Non-negative Matrix Factorization Methods for Face Recognition under Extreme Lighting Variations

Ioan Buciu*, Ioan Nafornita§
* Electronics Department
Faculty of Electrical Engineering and Information Technology
University of Oradea, 410087, Universitatii 1, Romania
Email: ibuciu@uoradea.ro
§Electronics and Communications Faculty
“Politehnica” University of Timisoara, Bd. Vasile Parvan, no.2, 300223 Timisoara, Romania
Email:ioan.nafornita@etc.upt.ro

Abstract— **Face recognition task is of primary interest in many computer vision applications, including access control for security systems, forensic or surveillance. Most commercial biometric systems based on face recognition are claimed to perform satisfactory when the enrollment and testing process takes place under controlled environmental conditions such as constant illumination, constant pose scale, non-occluded faces or frontal view. More or less deviation from those conditions might lead to poor recognition performances or even recognition system's failure when a test identity has to be recognized under new modified testing conditions. Three non-negative matrix factorization (NMF) methods, namely, the standard one, the local NMF (LNMF) and the discriminant NMF (DNMF) are employed in this paper where their robustness against extreme lighting variations are tested for the face recognition task. Principal Component Analysis (PCA) method was also chosen as baseline. Experiments revealed that the best recognition performance is obtained with NMF, followed by DNMF and LNMF.**

I. INTRODUCTION

Impressive interest was shown from the scientific community and commercial vendors for the face recognition task due to its large number and important applications related to visitor identification, building access control, security, surveillance or forensic. Face recognition is a topic researched since the 1960s [1]. A realistic scenario for acquiring image data in the enrolment procedure would involve unconstrained recording conditions, including uncontrolled varying illumination. The illumination setup in which recognition is performed is in most cases impractical to control, its physics difficult to accurately model and face appearance differences due to changing lighting are often larger than those differences between individuals. To cope with this important issue, Blanz and Vetter [2] proposed to use a gradient descent for minimizing the discrepancy between the predicted and observed facial appearance, thus recovering both shape and texture of a novel face. In the attempt of modeling illumination variation, PCA has been been applied [3] to images recorded under varying lighting conditions. The work showed that PCA can well approximates an image set by a low-dimensional linear subspace (five or six dimensional subspace) of the whole image set space. Other techniques include illumination cones [4] or generic shape-illumination manifold [5]. Recently, Wang et al. [6] proposed two strategies (adding or removing light) for illumination compensation and decomposed the image set using PCA, thus obtaining nine-dimensional face illumination subspace based on quotient image. While these methods work with faces acquired using standard visual spectra, other frameworks adopted infrared spectra. Thermal infrared face recognition systems are advantageous when there is little or no control over illumination. Socolinsky et al. [7], and more recently Li et al. [8] proposed an infrared face recognition system which appears to be superior to a face recognition system solely based on visible imagery.

Three methods for decomposing face image data into non-negative factors are employed in this paper to have an insight about their suitability for tackling the varying illumination issue. Standard non-negative matrix factorization (NMF) along with its two derived versions named Local NMF (LNMF) and Discriminant NMF (DNMF) are shortly described in Section II. We must notice that the application of those methods for the face recognition task is not novel. Li et al [9] already explored both NMF and LNMF techniques for data decomposition, while a simple Euclidean distance is used as classifier. Their experiments revealed the superiority of LNMF over the standard NMF for the ORL face database [10], especially for occluded faces. Guillamet and Vitrià [11] also applied NMF to a face recognition task, and, more recently, Buciu et al. [12] reported different results obtained with NMF, LNMF and DNMF when applied two different databases, i. e, ORL and YALE [13]. However, the databases involved in all those experiments contain face images recorded under uniform and constant illumination, with slightly lighting variation. Therefore, the framework described in this paper comes as a natural extension of the previous work, focusing on the methods' performance under extreme varying illumination conditions.

The remaining of the paper is structured as follows. As already mentioned Section II briefly describes the methods. Database involved in experiments is described in Section III. Experimental results are reported in Section IV and conclusions drawn in Section V end up the paper.

