International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 21, No. 8 (2007) 1353–1362
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION WITH APPLICATIONS TO FACIAL IMAGE PROCESSING

ZHIRONG YANG*, ZHIJIAN YUAN† and JORMA LAAKSONEN‡

Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 5400, FI-02015 HUT, Espoo, Finland
**zhirong.yang@hut.fi*
†*zhijian.yuan@hut.fi*
‡*jorma.laaksonen@hut.fi*

We propose a new variant of Non-negative Matrix Factorization (NMF), including its model and two optimization rules. Our method is based on positively constrained projections and is related to the conventional SVD or PCA decomposition. The new model can potentially be applied to image compression and feature extraction problems. Of the latter, we consider processing of facial images, where each image consists of several parts and for each part the observations with different lighting mainly distribute along a straight line through the origin. No regularization terms are required in the objective functions and both suggested optimization rules can easily be implemented by matrix manipulations. The experiments show that the derived base vectors are spatially more localized than those of NMF. In turn, the better part-based representations improve the recognition rate of semantic classes such as the gender or existence of mustache in the facial images.

Keywords: Non-negative matrix factorization; projective; facial image; principal component analysis.

1. Introduction

In image analysis, a compressive or compact representation of the high-dimensional input is often required. One of the standardly used methods is the Principal Component Analysis (PCA), which applies the Singular Value Decomposition (SVD) on the image covariance matrix and projects the input image on the resulting eigenvectors.

However, the base vectors of PCA fail to keep the non-negativity property of the input signals. Recently, Lee and Seung proposed a method called *Non-negative Matrix Factorization* (*NMF*),[3] which imposes the non-negativity constraints in learning the base images. Such constraints and the related multiplicative update rules seem to yield part-based representations. Nevertheless, NMF is sensitive to the initial values[7] and the additive parts learned by NMF are not necessarily localized.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Some variants of NMF (e.g. Refs. 2 and 3) have later been proposed. Most of them append one or more regularization terms to the objective function of NMF, and they are reported to generate more highly part-based features. The updating rules of these methods are albeit rather complicated and the localized representations can be obtained only if the trade-off parameter is properly chosen.

In our previous work,[8] a new variant of NMF, called *Projective Non-negative Matrix Factorization* (P-NMF), was proposed. The new method differs from NMF in that it replaces the weight matrix in NMF with the inner product of the base vectors and the input images. P-NMF does not involve any regularization terms or trade-off parameters, but is still able to learn more spatially localized, part-based representations of visual patterns. In the present work, we employ global normalization instead of separate optimizations for individual base vectors. In addition to recapitulate the model and algorithms of P-NMF, we also discuss the underlying reason that leads to high orthogonality or sparseness by P-NMF, and present both qualitative and quantitative comparison with NMF. Furthermore, the features obtained by P-NMF are used as an input to classifiers for facial images.

The remainder of the paper is organized as follows. We start with a brief review of NMF in Sec. 2. In Sec. 3, we first present the model of P-NMF and its connection to the PCA approach. Then we review and correct the associated optimization rules, with the discussion about the underlying reason of high orthogonality also included. Experimental results on facial images are shown in Sec. 4. Finally, Sec. 5 concludes the paper.

2. Non-Negative Matrix Factorization

Suppose that our non-negative data is given in the form of an $m \times n$ matrix $\mathbf{V}$. Its n columns are the data items, for example, a set of images that have been vectorized by row-by-row scanning. Then m is the number of pixels in any given image.

Given $\mathbf{V}$ and a constant r, the *Non-negative Matrix Factorization algorithm* (NMF)[3] finds a non-negative $m \times r$ matrix $\mathbf{Q}$ and another non-negative $r \times n$ matrix $\mathbf{H}$ such that they minimize the optimality problem

$$\min_{\mathbf{Q},\mathbf{H} \geq 0} \|\mathbf{V} - \mathbf{QH}\|. \tag{1}$$

This can be interpreted as follows: each column of the matrix $\mathbf{Q}$ contains a base vector while each column of $\mathbf{H}$ contains the weights needed to approximate the corresponding column in $\mathbf{V}$ using the base vectors from $\mathbf{Q}$. So the product $\mathbf{QH}$ can be regarded as a compressed form of the data in $\mathbf{V}$.

