

Learning Free Energy Kernel for Image Retrieval

Cungang Wang¹, Bin Wang^{2*}, Liping Zheng¹

¹ School of Computer Science, Liaocheng University
Liaocheng 252000, China

² College of Information, Mechanical and Electrical Engineering, Shanghai Normal University
Shanghai 200234, China

[email: binley.wang@gmail.com]

*Corresponding author: Bin Wang

*Received January 12, 2014; revised April 11, 2014; revised May 29, 2014; accepted June 19, 2014;
published August 29, 2014*

Abstract

Content-based image retrieval has been the most important technique for managing huge amount of images. The fundamental yet highly challenging problem in this field is how to measure the content-level similarity based on the low-level image features. The primary difficulties lie in the great variance within images, e.g. background, illumination, viewpoint and pose. Intuitively, an ideal similarity measure should be able to adapt the data distribution, discover and highlight the content-level information, and be robust to those variances. Motivated by these observations, we in this paper propose a probabilistic similarity learning approach. We first model the distribution of low-level image features and derive the free energy kernel (FEK), i.e., similarity measure, based on the distribution. Then, we propose a learning approach for the derived kernel, under the criterion that the kernel outputs high similarity for those images sharing the same class labels and output low similarity for those without the same label. The advantages of the proposed approach, in comparison with previous approaches, are threefold. (1) With the ability inherited from probabilistic models, the similarity measure can well adapt to data distribution. (2) Benefitting from the content-level hidden variables within the probabilistic models, the similarity measure is able to capture content-level cues. (3) It fully exploits class label in the supervised learning procedure. The proposed approach is extensively evaluated on two well-known databases. It achieves highly competitive performance on most experiments, which validates its advantages.

Keywords: Image feature, Free energy kernel, supervised similarity learning, GMM, image retrieval

This work was supported by Key Laboratory of Intelligent Information Processing and Network Security in Liaocheng University.

<http://dx.doi.org/10.3837/tiis.2014.08.019>

1. Introduction

Motivated by the rapid increasing amount of tremendous digital images, a number of techniques for image store, search and browse have been investigated during past decades [1,2]. To retrieve images that might be of interest from a huge database is highly challenging, which requires the service providers annotate the images beforehand. Traditional approaches for image retrieval usually annotate images manually or according to their surrounding text, and then resort to text-based techniques to perform image retrieval. However, these text-based image retrieval approaches are sensitive to the keywords input by the users [3]. To overcome the difficulties existing in text-based image retrieval techniques, content-based image retrieval (CBIR) is proposed [4]. CBIR systems aim to retrieve relevant images for a query image from a given image dataset based on the content-level similarity, which has attracted increasing attention in recent years and is widely applied to diverse areas, such as advertising, entertainment, fashion design and other industrial applications [5].

CBIR systems are typically comprised of two main components: (a) image representation and feature extraction; (b) image similarity measures defined on the feature space. Usually, for image representation, CBIR systems exploit low-level visual features, e.g., texture, color and shape of images [6,7,8,9,10]. Ideally, image features are expected to encode the semantic content of images. With the extracted image features, another important issue is how to measure the similarity between images [4]. Similarity measures target to merge the cues from low-level image features so as to build the content-level connection among images, which is well known as the problem of “reduce semantic gap”. [11] categorized similarity measures into four typical classes: non-parameter test statistic, e.g. χ^2 -statistics; heuristic distance such as Minkowski form distance L_p ; information theory divergence like Kullback-Leibler divergence. However, these similarity measures do not explicitly consider the broad diversities of different image databases. Consequently, their adaption abilities to the data distribution varying along with image databases are limited.

To build the content-level connections among images and adapt to image distribution, a promising and high-level perspective is to learn the similarity function from the dataset. Motivated by the above observations, a number of learning based approaches are proposed to attack the problem, where *distance metric learning* and *similarity metric learning* [12] are two types of representative approaches. Because distance metric learning can be converted to similarity metric learning, in this paper, we don't distinguish similarity metric learning from distance metric learning. And, for brevity, we refer to it as *similarity learning*. Depending on the availability of class label, these approaches fall into two classes: unsupervised similarity learning and supervised similarity learning. The approach proposed in this paper is a type of supervised similarity learning approach.

Unsupervised similarity learning typically constructs a manifold with low dimensionality, in which the geometric relationship between most observed data are largely preserved [12]. Some unsupervised learning approaches leverage eigen-decomposition to obtain a low dimensional embedding of data points that lie on non-linear manifold. Techniques under this criterion include Multiple Dimension Scaling (MDS), Laplacian eigenmap [13], principle component analysis (PCA) [14], ISOMAP [15] and local linear embedding (LLE) [16]. Multiple dimension scaling (MDS) [17] finds a rank projection that preserves the dissimilarity defined in pairwise distance matrix. PCA attempts to find a subspace, with the data variance preserved. In contrast with MDS and PCA, LLE can find non-linear structure of the data, under the principle that preserves the local rank relation between data in both the intrinsic space and the embedding space. [45] is an unsupervised similarity learning method used for image

retrieval, which is to project data points into a lower-dimensional space so as to exploit the advantage of multiple kd-trees over low-dimensional data.

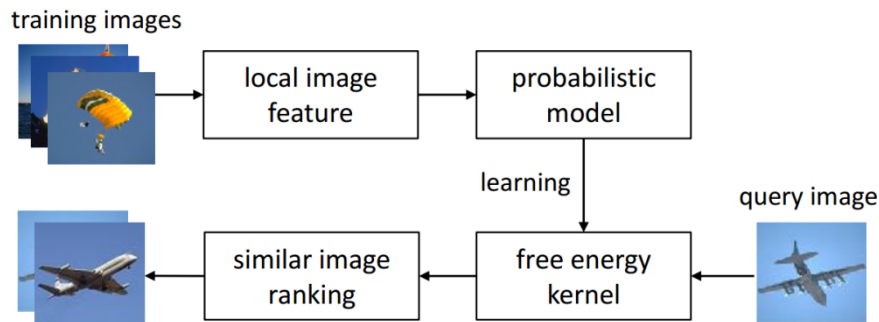


Fig. 1. The framework of the proposed approach

Supervised learning approaches learn similarity functions under the criterion that keeps data points within the same class close while separates data points of different classes far away [17,18]. The representative approaches include local Fisher discriminant analysis (LFDA) [18], relevant component analysis (RCA) [20] and local linear discriminative analysis (LLDA) [21]. Among these approaches, the most representative work is proposed by Xing et al. [22]. It formulates distance learning into a constrained convex programming problem, and learns the similarity metric through minimizing the distance between the data points in the equivalent constraints, subject to the constraint that the data points in the inequivalent constraints are well separated [17]. LFDA [19] extends the classical latent discriminant analysis (LDA) to the case that the form of side information is pairwise constraints. The large margin nearest neighbor (LMNN) [23] extends the neighborhood component analysis (NCA) [24] via a maximum margin framework. [43] is a supervised learning approach used to attack image retrieval problem, which learns a linear combination of a set of base kernels by optimising two objective functions that are commonly used in distance metric learning.

