

Research Article

Image Retrieval Based on Multiview Constrained Nonnegative Matrix Factorization and Gaussian Mixture Model Spectral Clustering Method

Qunyi Xie and Hongqing Zhu

School of Information Science & Engineering, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Hongqing Zhu; hqzhu@ecust.edu.cn

Received 15 June 2016; Accepted 3 November 2016

Academic Editor: Wanquan Liu

Copyright © 2016 Q. Xie and H. Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Content-based image retrieval has recently become an important research topic and has been widely used for managing images from repertoires. In this article, we address an efficient technique, called MNGS, which integrates multiview constrained nonnegative matrix factorization (NMF) and Gaussian mixture model- (GMM-) based spectral clustering for image retrieval. In the proposed methodology, the multiview NMF scheme provides competitive sparse representations of underlying images through decomposition of a similarity-preserving matrix that is formed by fusing multiple features from different visual aspects. In particular, the proposed method merges manifold constraints into the standard NMF objective function to impose an orthogonality constraint on the basis matrix and satisfy the structure preservation requirement of the coefficient matrix. To manipulate the clustering method on sparse representations, this paper has developed a GMM-based spectral clustering method in which the Gaussian components are regrouped in spectral space, which significantly improves the retrieval effectiveness. In this way, image retrieval of the whole database translates to a nearest-neighbour search in the cluster containing the query image. Simultaneously, this study investigates the proof of convergence of the objective function and the analysis of the computational complexity. Experimental results on three standard image datasets reveal the advantages that can be achieved with the proposed retrieval scheme.

1. Introduction

With the increasing abundance of digital images available from a variety of sources, content-based image retrieval (CBIR) from the huge databases has attracted a lot of attention in the past decade [1–3]. An effective CBIR system should search images by computing the similarity of the extracted features (views) between the user-defined query pattern and images in large-scale collections. Existing visual features include, but are not limited to, intensity, shape, colour, scene, texture, and local invariant. The early CBIR system calculates the similarity with only one feature [4, 5]. This would normally lead to undesirable retrieval results due to insufficient representation. That is, it is quite difficult to effectively distinguish all types of images using a single feature. Generally, to achieve proper results in the CBIR framework, appropriate features that can well capture

the meaningful contents of underlying images are usually integrated [6–8]. However, these visual features are often high dimensional and nonsparse, and direct manipulation of feature descriptors is the most time-consuming operation. To solve this limitation, reduction in dimensionality is widely used. Some popular dimensionality reduction techniques include linear discriminant analysis (LDA) [9], principal component analysis (PCA) [10], independent component analysis (ICA) [11], and singular value decomposition (SVD) [12]. Nonnegative matrix factorization (NMF), as a novel tool for source separation [13–15], can be an alternative way to reduce the dimensionality. It decomposes a nonnegative matrix into two small nonnegative matrices: basis matrix and coefficient matrix. The basis characteristic of NMF is that all elements are not negative, which distinguishes it from other conventional dimensionality reduction techniques. The coefficient matrix models the features of

images as an additive combination of a set of basis vectors. However, as is well known, the original NMF does not always yield a structure constraint for sparse representations of features during decomposition [16]. In recent years, various researches have been reported to extend the standard NMF by enforcing a structure preservation constraint on the objective function [17, 18]. Among them, a graph-embedding objective function of NMF encodes the graph information of the images into the sparse representation [19]. Liu et al. [20] introduced constrained nonnegative matrix factorization (CNMF) in which the label information considered as an additional hard constraint of semisupervised retrieval is directly incorporated into the original NMF. Another fashionable NMF algorithm, topographic NMF (TNMF), was proposed by Xiao et al. [21], in which they imposed a topographic constraint on the objective function to pool together structure-corrected features. The normalization strategies proposed for all above-mentioned NMF-based techniques pertain to the coefficient matrix of NMF decomposition. In fact, studies on the constraints of the basic matrix are still limited.

More recently, some works based on statistical model frameworks have been reported for image retrieval [22, 23]. For example, Zeng et al. [24] introduced an image description algorithm that characterizes a colour image by combining the spatiogram with Gaussian mixture model-(GMM-) based colour quantization. Similarly, another colour image indexing method through spatiochromatic multi-channel GMM was introduced by Piatek and Smolka [25]. Marakakis et al. [26] proposed a relevance feedback method for CBIR using GMM as image representations where Kullback-Leibler (KL) divergence is employed. Its retrieval capability mainly relies on the facts that mixture model-based techniques have provided powerful methodologies for data clustering [27, 28]. This type of technique has the capability to model the uncertainty in a statistical manner. Specifically, GMM fits different shapes of observed data using multivariable Gaussian distribution. A special virtue of the GMM is that it requires estimation of a small number of parameters.

As the aforementioned discussions, in this paper, we propose a novel technique combining multiview constrained NMF and GMM-based spectral clustering (MNGS) for image retrieval. It is noteworthy to highlight the following attractive characteristics of the proposed MNGS. First, multiple features are extracted from the underlying images, and then MNGS integrates these features to obtain a similarity-preserving matrix. Second, we incorporate two constrained terms into the original NMF objective function to represent latent feature information in a low-dimensional space. The first constrained term will help guarantee the basic matrix orthogonality as much as possible to reduce the redundancy. Therefore, this constraint will tend to obtain competitive sparse representations of the visual features. The remaining constraint allows us to consider the latent graph information of the images and satisfy the structure preservation requirement. More importantly, this study provides the proof of convergence of the objective function in detail to ensure that the algorithm converges to the local minima during

decomposition. Third, a multivariable GMM is embedded into the proposed MNGS to model the distribution of the sparse features in terms of the coefficient matrix of NMF. Consequently, images with sparse features belonging to the same Gaussian component are similar, so these images can be labelled with the same subcluster. Considering the complexity of images in repertories, it becomes natural for one to assign more components of GMM to label images. In general, the larger the component, the more accurate the indexing results using GMM. However, the computational cost is more expensive owing to the learning of parameters of GMM. To match the optimal number of components, inspired by the work of [29], finally, spectral clustering based on KL divergence is utilized to merge the GMM components and achieve the desired retrieval results. Specifically, by the eigendecomposition of the similarity matrix measured by KL divergence, similar GMM components can be grouped into several spectral components in a lower-dimensional spectral space. Thus, each sparse feature can be labelled by both a GMM component and a spectral component, which might lead to more accurate clustering results. With the label information of the query image using the mentioned statistical model, similar image retrieval can be effectively performed with the clustering results.

The rest of this paper is organized as follows. Section 2 introduces the related work, including the classical NMF and GMM. In Section 3, we describe the details of the proposed framework, followed by the proof of the convergence of this approach and the complexity analysis in Section 4. Retrieval experiments conducted on real-world image datasets are discussed in Section 5. Section 6 reports the concluding remarks and some suggestions for further research.

2. Preliminaries

This section briefly reviews the classical NMF and GMM. The former is reasonably attractive owing to its intrinsic advantage of providing a low-dimensional description for nonnegative data. The latter shows its accuracy and effectiveness in most clustering tasks.

2.1. Nonnegative Matrix Factorization. The NMF is commonly used to decompose a matrix into two nonnegative matrices under the condition that all its elements are non-negative. Mathematically, given a nonnegative data matrix $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{M \times N}$, NMF aims at finding two nonnegative matrices $B = [B_{id}] \in \mathbb{R}^{M \times D}$ and $G = [G_{jd}] \in \mathbb{R}^{N \times D}$ such that the original data matrix X can be well approximated by

$$X \approx BG^T, \quad (1)$$

where D is a new reduced dimension (inner dimension) and satisfies $D < \min\{M, N\}$. In this way, each column of G^T can be regarded as a sparse representation of the associated column vector in X . There are many criteria to solve the factoring problem and evaluate the quality of