3. Projective Non-Negative Matrix Factorization

3.1. *Model*

To improve locality of part-based representations, we incorporate the idea of compressive SVD by finding a subspace $\mathcal{B}$ of $\mathbb{R}^m$, and an $m \times m$ projection matrix $\mathbf{P}$

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

with given rank r, such that $\mathbf{P}$ projects the non-negative matrix $\mathbf{V}$ onto the subspace $\mathcal{B}$ and preserves the non-negativity property. Finally, it should minimize the difference $\|\mathbf{V} - \mathbf{PV}\|$.

We can write any symmetric positive semi-definite projection matrix of rank r in the form

$$\mathbf{P} = \mathbf{W}\mathbf{W}^T. \tag{2}$$

Here we require $\mathbf{W}$ to be an orthonormal $m \times r$ matrix. Thus, we can solve the problem by searching for a non-negative $\mathbf{W}$ as the solution to the following optimality problem

$$\min_{\mathbf{W} \geq 0} \|\mathbf{V} - \mathbf{W}\mathbf{W}^T\mathbf{V}\|, \tag{3}$$

where $\|\cdot\|$ is a matrix norm, e.g. the Frobenius norm. We call the new method *Projective Non-negative Matrix Factorization* (P-NMF).

The physical model of the objective function (3) is illustrated as follows. Suppose each observation $\mathbf{v}$ is composed of r nonoverlapped parts, i.e. $\mathbf{v} = \sum_{p=1}^{r} \mathbf{v}_p$. We model each part $\mathbf{v}_p$ by the scaling of a base vector $\mathbf{w}_p$ plus a noise vector $\boldsymbol{\epsilon}_p$:

$$\mathbf{v}_p = \alpha_p \mathbf{w}_p + \boldsymbol{\epsilon}_p. \tag{4}$$

If the base vectors are normalized so that $\mathbf{w}_p^T \mathbf{w}_q = 1$ for $q = p$ and 0 otherwise, then the reconstructed vector of this part is

$$\begin{aligned}
\sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \mathbf{v}_p &= \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T (\alpha_p \mathbf{w}_p + \boldsymbol{\epsilon}_p) \\
&= \sum_{q=1}^{r} \alpha_p \mathbf{w}_q \mathbf{w}_q^T \mathbf{w}_p + \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \boldsymbol{\epsilon}_p \\
&= \alpha_p \mathbf{w}_p + \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \boldsymbol{\epsilon}_p.
\end{aligned} \tag{5}$$

The norm of the reconstruction error is therefore bounded by

$$\begin{aligned}
\left\| \mathbf{v}_p - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \mathbf{v}_p \right\| &= \left\| \left(\mathbf{I} - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \right) \boldsymbol{\epsilon}_p \right\| \\
&\leq \left\| \left(\mathbf{I} - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \right) \right\| \cdot \|\boldsymbol{\epsilon}_p\| \\
&= \mathrm{Tr} \left(\left(\mathbf{I} - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \right)^T \left(\mathbf{I} - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \right) \right) \cdot \|\boldsymbol{\epsilon}_p\| \\
&= \mathrm{Tr} \left(\mathbf{I} - \sum_{q=1}^{r} \mathbf{w}_q \mathbf{w}_q^T \right) \cdot \|\boldsymbol{\epsilon}_p\|
\end{aligned}$$

$$= \left(\mathrm{Tr}(\mathbf{I}) - \sum_{q=1}^{r} \mathrm{Tr}\left(\mathbf{w}_q^T \mathbf{w}_q\right) \right) \cdot \|\boldsymbol{\epsilon}_p\|$$

