



Multiple Graph Regularized Non-negative Matrix Factorization Based on L_{2,1} Norm for Face Recognition

Minghai Yao^(✉), Jinsong Li, and Changyu Zhou

School of Information Science and Technology, BoHai University, Jinzhou 121001, China
yao_ming_hai@163.com

Abstract. In order to solve the problem that most non-negative matrix decomposition methods are sensitive to noise and outliers, resulting in poor sparsity and robustness, a multiple graph regularized non-negative matrix factorization based on L_{2,1} norm is proposed in this paper, and its performance is verified by face recognition. Firstly, the texture rich area is selected in the preprocessing stage. Secondly, L_{2,1} norm is used to improve the sparsity and robustness of decomposition results. Then, in order to better maintain the manifold structure of the data, the multi-graph constraint model is constructed. Furthermore, the corresponding multiplicative updating solution of the optimization framework is given, and the convergence proof is given. Finally, a large number of experimental results show that the superiority and effectiveness of the proposed approach.

Keywords: Non-negative matrix factorization · Graph regularization · Feature extraction · Face recognition

1 Introduction

As one of the most challenging classification tasks in computer vision and pattern recognition, face recognition have attracted much researchers' attentions [1–4]. Many face recognition techniques have been proposed in the past few decades. A face image of size pixels is usually represented by a dimensional vector. However, excessive high dimensionality has an impact on the efficiency and accuracy of face recognition. In the face of this situation, the method of feature extraction or data dimensionality reduction is generally adopted. The classic dimensionality reduction methods mainly include PCA, LDA, etc. Although these methods can reduce the dimension of the original data well, there is no non-negative requirement for the decomposition factor, so there are negative values in the decomposed components. These negative values have no physical significance in practical problems. Therefore, NMF is generated by operation.

In 1999, Non-negative matrix decomposition (NMF) algorithm was first proposed by Lee in Nature [5]. Its purpose is to approximate decompose the original nonnegative data matrix into the product of two matrices with nonnegative elements, namely basis matrix and coefficient feature matrix. This part-based, purely additive nature of the

NMF method can enhance the partial to overall interpretation of the data. Because of its non-negative constraints, sparse local expression and good interpretability, the NMF has been widely used in many real world problems such as face recognition, text clustering, feature recognition and information retrieval [6–8].

Many scholars have improved NMF and applied it in the field of face feature extraction and classification recognition. In 2002, Hoyer [9] integrated the idea of sparse coding into NMF, using L1 norm as sparse penalty term, and using Euclidean distance to calculate the error between the original data and the reconstructed data. To use the data geometric structure, Cai et al. [10] proposed a graph regularization non-negative matrix factorization (GNMF). In the GNMF algorithm, the geometrical structure of data is encoded by a k -nearest neighbor graph. In order to maintain the internal structure of the original data, Qing et al. proposed another variant of NMF, which is called a graph regularization non-negative matrix factorization (GNMF) method to describe the internal manifold structure of data points in the data matrix by constructing the neighborhood graph [11]. In order to improve the sparsity of the decomposition results and transmit the effective information, Jiang Wei et al. Proposed the sparse constraint graph NMF method (SGNMF) [12]. Zhou et al. Proposed the NMF algorithm based on Gabor wavelet by combining wavelet change and manifold [13].

With single figure to constraints of the internal structure of the original data, although to a certain extent to meet the demand of feature vector, but the results were not intellectual and meet the requirements of the single, while there are double or more constraints of the NMF algorithm, but while in measuring loss function based on L2 norm, existing algorithms are sensitive to noise and outliers caused by sparse decomposition results, and poor robustness problems. Therefore, this paper proposes a multi-graph regularized non-negative matrix factorization method based on L2,1 norm (L2,1-MGNMF). The method uses the line sparse property of L2,1 norm as the loss function. On the basis of the single graph regularization structure, the manifold structure of the original data is represented by fusing multiple graphs, and the constraints and decomposition results of the original structure of the original data are considered. Increased sparsity and robustness. The proposed L2,1-MGNMF method is finally tested on ORL and other face databases. A lot of experimental results show that our L2,1-MGNMF approach is effective and feasible, which surpasses some existing methods.

