

Group Sparse Non-negative Matrix Factorization for Multi-Manifold Learning

Xiangyang Liu^{1,2}

liuxy@sjtu.edu.cn

Hongtao Lu¹

htlu@sjtu.edu.cn

Hua Gu²

guhuasy@hhu.edu.cn

¹ Department of Computer Science and Engineering

Shanghai Jiao Tong University
Shanghai, China

² College of Science

Hohai University
Nanjing, China

Abstract

Many observable data sets such as images, videos and speech can be modeled by a mixture of manifolds which are the result of multiple factors (latent variables). In this paper, we propose a novel algorithm to learn multiple linear manifolds for face recognition, called Group Sparse Non-negative Matrix Factorization (GSNMF). Via the group sparsity constraint imposed on the column vectors of the coefficient matrix, we obtain multiple linear manifolds each of them belongs to a particular class. For a test image, we represent it as a linear combination of the learned multiple linear manifolds, and then the representation is naturally group sparse: only the coefficients corresponding to the same class are nonzero. We conduct extensive experiments to verify the proposed algorithm using the ORL database, the Yale database and the Extended Yale B database. Our evaluation shows that GSNMF achieves accurate recognition on face images with varying illuminations and expressions.

1 Introduction

Many observable data sets can be modeled by a mixture of manifolds [6]. For example, for handwritten digits, each digit forms its own manifold in the feature space. For human faces, the face images of the same person under different conditions lie on the same manifold and different persons are associated with different manifolds [15]. Many works focused on the case of hybrid linear modeling, i.e., one linear model for each homogeneous subset of data [4]. Vidal and Ma *et al.* proposed an algebraic geometric approach, called Generalized Principal Component Analysis (GPCA) [17] to estimate a mixture of subspaces. Chen and Lerman proposed Spectral Curvature Clustering (SCC) method to capture the curvature of a collection of points within an affine subspace for hybrid linear modeling [4]. Tipping *et al.* proposed a Mixture model of Probabilistic Principal Component Analyzers (MPPCA) [16] for face recognition. Recently there has been an interest in a classification model where the training samples of a particular class are assumed to form a linear basis for approximately representing any new test sample belonging to that class [18]. This assumption has led to the development of some new classifiers like the Sparse Classifier [18].

Nonnegative Matrix Factorization (NMF) [11] has been widely used due to its psychological and physiological interpretation of naturally occurring data whose representation may be parts-based in the human brain. NMF aims to find two nonnegative matrices whose product can well approximate the nonnegative data matrix, which naturally leads to parts-based and non-subtractive representation. Many variants of NMF have been proposed such as sparse NMF [8, 9], divergence-based NMF [5] and graph regularized NMF [3] for face recognition, and neighborhood preserving NMF [7] for clustering. Recently, many group sparsity regularization methods have been presented such as group lasso [14], group sparse coding [2]. Group sparsity has been considered for simultaneous sparse approximation and multi-task compressive sensing from the Bayesian hierarchical modeling point of view [10].

In this paper, we propose a novel algorithm to learn multiple linear manifolds based on group sparsity and non-negative matrix factorization. Via the group sparsity constraint imposed on the column vectors of the coefficient matrix, we obtain multiple linear manifolds each of them belongs to a particular class. We adopt the l_1/l_2 regularizer for the objective function to yield group sparsity. For a test image, we represent it as a linear combination of the learned linear manifolds, and then the representation is naturally group sparse: only the coefficients corresponding to the same class are nonzero. We derive multiplicative updating rules for the proposed algorithm with initialization. We conduct extensive experiments to verify the proposed algorithm using three face recognition databases. The evaluation shows that the proposed algorithm achieves good recognition rates on face images with varying illuminations and expressions.

The rest of the paper is organized as follows. Firstly, we give a review of NMF and its sparse variants. Then, we introduce our algorithm GNMF, infer the updating rules, and describe the experimental results. Finally, we provide conclusions and future work.

