Research Journal of Applied Sciences, Engineering and Technology 6(9): 1626-1631, 2013
DOI:10.19026/rjaset.6.3880
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: December 23, 2012 Accepted: January 25, 2013 Published: July 15, 2013

Research Article

Linear Projective Non-Negative Matrix Factorization

[1, 2, 3]Lirui Hu, [1, 3]Jianguo Wu and [1, 3]Lei Wang
[1]Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China
[2]School of Computer Science and Technology, Nantong University, Nantong 226019, China
[3]School of Computer Science and Technology, Anhui University, Hefei 230039, China

Abstract: In order to solve the problem that the basis matrix is usually not very sparse in Non-Negative Matrix Factorization (NMF), a method, called Linear Projective Non-Negative Matrix Factorization (LP-NMF), is proposed. In LP-NMF, from projection and linear transformation angle, an objective function of Frobenius norm is defined. The Taylor series expansion is used. An iterative algorithm for basis matrix and linear transformation matrix is derived and a proof of algorithm convergence is provided. Experimental results show that the algorithm is convergent; relative to Non-negative Matrix Factorization (NMF), the orthogonality and the sparseness of the basis matrix are better; in face recognition, there is higher recognition accuracy. The method for LP-NMF is effective.

Keywords: Face recognition, linear transformation, non-negative matrix factorization, projective

INTRODUCTION

Projective Non-negative Matrix Factorization (P-NMF) $X \approx WW^T X$ (Yuan and Oja, 2005) was proposed based on NMF (Lee and Seung, 1999). Since it was constructed from the projection angle, the basis matrix W was only computed in the algorithm for P-NMF. The computational complexity was lower for one iteration step for P-NMF, as only one matrix had to be computed instead of two for NMF.

Linear Projection-Based Non-negative Matrix Factorization (LPBNMF) $X \approx WQX$ (Li and Zhang, 2010a) was constructed from projection and linear transformation angle. In LPBNMF, a monotonic convergence algorithm was given and the orthogonality and the sparseness of the basis matrix were computed quantificationally.

On the basis of optimization rules for P-NMF and LPBNMF, the basis matrixes were all forced to tend to be orthogonal. So, relative to NMF, the orthogonality and the sparseness of the basis matrixes were better and then the methods for P-NMF and LPBNMF were more beneficial to the applications of data dimension reduction, pattern recognition, and so on. However, since the algorithm for P-NMF wasn't convergent, the method for LPBNMF was more beneficial to the application (Li and Zhang, 2010a).

In this study, another method is proposed based on LPBNMF $X \approx WQX$. We call it Linear Projective Non-negative Matrix Factorization (LP-NMF). Relative to the algorithm in the study (Li and Zhang, 2010a), the iterative formulae of this algorithm are simpler.

LINEAR PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION (LP-NMF)

Taking Frobenius norm as similarity measure, we consider an objective function:

$$F = \frac{1}{2}\|X - WQX\|_F^2 \tag{1}$$

where, $X \geq 0, W \geq 0, Q \geq 0$.

The mathematical model in NMF definition $X \approx WH$ is based on nonlinear projection. But, the basic idea for LP-NMF is that: firstly, we turn the data X into QW by a suitable linear transformation Q. Secondly, we may consider that QW is the projection of the sample space X onto a suitable subspace W. Finally, we minimize the objective function F in Eq. (1) to get W and Q. Here, we respectively call W basis matrix and Q linear transformation matrix.

The update rule for basis matrix W: For any element w_{ab} of W, let $F_{w_{ab}}$ stand for the part of F relevant to w_{ab} in Eq. (1). So, writing w instead of w_{ab} in the expression of $F_{w_{ab}}$, we may get a function $F_{w_{ab}}(w)$. Obviously, the first order derivative of $F_{w_{ab}}(w)$ at w_{ab} is the first order partial derivative of F with respect to w_{ab}. That is:

