6-20-2019

Sparse general non-negative matrix factorization based on left semi-tensor product

Zigang Chen

Lixiang Li

Haipeng Peng

Yuhong Liu
Santa Clara University, yhliu@scu.edu

Haihua Zhu

See next page for additional authors

Recommended Citation

Chen, Z., Li, L., Peng, H., Liu, Y., Zhu, H., & Yang, Y. (2019). Sparse General Non-Negative Matrix Factorization Based on Left Semi-Tensor Product. IEEE Access, 7, 81599–81611. https://doi.org/10.1109/ACCESS.2019.2924140

Received June 2, 2019, accepted June 8, 2019, date of publication June 20, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924140

Sparse General Non-Negative Matrix Factorization Based on Left Semi-Tensor Product

ZIGANG CHEN[1,2], LIXIANG LI[2], HAIPENG PENG[2], YUHONG LIU[3], HAIHUA ZHU[2], AND YIXIAN YANG[2]

[1]College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China
[2]Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053, USA

Corresponding author: Lixiang Li (lixiang@bupt.edu.cn)

This work was supported in part by the National Key R&D Program of China under Grant 2016YFB0800602, and in part by the National Natural Science Foundation of China under Grant 61771071 and Grant 61573067.

ABSTRACT The dimension reduction of large scale high-dimensional data is a challenging task, especially the dimension reduction of face data and the accuracy increment of face recognition in the large scale face recognition system, which may cause large storage space and long recognition time. In order to further reduce the recognition time and the storage space in the large scale face recognition systems, on the basis of the general non-negative matrix factorization based on left semi-tensor (GNMFL) without dimension matching constraints proposed in our previous work, we propose a sparse GNMFL/L (SGNMFL/L) to decompose a large number of face data sets in the large scale face recognition systems, which makes the decomposed base matrix sparser and suppresses the decomposed coefficient matrix. Therefore, the dimension of the basis matrix and the coefficient matrix can be further reduced. Two sets of experiments are conducted to show the effectiveness of the proposed SGNMFL/L on two databases. The experiments are mainly designed to verify the effects of two hyper-parameters on the sparseness of basis matrix factorized by SGNMFL/L, compare the performance of the conventional NMF, sparse NMF (SNMF), GNMFL, and the proposed SGNMFL/L in terms of storage space and time efficiency, and compare their face recognition accuracies with different noises. Both the theoretical derivation and the experimental results show that the proposed SGNMF/L can effectively save the storage space and reduce the computation time while achieving high recognition accuracy and has strong robustness.

INDEX TERMS Machine learning, unsupervised learning, semi-tensor product (STP), sparse general non-negative matrix factorization (SGNMF).

I. INTRODUCTION

Face recognition is an important research problem in computer vision, and it is widely used in banking, security, human-computer interaction and smart device apps. A face recognition system is a computer application capable of identifying or verifying a person from a digital image or a video frame. Given a face image, one of the most popular ways to recognize the face is using machine learning methods to identify similar images in a face database based on the selected face features.

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou.

"The trouble with facial recognition technology (in the real world)" on The Conversation on December 14, 2016 between Robin Kramer and Kay Ritchie states that, as of 2016, face recognition is still not effective for most applications even though the accuracy has been substantially improved. Although the systems are often advertised as having accuracy close to 100%, they usually use much smaller sample sizes than what would be necessary for large scale applications [1].

Large scale face recognition systems are still facing many challenges. One of the challenges is to achieve a certain level of recognition accuracy, the large scale face data often requires very high dimensional face features, causing large storage space and long recognition time. In addition, continuous improvements of image resolution provide even more information in images, resulting in further increment of high dimensional features. Therefore, the dimensionality reduction becomes very important for face data processing.

Non-negative matrix factorization (NMF) [2], [3] is widely accepted and proven to be effective to handle the dimensionality reduction, and can also be used to structure low-dimensional latent factor (LF) model/space [4]–[9] to build a low-rank approximation to the target matrix. It factorizes an input non-negative matrix into the product of two non-negative matrices with lower ranks, so that the major parts of the face can be learned more efficiently. When NMF is used for face recognition, the massive amounts of face training data are represented by a non-negative matrix $X \in R_+^{s\times t}$, where s represents the number of the selected features from one face image, and t is the number of the persons. NMF factorizes $X \in R_+^{s\times t}$ into a non-negative basis matrix $W \in R_+^{m\times n}$ and a coefficient matrix $H \in R_+^{p\times q}$, so that $X \approx WH$. Specifically, NMF requires the dimension matching constraints to be hold on the matrix multiplications, indicating that $m = s$, $n = p$, and $q = t$. Each column vector of H denotes the weights when approximating the corresponding column of X by using the bases from W. The coefficient matrix H then substitutes the original matrix X, and it can be used by machine learning schemes for face recognition. When $p \ll s$, we can achieve the face feature dimensional reduction.

The face recognition methods based on NMF [2], [10], incremental NMF (INMF) [11], [12] and Sparse NMF (SNMF) [13]–[18] are helpful to improve the speed or the accuracy of face recognition. However, none of them could further reduce the dimensionality of the basis matrix W and the coefficient matrix H, which plays an important role in influencing the recognition time (computation time) and the storage space.

In our previous work, we proposed a general non-negative matrix factorization (GNMF) [19], [20] which removed the dimension matching constraints on matrix multiplications, so that the dimensionality of the basis matrix and the coefficient matrix could be further reduced. Specifically, given a matrix $X^{s\times t}$, the proposed GNMF allowed flexible selections of the values of n and p, so that the number of columns in W did not need to match the number of rows in H. Then the values of m (i.e. the row number of W) and q (i.e. the column number of H) were further determined based on the value of the least common multiple l of n and p as $m = s \cdot n/l$ and $q = t \cdot p/l$. Through this process, we got $m \ll s$ and (or) $q \ll t$, which could effectively save the storage space of W or H.

In order to further reduce the recognition time and the storage space, inspired by the idea of SNMF, we propose the sparse GNMF to further reduce the computation time and improve the accuracy of face recognition in the large scale face recognition systems.

The major contributions of this work are given as follows:

1) We propose a sparse GNMF based on left STP (SGNMFL) to factorize a matrix $X \in R_+^{s\times t}$ into a sparse basis matrix $W \in R_+^{m\times n}$ and a coefficient matrix $H \in R_+^{p\times q}$, which removes the dimension matching constraints on the two factor matrices.
2) Our SGNMFL can help in saving the storage space and reducing the computation time by increasing the sparseness of the basis matrix. We perform lots of experiments on the JAFFE database and ORL database, so as to verify the effects of two hyper-parameters on the sparseness of the basis matrix factorized by SGNMFL/L, compare the performance of the conventional NMF, SNMFL, GNMFL and the proposed SGNMFL/L in terms of storage space and time efficiency, and compare their accuracies with different noises. Experimental results show that when achieving similar face recognition accuracy, the proposed SGNMFL/L can further save more storage space and more reduce the computation time.

The rest of this paper is organized as follows. Section II introduces the related works. Section III presents the proposed SGNMFL. Section IV presents the face recognition process based on GNMF and SGNMFL, followed by experiment results in Section V. Section VI concludes this paper.

II. RELATED WORKS

NMF decomposes the input non-negative matrix into two low-order non-negative matrices, both of which can learn part of the face. Since NMF was proposed, the non-negative matrix factorization algorithm has been applied to face feature dimensionality reduction and face recognition [2].

NMF factorizes the massive amounts of face training data X into two non-negative matrices W and H, so that $X \approx WH$. Given a target face image, the selected features of this image are represented by a 1-D column vector x with s elements. Then the dimensionality of this image can be reduced as $(W^TW)^{-1}W^Tx = h$, where h has p elements. Then machine learning schemes can be applied to recognize the most similar images from the database. For example, by computing the Euclid norm between h and each column of the corresponding coefficient matrix H, we can retrieve the image with the minimal Euclid norm. The we can identify the corresponding object in the training image.

The performance of NMF is reduced when NMF is used to deal with nonlinear data in complex environments, such as face images with changes in pose, illumination and face expression during face recognition. In order to overcome the limitations of the above-mentioned non-negative matrix decomposition algorithm, Chen *et al.* [10] proposed a supervised non-linear method to improve the classification ability of non-negative matrix decomposition in face recognition.

However, the conventional NMF is not designed to reduce storage space or computational complexity. Dynamic updating data sets will greatly increase the time and space overhead of NMF, which becomes a challenge issue in dynamically updating new samples through online training. The dynamic learning method based on incremental NMF (INMF) could solve this problem and be applied to the face recognition system. Yu *et al.* [11] proposed a subspace incremental method called incremental graph regularized nonnegative matrix factorization which imposed the manifold into INMF, so as to keep the geometric structure unchanged in spatial domain, save the running time and improve the face recognition rate. Chen *et al.* [12] provided a NMF face recognition method and system based on kernel machine learning in order to improve the accuracy of face recognition.