The remainder of the paper is organized as follows: Sect. 2 introduces the basic ideas of the existing NMF and GDNMF methods. Section 3 proposes our L2,1-MGNMF and gives theoretic analysis. Experimental comparisons are reported in Sect. 4. Finally, conclusions are given in Sect. 5.

2 A Brief Review of NMF

This section describes NMF and GNMF algorithms briefly. Let X be a training data matrix of $m \times n$ -dimensional samples x_1, x_2, \dots, x_n , *i.e.*, $x \in \mathcal{R}^{m \times n}$ with nonnegative entries. Each column of X represents a face image with m dimensions.

2.1 Non-negative Matrix Factorization(NMF)

NMF aims to approximately decomposes the training sample matrix X into a product of two non-negative matrices $A \in \mathcal{R}^{m \times r}$ and $S \in \mathcal{R}^{r \times n}$ ($r \ll \min(m, n)$), i.e., $X \approx AS$. Matrices A and S are called basis matrix and coefficient matrix, respectively. In general, NMF is based on minimizing the Euclidean distance between X and AS . The corresponding optimization problem is as follows:

$$\min_{A,S} \|X - AS\|_F^2, \text{ s.t. } A \geq 0, S \geq 0 \quad (1)$$

where $\|\bullet\|_F$ is the matrix Frobenius norm of a matrix. The optimization problem can be solved using gradient descent method. The well-known multiplicative update rules are as follows:

$$S^{(t+1)} \leftarrow S^{(t)} \frac{(A^{(t)T} X)}{(A^{(t)T} A^{(t)} S^{(t)})} \quad A^{(t+1)} \leftarrow A^{(t)} \frac{(X S^{(t)T})}{(A S^{(t)} S^{(t)T})} \quad (2)$$

The convergence of these multiplicative update rules have been proved in [14].

2.2 Graph Regularized Non-negative Matrix Factorization (GNMF)

To find a compact representation which uncovers the hidden semantics and simultaneously respects the intrinsic geometric structure, the graph regularized non-negative matrix factorization (GNMF) was proposed in [10]. GNMF solved the following optimization problem:

$$\min_{A,S} \|X - AS\|_F^2 + \lambda \text{Tr}(SLS^T), \text{ s.t. } A \geq 0, S \geq 0 \quad (3)$$

where $\text{Tr}(\bullet)$ is the trace of a matrix and $L = D - W$ is called the graph Laplacian matrix, regularization parameter $\lambda \geq 0$ controls the smoothness of the new representation. The W denotes the weight matrix, and D is a diagonal matrix. The corresponding multiplicative update rules for solving Eq. (3) are as follows:

$$S^{(t+1)} \leftarrow S^{(t)} \frac{(X^{(t)T} A + \lambda WS)}{(A^{(t)T} A^{(t)} S^{(t)})} \quad A^{(t+1)} \leftarrow A^{(t)} \frac{(X S^{(t)T})}{(A S^{(t)} S^{(t)T})} \quad (4)$$

3 Multiple Graph Regularized Non-negative Matrix Factorization Based on L2,1 Norm (L2,1-MGNMF)

Although the existing NMF-based improved methods have achieved good results, they still have some limitations. In this section, we will describe our Multiple Graph Regularized Non-negative Matrix Factorization based on L2,1 Norm (L2,1-MGNMF). Further we formulate our optimization problem by adding supervised label information to the objective function of L2,1-MGNMF. The definition and update rules of L2,1-MGNMF are given below.

3.1 L_{2,1}-MGNMF Model

In order to maintain the original structure of the data as much as possible, this paper uses the neighborhood weighting, weight weighting, and sparse weighting to constrain the structure of the original data. The three weight matrices are defined as follows:

1. Neighborhood weighting. $W_{ij}^N = 1$ if and only if sample j is the nearest neighbor of sample i . This is the simplest weighting method and is very easy to compute.
2. Weight weighting. If sample j is the neighbor of sample i , then

$$W_{ji}^W = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{5}$$

3. Sparse weighting. Sparse weight graphs can represent the sparse structure of the original data, and the sparse constraint is added to make the base image obtained by the decomposition represent the original image with as few features as possible, the weight matrix defined by:

$$W^S = s.t. \|x - D\varphi\|_p \leq \varepsilon, \tag{6}$$

where φ is the sparse coefficient.