Corresponding Author: Lirui Hu, School of Computer Science and Technology, Nantong University, Nantong 226019, China

$$F'_{w_{ab}}(w_{ab}) = \frac{\partial F}{\partial w_{ab}} = \frac{\partial(\frac{1}{2}\sum_{ij}[X_{ij}-(WQX)_{ij}]^2}{\partial w_{ab}}$$
$$= \sum_{ij} -(X_{ij}-(WQX)_{ij})\frac{\partial(WQX)_{ij}}{\partial w_{ab}}$$
$$= \sum_{ij}(-X_{ij}+(WQX)_{ij})(\frac{\partial(\sum_k W_{ik}(QX)_{kj})}{\partial w_{ab}})$$
$$= \sum_{ij}(-X_{ij}+(WQX)_{ij})(\sum_k \frac{\partial W_{ik}}{\partial w_{ab}}(QX)_{kj})$$
$$= \sum_{ij}(-X_{ij}+(WQX)_{ij})(\frac{\partial W_{ib}}{\partial w_{ab}}(QX)_{bj})$$
$$= \sum_j\left[\sum_i(-X_{ij}+(WQX)_{ij})(\frac{\partial W_{ib}}{\partial w_{ab}}(QX)_{bj})\right]$$
$$= \sum_j(-X_{aj}+(WQX)_{aj})(QX)_{bj}$$
$$= \sum_j - X_{aj}(QX)_{bj}+(WQX)_{aj}(QX)_{bj}$$
$$= -(XX^TQ^T)_{ab}+(WQXX^TQ^T)_{ab} \quad (2)$$

Similarly, the second order derivative of $F_{w_{ab}}(w)$ at w_{ab} is:

$$F''_{w_{ab}}(w_{ab}) = \frac{\partial(-(XX^TQ^T)_{ab}+(WQXX^TQ^T)_{ab})}{\partial w_{ab}}$$
$$= \frac{\partial(-(XX^TQ^T)_{ab})}{\partial w_{ab}} + \frac{\partial(WQXX^TQ^T)_{ab}}{\partial w_{ab}}$$
$$= (QXX^TQ^T)_{bb} \quad (3)$$

and other order derivatives of $F_{w_{ab}}(w)$ with respect to w are:

$$F^{(n)}_{w_{ab}}(w) = 0 \quad (4)$$

where, $n \geq 3$.
Thus, the Taylor series expansion of $F_{w_{ab}}(w)$ at w_{ab} is:

$$F_{w_{ab}}(w) = F_{w_{ab}}(w_{ab}) + F'_{w_{ab}}(w_{ab})(w-w_{ab}) + \frac{1}{2}F''_{w_{ab}}(w_{ab})(w-w_{ab})^2 \quad (5)$$

Meantime, to emphasize time of w_{ab} in numerical calculation, we write $w_{ab}^{(t)}$ instead of w_{ab} in the brackets of $F_{w_{ab}}(w)$. So, equation:

$$F_{w_{ab}}(w) = F_{w_{ab}}(w_{ab}^{(t)}) + F'_{w_{ab}}(w_{ab}^{(t)})(w-w_{ab}^{(t)}) + \frac{1}{2}F''_{w_{ab}}(w_{ab}^{(t)})(w-w_{ab}^{(t)})^2 = F_{w_{ab}}(w_{ab}^{(t)}) + F'_{w_{ab}}(w_{ab}^{(t)})(w-w_{ab}^{(t)}) + \frac{1}{2}(QXX^TQ^T)_{bb}(w-w_{ab}^{(t)})^2 \quad (6)$$

is gotten from Eq. (5).
Now, we define a function:

$$G_{w_{ab}}(w, w_{ab}^{(t)}) = F_{w_{ab}}(w_{ab}^{(t)}) + F'_{w_{ab}}(w_{ab}^{(t)})(w-w_{ab}^{(t)}) + \frac{1}{2}\frac{(WQXX^TQ^T)_{ab}}{w_{ab}^{(t)}}(w-w_{ab}^{(t)})^2 \quad (7)$$

Theorem 1: $G_{w_{ab}}(w, w_{ab}^{(t)})$ is an auxiliary function for $F_{w_{ab}}(w)$.