The above approaches exhibit a number of advantages in image retrieval. At the same time, probabilistic approaches [25-30] show promising performance in a wide range of applications, especially well known for their great adaption to data distribution. The so called probabilistic similarity learning methods derive the middle level feature and subsequently the similarity measures based on the probabilistic modeling of data. Therefore, they inherit the adaption abilities from probabilistic models, and are able to exploit hidden information inferred by Bayes inference. These approaches, Fisher kernel [26], probability product kernel [25], free energy score space (FESS) [27] and posterior divergence (PD) [28], can be unsupervised or supervised according to the availability of class label [27]. Nevertheless, we note that, these approaches can be further boosted by exploiting the class label when learning probabilistic models as well as similarity measures.

In this paper, we construct a free energy kernel based on the free energy score space [27], and then propose a similarity learning method for free energy kernel for CBIR. The framework of the proposed approach is graphically illustrated in Fig. 1. First, we model the probabilistic distribution of low-level image features using Gaussian mixture model (GMM). Second, based on GMM, we derive free energy kernel as a function of image features, mixture indicators and model parameters. At last, a supervised learning method is proposed for free energy kernel, so as to exploit class label. The learned free energy kernel measures the similarity between

images. The advantages of the proposed similarity learning approach are threefold: (1) it could fully exploit class label and hidden information while being adaptive to data distribution; (2) the learning method for free energy kernel is very efficient in computation because of the form of the free energy kernel; (3) the proposed learning approach shows highly competitive performance over a set of datasets in image retrieval. The kernel similarity learning approach proposed in this paper could be considered as a type of “metric learning” approach.

The remainder of this paper is organized as follows. Section 2 presents the proposed approach in details. We verify the effectiveness of the kernel learning approach in comparison with the state-of-the-art similarity learning approaches and image retrieval approaches in Section 3. Section 4 draws a conclusion.

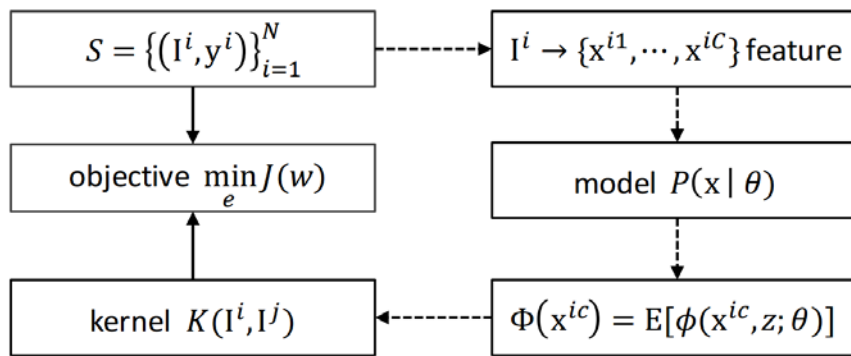


Fig. 2. The mathematical illustration of the proposed approach

2. Learning free energy kernels

This section will present the learning approach for the probabilistic kernel derived from Free Energy Score Space (FESS). We first employ Gaussian Mixture Model (GMM) to model the distribution of image features. The reason for using GMM is that the effectiveness of using GMM for image feature modeling [31] has been extensively verified. Second, we derive the FESS feature mapping based on GMM. Third, construct the free energy kernel based on the FESS feature mapping. Forth, we propose a learning approach for free energy kernel. The mathematical illustration can be found in Fig. 2. For readability, we make a summation of the involved notations in Table 1.

Table 1. The mathematical notation list

Notation	Description	Notation	Description
\mathbf{x}	observed variable	\mathbf{z}	hidden variable
\mathbf{y}	label vector	E	Expectation
D	dimension	d	index of dimension
K	#mixture center	k	index of center
N	#sample	i, j	index of sample
M	#local feature	c	index of local feature
\mathbf{u}	mean	Σ	covariance matrix
a	mixture prior	$K(\cdot, \cdot)$	kernel function
g	mixture posterior	Q	posterior distribution

2.1. Gaussian Mixture Model: A Generative Perspective

First, we introduce Gaussian Mixture Model (GMM) in the generative perspective. It is a probabilistic generative model with hidden variables, composing of multiple mixture centers each of which follows Gaussian distribution. It assumes a generation procedure that, to generate a sample, one first randomly chooses a mixture center and then draws the sample from a Gaussian distribution of this mixture center. It is widely used to model dimension-fixed real-valued data. Let $\mathbf{x} \in R^D$ be the observed variable of D -dimension. Specifically, \mathbf{x} is the local image feature in this work of image retrieval. Let $\mathbf{z} = (z_1, \dots, z_K)^T$ be the binary-valued hidden variable indicating which mixture center is selected to generate the samples. That is, $z_k = 1$ if the k -th mixture center is selected to generate the samples and $z_k = 0$ otherwise. Typically, the probabilistic distribution over \mathbf{z} is chosen to be Multinomial distribution,

$$P(\mathbf{z}) = \prod_{k=1}^K a_k^{z_k} \quad (1)$$

where $a_k = E_{P(\mathbf{z})}[z_k]$, $a_k \in [0, 1]$ and $\sum_k a_k = 1$. Let \mathbf{u}_k be the mean vector and Σ_k be the covariance matrix of the k -th Gaussian distribution for the k -th mixture center. Then the distribution over \mathbf{x} , conditioned on the hidden variable \mathbf{z} can be written as,

$$\begin{aligned} P(\mathbf{x}|\mathbf{z}) &= \prod_{k=1}^K N(\mathbf{x}; \mathbf{u}_k, \Sigma_k)^{z_k} \\ &= \prod_{k=1}^K \left[\frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{u}_k)^T \Sigma_k^{-1} (\mathbf{x} - \mathbf{u}_k) \right\} \right]^{z_k} \end{aligned} \quad (2)$$

Combining $P(\mathbf{z})$ and $P(\mathbf{x}|\mathbf{z})$, then the joint distribution of GMM can be expressed as,

$$\begin{aligned} P(\mathbf{x}, \mathbf{z}|\theta) &= P(\mathbf{x}|\mathbf{z})P(\mathbf{z}) \\ &= \prod_{k=1}^K \left[\frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{u}_k)^T \Sigma_k^{-1} (\mathbf{x} - \mathbf{u}_k) \right\} \right]^{z_k} \prod_{k=1}^K a_k^{z_k} \end{aligned} \quad (3)$$

where $\theta = \{\mathbf{u}_k, \Sigma_k, a_k\}_{k=1}^K$. For computational efficiency, we assume that the covariance matrixes Σ_k are diagonal, i.e., $\Sigma_k = \text{diag}(\sigma_{k1}^2, \dots, \sigma_{kD}^2)$. Note that, in real applications [31], this assumption will not bring negative effect to the performance of GMM.