$$= (m - r) \cdot \|\boldsymbol{\epsilon}_p\|, \tag{6}$$

if 2-norm is used, and similar bound can be derived for other types of norms. In other words, $\mathbf{w}_p \mathbf{w}_p^T \mathbf{v}_p$ reconstructs $\mathbf{v}_p$ well if the noise level $\boldsymbol{\epsilon}_p$ is small enough. According to this model, P-NMF can potentially be applied to signal processing problems where the global signals can be divided into several parts and for each part the observations mainly distribute along a straight line modeled by $\alpha_p \mathbf{w}_p$. This is closely related to Oja's PCA subspace rule,[4] which finds the direction of the largest variation, except that the straight line found by P-NMF has to pass through the origin.

If the columns of $\mathbf{W}$ are orthogonal to each other, the above derivation and non-negativity property can be easily extended to multiple parts because two non-negative vectors are orthogonal if and only if they do not have the same non-zero dimensions.

An alternative measurement for the difference (3) is the matrix divergence[3] of $\mathbf{V}$ from $\mathbf{U} = \mathbf{W}\mathbf{W}^T\mathbf{V}$, which is

$$D(\mathbf{V}\|\mathbf{U}) = \sum_{i,j} \left(V_{ij} \log \frac{V_{ij}}{U_{ij}} - V_{ij} + U_{ij} \right), \tag{7}$$

$D(\mathbf{V}\|\mathbf{U})$ is lower bounded by zero, and vanishes if and only if $\mathbf{V} = \mathbf{W}\mathbf{W}^T\mathbf{V}$. Such a divergence can be viewed as a variant of the Kullback–Leibler divergence and obtained as a negative log-likelihood under the assumption that each V_{ij} is generated by a Poisson distribution with parameter U_{ij}.[6]

3.2. *Optimization rules*

We first consider the Frobenius norm of (3). Define the function

$$F = \frac{1}{2} \sum_{i,j} [V_{ij} - (\mathbf{W}\mathbf{W}^T\mathbf{V})_{ij}]^2. \tag{8}$$

Then the unconstrained gradient of F for $\mathbf{W}$ is given by

$$\frac{\partial F}{\partial W_{ij}} = -2(\mathbf{V}\mathbf{V}^T\mathbf{W})_{ij} + (\mathbf{W}\mathbf{W}^T\mathbf{V}\mathbf{V}^T\mathbf{W})_{ij} + (\mathbf{V}\mathbf{V}^T\mathbf{W}\mathbf{W}^T\mathbf{W})_{ij}. \tag{9}$$

Using the gradient we can construct an additive update rule for minimization,

$$W_{ij} \leftarrow W_{ij} - \eta_{ij} \frac{\partial F}{\partial W_{ij}}, \tag{10}$$

where η_{ij} is a positive step size. To guarantee that the elements of W_{ij} remain non-negative, we choose the step size as

$$\eta_{ij} = \frac{W_{ij}}{(\mathbf{W}\mathbf{W}^T\mathbf{V}\mathbf{V}^T\mathbf{W})_{ij} + (\mathbf{V}\mathbf{V}^T\mathbf{W}\mathbf{W}^T\mathbf{W})_{ij}}. \tag{11}$$

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Then the additive update rule (10) can be formulated as a multiplicative update rule

$$W_{ij} \leftarrow W_{ij} \frac{2(\mathbf{V}\mathbf{V}^T\mathbf{W})_{ij}}{(\mathbf{W}\mathbf{W}^T\mathbf{V}\mathbf{V}^T\mathbf{W})_{ij} + (\mathbf{V}\mathbf{V}^T\mathbf{W}\mathbf{W}^T\mathbf{W})_{ij}}. \tag{12}$$

For the divergence measure (7), the gradient is

$$\begin{aligned}\frac{\partial D(\mathbf{V}\|\mathbf{W}\mathbf{W}^T\mathbf{V})}{\partial W_{ij}} = &\sum_k \left((\mathbf{W}^T\mathbf{V})_{jk} + \sum_l W_{lj}V_{ik} \right) \\ &- \sum_k V_{ik}(\mathbf{W}^T\mathbf{V})_{jk}/(\mathbf{W}\mathbf{W}^T\mathbf{V})_{ik} \\ &- \sum_k V_{ik} \sum_l W_{lj}V_{lk}/(\mathbf{W}\mathbf{W}^T\mathbf{V})_{lk}.\end{aligned} \tag{13}$$