Proof: $G_{w_{ab}}(w, w_{ab}^{(t)}) = F_{w_{ab}}(w)$ is obvious when $w_{ab}^{(t)} = w$. We need show that $G_{w_{ab}}(w, w_{ab}^{(t)}) \geq F_{w_{ab}}(w)$ when $w_{ab}^{(t)} \neq w$.
Because $W \geq 0, Q \geq 0, X \geq 0$,

$$(WQXX^TQ^T)_{ab} = \sum_k W_{ak}^{(t)}(QXX^TQ^T)_{kb} \geq W_{ab}^{(t)}(QXX^TQ^T)_{bb}$$

When $W_{ab}^{(t)} > 0$,

$$\frac{(WQXX^TQ^T)_{ab}}{W_{ab}^{(t)}} \geq (QXX^TQ^T)_{bb}$$

In fact, $W_{ab}^{(t)} = w_{ab}^{(t)}$. So,

$$\frac{(WQXX^TQ^T)_{ab}}{w_{ab}^{(t)}} \geq (QXX^TQ^T)_{bb}$$

And

$$G_{w_{ab}}(w, w_{ab}^{(t)}) \geq F_{w_{ab}}(w)$$

Thus, $G_{w_{ab}}(w, w_{ab}^{(t)})$ is an auxiliary function for $F_{w_{ab}}(w)$ according to the definition 1 of reference (Lee and Seung, 2001).

Theorem 2: $F_{w_{ab}}(w)$ is nonincreasing under the update:

$$w_{ab}^{(t+1)} = \arg\min_w G_{w_{ab}}(w, w_{ab}^{(t)})$$

Proof: Because:

$$F_{w_{ab}}(w_{ab}^{(t+1)}) \leq G_{w_{ab}}(w_{ab}^{(t+1)}, w_{ab}^{(t)}) \leq G_{w_{ab}}(w_{ab}^{(t)}, w_{ab}^{(t)}) = F_{w_{ab}}(w_{ab}^{(t)})$$

$F_{w_{ab}}(w)$ is nonincreasing.

Using the definition of auxiliary function and Theorem 2, we can get the local minimum of $F_{w_{ab}}(w)$ if only the local minimum of $G_{w_{ab}}(w, w_{ab}^{(t)})$ is gotten. To get a local minimum of $F_{w_{ab}}(w)$, we may calculate the first order partial derivative of $G_{w_{ab}}(w, w_{ab}^{(t)})$ with respect to w, and have:

$$\frac{\partial G_{w_{ab}}(w, w_{ab}^{(t)})}{\partial w} = F'_{w_{ab}}(w_{ab}^{(t)}) + \frac{(WQXX^TQ^T)_{ab}}{w_{ab}^{(t)}}(w - w_{ab}^{(t)})$$
$$= -(XX^TQ^T)_{ab} + (WQXX^TQ^T)_{ab} + \frac{(WQXX^TQ^T)_{ab}}{w_{ab}^{(t)}}(w - w_{ab}^{(t)}) = 0$$

and

$$w = w_{ab}^{(t)} \frac{(XX^TQ^T)_{ab}}{(WQXX^TQ^T)_{ab}}$$

So, the update rule of w_{ab} is:

$$w^{(t+1)} = w_{ab}^{(t)} \frac{(XX^TQ^T)_{ab}}{(WQXX^TQ^T)_{ab}} \tag{8}$$

Using this update rule, we may make the auxiliary function $G_{w_{ab}}(w, w_{ab}^{(t)})$ local minimum, and thus make the objective function $F_{w_{ab}}(w)$ local minimum. If all elements of W are updated by Eq. (8), the local minimum of the objective function F may be gotten. The algorithm converges after finite iterations. The Eq. (8) is the update rule for the basis matrix W.

The update rule for linear transformation matrix Q: Similarly, we can get a function $F_{q_{ab}}(q)$. All derivatives of $F_{q_{ab}}(q)$ are:

$$F'_{q_{ab}}(q_{ab}) = \frac{\partial F}{\partial q_{ab}} = \frac{\partial(\frac{1}{2}\sum_{ij}[X_{ij} - (WQX)_{ij}]^2}{\partial q_{ab}}$$
$$= \sum_{ij} -(X_{ij} - (WQX)_{ij})\frac{\partial (WQX)_{ij}}{\partial q_{ab}}$$
$$= \sum_{ij}(-X_{ij} + (WQX)_{ij})(\frac{\partial(\sum_k W_{ik}(QX)_{kj})}{\partial q_{ab}})$$
$$= \sum_{ij}(-X_{ij} + (WQX)_{ij})(\sum_k W_{ik}\frac{\partial (QX)_{kj}}{\partial q_{ab}})$$
$$= \sum_{ij}(-X_{ij} + (WQX)_{ij})(\sum_k W_{ik}\frac{\partial(\sum_l Q_{kl}X_{lj})}{\partial q_{ab}})$$
$$= \sum_{ij}(-X_{ij} + (WQX)_{ij})(\sum_k W_{ik}\frac{\partial Q_{kb}}{\partial q_{ab}}X_{bj})$$
$$= \sum_{ij}(-X_{ij} + (WQX)_{ij})(W_{ia}X_{bj})$$
$$= \sum_j\left[\sum_i(-X_{ij} + (WQX)_{ij})(W_{ia}X_{bj})\right]$$
$$= \sum_j\left[\sum_i(-X_{ij} + (WQX)_{ij})W_{ia}\right]X_{bj}$$
$$= \sum_j -(W^TX)_{aj}X_{bj} + (W^TWQX)_{aj}X_{bj}$$
$$= -(W^TXX^T)_{ab} + (W^TWQXX^T)_{ab}$$

and

$$F''_{q_{ab}}(q_{ab}) = \frac{\partial(-(W^TXX^T)_{ab} + (W^TWQXX^T)_{ab})}{\partial q_{ab}}$$
$$= \frac{\partial(-(W^TXX^T)_{ab})}{\partial q_{ab}} + \frac{\partial(W^TWQXX^T)_{ab}}{\partial q_{ab}}$$
$$= \frac{\partial(\sum_k (W^TW)_{ak}(QXX^T)_{kb})}{\partial q_{ab}}$$
$$= \sum_k (W^TW)_{ak}\frac{\partial(QXX^T)_{kb}}{\partial q_{ab}}$$
$$= \sum_k (W^TW)_{ak}\frac{\partial(\sum_l Q_{kl}(XX^T)_{lb})}{\partial q_{ab}}$$
$$= \sum_k (W^TW)_{ak}(\sum_l \frac{\partial Q_{kl}}{\partial q_{ab}}(XX^T)_{lb})$$
$$= \sum_k (W^TW)_{ak}(\frac{\partial Q_{kb}}{\partial q_{ab}}(XX^T)_{bb})$$
$$= (W^TW)_{aa}(XX^T)_{bb}$$

and

$$F^{(n)}_{q_{ab}}(q) = 0, \quad s.t. \quad n \geq 3$$

So, the Taylor series expansion of $F_{q_{ab}}(q)$ at q_{ab} is:

$$F_{q_{ab}}(q) = F_{q_{ab}}(q_{ab}) + F'_{q_{ab}}(q_{ab})(q - q_{ab}) + \frac{1}{2}F''_{q_{ab}}(q_{ab})(q - q_{ab})^2 \tag{9}$$

Meantime, when numerical calculation is considered, Eq. (9) is expressed through equation:

$$F_{q_{ab}}(q) = F_{q_{ab}}(q_{ab}^{(t)}) + F'_{q_{ab}}(q_{ab}^{(t)})(q - q_{ab}^{(t)}) + \frac{1}{2}(W^TW)_{aa}(XX^T)_{bb}(q - q_{ab}^{(t)})^2 \tag{10}$$

We define a function:

$$G_{q_{ab}}(q, q_{ab}^{(t)}) = F_{q_{ab}}(q_{ab}^{(t)}) + F'_{q_{ab}}(q_{ab}^{(t)})(q - q_{ab}^{(t)}) + \frac{1}{2}\frac{(W^TWQXX^T)_{ab}}{q_{ab}^{(t)}}(q - q_{ab}^{(t)})^2 \tag{11}$$

$G_{q_{ab}}(q, q_{ab}^{(t)}) = F_{q_{ab}}(q)$ is obvious when $q_{ab}^{(t)} = q$

$$(W^TWQXX^T)_{ab} = \sum_k (W^TW)_{ak}(QXX^T)_{kb}$$
$$\geq (W^TW)_{aa}(QXX^T)_{ab}$$
$$= (W^TW)_{aa}\sum_l Q_{al}(XX^T)_{lb}$$
$$\geq (W^TW)_{aa}Q_{ab}^{(t)}(XX^T)_{bb}$$

When $Q_{ab}^{(t)} > 0$

$$\frac{(W^TWQXX^T)_{ab}}{Q_{ab}^{(t)}} \geq (W^TW)_{aa}(XX^T)_{bb}$$

In fact, $Q_{ab}^{(t)} = q_{ab}^{(t)}$.