2 Review of NMF and Its Sparse Variants

In this section, we will briefly review NMF and its sparse variants. Given a non-negative data matrix $X = [x_1, x_2, \dots, x_n] \in \mathfrak{R}_+^{m \times n}$, each column of X is a data point. NMF [11] aims to find two non-negative matrices $W \in \mathfrak{R}_+^{m \times r}$ and $H \in \mathfrak{R}_+^{r \times n}$ which minimize the objective

$$\begin{aligned} O_1^{NMF} &= \|X - WH\|_F^2, \\ \text{s.t. } & W \geq 0, H \geq 0, \end{aligned} \quad (1)$$

where $\|\cdot\|_F$ is Frobenius norm. An alternative objective is the divergence between two matrices

$$\begin{aligned} O_2^{NMF} &= \sum_{i,j} \left(X_{ij} \ln \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right), \\ \text{s.t. } & W \geq 0, H \geq 0. \end{aligned} \quad (2)$$

Two updating algorithms have been derived for the above objective functions respectively [12]. NMF results in a reduced representation of the original data that can be seen either as a feature extraction or a dimensionality reduction technique. More importantly, NMF can be interpreted as a parts-based representation of the data due to the fact that only additive, no subtractive, combinations are allowed.

Sparsity constraints on either W or H can be similarly imposed. The notion of sparsity refers sometimes to a representational scheme where only a few features are effectively used

to represent data vectors [8]. Using theory inherited from Linear Sparse Coding, the direct sparseness constraint can be used as follows [8]

$$C_{SNMF} = \sum_{i,j} H_{ij}. \quad (3)$$

Hoyer proposed a novel sparseness constraints for NMF [9]

$$\text{Sparseness}(W_i) = S_w, \quad \forall_i, i = 1, \dots, r, \quad (4)$$

$$\text{Sparseness}(H_i) = S_h, \quad \forall_i, i = 1, \dots, r,$$

where W_i is the i th column of W , H_i is the i th row of H , S_w and S_h are the desired sparseness values for W and H , respectively. The sparseness measure [9] based on the relationship between the l_1 and l_2 norm of a vector \mathbf{x} is

$$\text{Sparseness}(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{n} - 1}, \quad (5)$$

where n is the dimensionality of the vector \mathbf{x} .

3 Group Sparse Non-negative Matrix Factorization (GSNMF)

This section presents the GSNMF model and the updating algorithm for multiple linear manifolds learning. The l_1/l_2 regularizer is used for the objective function to yield group sparsity on the coefficient matrix H . Via the group sparsity constraint imposed on the column vectors of the coefficient matrix, multiple linear manifolds are learned each of which belongs to a particular class.

3.1 Formulation

Given a data matrix $X = [x_1, x_2, \dots, x_n] \in \mathfrak{R}_+^{m \times n}$, Our GSNMF also aims to find two non-negative matrices $W \in \mathfrak{R}_+^{m \times r}$ and $H \in \mathfrak{R}_+^{r \times n}$. Assuming the number of manifolds is K , and the dimension of each manifold is p . Thus $r = K \times p$ in our scheme. For the j -th column vector h_j of H , it can be divide into K groups $\mathcal{G}_k, k = 1, \dots, K$, and each group has p coefficients. Given a grouping \mathcal{G} of the column vector h_j of H , the k th group norm $\|\mathcal{G}_k\|_2$ is given by

$$\|\mathcal{G}_k\|_2 = \left(\sum_{\alpha \in \mathcal{G}_k} H_{\alpha j}^2 \right)^{\frac{1}{2}}, \quad (6)$$

where $\|\cdot\|_2$ is l_2 norm. Then the group sparsity for the column $h_j (j = 1, \dots, n)$ of coefficient matrix H is defined by

$$\|h_j\|_1^{\mathcal{G}} = \sum_k \|\mathcal{G}_k\|_2 = \sum_k \left(\sum_{\alpha \in \mathcal{G}_k} H_{\alpha j}^2 \right)^{\frac{1}{2}}, \quad (7)$$

where $\|\cdot\|_1$ is l_1 norm. $\|h_j\|_1^{\mathcal{G}}$ is the l_1/l_2 regularizer of the vector h_j .

