IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 18, 2020, accepted July 9, 2020, date of publication July 17, 2020, date of current version July 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009261

Graph Regularized Constrained Non-Negative Matrix Factorization With L_p Smoothness for Image Representation

ZHENQIU SHU[1,2], ZONGHUI WENG[1], YUNMENG ZHANG[1], CONG-ZHE YOU[1], AND ZHEN LIU[2]

[1]School of Computer Engineering, Jiangsu University of Technology, Changzhou 231001, China
[2]Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 231001, China

Corresponding author: Zhenqiu Shu (shuzhenqiu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61603159, Grant 61806088, and Grant 61902160, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160293, in part by the Excellent Key Teachers of Qing-Lan Project in Jiangsu Province, in part by the Graduate Student Practice Innovation Foundation of Jiangsu Province under Grant 20_0930, in part by the Changzhou Science and Technology Support Plan (social development) under Grant CE20185044, and in part by the Changzhou Science and Technology Plan (applied basic research) under Grant CJ20190076.

ABSTRACT Nonnegative matrix factorization-based image representation algorithms have been widely applied to deal with high-dimensional data in the past few years. In this paper, we propose a graph regularized constrained nonnegative matrix factorization with L_p Smoothing (GCNMFS) for image representation. Specifically, the main contributions of the proposed GCNMFS method include as follows: firstly, the geometric manifold structure hidden in data is effectively exploited by adopting a graph regularizer. Secondly, the label information of labeled samples is incorporated into the model of NMF without additional parameters. Finally, the L_p smoothness constraint is used to constrain the basis matrix, and thus a smooth and more accurate solution is produced. Moreover, an effective optimization scheme is presented to solve the proposed model. Extensive experiments on several image datasets show the proposed GCNMFS method can achieve better performance than other state-of-the-art methods in clustering.

INDEX TERMS Image representation, manifold structure, graph, label information, L_p smoothness.

I. INTRODUCTION

Data representation has received considerable attention in practice for many years. Many popular data representation methods, such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2], independent component analysis (ICA) [3], singular value decomposition (SVD), [4] non-negative matrix factorization (NMF) [5], [6] and concept factorization (CF) [7], have been used to solve many problems in the real world.

In the past few decades, as a well-known data representation method, NMF seeks to approximate the data matrix using the product of two nonnegative low rank matrices. Since NMF only performs the additive operations, not subtraction operations, it is a parts-based representation method. Due to the strong interpretability in psychological and physiological [8], NMF has been widely applied in various fields, e.g., text analysis [9] and data clustering [10]–[12], image analysis [13]–[17] and face recognition [18]. Xu and Gong [7] proposed a concept factorization (CF) method to apply the document clustering. One of the advantages is that CF can deal with the data matrix containing negative elements due to noise. Therefore, it is more beneficial to apply real problems in practice. Recently, various constraints are imposed on the original NMF model, and its goal is to improve the performance by making full use of the prior knowledge of data. To consider the sparseness of coefficient, Liu *et al.* [19] proposed a sparse NMF method via imposing sparseness constraints explicitly. Qian *et al.* [20] proposed a sparsity-constrained NMF method for hyperspectral unmixing by imposing the $l_{1/2}$ sparsity constraint. Cai *et al.* [21] proposed a graph regularized non-negative matrix factorization (GNMF) method by constructing a nearest neighbor graph to model the

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim.

manifold structure of data. Sun *et al.* [22] proposed a sparse dual graph regularized NMF method that takes advantage of the dual manifold structure and the sparseness of the coefficient matrix. To consider the label information among data, some semi-supervised and supervised methods were proposed in recent years. Liu *et al.* [17] proposed a constrained non-negative matrix factorization (CNMF) that uses the label information of the labeled samples with parameter-free. In CNMF, the labeled samples from the same category are projected into the same point in the low-dimensional representation space. Therefore, it can effectively improve the representation power in clustering. To increase the discriminative ability, Lu *et al.* [23] proposed a nonnegative discriminant matrix factorization (NDMF) by simultaneously incorporating the nonnegativity, orthogonality constraints, and discriminative information. Sun *et al.* [24] proposed a graph regularized and sparse nonnegative matrix factorization with hard constraints method. The manifold structure and discriminative structure of data can be discovered by jointly integrating a graph regularizer and label information as well as sparseness constraint. However, the above-mentioned methods neglect the smoothness of the basis matrix. Many studies have shown that the smoothness assumption plays an important role in data representation [25]–[29]. To solve this issue, Leng *et al.* [28] proposed to constrain the basis matrix with the L_p smoothing. Therefore, it can generate a smooth and more accurate solution for the model. Salehani and Gazor [29] proposed to simultaneously explore the sparseness and smoothness of the abundance matrix in hyperspectral unmixing.

In this work, we propose a novel method, called graph regularized constrained nonnegative matrix factorization with L_p smoothing (GCNMFS), for data representation. GCNMFS takes advantage of more prior knowledge of data compared with other competitors. In addition, we present an efficient optimization algorithm based on the multiplicative updating algorithm to solve the proposed model. Experiments on several benchmark datasets show that our proposed method is superior to related state-of-the-art methods.

The main contributions of this work can be summarized as follows:

(1) Compared with traditional NMF methods, the proposed GCNMFS method effectively exploits more prior knowledge hidden in data. Specifically, the geometric manifold structure embedded in data is explored using regularization technology. The label information of known labeled samples is utilized with resort to the hard constraint without parameters. A more stable and accurate solution of the model is derived by imposing the L_p norm constraint on the basis matrix. Therefore, the proposed GCNMFS method can show more representation ability in clustering applications.

(2) We present an efficient optimization scheme to solve the proposed GCNMFS method, and prove their convergence. The extensive experimental results on several benchmark datasets manifest the effectiveness of the proposed method.

The remainder of this paper is organized as follows: Section 2 briefly reviews the related works. Section 3 introduces the proposed GCNMFS algorithm. Section 4 conducts the experiment results, and Section 5 concludes the paper.

II. RELATED WORKS

In this section, we briefly review some related works to our method.