So, $G_{q_{ab}}(q, q_{ab}^{(t)}) \geq F_{q_{ab}}(q)$ and $G_{q_{ab}}(q, q_{ab}^{(t)})$ is an auxiliary function for $F_{q_{ab}}(q)$. We have:

$$\frac{\partial G_{q_{ab}}(q, q_{ab}^{(t)})}{\partial q} =$$

$$F'_{q_{ab}}(q_{ab}^{(t)}) + \frac{(W^T WQXX^T)_{ab}}{q_{ab}^{(t)}}(q - q_{ab}^{(t)})$$

$$= -(W^T XX^T)_{ab} + (W^T WQXX^T)_{ab} + \frac{(W^T WQXX^T)_{ab}}{q_{ab}^{(t)}}(q - q_{ab}^{(t)}) = 0$$

and

$$q = q_{ab}^{(t)} \frac{(W^T XX^T)_{ab}}{(W^T WQXX^T)_{ab}}$$

Thus, the update rule of q_{ab} is:

$$q_{ab}^{(t+1)} = q_{ab}^{(t)} \frac{(W^T XX^T)_{ab}}{(W^T WQXX^T)_{ab}} \tag{12}$$

If all elements of Q are updated by Eq. (12), the local minimum of the objective function F may be gotten.

The Eq. (12) is the update rule for the linear transformation matrix Q.

Algorithm steps: Using Eq. (8) and Eq. (12), we may get an algorithm to compute the basis matrix W and the linear transformation matrix Q. As follows:

Step1: Initialize W, Q and X with non-negative data
Step2: Update W by Eq. (8)
Step3: Update Q by Eq. (12)
Step4: Repeat step2 and Step3 until algorithm converges

EXPERIMENTS AND ANALYSIS

In order to verify the convergence of the algorithm and the sparseness of the basis matrix W, we do an experiment. In the experiment, X consists of the first five images of each person in the ORL facial image database, a total of 200 data. We randomly initialize W and Q with non-negative data, and set the rank of the basis matrix W 80. In order to reduce the amount of computation and speed up operating speed, every image is reduced to half.

Algorithm convergence: In the experiment, the varied curve of the objective function values versus iteration steps is shown in Fig. 1. We can see that the algorithm is convergent, but the convergence speed is lower which the reason is that the algorithm is still an alternating optimization method.

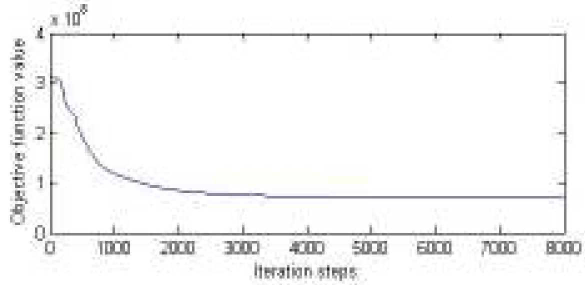


Fig. 1: Objective function values versus iteration steps when the basis matrix is initialized randomly with non-negative data

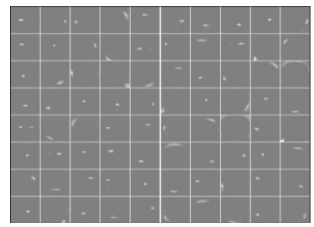
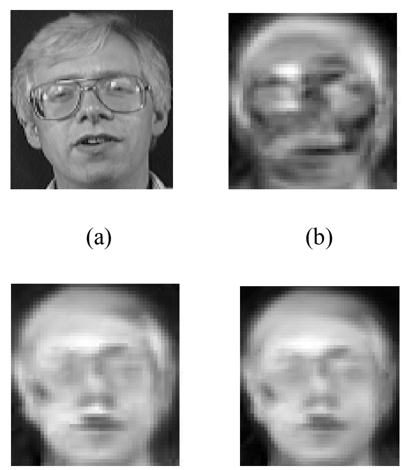