Later, the sparse NMF (SNMF) was presented and used to improve the accuracy of face recognition. To explicitly impose sparseness on the matrix W and H, Hoyer [14] proposed a formula to measure the sparseness of the vector x as shown in the following Equation (1).

$$sparseness(x) = \frac{\sqrt{r} - \|x\|_1/\|x\|_2}{\sqrt{r} - 1}, \tag{1}$$

where r is the dimensionality of x, $\|x\|_1$ and $\|x\|_2$ are the L_1-norm and the L_2-norm of the vector x, respectively, and $0 \leq sparseness(x) \leq 1$. Equation (1) evaluates to one if and only if x contains only a single non-zero component, and it takes a value of zero if and only if all components are non-zero and equal (up to signs).

NMF with sparse constraints is to add more constraints on the original objective function to obtain the decomposition information as sparse as possible. Pauca *et al.* [15] proposed a constrained NMF (CNMF). The objective function with the above constraints is given as follows:

$$\begin{aligned} min\, L_{CNMF}(W,H) &= \sum_{i=1}^{m}\sum_{j=1}^{q}([X_{ij} - (WH)_{ij}]^2) \\ &\quad + \alpha \sum_{i=1}^{m}\sum_{k=1}^{n} W_{ik}^2 + \beta \sum_{k=1}^{n}\sum_{j=1}^{q} H_{kj}^2, \\ s.t.\ W, H \geq 0,\ &\sum_{i=1}^{m} W_{ik} = 1, \end{aligned} \tag{2}$$

where $1 \leq k \leq n$. α and β in Equation (2) are regularization non-negative parameters to be determined. The sum of the elements of each vector in the non-negative matrix W is required to be one during the iterations.

To impose sparseness constraints on W, Kim and Park [16] proposed SNMF/L, where L denotes the sparseness imposed on the left factor. The sparse NMF formulation imposes the sparseness on a factor of NMF by utilizing L_1-norm minimization based on non-negativity constrained least squares. The detailed process is given as follows:

$$\begin{aligned} min\, L_{SNMF/L}(W,H) &= \sum_{i=1}^{m}\sum_{j=1}^{q}([X_{ij} - (WH)_{ij}]^2) \\ &\quad + \alpha \sum_{i=1}^{m}(\sum_{k=1}^{n}|W(i,k)|)^2 \\ &\quad + \beta \sum_{k=1}^{n}\sum_{j=1}^{q} H_{kj}^2, \\ s.t.\ W, \quad H \geq 0,\ &1 \leq k \leq n, \end{aligned} \tag{3}$$

where $\alpha > 0$ is a regularization parameter to balance the trade-off between the accuracy of the approximation and the sparseness of W, and $\beta > 0$ is a parameter to suppress $\|H\|_F^2$. The SNMF/L begins with the initialization of H with non-negative values.

Since the existing studies on NMF showed that when the constraints were added [13]–[17], the sparser factor matrices would lead to better decomposition quality. Liu *et al.* [13] combined sparse coding and NMF into SNMF, which could learn both parts-based representation and much sparser representation. Hoyer [14] combined NMF and sparse coding into Non-Negative Sparse Coding (NNSC), which made the decomposed coefficients much sparser. One drawback of the NNSC algorithm is that the basis vector is additively updated, which cannot keep non-negative characteristics well. Pauca *et al.* [15] proposed a sparse NMF algorithm by using the least square. Kim and Park [16] proposed SNMF/L and SNMF/R (where L and R denoted the sparseness imposed on the left and right factors, respectively). Shastri and Levine [18] used NNSC in the learning of face features for face recognition. Mairal *et al.* [21] focused on the large scale matrix factorization problem that includes learning the basics set in order to adapt to specific data based on sparse coding method. Liu *et al.* [22] proposed a group SNMF (GSNMF) algorithm to learn multiple linear manifolds for face recognition. An ensemble SNMF process was proposed in [23] to represent data instances in parts and partition the data space into localities, and then the individual classifiers in each locality were coordinated for final classification in videos concept detection. A face aging simulation method based on sparse-constrained method was proposed in [24] and then applied in the age-across face recognition.

With the successful application of NMF and SNMF, more advanced schemes were developed based on them, and were combined with other algorithms for face recognition, such as two-dimensional nonnegative principal component analysis [25], Supervised kernel NMF [10], discriminant non-negative graph embedding [26], Incremental NMF [27], NMF with bounded total variational regularization [28], large margin based NMF and partial least squares regression [29], fishers linear discriminant (FLD) and support vector machine (SVM) in NMF residual space [30], noise modeling and representation based NMF classification methods [31], convergent projective NMF with Kullback-Leibler Divergence [32], learning latent features by NMF combining similarity judgments [33], etc. Based on the clustering method, Xue *et al.* [34] discussed a structured NMF initialization scheme which achieved faster convergence while maintaining the data structure and also obtained good result for the face recognition task. Hu *et al.* [35] proposed a Newton-based algorithm for the conventional NMF in face image processing.

These methods based on NMF, INMF and SNMF are helpful to improve the speed or the accuracy of face recognition in face recognition systems. However, these methods cannot further reduce the dimension of the basis matrix / the coefficient matrix, which directly affects the recognition time (computation time) and the storage space of large scale face recognition system.

III. SPARSE GENERAL NON-NEGATIVE MATRIX FACTORIZATION

A. GENERAL NON-NEGATIVE MATRIX FACTORIZATION BASED ON LEFT SEMI-TENSOR

In this section, we give some necessary preliminaries on the semi-tensor product (STP) [36], [37] of matrices and the STP-based GNMF without dimensional matching constraints [19], [20] in our previous works.

Definition 1: Given two matrices $W \in R^{m\times n}$ and $H \in R^{p\times q}$, the variable l is the least common multiple of n and p. The left STP of matrices W and H is defined as

$$W \ltimes H = (W \otimes I_{l/n})(H \otimes I_{l/p}) \in R^{(m\cdot l/n)\times(l/p\cdot q)}, \tag{4}$$

where $\otimes$ represents the right Kronecker product [38].

Refs [36], [37] and Equation (4) show that the left STP can be applied to perform the conventional matrix product. That is, when $n = p$, $W \ltimes H = W \cdot H$.

Definition 2: Given a non-negative matrix $X \in R_+^{s\times t}$, the GNMF based on left STP (GNMFL) can find non-negative matrices $W \in R_+^{m\times n}$ and $H \in R_+^{p\times q}$, such that:

$$\begin{aligned} X_+^{s\times t} &\approx W_+^{m\times n} \ltimes H_+^{p\times q} \\ &= W_+^{m\times n} \ltimes H_+^{p\times q} + E^{s\times t}, \end{aligned} \tag{5}$$

where $E \in R^{s\times t}$ is a noise matrix, and it is part of the approximate linear mixture model for GNMFL [39].

As [36], [37] and Equation (4) show that the left STP is a general form of the conventional matrix product, GNMFL can also be considered as a general form of the conventional NMF.

Using the conventional gradient method, where $i = 1, 2, \cdots, s$, $j = 1, 2, \cdots, t$, $k = 1, 2, \cdots, l$, we get the following multiplicative update rules

$$\begin{cases} W^{(\tau+1)}_{\lfloor (i-1)/(l/n)\rfloor+1,\lfloor (k-1)/(l/n)\rfloor+1} \\ \leftarrow W^{(\tau)}_{\lfloor (i-1)/(l/n)\rfloor+1,\lfloor (k-1)/(l/n)\rfloor+1} \\ \cdot \dfrac{(X \ltimes H^T)_{ik}}{((W^{(\tau)} \ltimes H) \ltimes H^T)_{ik}} \\ H^{(\tau+1)}_{\lfloor (k-1)/(l/p)\rfloor+1,\lfloor (j-1)/(l/p)\rfloor+1} \\ \leftarrow H^{(\tau)}_{\lfloor (k-1)/(l/p)\rfloor+1,\lfloor (j-1)/(l/p)\rfloor+1} \\ \cdot \dfrac{(W^T \ltimes X)_{kj}}{(W^T \ltimes (W \ltimes H^{(\tau)}))_{kj}}. \end{cases} \tag{6}$$

Equation (6) not only shows that the left STP can be applied to perform the general matrix product, but also shows that GNMFL also can be applied to perform the conventional NMF.

According to Equation (6), when $l/n > 1$, the result of each iteration is $W \otimes I_{l/n} \in R^{m\cdot l/n\times l}$ in the multiplicative update rules. Then an element from each $l/n \times l/n$ sub-block matrix in $W \otimes I_{l/n} \in R^{m\cdot l/n\times l}$ is extracted as an element of the matrix W. When $l/p > 1$, the result of each iteration is $H \otimes I_{l/p} \in R^{l\times q\cdot l/p}$ in the multiplicative update rules. Then an element from each $l/p \times l/p$ sub-block matrix in $H \otimes I_{l/p} \in R^{l\times q\cdot l/p}$ is extracted as an element of the matrix H.