A. NMF

Consider a case that there is a non-negative matrix $X = [x_{ij}] \in R^{m\times n}$ from c categories. NMF aims to find two non-negative matrices $U = [u_{ik}] \in R^{m\times k}$ and $V = [v_{jk}] \in R^{n\times k}$ such that their product approximates the original data matrix X. Using the Euclidean distance to measure the approximation error, the objective function of NMF is expressed as follows:

$$\begin{aligned} O_{NMF} &= \sum_{i=1}^{N}\sum_{j=1}^{M}(X_{ij} - (UV^T)_{ij})^2 = \left\|X - UV^T\right\|_F^2 \\ &s.t.\ U \geq 0, \quad V \geq 0, \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm. It is arduous to solve the problem (1) for U and V together due to its non-convexity. The multiplicative iterative algorithm proposed by Lee and Seung [6] is used to solve the problem (1). Therefore, the updating rules of Eq. (1) can be derived as follows:

$$u_{ik} \leftarrow u_{ij}\frac{(XV)_{ij}}{(UV^TV)_{ij}}, \quad v_{jk} \leftarrow v_{ij}\frac{(X^TU)_{ij}}{(VU^TU)_{ij}}. \tag{2}$$

B. GNMF

According to manifold learning theory, Cai *et al.* [21] proposed the GNMF method, which takes the intrinsic geometric structure of the data as a regularization term to constrain the model of NMF. The graph regularization term is given as follows:

$$\begin{aligned} R &= \frac{1}{2}\sum_{j,l=1}^{N}\left\|v_j - v_l\right\|^2 W_{jl} \\ &= \sum_{j=1}^{N} v_j^T v_j D_{jj} - \sum_{j,l=1}^{N} v_j^T v_l W_{jj} \\ &= Tr(V^TDV) - Tr(V^TWV) \\ &= Tr(V^TLV), \end{aligned} \tag{3}$$

where $Tr(\cdot)$ is the trace of the matrix. v_i and v_j can be regarded as the new representation of these two points x_i and x_j in the new feature space. W is the affine matrix of the nearest neighbor graph, and $L = D - W$ is a Laplacian matrix, where D is a diagonal matrix, $D_{ij} = \sum_j W_{ij}$.

By imposing the graph regularization term constraint on the model of the original NMF, the objective function of GNMF is given as follows:

$$\begin{aligned} \text{OGNMF} = &\left\| X - UV^T \right\|_F^2 + \lambda Tr(V^T LV) \\ &s.t.\ U \geq 0, \quad V \geq 0. \end{aligned} \tag{4}$$

Using a similar optimization scheme, the updating rules of the problem (4) are derived as follows:

$$U_{ij} \leftarrow U_{ij} \frac{(XV)_{ij}}{(UV^T V)_{ij}}, \quad V_{ij} \leftarrow V_{ij} \frac{(X^T U + \lambda WV)_{ij}}{(VU^T U + \lambda DV)_{ij}}. \tag{5}$$

III. THE PROPOSED METHOD

In this section, we introduce the proposed GCNMFS method in detail.

A. CONSTRUCTION OF AUXILIARY MATRIX

Given a non-negative matrix $\{x_i\}_{i=1}^n$ from c categories, the first l samples are labeled and the remaining $n-l$ samples are unlabeled. The labeled samples are all marked as one of these clusters. Assume x_i is marked as the j-th cluster, then $m_{ij} = 1$, otherwise $m_{ij} = 0$. Therefore, the indicator matrix A is constructed as follows:

$$A = \begin{pmatrix} M_{l\times c} & 0 \\ 0 & I_{n-l} \end{pmatrix}, \tag{6}$$

where I_{n-l} is an identity matrix. To utilize the label information of the labeled samples, a label constraint is imposed by introducing an auxiliary matrix Z. The Eq. (7) is given as follows:

$$V = AZ. \tag{7}$$

From Eq. (7), we can see that the high-dimensional data from the same category are projected at the same point in low-dimensional representation space.

B. OBJECTIVE FUNCTION OF GSNMFS

To make full use of the prior knowledge, in this paper, we propose the GCNMFS method to deal with high-dimensional data. The objective function of the proposed GCNMFS method is given as follows:

$$\begin{aligned} O_{GCNMFS} = &\parallel X - UZ^T A^T \parallel_F^2 + \lambda Tr(Z^T A^T LAZ) \\ &+ 2\mu \parallel U \parallel^p, \end{aligned} \tag{8}$$

where λ and μ are two nonnegative parameters, respectively. The first term is the reconstruction error. The second term is the graph regularization, and the third term uses L_p norm to constrain the basis matrix.

C. OPTIMIZATION

It is obvious that model (8) is non-convex, and thus cannot find the global optimal solution. The multiplicative iterative algorithm is used to solve the model (8). Therefore, we can achieve a local minimum of Eq. (8). Then Eq. (8) can be further rewritten as follow:

$$\begin{aligned} &O_{GCNMFS} \\ &= \parallel X - UZ^T A^T \parallel_F^2 + \lambda Tr(Z^T A^T LAZ) + 2\mu \parallel U \parallel^p \\ &= Tr((X - UZ^T A^T)(X - UZ^T A^T)^T) \\ &\quad + \lambda Tr(Z^T A^T LAZ) + 2\mu \parallel U \parallel^p \\ &= Tr(XX^T) - 2Tr(XAZU^T) + Tr(UZ^T A^T AZU^T) \\ &\quad + \lambda Tr(Z^T A^T LAZ) + 2\mu \parallel U \parallel^p \end{aligned} \tag{9}$$

Let φ_{ij} and ϕ_{ij} be the Lagrange multiplier for constraints $u_{ij} \geq 0, z_{ij} \geq 0$, and $\Psi = [\varphi_{ij}]$, $\Phi = [\phi_{ij}]$, the Lagrange function ℓ is given as follows:

$$\ell = O_{\text{GCNMFS}} + Tr(\Psi U^T) + Tr(\Phi Z^T) \tag{10}$$

Taking the partial derivatives of U and Z for ℓ, we have

$$\frac{\partial \ell}{\partial U} = -2XAZ + 2UZ^T A^T AZ + 2\mu PU^{P-1} + \Psi \tag{11}$$

$$\frac{\partial \ell}{\partial Z} = -2A^T X^T U + 2A^T AZU^T U + 2\lambda A^T LAZ + \Phi \tag{12}$$

According to KKT conditions $\psi_{ij} u_{ij} = 0$ and $\phi_{ij} v_{ij} = 0$, Eqs. (11) and (12) can be further rewritten as follows:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(XAZ)_{ij}}{(UZ^T A^T AZ + \mu PU^{P-1})_{ij}}, \tag{13}$$

$$z_{ij}^{t+1} \leftarrow z_{ij}^t \frac{(A^T X^T U + \lambda A^T WAZ)_{ij}}{(A^T AZU^T U + \lambda A^T DAZ)_{ij}}. \tag{14}$$

D. CONVERGENCE ANALYSIS

In this section, we give the convergence proof of the proposed model (8) using the updating rules (13) and (14). Before giving the convergence proof, we introduce some related definitions and lemmas.