2.2. Variational Inference and Parameter Estimation

It is worth noting that the log likelihood function $P(\mathbf{x}|\theta) = \sum_{\mathbf{z}} P(\mathbf{x}, \mathbf{z}|\theta)$ is difficult to be maximized. A more sophisticated approach is the Variational Expectation Maximization (VEM) algorithm which alternatively maximizes the log likelihood function over the training set with respect to the posterior distribution of \mathbf{z} (E-step or inference step) and the parameters

(M-step or parameter estimation step). Let $Q^c(\mathbf{z})$ be the posterior approaching to $P(\mathbf{z}|\mathbf{x}^c)$, then we have the following variational lower bound,

$$\begin{aligned} \sum_c \log P(\mathbf{x}^c|\theta) &\geq \sum_c E_{Q^c(\mathbf{z})}[\log P(\mathbf{x}^c, \mathbf{z}|\theta) - \log Q^c(\mathbf{z})] \\ &= \sum_c E_{Q^c(\mathbf{z})} \left[\sum_k z_k \log N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k) + \sum_k z_k \log \frac{a_k}{g_k^c} \right] \\ &= \sum_c \left[\sum_k g_k^c \log N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k) + \sum_k g_k^c \log \frac{a_k}{g_k^c} \right] = L \end{aligned} \quad (4)$$

Assuming that the posterior for the sample \mathbf{x}^c takes the same form with its prior but with different parameter $Q^c(\mathbf{z}) = \prod_{k=1}^K (g_k^c)^{z_k}$ [28], the E-step updates the posterior of the hidden variable, for each observed sample \mathbf{x}^c of the training set $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$,

$$\begin{aligned} \max_{g^c} L(g^c), \text{ s.t. } \sum_k g_k^c = 1 &\Rightarrow \max_{g^c, \lambda} f(g^c, \lambda) = L(g^c) + \lambda(\sum_k g_k^c - 1) \\ \frac{\partial f}{\partial g_k^c} = 0 &\Rightarrow \log N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k) + \log \frac{a_k}{g_k^c} - 1 + \lambda = 0 \Rightarrow g_k^c = \frac{a_k N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k)}{\exp(1 - \lambda)} \\ \frac{\partial f}{\partial \lambda} = 0 &\Rightarrow \sum_k g_k^c - 1 = 0 \Rightarrow \exp(1 - \lambda) = \sum_k \alpha_k N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k) \end{aligned}$$

Then we have,

$$g_k^c = \frac{\alpha_k N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k)}{\sum_k \alpha_k N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k)} \quad (5)$$

where λ is a multiplier. The M-step updates the parameters of GMM,

$$\frac{\partial L}{\partial a_k} = 0, \text{ s.t. } \sum_k a_k = 1 \Rightarrow a_k = \frac{1}{N} \sum_{c=1}^N g_k^c \quad (6)$$

The expression of a_k is actually the average value of posterior probabilities g_k^c across samples. Similarly we have,

$$\mathbf{u}_k = \frac{1}{N} \sum_{c=1}^N g_k^c \mathbf{x}^c, \quad \sigma_{kd}^2 = \frac{1}{N} \sum_{c=1}^N g_k^c (x_d^c - u_{kd})^2. \quad (7)$$

Here, \mathbf{u}_k and σ_{kd}^2 are the weighted mean and variance, where g_k^c weights the contribution of the sample \mathbf{x}^c to the k -th mixture center. The learning algorithm for GMM is the iteration of the E-step and M-step, which is summarized in Algorithm 1.

Algorithm 1. Variational learning of GMM

- 1: **input:** training data set $X = \{\mathbf{x}^c\}_{c=1}^N$
 - 2: initialize parameters θ
 - 3: **repeat**
 - 4: **for** $c=1$ to N **do**
 - 5: compute $\{g_k^c, \mathbf{L}, g_K^c\}$ using Eq. (5)
 - 6: **end for**
 - 7: update θ using Eq. (7)
 - 8: **until** convergence
 - 6: **output:** θ
-

2.3. Free energy feature mapping

We now proceed to derive the free energy score function [27] based on GMM and then the kernel based on the score function. Having the lower bound F^c for $\log P(\mathbf{x}^c | \theta)$ in Eq. (4), we have the following decomposition,

$$\begin{aligned}
 F^c &= \sum_k \underbrace{g_k^c \log N(\mathbf{x}^c; \mathbf{u}_k, \Sigma_k)}_{\varphi_{k,fit}^{c,x}} + \sum_k \underbrace{g_k^c \log a_k}_{\varphi_{k,fit}^{c,z}} - \sum_k \underbrace{g_k^c \log g_k^c}_{\varphi_{k,ent}^{c,z}} \\
 &= \sum_k g_k^c \left[-\log(2\pi)^{\frac{D}{2}} |\Sigma_k|^{-\frac{1}{2}} - \frac{1}{2} \sum_d (x_d^c - u_d^c)^2 \sigma_k^{-2} \right] + \sum_k \underbrace{g_k^c \log a_k}_{\varphi_{k,fit}^{c,z}} - \sum_k \underbrace{g_k^c \log g_k^c}_{\varphi_{k,ent}^{c,z}} \quad (8) \\
 &= \sum \sum_{k,d} \underbrace{\left[-\frac{1}{2} g_k^c x_d^c \sigma_k^{-2} x_d^c + g_k^c x_d^c \sigma_k^{-2} u_d^c \right]}_{\varphi_{k,d,fit}^{c,x}} + \sum_k \underbrace{g_k^c \log a_k}_{\varphi_{k,fit}^{c,z}} - \sum_k \underbrace{g_k^c \log g_k^c}_{\varphi_{k,ent}^{c,z}} + \text{const}
 \end{aligned}$$

The elements of free energy score function are the summation terms of the above variational lower bound, and can be divided into three groups [27],

$$\begin{aligned}
 \varphi_{fit}^{c,x} &= \left(\frac{1}{2} g_1^c x_1^c \sigma_1^{-2} x_1^c, g_1^c x_1^c \sigma_1^{-2} u_1^c, \dots, \frac{1}{2} g_K^c x_D^c \sigma_K^{-2} x_D^c, g_K^c x_D^c \sigma_K^{-2} u_D^c \right)^T \\
 \varphi_{fit}^{c,z} &= (g_1^c \log a_1, \dots, g_K^c \log a_K)^T \\
 \varphi_{ent}^{c,z} &= (g_1^c \log g_1^c, \dots, g_K^c \log g_K^c)^T
 \end{aligned} \quad (9)$$

where the *fit* group measures how well the sample fits the model, and the *ent* group measures the uncertainty in the fitting. We note that the elements of free energy score function are the expectation of the functions over the observed variable \mathbf{x} , hidden variables \mathbf{z} and model parameters θ . The hidden variables enable free energy kernel the ability to exploit hidden information, and model parameters enable it ability to adapt to data distribution. The free energy score function is the combination of the above functions,

$$\Phi(\mathbf{x}^c) = (\varphi_{fit}^{c,x T}, \varphi_{fit}^{c,z T}, \varphi_{ent}^{c,z T})^T \quad (10)$$

