

A novel relevance feedback method for CBIR

Yunbo Rao¹ · Wei Liu¹  · Bojiang Fan¹ · Jiali Song¹ · Yang Yang²

Received: 2 July 2017 / Revised: 12 November 2017 / Accepted: 13 December 2017 /
Published online: 5 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract In this paper, we address the challenge about insufficiency of training set and limited feedback information in each relevance feedback (RF) round during the process of content based image retrieval (CBIR). We propose a novel active learning scheme to utilize the labeled and unlabeled images to build the initial Support Vector Machine (SVM) classifier for image retrieving. In our framework, two main components, a pseudo-label strategy and an improved active learning selection method, are included. Moreover, a feature subspace partition algorithm is proposed to model the retrieval target from users by the analysis from relevance labeled images. Experimental results demonstrate the superiority of the proposed method on a range of databases with respect to the retrieval accuracy.

Keywords Relevance feedback · Content based image retrieval · Pseudo-labeling strategy · Active learning · SVM

1 Introduction

Content-based image retrieval (CBIR) is a hot research field in the background of rapid grows in image database [26, 27, 39, 40]. Different from traditional technologies of retrieving, CBIR allow users to organize and retrieve an image in the database by analyzing its visual features [8, 51]. Current CBIR systems are usually based on building a visual model

This article belongs to the Topical Collection: *Special Issue on Deep Mining Big Social Data*
Guest Editors: Xiaofeng Zhu, Gerard Sanroma, Jilian Zhang, and Brent C. Munsell

✉ Wei Liu
l.weiu@sina.com

¹ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, People's Republic of China

² School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, People's Republic of China

to capture the variability of a set of images. In the earlier days, image retrieval systems were based on the comparison of global feature descriptors, such as color or texture histogram [24, 36]. This method is unable to get a satisfied performance on condition that image set have a relative complex content, such as human face and zebras. Recent approaches learn visual model by means of supervised learning method, e.g., a two-layer image description based on “generic” descriptors and its neighborhood distribution [31], saliency-aware foreground annotation method by analyzing image-level label [2, 5, 9, 21, 22, 29, 50], clustering analysis technologies [20, 46, 47], and hashing approaches [10, 49].

Although those methods have achieve some impressive result, a problem is still far from solve due to the ‘semantic gap’ existing between the low-level image visual features description and high-level concepts. RF is an effective but commonly used method to bridge ‘semantic gap’ and to scale up the retrieval performance in CBIR system. RF, originally proposed for text retrieval system, refers to a set of query optimization approaches learning from the results of analysis of users’ browsing behaviors in retrieving processes. The main idea of RF is letting the user participate in CBIR system and ameliorating image retrieval result. In general, RF is a supervised learning technique which focuses on the interaction between users and search engine by utilizing manual label images with positive or negative tag. User can get better retrieval results by constantly modifying the query expression, which is also called query reformulation [16]. In the process of RF, CBIR systems are able to better grasp the query requirement from user by constructing a better query condition or modify the parameters in the search strategy. The conventional processes of RF contain two steps. First, the users label a number of retrieved images with positive feedbacks or negative feedbacks. Then, CBIR system refine its query strategy by analysis those feedback samples. All those above steps are carried out iteratively until satisfactory performance is obtained.

There are several methods to implement an RF mechanism. Rocchio algorithm is a famous RF method used in information retrieval [28]. Like most retrieval technologies, the Rocchio RF approach was developed based on the vector space model and the assumption that most users have a general conception of which documents should be denoted as relevant or irrelevant. After a series of recursive process, a new query is generated from previous rounds, which can be considered as a method about query reformulation. In addition, deep learning driven method start to show its power in RF by the backpropagation algorithm in networks of artificial neurons [17, 30, 37, 44]. It is also a query reformulation.

In recent years, some methods often treat RF as classification problem [33, 36, 38] in which supervised classification based methods are used in CBIR by training the classifier with already annotated images [6, 19, 34]. In the approaches of classification, RF is considered as a binary classification problem with relevant and irrelevant [16, 28, 31, 36]. Some machine learning technologies are employed to train a classifier, which then is used to classify the target images set into classes that is relevant to the query or not [28, 36]. Some approaches for RF use SVM to learn a decision boundary to separating relevant and irrelevant images collected from the annotated information [7, 12, 23, 25, 38]. Compared with other machine learning algorithms [11, 13, 45], SVM have its advantages over generalization ability, without restrictive assumptions regarding the data, fast speed of processing and flexible kernels that can be easily tuned. However, existing techniques tend to perform poorly due to the fact that the feedback information in each cycle of retrieving is limited. Moreover, labeling images as relevant or irrelevant is time-consuming and costly, especially when huge archives of images are considered. Thus, it’s unrealistic to expect users to do a lot of works of interactions in practice. There have

been several attempts to overcome this. One of the representative ways is use several weak SVM classifier to create a stable and accurate strong classifier, such as an Asymmetric Bagging and Random Subspace SVM algorithm proposed by S. Jones et al. [35] which is designed for image retrieving system with just a very few positive samples and a random sampling SVM model proposed in [48] for query expansion. Some approaches want to use a fuzzy or changeable kernel function to build a SVM-based RF system. K. Wu et al. [43] presented a fuzzy support vector machine to make the CBIR system is more robust to some usual problems of the small size of training samples, biased hyperplane and over-fitting. Wang et al. [41] proposed to use probabilistic features and weighted kernel function to build a SVM-based RF method in which a relief feature weighting algorithm is applied to compute feature's weight values and the SVM kernel function is modified dynamically according to the feedback from the feature weights. The main drawback of those works is different SVM kernels would have different output and the difficulty in selection of a proper membership function when users have different retrieval actions.