Definition 1: $G(x, x')$ is an auxiliary function of $F(x)$, if $G(x, x')$ satisfies the conditions $G(x, x') \geq F(x)$ and $G(x, x) = F(x)$.

Now we use lemma 1 to give auxiliary functions.

Lemma 1: If G is an auxiliary function of F, then F is a non-increasing function under the update rule

$$x^{t+1} = \arg\min_x G(x, x^t). \tag{15}$$

Proof: From Definition 1, we have

$$F(x^{t+1}) \leq G(x^{t+1}, x^t) \leq G(x^t, x^t) = F(x^t). \tag{16}$$

Then we have

$$F(x^{t+1}) \leq F(x^t)$$

$F_{u_{ab}}$ denotes the part of the objective function O_{GCNMFS} which is only relevant to element U_{ab} in U. Thus, we have

$$\begin{aligned} F'_{u_{ab}} &= \left(\frac{\partial O_{GCNMFS}}{\partial U}\right)_{ab} \\ &= -2(XAZ + 2UZ^T A^T AZ + 2\mu PU^{P-1})_{ab}, \end{aligned} \tag{17}$$

$$\begin{aligned} F''_{u_{ab}} &= \left(\frac{\partial^2 O_{GCNMFS}}{\partial U^2}\right)_{ab} \\ &= 2(Z^T A^T AZ)_{bb} + 2\mu P(P-1)(U^{P-2})_{ab}. \end{aligned} \tag{18}$$

Lemma 2: The function

$$G(u, u_{ab}^t) = F_{u_{ab}}(u_{ab}^t) + F_{u_{ab}}^{'}(u_{ab}^t)(u - u_{ab}^t) + \frac{(Z^T A^T AZ)_{ab} + \mu P(U^{P-1})_{ab}}{U_{ab}^t}(U - U_{ab}^t)^2. \tag{19}$$

is an auxiliary function for $F_{u_{ab}}$.

Proof: Since $G(u, u) = F_{uab}(u)$ is obvious, we only need to prove that $G(u, u_{ab}^t) \geq F_{u_{ab}}(u)$. To achieve this goal, we expand the Taylor series expansion of $F_{u_{ab}}(u)$ as follows:

$$\begin{aligned} F_{u_{ab}}(u) &= F_{u_{ab}}(u_{ab}^t) + F_{u_{ab}}^{'}(u - u_{ab}^t) + \frac{1}{2}F_{u_{ab}}^{''}(u - u_{ab}^t) \\ &= F_{u_{ab}}(u_{ab}^t) + F_{u_{ab}}^{'}(u - u_{ab}^t) \\ &\quad + \left[(Z^T A^T AZ)_{bb} + \mu P(P-1)(U^{P-2})_{ab}\right] \\ &\quad (u - u_{ab}^t)_{ab}^2. \end{aligned} \tag{20}$$

Using Eq. (19) to prove $G(u, u_{ab}^t) \geq F_{uab}(u)$ is equivalent to:

$$\begin{aligned} &\frac{(UZ^T A^T AZ)_{ab} + \mu P(U^{P-1})}{u_{ab}^t} \\ &\geq \frac{1}{2}F_{u_{ab}}^{''} \\ &= (Z^T A^T AZ)_{bb} + \mu P(P-1)(U^{P-2}). \end{aligned} \tag{21}$$

We have

$$\begin{aligned} (UZ^T A^T AZ)_{ab} &= \sum_{l=1}^{k}(U)_{al}(UZ^T A^T AZ)_{lb} \\ &\geq u_{ab}^t(Z^T A^T AZ)_{bb}. \end{aligned} \tag{22}$$

The second term is obvious, and thus (21) holds.

Theorem 1: The objective function O_{GCNMFS} in Eq. (9) is non-increasing under the update rules (13) and (14).

Proof: Replacing $G(u, u_{ab}^t)$ in (15) by (19), we have

$$\begin{aligned} u_{ab}^{t+1} &= u_{ab}^t - u_{ab}^t \frac{F_{ab}^{'}(u_{ab}^t)}{2(UAZZ^T A^T + \mu PU^{P-1})_{ab}} \\ &= u_{ab}^t \frac{(XZ^T A^T)_{ab}}{(UAZZ^T A^T + \mu PU^{P-1})_{ab}}. \end{aligned} \tag{23}$$

Since (19) is an auxiliary function for $F_{u_{ab}}$, $F_{u_{ab}}$ is non-increasing under this update rule. Similarly, we have

$$v_{ab}^{t+1} = v_{ab}^t \frac{(U^T X + \lambda AZW)_{ab}}{(U^T UAZ + \lambda AZD)_{ab}}. \tag{24}$$

In summary, the convergence of the model (9) can be guaranteed using the updating rules (13) and (14).

E. COMPUTATIONAL COMPLEXITY ANALYSIS

In this subsection, we discuss the computation complexity analysis of the optimization scheme of the proposed model. The big O notation is used to denote the complexity of the algorithm. Tables 1 summarized the parameters used in this work.

TABLE 1. Parameters used in complexity analysis.

Parameters	Description
m	number of features for each data point
n	number of data points
k	number of factors
l	number of labeled data points
c	number of classes
p	the number of nearest neighbors, $p = n$

We can count the operation times of each iteration of all methods. Table 2 summarized the results of the proposed GCNMFS and other competitors. From Table 2, it can be seen that the overall costs of our proposed method are $O(mnk)$. Although the proposed GCNMFS considers more prior knowledge of data compared with other methods, it still maintains the same overall computational complexity.