The L_{2,1}-MGNMF solves the following optimization problem:

$$\begin{aligned} \min_{A,S} & \|X - AS\|_{2,1} + \alpha \sum_{x=1}^X \mu_x \left(\sum_{i,j=1}^m \|s_i - s_j\|_2^2 w_{ij}^x \right) + \beta \|\mu\|_2^2, \\ s.t. & A \geq 0, S \geq 0, \sum_{x=1}^X \mu_x = 1, \mu \geq 0 \end{aligned} \tag{7}$$

where, W^X represents the x -th weight graph, and α and β are balance parameters. The discrimination ability of different graphs is very different. The weight μ should be set according to the graph, and the balance parameter α determines the influence of the integrated manifold structure on the objective function.

3.2 The Update Rules of L_{2,1}-MGNMF

Though the objective function in Eq. (7) is not jointly convex in the pair (A, S, μ) , it is convex with respect to one variable in the (A, S, μ) while fixing the others. Therefore it is unrealistic to expect an algorithm to find the global minima. In the following, we can use the iterative solution of fixing two variables to update another variable which can achieve local minima. The solution process is as follows:

1. Fix μ and S , update A . To remove the irrelevant items, the optimization problem of A can be transformed into the following:

$$\min \|X - AS\|_{2,1} = \text{tr}((X - AS)D(X - AS)^T)$$

$$= \text{tr}(XDX^T - 2ASDX^T + ASDS^T A^T) \quad (8)$$

s.t. $A \geq 0$

where D is a diagonal matrix $d_{ii} = 1/\|x_i - As_i\|$.

2. Let Λ be the Lagrange multiplier, the Lagrange ℓ is:

$$\ell(A, \Lambda) = \text{tr}(XDX^T) - 2\text{tr}(ASDX^T) + \text{tr}(ASDS^T A^T) + \lambda \text{tr}(\Lambda A), \quad (9)$$

3. The partial derivatives of ℓ with Eq. (9) is:

$$\frac{\partial \ell(A, \Lambda)}{\partial A} = -2XDS^T + 2ASDS^T + \lambda \Lambda = 0 \quad (10)$$

4. Using the KKT conditions $\Lambda_{ij}A_{ij} = 0$, we get the following equations:

$$\left(-2XDS^T + 2ASDS^T\right)A_{ij} = 0 \quad (11)$$

5. Fix μ and A, update S. To remove the irrelevant items, the optimization problem of S can be transformed into the following:

$$\min \|X - AS\|_{2,1} + \alpha \sum_{x=1}^X \mu_x \left(\sum_{i,j=1}^m \|s_i - s_j\|_2^2 w_{ij}^x \right). \quad \text{s.t. } S \geq 0 \quad (12)$$

6. Similarly, we get the following equations:

$$\left(-2A^T XD + 2A^T ASD + \alpha SL\right)S_{ij} = 0 \quad (13)$$

Where $L = \sum_{x=1}^X \mu_x L^x$.

7. Therefore, according to Eq. (11) and Eq. (13), we have the following updating rules:

$$A^{(t+1)} \leftarrow A^{(t)} \frac{(XDS^T)^{(t)}}{(ASDS^T)^{(t)}} \quad (14)$$

$$S^{(t+1)} \leftarrow S^{(t)} \frac{(A^T XD)^{(t)}}{(A^T ASDS + \alpha SL)^{(t)}} \quad (15)$$

Theorem 1: The objective function O_1 in Eq. (7) is non-increasing under the updating rules in Eq. (14), and (15).

We can iteratively update A and S until the objective value of O_1 does not change or the number of iteration exceed the maximum value. Theorem 1 also guarantees that the multiplicative updating rules in Eq. (14) and (15) converge to a local optimum.

We summarize our Multiple Graph Regularized Non-negative Matrix Factorization based on $L_{2,1}$ Norm ($L_{2,1}$ -MGNMF) algorithm in Table 1.

4 Experimental Results

In this section, we compare the proposed $L_{2,1}$ -MGNMF with four representative algorithms, which are NMF, SGNMF, LGNMF [15], and SPGNMF [16], on three face datasets, i.e., ORL [17], Yale [18], and PIE [19].