Using the gradient, the additive update rule becomes

$$W_{ij} \leftarrow W_{ij} - \zeta_{ij} \frac{\partial D(\mathbf{V}\|\mathbf{W}\mathbf{W}^T\mathbf{V})}{\partial W_{ij}}, \tag{14}$$

where ζ_{ij} is a step size. Choosing this step size as

$$\zeta_{ij} = \frac{W_{ij}}{\sum_k ((\mathbf{W}^T\mathbf{V})_{jk} + \sum_l W_{lj}V_{ik})}, \tag{15}$$

we obtain the multiplicative update rule

$$W_{ij} \leftarrow W_{ij} \frac{\sum_k V_{ik}(\mathbf{W}^T\mathbf{V})_{jk}/(\mathbf{W}\mathbf{W}^T\mathbf{V})_{ik} + \sum_k V_{ik} \sum_l W_{lj}V_{lk}/(\mathbf{W}\mathbf{W}^T\mathbf{V})_{lk}}{\sum_k ((\mathbf{W}^T\mathbf{V})_{jk} + \sum_l W_{lj}V_{ik})}. \tag{16}$$

Please notice that (14)–(16) correct the errors in Ref. 8 which resulted in reversed order of the numerator and denominator in the multiplicative update rule.

The numerators in (12) and (16) originate from the negative terms in the partial derivative while the denominators from the positive ones. Iteratively applying the multiplicative update rules actually implements a kind of Hebbian learning. The entries of $\mathbf{W}$ with the partial derivative larger than zero will be awarded an amplified factor, whereas those with negative partial derivative will be squeezed to zero. In the terms of neural networks, this can be interpreted as a competition between the elements, both within a neuron and across the neurons. In the matrix case, more than one neuron compete for the energy from the objective function, and the normalized Hebbian learning forces that only one of them wins over all the others. This leads to high orthogonality between the learned base vectors. Therefore, although the optimization of P-NMF does not involve an explicit orthogonalization step, the resulting matrix $\mathbf{W}$ has high sparsity, locality and orthogonality.

However, the plain Hebbian learning may cause some entries of $\mathbf{W}$ to blow up after a large number of iterations, and stabilizing the basis is therefore required.

Suppose $\mathbf{w}_i$, $i = 1, \ldots, r$, are the base vectors of P-NMF. We normalize them after each multiplicative update (12) or (16) by

$$\mathbf{W} \leftarrow \mathbf{W} / \max_i \{\|\mathbf{w}_i\|\}. \tag{17}$$

The norms of all base vectors are not necessarily unitary after each iteration, but we will show later that most base vectors will get close to the unit sphere after sufficient learning.

4. Experiments

We have used the FERET database of facial images.[5] After face segmentation, 2409 frontal facial images (poses "fa" and "fb") of 867 subjects were stored in the database for the experiments. For the study we have obtained the coordinates of the eyes from the ground truth data of the FERET collection and calibrated the head rotation so that all faces are upright. All face boxes were normalized to the size of 32×32, with fixed locations for the left eye (26,9) and the right eye (7,9). We used the optimization rules (16) and (17) with the Kullback–Leibler divergence norm in all the following experiments.

4.1. *Learning base components*

The base images of NMF and P-NMF with $r = 16$ are shown in Fig. 1. Each base component consists of 32×32 pixels and corresponds to a column in the resulting matrix $\mathbf{W}$. Brighter pixels correspond to larger values in the basis. All the images are displayed with the matlab command "imagesc" without any extra processing.