IV. EXPERIMENT ANALYSIS

In this section, we carried out extensive experiments on PIE, MNIST and COIL20 datasets, and verified the effectiveness of our proposed method compared with other methods including k-means (KM), CF, NMF, GNMF, GCNMF and GSNMF. In our experiments, two popular metrics including accuracy (AC) and normalized mutual information (NMI) were used to evaluate the performances of all methods.

A. EVALUATION METRICS

AC is used to measure the proportion of samples with the correct category information in the sample. Suppose given a sample set, this includes n samples. For each sample, l_i is the class label obtained from the experimental prediction, and r_i is the correct label provided by the real dataset. Therefore, we give the definition of AC as follows:

$$AC = \frac{\sum_{i=1}^{n}\delta(r_i, map(l_i))}{n}, \tag{25}$$

where $\delta(x, y)$ is a function equal to 1 when $x = y$, otherwise it is set to 0. $map(li)$ is a mapping function that maps each cluster label r_i to a given equivalent label.

NMI is adopted as the similarity measure between the test sample set and the original sample set. Suppose two clusters C and C' are given, then the mutual information $MI\left(C, C'\right)$ can be defined as follows:

$$MI(C, C') = \sum_{c_i \in G, c_i' \in G'} p(c_i, c_i') \cdot \log \frac{p(c_i, c_i')}{p(c_i) \cdot p(c_i')}, \tag{26}$$

where $p(c_i)$ and $p(c_i')$ denote the possibilities that an image selected arbitrarily from the dataset belongs to the clusters c_i and c_i', respectively, and $p(c_i, c_i')$ denotes the joint possibility that this arbitrarily selected image belongs to the cluster c_i as well as c_i' at the same time. NMI is formulated as

TABLE 2. Computational operation counts for each iteration.

	fladd	flmlt	fldiv	overall
NMF	$2mnk+2(m+n)k^2$	$2mnk+2(m+n)k^2+(m+n)k$	$(m+n)k$	$O(mnk)$
GNMF	$2mnk+2(m+n)k^2+n(p+3)k$	$2mnk+2(m+n)k^2+(m+n)k+n(p+1)k$	$(m+n)k$	$O(mnk)$
CNMF	$(2m+2n-l+c)(n-l+c)k+2(m+n)k^2$	$(2m+2n-l+c)(n-l+c)k+2(m+n)k^2+(m+n-l+c)k$	$(m+n-l+c)k$	$O(mnk)$
GCNMF	$(2m+2n-l+c)(n-l+c)k+2(m+n)k^2+n(p+3)k$	$(2m+2n-l+c)(n-l+c)k+2(m+n)k^2$ $+(m+n-l+c)k+(m+n)k+n(p+1)k$	$(m+n-l+c)k$	$O(mnk)$
GCNMFS	$(2m+2n-l+c)(n-l+c)k$ $+2(m+n)k^2+n(p+3)k+mk$	$(2m+2n-l+c)(n-l+c)k+2(m+n)k^2+(m+n-l+c)k$ $+(m+n)k+n(p+1)k+mk$	$(m+n-l+c)k$	$O(mnk)$

follows:

$$\mathrm{NMI}(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))}. \tag{27}$$

B. PIE FACE DATABASE

PIE face database contains 41,368 multi-posture, light, and expression facial images of 68 individuals. We selected 42 facial images with different light and illumination conditions from each person. All grayscale face images were resized as 32×32 pixies, and each image can be represented by a 1024-dimensional vector. Fig.1 shows some samples from the PIE database.

FIGURE 1. Some samples from PIE dataset.

TABLE 3. The AC of different methods on PIE dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
10	0.208	0.378	0.219	0.404	0.433	0.528	0.530	**0.547**
15	0.227	0.316	0.191	0.463	0.391	0.530	0.486	**0.537**
20	0.211	0.289	0.186	0.426	0.430	0.542	0.471	**0.564**
25	0.201	0.247	0.156	0.466	0.382	**0.521**	0.490	0.517
30	0.220	0.230	0.133	0.435	0.361	0.500	0.477	**0.514**
35	0.194	0.208	0.111	0.389	0.367	0.467	0.447	**0.474**
40	0.200	0.210	0.113	0.482	0.353	**0.515**	0.489	0.503
45	0.191	0.194	0.101	0.423	0.378	0.429	0.419	**0.507**
avg	0.206	0.259	0.151	0.436	0.387	0.504	0.476	**0.520**

In this experiment, we randomly selected T categories samples as the experimental sub dataset to evaluate the proposed method. For each value of T, the average performances were recorded as the final result after all methods were repeated ten times. Tables 3 and 4 show the performances of all methods. Tables 3 and 4, it can be seen that the average performances of GNMF, CNMF and GSNMF are significantly improved over NMF. This is because these three methods consider more prior knowledge of data than traditional NMF. Specifically, GNMF effectively discovers the manifold structure of data using the graph regularizer. CNMF is a semi-supervised learning algorithm and thus makes full use of the label information among data. GSNMF takes advantage of the properties of the solution. However, our proposed GCNMFS method integrates the merits of the above-mention three algorithms. It not only utilizes the geometric manifold structure of data and the label information of known labeled samples, but also adopts the LP smoothing constraint to obtain a smooth and more accurate solution.

TABLE 4. The NMI of different methods on PIE dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
10	0.243	0.401	0.191	0.572	0.485	0.654	0.648	**0.682**
15	0.254	0.381	0.152	0.540	0.485	0.629	0.574	**0.634**
20	0.303	0.395	0.181	0.616	0.530	0.663	0.638	**0.665**
25	0.297	0.389	0.179	0.590	0.535	**0.654**	0.618	0.651
30	0.333	0.376	0.183	0.577	0.532	0.636	0.608	**0.641**
35	0.339	0.373	0.174	0.541	0.550	**0.633**	0.592	0.632
40	0.362	0.394	0.205	0.643	0.556	0.666	0.618	**0.667**
45	0.358	0.373	0.198	0.613	0.572	0.627	0.614	**0.681**
avg	0.311	0.385	0.182	0.586	0.531	0.645	0.613	**0.657**

C. MNIST DATABASE

MNIST database includes a total of 70,000 handwritten digit samples. The gray level of each sample in the database is 8, and each sample can be represented by a vector of size 784. We randomly selected 50 training images from each category samples as the data subset for experiments. Some images from the MNIST dataset are shown in Figure. 2.

