

Sparse Non-Negative Matrix Factorization based on Spatial Pyramid Matching for Face Recognition

Xianzhong Long, Hongtao Lu, Yong Peng
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, 200240, PR China
 {lxz85, htlu, pengyong851012}@sjtu.edu.cn

Abstract—The non-negative matrix factorization (NMF) is a part-based image representation method which allows only additive combinations of non-negative basis components. NMF has been widely used as a dimensionality reduction technique to solve problems in computer vision and pattern recognition fields. The sparse representation and spatial information of image are also important, however, existing NMF methods do not take these two aspects into consideration simultaneously. In this paper, we propose a novel NMF method with spatial information for face recognition, which is called sparse non-negative matrix factorization based on spatial pyramid matching (SNMFSPM). Experimental results on several benchmark databases show that the proposed scheme outperforms some classical methods.

Keywords—Spatial Pyramid Matching, Sparse Non-Negative Matrix Factorization, Scale Invariant Feature Transform, Face Recognition

I. INTRODUCTION

Face recognition is one of the most challenging tasks in computer vision and pattern recognition fields. We usually represent a face image of size $m \times n$ pixels by an $m \times n$ dimensional vector. However, these $m \times n$ dimensional vectors are too large to allow fast processing. In order to resolve this problem, many dimensionality reduction techniques have been proposed, such as Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Non-negative Matrix Factorization (NMF) [3] and Locality Preserving Projections (LPP) [4] etc. Some corresponding projection matrices are generated after using these methods mentioned above. Each column of these projection matrices is a basis image, so the dimensionality reduction techniques are used to learn the representation of a face as linear combination of basis images. The basis images of PCA are orthogonal and have a statistical interpretation as the directions of the largest variance of data. LDA tries to find a linear transformation that can maximize the between-class scatter matrix and meanwhile minimize the within-class scatter matrix. NMF factorizes training data matrix into a product of two matrix factors whose entries are all non-negative, which produces a part-based representation of images because it allows only additive, not subtractive,

combinations of basis images. LPP seeks to preserve the intrinsic geometry of the data and local structure.

In this paper, we propose a sparse non-negative matrix factorization method based on spatial pyramid matching (SNMFSPM) for face recognition. There are three main reasons for presenting our method. First, sparse representation of image has been proposed recently in machine learning and many competitive results have been obtained [5]. The second reason is that PCA and LDA try to preserve the global structure, however, in many real-world applications, the local structure is more important. NMF is local approach which represents the original image using a set of basis images. Some experimental results showed that when we consider local effects, such as occlusions, changes in expressions or changes in illumination, PCA is not able to represent them as well as NMF [6]. The third one is that the spatial pyramid matching (SPM) model is a good method considering the spatial information of images, which has been successfully applied in general image classification tasks and some very good results have been achieved [7], [8]. However, to our knowledge, the spatial pyramid matching model has not been applied in the face recognition.

II. SPARSE NON-NEGATIVE MATRIX FACTORIZATION BASED ON SPATIAL PYRAMID MATCHING (SNMFSPM)

Let \mathbf{X} be a data matrix of n m -dimensional samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, i.e., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$. Each column of \mathbf{X} represents a face image with m dimensions. Usually, the value of m is very large and this may lead to slow computation speed and low recognition accuracy. NMF [3] decomposes a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ into a product of two matrices $\mathbf{W} \in \mathbb{R}^{m \times r}$ and $\mathbf{H} \in \mathbb{R}^{r \times n}$ ($r \ll \min(m, n)$), i.e., $\mathbf{X} \approx \mathbf{WH}$. In our paper, we name the multiplicative update rules based on minimizing Euclidean distance as ED_NMF and name the multiplicative update rules based on Kullback-Leibler divergence as KL_NMF respectively. In this section, a new additional constraint is imposed on the coefficient matrix \mathbf{H} , we first introduce the sparse non-negative matrix factorization and then combine it with the spatial pyramid matching model.

