic and local representation methods have been studied and applied to classify facial actions units by Donato et. al [8]. Best recognition results have been reported by using ICA and Gabor filters.

Furthermore, it is still unclear whether different face processing tasks require different types of processing, as expressed in [3] and [6]. For example, there is some evidence that face recognition and facial expression recognition are two tasks that rely on different representations and processing mechanisms. This conclusion is drawn based on the dissociation between these two process found in brain damaged patients, leading to the hypothesis that there may be multiple representations of faces in the visual cortex.

Another two part-based image representation such as NMF and LNMF first introduced by Lee et. al [11] and Li et al. [12], respectively, have been already 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 better classification performance than PCA. Chen et al. [5] successfully applied LNMF for face detection. NMF and LNMF have not been applied so far for facial expression recognition.

2. PCA, NMF and LNMF

The six basic facial expressions plus the neutral pose form seven classes. Each image $\mathbf{X}$ must be assigned to a facial expression class $\mathcal{L}$ and get a label $l(\mathbf{X}) \in \{1, \ldots, 7\}$. Let us assume that the training face images are $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$, taking values in an n-dimensional feature space (i.e. $\mathbf{X} \in \Re^n$). Once the system is trained and we present it a new image $\mathbf{X}_{test}$ it should produce the correct facial expression label $l(\mathbf{X}_{test})$. Principal component analysis uses a linear dimensionality reduction transformation that maximizes the scatter of all projected images and decomposes the images in terms of basis images (eigenimages) [1]. The average face is $\mathbf{\Psi} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}_k$, and the variation of each face from the average is $\mathbf{\Phi}_k = \mathbf{X}_k - \mathbf{\Psi}$. Then, the eigenvectors $\mathbf{W} \in \Re^{n\times m}$ of the total scatter matrix $\mathbf{S}_T = \sum_{k=1}^{N}\mathbf{\Phi}_k\mathbf{\Phi}_k^T$ are a set of eigenfaces with $m < n$. Each training face image $\mathbf{X}_k$ can be projected to the eigenfaces $\mathbf{\Omega}_k = \mathbf{W}^T(\mathbf{X}_k - \mathbf{\Psi})$ forming a new feature vector $\mathbf{\Omega}_k \in \Re^m$. The original images can be reconstructed as linear combinations of the basis images $\mathbf{W}$ as $\mathbf{X}^{rec} = \mathbf{W}\mathbf{\Omega}$. The entries of $\mathbf{W}$ and $\mathbf{\Omega}$ are of arbitrary sign.

Contrary to PCA, non-negative matrix factorization (NMF) does not allow negative entries in the matrix factors $\mathbf{W}$ and $\mathbf{\Omega}$ [11]. The non-negative constraints are imposed to be consistent with the neurophysiological fact that the neural firing rate is non-negative. NMF attempts to find an approximate factorization for $\mathbf{X}^{rec} \approx \mathbf{X}$ that minimizes the divergence D between $\mathbf{X}$ and $\mathbf{X}^{rec}$ subject to $\mathbf{W}, \mathbf{\Omega} \geq 0$ and $\sum_{j=1}^{n}\mathbf{w}_{jk} = 1$, $\forall k$ [11]. The cost function D to be minimized is given explicitly by:

$$D(\mathbf{X}||\mathbf{X}^{rec}) = \sum_{j,k}\left(x_{jk}\log\frac{x_{jk}}{x_{jk}^{rec}} - x_{jk} + x_{jk}^{rec}\right) \quad (1)$$

The cost function can be minimized through an iterative process by applying an auxiliary function similar to that used in the EM algorithm [7]. When the minimum is found, the basis images $\mathbf{W}$ should contain parts-based image features.