FIGURE 2. Some samples from the MNIST dataset.

In this experiment, T categories samples were randomly selected as the data subset, and then mixed them

TABLE 5. The AC of different methods on MNIST dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
4	0.573	0.540	0.523	0.602	0.610	0.633	0.597	**0.638**
5	0.591	0.597	0.564	0.664	0.610	0.726	0.668	**0.726**
6	0.574	0.559	0.539	0.655	0.611	0.690	0.655	**0.692**
7	0.566	0.549	0.540	0.633	0.609	0.672	0.619	**0.674**
8	0.561	0.533	0.533	0.623	0.607	0.648	0.613	**0.654**
9	0.559	0.522	0.521	0.611	0.601	0.624	0.603	**0.629**
10	0.554	0.515	0.508	0.591	0.593	0.622	0.587	**0.627**
avg	0.568	0.545	0.532	0.625	0.606	0.659	0.620	**0.663**

TABLE 6. The NMI of different methods on MNIST dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
4	0.477	0.438	0.426	0.543	0.492	0.554	0.545	**0.559**
5	0.489	0.481	0.452	0.583	0.489	0.622	0.593	**0.623**
6	0.481	0.457	0.438	0.576	0.494	0.587	0.578	**0.590**
7	0.481	0.450	0.442	0.567	0.492	0.585	0.567	**0.587**
8	0.476	0.437	0.435	0.560	0.494	0.570	0.560	**0.573**
9	0.479	0.434	0.430	0.557	0.594	0.555	0.558	**0.559**
10	0.480	0.431	0.423	0.545	0.492	0.554	0.552	**0.558**
avg	0.480	0.447	0.435	0.561	0.507	0.575	0.565	**0.578**

for clustering. All methods were repeated ten times, and the average performances were recorded as the final result. Tables 5 and 6 show the performances of all methods on the MNIST dataset. It is obvious that the average performance of our proposed GCNMFS method shows the superiority in comparison to other methods. The main reason is that our proposed method effectively learns the manifold preserving representation of high-dimensional data, and simultaneously considers the label information of labeled samples and the smoothness of the solution. However, other competitors fail to make full use of the prior knowledge hidden in data compared with the proposed GCNMFS method. Therefore, the proposed GCNMFS method has shown the most representation ability among all methods.

D. COIL20 DATABASE

COIL20 database includes a total of 1440 images from 20 categories. Each image was resized as a 32×32 pixel gray image. Thus, the total samples of the COIL20 database can be represented by a 1024×1440 matrix. Some samples from COIL20 database are shown in Figure 3.

Similarly, we randomly picked out T categories samples from COIL20 database to evaluate the proposed GCNMFS method. The average results of all methods were reported after they were run ten times. Tables 7 and 8 show the performances of all methods in clustering. It can be seen that GCNMF outperforms GNMF and CNMF in terms of the average AC and NMI. The probable reason is that both GNMF and CNMF utilize only one of the prior knowledge of data, and GCNMF simultaneously considers the manifold structure of

FIGURE 3. Some samples from the COIL20 dataset.

TABLE 7. The AC of different methods on COIL20 dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
5	0.684	0.686	0.635	0.828	0.686	0.824	**0.836**	0.826
7	0.585	0.619	0.631	0.821	0.587	0.898	0.823	**0.904**
9	0.745	0.731	0.640	0.889	0.705	0.912	0.878	**0.923**
11	0.689	0.732	0.688	0.812	0.650	0.805	0.811	**0.934**
13	0.707	0.780	0.629	0.857	0.783	0.851	0.855	**0.934**
15	0.694	0.702	0.647	0.839	0.696	0.825	**0.840**	0.824
17	0.607	0.685	0.576	0.815	0.642	0.807	**0.817**	0.809
19	0.632	0.640	0.594	0.754	0.711	0.759	0.751	**0.808**
avg	0.668	0.700	0.630	0.827	0.682	0.835	0.826	**0.870**

TABLE 8. The NMI of different methods on COIL20 dataset.

T	KM	NMF	CF	GNMF	CNMF	GCNMF	GSNMF	GCNMFS
5	0.730	0.722	0.673	0.876	0.721	0.881	0.884	**0.886**
7	0.584	0.590	0.670	0.800	0.568	0.856	0.803	**0.9861**
9	0.731	0.729	0.624	0.879	0.691	0.915	0.871	**0.922**
11	0.771	0.730	0.696	0.878	0.680	0.873	0.878	**0.928**
13	0.769	0.789	0.718	0.914	0.783	0.920	0.914	**0.944**
15	0.727	0.718	0.671	0.878	0.717	0.878	0.881	**0.882**
17	0.703	0.746	0.627	0.865	0.714	0.885	0.867	**0.884**
19	0.725	0.716	0.675	0.851	0.768	0.862	0.857	**0.876**
avg	0.717	0.717	0.669	0.868	0.705	0.883	0.869	**0.913**

data and the label information of the labeled samples. It can be observed that our proposed GCNMFS method is superior to other competitors, mainly because it takes advantage of the most prior knowledge of data, such as label information, manifold structure information and smoothness of solution, compared with other methods in clustering.