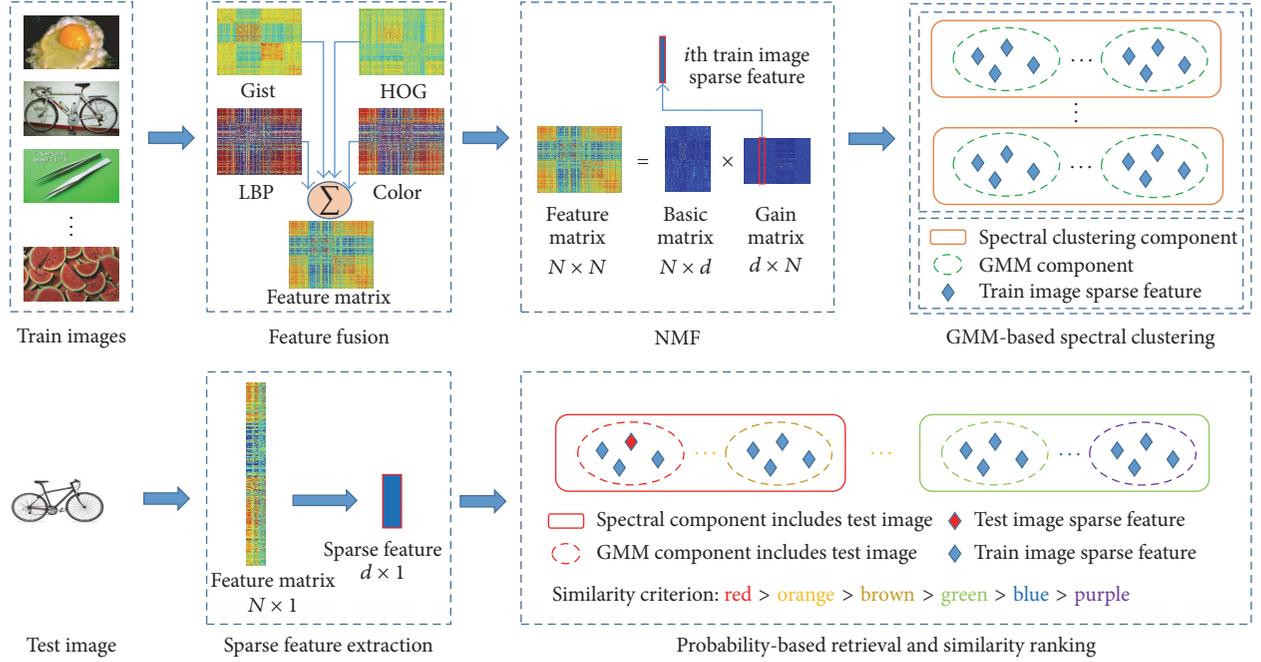


FIGURE 1: Flowchart of the proposed framework. During the training period, raw image features are extracted, whose sparse representations acquired by NMF are labelled via GMM and spectral clustering. Utilizing the sparse features of the query image and the trained GMM, the proposed method then outputs the similarity ranking based on the probabilities that the images belong to the clusters.

the decomposition. Generally, the Euclidean is utilized to construct the objective function; for example,

$$\begin{aligned} \arg \min_{B, G} \quad & \|X - BG^T\|_F^2 \\ \text{s.t.} \quad & B, G \geq 0, \end{aligned} \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. This objective function is proved to be nonconvex and nonincreasing [16]. To minimize the above objective function, Lee and Seung [30] derived the multiplicative updates of the basic matrix B and coefficient matrix G , respectively, as follows:

$$\begin{aligned} B_{id} &\leftarrow B_{id} \frac{(XG)_{id}}{(BG^T G)_{id}}, \\ G_{jd} &\leftarrow G_{jd} \frac{(X^T B)_{jd}}{(GB^T B)_{jd}}. \end{aligned} \quad (3)$$

The above multiplicative updates can ensure convergence to a local optimal solution. Thus, the iteration stops when the objective function converges or the maximum number of iterations is reached.

2.2. Gaussian Mixture Model. The classical Gaussian mixture model assumes that each observed data point is dependent on the label C_j . The density function of GMM at each observation y_i can be described by

$$f(y_i | \Pi, \mathbf{C}) = \sum_{j=1}^K \pi_j \Phi(y_i | C_j), \quad (4)$$

where $\Pi = \{\pi_1, \pi_2, \dots, \pi_K\}$, and the prior distribution π_j indicates the probability of each component belonging to GMM, which satisfies the constraints

$$\sum_{j=1}^K \pi_j = 1, \quad 0 \leq \pi_j \leq 1. \quad (5)$$

Expression (4) can be regarded as a linear combination of several Gaussian components. Each component is a Gaussian distribution that has its own covariance σ_j and mean μ_j , defined as

$$\Phi(y_i | C_j) = \frac{\exp\left\{-\frac{1}{2}(y_i - \mu_j)^T \sigma_j^{-1} (y_i - \mu_j)\right\}}{\sqrt{2\pi} |\sigma_j|}. \quad (6)$$

To optimize the parameter set $C_j = \{\mu_j, \sigma_j\}$, the log-likelihood function of (4) must be maximized by the expectation maximization (EM) algorithm [31], which is expressed as

$$L(\mathbf{C}, \Pi | Y) = \sum_{i=1}^N \log \left(\sum_{j=1}^K \pi_j \Phi(y_i | C_j) \right). \quad (7)$$

3. The Proposed Method

In this section, we introduce an image retrieval approach, called MNGS, which consists of four major parts: feature extraction, multiview constrained NMF, GMM-based spectral clustering, and similarity ranking. Figure 1 shows the overall framework of the proposed MNGS method.

3.1. Feature Extraction. Images may be represented in terms of different features, each of which implies different information, such as orientation, texture, intensity, scene, colour, and spatial relation. To construct the image feature space, this paper utilizes several visual descriptors for the detection of objects.

Scene characteristics play an important role in the description of the object. To obtain the scene information, in this study, the energy spectrum of the grey-scaled image is first filtered by 32 Gabor filters in 4 frequency bands with 8 orientations. Each filtered spectrum image is then decomposed into 4×4 grid subregions leading to a 512-dimensional Gist [32] feature.

Orientation information is a common feature for each object. Here, the grey-scaled image with Gamma correction is divided by 3×3 blocks, which consist of 4×4 cells. By calculating 8 gradient histograms of each cell, the HOG feature [33] with 1152-dimensionality is obtained.

The texture feature is one of the most important characteristics for image retrieval strategy. This is because each object has its own texture in nature. According to [34], LBP is highly suitable to characterize the texture of an object; therefore, this paper adopts a 512-dimensional LBP feature. More specifically, LBP redefines the grey value of each pixel by thresholding a 3×3 neighbourhood, and the histogram of this new grey-scaled image is defined as the LBP feature.

Generally, features extracted from the natural image include colour information. Therefore, it is very important for the image retrieval system to select a colour descriptor that is robust to illumination, colour, and hue. This paper computes a 64-bin histogram of each RGB channel and lists them together, leading to a 192-dimensional ColorHist [35].

Thus, each image can be simultaneously described by four visual features, such as Gist, HOG, LBP, and ColorHist.

3.2. Multiview Constrained NMF. This paper seeks a novel constrained objective function to make it converge to a local minimum. We thus first consider incorporating the structure constraint into the conventional NMF framework for similarity preservation.

Let $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \dots, x_N^{(v)}] \in \mathbb{R}^{D_v \times N}$ be D_v -dimensional feature vectors of the given images in v th view. We then adopt a Gaussian-like heat kernel weighting as follows:

$$s_{ij}^{(v)} = \exp\left(\frac{-\|x_i^{(v)} - x_j^{(v)}\|_F^2}{2(\lambda^{(v)})^2}\right), \quad i, j \in [1, N], \quad (8)$$

where the scalable parameter $\lambda^{(v)}$ stands for the median of all paired distances measured by the Frobenius form. We can then construct the following feature matrix to fuse the different views:

$$S = \frac{1}{N_v} \sum_{v=1}^{N_v} s^{(v)}, \quad (9)$$

where N_v is the view number; in the current study, $N_v = 4$. In classical NMF, the coefficient matrix cannot reflect the latent similar contents shared by different views, which is meaningful to the image retrieval system. To solve this problem, a symmetric similarity measurement matrix is introduced to describe the closeness of the considered images, which is expressed as

$$W = \frac{1}{N_v} \sum_{v=1}^{N_v} \left(\frac{s^{(v)} - I_N}{\sum_{i \neq j} s_{ij}^{(v)}} \right), \quad (10)$$

where I_N is a unit matrix with size $N \times N$. With the similarity matrix defined above, we can enforce the structure constraint into the matrix factorization process by introducing the following regularization:

$$R_G = \text{tr}(G^T \bar{L} G), \quad (11)$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix, \bar{L} is a Laplacian matrix $\bar{L} = \Lambda - W$ [36], and Λ represents a diagonal matrix whose element corresponds to $\Lambda_{jj} = \sum_i W_{ij}$.

In addition, considering the redundancy of the basic matrix, the current paper imposes an orthogonality constraint on each basic vector and expects this orthogonal regularization to sharply reduce the redundancy and make the basic matrix near-orthogonal. Here, the regularization for the basic matrix is defined as

$$R_B = \|B^T B - I\|_F^2. \quad (12)$$

Based on the discussions above, by introducing the structure and orthogonality constraints into the objective function of a classical NMF framework, we obtain the proposed objective function formulated by

$$\begin{aligned} \arg \min_{B, G} \quad & \|S - BG^T\|_F^2 + \alpha \text{tr}(G^T \bar{L} G) + \beta \|B^T B - I\|_F^2 \\ \text{s.t.} \quad & B, G \geq 0, \end{aligned} \quad (13)$$

where α and β are two nonnegative parameters for balancing the factorization error and regularized constraints.