Local non-negative matrix factorization (LNMF) has been developed by Li et al [12]. This technique is a version of NMF which imposes more constraints on the cost function related to spatial locality. Therefore, the locality of the learned features from the image is improved by imposing these additional constraints. The new cost function is now:

$$\begin{aligned} D(\mathbf{X}||\mathbf{X}^{rec}) = \sum_{j,k}\left(x_{jk}\log\frac{x_{jk}}{x_{jk}^{rec}} - x_{jk} + x_{jk}^{rec}\right) + \\ + \alpha\sum_{jk}u_{jk} - \beta\sum_{j}v_{jj} \end{aligned} \quad (2)$$

where $[\mathbf{u}_{jk}] = \mathbf{U} = \mathbf{W}^T\mathbf{W}$, $[\mathbf{v}_{jk}] = \mathbf{V} = \mathbf{\Omega}\mathbf{\Omega}^T$ and $\alpha,\ \beta > 0$ are constants. Therefore, 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)$max\sum_j \mathbf{v}_{jj}$. A solution for the minimization of relation (2) can be found in [12].

3 Facial expressions classification

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) [13]. A second database has been derived from Cohn-Kanade AU-coded facial expression database [10] and the facial action (action units) have been converted into emotions according to [14]. A number of 234 images was used. Each original image has been aligned with respect to its upper left corner. We aligned the images, since this preprocessing step is critical in achieving good classifier performance for PCA. Each original image from both databases is cropped and downsamped in a such way that the final image size is 40×30 pixels.

In the classical classification problem, we construct a classifier where the output (predicted value) of the classifier for a test sample $\mathbf{\Omega}_{test}$ is p_j. Since $\mathbf{X} = \mathbf{W}\mathbf{\Omega}$, the feature vectors used for classification are formed as $\mathbf{\Omega} = \mathbf{W}^{-1}(\mathbf{X} - \mathbf{\Psi})$. A new test feature vector $\mathbf{\Omega}_{test}$ is then formed as $\mathbf{\Omega}_{test} = \mathbf{W}^{-1}(\mathbf{X}_{test} - \mathbf{\Psi})$. The classifier accuracy is defined as the percentage of the correctly classified images when $\{l(\mathbf{\Omega}_{test}) = p(\mathbf{\Omega}_{test})\}$. Once we have formed 7 classes of new feature vectors (or prototype samples) a nearest neighbor classifier is employed to classify the new test sample, by using the following similarity measures:

1. *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 $CSM = \mathrm{argmin}_{j\in\mathcal{L}_j}\{d_j\}$, where $d_j = \frac{\mathbf{\Omega}_{test}\mathbf{\Omega}_j^T}{\|\mathbf{\Omega}_{test}\|\|\mathbf{\Omega}_j\|}$ and d_j is the cosine of the angle between a test feature vector $\mathbf{\Omega}_{test}$ and the prototype one $\mathbf{\Omega}_j$.

2. *Maximum correlation classifier* (MCC). The second classifier is a minimum Euclidean distance classifier. The Euclidean distance from $\mathbf{\Omega}_{test}$ to $\mathbf{\Omega}_j$ is expressed as $\|\mathbf{\Omega_{test}} - \mathbf{\Omega}_j\|^2 = -2h_j(\mathbf{\Omega}_{test}) + \mathbf{\Omega}_{test}\mathbf{\Omega}_{test}^T$, where $h_j(\mathbf{\Omega}_{test}) = \mathbf{\Omega}_j\mathbf{\Omega}_{test}^T - \frac{1}{2}\|\mathbf{\Omega}_j\|^2$ is a linear discriminant function of $\mathbf{\Omega}_{test}$. A test image is classified by this classifier by computing seven linear discriminant functions and choosing $MCC = \mathrm{argmax}_{j\in\mathcal{L}_j}\{h_j(\mathbf{\Omega}_{test})\}$.