E. DISCUSSION ON PARAMETER SETTING

The proposed model contains three parameters λ, μ and P. We carried out some experiments on PIE, MNIST and COIL20 datasets to investigate the parameter sensitivity of the proposed method. Specifically, one parameter is varied when other parameters are fixed. The performance of our

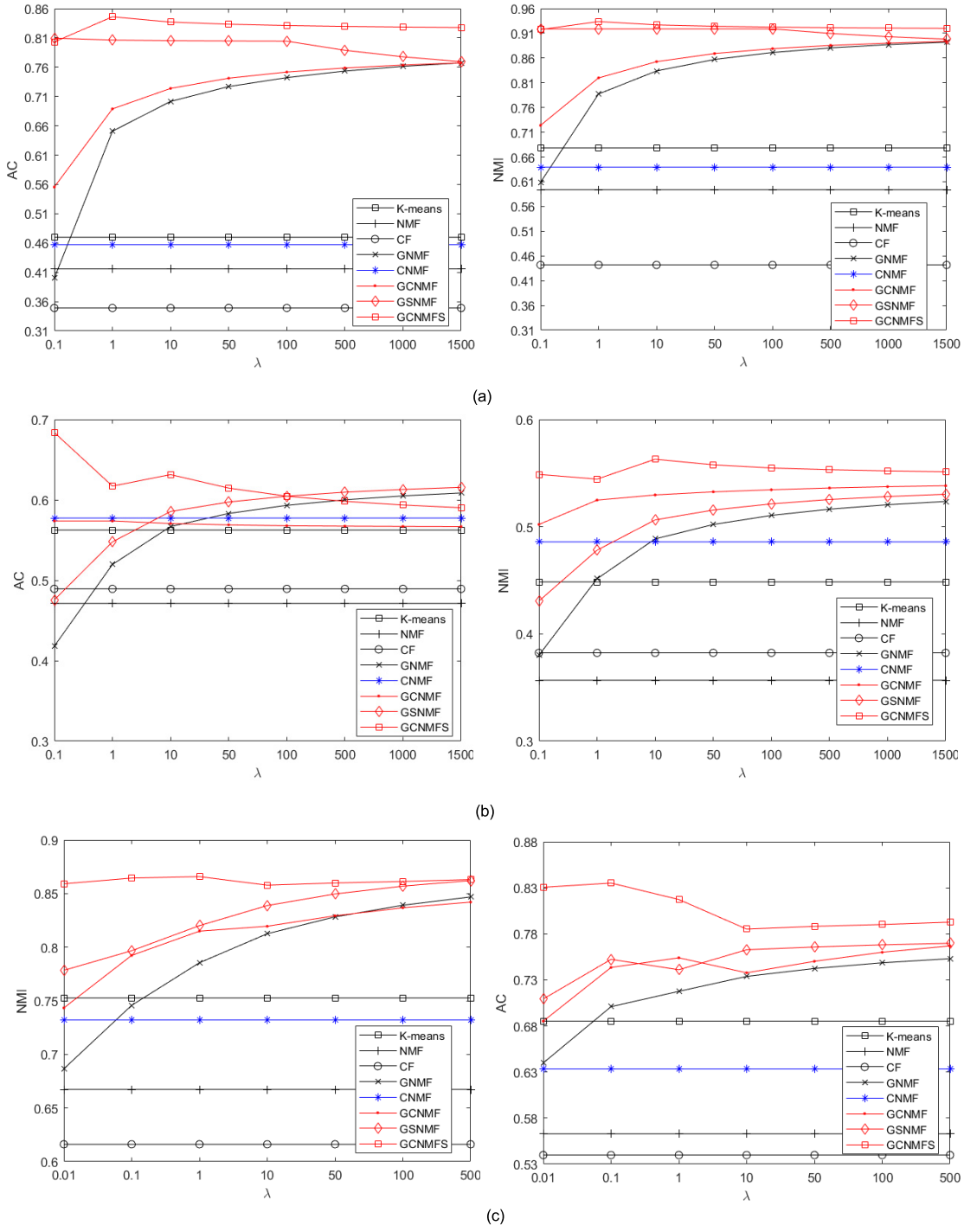

FIGURE 4. Clustering performance versus parameter λ with different values: (a) PIE; (b) MNIST; (c) COIL20.

proposed method with different values of parameter λ is shown in Figure 4. We can see that the proposed GCNMFS method consistently outperforms other competitors when the values of parameter λ is in most cases. The result of our proposed method among various configurations of the parameter μ are illustrated in Figure 5. We find that the proposed method is relatively consistent over a wide range of parameter μ. Figure 6 shows the clustering result of our proposed method varies with the parameter P. It can be seen that GCNMFS delivers stable results where P varied from 1.1 to 1.9.

F. EFFICIENCY ANALYSIS

In this subsection, we evaluate the efficiency of the proposed GCNMFS method with other methods on different datasets.