Because the objective function of the proposed regularized NMF is not convex, it is unrealistic to search for a global solution with respect to both B and G . To solve the optimization problem, instead, a novel multiplicative update scheme for B and G is introduced in this paper to achieve a local minimum.

Considering the nonnegativity of the two variables B and G , the Lagrange multipliers $\Phi = [\phi_{id}] \in \mathbb{R}^{N \times D}$ and $\Psi = [\psi_{jd}] \in \mathbb{R}^{N \times D}$ are utilized to optimize the objective function

(13). This yields the corresponding Lagrange function written by

$$\begin{aligned}
\mathcal{L} &= \|S - BG^T\|_F^2 + \alpha \operatorname{tr}(G^T \bar{L}G) + \beta \|B^T B - I\|_F^2 \\
&\quad + \operatorname{tr}(\Phi B^T) + \operatorname{tr}(\Psi G^T) \\
&= \operatorname{tr}\left((S - BG^T)(S - BG^T)^T\right) \\
&\quad + \beta \operatorname{tr}\left((B^T B - I)(B^T B - I)^T\right) + \operatorname{tr}(\Phi B^T) \\
&\quad + \operatorname{tr}(\Psi G^T) + \alpha \operatorname{tr}(G^T \bar{L}G) \\
&= \operatorname{tr}(SS^T) + \operatorname{tr}(BG^T G B^T) - 2 \operatorname{tr}(SGB^T) \\
&\quad + \beta \operatorname{tr}(B^T B B^T B) + \beta \operatorname{tr}(I) - 2\beta \operatorname{tr}(B^T B) \\
&\quad + \operatorname{tr}(\Phi B^T) + \operatorname{tr}(\Psi G^T) + \alpha \operatorname{tr}(G^T \bar{L}G).
\end{aligned} \tag{14}$$

To find B and G that minimize \mathcal{L} , we take the partial derivatives of \mathcal{L} with respect to B and G on both sides

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial B} &= 2BG^T G - 2SG + 4\beta BB^T B - 4\beta B + \Phi = 0, \\
\frac{\partial \mathcal{L}}{\partial G} &= 2GB^T B - 2S^T B + 2\alpha(\Lambda - W)G + \Psi = 0.
\end{aligned} \tag{15}$$

Applying the Karush-Kuhn-Tucker condition [37] $\phi_{id} B_{id} = 0$ and $\varphi_{jd} G_{jd} = 0$ lead to the following update formula for B and G :

$$(2BG^T G + 4\beta BB^T B) B_{id} - (2SG + 4\beta B) B_{id} = 0, \tag{16}$$

$$(2GB^T B + 2\alpha \Lambda G) G_{jd} - (2S^T B + 2\alpha W G) G_{jd} = 0. \tag{17}$$

Based on this condition, it is easy to derive the ultimate update rules as follows:

$$B_{id} \leftarrow B_{id} \frac{(SG + 2\beta B)_{id}}{(BG^T G + 2\beta BB^T B)_{id}}, \tag{18}$$

$$G_{jd} \leftarrow G_{jd} \frac{(S^T B + \alpha W G)_{jd}}{(GB^T B + \alpha \Lambda G)_{jd}}. \tag{19}$$

Regarding the convergence of the update rules, we have the following theorem.

Theorem 1. *The objective function in (13) is nonincreasing under the updating rules in (18) and (19).*

This theorem guarantees that the proposed objective function can converge to a local optimum, and its proof is given via an auxiliary function in Section 4.

3.3. GMM-Based Spectral Clustering. The coefficient matrix of the proposed multiview NMF can be regarded as a low-rank representation of the latent sparse features from various sources of data. Specifically, each row of the coefficient matrix

represents the feature of the associated training sample. Thus, once we obtain the coefficient matrix, clustering of the training data is realized by using the common clustering method, such as K -means or fuzzy C -means (FCM) on the sparse features. In our method, to achieve this, we applied GMM to analyse the data of the coefficient matrix and cluster $G = [g_1, g_2, \dots, g_N]^T$ into K labels, where g_i is a D -dimensional feature vector. Labels are denoted by (C_1, C_2, \dots, C_K) . The proposed GMM-based spectral clustering strategy assumes that it is possible to fit the distribution of different features of the coefficient matrix using a GMM with parameter $\Xi = \{\pi_j, C_j = \{\mu_j, \sigma_j\}\}$. In light of (7), it is clear that the EM algorithm cannot be directly adopted to optimize these parameters. To solve this problem, Jensen's inequality in the form of $\log(\sum_{j=1}^K z_{ij} s) \geq \sum_{j=1}^K z_{ij} \log(s)$ is introduced. Thus, we can rewrite the log-likelihood function (7) as follows:

$$J(\Xi) = \sum_{i=1}^N \sum_{j=1}^K c_{ij} \log(\pi_j \Phi(g_i | C_j)). \tag{20}$$

With the Bayesian theory, the posterior probability c_{ij} indicates the possibility of g_i belonging to the j component by

$$c_{ij} = \frac{\pi_j \Phi(g_i | C_j)}{\sum_{l=1}^K \pi_l \Phi(g_i | C_l)}. \tag{21}$$

Considering the constraints $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^K \pi_j = 1$, the prior distribution π_j can be estimated by setting the partial derivative of the objective function $J(\Xi)$ over it to zero:

$$\frac{\partial}{\partial \pi_j} \left[J(\Xi) - \gamma \left(\sum_{k=1}^K \pi_k - 1 \right) \right] = 0, \tag{22}$$

where γ is the Lagrange multiplier. Equation (22) can yield the following form at step $(t + 1)$:

$$\pi_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^N c_{ij}^{(t)}. \tag{23}$$

Similarly, taking the derivative of the objective function with respect to the mean μ_j and covariance σ_j as zero, we obtain their estimations as follows:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^N c_{ij}^{(t)} g_i}{\sum_{i=1}^N c_{ij}^{(t)}}, \tag{24}$$

$$\sigma_j^{(t+1)} = \frac{\sum_{i=1}^N c_{ij}^{(t)} (g_i - \mu_j^{(t+1)}) (g_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^N c_{ij}^{(t)}}. \tag{25}$$

According to the posterior probability, the final labelling result is determined by

$$g_i \in C_j; \text{ IF } c_{ij} \geq c_{il}, \quad l, j = 1, 2, \dots, K. \tag{26}$$

A crucial problem in GMM-based clustering is how to select a proper number of Gaussian components (sub-clusters). Owing to the complexity of the training samples,

generally, the number of components in GMM is expected to still be larger than the number of artificial labels. The subclusters consisting of sparse features labelled by similar GMM components are then regarded as similar. Finally, we merge the subclusters using spectral clustering to match the number of artificial labels. In our method, two main successive steps are considered. First, the similarity measurement of different Gaussian components is implemented using KL divergence [38], defined as

$$\mathcal{D}(\Phi_p \parallel \Phi_q) = \sum_h \Phi_p(g_h \mid \mathbb{C}_p) \frac{\Phi_p(g_h \mid \mathbb{C}_p)}{\Phi_q(g_h \mid \mathbb{C}_q)}. \quad (27)$$

In particular, according to (27), the explicit expression of KL divergence between two Gaussian distributions can be obtained by

$$\begin{aligned} \mathcal{D}(\Phi_p \parallel \Phi_q) = & \frac{1}{2} \left(\log \frac{|\sigma_q|}{|\sigma_p|} + \text{tr}(\sigma_p(\sigma_q^{-1} - \sigma_p^{-1})) \right. \\ & \left. + (\mu_p - \mu_q)^T \sigma_q^{-1} (\mu_p - \mu_q) \right). \end{aligned} \quad (28)$$

After similarity measurement has been completed, second, the generation of a symmetric similarity matrix is critical to the success of spectral clustering. We then transform the subclusters obtained by GMM into a two-dimensional eigenspace by utilizing eigenvalues and eigenvectors of the symmetric similarity matrix defined by

$$\mathcal{S}_{pq}^{\text{GMM}} = \frac{\mathcal{D}(\Phi_p \parallel \Phi_q) \times \mathcal{D}(\Phi_q \parallel \Phi_p)}{\mathcal{D}(\Phi_p \parallel \Phi_q) + \mathcal{D}(\Phi_q \parallel \Phi_p)}. \quad (29)$$