In addition, Active Learning (AL) is another effective approach used in RF, which can reduce the effort of annotation by carefully selecting samples shown to users for annotation to achieve maximal information feedback [3, 4, 12, 36]. In general, the unlabeled images with highly uncertainty and diversity to each other are usually considered as the most informative samples for image classification. The uncertainty of a sample is related to the condence of its correctly classifying by supervised learning algorithm and the diversity among samples is associated to their correlation in the feature space. However, some SVM-based AL, such as those works proposed in [4, 36], simply select uncertain images close to current separating hyperplane by means of Margin Sampling (MS) [32]. An obvious shortcoming of MS is that it does not evaluate the diversity attribute with representative images. To solve this issue, B. Demir et al. [3] proposed a triple criteria AL method for the selection of most informative and representative unlabeled images to be annotated. The main idea of B. Demir et al.'s method is put the selection of representative images into three step, firstly, uncertain images are selected by MS strategy, then getting the diverse image from the highest density region, and the final step is achieved by clustering-based strategy in element-element clustering to evaluate the density and the diversity of unlabeled images in image feature space to drive the selection of images to be annotated. Liu et al. [18] presented another representative image selection scheme in CBIR, in which information measure of unlabeled data is applied by a dynamic clustering process and one image is selected from each cluster for labeling.

Although SVM-based active learning have large improvement to the traditional SVM-based RF approaches in retrieval performance. But there are some drawbacks that cannot be ignored. Firstly, the number of labeled images is a major influence for the performance of SVM-based active learning system. It's hard to build an effective SVM classifier using the limited feedback information in the classifier training. Moreover, how to efficiently select the most informative images for users annotation.

To solve those problems, we propose a novel SVM-based active learning scheme using the labeled and unlabeled images to build a SVM classifier, moreover an improved active selection criterion is proposed to increase the feedback information in each retrieval round. The main contribution of this paper include: Firstly, we propose a feature subspace partition algorithm to partition image feature into several subspaces by analysis the characteristic of the images in the first feedback and give a better representation of users' retrieval behavior. Secondly, a pseudo-label strategy is proposed to build the SVM classifier by using labeled

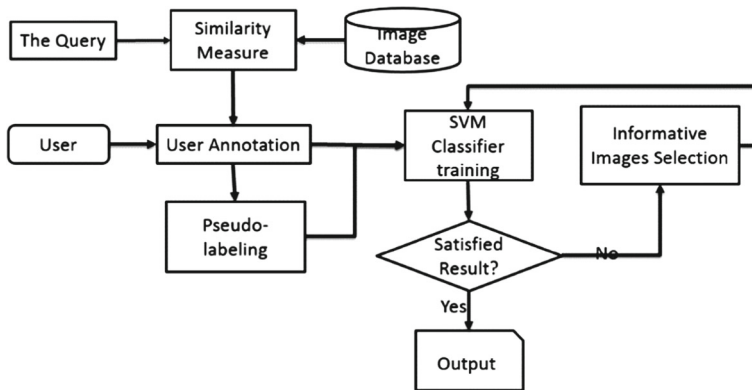


Figure 1 Basic Scheme of relevant feedback in CBIR system

and unlabeled images. At last, we propose an active representative image selection method. All of our work will give a detail description in the following section.

The rest of the paper is organized as follows: Section 2 gives a detail description about our proposed work. In Section 3, image feature extraction and experiment results are given. Finally, conclusions are presented in Section 4.

2 Proposed method

2.1 Overview of the proposed framework

This section presents our approach for CBIR and its RF. As shown in Figure 1, the overview of the proposed framework is outlined as follows:

- (i) Given a query, the system performs the search in the database using the Euclidean distance for similarity matching. The top l similar images are shown to users for labeling.
- (ii) The users label the l images as either relevance or non-relevance. Based on the relevant images, system performs feature subspace partition algorithm to partition the image feature into several subspaces and gives each subspace different weight.
- (iii) The system performs a clustering analysis strategy in the image set. The label (relevant or irrelevant) of labeled images is then propagated to the some nearest unlabeled neighbors in a selective strategy.
- (iv) The user-labeled images combined with the images with a pseudo-label are used for SVM classifier training.
- (v) The most informative images is sent to users for labeling.
- (vi) Repeat steps from (iv) to (v) until the retrieval performance is satisfactory.