After the l_1/l_2 regularizer on the column vectors of the coefficient matrix H , the corresponding basis matrix W is expected to be composed of multiple linear manifolds each of which belongs to a particular class. GSNMF minimizes the distance objective function combining the l_1/l_2 regularizer as follows

$$\begin{aligned} O_1 &= \|X - WH\|_F^2 + \lambda \sum_{j=1}^n \|h_j\|_1^{\mathcal{G}}, \\ \text{s.t. } \quad &W \geq 0, H \geq 0, \end{aligned} \quad (8)$$

or, minimizes the divergence between the two matrices with the l_1/l_2 regularizer

$$\begin{aligned} O_2 &= \sum_{i,j} \left(X_{ij} \ln \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij} \right) + \lambda \sum_{j=1}^n \|h_j\|_1^{\mathcal{G}}, \\ \text{s.t. } \quad &W \geq 0, H \geq 0, \end{aligned} \quad (9)$$

where the regularization parameter λ controls the smoothness of the new representation.

3.2 Updating Rules

The objective function O_1 and O_2 of GSNMF in Eq. (8) and Eq. (9) are not convex in both W and H together. Therefore it is unrealistic to expect an algorithm to find the global minima for O_1 and O_2 . In the following, we introduce two iterative algorithms which can achieve local minima for the objective function O_1 and O_2 , respectively.

We first discuss how to minimize the objective function O_1 . Since $W \geq 0, H \geq 0$, we introduce the Lagrangian multiplier $\phi \in \mathfrak{R}^{m \times r}$ and $\varphi \in \mathfrak{R}^{r \times n}$, thus the Lagrangian function is

$$L = \|X - WH\|_F^2 + \lambda \sum_{j=1}^n \|h_j\|_1^{\mathcal{G}} + \text{tr}(\phi W^T) + \text{tr}(\varphi H^T), \quad (10)$$

where $\text{tr}(\cdot)$ is the trace of a matrix.

Setting $\frac{\partial L}{\partial W_{it}} = 0$ and $\frac{\partial L}{\partial H_{tj}} = 0$, we obtain

$$2(WHH^T - XH^T)_{it} + \phi_{it} = 0, \quad (11)$$

$$2(W^TWH - W^TX)_{tj} + \lambda \frac{H_{tj}}{\sqrt{\sum_{t,\alpha \in \mathcal{G}_k} H_{\alpha j}^2}} + \varphi_{tj} = 0. \quad (12)$$

Using the KKT condition $\phi_{it}W_{it} = 0$ and $\varphi_{tj}H_{tj} = 0$, we get

$$(WHH^T - XH^T)_{it}W_{it} = 0, \quad (13)$$

$$\left((W^TWH - W^TX)_{tj} + \frac{\lambda H_{tj}}{2\sqrt{\sum_{t,\alpha \in \mathcal{G}_k} H_{\alpha j}^2}} \right) H_{tj} = 0. \quad (14)$$

These equations lead to the following updating rules

$$H_{ij} \leftarrow H_{ij} \frac{(W^T X)_{ij}}{(W^T W H)_{ij} + \frac{\lambda H_{ij}}{2 \sqrt{\sum_{t, \alpha \in \mathcal{G}_k} H_{\alpha j}^2}}}, \quad (15)$$

$$W_{it} \leftarrow W_{it} \frac{(X H^T)_{it}}{(W H H^T)_{it}}. \quad (16)$$

For the divergence formulation of GSNMF, we also have two updating rules which can achieve a local minimum of Eq. (9)

$$H_{ij} \leftarrow H_{ij} \frac{\sum_i W_{it} X_{ij} / (W H)_{ij}}{\sum_i W_{it} + \frac{\lambda H_{ij}}{\sqrt{\sum_{t, \alpha \in \mathcal{G}_k} H_{\alpha j}^2}}}, \quad (17)$$

$$W_{it} \leftarrow W_{it} \frac{\sum_j H_{ij} X_{ij} / (W H)_{ij}}{\sum_j H_{ij}}. \quad (18)$$

The convergence of the two updating formula can be proved using the similarity method in [12] with the auxiliary function.