II. METHODS

Non-negative Matrix Factorization was proposed by Lee and Seung [14] as a method for decomposing a data matrix $\mathbf{X}$ of dimension $m \times n$ into two factors $\mathbf{W}$ and $\mathbf{H}$, of dimension $m \times p$ and $p \times n$, respectively. The particular characteristic of this decomposition is the constraints imposed on both factors such to have only non-negative entries. Since Lee and Seung published their work in *Nature*, more and more NMF - derived methods were proposed due to its simplicity and consistency with physiological principles and findings. The authors motivated the NMF developing due to partly biological reasons: the non-negative matrix $\mathbf{H}$ can be sought as encoding biological firing rates. Generally, the algorithm starts with random values for both factors $\mathbf{W}$ and $\mathbf{H}$, and, iteratively, the factors entry is updated so that a cost function is minimized. Once a minimum in the cost function is found, the algorithm stops. In principle, finding a minimum is equivalent to have a factors product $\mathbf{W} \cdot \mathbf{H}$ as close as possible to the original data matrix $\mathbf{X}$. Mathematically, $\widetilde{\mathbf{X}} = \mathbf{W} \cdot \mathbf{H} \approx \mathbf{X}$. When data comprise images, those images are usually stored in the columns of $\mathbf{X}$, $\mathbf{W}$ refers to basis images and $\mathbf{H}$ refers to encoding or coefficients matrix. When $p < n$, as usually chosen, NMF compresses the whole data represented by all n images into its p decomposition factors. Each original image can then be reconstructed, in mathematical terminology, as $\mathbf{x}_j = \mathbf{W}\mathbf{h}_j$, where $j = 1 \ldots n$. The quality of reconstruction depends on the cost function associated to the decomposition. Two cost functions have been proposed: Kullback-Leibler divergence $KL(\mathbf{x}||\mathbf{W}\mathbf{h})_{NMF} = \sum_i \left(x_i \log \frac{x_i}{\sum_k W_{ik} h_k} - x_i + \sum_k W_{ik} h_k \right)$ and squared Euclidean distance $D(\mathbf{x}||\mathbf{W}\mathbf{h})_{NMF} = \sum_i \|x_i - \sum_k W_{ik} h_k\|^2$ between $\mathbf{x}$ and its decomposition $\mathbf{W}\mathbf{h}$, for $i = 1 \ldots m$ and $k = 1 \ldots p$.

Li et al [9] modified the NMF algorithm to obtain sparser image features while eliminating redundant information. Their method named *Local Non-negative Matrix Factorization* was successfully applied to face recognition and the experiments indicated the LNMF is more robust to the occlusion leading to superior performance when compared to the standard NMF. The associated LNMF cost function that has to be minimized is provided by:

$$\begin{aligned} D(\mathbf{X}||\mathbf{W}\mathbf{H})_{LNMF} = KL(\mathbf{X}||\mathbf{W}\mathbf{H})_{NMF} + \\ + \alpha \sum_{ik} u_{ik} - \beta \sum_k v_{kk} \end{aligned} \quad (1)$$

where $[u_{jk}] = \mathbf{U} = \mathbf{W}^T\mathbf{W}$, $[v_{jk}] = \mathbf{V} = \mathbf{H}\mathbf{H}^T$ and α, $\beta > 0$ are constants.

LNMF was extended by Buciu et al [15] who proposed the *Discriminant Non-negative Matrix Factorization* which greatly improved the performance in classifying facial expressions. Unlike NMF and LNMF, this technique is a supervised decomposition technique which allows class information encoding in its coefficient matrix. Considering we have $\mathcal{Q}$ distinctive image classes and denoting by n_c the number of samples in class c, $c = 1, \ldots, \mathcal{Q}$, each image from the image database (corresponding to one column of matrix $\mathbf{X}$) pertains to one of those classes. Each column of $\mathbf{H}$ can be expressed as the image representation coefficients vector $\mathbf{h}_{cl}$, where $c = 1, \ldots, \mathcal{Q}$ and $l = 1, \ldots, n_c$. The total number of coefficient vectors is $n = \sum_{c=1}^{\mathcal{Q}} n_c$. We denote the mean coefficient vector of class c by $\boldsymbol{\mu}_c = \frac{1}{n_c} \sum_{l=1}^{n_c} \mathbf{h}_{cl}$ and the global mean coefficient vector by $\boldsymbol{\mu} = \frac{1}{n} \sum_{c=1}^{\mathcal{Q}} \sum_{l=1}^{n_c} \mathbf{h}_{cl}$. Denoting the within-class scatter matrix by $\mathbf{S}_w = \sum_{c=1}^{\mathcal{Q}} \sum_{l=1}^{n_c} (\mathbf{h}_{cl} - \boldsymbol{\mu}_c)(\mathbf{h}_{cl} - \boldsymbol{\mu}_c)^T$ and the between-class scatter matrix by $\mathbf{S}_b = \sum_{c=1}^{\mathcal{Q}} (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^T$, the cost function associated with DNMF algorithm is written as [15]:

$$\begin{aligned} D(\mathbf{X}||\mathbf{W}\mathbf{H}) = KL(\mathbf{X}||\mathbf{W}\mathbf{H})_{NMF} + \alpha \sum_{i,k} u_{ik} - \beta \sum_k v_{kk} + \\ + \gamma \mathbf{S}_w(h) - \delta \mathbf{S}_b(h), \end{aligned} \quad (2)$$

subject to $\mathbf{W}, \mathbf{H} \geq 0$. Here γ and δ are constants. Both $\mathbf{S}_w(h)$ and $\mathbf{S}_b(h)$ are associated to the coefficient matrix.