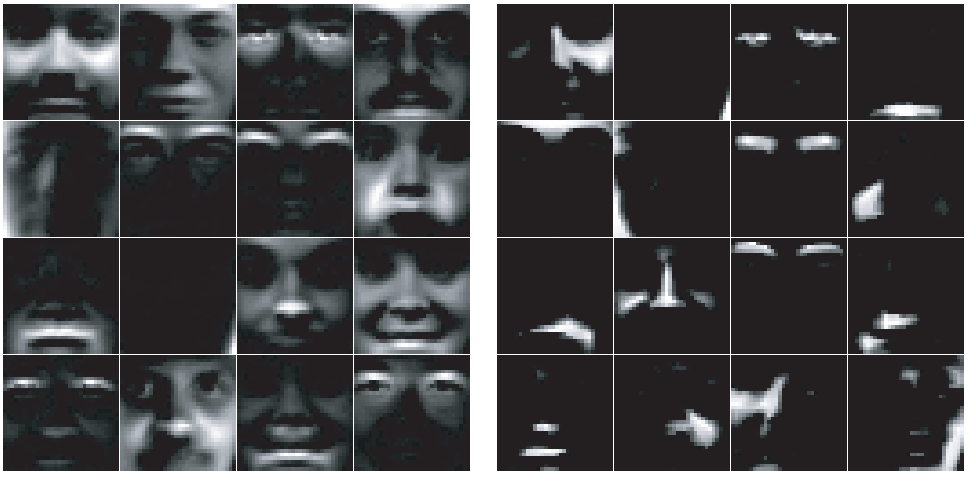


Fig. 1. NMF (left) and P-NMF (right) bases of 16 dimensions.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

The base components of P-NMF are spatially more localized and nonoverlapped. P-NMF clearly divides a facial image into several facial parts such as the lips and eyebrows. In contrast, the base images of NMF are more holistic and their visible parts are clearly overlapped.

4.2. *Orthonormality tests*

Two non-negative vectors are orthogonal if and only if they do not have same non-zero dimensions. Hence, the orthogonality between the learned base vectors reveals the sparsity of the resulting representations and, in turn, the localization for facial images. Suppose the normalized inner product between two base vectors $\mathbf{w}_i$ and $\mathbf{w}_j$ is

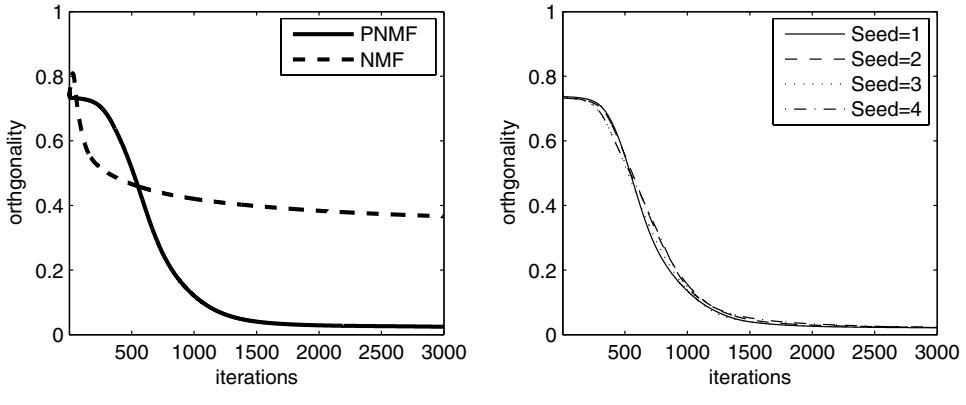
$$\mathbf{R}_{ij} = \frac{\mathbf{w}_i^T \mathbf{w}_j}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|}. \tag{18}$$

Then the orthogonality of the basis can be quantified by the following ρ measurement:

$$\rho = \|\mathbf{R} - \mathbf{I}\|/(r(r-1)), \tag{19}$$

where $\|\cdot\|$ refers to the Frobenius matrix norm. Smaller ρ's indicate higher orthogonality and ρ reaches 0 when the columns of $\mathbf{W}$ are completely orthogonal.