3.3 GSNMF Initialization

In theory, NMF unavoidably converges to local minima [12], and the solution to Eq. (8) and Ep. (9) is not unique under such constraints. We will give a proper initialization in order to improve the performance of GSNMF either for consideration of computational complexity or for the learning of multiple linear manifolds. We aim to learn multiple linear manifolds to represent face images. Thus, the initialization is given as follows. For each class, we learn a basis matrix $W_i \in \mathfrak{R}^{m \times p}$, ($i = 1, \dots, K$) from the training data by NMF. Then W can be initialized by

$$W = [W_1, W_2, \dots, W_K], \quad (19)$$

where K is the number of manifolds. At the same time, in order to enforce the multiple linear manifolds learning, the corresponding coefficient matrix is initialized by

$$H_{ij} = \begin{cases} \delta/p & [i/p] = [j/q] \\ (1 - \delta)/(r - p) & otherwise \end{cases}, \quad (20)$$

where $\lceil x \rceil$ is the ceiling function which rounds x to the nearest integers towards infinity, p is the dimension of each manifold, q is the number of the training data for each class, $r = K \times p$ and $\delta > 0$ is close to 1. Experiments show that GSNMF with the proposed initialization has good convergence for multiple linear manifolds learning.

4 Experiments

In this section, we investigate the performance of our proposed GSNMF method for face recognition. In addition, we experimentally verify our claim that multiple linear manifolds can be learned by the group sparsity constraint.

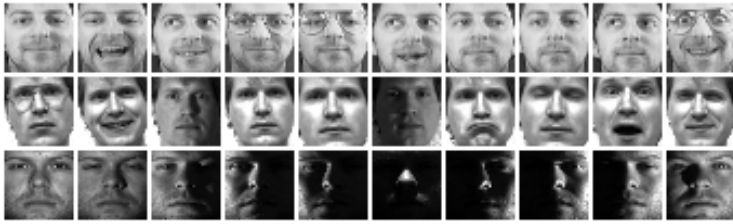


Figure 1: Some sample face images. The top row is from ORL database, the middle row is from Yale database, and the bottom row is from Extended Yale B database.

Table 1: Face Recognition accuracy on the ORL database (mean \pm std-dev%)

Methods	2Train	3Train	4Train	6Train	7Train	8Train
PCA	66.9 \pm 3.5	76.6 \pm 2.3	82.1 \pm 2.2	88.6 \pm 2.3	91.3 \pm 2.4	92.6 \pm 2.4
NMF	65.8 \pm 3.6	72.3 \pm 2.7	76.1 \pm 3.2	81.7 \pm 2.7	84.6 \pm 3.5	87.2 \pm 3.5
SNMF	64.0 \pm 3.1	70.7 \pm 2.7	75.2 \pm 3.0	81.8 \pm 2.9	84.8 \pm 3.0	87.7 \pm 3.0
GSNMF	77.3\pm2.9	85.4\pm2.2	89.5\pm2.0	93.1\pm2.3	94.5\pm1.9	95.3\pm2.3

4.1 Data Set

Three real face databases are used in the experiments: the ORL face database, the Yale face database and the Extended Yale B database. The important statistics of these databases are summarized below.

The ORL face database¹: There are 10 images for each of the 40 human subjects, which were taken at different times, varying the lightings, facial expressions and facial details. The original images have size 92×112 , which are cropped and resized to 32×32 for efficiency; The Yale face database² [1]: It contains 11 gray scale images for each of the 15 individuals. The images demonstrate variations in lighting condition, facial expression and with/without glasses. The images were resized to 32×32 ; The Extended Yale B face database [13]: It contains 2414 front-view face images of 38 individuals. For each individual, about 64 images were taken under various laboratory controlled lighting conditions. The images are also cropped and resized to 32×32 for efficiency.