The eigenvalues and eigenvectors can be obtained by eigen-decomposition of Laplacian matrix $L^{\text{GMM}} = \Lambda^{\text{GMM}} - \mathcal{S}^{\text{GMM}}$, where Λ^{GMM} is a diagonal matrix with $\Lambda_{pp}^{\text{GMM}} = \sum_q \mathcal{S}_{pq}^{\text{GMM}}$. According to graph Laplacian theory [39], the crucial feature information of the subclusters is contained in the eigenvectors corresponding to the second and third minimal eigenvalues. Assume that two chosen eigenvectors are $v^1 = [v_1^1, v_2^1, \dots, v_K^1]^T$ and $v^2 = [v_1^2, v_2^2, \dots, v_K^2]^T$; the j th subcluster is uniquely characterized by (v_j^1, v_j^2) . Some conventional clustering approaches, such as K -means and FCM, are then conducted on the points $\{(v_j^1, v_j^2) \mid j = 1, \dots, K\}$ to obtain the clustering results. Finally, if some subclusters belong to the same labels in the spectral space $(\Omega_1, \Omega_2, \dots, \Omega_C)$, where C is the number of artificial labels, the samples belonging to each subcluster are treated as similar and can be merged as a new spectral component Ω_C . The framework of the MNGS method is outlined as follows.

3.4. Similarity Ranking. The related kernel matrix of the similarity measurement between a given query image and training samples can be expressed by

$$s_{\text{test},i} = \frac{1}{N_v} \sum_{v=1}^{N_v} \exp \left(- \frac{\|x_{\text{test}}^{(v)} - x_i^{(v)}\|_F^2}{2(\lambda^{(v)})^2} \right), \quad i \in [1, N], \quad (30)$$

where $x_{\text{test}}^{(v)}$ represents the v th feature of the query image and $x_i^{(v)}$ corresponds to the same type of feature about the training sample. Meanwhile, using a linear projection matrix ρ ,

$$\rho = (B^T B)^{-1} B^T. \quad (31)$$

We can directly project the kernel matrix (30) into the low-dimensional space and finally obtain the sparse representation of the test sample as follows:

$$g_{\text{test}} = \rho s_{\text{test}}. \quad (32)$$

To rank the training samples based on its similarity against the query image in descending order, the probability that the query image belongs to different GMM components should be computed first. After all probabilities $\{\Phi_1(g_{\text{test}} \mid \mathbb{C}_1), \Phi_2(g_{\text{test}} \mid \mathbb{C}_2), \dots, \Phi_K(g_{\text{test}} \mid \mathbb{C}_K)\}$ have been obtained, the next objective is to retrieve the most similar samples for the given query image. The whole process follows the following three steps: (1) for the samples belonging to each component of GMM, the similarity is compared by ranking the probabilities $\{\Phi_j(g_i \mid \mathbb{C}_j) \mid g_i \in \mathbb{C}_j\}$ of participants in descending order; (2) we compare the probabilities $\{\Phi_j(g_{\text{test}} \mid \mathbb{C}_j) \mid \mathbb{C}_j \in \Omega_c\}$ in each spectral component to distinguish the similarity of different subclusters against the query image with descending order; (3) the last step sorts the probabilities $\{\max\{\Phi_j(g_{\text{test}} \mid \mathbb{C}_j), \mathbb{C}_j \in \Omega_c\} \mid c = 1, 2, \dots, C\}$ to compare the similarity between different spectral components against the query. With the rules and similarity measurements mentioned above, all training samples can be sorted in a reverse order (e.g., the third step, the second step, and the first step). Finally, for each query, we can retrieve the best matches for it using this queue.

Algorithm 1 (multiview constrained NMF followed by GMM spectral clustering (MNGS)).

Input. Multiview feature fusion matrix S using (9), the inner dimension D , the regularized parameters (α, β) , the components K of GMM, and the image dataset.

Output. The trained GMM $\{\pi_j, \mu_j, \sigma_j \mid j = 1, 2, \dots, K\}$; the GMM and spectral component labels for each training sample.

- (1) Calculate the similarity measurement matrix W using (10); initialize the basic matrix B and the coefficient matrix G ;
- (2) *Repeat*
- (3) Update the basic matrix B using (18) and normalize each column of matrix B ;
- (4) Update the coefficient matrix G using (19);
- (5) *Until* the objective function (13) converges
- (6) Initialize the GMM parameter set $\Xi = \{\pi_j, \mathbb{C}_j = \{\mu_j, \sigma_j\} \mid j = 1, 2, \dots, K\}$;
- (7) Calculate the posterior distribution with (21); then update the prior distribution, the mean vector, and the covariance matrix using (23), (24), and (25), respectively;

- (8) Repeat step 7 until reaching the maximum number of iterations, and subclusters can be obtained by (26);
- (9) Calculate the symmetric similarity matrix containing the KL divergence between different GMM components using (29), and spectral clustering is performed to further merge the GMM components.

4. Convergence and Computational Complexity Analysis

4.1. Proof of Convergence. To prove the convergence of the proposed update rules for B and G , we must demonstrate that the following objective function $J(B, G)$ is nonincreasing under the update rules:

$$J(B, G) = \|S - BG^T\|_F^2 + \alpha \operatorname{tr}(G^T \bar{L} G) + \beta \|B^T B - I\|_F^2. \quad (33)$$

The convergence proof of the proposed method benefits from an auxiliary function, which is characterized by the following lemma.

Lemma 2. *If h is an auxiliary function of J and the conditions $h(x, x') \geq J(x)$ and $h(x, x) = J(x)$ are satisfied, then J will be convergent under the update*

$$x^{t+1} = \arg \min_x h(x, x^t). \quad (34)$$

Proof. Obviously, according to the conditions, we have

$$J(x^{t+1}) \leq h(x^{t+1}, x^t) \leq h(x^t, x^t) = J(x^t). \quad (35)$$

The equality $J(x^{t+1}) = J(x^t)$ holds only if x^t is the local minimum of $h(x, x^t)$.

Because the update operations defined by (18) and (19) are element-wise in nature, if we let G be a constant, it is sufficient to verify that $J(B, G) = J(B)$ is convergent for any element B_{id} in B . To accomplish this, this paper defines an auxiliary function regarding B_{id}^t as follows:

$$h(b, B_{id}^t) = J_{id}(B_{id}^t) + J'_{id}(B_{id}^t)(b - B_{id}^t) + \frac{(BG^T G)_{id} + (2\beta BB^T B)_{id} + (2\beta B)_{id}}{B_{id}^t} (b - B_{id}^t)^2. \quad (36)$$

From (36), it is easy to find $h(B_{id}^t, B_{id}^t) = J(B_{id}^t)$; thus, the problem is reduced to prove $h(b, B_{id}^t) \geq J_{id}(b)$. To achieve this, the auxiliary function in (36) is compared with the Taylor series expansion of $J_{id}(b)$, given by

$$J_{id}(b) = J_{id}(B_{id}^t) + J'_{id}(B_{id}^t)(b - B_{id}^t) + \frac{1}{2} J''_{id}(B_{id}^t)(b - B_{id}^t)^2, \quad (37)$$

where J'_{id} and J''_{id} represent the first and second derivatives, respectively. Using (33), we have

$$J'_{id}(B_{id}^t) = (2BG^T G - 2SG + 4\beta BB^T B - 4\beta B)_{id},$$

$$J''_{id}(B_{id}^t) = (2(G^T G)_{dd} + 4\beta(2(B^T B)_{dd} + (BB^T)_{ii}) - 4\beta). \quad (38)$$

Consequently, comparing (36) to (37) by employing (38), the problem of $h(b, B_{id}^t) \geq J_{id}(b)$ can be equivalently transformed to prove

$$\frac{(BG^T G)_{id} + (2\beta BB^T B)_{id} + (2\beta B)_{id}}{B_{id}^t} \geq (G^T G)_{dd} + 2\beta(2(B^T B)_{dd} + (BB^T)_{ii} - 1). \quad (39)$$

With the definition of matrix multiplication, the left hand of the inequality above is certified as

$$(BG^T G)_{id} = \sum_{l=1}^D B_{il}^t (G^T G)_{ld} \geq B_{id}^t (G^T G)_{dd}. \quad (40)$$

Additionally, as mentioned in Algorithm 1, each column of $B = [b_1, b_2, \dots, b_D]$ is normalized during each iteration. This means that $(B^T B)_{dd} = b_d^T b_d = 1$. With this conclusion, the remaining part in (39) can be equivalently proved by

$$2\beta(BB^T B + B)_{id} = 2\beta \left(\sum_n (BB^T)_{in} B_{nd}^t + B_{id}^t \right) \geq 2\beta((BB^T)_{ii} B_{id}^t + B_{id}^t) = 2\beta((BB^T)_{ii} + 2(B^T B)_{dd} - 1) B_{id}^t. \quad (41)$$