The classifier accuracy is measured by forming a training and a test set from the database and by applying the classification procedure. To form the training set a number of 164 and 150 samples were randomly chosen from Cohn-Kanade derived database and JAFFE database, respectively. The remaining samples were used for testing. The classification accuracy for the Cohn-Kanade database and both classifiers are presented in Figures 1 and 2 in percentage form, as function of the number of basis components. For this database the highest accuracy is achieved by LNMF with 81.42% corresponding to 36 basis and when MCC is applied. LNMF clearly outperforms NMF for each number of basis components and for both CSM and MCC classifiers. The results for JAFFE database are much worse for the all three methods as compared to Cohn-Kanade database as can be seen from Figures 3 and 4. This is caused by the fact that the expressers posing for this database are not as expressive as those from Cohn-Kanade, making this way the expressions harder to be classified. As the number of basis increases, NMF outperforms both LNMF and PCA when CSM is applied. For MCC, the accuracy variation in the case of NMF is not as smooth as that of CSM. NMF does not outperforms the other two methods which performs equally the same as it can be seen from Figure 4. In order to increase the classifier accuracy for this database, for each subject we have subtracted each expression image from its corresponding neutral pose, thus obtaining difference images and the experiments were repeated. The accuracy for this case is presented in Figures 5 and 6 for CSM and MCC, respectively. By running the experiments on difference images, the accuracy was significantly improved up to 91% and 89.28% for PCA and LNMF, while the accuracy gets worse for NMF, the highest value obtained by it being of only 69.64% achieved at the largest number of basis components. Unfortunately, in many practical cases we do not possess a neutral image to take the difference. Furthermore, image registration should be used before difference calculation.

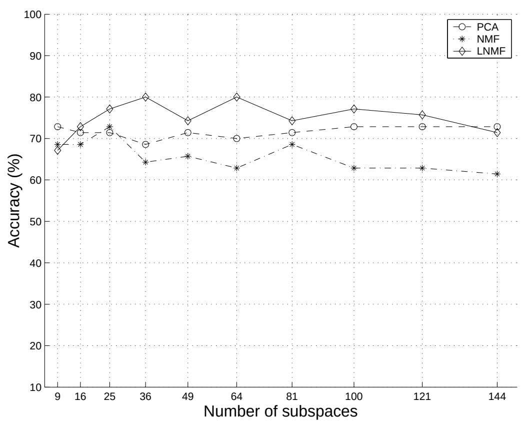

Figure 1. Accuracy for CK database and CSM classifier.

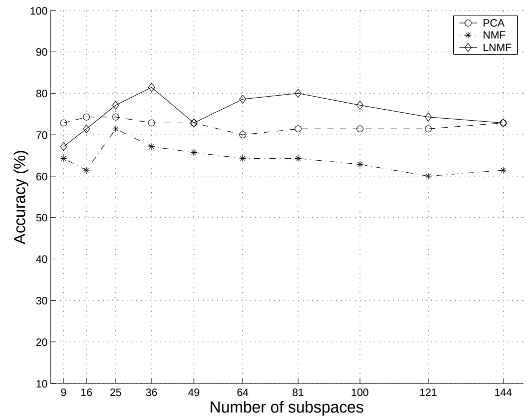

Figure 2. Accuracy for CK database and MCC classifier.

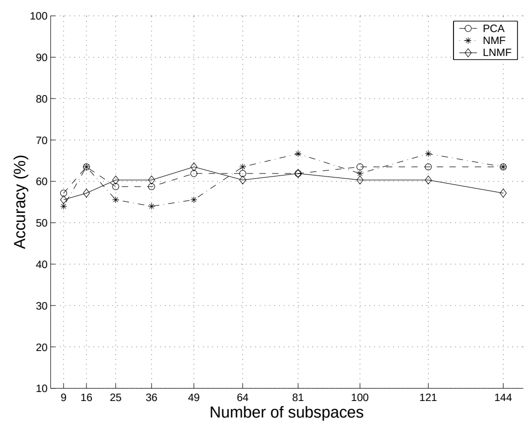

Figure 3. Accuracy for JAFFE database (with neutral pose) and CSM classifier.

4 Conclusion

Two methods (NMF and LNMF) for extracting facial features from two images database representing facial expressions and two classifiers (CSM and MCC) for recognizing these expressions were investigated in this paper. As a baseline, PCA was employed as well. LNMF clearly outperforms NMF for Cohn-Kanade database and JAFFE difference images whilst, in the case of Cohn-Kanade, its performance is superior to PCA. For JAFFE, LNMF and PCA performs approximately the same. The only case when NMF behaves better than PCA and LNMF is for the JAFFE database including neutral pose and for CSM, but it needs the maximum basis components to achieve the highest value of accuracy. As far as the classifier is concerned, overall, we found CSM more reliable than MCC, yielding a higher recognition accuracy.

