

DOUBLY WEIGHTED NONNEGATIVE MATRIX FACTORIZATION FOR IMBALANCED FACE RECOGNITION

Jiwen Lu and Yap-Peng Tan

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

We propose in this paper a novel doubly weighted nonnegative matrix factorization (DWNMF) method for imbalanced face recognition. Motivated by the fact that some face samples and certain parts of each face sample are more useful for recognition, we construct two weighted matrices based on the pairwise similarity of face samples in the same class and the discriminant score of each face pixel. Compared with the existing NMF algorithm, the proposed DWNMF method can more effectively exploit the discriminative and geometrical information of face samples, and it is especially suitable for imbalanced face recognition. Experimental results are presented to demonstrate the efficacy of the proposed method.

Index Terms— Face recognition, manifold structure, nonnegative matrix factorization, subspace learning.

1. INTRODUCTION

Subspace learning-based face recognition has been widely studied and numerous algorithms have been proposed during the past two decades. The most representative such algorithms include principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], and locality preserving projections (LPP) [3]. While these algorithms have achieved good success in face recognition, they mainly make use of face features as a whole and don't explicitly emphasize face samples and parts differently. There are psychological and physiological evidences that human recognize objects through part-based representations in the brain [4]. Based on this belief, a new subspace learning method called nonnegative matrix factorization (NMF) [4, 5] has been recently proposed in the literature to impose nonnegative constraints in the basis and representation coefficients, leading to better representation of part-based face features such as mouth, nose and eyes, which have more discriminative information than other parts such as cheek and forehead as indicated in [9]. A number of NMF extensions, such as DNMF [6], LNMF [7], and TPNMF [8], have also been proposed to enhance the representation capability and classification performance of NMF.

Existing NMF-based algorithms, however, consider each training sample contributing equally in the learning phase and ignore the fact that the high similarity among some training samples will result in large redundancy in the basis. In real-world applications of face recognition, it is common to have different number of useful training face samples due to pose and illumination variations. Hence, a recognition algorithm that can deal with imbalanced face samples and place proper emphasis on different face samples and their parts is of great interest and significance.

Motivated by aforementioned observations, we propose in this paper a doubly weighted nonnegative matrix factorization (DWNMF) method that can impose larger weights to face samples which have low similarity with others and apply different emphases on different face parts. Being able to weight the samples and their parts differently according to their importance on recognition allows the proposed DWNMF method to handle imbalanced training samples more effectively. Our empirical results show that the proposed method can better exploit the discriminative and geometrical information of face features, resulting in better face recognition performance, particularly for cases with imbalanced training samples.

2. BRIEF REVIEW OF NMF

Given a nonnegative $n \times m$ matrix V and a rank parameter r , the NMF model finds a nonnegative $n \times r$ matrix W and another nonnegative $r \times m$ matrix H such that their product WH approximates to V , i.e., $V \approx WH = \sum_{k=1}^r W_{ik}H_{kj}$, where $i = 1, 2, \dots, n, j = 1, 2, \dots, m$. Usually, r is chosen to be smaller than $\min\{m, n\}$. To find a good approximate factorization, one can minimize the total square error between V and WH with nonnegative constraints defined as follows:

$$\begin{aligned} \min_{W, H} \{ \|V - WH\|^2 &= \sum_{i=1}^n \sum_{j=1}^m (V_{ij} - (WH)_{ij})^2 \} \\ \text{s.t. } W, H &\geq 0 \end{aligned} \quad (1)$$

To solve this nonconvex optimization problem, the classical expectation maximization (EM) algorithm can be applied to find the local minima and more details can be found in [5].

3. THE PROPOSED DWNMF METHOD

The proposed DWNMF method makes use of two weighting matrices to reflect the discriminative and geometrical information of face features. The first matrix, referred to as between-sample weighting matrix A_b , is to impose different weights on different face samples to reflect their respective importance in the learning phase. The other matrix, referred to as within-sample weighting matrix A_w , is to give appropriate weights to pixels at different face parts. As both between-sample and within-sample features are weighted simultaneously, we therefore name our proposed algorithm a “doubly weighted” method.