Substituting $h(b, B_{id}^t)$ of (36) into (34), we can arrive at the local minimum of the auxiliary function denoted by

$$B_{id}^{t+1} = \arg \min_b h(b, B_{id}^t) = B_{id}^t \frac{(SG + 4\beta B)_{id}}{(BG^T G + 2\beta BB^T B + 2\beta B)_{id}}. \quad (42)$$

We rewrite (42) as

$$(2BG^T G + 4\beta BB^T B + 4\beta B) B_{id} - (2SG + 8\beta B) B_{id} = 0. \quad (43)$$

It is obvious that (43) is equivalent to the expression of (16); therefore, the minimum solution (42) is equivalent to the update rule for the basis matrix in (18). Now, combining (36), (37), (39), and (42), we can obtain

$$J(B_{id}^{t+1}) \leq h(B_{id}^{t+1}, B_{id}^t) \leq h(B_{id}^t, B_{id}^t) = J(B_{id}^t). \quad (44)$$

Next, following a procedure similar to that described above, fixing B as a constant, the update rule for G will be

proved to be nonincreasing. Let $J_{jd}(g)$ represent the element-based objective function. Similarly, the proof begins with the introduction of an auxiliary function with respect to G_{jd}^t , defined by

$$h(g, G_{jd}^t) = J_{jd}(G_{jd}^t) + J'_{jd}(G_{jd}^t)(g - G_{jd}^t) + \frac{(GB^T B)_{jd} + (\alpha\Lambda G)_{jd}}{G_{jd}^t} (g - G_{jd}^t)^2. \quad (45)$$

Because $h(G_{jd}^t, G_{jd}^t) = J(G_{jd}^t)$ is obvious, the Taylor series expansion of $J_{jd}(g)$ is then utilized to prove the inequality $h(g, G_{jd}^t) \geq J_{jd}(g)$, expressed as

$$J_{jd}(g) = J_{jd}(G_{jd}^t) + J'_{jd}(G_{jd}^t)(g - G_{jd}^t) + \frac{1}{2}J''_{jd}(G_{jd}^t)(g - G_{jd}^t)^2. \quad (46)$$

According to (33), the corresponding first- and second-order derivatives of $J_{jd}(g)$ regarding G_{jd}^t can be formulated by

$$J'_{jd}(G_{jd}^t) = (2GB^T B - 2S^T B + 2\alpha\bar{L}G)_{jd}, \quad (47)$$

$$J''_{jd}(G_{jd}^t) = (2(B^T B)_{dd} + 2\alpha\bar{L}_{jj}).$$

We could find without difficulty from (45) and (46) that $h(g, G_{jd}^t) \geq J_{jd}(g)$ yields

$$\frac{(GB^T B)_{jd} + (\alpha\Lambda G)_{jd}}{G_{jd}^t} \geq ((B^T B)_{dd} + \alpha\bar{L}_{jj}). \quad (48)$$

Obviously, we have the following inequality:

$$(GB^T B)_{jd} = \sum_{l=1}^D G_{jl}^t (B^T B)_{ld} \geq G_{jd}^t (B^T B)_{dd},$$

$$(\alpha\Lambda G)_{jd} = \sum_n \alpha\Lambda_{jn} G_{nd}^t \geq \alpha\Lambda_{jj} G_{jd}^t \quad (49)$$

$$\geq \alpha(\Lambda_{jj} - W_{jj}) G_{jd}^t = \alpha\bar{L}_{jj} G_{jd}^t.$$

Substituting $h(g, G_{jd}^t)$ into (34), the update rule of G in (19) can be obtained as a local optimum of the auxiliary function (45):

$$G_{jd}^{t+1} = \arg \min_g h(g, G_{jd}^t) = G_{jd}^t \frac{(S^T B + \alpha W G)_{jd}}{(GB^T B + \alpha\Lambda G)_{jd}}. \quad (50)$$

Again combining (45), (46), (48), and (50), the following inequality is inferred as

$$J(G_{jd}^{t+1}) \leq h(G_{jd}^{t+1}, G_{jd}^t) \leq h(G_{jd}^t, G_{jd}^t) = J(G_{jd}^t). \quad (51)$$

With (44) and (51), the objective function $J(B, G)$ can be proved to be nonincreasing with (18) and (19) as

$$J(B^{t+1}, G^{t+1}) \leq J(B^t, G^{t+1}) \leq J(B^t, G^t). \quad (52)$$

Thus, we have proved the convergence of Theorem 1. \square

4.2. Complexity Analysis. This subsection mainly analyses the computational complexity of the proposed MNGS algorithm using the large \mathcal{O} notation. The computational cost of MNGS primarily comprises three parts. First, we need $\mathcal{O}(2N^2(\sum_{v=1}^{N_v} D_v))$ float point operations to obtain the multi-view feature fusion matrix S and the similarity measurement matrix W in terms of (9) and (10), respectively. Based on the updating rules reported by (18) and (19), the cost required for updating matrices B and G is then $\mathcal{O}(N^2 D)$. Additionally, the proposed MNGS requires $\mathcal{O}(D^3)$ to work with the matrix operation for each component of GMM and $\mathcal{O}(D^2)$ to label the membership for each sample of the subclusters. Thus, assuming that subclusters K , samples N , and iterations m have been considered, the total time complexity of the EM algorithm is $\mathcal{O}(mKND^2 + mKD^3)$. Finally, concentrating on the spectral clustering represented by (28) and (29), the cost spent on the similarity matrix of GMM components is $\mathcal{O}(K^2 D^3)$. Except for this cost, the computational complexity of clustering the similar components is $\mathcal{O}(K^3)$. Considering the relationships $N > K$ and $N > D$, the time cost of the spectral clustering can be ignored. Thus, from the foregoing complexity analysis, if the algorithm terminates after t iterations, the maximum overall cost of the proposed MNGS is $\mathcal{O}(2N^2(\sum_{v=1}^{N_v} D_v) + tN^2 D + mKND^2 + mKD^3)$.

5. Experimental Results

This section presents a series of experiments to verify the accuracy and effectiveness of the proposed MNGS algorithm. In the process of training, the searching performance evaluates the potential clustering capability held by multiview constrained NMF. Once the clustering result is obtained using GMM-based spectral clustering, all N training images in the dataset are ranked according to the probabilities that the images and the query belong to the clustering components in the retrieval phase. A returned image set with retrieval length \bar{L} in the form of a percentage is then viewed as the $N \times \bar{L}$ images nearest the query. Three publicly available datasets used for the present study include Calth-256 (Calth-256: http://www.vision.caltech.edu/Image_Datasets/Caltech256/) [40], CIFAR-10 (CIFAR-10/100: <http://www.cs.toronto.edu/~Ekriz/cifar.html>) [41], and CIFAR-100 (CIFAR-10/100: <http://www.cs.toronto.edu/~kriz/cifar.html>) [41]. The evaluations are carried out on a general purpose computer with an Intel Core 2 Duo 2.1GHz CPU (T6570) and 3GB of RAM under the Windows 7 environment. All algorithms have been implemented using the MATLAB 2010b software application.

5.1. Preparation for Datasets. The Calth-256 dataset includes 30,607 images and consists of 257 categories; each category contains more than 80 colour images. These images cover a wide variety of nature scenes and artificial objects. The experiment randomly selects four categories, named fried-egg, touring-bike, tweezer, and watermelon, with a total of 415 images as training samples, and forty images are randomly chosen from these four categories as query images.

CIFAR-10 and CIFAR-100 are databases of 60,000 images with 10 groups and 100 groups, respectively. For the former dataset, we randomly consider 541 images selected from 10 groups as the training set, and 10 images from each group are randomly selected as query samples. We also perform the experiment over the CIFAR-100 dataset, in which we select 400 images from four categories: beetle, bicycle, chair, and mountain. Similarly, the testing set consists of forty images randomly selected from the four categories. In this work, the images are resized to the resolution of 128×128 pixels for convenience of feature extraction.