B. SPARSE GENERAL NON-NEGATIVE MATRIX FACTORIZATION BASED ON LEFT STP

In order to further reduce recognition time and storage space in face recognition, inspired by the idea of SNMF, we discuss the sparse GNMFL in this section.

Referring to Equation (3), we propose SGNMFL/L, where L denotes that the sparseness is imposed on the left factor, and the sparse NMF formulation imposes sparseness on a factor of NMF by utilizing L_1-norm minimization based on alternating non-negativity constrained least squares. The detailed process is given as follows:

$$\begin{aligned} min\ L_{SGNMFL/L}(W, H) = &\sum_{i=1}^{m}\sum_{j=1}^{q}([X_{ij} - (W \ltimes H)_{ij}]^2) \\ &+ \alpha \sum_{i=1}^{m}(\sum_{k_1=1}^{n} |W(i, k_1)|)^2 \\ &+ \beta \sum_{k_2=1}^{p}\sum_{j=1}^{q} H_{k_2 j}^2, \\ s.t.\ W, H \geq 0, 1 \leq k \leq n. \end{aligned} \tag{7}$$

where $\alpha > 0$ is a regularization parameter to balance the trade-off between the accuracy of the approximation and the sparseness of W, and $\beta > 0$ is a parameter to suppress $\|H\|_F^2$.

The SGNMFL/L begins with the initialization of H with non-negative values. Then, it iterates the following alternating non-negativity constrained least squares until the convergence is reached:

$$\min_H \left\| \begin{pmatrix} W \\ \sqrt{\beta} I_n \end{pmatrix} \ltimes H - \begin{pmatrix} X \\ 0_{l\times t} \end{pmatrix} \right\|_F^2, \quad s.t.\ H \geq 0, \tag{8}$$

where I_n is an identity matrix with size $n \times n$, $0_{l\times t}$ is a zero matrix with size $l \times t$.

The SGNMFL/L begins with the initialization of W with non-negative values. Then, it iterates the following alternating non-negativity constrained least squares until the convergence is reached:

$$\min_W \left\| \begin{pmatrix} H^T \\ \sqrt{\alpha} e_{1\times p} \end{pmatrix} \ltimes W^T - \begin{pmatrix} X^T \\ 0_{1\times s} \end{pmatrix} \right\|_F^2, \quad s.t.\ W \geq 0, \tag{9}$$

where $e_{1\times p} \in R^{1\times p}$ is a row vector whose components all equal to one, $0_{1\times s} \in R^{1\times s}$ is a zero vector, and Equation (9) minimizes L_1-norm of rows of $W \in R^{m\times n}$, which imposes sparseness on W.

Using the conventional gradient method, where $i = 1, 2, \cdots, s$, $j = 1, 2, \cdots, t$, $k = 1, 2, \cdots, l$, we get the multiplicative update rules as follows

$$\begin{cases} W^{(\tau+1)}_{\lfloor (i-1)/(l/n)\rfloor+1,\lfloor (k-1)/(l/n)\rfloor+1} \\ \leftarrow W^{(\tau)}_{\lfloor (i-1)/(l/n)\rfloor+1,\lfloor (k-1)/(l/n)\rfloor+1} \\ \cdot \dfrac{(X \ltimes H^T)_{ik}}{((W^{(\tau)} \ltimes H) \ltimes H^T + \alpha \times (W^{(\tau)} \ltimes (e^T_{1\times p})))_{ik}} \\ H^{(\tau+1)}_{\lfloor (k-1)/(l/p)\rfloor+1,\lfloor (j-1)/(l/p)\rfloor+1} \\ \leftarrow H^{(\tau)}_{\lfloor (k-1)/(l/p)\rfloor+1,\lfloor (j-1)/(l/p)\rfloor+1} \\ \cdot \dfrac{(W^T \ltimes X)_{kj}}{(W^T \ltimes (W \ltimes H^{(\tau)}) + \beta \times (I_n \ltimes H^{(\tau)}))_{kj}}, \end{cases} \quad (10)$$

where $e_{1\times p} \in R^{1\times p}$ is a row vector whose components all equal to one, I_n is an identity matrix with size $n \times n$, and the variable l is the least common multiple of n and p. According to Equation (10), when $l/n > 1$, the result of each iteration is $W \otimes I_{l/n} \in R^{m\cdot l/n \times l}$ in the multiplicative update rules. Then an element from each $l/n \times l/n$ sub-block matrix in $W \otimes I_{l/n} \in R^{m\cdot l/n\times l}$ is extracted as an element of the matrix W. When $l/p > 1$, the result of each iteration is $H \otimes I_{l/p} \in R^{l\times q\cdot l/p}$ in the multiplicative update rules. Then an element from each $l/p \times l/p$ sub-block matrix in $H \otimes I_{l/p} \in R^{l\times q\cdot l/p}$ is extracted as an element of the matrix H.

Based on the above derivation, we design the iterative updating algorithm for SGNMFL/L in ALGORITHM 1.

C. THE COMPUTATIONAL COMPLEXITY

Next, the computational complexity of the proposed SGNMFL/L is theoretically analyzed. The computational complexity of the algorithm is represented by the number of three arithmetic operations, including addition, multiplication, and division on floating-point numbers.

According to Eqs. (7) and (10), SGNMFL/L decomposes the matrix $X^{s\times t}$ into two non-negative matrices $W^{m\times n}$ and $H^{p\times q}$, such that $X^{s\times t} \approx W^{m\times n} \ltimes H^{p\times q}$, where $s = m \times l/n$, $t = q \times l/p$, and the variable l is the least common multiple of n and p. The three arithmetic operations of SGNMFL/L are as follows:

i) $W^T \ltimes X$ needs $n \times s \times t$ floating-point additions and $n \times m \times t$ floating-point multiplications;

ii) $W^T \ltimes (W \ltimes H)$ needs $n \times s \times t + p \times s \times t$ floating-point additions. if $n \le p$, $W^T \ltimes (W \ltimes H)$ needs $n \times m \times t + s \times n \times t$ floating-point multiplications. if $n > p$, $W^T \ltimes (W \ltimes H)$ needs $n \times m \times t + s \times p \times t$ floating-point multiplications. $\alpha \times (W \ltimes e^T_{1\times p})$ needs $min\{m \times n, m \times p\}$ floating-point additions and $min\{m \times n, m \times p\}$ floating-point multiplications. In addition, $W^T \ltimes (W \ltimes H) + \alpha \times (W \ltimes e^T_{1\times p})$ also needs $s \times l$ floating-point additions;

iii) $\frac{X \ltimes H^T}{W \ltimes H \ltimes H^T + \alpha \times (W \ltimes (e^T_{1\times p}))}$ needs $s \times l$ floating-point divisions;

iv) $X \ltimes H^T$ needs $p \times s \times t$ floating-point additions and $s \times q \times p$ floating-point multiplications;

v) $(W \ltimes H) \ltimes H^T$ needs $n \times s \times t + p \times s \times t$ floating-point additions. if $n \le p$, $(W \ltimes H) \ltimes H^T$ needs $s \times q \times p + s \times n \times t$ floating-point multiplications. if $n > p$, $(W \ltimes H) \ltimes H^T$ needs $s \times q \times p + s \times p \times t$ floating-point multiplications. $\beta \times (I_n \ltimes H)$ needs $p \times q$ floating-point multiplications. In addition, $(W \ltimes H) \ltimes H^T + \beta \times (I_n \ltimes H)$ also needs $l \times t$ floating-point additions;

vi) $\frac{W^T \ltimes X}{W^T \ltimes (W \ltimes H) + \beta \times (I_n \ltimes H)}$ needs $l \times t$ floating-point divisions.

Algorithm 1 The Iterative Updating Algorithm for SGNMFL/L

Input: X, n, p, τ, α, and β.
/* X is the input data. */
/* n is the column number of W, and p is the row number of H. */
/* τ is an appropriate number of iterations, such as 300. */
/* α is a regularization parameter to balance the trade-off between the accuracy of the approximation and the sparseness of W, and $\alpha \in [0, 1]$. */
/* β is a parameter to suppress $\|H\|_F^2$, and $\beta \in [0, 1]$. */
Initialize: $E \leftarrow 1e-9$ /* According to Equation (5).*/
Output: The optimal matrices of W and H.