The preprocessed databases are directly downloaded from the web³. Figure 1 shows some face images from the three face databases used here. For each database, we randomly divide it into training and testing sets, and evaluate the recognition accuracy on the testing set. In detail, for each individual in the ORL and Yale databases, $t = 2, 3, 4, 6, 7, 8$ images are randomly selected as training samples, and the rest are used for testing, while for each individual in the Extended Yale B database, $t = 5, 10, 20, 30, 40, 50$ images are randomly selected as training samples. The training set is used to learn multiple linear manifolds, and the recognition is performed in the multi-manifolds using the simplest nearest neighbor (1-NN) classifier. Since the training set is randomly chosen, for each t Train, we repeat each experiment 50 times and calculate the average recognition accuracy.

¹<http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data>

²<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

³<http://www.cs.uiuc.edu/homes/dengcai2>

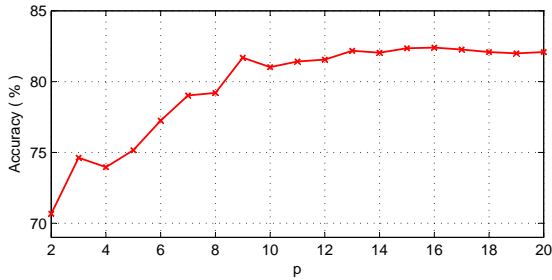


Figure 2: The performance of GSNMF vs. the parameter p on the Yale database ($t=8$). (p is the dimension of each manifold, and $\lambda=10$.)

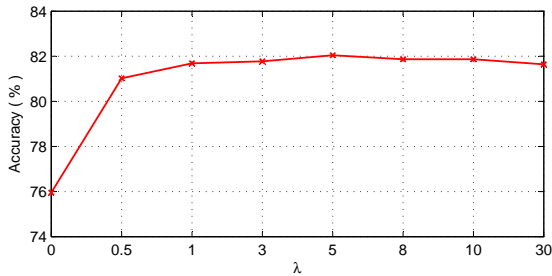


Figure 3: The performance of GSNMF vs. the parameter λ on the Yale database ($t=8$). (λ is the parameter of the group sparsity regularization, and $p=10$.)

4.2 Parameter Settings

There are two main parameters in our GSNMF approach: the dimension of each manifold p and the regularization parameter λ . Figure 2 and Figure 3 show how the average performance of GSNMF varies with the parameters p and λ on the Yale database ($t=8$), respectively. As we can see, when the dimension p of each manifold is high enough, the performance of GSNMF is stable. GSNMF achieves consistently good performance when p is higher than about 10 on the Yale database. At the same time, the performance of GSNMF is very stable with respect to the parameter λ . GSNMF achieves consistently good performance when λ varies from 1 to 30 on the Yale database. For later experiments, we set the dimension p of each manifold to 10, the value of the regularization parameter λ to 10. To demonstrate how the recognition performance can be improved by our method, we compared GSNMF with other three popular dimensionality reduction algorithms: Principal Component Analysis (PCA), the original Nonnegative Matrix Factorization (NMF) [11] and the Sparse NMF (SNMF) [9]. The multiplicative update rule for GSNMF converges fast, and we run 200 iterations including NMF and SNMF.

4.3 Classification Results

After the matrix factorization, we gain a nonnegative basis matrix W and the coefficient matrix H . Ideally, each column in the matrix W should represent a human subject. We visualize these basis learned by PCA, NMF, SNMF and GSNMF in Figure 4. We can easily see that GSNMF learns private representation for each class varying in facial expression,