5 Aknowledgement

The authors would like to thank Stan Z. Li for helpful discussions. This work was supported by the European Union Research Training Network “Multi-modal Human-Computer Interaction (HPRN-CT-2000-00111)

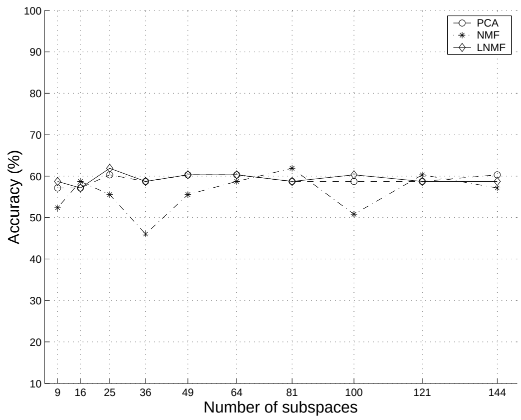

Figure 4. Accuracy for JAFFE database (with neutral pose) and MCC classifier.

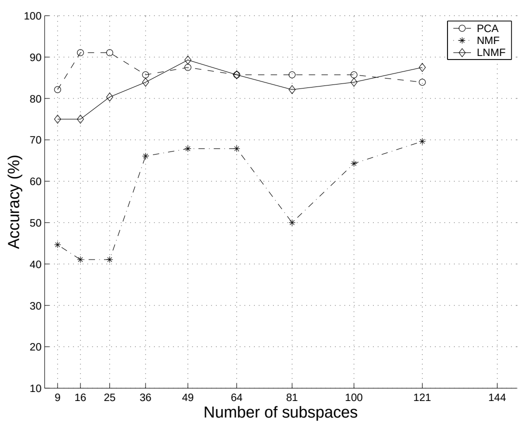

Figure 5. Accuracy for JAFFE database (difference images) and CSM classifier.

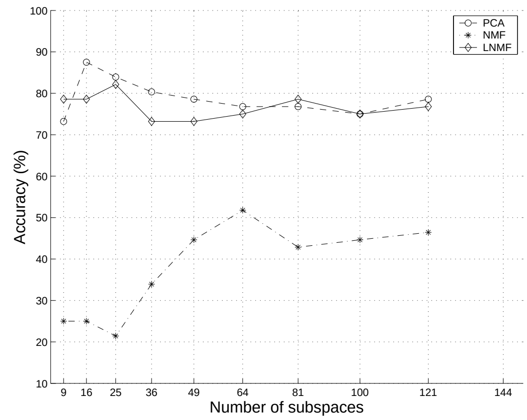

Figure 6. Accuracy for JAFFE database (difference images) and MCC classifier.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[2] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

[3] I. Biederman and P. Kalocsai. Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London*, 352(10):1203–1219, 1997.

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

[5] 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*, 2001.

[6] G. W. Cottrell, M. N. Dailey, C. Padgett, and R. Adolphs. Is all face processing holistic ? *Philosophical Transactions of the Royal Society of London*, 352(10):1203–1219, 1997.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Royal Stat. Soc.*, pages 1–38, 39 2001.

[8] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.

[9] M. J. Farah, K. D. Wilson, M. Drain, and J. N. Tanaka. What is special “about” facial perception ? *Psychological Review*, 105(3):482–498, 1998.

[10] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *in Proc. IEEE Inter. Conf. on Face and Gesture Recognition*, pages 46–53, March 2000.

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

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

[13] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. *in Proc. Third IEEE Inter. Conf. on Automatic Face and Gesture Recognition*, pages 200–205, April 1998.

[14] M. Pantic and L. J. M. Rothkrantz. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, (18):881–905, March 2000.

[15] E. Wachsmuth, M. W. Oram, and D. I. Perrett. Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cereb. Cortex*, 22(4):509–522, December 1994.