A. Sparse Non-Negative Matrix Factorization

Sparse non-negative matrix factorization based on minimizing Euclidean distance is as follows:

$$\min_{\mathbf{W}, \mathbf{H}} \|\mathbf{X} - \mathbf{WH}\|_F^2 + \lambda \|\mathbf{H}\|_1 \quad \text{s.t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0 \quad (1)$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm and $\|\mathbf{H}\|_1$ denotes the L_1 norm of \mathbf{H} , which is the sum of the absolute values of matrix entries, $\lambda \geq 0$ is a parameter controlling the strength of sparse term. The nonnegativity constraint in (1) makes the L_1 norm of \mathbf{H} simplified to the sum of its entries, i.e., $\|\mathbf{H}\|_1 = \sum_{s=1}^r \sum_{j=1}^n \mathbf{H}_{sj}$, which is continuously differentiable. Thus, we can calculate the gradient of the objective functions in (1) with respect to \mathbf{H}_{sj} and derive additive update rules based on the conventional gradient descent approach, then we choose specific stepsizes in a similar way as in [9] to derive the multiplicative update rules, the final update rules of sparse non-negative matrix factorization are as follows:

$$\mathbf{H}_{sj} \leftarrow \mathbf{H}_{sj} \frac{(\mathbf{W}^T \mathbf{X})_{sj}}{(\mathbf{W}^T \mathbf{WH})_{sj} + \lambda} \quad \mathbf{W}_{is} \leftarrow \mathbf{W}_{is} \frac{(\mathbf{XH}^T)_{is}}{(\mathbf{WHH}^T)_{is}} \quad (2)$$

Another sparse non-negative matrix factorization based on minimizing KL divergence is as follows:

$$\min_{\mathbf{W}, \mathbf{H}} D(\mathbf{X} \parallel \mathbf{WH}) + \lambda \sum_{s=1}^r \sum_{j=1}^n \mathbf{H}_{sj} \quad (3)$$

$$\text{s.t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0$$

where $D(\mathbf{X} \parallel \mathbf{WH}) = \sum_{i=1}^m \sum_{j=1}^n (\mathbf{X}_{ij} \ln \frac{\mathbf{X}_{ij}}{(\mathbf{WH})_{ij}} - \mathbf{X}_{ij} + (\mathbf{WH})_{ij})$ is the Kullback-Leibler divergence between matrices \mathbf{X} and \mathbf{WH} . Similarly, we can get the corresponding multiplicative update rules of (3):

$$\mathbf{H}_{sj} \leftarrow \mathbf{H}_{sj} \frac{\sum_{i=1}^m \mathbf{W}_{is} \mathbf{X}_{ij} / (\mathbf{WH})_{ij}}{\sum_{i=1}^m \mathbf{W}_{is} + \lambda} \quad (4)$$

$$\mathbf{W}_{is} \leftarrow \mathbf{W}_{is} \frac{\sum_{j=1}^n \mathbf{H}_{sj} \mathbf{X}_{ij} / (\mathbf{WH})_{ij}}{\sum_{j=1}^n \mathbf{H}_{sj}}$$

In our paper, we name the multiplicative update rules (2) and (4) as ED_SNMFSMP and KL_SNMFSMP respectively.

B. Spatial Pyramid Matching Model

The spatial pyramid matching (SPM) model was proposed by Lazebnik et al. [7], before it was proposed, the Bag of Features (BoF) model has been widely used in image representation. However, BoF model represents an image as an orderless collection of local features, ignoring the spatial information of images. So it does not possess enough descriptive capability. The SPM model is an extension of BoF model which takes the spatial information into account, and has recently been proved to achieve better classification than BoF for object classification and scene recognition

tasks [7], [8]. In the traditional SPM model, there are five steps for the framework of object recognition, i.e., local descriptors extraction, dictionary learning, feature coding, spatial pooling and classification.