1) $l \leftarrow lcm(n, p)$
/* l is the least common multiple of n and p. */
2) $(s, t) \leftarrow size(X)$
/* s and t are the row number and the column number of X respectively. */
3) $m \leftarrow s \times n/l$ /* m is the row number of W. */
4) $q \leftarrow t \times p/l$ /* q is the column number of H. */
5) $W \leftarrow abs(rand(m, n))$
/* Random initialization of W. */
6) $H \leftarrow abs(rand(p, q))$
/* Random initialization of H. */
7) $e_{1\times p} \leftarrow (1, 1, \cdots, 1)$
/* $e_{1\times p}$ is a row vector whose components all equal to one. */
8) **for** $i = 1 : \tau$
9) $W \leftarrow W \cdot \frac{X \ltimes H^T}{(W \ltimes H) \ltimes H^T + \alpha \times (W \ltimes (e^T_{1\times p})) + E}$
/* According to Equation (10).*/
10) $H \leftarrow H \cdot \frac{W^T \ltimes X}{W^T \ltimes (W \ltimes H) + \beta \times (I_n \ltimes H) + E}$
/* According to Equation (10).*/
11) **for** $j = 1 : m$
12) $W(j, :) \leftarrow W(j, :)./(sum(W(j, :) + E))$
/* Normalizing each row of the basis matrix. */
13) **end**
14) **end**

IV. FACE RECOGNITION PROCESS BASED ON GNMFL AND SGNMFL/L

In order to understand the face recognition process based on GNMFL and SGNMFL/L better, we need to first recapitulate the face recognition process based on NMF and SNMF/L.

Let

$$X = WH, \quad (11)$$

then we can get

$$(W^T W)^{-1} W^T X = (W^T W)^{-1} W^T WH = H. \quad (12)$$

A face image to be tested is represented by a 1-D column vector x. Face recognition process based on conventional NMF and SNMF/L is given as follows. According to Equation (12), we can first obtain the coefficient vector h of the testing image as $h = (W^T W)^{-1} W^T x$. Then by computing the Euclid norm between h and each column of the corresponding coefficient matrix H, we can identify the column in H with the minimum Euclid norm, representing the most similar image to the testing image.

Similar to the above NMF and SNMF/L processes, we discuss the face recognition process based on the proposed GNMFL and SGNMFL/L as follows. First, let

$$X^{(m\cdot l/n)\times(q\cdot l/p)} = W^{m\times n} \ltimes H^{p\times q}, \quad (13)$$

where, the variable l is the least common multiple of n and p.

Based on [36], [37] and Equation (4), we get

$$X^{(m\cdot l/n)\times(q\cdot l/p)} = (W^{m\times n} \otimes I^{l/n})(H^{p\times q} \otimes I^{l/p}). \quad (14)$$

According to basic properties of Kronecker product and Equation (4), Equation (14) is converted to the following equation,

$$(((W^{m\times n})^T W^{m\times n})^{-1}(W^{m\times n})^T) \ltimes X^{(m\cdot l/n)\times(q\cdot l/p)} = H^{p\times q} \otimes I^{l/p}. \quad (15)$$

When $n = p$, the GNMFL becomes the same as the conventional NMF, and Equation (15) becomes the same as Equation (12).

In order to further reduce recognition time and storage space in face recognition, n (i.e. the column number of W) should be less than p (i.e. the row number of H) and $mod(p, n) = 0$, then $l = p$, thus Equation (15) becomes

$$(((W^{m\times n})^T W^{m\times n})^{-1}(W^{m\times n})^T) \ltimes X^{(m\cdot l/n)\times q} = H^{p\times q}. \quad (16)$$

A face image to be tested is represented by a 1-D column vector x.

According to Equation (16), we have

$$h = (((W^{m\times n})^T W^{m\times n})^{-1}(W^{m\times n})^T) \ltimes x. \quad (17)$$

Then by computing the Euclid norm between h and each column of the corresponding coefficient matrix H, we can identify the column in H with the minimum Euclid norm, representing the most similar image to the testing image. The corresponding object in the testing image is then identified.

V. EXPERIMENTS AND ANALYSES

In order to reduce recognition time (computation time) and save the storage space of the basis matrix in large scale face recognition systems, we aim to reduce m and n of W by using the proposed GNMFL and SGNMFL/L in the case that p and q of H do not change.

In this section, two sets of experiments are conducted to show the effectiveness of our SGNMFL/L on two databases. Experiments are mainly designed to compare the performance of the conventional NMF, SNMF/L, GNMFL and the proposed SGNMFL/L in terms of storage space and time efficiency. The JAFFE database and the ORL database are used as testing data sets.

A. COMPARISON OF STORAGE SPACE AND TIME EFFICIENCY ON JAFFE DATABASE

In this subsection, we conduct face recognition experiments on JAFFE database based on NMF, SNMF/L, GNMFL and SGNMFL/L, respectively, aiming to compare the storage space and sparseness of the basis matrix in the training process, and time efficiency in face recognition process.

Face major component (120×120) is extracted from each original 213 face images (256×256) in the JAFFE database, and it is represented by a 1-D column vector (14400×1). Then we construct the training data $X \in R_+^{14400\times 70}$ through extracting 70 face expression images posed by 10 female models (Each person has 7 expressions).

The training data $X \in R_+^{14400\times 70}$ are factorized into two factor matrices $W \in R_+^{14400\times 60}$ and $H \in R_+^{60\times 70}$ based on the conventional NMF and SNMF/L, and the iteration is set to be 300. These factorized 60 columns of matrix $W \in R_+^{14400\times 60}$ are called basis images (i.e. basis matrix). FIGUREs 1-2 show the experimental results based on the conventional NMF and SNMF/L (where $\alpha = 1$ and $\beta = 1$), respectively. Each 1-D column vector (60×1) of $H \in R^{60\times 70}$ is the coefficient of the corresponding 1-D column vector (14400×1) of $X \in R_+^{14400\times 70}$ based on W. FIGURE 2 shows that the basis images factorized by SNMF/L are sparser.

Next, the training data matrix $X \in R_+^{14400\times 70}$ is factorized into two non-negative matrices $W \in R_+^{3600\times 15}$ and $H \in R_+^{60\times 70}$ based on GNMFL and SGNMFL/L, and the iteration is set to be 300. These factorized 15 columns of matrices $W \in R_+^{3600\times 15}$ are called basis images (i.e. basis matrix).

In order to further verify the effect of two hyper-parameters (i.e. α and β) on the sparseness of W factorized by SGNMFL/L, these two parameters α (i.e. x axis in FIGURE 3) and β (i.e. y axis in FIGURE 3) are adjusted from 0 to 2, and the step size is 0.1. The corresponding experimental results are shown in FIGURE 3, where *sparseness* (i.e. z axis) represents the sparseness of W, which is calculated from Equation (1). FIGURE 3 shows that the *sparseness* increases as α and β get bigger. However, when $\alpha > 1$ or $\beta > 1$, the simulation software matlab prompts the warning "the matrix has singular accuracy" in the face training process. Therefore, $0 \le \alpha \le 1$ and $0 \le \beta \le 1$.

FIGURE 1. Basis images factorized by conventional NMF when $W \in R_{+}^{14400\times 60}$, and *sparseness* = 0.7929.

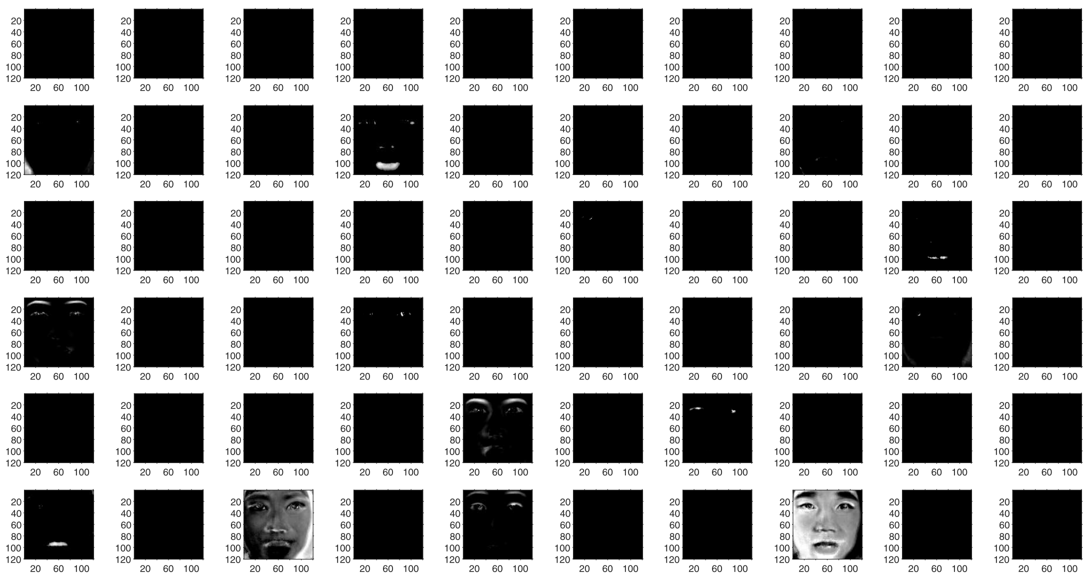


FIGURE 2. Basis images factorized by SNMF/L ($\alpha = 1$ and $\beta = 1$) when $W \in R_{+}^{14400\times 60}$, and *sparseness* = 0.9732.