3.1. Between-Sample Weighting Matrix

As high similarity among the training samples will result in much redundancy in the basis of NMF (W), introducing a proper weighting matrix is an effective way to reduce this redundancy. The basic rule should be that “the higher the similarity between two samples, the smaller the weight should be given”. Figure 1 provides one toy example to illustrate this rule. The figure shows three classes of samples denoted by squares, circles and triangles, and each class has different number of samples. It can be seen that the number samples in the circle or square class is less than that in the triangle class, and there exists higher correlation among samples in the triangle class. Hence, smaller weights should be assigned to samples in the circle class. Furthermore, different weights should also be assigned to the samples in the same class. For example, in the circle class, as the distance between sample S_1 and sample S_3 is much smaller than that of S_1 and S_2 , smaller weights should be given to S_1 and S_3 compared with that of S_1 and S_2 . To effectively characterize the similarity and structure of the samples, we devise the following algorithm to construct the between-sample weighting matrix.

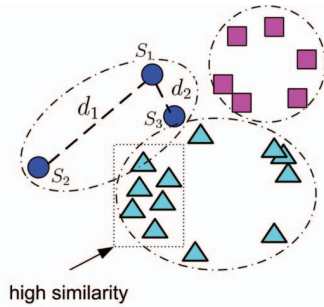


Fig. 1. Toy example showing the different similarity and structure in each class.

Given a set of properly normalized $w \times h$ face images, we construct one training set of column image vectors X_{ij} , where $X_{ij} \in \mathbb{R}^{d=wh}$, by lexicographic ordering the pixels of the j th image of the i th subject. Assume that there are n different

subjects and the c th subject has n_c samples. The algorithm for constructing A_b is described below:

- (1) For each sample in the c th class, $c = 1, 2, \dots, n$, calculate the similarity between each pair of samples in the class;

$$s_{i,j}^c = S(X_{ci}, X_{cj}), i \neq j \quad (2)$$

We use the following similarity metric:

$$S(X_{ci}, X_{cj}) = \exp^{-\|X_{ci} - X_{cj}\|^2 / \sigma} \quad (3)$$

where σ is a suitable constant and the elements of $S(X_{ci}, X_{cj})$ are all in range $[0, 1]$.

- (2) Find the maximum similarity value S_M for X_{ij} among all pairwise similarities between X_{ij} and other samples in the i th class,

$$S_M(X_{ij}, :) = \max\{s_{i,j}^c\} \quad (4)$$

where $j = 1, 2, \dots, n_i$ and $j \neq i$.

- (3) Assign weight a_{ij} to sample X_{ij} as follows:

$$a_{ij} = 1 - S_M(X_{ij}, :) \quad (5)$$

- (4) Normalize a_{ij} to make all the weights sum to unity, i.e., $\sum_{i=1}^n \sum_{j=1}^{n_i} a_{ij} = 1$:

$$a_{ij} = \frac{1}{n_i} \frac{a_{ij}}{\sum_{i=1}^n \sum_{j=1}^{n_i} a_{ij}} \quad (6)$$

- (5) Construct an $N \times N$ diagonal weighting matrix A_b with diagonal element a_{ij} , where $N = \sum_{i=1}^n n_i$ and

$$A_b = \text{diag}(a_{11}, a_{12}, \dots, a_{1n_1}, a_{21}, \dots, a_{nn_n}).$$

3.2. Within-Sample Weighting Matrix

Owing to their different influences on face recognition, different parts of face features should be assigned different weights to achieve better discrimination performance. Intuitively, some face parts such as eyes, nose and mouth may have more discriminant information than other parts such as cheek and forehead. We propose as follows a pixel-region fusion weighting method to exploit the discriminant information of different face parts:

- (1) Apply a feature selection scheme based on a new manifold-based maximum margin (MMM) criterion [10], described below, to calculate the discriminant capability of each pixel in the face samples.

Let f_{ri} denote the r th feature of the i th face sample x_i , where $i = 1, 2, \dots, N$ and $r = 1, 2, \dots, m$, and let L_r denote the MMM score of the r th feature. We first construct a graph G to describe the locality structure of the data

set. For each sample x_i , we find its k nearest neighbors and put an edge between x_i and its neighbors. Let $N(x_i) = \{x_i^1, \dots, x_i^k\}$ be the set of the k nearest neighbors of x_i . It follows that G can be defined as

$$G_{ij} = \begin{cases} 1 & \text{if } x_i \in N(x_j) \text{ or } x_j \in N(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

To better characterize the locality and discriminant information of the data set, we split $N(x_i)$ into two subsets: $N_b(x_i)$ and $N_w(x_i)$, s.t., $N_b(x_i) \cap N_w(x_i) = \emptyset$ and $N_b(x_i) \cup N_w(x_i) = N(x_i)$, where $N_w(x_i)$ contains the neighbors with labels the same as that of x_i and $N_b(x_i)$ contains the neighbors with different labels. Then, derive two matrices G^c and G^p as follows:

$$G_{ij}^c = \begin{cases} \rho & \text{if } x_i \in N(x_j) \text{ or } x_j \in N(x_i) \text{ and } l(x_i) = l(x_j) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

and

$$G_{ij}^p = \begin{cases} 1 & \text{if } x_i \in N(x_j) \text{ or } x_j \in N(x_i) \text{ and } l(x_i) \neq l(x_j) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where $l(x_i)$ and $l(x_j)$ denote the class labels of x_i and x_j , and the values of k and ρ are empirically set to 10 and 50, respectively.