III. EXPERIMENTAL RESULTS

A. Experiment Settings

In the classical methods, such as PCA, LDA, LPP and NMF, gray values were used to represent each image. All the uncropped face images used in our experiments are first manually resized to a resolution of 32×32 , with 256 gray levels per pixel. The pixel values are then scaled to $[0, 1]$. Each face image is represented as a 1024-dimensionality vector. However, in the SNMFSPM methods, SIFT features are used to describe an image. Specifically, we use the 128 dimensional SIFT descriptor which densely extracted from image patches on a grid with step size of 6 pixels with patch size 16×16 . We resize the maximum size (length or width) of each image to 300 pixels. The sparse NMF in the SNMFSPM is used to learn dictionary and code feature. It is noteworthy that the respective eigenvectors are used as a projection matrix in the PCA, LDA and LPP method respectively, but the projection matrix in NMF is $(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T$ rather than the \mathbf{W} [10]. In our SNMFSPM, we use the multiplicative update algorithms (2) and (4) to learn dictionary and code feature where the parameters $\lambda = 1.5$ and $r = 30$. For the SPM model, we use 3 layers and get total 21 subregions from an image. At last, we use the same linear SVM as classifier in the PCA, LDA, LPP, NMF and SNMFSPM respectively. We randomly select some images per class as training data and use the rest as testing data. For getting a more stable estimation of recognition accuracy, all the results for each group of training data and testing data are repeated 50 times. The average accuracy and the standard deviation are reported. All experiments are conducted in MATLAB, which is executed on a server with an Intel X5650 CPU (2.66GHz and 12 cores) and 32GB RAM.

B. Databases Description

- The ORL face database¹ consists 400 images of 40 different subjects in PGM format. Each subject has 10 images.
- The Yale face database² contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject.
- The Georgia Tech face database³ contains 750 images of 50 different subjects and is stored in JPG format. For each individual, there are 15 color images.

¹<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

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

³http://www.anefian.com/face_reco.htm

Table I
RECOGNITION ACCURACY ON THE ORL FACE DATABASE (MEAN \pm STD-DEV)%

Train number Method	3 Train	4 Train	5 Train	6 Train	7 Train	8 Train
PCA	89.70 \pm 2.4	93.42 \pm 1.7	95.30 \pm 1.7	95.81 \pm 1.5	97.02 \pm 1.4	97.65 \pm 1.7
LDA	86.64 \pm 2.2	89.05 \pm 2.0	92.60 \pm 2.1	92.21 \pm 1.9	92.57 \pm 2.2	92.70 \pm 2.7
LPP	70.36 \pm 3.7	74.92 \pm 2.6	79.22 \pm 3.2	84.01 \pm 2.9	86.52 \pm 3.0	87.65 \pm 2.9
ED_NMF	85.06 \pm 2.3	88.21 \pm 2.1	89.91 \pm 2.3	91.96 \pm 2.4	93.03 \pm 2.2	93.55 \pm 2.4
KL_NMF	85.32 \pm 2.1	88.90 \pm 2.4	90.08 \pm 2.0	91.91 \pm 2.2	92.92 \pm 2.0	93.63 \pm 2.8
ED_SNMFSPPM	92.97 \pm1.8	95.69 \pm1.8	96.97 \pm1.6	98.46 \pm1.0	98.65 \pm1.0	98.85\pm1.1
KL_SNMFSPPM	93.61 \pm1.9	95.58 \pm2.0	97.40 \pm1.1	98.53 \pm0.9	98.78 \pm1.0	98.90\pm1.2