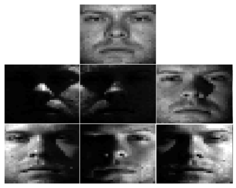

Fig. 1. Seven samples from the training set pertaining to the same subject. Top image represents the free-illumination sample, while the remaining samples correspond to the following light direction: $(A, E) = \{(-110, +65), (+110, +65), (+60, -20), (+60, +20), (-60, -20), (-60, +20)\}$

III. DATABASE DESCRIPTION

The experiments were carried out using the Extended Yale Face Database B [16], [17]. The database contains 38 subjects under 9 poses and 64 illumination conditions. However, we have used the cropped version of the database that only comprises frontal pose, yielding a set of 2462 image samples. The cropped images are re-sized to 168×192 pixels and further to 32×32 pixels to reduce the computational load. We have formed the training and test set as follows. For each subject we picked up the free-illumination pose plus six samples with extreme light direction, i.e., $(A, E) = \{(-110, +65), (+110, +65), (+60, -20), (+60, +20), (-60, -20), (-60, +20)\}$, where A and E stands for the azimuth and elevation, respectively, of a single light source direction. A positive azimuth implies that the light source was to the right of the subject while negative azimuth value refers to the left part. Positive elevation implies above the horizon, while negative implies below the horizon. Therefore, seven samples per subject, as depicted in Figure 1, form the training set for a total of 266 samples. This set up guarantees that the basis images built after applying the decomposition method interpolates between the illumination-free sample and the extreme light variation (in direction) samples, capturing, hopefully, relevant and discriminant information. The remaining images are collected for the test set, comprising 2166

Fig. 2. 57 image samples (out of 64) included in the test set, for the same subject depicted in Figure 1. The illumination changes drastically in orientation and magnitude, so face appearance, making the face recognition task very challenging.

samples. We deliberately form a small training set (7 samples per subject for training and 57 per subject samples for the test set) in order to tackle the small sample size issue too. Figure 2 shows the 57 samples per subject included in the test set.

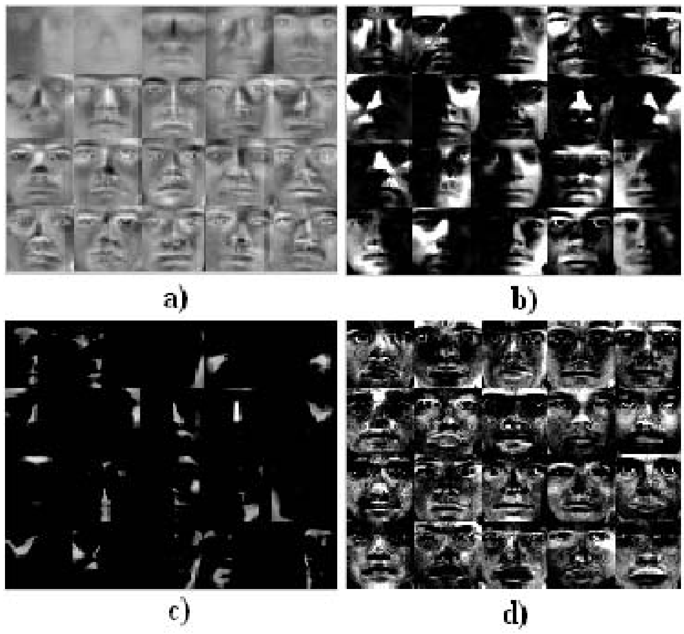

Fig. 3. All training image samples were compressed here onto 20 basis images corresponding to a) PCA, b) NMF, c) LNMF, and d) DNMF, respectively. Notice their different degree of sparseness and representation. Also, notice how the illumination information is accurately encoded into the basis images retrieved by the NMF algorithm.

IV. Experimental results

The full data set of n face images is split into a training set $n^{(tr)}$ and a disjoint test set $n^{(te)}$ with the corresponding matrices $\mathbf{X}^{(tr)}$ and $\mathbf{X}^{(te)}$, respectively. The training images $\mathbf{X}^{(tr)}$ are used for evaluating the decomposition factors. The training procedure refers to finding the decomposition factors for the all three NMF methods. Figure 3 shows 20 basis images found by the three NMF methods and the eigenvectors associated to the PCA decomposition. The basis images greatly differ in their appearance. While holistic basis image representation was found for PCA and NMF, moderate sparse basis images were retrieved by DNMF. Unlike NMF, LNMF algorithm leaded to local facial features.