(2) Define two diagonal matrices D^c and D^p , where $D_{ii}^c = \sum_j G_{ij}^c$ and $D_{ii}^p = \sum_j G_{ij}^p$, and calculate the MMM matrices as $H^c = D^c - G^c$ and $H^p = D^p - G^p$. Then, compute the pixel-based MMM score for the r th feature:

$$L_r = \frac{f_r^T H^c f_r}{f_r^T H^p f_r} \quad (10)$$

where $r = 1, 2, \dots, m$.

(3) For each face image x_i with size of $w \times h$, partition it into P $k \times l$ non-overlapping small patches $x_i = (x_i^1, x_i^2, \dots, x_i^P)$, as shown in Figure 2.

Let $m_q = (m_q^1, m_q^2, \dots, m_q^N)^T$ be the average of the q th patch of all the samples. Apply the similar approach to calculate the patch-based MMM score for each q patch:

$$R_q = \frac{m_q^T H^c m_q}{m_q^T H^p m_q} \quad (11)$$

where $q = 1, 2, \dots, P$.

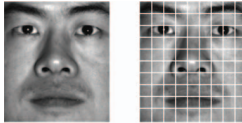


Fig. 2. Partitioning a face image into patches.

(4) For each position (i, j) in the face, we find its pixel-based MMM score L_{ij} and patch-based score R_{ij} , where

$L_{ij} = L(x)$, $R_{ij} = R(y)$, $x = (i - 1) \times w + j$, and $y = ([(i - 1)/k] + 1) \times k + ([(j - 1)/l] + 1)$. Operator $[c]$ obtains the integer part of c .

(5) For each pixel $f(i, j)$ in each face image, we combine the pixel-based and patch-based MMM scores to obtain a fused score as follows:

$$t_{i,j} = L_{ij} R_{ij} \quad (12)$$

where $1 \leq i \leq w$ and $1 \leq j \leq h$.

(6) Construct a $wh \times wh$ matrix A_w with diagonal elements t_{ij} 's, i.e., $A_w = \text{diag}(t_{1,1}, t_{2,1}, \dots, t_{w,1}, \dots, t_{w,h})$.

3.3. The Algorithm

We formulate the proposed DWNMF algorithm by integrating the between-sample and within-sample weighting matrices into the original NMF method as follows:

(1) Given N training samples X_{ij} containing n different subjects, and each subject has n_i samples of size $w \times h$. Organize the given training samples into one $N \times wh$ matrix V , with each column of which containing one training sample.

(2) Construct the $N \times N$ between-sample weighting matrix A_b for V . Multiply V with A_b to obtain V' , i.e.,

$$V' = V A_b \quad (13)$$

(3) Perform NMF for V' to obtain the $wh \times r$ bases W and $r \times N$ coefficients H , respectively.

(4) Construct the $wh \times wh$ within-sample weighting matrix A_w , and then multiple A_w with W to obtain the new bases W' :

$$W' = A_w W \quad (14)$$

(5) Project the training and testing samples on W' and extract their feature representations, respectively.

4. EXPERIMENTS AND RESULTS

We have used the Extended Yale B [11] and CMU PIE [12] face databases to evaluate the effectiveness of our proposed DWNMF method for face recognition. The Extended Yale B database consists of 2432 gray images of 38 human subjects with different 64 illumination conditions, while the PIE database comprises 68 subjects with 41368 face images of different poses and illumination conditions. In our experiments, we selected 45 frontal images of each subject in the PIE database.

We constructed imbalanced training sets for experiments on imbalanced face recognition. For the Extended Yale B database, we randomly divided the 38 subjects into four non-overlapping groups—two with 10 subjects and two with 9 subjects—and randomly selected 5, 5, 10, and 10 images, respectively, for each subject in the four different groups as the training set. For the PIE database, we similarly divided the database into four different groups of 17 subjects each. For

each group, we randomly selected 10, 15, 20 and 25 images of each subject to construct the training set. The remaining images in the two databases were used to construct the testing set. The nearest neighbor classifier is applied for the recognition and the experiments were run randomly 20 times to obtain the average recognition accuracy.