Fig. 2: Basis matrix image



(a) (b)

(c) (d)

Fig. 3: (a) Original image x; (b) $W(W^T x)$; (c) $W(W^T W)^{-1} W^T x$; (d) $W(Qx)$

Analysis of the basis matrix: Meantime, the basis matrix image is shown in Fig. 2. We respectively take the vector $W^T x$, $(WTW)^{-1} W^T x$ and Qx as the feature vector of the data x and reconstruct x, and reconstructed results are respectively shown in the (b) image, the (c) image and the (d) image in Fig. 3.

From the basis matrix image, we can see that the basis matrix is very sparse. This shows that the basis matrix W is forced to tend to be orthogonal by optimizing the objective function F.

From the reconstructed images, we can see that three reconstructed images are all effective, and this shows that the basis matrix W is effective; getting the reconstructed image of x is better by $W(W^T W)^{-1} W^T x$ or $W(Qx)$ than by $W(W^T x)$, and this shows that the basis

Table 1: Comparison of recognition accuracy for LP-NMF (%)

The rank of W/ Template library	20	40	60	80	100	120	140	160
$(W^TW)^{-1} W^TX$	86.5	89.5	89.5	91.5	90	89.5	54	25
QX	84.5	88	89	88.5	89.5	87.5	90.5	90

matrix W is still approximately orthogonal; however, $W(W^TW)^{-1}W^Tx$ image almost is the same as $W(Qx)$ image, so it is question that how to select the feature vector of data x from $(W^TW)^{-1}W^Tx$ and Qx. It will be answered in next section.

In addition, the orthogonality and the sparseness of the basis matrix may be computed quantificationally (Li and Zhang, 2010a; Yang *et al.*, 2007). Without doubt, because this method is still based on the objective function in Eq. (1) for optimization, the orthogonality and the sparseness of the basis matrix are still better. Here, we don't repeat them.

RESULTS OF FACE RECOGNITION AND ANALYSIS

In learning phase, X consists of the first five images of each person in the ORL facial image database, a total of 200 data. In order to reduce the amount of computation, and speed up the operating speed, each image is reduced to a quarter of the original. We initialize randomly W and Q with non-negative data. After the algorithm converges, we get the basis matrix W, matrix Q, $(W^TW)^{-1}W^TX$, and QX.

In the pattern recognition test phase, we take the after five images of each person in the ORL facial image database, a total of 200 data, as test data, and reduce every image to a quarter of the original.

We first decide the feature vector of data x from $(W^TW)^{-1}W^Tx$ and Qx in LP-NMF by experiments. In these experiments, we respectively take $(W^TW)^{-1}W^TX$ and QX as template library, and use the nearest neighbor rule for face recognition. When the ranks of the basis matrix W are set different values, the results of the face recognition are shown in Table 1. From the Table 1, the template library $(W^TW)^{-1}W^TX$ used, the recognition accuracy is very low when the rank of the basis matrix is greater than or equal to 140. On the contrary, the recognition accuracy is higher when the template library QX is used. But, when the rank of the basis matrix is smaller than 140, the recognition accuracy is slightly higher using $(W^TW)^{-1}W^TX$ than using QX. Therefore, the linear transformation matrix Q has some important information.