FIGUREs 4 and 5 show that the experimental results based on GNMFL and SGNMFL/L ($\alpha = 1$ and $\beta = 1$), respectively. Each 1-D column vector (60×1) of $H \in R_{+}^{60\times 70}$ is the coefficient of the corresponding 1-D column vector (14400×1) of $X \in R_{+}^{14400\times 70}$ based on W. FIGURE 5 shows that the basis images factorized by SGNMFL/L are sparser.

Based on the conventional NMF, SNMF/L, GNMFL and SGNMFL/L, the basis matrix and the elapsed times in face training are shown in Table 1. From Table 1, we can see that the elapsed time is greater based on GNMFL and SGNMFL/L ($\alpha = 1$ and $\beta = 1$) than based on NMF and SNMF/L ($\alpha = 1$ and $\beta = 1$), but the basis images factorized by GNMFL and SGNMFL/L are only $3600 \times 15 \ll 14400 \times 60$.

TABLE 1. Comparison of storage space and time efficiency in face training on JAFFE database.

Factorization method	W	Elapsed time(s)
NMF	$W_{+}^{14400\times 60}$	26.431834
SNMF/L	$W_{+}^{14400\times 60}$	26.444202
GNMFL	$\boldsymbol{W}_{+}^{3600\times 15}$	81.988792
SGNMFL/L	$\boldsymbol{W}_{+}^{3600\times 15}$	81.596363

All of 213 face images are test samples. In Table 2, the coefficient matrix $H \in R^{60\times 70}$, and the elapsed time is the average elapsed time of 213 times $(W^T W)^{-1} W^T X_j = H_j$ or $(W^T W)^{-1} W^T \ltimes X_j = H_j$, where, X_j is j-th column

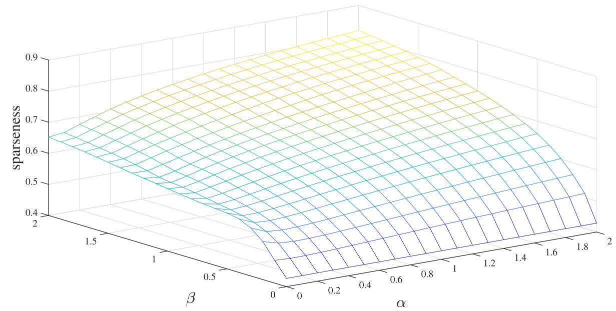


FIGURE 3. The *sparseness* of the basis images factorized by SGNMFL/L when $W \in R_{+}^{3600\times 15}$.

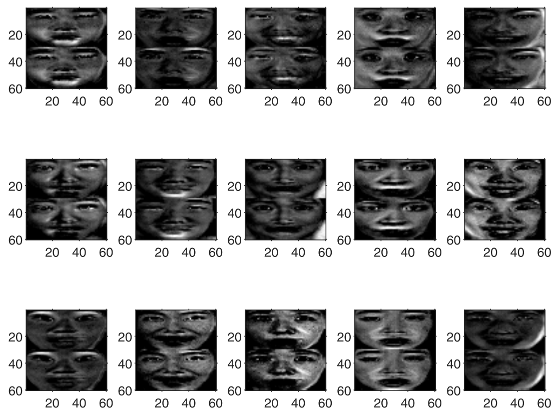


FIGURE 4. Basis images factorized by GNMFL when $W \in R_{+}^{3600\times 15}$, and *sparseness* = 0.4266.

TABLE 2. Comparison of storage space and time efficiency in face recognition on JAFFE database.

Factorization method	Basis matrix	Accuracy	Elapsed time(s)
NMF	$W_{+}^{14400\times 60}$	0.8310	0.0219
SNMF/L	$W_{+}^{14400\times 60}$	0.0093	0.0218
GNMFL	$\mathbf{W_{+}^{3600\times 15}}$	0.8920	**0.0083**
SGNMFL/L	$\mathbf{W_{+}^{3600\times 15}}$	**0.8967**	**0.0081**

vector of X, H_j is j-th column vector of H, and $j \in [1\ 213]$. The accuracy and the elapsed time of several factorization methods in face recognition are shown in Table 2. From Table 2, we can see that compared with NMF, SNMF/L ($\alpha = 1$ and $\beta = 1$) and GNMFL, the proposed SGNMFL/L ($\alpha = 1$ and $\beta = 1$) achieves the highest computational efficiency. Meanwhile, it also achieves a higher face recognition accuracy than GNMFL. Besides, the simulation software matlab prompts the warning "the matrix has singular accuracy" in the face recognition process based on SNMF/L ($\alpha = 1$ and $\beta = 1$). As a result, the accuracy of face recognition is very low.

In order to analyze the robustness of the proposed SGNMFL/L ($\alpha = 1$ and $\beta = 1$), we artificially add four different noises (i.e. Gaussian white noise, Poisson noise, salt & pepper noise, speckle noise) into face images in the processes of face training and recognition. And the parameters of each kind of noise are default parameters. Comparison results of face recognition accuracy with different noises on JAFFE database are shown in Table 3. From Table 3, we can see that the accuracies of face recognition with different noises are close to that of face recognition without noise. That is to say, the proposed SGNMFL/L is robust.

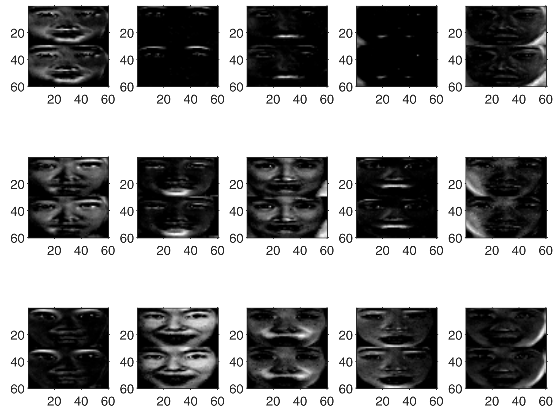


FIGURE 5. Basis images factorized by SGNMFL/L ($\alpha = 1$ and $\beta = 1$) when $W \in R_{+}^{3600\times 15}$, and *sparseness* = 0.6810.

TABLE 3. Comparison of face recognition accuracy with different noises on JAFFE database.

Factorization method	Noise type	Accuracy
SGNMFL/L	Gaussian	0.8175
SGNMFL/L	Poisson	0.8175
SGNMFL/L	Salt & pepper	0.8425
SGNMFL/L	Speckle	0.8967

We also compare the minimum Euclid distance of conventional NMF ($W \in R_{+}^{14400\times 60}$), SNMF/L ($W \in R_{+}^{14400\times 60}$), GNMFL ($W \in R_{+}^{3600\times 15}$) and SGNMFL/L ($W \in R_{+}^{3600\times 15}$). The results are shown in FIGURE 6. According to FIGURE 6, we observe that minimum Euclid norms based on NMF and GNMFL are relatively smaller, and minimum Euclid norms based on SNMF/L and SGNMFL/L are greater. The latter is easier to be recognized, which can effectively reduce the rate of error recognition.

B. COMPARISON OF STORAGE SPACE AND TIME EFFICIENCY ON ORL DATABASE

In this subsection, we conduct face recognition experiments on ORL database by using NMF, SNMF/L, GNMFL and SGNMFL/L respectively, aiming to compare the storage space and the sparseness of the basis matrix in the training process, and time efficiency in face recognition process.

Face major component (34×34) is extracted from each original 400 face images in the ORL face database (40 individuals, each 10 images), and it is represented by a 1-D column vector (1156×1). Then we construct the training data $X \in R_{+}^{1156\times 200}$ through extracting 5 face images posed by each one.