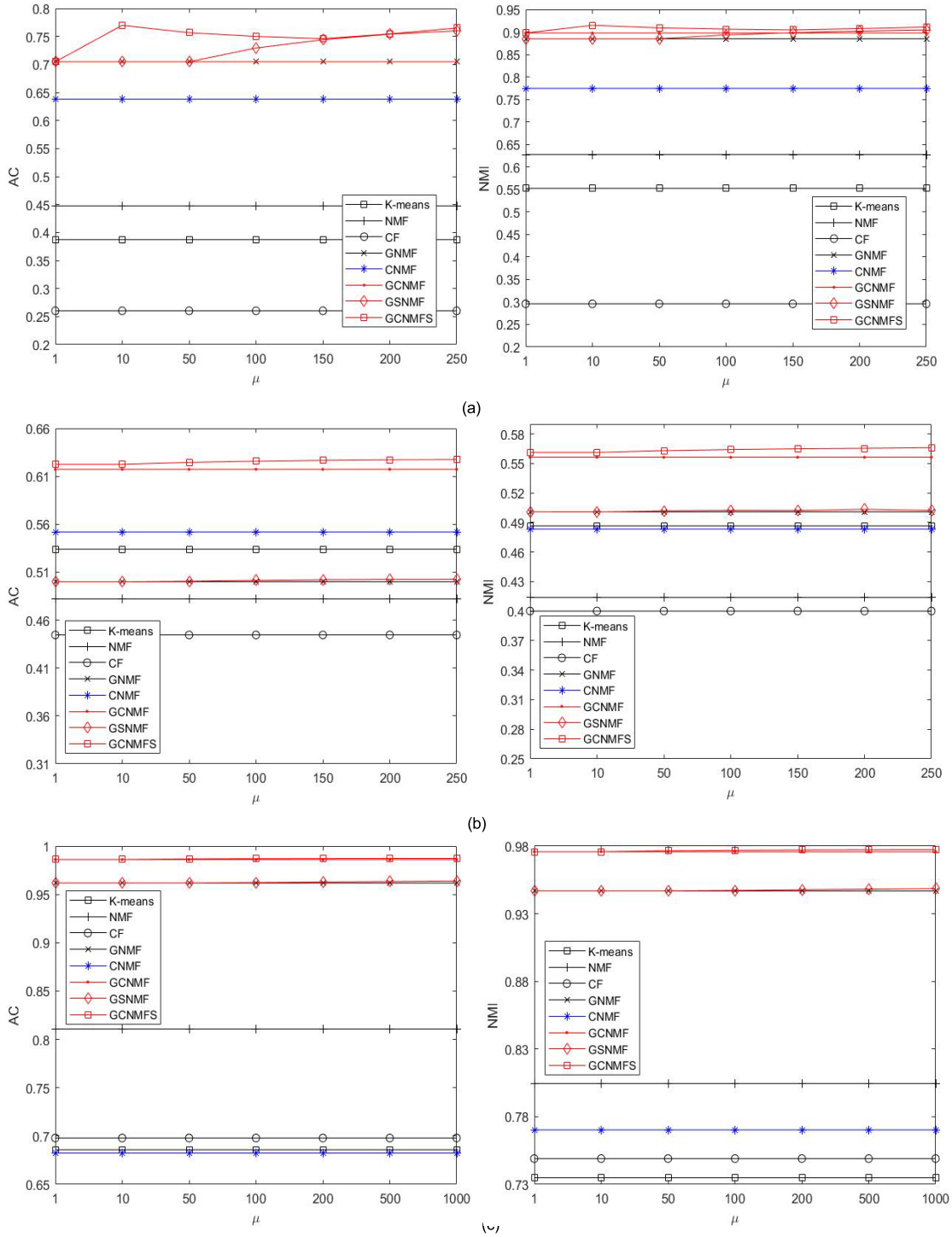

FIGURE 5. Clustering performance versus parameter μ with different values: (a) PIE; (b) MNIST; (c) COIL20.

The experiments were carried out on a Windows 10 machine with Intel Core 2.4 GHz CPU and 16 GB RAM. We randomly sampled 50, 8 and 19 categories from PIE, MNIST and COIL20 datasets as the sub-datasets. We run all methods ten times and reported their average time. The running time of all methods is shown in Table 9. Obviouly, we can see that NMF, CF and GNMF take similar running time in three datasets. It can be observed that both GCNMFS and GCNMF methods take more running time than other methods in each iteration than other methods. This is because they employ more constraints to make full use of the prior knowledge hidden in data.

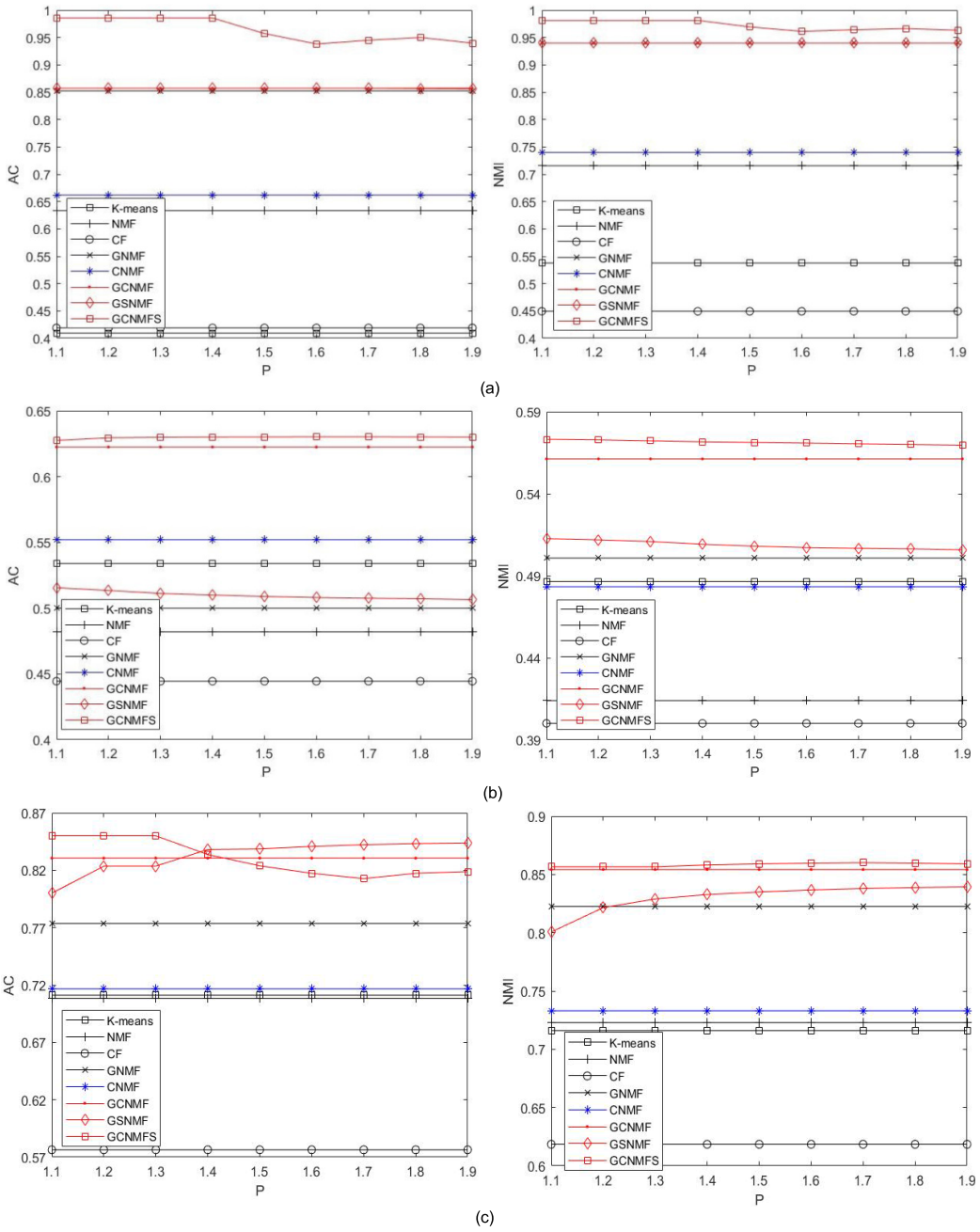

FIGURE 6. Clustering performance versus parameter P with different values: (a) PIE; (b) MNIST; (c) COIL20.