So, in next experiments, we take the matrix QX as a template library, and use Qx to compute the feature vector of test image x by the matrix Q obtained in the learning phase, and use the nearest neighbor rule for

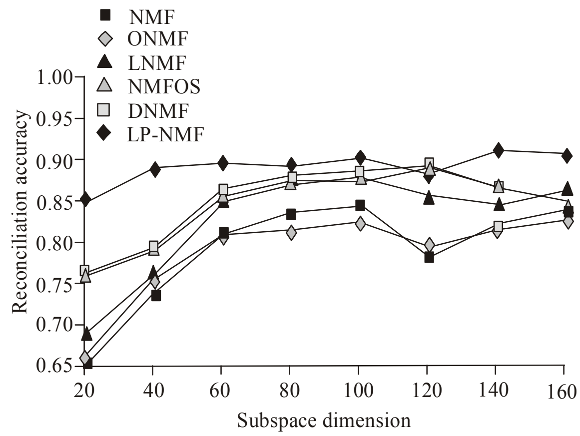


Fig. 4: Comparison of the results of face recognition in the ORL

face recognition. We compare this method with the methods of NMF, LNMF (Li *et al.*, 2001), NMFOS (Li *et al.*, 2010b), ONMF (Yoo and Choi, 2010), and DNMF (Buciu and Nafornita, 2009). When the ranks (i.e., the feature subspace dimensions) of the basis matrix are set different values, the results of the face recognition are shown in Fig. 4.

As can be seen from the Fig. 4, the recognition accuracy is obviously higher using LP-NMF than using NMF or ONMF. The cause is that the basis matrix W is forced to tend to be orthogonal by the objective function for LP-NMF in Eq. (1) so that the basis matrix is more orthogonal in LP-NMF than in NMF. So the discriminative power of the feature vector Qx for LP-NMF is better. Meantime, when the rank of the basis matrix is greater than or equal to 60, the recognition accuracy is slightly higher using LP-NMF than using LNMF or NMFOS. This is because there are also approximately orthogonal constraints for the basis matrixes in the objective functions for LNMF and NMFOS so that the discriminative power of the feature vectors is also good. But the discriminative power of the feature vector Qx for LP-NMF is better. Finally, since the class information is taken into account in DNMF, there is also higher recognition accuracy.

In addition, when the rank of the basis matrix of LP-NMF is between 40 and 160, the recognition accuracy becomes more stable. This is because the orthogonality and the sparseness of the basis matrix for the LP-NMF are always better so that the recognition accuracy is less affected by the number of the rank of basis matrix.

CONCLUSION

In this study, we propose a method, called Linear Projective Non-negative Matrix Factorization (LP-NMF). In LP-NMF, the algorithm steps are given. Relative to LPBNMF, the iterative formulae are simpler. Relative to NMF, the orthogonality and the sparseness of the basis matrix are better. Relative to NMF and some extended NMF, there is higher recognition accuracy in face recognition.

ACKNOWLEDGMENT

This study was supported by Key Technologies Research and Development Program of Chinese Anhui Province under grant No. 07010202057.

REFERENCES

Buciu, I. and I. Nafornita, 2009. Non-negative matrix factorization methods for face recognition under extreme lighting variations. Proceeding of the International Symposium on Signals, Circuits and Systems, Electron. Dept., Univ. of Oradea, Oradea, Romania, pp: 125-128.

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

Lee, D.D. and H.S. Seung, 2001. Algorithms for Non-Negative Matrix Factorization. Advances in Neural Information Processing Systems, MIT Press, Cambridge, Mass.

Li, S.Z., X.W. Hou, H.J. Zhang and Q.S. Cheng, 2001. Learning spatially localized, parts-based representation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Beijing Sigma Center, Microsoft Res. China, Beijing, China, pp: 1-6.

Li, L. and Y.J. Zhang, 2010a. Linear projection-based non-negative matrix factorization. Acta Automat. Sinica, 36(1): 23-39.

Li, Z., X. Wu and H. Peng, 2010b. Non-negative matrix factorization on orthogonal subspace. Patt. Recognit. Lett., 31(9): 905-911.

Yang, Z.R., Z.J. Yuan and J. Laaksonen, 2007. Projective nonnegative matrix factorization with applications to facial image processing. Int. J. Patt. Recognit. Artificial Intell., 21(8): 1353-1362.

Yoo, J. and S. Choi, 2010. Orthogonal non-negative matrix factorization for co-clustering: Multiplicative updates on stiefel manifolds. Inform. Proc. Manag., 46(5): 559-570.

Yuan, Z.J. and E. Oja, 2005. Projective nonnegative matrix factorization for image compression and feature extraction. Proceeding of the 14th Scandinavian Conference on Image Analysis, pp: 333-342.