The training data $X \in R_{+}^{1156\times 200}$ is factorized into two factor matrices $W \in R^{1156\times 36}$ and $H \in R^{36\times 200}$ based on the

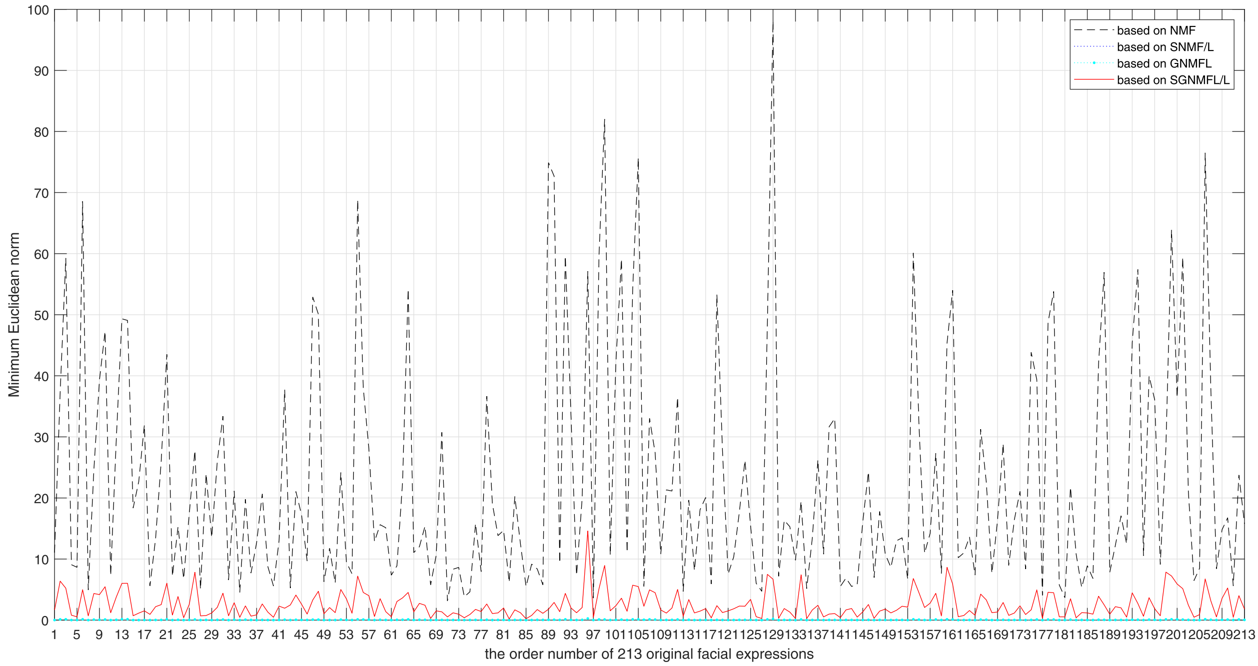


FIGURE 6. Comparison in minimum Euclid norms by using the conventional NMF, SNMF/L, GNMFL and SGNMFL/L on JAFFE database.

FIGURE 7. Basis images factorized by the conventional NMF when $W \in R^{1156\times 36}$, and *sparseness* = 0.6857.

conventional NMF and SNMF/L (where $\alpha = 1$ and $\beta = 1$), and the iteration number is set to be 300. FIGUREs 7 and 8 show that the experimental results based on the conventional NMF and SNMF/L, respectively. Each 1-D column vector (36×1) of H is the coefficient of the corresponding 1-D column vector (1156×1) of X based on W. FIGURE 8 shows that that the basis images factorized by SNMF/L are sparser.

The training data $X \in R_+^{1156\times 200}$ is factorized into two non-negative matrices $W \in R_+^{289\times 9}$ and $H \in R_+^{36\times 200}$ based on GNMFL and SGNMFL/L, and the iteration number is set to be 300.

In order to further verify the effect of two hyper-parameters (i.e. α and β) on the sparseness of the basis matrix decomposed by SGNMFL/L, these two parameters α (i.e. x axis in FIGURE 9) and β (i.e. y axis in FIGURE 9) are adjusted from 0 to 2, and the step size is 0.1. The experimental results are shown in FIGURE 9, where *sparseness* (i.e. z axis) represents the sparseness of W, which is calculated by Equation (1). FIGURE 9 shows that the sparseness increases as α and β get

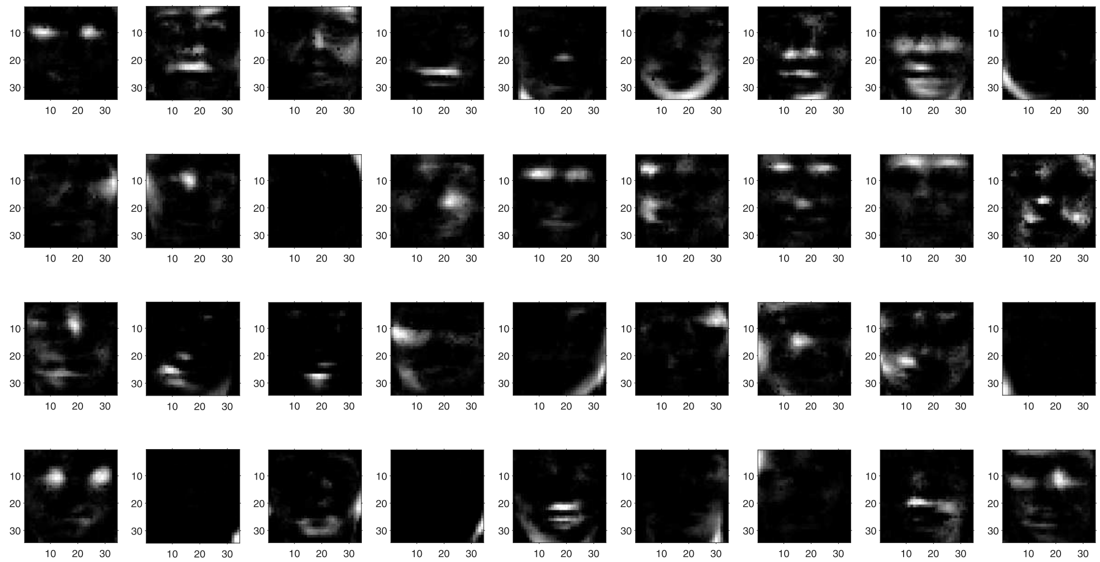


FIGURE 8. Basis images factorized by SNMF/L ($\alpha = 1$ and $\beta = 1$) when $W \in R^{1156\times 36}$, and *sparseness* = 0.8563.

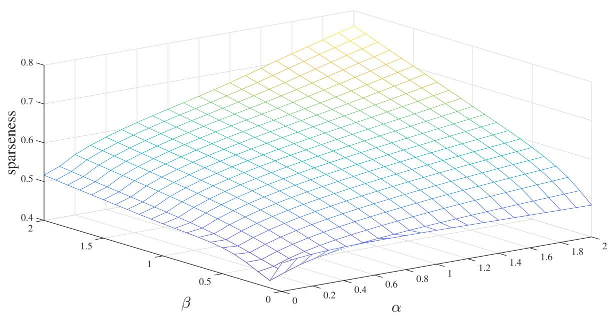


FIGURE 9. The *sparseness* of the basis images factorized by SGNMFL/L when $W \in R_{+}^{289\times 9}$.

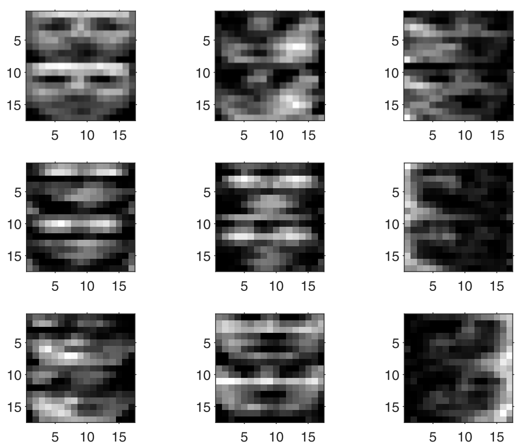


FIGURE 10. Basis images factorized by GNMFL when $W \in R_{+}^{289\times 9}$, and *sparseness* = 0.4730.

bigger. However, when $\alpha > 1$ or $\beta > 1$, the simulation software matlab prompts the warning "the matrix has singular accuracy" in the face training process. Therefore, $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$.

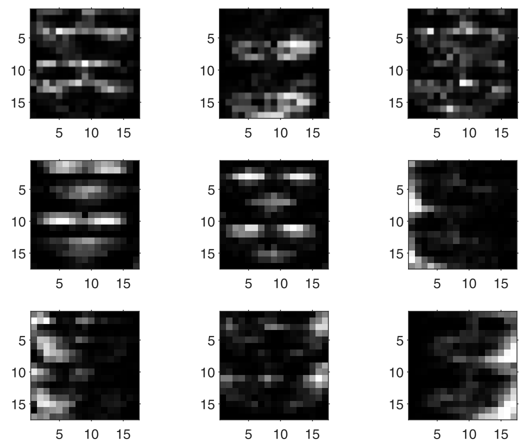


FIGURE 11. Basis images factorized by SGNMFL/L ($\alpha = 1$ and $\beta = 1$) when $W \in R_{+}^{289\times 9}$, and *sparseness* = 0.5868.

FIGUREs 10 and 11 show that the experimental results based on GNMFL and SGNMFL/L ($\alpha = 1$ and $\beta = 1$). Each 1-D column vector (36×1) of H represents the coefficients of the corresponding 1-D column vector (1156×1) of X based on W. FIGURE 11 shows that the basis images factorized by SGNMFL/L are sparser.