Table II
RECOGNITION ACCURACY ON THE YALE FACE DATABASE (MEAN \pm STD-DEV)%

Train number Method	3 Train	4 Train	5 Train	6 Train	7 Train	8 Train
PCA	85.90 \pm 3.8	88.72 \pm 3.8	91.27 \pm 3.3	93.01 \pm 3.0	93.50 \pm 3.3	94.27 \pm 3.5
LDA	89.05 \pm 3.1	92.80 \pm 2.0	94.31 \pm 3.4	96.16 \pm 1.9	96.07 \pm 2.2	97.16 \pm 2.6
LPP	83.13 \pm 4.3	86.38 \pm 3.9	87.58 \pm 3.3	88.32 \pm 2.9	89.57 \pm 3.6	89.38 \pm 3.0
ED_NMF	79.48 \pm 3.5	83.05 \pm 4.4	84.82 \pm 4.8	86.43 \pm 3.8	88.20 \pm 4.2	90.31 \pm 4.3
KL_NMF	80.80 \pm 6.0	84.57 \pm 3.5	87.47 \pm 3.4	87.79 \pm 4.0	90.33 \pm 3.7	90.62 \pm 4.5
ED_SNMFSPPM	92.03 \pm2.1	94.70 \pm2.3	95.82 \pm2.0	96.83 \pm1.7	97.17 \pm2.1	98.49\pm1.6
KL_SNMFSPPM	93.68 \pm2.6	95.37 \pm2.5	96.71 \pm2.1	97.81 \pm1.9	98.37 \pm1.6	98.67\pm1.7

Table III
RECOGNITION ACCURACY ON THE GEORGIA TECH FACE DATABASE (MEAN \pm STD-DEV)%

Train number Method	7 Train	8 Train	9 Train	10 Train	11 Train	12 Train
PCA	74.94 \pm 2.1	76.45 \pm 1.8	77.35 \pm 2.4	78.17 \pm 2.6	80.20 \pm 2.4	80.75 \pm 3.3
LDA	49.09 \pm 1.9	48.19 \pm 2.0	47.83 \pm 2.3	46.92 \pm 2.7	45.77 \pm 2.7	44.03 \pm 3.9
LPP	50.88 \pm 2.3	54.14 \pm 1.8	57.67 \pm 2.4	59.31 \pm 2.4	62.98 \pm 2.9	63.97 \pm 2.9
ED_NMF	65.43 \pm 2.1	67.15 \pm 2.1	68.39 \pm 2.2	69.36 \pm 2.8	71.26 \pm 2.9	71.33 \pm 3.2
KL_NMF	66.07 \pm 2.2	67.55 \pm 2.3	68.87 \pm 2.8	70.98 \pm 2.5	71.33 \pm 2.8	72.92 \pm 3.3
ED_SNMFSPPM	86.37 \pm1.7	87.66 \pm1.7	88.76 \pm1.6	90.22 \pm1.5	91.00 \pm1.8	91.52\pm2.4
KL_SNMFSPPM	87.07 \pm1.4	88.25 \pm1.8	89.85 \pm1.8	90.30 \pm1.7	90.87 \pm1.9	91.77\pm1.9

Table IV
RECOGNITION ACCURACY ON THE CALTECH FACE DATABASE (MEAN \pm STD-DEV)%

Train number Method	2 Train	5 Train	8 Train	11 Train	14 Train	17 Train
PCA	25.66 \pm 2.6	40.06 \pm 3.4	47.92 \pm 3.2	52.50 \pm 3.1	56.59 \pm 3.9	59.14 \pm 5.2
LDA	28.15 \pm3.4	42.73 \pm 2.5	48.81 \pm 3.1	51.99 \pm 2.8	54.01 \pm 4.2	56.49 \pm 5.2
LPP	24.49 \pm 2.6	23.88 \pm 2.6	27.05 \pm 2.5	29.39 \pm 3.1	31.42 \pm 3.8	35.82 \pm 5.2
ED_NMF	25.01 \pm 3.0	32.76 \pm 2.6	37.98 \pm 2.3	41.78 \pm 2.8	43.80 \pm 3.7	47.98 \pm 5.1
KL_NMF	24.47 \pm 3.1	32.70 \pm 3.1	38.79 \pm 3.0	40.53 \pm 3.0	43.49 \pm 4.2	45.81 \pm 5.2
ED_SNMFSPPM	27.51 \pm 2.4	47.08 \pm3.0	59.50 \pm3.0	67.09 \pm3.1	71.82 \pm3.7	76.17 \pm4.3
KL_SNMFSPPM	28.59 \pm2.1	46.57 \pm2.7	59.39 \pm2.8	66.62 \pm3.2	71.70 \pm3.2	75.21 \pm5.1