Table 1. The algorithm of L2,1-MGNMF

Input: Data matrix $X \in \mathfrak{R}^{m \times n}$, balance parameters α and weight μ , convergence condition

Initialization: Randomly initialize two non-negative matrices $A \in \mathfrak{R}^{m \times r}$ and $S \in \mathfrak{R}^{r \times n}$

Repeat: 1. Update A by rule(14)
2. Update S by rule(15)

Until Convergence

Output: Basis matrix A

4.1 Dataset

Three datasets are used in the experiment. The important statistics of these data sets are summarized in Table 2. Figure 1 shows example images of ORL, Yale, and PIE datasets. Before the experiment, face image should be preprocessed. We think that the texture region contains more information, and the rest of the region is redundant. Therefore, the texture rich region is reserved for future experiments (see Fig. 2). Then, each face image is represented as a column vector and the features (pixel values) are then scaled to $[0, 1]$ (divided by 255).

For each sample, we randomly select half of the images as the training set and the rest as the test set. Repeat random selection 1 times to ensure that 95% of the images have participated in the training and testing. The performance of L2,1-MGNMF is measured by Accuracy (ACC) and False Acceptance Rate (FAR).

Table 2. Statistics of the four datasets

Dataset	Number of samples (P)	Dimensionality (N)	Number of classes (K)
ORL	400	112*92	40
Yale	165	64*64	15
PIE	1632	64*64	68

4.2 Parameter Selection

The main parameters in the algorithm model are the dimension rafter dimension reduction, the iteration number L_{it} which affects the convergence speed of the algorithm, and balance parameters α which affects the decomposition error.



(a)



(b)



(c)

Fig. 1. Face image examples of the (a) ORL, (b) Yale, and (c) PIE datasets



Fig. 2. Pre-processed image examples

Figure 3 shows the accuracy curve when Lit values 100, 300, 500, 800 and 1500 respectively. It can be seen from the figure that the performance of the algorithm increases with the number of iterations. Considering the relationship between accuracy and computational efficiency, Lit is set to 1500.

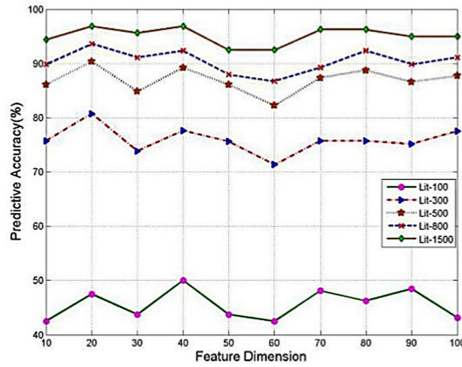


Fig. 3. The ACC with different Lit

It can be seen from the figure that the performance of the algorithm increases with the decrease of the balance factor α . when $\alpha < 0.01$, the performance of the algorithm fails to further improve, so the balance parameters in this paper is 0.01 (Fig. 4).

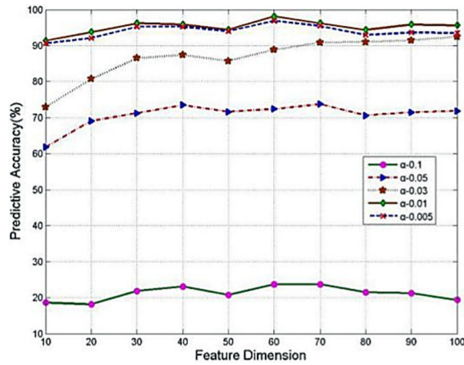


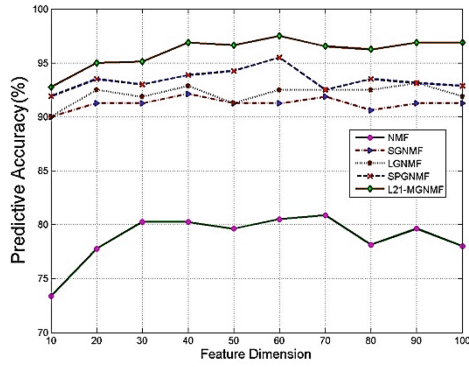
Fig. 4. The ACC with different α

4.3 Compared Algorithms

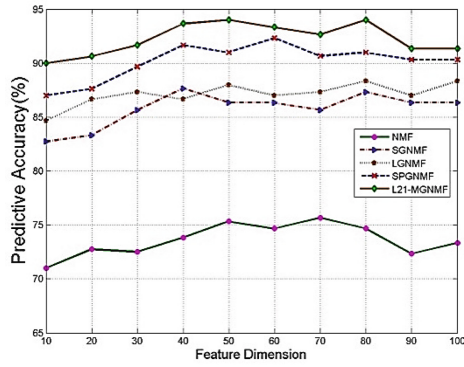
To demonstrate how our approach improves performance, we compared the following four popular algorithms: NMF, SGNMF, LGNMF, and SPGNMF.