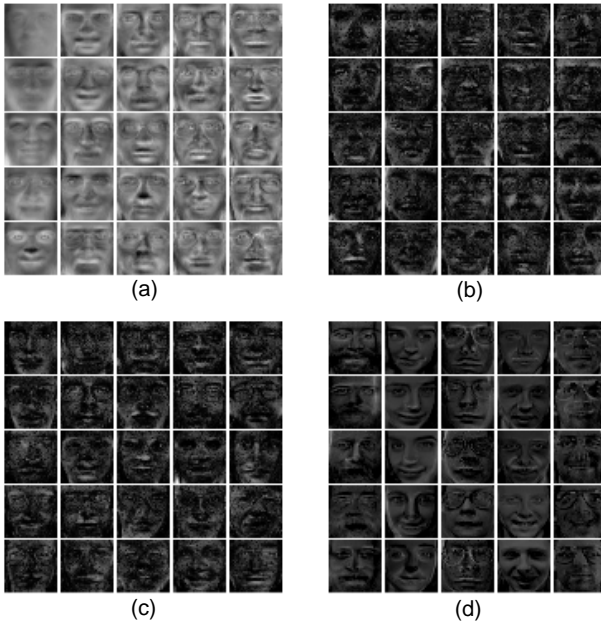


Figure 4: Basis vectors learned from the face images in ORL database. (a) shows the first 25 Eigenfaces, (b) shows 25 NMF faces, (c) shows 25 SNMF faces, and (d) shows 25 GSNMF faces (Each column belongs to a manifold varying in facial expression, lighting and so on.)

Table 2: Face Recognition accuracy on the Yale database (mean \pm std-dev%)

Methods	2Train	3Train	4Train	6Train	7Train	8Train
PCA	43.4 \pm 3.9	49.4 \pm 4.2	52.7 \pm 4.0	58.7 \pm 4.7	60.2 \pm 4.9	63.6 \pm 5.1
NMF	39.9 \pm 3.7	44.6 \pm 3.7	45.9 \pm 4.5	48.9 \pm 4.8	50.7 \pm 6.5	52.1 \pm 5.8
SNMF	38.9 \pm 3.9	43.7 \pm 3.9	44.4 \pm 5.8	49.4 \pm 5.0	50.3 \pm 5.6	51.4 \pm 4.9
GSNMF	56.6\pm4.4	67.8\pm3.6	72.5\pm4.6	79.1\pm4.4	79.9\pm3.7	82.2\pm5.2

eye wear, pose and lighting (Each column in Figure 4 (d) belongs to a manifold). The main reason is that group sparsity constraint imposed on the coefficient matrix H leads to multiple linear manifolds in W .

The face recognition accuracy of different algorithms on the ORL, Yale and Extended Yale B databases are reported on the Table (1), (2) and (3) respectively. For each t Train, we average the results over 50 random splits and report the mean as well as the standard deviation. The performance usually varies with the number of dimensions. In our experiments, we set the same dimension of GSNMF (the whole dimension of multiple linear manifolds) with PCA, the traditional NMF and SNMF. From the results in the Table (1), (2) and (3), we can see that our algorithm GSNMF outperforms PCA, the traditional NMF and SNMF.

4.4 Group Sparsity in H

NMF only allows additive combinations between the basis vectors and it is believed that this property enables NMF to learn a parts-based representation [11]. SNMF adds the sparseness constraints to the classical NMF to obtain sparse data representation [9]. GSNMF imposes

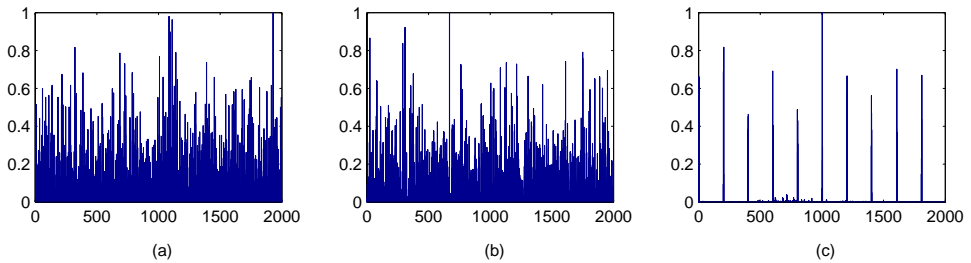


Figure 5: The first 2000 values of the coefficient matrix H learned by (a) NMF, (b) SNMF, (c) GSNMF (Good group sparsity in H is corresponding to the good performance of multiple linear manifolds learning.)