5.2. Parameter Selection. It can be seen from Section 3 that there are four parameters in the proposed MNGS: inner dimension D of NMF, component numbers K of GMM, and regularization parameters α , β . It is natural to think that the retrieval accuracy might be affected by these parameters; therefore, this subsection first discusses the parameters we choose for the proposed method. The evaluation is performed based on the two most commonly used metrics: precision and recall rate [42]. The precision is defined as the ratio of the relevant images in all retrieval images:

Precision

$$= \frac{\text{Number of relevant images with retrieval length}}{\text{Total number of images with retrieval length}} \quad (53)$$

The recall represents the ratio of the retrieved relevant image to all relevant images in the dataset, defined as

Recall

$$= \frac{\text{Number of relevant images with retrieval length}}{\text{Total number of relevant images in the dataset}} \quad (54)$$

The NMF-based method has shown great advantage in working with high-dimensional data because NMF can reduce the dimensionality of considered data while preserving the characteristics of the underlying data. It is clear that NMF with a small inner dimension D can efficiently reduce the complexity of decomposition. However, an excessively small D could lead to an unfavourable retrieval result because of the lost information of the data. Conversely, having a large inner dimension D may incur extra computational cost. Thus, one key issue is that D can yield a “meaningful” retrieval result. Following a similar consideration to [43], the work first discusses the influence of D , for example, $D = 8, 16, 32, 64, 96, 128$, on the retrieval results when parameters $\alpha = 0.15$ and $\beta = 0.325$ are fixed. Table 1 summarizes the corresponding results under different choices of inner dimension D . As we can observe from this table, for the proposed framework, its performance at $D = 64$ for datasets Calth-256 and CIFAR-100 and $D = 32$ for CIFAR-10 is better than that at other D values.

Next, we test the performance of the proposed method with varying components K of GMM. As mentioned before, each artificially selected category can be further subdivided

TABLE 1: Comparison retrieval performance in terms of D by fixing retrieval length ($\bar{L} = 0.1$), regularization parameters ($\alpha = 0.15$, $\beta = 0.325$), and GMM component $K = 12, K = 10$, and $K = 6$ for Calth-256, CIFAR-10, and CIFAR-100, respectively.

D	Calth-256		CIFAR-10		CIFAR-100	
	Recall rate	Precision	Recall rate	Precision	Recall rate	Precision
8	0.1332	0.3561	0.1397	0.1411	0.1401	0.3469
16	0.1424	0.3835	0.1564	0.1574	0.1513	0.3706
32	0.1421	0.3726	0.1611*	0.1619*	0.1522	0.3750
64	0.1491*	0.3915*	0.1464	0.1467	0.1554*	0.3850*
96	0.1466	0.3805	0.1449	0.1467	0.1523	0.3719
128	0.1263	0.3232	0.1443	0.1459	0.1480	0.3600

* Bold values indicate the best result of the column.

into several subclusters owing to the abundant images of the category. Considering that the number of subclusters labelled by GMM is always larger than the number of its artificial categories, the experiment assigns the subclusters (component numbers) of GMM to ($K = 6, 8, 10, 12, 14$, and 16) for datasets Calth-256 and CIFAR-100 and ($K = 12, 14, 18, 22, 26$, and 30) for CIFAR-10. Table 2 records the corresponding retrieval results for the three datasets with different component numbers. Based upon the results of precision and recall rate analysis, it is obvious that $K_1 = 6$ for Calth-256 and CIFAR-100 and $K_2 = 12$ for CIFAR-10 provide higher retrieval precision. Thus, this experiment indicates that the resulting subclusters near the artificial class labels and an excessively large component number K might reduce the retrieval precision.

In the proposed MNGS, there are two regularization coefficients α, β , which represent the trade-off between factorization errors and regularized constraints. The following values for regularization coefficient α are considered first: 0.05, 0.15, 0.35, 0.55, 0.75, and 0.95. Table 3 illustrates the comparison using different datasets with a fixed coefficient $\beta = 0.325$, and the retrieval results indicate that the best performance is obtained using $\alpha = 0.15$. Similarly, we investigate the effect of parameter β on the retrieval performance, and experimental results using different values of β are depicted in Table 4. In this case, we can see from the table that using $\beta = 0.325$ obviously leads to both the highest precision and recall rate for all three datasets. From these results, we find if the parameters are set to $\alpha = 0.15$ and $\beta = 0.325$, our approach provides the best retrieval results.

To investigate the influence of feature dimensions on the retrieval results and execution time, the proposed method adopts two different feature dimensions. One is mentioned, which we refer to as a higher dimension case. Another case we just call the lower-dimensional features where Gist feature, HOG feature, LBP feature, and ColorHist feature are 256, 576, 256, and 96, respectively. We can observe from Table 5 that the reduction of feature dimensions helps to reduce the time consumption. But then, it is accompanied with

TABLE 2: Comparison retrieval performance in terms of K by fixing retrieval length ($\bar{L} = 0.1$), regularization parameters ($\alpha = 0.15, \beta = 0.325$), and inner dimension $D = 64, D = 32$, and $D = 64$ for Calth-256, CIFAR-10, and CIFAR-100, respectively.

K	Calth-256		K	CIFAR-10		K	CIFAR-100	
	Recall rate	Precision		Recall rate	Precision		Recall rate	Precision
6*	0.1754*	0.4585*	12*	0.1726*	0.1724*	6*	0.1550*	0.3812*
8	0.1604	0.4238	14	0.1706	0.1711	8	0.1158	0.2844
10	0.1327	0.3476	18	0.1565	0.1580	10	0.1172	0.2900
12	0.1550	0.4073	22	0.1471	0.1485	12	0.1051	0.2650
14	0.1545	0.4079	26	0.1544	0.1559	14	0.1037	0.2600
16	0.1454	0.3866	30	0.1477	0.1485	16	0.1030	0.2612

*Bold values indicate the best result of the column.

TABLE 3: Comparison retrieval performance in terms of regularization parameter α by fixing retrieval length ($\bar{L} = 0.1$), regularization parameter ($\beta = 0.325$), inner dimension $D = 64$, and GMM components $K = 6$ for Calth-256 and CIFAR-100 datasets and $D = 32$ and $K = 12$ for CIFAR-10 dataset.

α	Calth-256		CIFAR-10		CIFAR-100	
	Recall rate	Precision	Recall rate	Precision	Recall rate	Precision
0.05	0.1791	0.4634	0.1587	0.1581	0.1283	0.3162
0.15*	0.1862*	0.4896*	0.1776*	0.1792*	0.1508*	0.3802*
0.35	0.1697	0.4463	0.1554	0.1552	0.1316	0.3200
0.55	0.1613	0.4415	0.1475	0.1470	0.1459	0.3613
0.75	0.1594	0.4384	0.1598	0.1602	0.1453	0.3556
0.95	0.1538	0.4085	0.1596	0.1604	0.1370	0.3394

*Bold values indicate the best result of the column.

moderately decreasing of precision and recall rate. Unless otherwise specified, the first feature dimensions are used in our method.

5.3. Performance and Comparison. This subsection evaluates the retrieval performance of the proposed MNGS algorithm. Considering some random initialization of the proposed method, repeated retrieval performance should be considered on the same dataset to obtain a relatively stable result. More precisely, taking Calth-256 and CIFAR-100 datasets as an example, for each retrieval task, the training process is conducted once, yet 40 iterations of testing for 40 query images that are selected from the corresponding dataset would be carried out to assess the retrieval accuracy. As mentioned before, the datasets used in this experiment are parts of the complete datasets but are still named Calth-256, CIFAR-10, and CIFAR-100 for convenience in the following description. The experiments will assign parameters D, K, α, β in terms of the above discussion. We perform the comparison with three state-of-the-art approaches on these datasets. The first method measures the similarity by modelling the RGB image with Gaussian distributions and calculating the KL divergence between the two GMMs, which we call KL-GMM [44] in

our experiment. The second approach detects the resembled images by labelling the dataset via multiview joint nonnegative matrix factorization (JNMF) [45]. The third algorithm searches the nearest neighbour of the query image in a low-dimensional feature space via multiview alignment hashing (MAH) [43]. The precision and recall rate versus different retrieval lengths are adopted to compare the behaviour of different approaches, and the results are exhibited in Figure 2. For the precision curve, we can observe that, compared with using the CIFAR-10 dataset, all approaches achieve higher-precision values while using Calth-256 and CIFAR-100. We believe that this is because the samples of CIFAR-10 present more complex details. Another interesting finding is that the plots of precision tend to a constant m/N with increasing retrieval length values, where m represents the total number of images in each category. This is because most images that heavily match the query assemble in the former queue owing to the similarity ranking of the training set. Additionally, with increasing retrieval length, as a common phenomenon, the recall rates of all algorithms exhibit an ascending trend. It is noteworthy that the algorithms with feature extraction always achieve higher retrieval accuracy, such as JNMF, MAH, and MNGS. In addition, in the proposed method, the GMM-based spectral clustering scheme achieves better performance than regression-based hashing of MAH. Therefore, compared with JNMF, MAH, and KL-GMM shown in Figure 2, it is obvious that MNGS outperforms its competitors in terms of assessment criteria. Table 6 presents the top six images retrieved from the respective categories corresponding to the query images. The advantage of the algorithm can be found from the number of correctly matched images against query images. It can be observed that the retrieval accuracy for KL-GMM is slightly poorer than the other methods. This is consistent with the results depicted by Figure 2. The proposed method is proved again to moderately improve the retrieval performance.