Figure 2(a) shows the resulting ρ's of P-NMF and NMF with 16 dimensions. NMF converges to a local minimum with $\rho = 0.367$ while P-NMF learns $\mathbf{W}$ with $\rho = 0.022$ after 3000 iterations. We also trained P-NMF with different random seeds for the initial values of $\mathbf{W}$ and the results are shown in Fig. 2(b). It can be seen that P-NMF converges with very similar curves. That is, the high orthogonality obtained by P-NMF does not take place by accident and is not sensitive to the initial values.



(a) (b)

Fig. 2. ρ values of (a) P-NMF and NMF with 16 dimensions and (b) P-NMF with four different random seeds.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

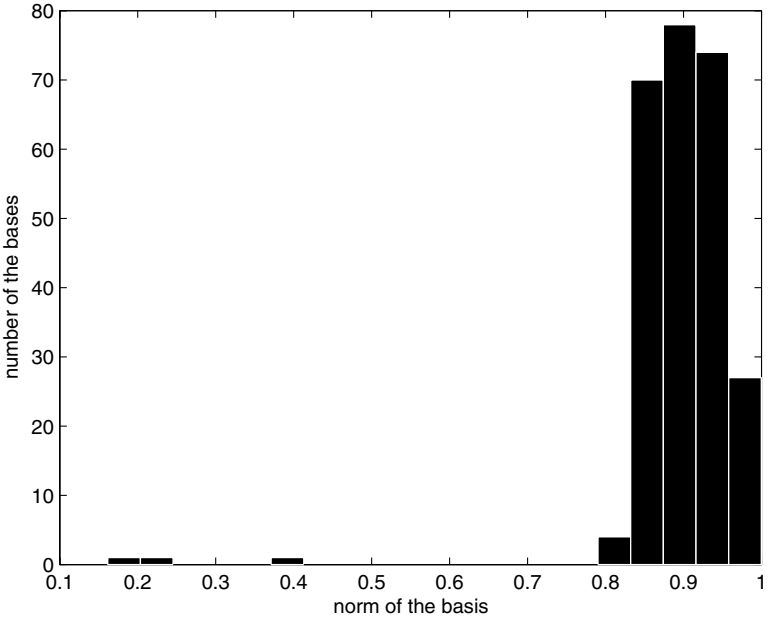


Fig. 3. Histogram of the norms of the P-NMF base vectors after 20,000 iterations.

Next, we demonstrate that most base vectors approach the unit norm after sufficient training. We have again used the same faces, but set $r = 256$ for convenient histogram plotting. The results are shown in Fig. 3, from which we can clearly see that most norms are between 0.8 and 1. The simple normalization rule (17) guarantees that the base vectors do not blow up, and together with the multiplicative updates, finally forces the base vectors close to the unit sphere.

4.3. *Classification*

In this section we demonstrate an application of P-NMF and NMF which serves as a preprocessing step before discriminative learning. The resulting coefficients of P-NMF and NMF are further projected to a one-dimensional subspace created by Fisher's Linear Discriminant Analysis (LDA). 1204 images are used in the LDA training and 1,205 for testing. LDA outputs only scalar values to discriminate whether the person in the image has *mustache* (256 images) or not (2153 images). Afterwards, a threshold is used to divide the values into two classes such that the misclassification error rates on both classes are equal. The resulting *equal error rate* (EER) for the procedure using P-NMF is 16.06% while for NMF it is 19.18%.

We also found that in some cases P-NMF can approach or even outperform PCA. Because the features generated by P-NMF with different r's represent the

information of the image in different resolutions, we can follow the idea of wavelet decomposition by concatenating the coefficients from multiple levels before input to LDA. We tested the classification of *gender* using P-NMF with $\{r = 9, 16, 25, 36, 49, 64, 81\}$ coefficients and the compared it with PCA of $9 + 16 + 25 + 36 + 49 + 64 + 81 = 280$ principal components. The EER of the gender classification is 15.27% with P-NMF features and 16.72% with PCA.