Once $\mathbf{W}$ and $\mathbf{H}$ are found, and, since $\mathbf{X}^{(tr)} = \mathbf{W}\mathbf{H}$, the feature vectors used for classification are formed as $\mathbf{h}^{(tr)} = \mathbf{W}^{-1}\mathbf{x}^{(tr)}$, where $\mathbf{x}^{(tr)}$ is now a zero mean training face. A new test feature vector $\mathbf{h}^{(te)}$ is then formed as $\mathbf{h}^{(te)} = \mathbf{W}^{-1}\mathbf{x}^{(te)}$, where $\mathbf{x}^{(te)}$ is a zero mean test face. In the classical classification problem, we construct a classifier where the output (predicted value) of the classifier for a test image $\mathbf{x}^{(te)}$ is $\widetilde{l}$. The *recognition error* is defined as the percentage of misclassified face images when $\{\widetilde{l}(\mathbf{h}^{(te)}) \neq l(\mathbf{h}^{(te)})\}$. Thus, the *recognition rate* can be defined as *1 - recognition error*. Once we have formed Q classes of new feature vectors two simple classifiers namely, *cosine similarity measure* (CSM) and *maximum correlation classifier* (MCC) [12] are employed to classify a new test image. CSM is based on the nearest neighbor rule and uses as similarity the angle between a test feature vector and a training one, while MCC is in fact a minimum Euclidean distance classifier. The experiments were carried out for various number of basis images, more precisely, for $p = \{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150\}$. We also used PCA for comparison purpose. The results for all four methods corresponding to CSM and MCC are illustrated in Figure 4 and Figure 5, respectively. As expected, the recognition rate greatly improves and increases monotonically with an increase in the number of basis images, especially up to 50 basis images. From 60 to 150 basis images the performance continues to get higher but smoother. Comparing the methods, the highest recognition rate is attributed to the standard NMF, closely followed by DNMF. The third place is taken by PCA with LNMF in the last position which performs the poorest despite retrieving local facial features. Regarding the two classifiers, Figures 4 and 5 indicate superior performance for the CSM classifier compared to the MCC one.

Table I tabulates the maximum recognition rate in percentage corresponding to the four methods and the two classifiers employed in experiments, where the best results are in bold. An impressive difference in the performance was noticed when CSM and MCC are compared. For CSM, the recognition rate greatly improves with over 6 % for all methods except the DNMF algorithm where the increase is of only 1 %.

V. Conclusion

Face recognition under extreme illumination changes was addressed in this paper. Three techniques, NMF, LNM and DNMF, relying on non-negative matrix decomposition were

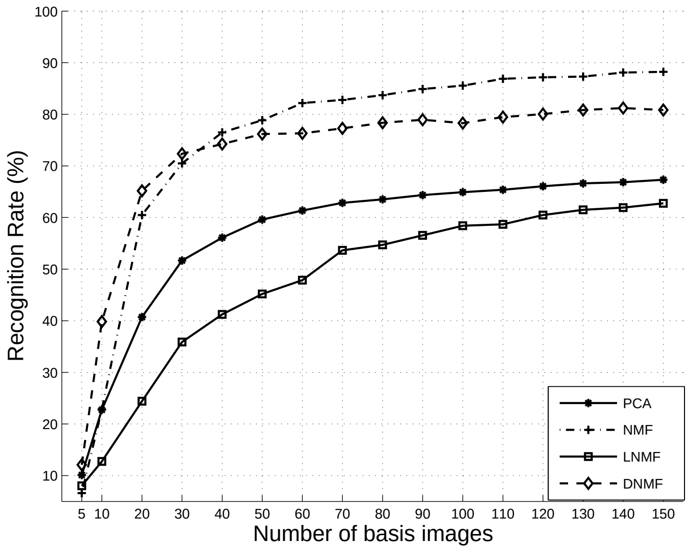

Fig. 4. Recognition rate in percentage (%) for CSM classifier and various number (p) of basis images.