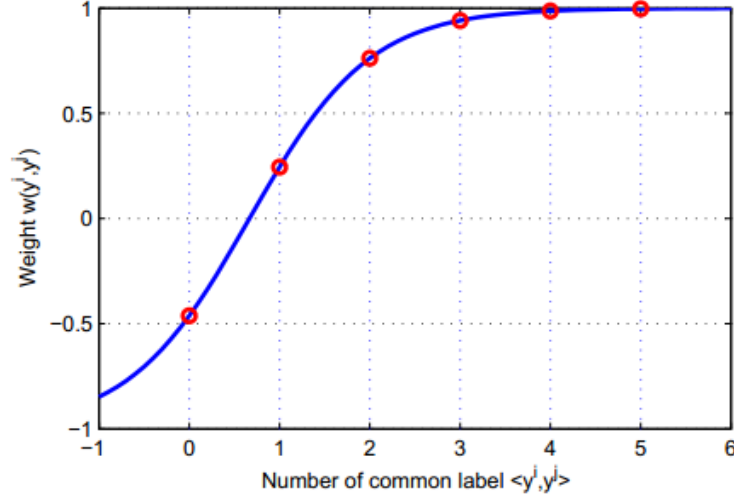


Fig. 3. Illustration of the weight function $w(\mathbf{y}^i, \mathbf{y}^j)$, where $a = 1.5, b = 1, u = 2, v = 1$.

2.4. Learning free energy kernel

Having the score functions or feature mappings for image patches, we now proceed to define the kernel similarity function for images. The above modeling the score function works with image patches. Note that, each image contains a set of image patches, each of which has a corresponding free energy score feature. The distribution of these score features for an image encodes the information of the image and is able to identify the image. We follow an effective and widely used strategy [31] that uses the first order statistics, i.e., the mean of these score features, as the feature of images.

$$\bar{\Phi}(\mathbf{I}^i) = \frac{1}{M} \sum_{c=1}^M \Phi(\mathbf{x}^{ic}) \quad (11)$$

where $\Phi(\mathbf{x}^{ic})$ is the feature mapping for the c -th patch of i -th image. Let $\mathbf{y}^i = (y_1^i, \dots, y_C^i)^T$ be the label vector for the image \mathbf{I}^i , where $y_c^i = 1$ iff the c -th label of all C ones belongs to the image \mathbf{I}^i and $y_c^i = 0$ otherwise. Then the kernel similarity of two images, simultaneously considering image and its corresponding label, can be defined as follows,

$$K(\mathbf{I}^i, \mathbf{I}^j) = w(\mathbf{y}^i, \mathbf{y}^j) K_I(\mathbf{I}^i, \mathbf{I}^j) = w(\mathbf{y}^i, \mathbf{y}^j) \bar{\Phi}(\mathbf{I}^i)^T \bar{\Phi}(\mathbf{I}^j) \quad (12)$$

where $K_I(\mathbf{I}^i, \mathbf{I}^j)$ is the kernel similarity without taking class label into account; $w(\mathbf{y}^i, \mathbf{y}^j)$ is a weight function depending on the similarity of the two label vectors, and is expected to take a positive value if they have shared labels and to take a negative value if they have no shared label. Here we choose the following sigmoid based function:

$$w(\mathbf{y}^i, \mathbf{y}^j) = \frac{u}{1 + \exp(-a \mathbf{y}^i{}^T \mathbf{y}^j + b)} - v \quad (13)$$

where $\mathbf{y}^i \mathbf{y}^j$ is the number of labels shared by image \mathbf{I}^i and \mathbf{I}^j ; a, b, u, v are parameters to be determined. The function is illustrated in Fig. 3. In the following part, we will show how to determine these parameters.

We consider the 1-nearest neighbor criterion [32] which favors higher similarity for images with more shared class label and favors lower similarity for images with less shared labels. The the objective function can be expressed as,

$$J = \sum_i \sum_{j \neq i} w(\mathbf{y}^i, \mathbf{y}^j) K_l(\mathbf{I}^i, \mathbf{I}^j) \quad (14)$$

The above objective function can be maximized using gradient descend algorithms. The gradient of O with respect to a, b, u, v, θ are as follows,

$$\frac{\partial J}{\partial a} = \sum_i \sum_{j \neq i} \frac{u \exp(-a \mathbf{y}^i \mathbf{y}^j + b) (-\mathbf{y}^i \mathbf{y}^j)}{[1 + \exp(-a \mathbf{y}^i \mathbf{y}^j + b)]^2} K_l(\mathbf{I}^i, \mathbf{I}^j) \quad (15)$$

$$\frac{\partial J}{\partial b} = \sum_i \sum_{j \neq i} \frac{u \exp(-a \mathbf{y}^i \mathbf{y}^j + b)}{[1 + \exp(-a \mathbf{y}^i \mathbf{y}^j + b)]^2} K_l(\mathbf{I}^i, \mathbf{I}^j) \quad (16)$$

$$\frac{\partial J}{\partial u} = \sum_i \sum_{j \neq i} \frac{K_l(\mathbf{I}^i, \mathbf{I}^j)}{1 + \exp(-a \mathbf{y}^i \mathbf{y}^j + b)} \quad (17)$$

$$\frac{\partial J}{\partial v} = \sum_i \sum_{j \neq i} -K_l(\mathbf{I}^i, \mathbf{I}^j) \quad (18)$$

$$\frac{\partial J}{\partial \theta} = \sum_i \sum_{j \neq i} w(\mathbf{y}^i, \mathbf{y}^j) \left[\left(\frac{1}{M} \sum_{c=1}^M \frac{\partial \Phi(\mathbf{x}^{ic})}{\partial \theta} \right) \left(\frac{1}{M} \sum_{c=1}^M \Phi(\mathbf{x}^{jc}) \right) + \left(\frac{1}{M} \sum_{c=1}^M \frac{\partial \Phi(\mathbf{x}^{jc})}{\partial \theta} \right) \left(\frac{1}{M} \sum_{c=1}^M \Phi(\mathbf{x}^{ic}) \right) \right] \quad (19)$$

The learning approach is an iteration procedure, which is summarized in Algorithm 2.