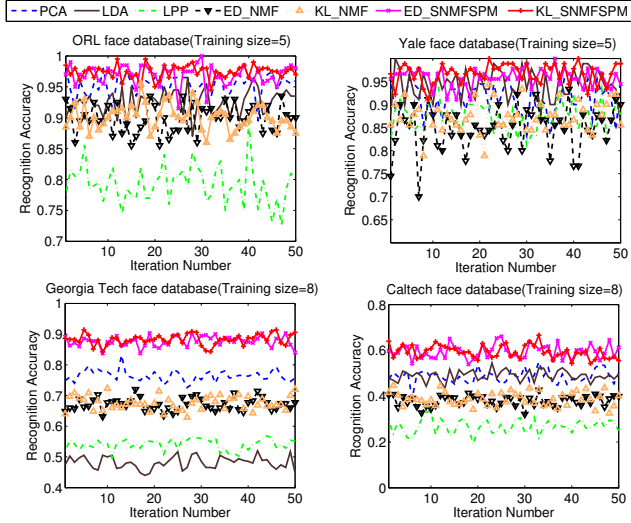


Figure 1. Recognition accuracy comparison of PCA, LDA, LPP, NMF and SNMFSPM on four well-known face databases. The x-axis denotes 50 different iterations and the y-axis is the corresponding recognition accuracy with respect to 50 different iterations

- The Caltech face database⁴ contains 450 images in JPG format. We select a subset from this database to test all algorithms proposed in our paper, which contains 412 images of 19 different subjects with different lighting, expressions and backgrounds.

C. Performance Evaluations and Comparisons

We report the mean accuracy and standard deviation of the 50 different runs for each database with different training numbers from Table 1 to Table 4. In all the tables, the ‘Train number’ denotes the number of training samples which are randomly selected from each class. It can be concluded that our SNMFSPM achieves the best performance. Especially for Caltech Face database, our SNMFSPM still outperforms other methods although the recognition accuracy of all the methods are low because the backgrounds in this database are cluttered. It is worth noting that with increase of training number, the recognition accuracy of our SNMFSPM increases more significantly and achieves much better results than the other methods. On the Georgia database, the SNMFSPM achieves more than 10% improvement in all cases over the best of the other methods. Furthermore, the standard deviation of SNMFSPM is lower than other methods, which implies that our SNMFSPM is more stable than other four methods. Figure 1 illustrates the recognition accuracy comparisons of PCA, LDA, LPP, NMF and SNMFSPM on the four face databases mentioned above.

⁴<http://www.vision.caltech.edu/html-files/archive.html>

IV. CONCLUSIONS

In this paper, an efficient sparse non-negative matrix factorization method based on spatial pyramid matching is proposed. We have compared our scheme for face recognition with four classical methods, including PCA, LDA, LPP and NMF. Experiments on four databases demonstrate that recognition accuracy of our SNMFSPM is much better than the previous classical methods.

ACKNOWLEDGMENT

This work is supported in part by the National Basic Research Program of China (973 program) under Grant 2009CB320901, the National Natural Science Foundation of China under Grant 61272247, and the Innovation Ability Special Fund of Shanghai Jiao Tong University under Grant Z030026.

REFERENCES

- [1] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
- [2] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [3] D. Lee and H. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [4] P. Niyogi, “Locality preserving projections,” *Advances in neural information processing systems*, vol. 16, pp. 153–160, 2004.
- [5] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Verlag, 2010.
- [6] D. Guillamet and J. Vitrià, “Non-negative matrix factorization for face recognition,” *Topics in Artificial Intelligence*, pp. 336–344, 2002.
- [7] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 2169–2178.
- [8] J. C. Yang, K. Yu, Y. H. Gong, and T. Huang, “Linear spatial pyramid matching using sparse coding for image classification,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1794–1801.
- [9] D. Lee and H. Seung, “Algorithms for non-negative matrix factorization,” *Advances in neural information processing systems*, vol. 13, 2001.
- [10] N. Guan, D. Tao, Z. Luo, and B. Yuan, “Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent,” *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2030–2048, 2011.