5. Conclusions

We have proposed a new variant of NMF, which differs from the original method in two aspects. Firstly, the product $\mathbf{QH}$ is replaced by $\mathbf{WW}^T\mathbf{V}$. This corresponds to an easily interpreted model and implies its suitable application range. Secondly, $\mathbf{W}$ is normalized by a single scalar in each iteration. Unlike NMF, where the base vectors are individually normalized by their vector sum, our scaling method does not break the relationship between the base vectors during the optimization.

The empirical results have shown that P-NMF is able to learn highly localized and part-based representations of facial images. The optimization turned out to be insensitive to the initial values. Compared with other variants of NMF, our derivation is more straightforward and the algorithm can be implemented more efficiently. In addition, the model is ready to be extended to nonlinear cases, for example, by generalizing the dot product between base vectors and input signals to other kernel functions.

Acknowledgments

The work that constitutes this paper is supported by the Academy of Finland in the projects "Neural methods in information retrieval based on automatic content analysis and relevance feedback" and Finnish Centre of Excellence in Adaptive Informatics Research.

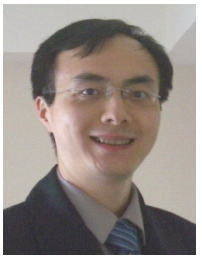
References

1. I. Kotsia, S. Zafeiriou and I. Pitas, A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems, *IEEE Trans. Inform. Forens. Secur.* **2**(3) (2007) 588–595.
2. H. Laurberg and L. K. Hansen, On affine non-negative matrix factorization, *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* **2** (2007), pp. 653–656.
3. D. D. Lee and H. S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* **401** (1999) 788–791.
4. E. Oja, Principal components, minor components, and linear neural networks, *Neural Networks* **5** (1992) 927–935.
5. P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. Patt. Anal. Mach. Intell.* **22** (2000) 1090–1104.
6. P. Sajda, S. Du and L. C. Parra, Recovery of constituent spectra using non-negative matrix factorization, *Proc. SPIE* **5207** (November 2003), pp. 321–331.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.

7. S. Wild, J. Curry and A. Dougherty, Improving non-negative matrix factorizations through structured initialization, *Patt. Recogn.* **37**(11) (2004) 2217–2232.
8. Z. Yuan and E. Oja, Projective nonnegative matrix factorization for image compression and feature extraction, *Proc. 14th Scandinavian Conf. Image Analysis (SCIA 2005)*, Joensuu, Finland (June 2005), pp. 333–342.

Int. J. Patt. Recogn. Artif. Intell. 2007.21:1353-1362. Downloaded from www.worldscientific.com
by UNIVERSITY OF AARHUS on 06/02/19. Re-use and distribution is strictly not permitted, except for Open Access articles.



Zhirong Yang received his Bachelor's and Master's degrees in computer science from Sun Yat-Sen University, Guangzhou, China, in 1999 and 2002, respectively. Presently he is a doctoral candidate at the Computer and Information Science Laboratory of Helsinki University of Technology.

His research interests include machine learning, pattern recognition, computer vision and multimedia retrieval.

Zhijian Yuan received the M. S. degree in mathematics from Beijing Institute of Technology, Beijing, China in 1992, and Licenciate degree in mathematics from Helsinki University of Technology, Helsinki, Finland in 2002. He is currently working toward the Ph.D. degree at the Laboratory of Information and Computer Science, Helsinki University of Technology, Helsinki, Finland.

Jorma Laaksonen received his Dr. of Science in Technology in 1997 from Helsinki University of Technology, Finland, where he is presently Academy Research Fellow of Academy of Finland at the Laboratory of Computer and Information Science.

Dr. Laaksonen is an IEEE senior member, a founding member of the SOM and LVQ Programming Teams and the PicSOM Development Group, and a member of the International Association of Pattern Recognition (IAPR) Technical Committee 3: Neural Networks and Machine Learning.

His research interests are in content-based information retrieval and recognition of handwriting. He is the author of several journal and conference papers on pattern recognition, statistical classification, and neural networks.