Algorithm 2. Learning free energy kernel

- 1: input: training set $S = \{(\mathbf{x}^i, y^i)\}_{i=1}^{N_t}$; threshold ε ; learning rate γ
- 2: initialize parameters $a^{(0)}, b^{(0)}, u^{(0)}, v^{(0)}, \theta^{(0)}$
- 3: $t = 1$
- 4: repeat

$$5: \quad a^{(t)} \leftarrow a^{(t-1)} - \frac{\partial J}{\partial a} \Big|_{(a^{(t-1)}, b^{(t-1)}, u^{(t-1)}, v^{(t-1)}, \theta^{(t-1)})} \quad (\text{Eq. (15)})$$

$$6: \quad b^{(t)} \leftarrow b^{(t-1)} - \frac{\partial J}{\partial b} \Big|_{(a^{(t)}, b^{(t-1)}, u^{(t-1)}, v^{(t-1)}, \theta^{(t-1)})} \quad (\text{Eq. (16)})$$

$$7: \quad u^{(t)} \leftarrow u^{(t-1)} - \frac{\partial J}{\partial u} \Big|_{(a^{(t)}, b^{(t)}, u^{(t-1)}, v^{(t-1)}, \theta^{(t-1)})} \quad (\text{Eq. (17)})$$

$$8: \quad v^{(t)} \leftarrow v^{(t-1)} - \frac{\partial J}{\partial v} \Big|_{(a^{(t)}, b^{(t)}, u^{(t)}, v^{(t-1)}, \theta^{(t-1)})} \quad (\text{Eq. (18)})$$

$$9: \quad \theta^{(t)} \leftarrow \theta^{(t-1)} - \frac{\partial J}{\partial \theta} \Big|_{(a^{(t)}, b^{(t)}, u^{(t)}, v^{(t)}, \theta^{(t-1)})} \quad (\text{Eq. (19)})$$

$$10: \quad t \leftarrow t + 1$$

$$11: \quad \text{until } |J^{(t)} - J^{(t-1)}| \leq \varepsilon$$

$$12: \quad \text{Output: } a, b, u, v, \theta$$

Here we make a summary. In the training step, the label vectors for training samples are available, and thus $w(\mathbf{y}^i, \mathbf{y}^j)$ is available. This model can be trained using the approach described above. In the test step, the label vectors for test samples are no longer available, and thus $w(\mathbf{y}^i, \mathbf{y}^j)$ is unavailable. In this situation, we treat $w(\mathbf{y}^i, \mathbf{y}^j) = 1$ and run the regular retrieval using the parameters learned in the training procedure. It is worth noting that, the reason for introducing $w(\mathbf{y}^i, \mathbf{y}^j)$ is to exploit label information when learning the generative model θ , which is essentially a discriminative learning approach (supervised) for generative model as well as FESS, differing from the native FESS where the label information is absent (unsupervised). Namely, θ is determined by x, y in our proposed approach and is determined by x in FESS.

3. Experiments

This section will experimentally validate the effectiveness of our proposed similarity learning approach, by comparing our approach with state-of-the-art approaches for CBIR over two popular databases across different evaluation criteria.

3.1. Databases

Two popular databases, Wang's [33,34] and Caltech 101 [35,36], are chosen for experimental evaluation.



Fig. 4. Sample images of Wang's database.

Wang's database¹ [33,34] contains 1,000 challenging images selected from Corel database. The database is composed of images with various contents, ranging from natural images to animals. It contains images with the size of 256×384 and 384×256 . The database is divided into 10 groups each of which contains 100 images. The images in the same group are considered to be similar. The group names are respectively African people village, beach, Building, buses, dinosaurs, elephants, flowers, horses, mountains and glaciers, and food. Some sample images from all the 10 categories in the Wang's database are shown in Fig. 4.

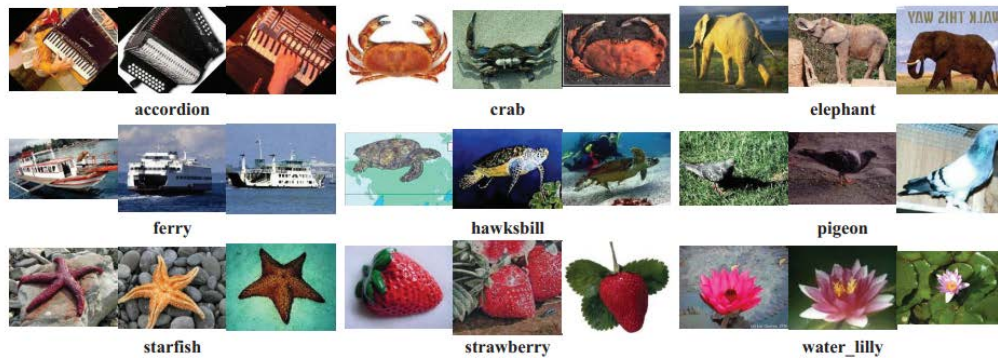


Fig. 5. Sample images of Caltech-101 database.

Caltech-101 database [35] is composed of 9,196 images, which is often used for larger scale experiments. The images in the database are categorized into 101 categories, for example, beaver, ant, crayfish, dolphin and llama, etc. The number of images in the database varies along category from 31 to 800. Most of the images are medium resolution, about 300×300 pixels [37]. The Caltech-101 database is probably the most diverse database available today. Some sample images from certain categories of Caltech-101 database are shown in Fig. 5.

It is worth noting that, in Wang's dataset, each image has multiple labels. For this dataset, images belonging to a certain category not necessarily have the same label vector and $w(\mathbf{y}^i, \mathbf{y}^j)$ not necessarily equals 1. On contrary, in Caltech 101, each image has only one label. For this dataset, images belonging to a certain category have the same vector and $w(\mathbf{y}^i, \mathbf{y}^j) = 1$.

The most important parameter in the proposed approach is the number of mixture centers of GMM. In general, GMM with small number of mixture centers tends to lose information and discrimination ability because it in this case is not capable enough to model the distribution of data. On the other hand, GMM with large number of mixture centers tends to procedure high dimension feature space that leads to poor generalization ability according to generalization theory, and therefore suffers from the so called "curse of dimensionality". Subsequently, it is of great importance to determine an appropriate number of mixture centers. In this paper, we use cross validation over the range of [20,260] with a step of 20 to choose the parameter, and find that, a wide range of about [60,160] could produce satisfied results. Moreover, we also found that two primary factors dominate the number of mixture centers, (1) the number of mixture centers is generally proportional to the number of categories of the dataset in certain range; (2) the number of training samples in the feature spaces.

¹available at <http://wang.ist.psu.edu/docs/related/>

3.2. Image Representation

To cover the diverse visual attributes within images, we use a set of comprehensive features to represent the images. More specifically, we use multiple color SIFT descriptors for representation, due to their state-of-the-art performance in image retrieval and recognition [38]. Following the recommendation by [38], we use OpponentSIFT, C-SIFT, rgSIFT and RGB-SIFT. These color SIFT descriptors are extracted from the patches sampled by dense sampling and Harris-Laplace point sampling, followed by spatial pyramid. For dense sampling, descriptors are extracted around the points of a grid with the step size of 4 pixels. These descriptors are computed from three different scales: 16×16 , 24×24 and 32×32 .