We compared our proposed DWNMF method with other popular subspace learning algorithms including PCA [1], LDA [2], LPP [3] and NMF [4, 5]. The best results of the comparing algorithms were obtained by exploring all possible feature dimensions, and the results along with their standard deviation are tabulated in Table 1. Figure 3 shows the recognition performance versus different dimension in the reduced subspace.

Table 1. Comparison on recognition performance using the Extended Yale B and the PIE database (mean \pm std).

Method	Extended Yale B		PIE	
	Accuracy (%)	Dim	Accuracy (%)	Dim
PCA	40.97 \pm 3.4	340	62.39 \pm 2.5	140
LDA	59.09 \pm 2.8	37	67.52 \pm 1.8	67
LPP	57.06 \pm 2.3	220	66.55 \pm 1.6	120
NMF	55.94 \pm 2.7	180	60.75 \pm 2.1	100
DWNMF	63.87 \pm 1.8	200	67.72 \pm 1.9	100

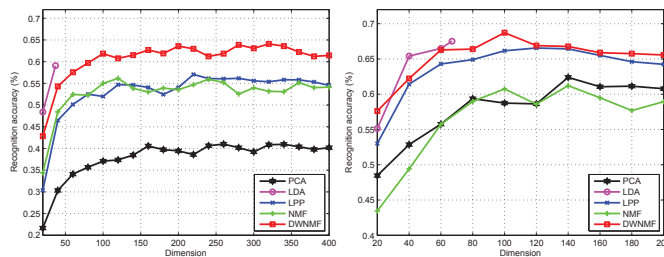


Fig. 3. Average recognition accuracy versus reduced dimension. The left is on the Extended Yale B database and the right is on the PIE database.

We have made the following observations from the experimental results:

1) The proposed DWNMF notably outperforms PCA and LPP on all experiments. The superiority of DWNMF stems from the fact that both discriminative and geometric information are more effectively exploited in DWNMF to improve the recognition performance.

2) The proposed DWNMF also outperforms LDA, which may due to the fact that DWNMF doesn't need the assumption of Gaussian distribution on the samples as required by LDA.

3) Compared with NMF, DWNMF also has better recognition performance. The reason may be that DWNMF can effectively reduce the redundancy of the basis generated from similar training samples, a property that makes DWNMF particularly suitable for imbalanced face recognition.

5. CONCLUSIONS

We have proposed in this paper a novel NMF-based subspace learning algorithm, called doubly weighted nonnegative matrix factorization (DWNMF), for efficient face recognition with imbalanced number of training samples. The DWNMF method works by using two weighting matrices, between-sample and within-sample weighting matrices, to better exploit the discriminative and geometric information of samples for imbalanced face recognition. The proposed doubly weighed technique can likely be extended to other subspace learning algorithms.

6. REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [2] P. N. Belhumeur, J. P. Hefanah, and D. J. Kriegman, "Eigenfaces vs. fisherface: recognition using class specific linear projection," *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 711-720, 1997.
- [3] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianface," *IEEE Trans. on PAMI*, vol. 27, no. 3, pp. 328-340, 2005.
- [4] D. D. Lee, and S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401: pp. 788-791, 1999.
- [5] D. D. Lee, and S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS'2001*, pp. 556-562, 2001.
- [6] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with applications to frontal face verification," *IEEE Trans. on Neural Networks*, vol. 17, no. 3, pp. 683-695, 2006.
- [7] S. Z. Li, X. Hou, H. J. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proc. CVPR'2001*, pp. 207-212, 2001.
- [8] T. Zhang, B. Fang, Y. Y. Tang, G. He, and J. Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE Trans. on Image Processing*, vol. 17, no. 4, pp. 574-584, 2008.
- [9] X. Xie, and K. M. Lam, "Gabor-based kernel PCA with doubly nonlinear mapping for face recognition with a single face image," *IEEE Trans. on Image Processing*, vol. 15, no. 9, pp. 2481-2492, 2006.
- [10] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. on Knowledge and Data Engineering*, vol. 20, no. 2, pp. 189-201, 2008.
- [11] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on PAMI*, vol. 23, no. 6, pp. 643-660, 2001.
- [12] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illuminant, and expression database," *IEEE Trans. on PAMI*, vol. 25, no. 9, pp. 1615-1618, 2003.