TABLE 9. The running time of each iteration (in milliseconds) of different methods.

	PIE	MNIST	COIL20
NMF	1.9	2.3	2.1
CF	3.6	8.2	5.6
GNMF	5.8	2.0	3.4
CNMF	51.5	112.7	86.9
GCNMF	78.3	299.9	155.1
GCNMFS	89.4	264.7	139.6

V. CONCLUSION

In this paper, we proposed a graph regularized constrained non-negative matrix factorization with L_P smoothing (GCNMFS) method for image representation. Compared with traditional methods, the advantage of the proposed GCNMFS method effectively explores the prior knowledge

including the label information of labeled samples, the manifold structure of data and the smoothness of the solution. In addition, the efficient updating rules are provided to solve the model of GCNMFS. Experimental results on benchmark datasets demonstrate that the proposed GCNMFS method outperforms other state-of-the-art methods for clustering.

However, there are still some aspects of the proposed GCNMFS method that deserve further study. On the one hand, it is an open problem how to consider the prior knowledge of data with fewer parameters. On the other hand, we try to seek a more efficient optimization algorithm to solve the proposed model.

REFERENCES

[1] I. Jolliffe, *Principle Component Analysis*. New York, NY, USA: Springer, 1986.
[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[3] A. Hyvarinen, J. Hurri, and P. O. Hoyer, "Independent component analysis," in *Natural Image Statistics*. London, U.K.: Springer, 2009.
[4] G. Strang, *Introduction to Linear Algebra*. Cambride Boston, MA, USA: Wellesley, 2009.
[5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999, doi: 10.1038/44565.
[6] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," in *Proc. Adv. Neural Inf. Process.*, 2000, pp. 556–562.
[7] W. Xu and Y. Gong, "Document clustering by concept factorization," in *Proc. 27th ACM/SIGIR Int. Symp. Inf. Retr.*, Sheffield, U.K., Jul. 2004, pp. 202–209.
[8] S. E. Palmer, "Hierarchical structure in perceptual representation," *Cognit. Psychol.*, vol. 9, no. 4, pp. 441–474, Oct. 1977, doi: 10.1016/0010-0285(77)90016-0.
[9] A. R. Willse, "Text analysis technique," U.S. Patent 7 158 983, Feb. 2, 2007.
[10] Z. Shu, X. Wu, H. Fan, P. Huang, D. Wu, C. Hu, and F. Ye, "Parameter-less auto-weighted multiple graph regularized nonnegative matrix factorization for data representation," *Knowl.-Based Syst.*, vol. 131, pp. 105–112, Sep. 2017, doi: 10.1016/j.knosys.2017.05.029.
[11] Z. Shu, C. Zhao, and P. Huang, "Local regularization concept factorization and its semi-supervised extension for image representation," *Neurocomputing*, vol. 158, pp. 1–12, Jun. 2015, doi: 10.1016/j.neucom.2015.02.014.
[12] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Inf. Process. Manage.*, vol. 42, no. 2, pp. 373–386, Mar. 2006, doi: 10.1016/j.ipm.2004.11.005.
[13] Y. Tian, X. Li, K. Wang, and F.-Y. Wang, "Training and testing object detectors with virtual images," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 539–546, Mar. 2018, doi: 10.1109/JAS.2017.7510841.
[14] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017, doi: 10.1109/tgrs.2017.2683719.
[15] C. Leng, G. Cai, D. Yu, and Z. Wang, "Adaptive total-variation for non-negative matrix factorization on manifold," *Pattern Recognit. Lett.*, vol. 98, pp. 68–74, Oct. 2017, doi: 10.1016/j.patrec.2017.08.027.
[16] J. Gu, H. Hu, and H. Li, "Local robust sparse representation for face recognition with single sample per person," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 547–554, Mar. 2018, doi: 10.1109/JAS.2017.7510658.
[17] H. Liu, Z. Wu, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1299–1311, Jul. 2012, doi: 10.1109/TPAMI.2011.217.
[18] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010, doi: 10.1109/TIP.2010.2042645.
[19] W. Liu, N. Zheng, and X. Lu, "Non-negative matrix factorization for visual coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2003, p. 293.
[20] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011, doi: 10.1109/TGRS.2011.2144605.
[21] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011, doi: 10.1109/TPAMI.2010.231.
[22] J. Sun, Z. Wang, F. Sun, and H. Li, "Sparse dual graph-regularized NMF for image co-clustering," *Neurocomputing*, vol. 316, pp. 156–165, Nov. 2018.
[23] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, "Nonnegative discriminant matrix factorization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 7, pp. 1392–1405, Jul. 2017, doi: 10.1109/TCSVT.2016.2539779.
[24] F. Sun, M. Xu, X. Hu, and X. Jiang, "Graph regularized and sparse nonnegative matrix factorization with hard constraints for data representation," *Neurocomputing*, vol. 173, pp. 233–244, Jan. 2016.
[25] G. A. Wood and L. S. Jennings, "On the use of spline functions for data smoothing," *J. Biomech.*, vol. 12, no. 6, pp. 477–479, Feb. 1979, doi: 10.1016/0021-9290(79)90033-2.
[26] J. W. Lyons, "Differentiation of solutions of nonlocal boundary value problems with respect to boundary data," *Electron. J. Qual. Theory Differ. Equ.*, vol. 51, pp. 1–11, Jul. 2011, doi: 10.14232/ejqtde.2011.1.51.
[27] L. Xu, "Data smoothing regularization, multi-sets-learning, and problem solving strategies," *Neural Netw.*, vol. 16, no. 5, pp. 817–825, Jun. 2003, doi: 10.1016/S0893-6080(03)00119-9.
[28] C. Leng, H. Zhang, and G. Cai, "A novel data clustering method based on smooth non-negative matrix factorization," in *Proc. ICSM*, Toulon, France, 2018, pp. 24–26.
[29] Y. E. Salehani and S. Gazor, "Smooth and sparse regularization for NMF hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3677–3692, Aug. 2017, doi: 10.1109/jstars.2017.2684132.