3.3 Evaluation criteria

To comprehensively evaluate the proposed approach, we use the following criteria to measure image retrieval approaches:

Average Precision (AP) [33,39]: the average of the precision values at the ranks where relevant images appear. Specifically, for a query image I^q , the precision (P) and recall (R), as two most commonly used criteria in image retrieval system, can be defined as follows: $P(I^q) = n^q / L$ and $R(I^q) = n^q / N$, where L is the number of retrieved images; n^q is the number of images relevant to the query image in the retrieved images; N is the number of relevant images in the database. Finally, the average precision (AP) and average recall (AR) are computed over all reference images.

Average Retrieval Precision (ARP) [33]: the average precision of the retrieval results of the various images with the number of returned images. It is worth noting that ARP is obtained by means of computing the average precision versus the number of searched images. That is to say, to obtain ARP graph, we calculate the precision for different numbers of retrieved images [34,40].

Average Retrieval Rate (ARR) [33]: the average recall of the retrieval results of the various images with the number of returned images. Similar with ARP graph, to obtain ARR graph, recall values are calculated for varied number of retrieved images [34,40].

3.4 Experimental results

3.4.1. Experiments on Wang's database

The first experiment is performed on Wang's database. This database is thought to meet all the requirements of evaluating the image retrieval systems, because of its diversity in content. The performance criteria in this experiment include average precision, average recall and average retrieval rate. The detailed definition can be found in [33].

Table 2. The average precision (%) of comparison approaches on Wang's database

Category	Euclid.	MCM [41]	LMNN [23]	CTCHIRS[34]	SS[42]	MKL[43]	FESS [27]	Ours
Africans.	40.8	45.3	55.6	68.3	46.2	70.6	73.4	76.7
Beaches	42.9	39.8	44.7	54.0	53.1	51.2	52.2	52.8
Buildings	35.8	37.4	42.5	56.2	49.5	57.3	58.3	60.4
Buses	64.3	74.1	76.6	88.8	90.2	86.1	87.7	89.6
Dinosaurs	60.1	91.5	93.5	99.3	93.7	96.3	95.4	98.1
Elephants	30.7	30.4	32.6	65.8	59.5	61.1	62.4	62.5
Flowers	76.4	85.2	87.3	89.1	90.6	90.8	91.2	92.0

Horses	49.3	56.8	73.6	80.3	85.5	86.4	87.6	90.4
Mountains	31.2	29.3	35.3	52.2	48.5	45.3	46.2	46.8
Food	33.3	37.0	46.7	73.3	72.2	69.2	70.1	71.2
Total	46.5	52.6	58.8	72.7	68.9	71.4	72.5	74.1

Table 3. The average recall (%) of comparison approaches on Wang's database

Category	Euclid.	MCM [41]	LMNN[23]	CTCHIRS [34]	SS[42]	MKL[43]	FESS[27]	Ours
Africans.	20.2	19.5	23.9	42.1	41.3	49.8	55.0	56.2
Beaches	24.6	27.9	33.1	32.4	32.6	33.3	34.8	35.4
Buildings	23.3	20.5	23.7	36.5	35.9	30.4	38.4	36.9
Buses	19.4	25.9	32.5	61.7	62.2	65.7	66.8	75.3
Dinosaurs	23.6	40.3	48.4	96.1	95.4	96.0	97.1	92.2
Elephants	25.9	32.5	38.6	33.1	34.1	38.7	35.8	33.9
Flowers	36.9	48.1	50.2	77.5	68.8	70.0	65.5	67.3
Horses	39.7	40.8	42.1	47.6	48.7	49.4	50.6	59.7
Mountains	22.3	28.5	24.2	27.7	27.9	28.3	29.2	30.8
Food	16.1	24.0	35.3	49.0	57.0	50.9	48.7	51.1
Total	25.2	30.8	35.2	50.3	50.4	51.3	52.2	53.9

In each round of experiment, 20% samples are randomly chosen from the database to form the training set and the rest 80% samples to form the test set. In our experiment, each image is used as a query image for evaluation. We firstly carry out the experiment to compute the precision P of every query image with setting the number L of returned retrieved images as 20, and finally obtain the average precision. The total average recall is obtained in the same manner with the number of returned images set to be 100. In this experiment, the Euclidean distance is still a baseline method. Other comparison approaches include motif cooccurrence matrix (MCM) [41], large margin nearest neighbor classification (LMNN) [23], CTCHIRS [34], semi-supervised distance metric learning (defined as SS) [42], multiple kernel learning via distance metric learning (defined as MKL) [43], and FESS [27]. Among these approaches, MCM and CTCHIRS are two state-of-the-art image retrieval methods, SS and MKL are two distance metric learning-based approaches used for image retrieval, LMNN is a supervised similarity learning approach, FESS is the probabilistic similarity learning methods closely related to our approach. The experimental results are summarized in Table 2 and Table 3. We find that, MCM and LMNN gain significant improvement over the baseline method. Due to the adoption of an optimal feature selection technique, CTCHIRS obtains a better performance for image retrieval than MCM and LMNN. As two distance metric learning approaches, the performance of MKL is better than SS. The reason is that MKL learns a linear combination of a set of base kernels by optimising two objective functions that are commonly used in distance metric learning [43]. From Table 2 and Table 3, we can observe that the proposed approach obtains the best average precision than other comparison methods. Specifically, comparing with SS, which is a representative distance metric learning approach used for image retrieval, the proposed method achieves 5.2% improvement in AP and 3.5% improvement in AR respectively. Moreover, in comparison with FESS which is most closely related to our approach, we achieve 1.6% improvement in AP and 1.7% improvement in AR respectively. The reason for this observation is that our approach fully exploits class label which is very informative for image retrieval task. Fig. 6 illustrates the average precision and the average recall of the retrieval results of various images with the number of retrieved images

respectively. The experimental results clearly present that, for the first 20 to 100 retrieved images of the 1000 ten-category image database, our approach consistently outperforms the other methods. In the average recall experiment (ARR), the precision of image retrieval increases with the number of retrieved images. So, our approach is superior to other models.