Based on the conventional NMF, SNMF/L, GNMFL and SGNMFL/L, the basis matrix and the elapsed times of several factorization algorithms in face training are shown in Table 4. Table 4 shows that the basis images factorized by GNMFL and SGNMFL/L are only $289 \times 9 \ll 1156 \times 36$.

All of 400 face images are test samples. In Table 2, the coefficient matrix $H \in R^{36\times 200}$, the elapsed time is the average elapsed time of 400 times $(W^TW)^{-1}W^TX_j = H_j$ or $(W^TW)^{-1}W^T \ltimes X_j = H_j$, where, X_j is j-th column vector of X, H_j is j-th column vector of H, and $j \in [1, 400]$. Table 5 shows the accuracy and the elapsed time of several

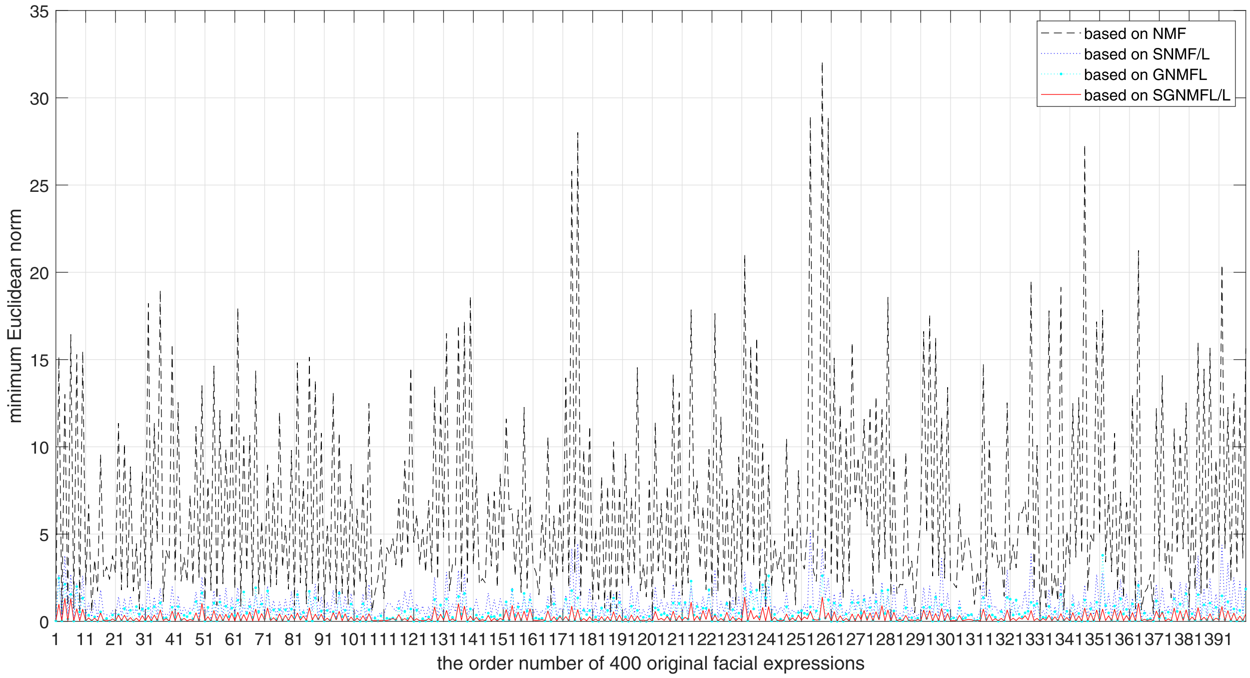


FIGURE 12. Comparison in minimum Euclid norms by using the conventional NMF, SNMF/L, GNMFL and SGNMFL/L on ORL database.

TABLE 4. Comparison of storage space and time efficiency in face training on ORL database.

Factorization method	W	Elapsed time(s)
NMF	$W_{+}^{1156\times36}$	1.838941
SNMF/L	$W_{+}^{1156\times36}$	1.800865
GNMFL	$\boldsymbol{W}_{+}^{\boldsymbol{289\times9}}$	7.323528
SGNMFL/L	$\boldsymbol{W}_{+}^{\boldsymbol{289\times9}}$	7.247721

TABLE 5. Comparison of storage space and time efficiency in face recognition on ORL database.

Factorization method	Basis matrix	Accuracy	Elapsed time(s)
NMF	$W_{+}^{1156\times36}$	0.8550	6.1601e-04
SNMF/L	$W_{+}^{1156\times36}$	0.8900	6.1432e-04
GNMFL	$\boldsymbol{W}_{+}^{\boldsymbol{289\times9}}$	0.8300	**5.9435e-04**
SGNMFL/L	$\boldsymbol{W}_{+}^{\boldsymbol{289\times9}}$	**0.8725**	**5.9104e-04**

factorization methods about NMF, SNMF/L ($\alpha = 1$ and $\beta = 1$), GNMFL and SGNMF/L ($\alpha = 1$ and $\beta = 1$) in face recognition. We can see from Table 5 that the proposed SGNMFL/L ($\alpha = 1$ and $\beta = 1$) achieves the highest computational efficiency. Meanwhile, it also achieves a higher face recognition accuracy than GNMFL.

In order to further analyze the robustness of the proposed SGNMFL/L ($\alpha = 1$ and $\beta = 1$), we artificially add four different noises into face images in the processes of face training and recognition. The parameters of each kind of noise are default parameters. Comparison results of face recognition accuracy with different noises on ORL database are shown in Table 6. We can observe from Table 6 that

TABLE 6. Comparison of face recognition accuracy with different noises on ORL database.

Factorization method	Noise type	Accuracy
SGNMFL/L	Gaussian	0.8475
SGNMFL/L	Poisson	0.8200
SGNMFL/L	Salt & pepper	0.8350
SGNMFL/L	Speckle	0.8025

the accuracies of face recognition with different noises are close to that of face recognition without noise. That is to say, the proposed SGNMFL/L is robust.

We also compare the minimum Euclid distance based on the conventional NMF ($W \in R_{+}^{1156\times36}$), SNMF/L ($W \in R_{+}^{1156\times36}$), GNMFL ($W \in R_{+}^{289\times9}$) and SGNMFL/L ($W \in R_{+}^{289\times9}$). The results are shown in FIGURE 12. According to FIGURE 12, we observe that the minimum Euclid norms based on NMF and GNMFL are relatively smaller, and minimum Euclid norms based on SNMF/L and SGNMFL/L are greater. The latter are easier to be recognized, which can effectively reduce the rate of error recognition.

In conclusion, the proposed SGNMFL/L performs the best in terms of saving storage space and reducing computation time in face recognition.

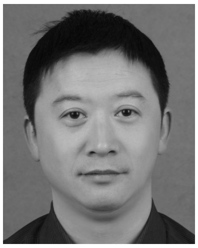
VI. CONCLUSIONS

In this paper, the proposed SGNMFL/L is presented in details and applied in face recognition. We perform two sets of experiments on the JAFFE database and the ORL database, respectively, so as to verify the effects of two hyper-parameters on the sparseness of basis matrix factorized by

SGNMFL/L, compare the performance of the conventional NMF, SNMF/L, GNMFL and the proposed SGNMFL/L in terms of storage space and time efficiency, and compare the face recognition accuracy with different noises. Experimental results show that the proposed SGNMF/L can significantly save the storage space ($3600 \times 15 \ll 14400 \times 60$ and $289 \times 9 \ll 1156 \times 9$), reduce the computation time in face recognition, and it has strong robustness. In the near future, we will add graph regularized constraint [40] to the proposed methods to improve the accuracy. We expect that the proposed SGNMFL/L will be improved for broad applications in science and engineering fields in the future.