Table 3: Face Recognition accuracy on the Extended Yale B database (mean \pm std-dev%)

Methods	5Train	10Train	20Train	30Train	40Train	50Train
PCA	36.4 \pm 1.6	53.6 \pm 1.1	69.6 \pm 1.1	77.4 \pm 1.2	81.7 \pm 1.0	84.2 \pm 1.5
NMF	69.3 \pm 2.3	83.4 \pm 1.0	89.6 \pm 0.7	91.7 \pm 0.8	92.6 \pm 0.7	93.2 \pm 1.1
SNMF	68.6 \pm 2.2	83.6 \pm 1.1	90.4 \pm 0.6	92.6 \pm 0.6	93.5 \pm 0.7	93.9 \pm 0.8
GSNMF	74.8\pm2.2	86.5\pm1.1	92.0\pm0.8	94.0\pm0.7	94.9\pm0.6	95.1\pm0.9

the group sparsity constraint on H to learn multiple linear manifolds for face recognition. Figure 5 shows how the sparse performance of H learned on ORL database by NMF, SNMF and GSNMF, respectively. For the proposed GSNMF, good group sparsity in H is corresponding to the good performance of multiple linear manifolds learning. In Figure 5 (c), we can see that the proposed GSNMF gets good group sparsity in H .

5 Conclusions and Future Work

We have presented a novel method for multiple linear manifolds learning, called Group Sparse Non-negative Matrix Factorization (GSNMF). GSNMF models the data space as hybrid linear modeling by group sparsity and non-negative matrix factorization. It has more better performance for face recognition than PCA, NMF and SNMF on the three public face recognition databases. Experimental results show that GSNMF achieves accurate recognition on face images with varying illumination and expression. One question remains to be investigated in our future work: Sparse representation based on learned multiple manifolds can be implemented in recognition phase to improve accuracy [18]. Another question is to theoretically analyze the influence of the two parameters in GSNMF.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61001139 and No. 60873133), 863 Program (No. 2008AA02Z310), 973 Program (No. 2009CB320901), Fundamental Research Funds for the Central Universities (No. 2009B27214), Hohai University Funds (No. 2084/409279).

References

- [1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 2002.
- [2] S. Bengio, F. Pereira, Y. Singer, and D. Strelow. Group sparse coding. In *Advances in Neural Information Processing Systems*, 2009.
- [3] D. Cai, X. He, and J. Han. Graph regularized non-negative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.
- [4] G. Chen and G. Lerman. Spectral curvature clustering (SCC). *International Journal of Computer Vision*, 81(3):317–330, 2009. ISSN 0920-5691.
- [5] A. Cichocki and S. Choi. Nonnegative tucker decomposition with alpha-divergence. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 1829–1832, 2008.
- [6] Ronald A. DeVore, Tamara G. Kolda, Gilad Lerman, and Guillermo R. Sapiro. Hot topics workshop: Multi-manifold data modeling and applications, 2008. URL <http://www.ima.umn.edu/2008-2009/SW10.27-30.08/>.
- [7] Q. Gu and J. Zhou. Neighborhood preserving nonnegative matrix factorization. In *Proceedings of the 20th British Machine Vision Conference*, 2009.
- [8] PO Hoyer. Non-negative sparse coding. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 557–565, 2002. ISBN 0780376161.
- [9] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004. ISSN 1532-4435.
- [10] J. Huang and T. Zhang. The benefit of group sparsity. *The Annals of Statistics*, 38(4): 1978–2004, 2010. ISSN 0090-5364.
- [11] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. ISSN 0028-0836.
- [12] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, 2001.
- [13] K.C. Lee, J. Ho, and D.J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 684–698, 2005.
- [14] L. Meier, S. van de Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B*, 70(1):53–71, 2008. ISSN 1369-7412.
- [15] B. Shen and L. Si. Nonnegative matrix factorization clustering on multiple manifolds. In *Proceedings of the 24th Conference on Artificial Intelligence*, pages 575–580, 2010.
- [16] M.E. Tipping and C.M. Bishop. Mixtures of probabilistic principal component analyzers. *Journal of Neural Computation*, 11(2):443–482, 1999.

-
- [17] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.
- [18] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2008.