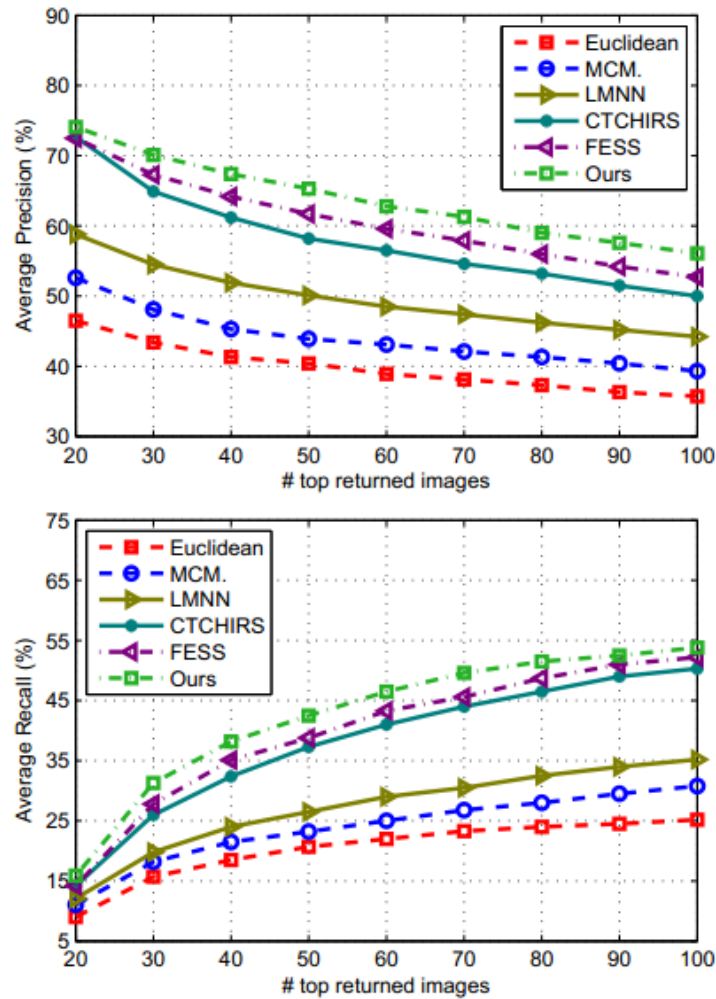


Fig. 6. The average retrieval precision (top) and average retrieval rate (bottom) of these approaches on Wang's dataset.

3.4.2. Experiments on Caltech-101 database

To further validate the abilities of our approach in adapting different databases and in scaling to larger database, we further evaluate the proposed approach on Caltech101 database. For confident conclusion, we repeatedly run the experiment and report the average results. In each round of experiment, 20% samples are randomly chosen from the database to form the training set and the rest 80% samples to form the test set. It is worth noting that, the training set is used to learn GMM as well as free energy kernel.

The experiment is performed using each image of each category as a query image. We carry out the experiment with setting the number of returned images as 20 to calculate the precision P for each query, and finally get the average precision P / N_c (N_c images per category). The experimental results over Caltech101 database are reported in **Table 4**. Different from that on Wang's database, we compare with Xing's method [22], DML-eig [44], large margin nearest neighbor (LMNN) [23], semi-supervised distance metric learning(defined as SS) [42], multiple kernel learning via distance metric learning (defined as MKL) [43], and free energy score space (FESS) [27]. Euclidean distance is still included as a baseline method here. It's worth noting that FESS is the basis of our approach. It is of interest to find that, the relative comparison results are close to that over Wang's database, which indicates that the results are stable across two databases. As shown in **Table 4**, FESS, as an unsupervised similarity learning approaches derived from probabilistic models, shows highly competitive performance against other comparison methods. Our proposed approach again achieves improvement over all the other compared approaches with distinct methodologies. A reason accounting for the results is that, it incorporates different content level information together to form a comprehensive similarity for image retrieval.

Table 4. The average precision of these approaches on Caltech-101dataset.

Algorithm	Average Precision (%)
Euclidean	15.5
Xing's [22]	17.5
DML-eig [44]	17.9
LMNN [23]	18.0
SS [42]	16.7
MKL [43]	18.4
FESS [27]	18.6
Ours	21.2

4. Conclusions

In this paper, we propose a free energy kernel based on the well-known free energy score space (FESS), and then learn the derived kernel in a supervised manner. Specifically, we first model the distribution of image features using GMM. Second, we derive a free energy kernel from GMM, which is a function of image feature, mixture indicator and model parameter. Third, we propose a supervised learning approach for the free energy kernel to exploit label information. The experimental results on two databases demonstrate that the proposed approach is superior to other comparison approaches for the content-based image retrieval task.

References

- [1] D. Cerra and M. Datcu, "A fast compression-based similarity measure with applications to content-based image retrieval," *Journal of Visual Communication and Image Representation*, 23(2), pp.293–302, 2012. [Article \(CrossRef Link\)](#).
- [2] B. Wang, Y. Shen, and Y. Liu, "Integrating distance metric learning into label propagation model for multi-label image annotation," in *Proc. of IEEE International Conference on Image Processing*, IEEE, pp.3649–3652, 2011. [Article \(CrossRef Link\)](#).

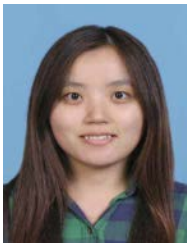
- [3] M. E. ElAlami, "A novel image retrieval model based on the most relevant features," *Knowledge-Based Systems*, 24(1), pp.23–32, 2011. [Article \(CrossRef Link\)](#).
- [4] D. Ziou, T. Hamri, and S. Boutemedjet, "A hybrid probabilistic framework for content-based image retrieval with feature weighting," *Pattern Recognition*, 42(7), pp.1511–1519, 2009. [Article \(CrossRef Link\)](#).
- [5] M. Arevalillo-Herr´aez, F. Ferri, and J. Domingo, "A naive relevance feedback model for content-based image retrieval using multiple similarity measures," *Pattern Recognition*, 43(3), pp.619–629, 2010. [Article \(CrossRef Link\)](#).
- [6] W. Bian and D. Tao, "Biased discriminant Euclidean embedding for content-based image retrieval," *IEEE Transactions on Image Processing*, 19(2), pp.545–554, 2010. [Article \(CrossRef Link\)](#).
- [7] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, 29(8), pp.1233–1244, 1996. [Article \(CrossRef Link\)](#).
- [8] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, 42(3), pp.145–175, 2001. [Article \(CrossRef Link\)](#).
- [9] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, 7(1), pp.11–32, 1991. [Article \(CrossRef Link\)](#).
- [10] K. Kim, M. Hasan, J. Heo, Y. Tai, and S. Yoon, "Probabilistic cost model for nearest neighbor search in image retrieval," *Computer Vision and Image Understanding*, 2012. [Article \(CrossRef Link\)](#).
- [11] J. Puzicha, J. M. Buhmann, Y. Rubner, and C. Tomasi, "Empirical evaluation of dissimilarity measures for color and texture," in *Proc. of Proceedings of IEEE International Conference on Computer Vision*, IEEE, pp.1165–1172, 1999. [Article \(CrossRef Link\)](#).
- [12] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. Hoi, and M. Satya-narayanan, "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1), pp.30–44, 2010. [Article \(CrossRef Link\)](#).
- [13] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, 15(6), pp.1373–1396, 2003. [Article \(CrossRef Link\)](#).
- [14] A. Webb, *Statistical pattern recognition*, Wiley, 2003. [Article \(CrossRef Link\)](#).
- [15] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for non-linear dimensionality reduction," *Science*, 290(5500), pp.2319–2323, 2000. [Article \(CrossRef Link\)](#).
- [16] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *The Journal of Machine Learning Research*, 4, pp.119–155, 2003. [Article \(CrossRef Link\)](#).
- [17] L. Yang and R. Jin, "Distance metric learning: A comprehensive survey," *Michigan State University*, pp.1–51, 2006. [Article \(CrossRef Link\)](#).
- [18] B. Wang and Y. Liu, "Collaborative similarity metric learning for semantic image annotation and retrieval," *KSII Transactions on Internet & Information Systems*, 7(5), 2013. [Article \(CrossRef Link\)](#).
- [19] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proc. of International Conference on Machine learning*, pp.905–912, ACM, 2006. [Article \(CrossRef Link\)](#).
- [20] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning distance functions using equivalence relations," in *Proc. of International Conference on Machine Learning*, 20(1), 11, 2003. [Article \(CrossRef Link\)](#).
- [21] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6), pp.607–616, 1996. [Article \(CrossRef Link\)](#).
- [22] E. Xing, A. Ng, M. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," *Advances in Neural Information Processing Systems*, 15, pp.505–512, 2002. [Article \(CrossRef Link\)](#).