5.4. Time Consumed. This subsection compares the computational time for each of the aforementioned approaches. The runtimes including the training time and testing time of the different algorithms are illustrated in Figure 3. Owing to the feature extraction steps of the JNMF, MAH, and MNGS

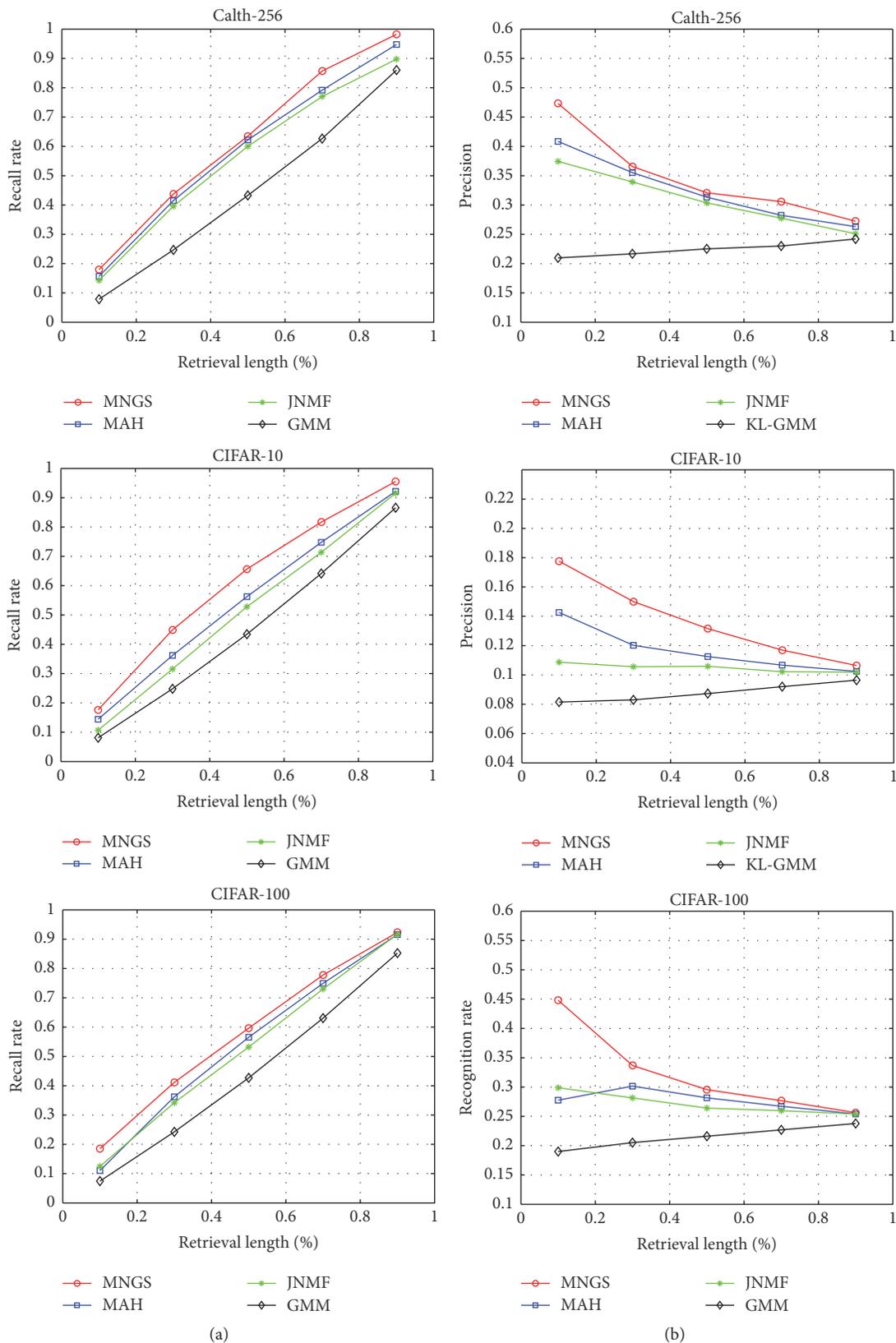


FIGURE 2: The comparison of image retrieval results on the three datasets in terms of recall rate (a) and precision (b).

TABLE 4: Comparison retrieval performance in terms of regularization parameter β by fixing retrieval length ($\bar{L} = 0.1$), regularization parameter ($\alpha = 0.15$), inner dimension $D = 64$, and GMM components $K = 6$ for Calth-256 and CIFAR-100 datasets and $D = 32$ and $K = 12$ for CIFAR-10 dataset.

β	Calth-256		CIFAR-10		CIFAR-100	
	Recall rate	Precision	Recall rate	Precision	Recall rate	Precision
0.125	0.1754	0.4598	0.1707	0.1711	0.1293	0.3206
0.225	0.1548	0.4085	0.1560	0.1559	0.1376	0.3394
0.325*	0.1822*	0.4793*	0.1768*	0.1776*	0.1531*	0.3827*
0.425	0.1526	0.4018	0.1751	0.1761	0.1392	0.3450
0.525	0.1633	0.4280	0.1552	0.1552	0.1536	0.3781
0.825	0.1628	0.4280	0.1746	0.1752	0.1386	0.3406

*Bold values indicate the best result of the column.

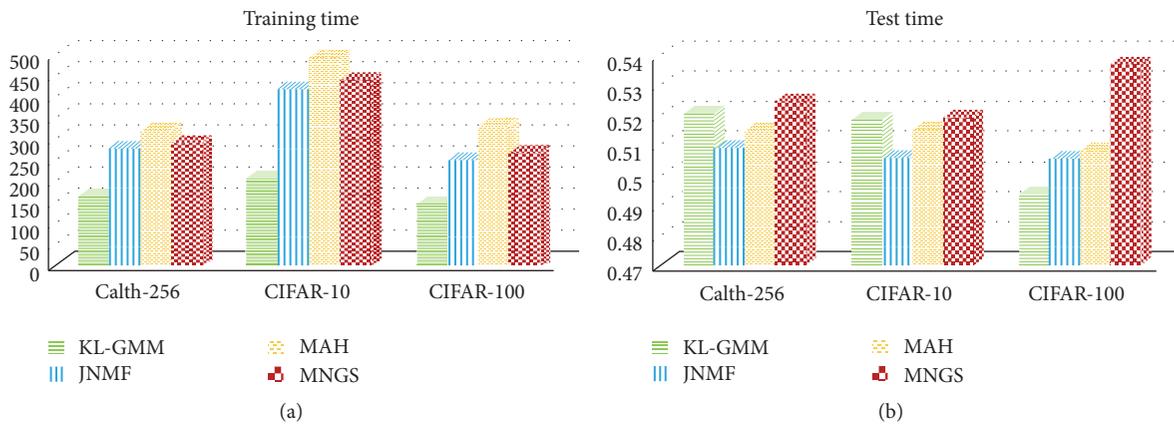


FIGURE 3: Average runtime of the four algorithms on three datasets, (a) training times and (b) test times.

algorithms, the computational times required in the training phase could have been increased. In contrast, KL-GMM costs remarkably less training time. However, for the average CPU time in the testing phase, as we can see from this figure, no significant difference could be found for all approaches; only KL-GMM and MNGS spend slightly more time than the other methods on Calth-256 and CIFAR-10, as does MNGS on CIFAR-100.

6. Conclusions

In this paper, a novel retrieval algorithm called MNGS was presented, in which the sparse representations of the training data were precisely learnt via constrained NMF with pooling of multiview features. The main contribution of this work includes the following four aspects: (1) the proposed method imposed structure constraints and orthogonal regularization into the standard NMF framework and demonstrated its convergence; (2) by incorporating multiple structure-correlated features together, the coefficient matrix of MNGS preserved features from all images in the low-dimensional space; (3) the EM algorithm was incorporated to estimate the GMM components for the sparse features; and

TABLE 5: The influence of feature dimensions on retrieval results and execution time.