Figure 5 gives that the average accuracy versus the dimension of the subspace. Figure 6 gives that the average FAR versus the dimension of the subspace. According to Fig. 5 and 6, it can be seen that the performance of the proposed method on three face databases is significantly higher than that of other methods.

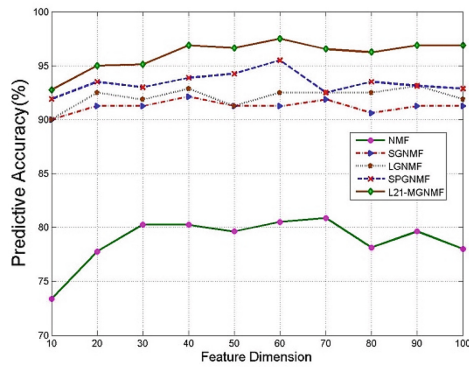
Finally, we compared the sparsity of the matrix decomposition results of these five algorithms, and selected the base image obtained by the decomposition of the original data X. The feature dimension of the base image was set at 30, and the sparsity is defined



(a) ORL

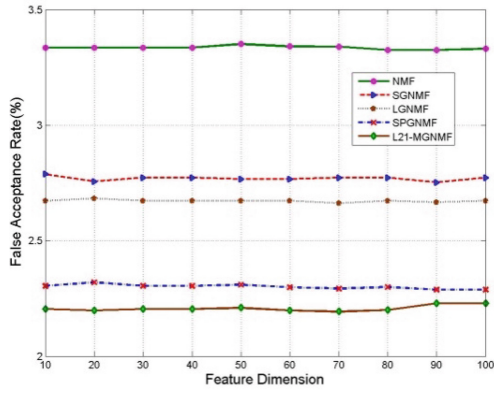


(b) Yale

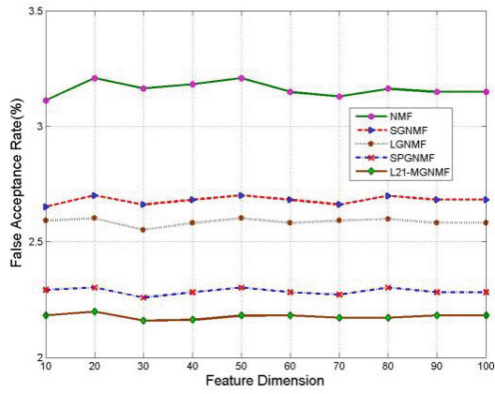


(c) PIE

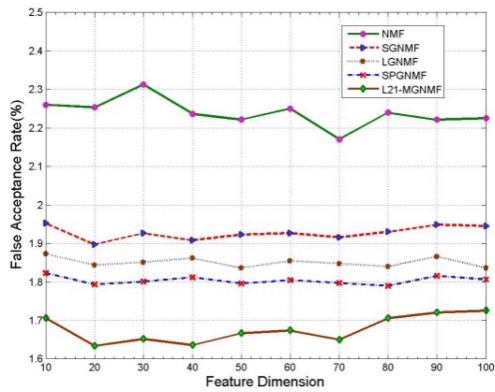
Fig. 5. Face predictive accuracy on the (a) ORL, (b) Yale, and (c) PIE datasets



(a) ORL



(b) Yale



(c) PIE

Fig. 6. The FAR on the (a) ORL, (b) Yale, and (c) PIE datasets

by:

$$SP(x) = \frac{\sqrt{m} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1} \quad (16)$$

where m is the dimension of the vector. When there is only one non-zero element, $SP = 1$, the smaller the value of SP , the denser the vector x , otherwise, the sparser. The experimental results are as follows (Table 3):

Table 3. The SP of different algorithm

Algorithm	ORL	Yale	PIE
NMF	0.3546	0.3789	0.3762
SGNMF	0.4632	0.4762	0.4785
LGNMF	0.4469	0.4618	0.4539
SPGNMF	0.4793	0.5149	0.5278

It can be seen from the table that compared with the sparsity results of these algorithms, the NMF algorithm has the worst sparsity. In this paper, L2,1-mgnmf algorithm has the highest result than other algorithms, and the decomposed base image is the most sparse with better local expression ability.