REFERENCES

[1] D. White, J. D. Dunn, A. C. Schmid, and R. I. Kemp, "Error rates in users of automatic face recognition software," *PLoS One*, vol. 10, no. 10, Oct. 2015, Art. no. e0139827.
[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, Oct. 1999.
[3] Z. Yang, Z. Yu, Y. Wei, X. Yong, and S. Xie, "A fast non-smooth nonnegative matrix factorization for learning sparse representation," *IEEE Access*, vol. 4, pp. 5161–5168, 2017.
[4] X. Luo, M. Zhou, S. Li, Z. You, Y. Xia, and Q.-S. Zhu, "A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 579–592, Mar. 2016.
[5] X. Luo, M. Zhou, Y. Xia, Q. Zhu, A. C. Ammari, and A. Alabdulwahab, "Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 524–537, Mar. 2016.
[6] X. Luo, M. Zhou, S. Li, Y. Xia, Z. H. You, Q. Zhu, and H. Leung, "Incorporation of efficient second-order solvers into latent factor models for accurate prediction of missing QoS data," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1216–1228, Apr. 2018.
[7] M. Shang, X. Luo, Z. Liu, J. Chen, Y. Yuan, and M. Zhou, "Randomized latent factor model for high-dimensional and sparse matrices from industrial applications," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 1, pp. 131–141, Jan. 2019.
[8] X. Luo, H. Wu, H. Yuan, and M. Zhou, "Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors," *IEEE Trans. Cybern.*, to be published.
[9] X. Luo, M. Zhou, S. Li, L. Hu, and M. Shang, "Non-negativity constrained missing data estimation for high-dimensional and sparse matrices from industrial applicationss," *IEEE Trans. Cybern.*, to be published.
[10] W.-S. Chen, Y. Zhao, B. Pan, and B. Chen, "Supervised kernel nonnegative matrix factorization for face recognition," *Neurocomputing*, vol. 205, pp. 165–181, Sep. 2016.
[11] Z.-Z. Yu, Y.-H. Liu, B. Li, S.-C. Pang, and C.-C. Jia, "Incremental graph regulated nonnegative matrix factorization for face recognition," *J. Appl. Math.*, May 2014, Art. no. 928051.
[12] W. Chen, Y. Zhao, B. Chen, and P. Binbin, "Non-negative matrix factorization face recognition method and system based on kernel machine learning," U.S. Patent 15 769 741, Oct. 25, 2018,
[13] W. Liu, N. Zheng, and X. Lu, "Non-negative matrix factorization for visual coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Apr. 2003, p. III–293.
[14] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, no. 1, pp. 1457–1469, 2004.
[15] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 29–47, 2006.
[16] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis," *Bioinformatics*, vol. 23, no. 12, pp. 1495–1502, May 2007.
[17] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 155–173, Sep. 2007.
[18] B. J. Shastri and M. D. Levine, "Face recognition using localized features based on non-negative sparse coding," *Mach. Vis. Appl.*, vol. 18, no. 2, pp. 107–122, Apr. 2007.
[19] Z. Chen, L. Li, H. Peng, Y. Liu, and Y. Yang, "Attributed community mining using joint general non-negative matrix factorization with graph laplacian," *Phys. A, Stat. Mech. Appl.*, vol. 495, pp. 324–335, Apr. 2018.
[20] Z. Chen, L. Li, H. Peng, Y. Liu, and Y. Yang, "A novel digital watermarking based on general non-negative matrix factorization," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 1973–1986, Aug. 2018.
[21] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 19–60, Jan. 2010.
[22] X. Liu, H. Lu, and H. Gu, "Group sparse non-negative matrix factorization for multi-manifold learning," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2011, pp. 56.1–56.11.
[23] S. Tang, Y. T. Zheng, Y. Wang, and T. S. Chua, "Sparse ensemble learning for concept detection," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 43–54, Feb. 2012.
[24] J.-X. Du, C.-M. Zhai, and Y.-Q. Ye, "Face aging simulation based on nmf algorithm with sparseness constraints," *Neurocomputing*, vol. 116, no. 10, pp. 250–259, 2013.
[25] P. Ma, Y. Qu, D. Yang, Y. Ge, and X. Zhang, "Face recognition using two-dimensional nonnegative principal component analysis," *Proc. SPIE*, vol. 21, no. 3, Aug. 2012, Art. no. 033011.
[26] J. Cui, J. Wen, Z. Li, and B. Li, "Discriminant non-negative graph embedding for face recognition," *Neurocomputing*, vol. 149, pp. 1451–1460, Feb. 2015.
[27] W.-S. Chen, B. Pan, B. Fang, M. Li, and J. Tang, "Incremental nonnegative matrix factorization for face recognition," *Math. Problems Eng.*, May 2008, Art. no. 410674.
[28] H. Yin and H. Liu, "Nonnegative matrix factorization with bounded total variational regularization for face recognition," *Pattern Recognit. Lett.*, vol. 31, no. 16, pp. 2468–2473, Dec. 2010.
[29] J.-Y. Pan and J.-S. Zhang, "Large margin based nonnegative matrix factorization and partial least squares regression for face recognition," *Pattern Recognit. Lett.*, vol. 32, no. 14, pp. 1822–1835, Oct. 2011.
[30] C. Zhou, X. Wei, Q. Zhang, and X. Fang, "Fisher's linear discriminant (FLD) and support vector machine (SVM) in non-negative matrix factorization (NMF) residual space for face recognition," *Optica Applicata*, vol. 40, no. 3, pp. 693–704, Jan. 2010.
[31] Z. Zhang, L. Wang, Q. Zhu, Z. Liu, and Y. Chen, "Noise modeling and representation based classification methods for face recognition," *Neurocomputing*, vol. 148, pp. 420–429, Jan. 2015.
[32] L. Dai, L. Dai, and J. Wu, *Convergent Projective Non-negative Matrix Factorization with Kullback-Leibler Divergence*. Amsterdam, The Netherlands: Elsevier, 2014.
[33] J.-S. Zhang, C.-P. Wang, and Y.-Q. Yang, "Learning latent features by nonnegative matrix factorization combining similarity judgments," *Neurocomputing*, vol. 155, pp. 43–52, May 2015.
[34] Y. Xue, C. S. Tong, Y. Chen, and W. S. Chen, "Clustering-based initialization for non-negative matrix factorization," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 525–536, Nov. 2008.
[35] L.-Y. Hu, G.-D. Guo, and C.-F. Ma, "Image processing using newton-based algorithm of nonnegative matrix factorization," *Appl. Math. Comput.*, vol. 269, pp. 956–964, Oct. 2015.
[36] C. Daizhan, "Semi-tensor product of matrices and its application to Morgen's problem," *Sci. China Inf. Sci.*, vol. 44, no. 3, pp. 195–212, Jun. 2001.
[37] D. Cheng, H. Qi, and Y. Zhao, *An Introduction to Semi-tensor Product of Matrices and Its Applications*. Singapore: World Scientific, 2012.
[38] R. Bellman, *Introduction to Matrix Analysis*, vol. 56, no. 3, 2nd ed., New York, NY, USA: McGraw-Hill, 1970, pp. 225–226.
[39] P. Sajda, S. Du, and L. C. Parra "Recovery of constituent spectra using non-negative matrix factorization," *Proc. SPIE*, vol. 5207, pp. 321–331, Nov. 2003.
[40] C. Wang, X. Song, and J. Zhang, "Graph regularized nonnegative matrix factorization with sample diversity for image representation," *Eng. Appl. Artif. Intell.*, vol. 68, pp. 32–39, Feb. 2018.



ZIGANG CHEN received the M.S. degree in computer technology from the Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2019. He is currently an Associate Professor with the College of Artificial Intelligence, Tianjin University of Science and Technology, China. His major research interests include image processing and pattern recognition.

LIXIANG LI received the M.S. degree in circuit and system from Yanshan University, Qinhuangdao, China, in 2003, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006, where she is currently a Professor with the School of Cyberspace Security. She visited the Potsdam Institute for Climate Impact Research, Germany, from 2011 to 2012. She has published more than 100 papers and a monograph and being supported by over 10 national/provincial and ministerial level projects in last five years. Her research interests include swarm intelligence and network security. She was the National Excellent Doctoral theses winner, the New Century Excellent Talents in University, and the winner of Henry Folk Education Foundation, the Hong Kong Scholar Award, and the Beijing higher education program for young talents.

HAIPENG PENG received the M.S. degree in system engineering from the Shenyang University of Technology, Shenyang, China, in 2006, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010, where he is currently a Professor with the School of Cyberspace Security. He is the coauthor of 50 scientific papers. His research interests include information security, network security, complex networks, and control of dynamical systems.

YUHONG LIU received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2007, respectively, and the Ph.D. degree from The University of Rhode Island, in 2012. She is currently an Assistant Professor with the Department of Computer Engineering, Santa Clara University. With expertise in trustworthy computing and cyber security, her research interests include developing trust models and applying them on emerging applications, such as online social media, cyber-physical systems, and cloud computing. She is a recipient of the 2013 University of Rhode Island Graduate School Excellence in Doctoral Research Award, and her work on securing online reputation systems received the Best Paper Award at the IEEE International Conference on Social Computing 2010 (acceptance rate = 13%).

HAIHUA ZHU received the M.S. degree in computer application technology from the Wuhan University of Technology, Wuhan, China, in 2007. She is currently pursuing the Ph.D. degree in information security with the Beijing University of Posts and Telecommunications, Beijing, China. Her major research interests include image processing and information security.

YIXIAN YANG received the M.S. degree in applied mathematics and the Ph.D. degree in electronics and communication systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 1986 and 1988, respectively, where he is currently the Managing Director of the Information Security Center. He has authored more than 40 national and provincial key scientific research projects and published more than 300 high-level papers and 20 monographs. His major research interests include coding and cryptography, information and network security, and signal and information processing. He is a Yangtze River Scholar Program Professor, received the National Outstanding Youth Fund winners, and recognized as the National Teaching Masters.