- [23] J. Blitzer, K. Q. Weinberger, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Advances in Neural Information Processing Systems*, pp.1473–1480, 2005. [Article \(CrossRef Link\)](#).
- [24] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," *Advances in Neural Information Processing Systems*, 2004. [Article \(CrossRef Link\)](#).
- [25] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *The Journal of Machine Learning Research*, 5, pp.819–844, 2004. [Article \(CrossRef Link\)](#).
- [26] T. Jaakkola, D. Haussler, et al., "Exploiting generative models in discriminative classifiers," *Advances in Neural Information Processing Systems*, pp.487–493, MIT, 1999. [Article \(CrossRef Link\)](#).
- [27] A. Perina, M. Cristani, U. Castellani, V. Murino, and N. Jojic, "Free energy score spaces: Using generative information in discriminative classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7), pp.1249–1262, 2012. [Article \(CrossRef Link\)](#).
- [28] X. Li, T. Lee, and Y. Liu, "Hybrid generative-discriminative classification using posterior divergence," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.2713–2720, 2011. [Article \(CrossRef Link\)](#).
- [29] X. Li, Y. Liu, and T. Lee, "Stochastic feature mapping for PAC-Bayes classification," *arXiv:1204.2609*, 2012. [Article \(CrossRef Link\)](#).
- [30] X. Li, X. Zhao, Y. Fu, and Y. Liu, "Bimodal gender recognition from face and fingerprint," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.2590–2597, 2010. [Article \(CrossRef Link\)](#).
- [31] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an e-evaluation of recent feature encoding methods," in *Proc. of British Machine Vision Conference*, 2011. [Article \(CrossRef Link\)](#).
- [32] Laurens van der Maaten, "Learning discriminative fisher kernels," in *Proc. of International Conference on Machine Learning*, pp.217–224, 2011. [Article \(CrossRef Link\)](#).
- [33] M. Subrahmanyam, R. Maheshwari, and R. Balasubramanian, "Expert system design using wavelet and color vocabulary trees for image retrieval," *Expert Systems with Applications*, 39(5), pp.5104–5114, 2012. [Article \(CrossRef Link\)](#).
- [34] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image and Vision Computing*, 27(6), pp.658–665, 2009. [Article \(CrossRef Link\)](#).
- [35] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, 106(1), pp.59–70, 2007. [Article \(CrossRef Link\)](#).
- [36] A. D. Holub, M. Welling, and P. Perona, "Combining generative models and fisher kernels for object recognition," in *Proc. of IEEE International Conference on Computer Vision*, 1, pp.136–143, IEEE, 2005. [Article \(CrossRef Link\)](#).
- [37] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2, pp.2169–2178, 2006. [Article \(CrossRef Link\)](#).
- [38] K. Van De Sande, T. Gevers, C. Snoek, Evaluating color descriptors for object and scene recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), pp.1582–1596, 2010. [Article \(CrossRef Link\)](#).
- [39] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, and T. Han, "Contextual weighting for vocabulary tree based image retrieval," in *Proc. of IEEE International Conference on Computer Vision*, pp.209–216, 2011. [Article \(CrossRef Link\)](#).
- [40] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems," *Knowledge-Based Systems*, 39, pp.85–94, 2013. [Article \(CrossRef Link\)](#).
- [41] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, "Content based image retrieval using motif co-occurrence matrix," *Image and Vision Computing*, 22(14), pp.1211–1220, 2004. [Article \(CrossRef Link\)](#).

- [42] S Zhang, M Yang, T Cour, K Yu, DN Metaxas, “Semi-supervised distance metric learning for collaborative image retrieval and clustering,” *ACM Transactions on Multimedia Computing, Communications and Applications*, 2010. [Article \(CrossRef Link\)](#).
- [43] X. He, “Multiple kernel learning via distance metric learning for interactive image retrieval,” *Multiple Classifier Systems*, 6713, pp.147-156, 2011. [Article \(CrossRef Link\)](#).
- [44] Y. Ying and P. Li, “Distance metric learning with eigenvalue optimization,” *The Journal of Machine Learning Research*, 13, pp.1–26, 2012. [Article \(CrossRef Link\)](#).
- [45] P Wu, SCH Hoi, DD Nguyen, Y He, “Randomly Projected KD-Trees with Distance Metric Learning for Image Retrieval,” *Advances in Multimedia Modeling*, pp.371–382, 2011. [Article \(CrossRef Link\)](#).



Cungang Wang received his B.E. degree in educational technology from Liaocheng University, Liaocheng, China, in 2000, and his M.S. degree in computer science and technology from Ocean University of China, Qingdao, China, in 2006. He worked as a lecturer in the School of Computer Science at Liaocheng University since 2008. His research interests include machine learning, image processing, pattern recognition.



Bin Wang received her BE degree in information and computational science from Shandong Normal University, Ji’nan, China, in 2006; and the MS degree in applied mathematics from University of Science and Technology Beijing, Beijing, China. She received her PhD degree in 2014 from Department of Automation, Shanghai Jiao Tong University, Shanghai, China. She is currently an assistance professor in the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University. Her research interests include computer vision, machine learning, image processing, multimedia analysis.



Liping Zheng received her Ph.D. degree from Tongji University, Shanghai, China. She is now an associate professor in School of Computer Science, Liaocheng University, Shandong, China. Her Research directions include image processing and artificial intelligence.