5 Conclusions

We have presented an efficient method for matrix factorization, called Multi-graph regularized Non-negative Matrix Factorization method based on L2,1 norm (L2,1-MGNMF). As a result, L2,1-MGNMF can have more discriminative power than the conventional NMF and its several variants. Further, we show the corresponding multiplicative update rules and convergence studies. Evaluations on three face datasets have revealed both higher recognition accuracy, sparsity and lower false acceptance rate of the proposed algorithm in comparison to those of the state-of-the-art algorithms. But how to build a multi-graph fusion model is still the focus of future research.

Acknowledgements. This work was supported by Natural Science Fund of Liaoning Province (No. 2019-ZD-0503) and Science and Technology Research project of Liaoning Province Education Department (No. LQ2017004).

References

1. Luo, X., Xu, Y., Yang, J.: Multi-resolution dictionary learning for face recognition. *Pattern Recogn.* (93), 283–292 (2019)

2. Li, S., Liu, F., Liang, J.Y., Cai, Z.H., Liang, Z.Y.: Optimization of face recognition system based on azure IoT edge. *Comput. Mater. Continua* **3**(61), 1377–1389 (2019)
3. Wang, X., Xiong, C., Pei, Q.Q., Qu, Y.Y.: Expression preserved face privacy protection based on multi-mode discriminant analysis. *Comput. Mater. Continua* **1**(57), 107–121 (2018)
4. Yu, X., Tian, Z.H., Qiu, J., Su, S., Yan, X.R.: An intrusion detection algorithm based on feature graph. *Comput. Mater. Continua* **1**(61), 255–274 (2019)
5. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
6. Li, S.Z., Hou, X.W., Zhang, H.J.: Learning Spatially Localized, Parts-Based Representation. *IEEE Computer Society* (2001)
7. Yu, N., Gao, Y.L., Liu, J.X., Shang, J., Zhu, R., Dai, L.Y.: Co-differential gene selection and clustering based on graph regularized multi-view NMF in cancer genomic data. *Genes* **9**(12), 586 (2018)
8. Su, Y.R., Niu, X.X.: An improved random seed searching clustering algorithm based on shared nearest neighbor. *Appl. Mech. Mater.* **3749**(719), 1160–1165 (2015)
9. Hoyer, P.O.: Non-negative sparse coding. In: *Proceedings of Neural Networks for Signal Processing*, pp. 557–565 (2002)
10. Cai, D., He, X., Han, J., Huang, T.: Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1548–1560 (2011)
11. Liao, Q., Zhang, Q.: Local coordinate based graph-regularized NMF for image representation. *Signal Process.* **124**, 103–114 (2016)
12. Zhou, J., Zhang, S., Mei, H., Wang, D.: A method of facial expression recognition based on Gabor and NMF. *Pattern Recogn. Image Anal.* **26**(1), 119–124 (2016). <https://doi.org/10.1134/S1054661815040070>
13. Yang, Y.S., Ming, A.B., Zhang, Y.Y., Zhu, Y.S.: Discriminative non-negative matrix factorization (DNMF) and its application to the fault diagnosis of diesel engine. *Mech. Syst. Signal Process.* **3**(26), 158–171 (2017)
14. Lee, D., Seung, H.: Algorithms for non-negative matrix factorization. *Adv. Neural. Inf. Process. Syst.* **13**, 556–562 (2001)
15. Jia, X., Sun, F., Li, H.J.: Image feature extraction method based on improved nonnegative matrix factorization with universality (2018)
16. Jiang, J.H., Chen, Y.J., Meng, X.Q., Wang, L.M., Li, K.Q.: A novel density peaks clustering algorithm based on k-nearest neighbors for improving assignment process. *Physica A: Stat. Mech. Appl.* **3**(12), 702–713 (2019)
17. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pp. 138–142 (1994)
18. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
19. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1615–1618 (2003)