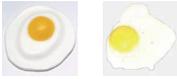
Datasets	Calth-256		CIFAR-10		CIFAR-100	
	High	Low	High	Low	High	Low
Recall rate	0.1822	0.1622	0.1768	0.1530	0.1531	0.1326
Precision	0.4793	0.4226	0.1776	0.1628	0.3827	0.3562
Training Time	285.5	210.3	425.6	320.1	250.3	173.6
Test time	0.522	0.391	0.515	0.348	0.534	0.355

(4) spectral clustering based on the KL divergence strategy was designed to obtain the final retrieval results. Experiments on the three standard datasets indicated that, when appropriate parameters were used, the desired image retrieval could be achieved in terms of searching accuracy and effectiveness.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

TABLE 6: The top 6 images retrieved by using different algorithms on the Calth-256 dataset. The first column is the query image, followed by the retrieved images listed in the form of columns.

	KL-GMM	JNMF	MAH	MNGS
	 	 	 	 
	  	  	  	  
	  	  	  	  
	  	  	  	  

Acknowledgments

This work was supported by the National Nature Science Foundation of China under Grant 61371150.

References

[1] B. Wang, X. Zhang, X.-Y. Zhao, Z.-D. Zhang, and H.-X. Zhang, "A semantic description for content-based image retrieval," in *Proceedings of the 7th International Conference on Machine Learning and Cybernetics (ICMLC '08)*, pp. 2466–2469, Kunming, China, July 2008.

[2] K.-H. Thung, R. Paramesran, and C.-L. Lim, "Content-based image quality metric using similarity measure of moment vectors," *Pattern Recognition*, vol. 45, no. 6, pp. 2193–2204, 2012.

[3] J.-M. Guo and H. Prasetyo, "Content-based image retrieval using features extracted from halftoning-based block truncation coding," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1010–1024, 2015.

[4] L. Kotoulas and I. Andreadis, "Colour histogram content-based image retrieval and hardware implementation," *IEE Proceedings: Circuits, Devices and Systems*, vol. 150, no. 5, pp. 387–393, 2003.

[5] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 6, pp. 1168–1178, 2005.

[6] J. Yu, Y. Rui, and B. Chen, "Exploiting click constraints and multi-view features for image re-ranking," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 159–168, 2014.

- [7] X. Lyu, H. Li, and M. Flierl, "Hierarchically structured multi-view features for mobile visual search," in *Proceedings of the Data Compression Conference (DCC '14)*, pp. 23–32, March 2014.
- [8] J. Yu, D. Liu, D. Tao, and H. S. Seah, "On combining multiple features for cartoon character retrieval and clip synthesis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 5, pp. 1413–1427, 2012.
- [9] M. Braun, K. Hubay, E. Magyari, D. Veres, I. Papp, and M. Bálint, "Using linear discriminant analysis (LDA) of bulk lake sediment geochemical data to reconstruct lateglacial climate changes in the South Carpathian Mountains," *Quaternary International*, vol. 293, no. 8, pp. 114–122, 2013.
- [10] K. Y. Yeung and W. L. Ruzzo, "Principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.
- [11] X. Shi, "Independent component analysis," *IEEE Transactions on Neural Networks*, vol. 15, pp. 735–741, 2001.
- [12] V. C. Klema and A. J. Laub, "The singular value decomposition: its computation and some applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.
- [13] S. Ewert and M. Muller, "Using score-informed constraints for NMF-based source separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)*, pp. 129–132, Kyoto, Japan, March 2012.
- [14] S. Babji and A. K. Tangirala, "Source separation in systems with correlated sources using NMF," *Digital Signal Processing*, vol. 20, no. 2, pp. 417–432, 2010.
- [15] E. M. Grais and H. Erdogan, "Regularized nonnegative matrix factorization using Gaussian mixture priors for supervised single channel source separation," *Computer Speech and Language*, vol. 27, no. 3, pp. 746–762, 2013.
- [16] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, no. 6, pp. 556–562, 2001.
- [17] J.-L. Mao, "Geometric and Semantic similarity preserved non-negative matrix factorization for image retrieval," in *Proceedings of the International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP '12)*, pp. 140–144, Chengdu, China, December 2012.
- [18] M. Babaei, R. Bahmanyar, G. Rigoll, and M. Datcu, "Farness preserving nonnegative matrix factorization," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '14)*, pp. 3023–3027, IEEE, Paris, France, October 2014.
- [19] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.
- [20] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299–1311, 2012.
- [21] Y. Xiao, Z. Zhu, Y. Zhao, Y. Wei, S. Wei, and X. Li, "Topographic NMF for data representation," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1762–1771, 2014.
- [22] Y. D. Khan, F. Ahmad, and S. A. Khan, "Content-based image retrieval using extroverted semantics: a probabilistic approach," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1735–1748, 2014.
- [23] A. Padala, S. Yarramalla, and K. P. Mhm, "An approach for effective image retrievals based on semantic tagging and generalized gaussian mixture model," *International Journal of Information Engineering and Electronic Business*, vol. 7, no. 3, pp. 39–44, 2015.
- [24] S. Zeng, R. Huang, H. Wang, and Z. Kang, "Image retrieval using spatiograms of colors quantized by gaussian mixture models," *Neurocomputing*, vol. 171, no. 1, pp. 673–684, 2015.
- [25] M. L. Piatek and B. Smolka, "Color image retrieval based on spatiochromatic multichannel gaussian mixture modelling," in *Proceedings of the 8th International Symposium on Image and Signal Processing and Analysis (ISPA '13)*, pp. 130–135, Trieste, Italy, September 2013.
- [26] A. Marakakis, N. Galatsanos, A. Likas, and A. Stafylopatis, "Stafylopatis, application of relevance feedback in content based image retrieval using Gaussian mixture models," in *Proceedings of the IEEE International Conference on Tools with Artificial Intelligence*, vol. 1, pp. 141–148, 2008.
- [27] X. He, D. Cai, Y. Shao, H. Bao, and J. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 9, pp. 1406–1418, 2011.
- [28] A. Roy, S. K. Parui, and U. Roy, "SWGMM: a semi-wrapped Gaussian mixture model for clustering of circular-linear data," *Pattern Analysis and Applications*, vol. 19, no. 3, pp. 631–645, 2016.
- [29] S. Zeng, R. Huang, Z. Kang, and N. Sang, "Image segmentation using spectral clustering of Gaussian mixture models," *Neurocomputing*, vol. 144, pp. 346–356, 2014.
- [30] 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, 1999.
- [31] X. Yan, W. Xiong, L. Hu, F. Wang, and K. Zhao, "Missing value imputation based on gaussian mixture model for the internet of things," *Mathematical Problems in Engineering*, vol. 2015, Article ID 548605, 8 pages, 2015.
- [32] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [33] N. D. B. Triggs, "Histograms of oriented gradients for human detection," *CVPR Journals*, vol. 1, no. 12, pp. 886–893, 2005.
- [34] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," in *Proceedings of the 8th European Conference on Computer Vision*, vol. 3021 of *Lecture Notes in Computer Science*, pp. 469–481, Prague, Czech Republic, May 2004.
- [35] J. Han and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 944–952, 2002.
- [36] J.-S. Li and Y.-L. Pan, "A note on the second largest eigenvalue of the Laplacian matrix of a graph," *Linear and Multilinear Algebra*, vol. 48, no. 2, pp. 117–121, 2000.
- [37] Boyd, Vandenberghe, and Foybusovich, "Convex optimization," *IEEE Transactions on Automatic Control*, vol. 51, no. 11, p. 1859, 2006.
- [38] S. G. J. Goldberger and H. Greenspan, "An efficient image similarity measure based on approximations of kl-divergence between two Gaussian mixtures," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '03)*, vol. 1, pp. 487–493, Nice, France, October 2003.
- [39] A. Banerjee and J. Jost, "On the spectrum of the normalized graph Laplacian," *Linear Algebra and Its Applications*, vol. 428, no. 11-12, pp. 3015–3022, 2008.
- [40] G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, California Institute of Technology, 2007.

- [41] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Tech. Rep. TR-2009, Department of Computer Science, University of Toronto, Toronto, Canada, 2009.
- [42] Y. R. Charles and R. Ramraj, "A novel local mesh color texture pattern for image retrieval system," *AEU—International Journal of Electronics and Communications*, vol. 70, no. 3, pp. 225–233, 2016.
- [43] L. Liu, M. Yu, and L. Shao, "Multiview alignment hashing for efficient image search," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 956–966, 2015.
- [44] S. Cui and M. Datcu, "Comparison of kullback-leibler divergence approximation methods between gaussian mixture models for satellite image retrieval," in *Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS '15)*, pp. 3719–3722, Milan, Italy, July 2015.
- [45] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, pp. 252–260, Austin, Tex, USA, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