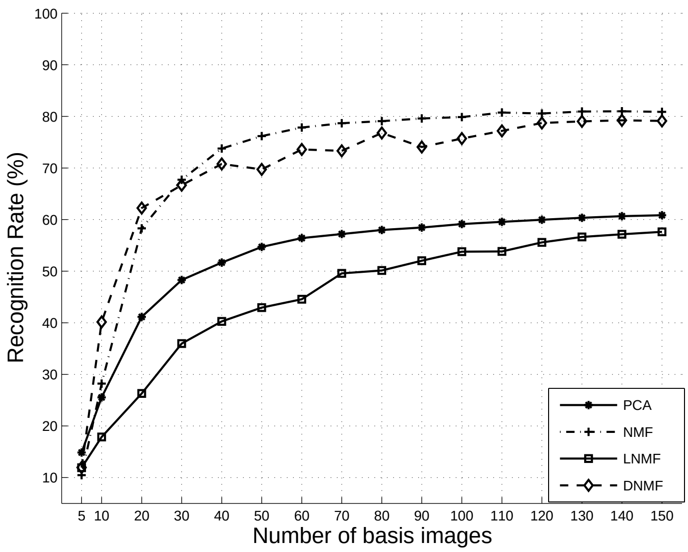

Fig. 5. Recognition rate in percentage (%) for MCC classifier and various number (p) of basis images.)

employed to extract robust facial features. The features extracted by those decomposition methods are further classified using two classifiers, CSM and MCC. Summarizing, CSM is preferred as it exhibits significant superior performance when combined with any decomposition method. As far as the feature extraction method is concerned, NMF seems to retrieve the most robust features against extremely varying illumination conditions followed by DNMF. Correlating the results and the image representation, intuitively, a more holistic image representation favors the recognition performance, fact that could explain the poor behavior of the LNMF approach. Although both PCA and NMF conduct to holistic features, it was the NMF only which incorporates illumination appearance

TABLE I

MAXIMUM RECOGNITION RATE EXPRESSED IN PERCENTAGE (%) FOR ALL FOUR METHODS INVESTIGATED IN THE PAPER AND THE TWO CLASSIFIERS.

Classifier	Method			
	PCA	NMF	LNMF	DNMF
CSM	67.31	**88.22**	62.74	80.22
MCC	60.84	**80.97**	57.64	79.22

and information in its basis images as depicted in Figure 3. This might explain why NMF leaded to the highest recognition rate providing illumination invariant facial features.

ACKNOWLEDGMENT

This work was supported by a grant from the Romanian National University Research Council, PNCDI Programe ID-668/2009.

REFERENCES

[1] Stan Z. Li and Anil K. Jain, *Handbook of face recognition*, Springer-Verlag, 2005.

[2] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3D faces,” *IEEE Conf. on Computer Graphics*, pp. 187194, 1999.

[3] P. Hallinan, “A low-dimensional representation of human faces for arbitrary lighting conditions,” *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 995999, 1994.

[4] P. N. Belhumeur and D. J. Kriegman, “What is the set of images of an object under all possible illumination conditions?,” *International Journal of Computer Vision*, 28(3), pp. 245 - 260, 1998.

[5] O. Arandjelovic and R. Cipolla, “Face recognition from video using the generic shape-illumination manifold,” *9th European Conference on Computer Vision*, pp. 7 - 13, May 2006.

[6] Y. - H. Wang, X. J. Ning, C. X. Yang, and Q. F. Wang, “A method of illumination compensation for human face image based on quotient image,” *Informational Sciences*, vol. 178, no. 12, pp. 2705–2721, 2008.

[7] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, “Illumination invariant face recongition using thermal infrared imagery,” *Comput. Vision Pattern Recognition*, 1, pp. 527–534, 2001.

[8] S. Z. Li, R. F. Chu, S-C. Liao, and L. Zhang “Illumination invariant face recognition using near-infrared images,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.

[9] 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.

[10] http://www.uk.research.att.com/

[11] D. Guillamet and Jordi Vitrià, “Non-negative matrix factorization for face recognition,” *Topics in Artificial Intelligence*, Springer Verlag Series: Lecture Notes in Artificial Intelligence, vol. 2504, pp. 336–344, 2002.

[12] I. Buciu, N. Nikolaidis, and I. Pitas, “A comparative study of NMF, DNMF, and LNMF algorithms applied for face recognition,” *2006 Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, 2006.

[13] http://cvc.yale.edu

[14] 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.

[15] I. Buciu and I. Pitas, “A new sparse image representation algorithm applied to facial expression recognition,” in Proc. *IEEE Workshop on Machine Learning for Signal Processing*, pp. 539–548, 2004.

[16] A. Georghiades, P. Belhumeur, and D. Kriegman, “From few to many: generative models for recognition under variable pose and illumination,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[17] K.C. Lee, J. Ho and D. Kriegman, “Acquiring Linear Subspaces for Face Recognition under Variable Lighting,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 648–698, 2005.