2.2 Image feature subspace partition

Images are usually qualified into a high dimensional feature vector in CBIR. However, users may focus on different feature subspaces which are not equal to each other. For example, one will focus on blue color when searching for a “sky image” but paying more attention to shape feature when searching a “banana image”. To better model user’s retrieval behavior,

we propose a Feature Subspace Partitioning (FSP) algorithm to partition the image feature into several subspaces, and assign each subspaces different weights.

Our assumption for subspace partitioning is that the positive labeled images can represent user's retrieval behavior and image feature can be partitioned into several subspaces in which features have some statistical consistency. In general, the statistical characteristic of the labeled images will highly affect the performance of FSP. When there are little consistency among the training set, the learning algorithm would encounter a great difficulty in training a reasonable subspace partition. In order to make the training more smoothly, we get training set from user's feedback. Given a query, the system firstly performs feature similarity matching in the target database and return several images in the first retrieval round. We select positive labeled image as training set after user's feedback.

Algorithm 1 Feature subspace partition (FSP)

Require: $L = \{I_1, I_2, \dots, I_l\}$

Ensure: The subspace partition for the d -dimensional feature vector

- 1: Construct a correlation matrix between any two features to get a $d \times d$ matrix. Initially, treat each feature as a single group.
 - 2: Compute the correlation coefficient r between any two groups.
 - 3: If the coefficient r between any two features satisfies $r \geq 0.6$, merge those two features in to one subspace.
 - 4: Repeat steps from 2 to 3 until any two group are independent.
-

To model this, we treat every image feature (e.g. one dimension of color histogram) as a discrete random variable. Let $L = \{I_1, I_2, \dots, I_l\}$ denotes the set of images labeled with relevance, where $I_i \in R^d$ represents an image with d dimensional feature vector. We use Pearson correlation coefficient to test the statistical dependency between any two images. For two feature x_i and x_j , the correlation coefficient can be represented as follow:

$$r = \frac{n \sum x_i x_j - \sum x_i \sum x_j}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum x_j^2 - (\sum x_j)^2}}, \quad (1)$$

here, $r \in [-1, 1]$. In the viewpoint of statistics, $|r| \leq 0.6$ (which was used in literature [42]) represents a low correlation between x_i and x_j . Thus, they are regard as independence and put into different subspace when $|r| \leq 0.6$ and put together otherwise.

As mentioned above, different subspace may not equally important in image retrieval. To solve this problem, we introduce weights to address this point. Let $\{S^1, S^2, \dots, S^k\}$ represents K subspaces. P_i is a set to denotes the feature projection from all relevant labeled images (labeled by users) to the i th subspace $\{S^k\}$. If P_i is empty, set its weight as $\frac{1}{K}$; else calculate the subspace weights of P_i by

$$\omega_i = \begin{cases} \frac{1}{K} & \sigma_i = 0 \\ \frac{\sigma_i}{\sum_{i=1}^k \frac{1}{\sigma_i}} & \text{Otherwise} \end{cases}, \quad (2)$$

where σ_i denotes variance in P_i .

2.3 Exploiting unlabeled images

In the relevant feedback of CBIR system, the number of feedback is not user-friendly, especially when users cannot label too many images.

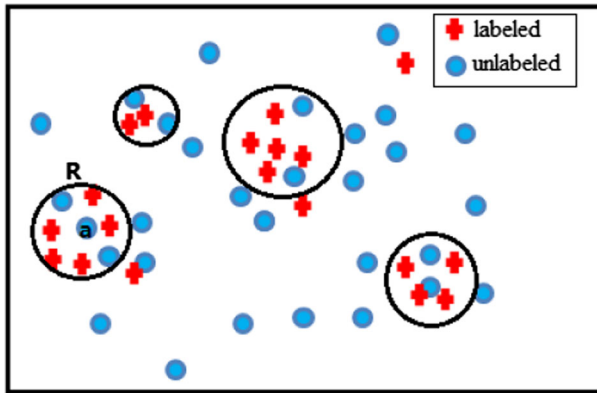


Figure 2 Pseudo-labeling estimation

In our work, we propose a pseudo-labeling method to exploit the characteristic of the labeled image to select the unlabeled images for pseudo-labeling. The basic idea is that images with the same label usually exhibit some similarity in image feature. By using this characteristic, we can enlarge the training dataset. The labeled samples are clustered according to their types: relevant or irrelevant respectively. We firstly group the labeled images into a set of clusters and then identify some unlabeled images which are close to the labeled images to perform pseudo-labeling.

To utilize the property of image subspace feature, we choose cluster images in each subspace separately (suppose we have gotten image subspace feature partitioning) and set the number of clusters in each subspace clustering analysis is proportional to its weight. For example, we have gotten two feature subspace with weight 0.7 and 0.3 separately beforehand. First, we partition the training set into seven clusters in the first subspace and three clusters in the second subspace. Second, images which are nearest to the centroid of each cluster will be used for pseudo-labeling. In order to exploit the local characteristics of image similarity, we introduce Iterative Self-Organizing Data Analysis Technique (ISODATA) method proposed in [1] for clustering analysis. ISODATA is an efficient clustering algorithm which permits diverse amount of clusters to be specified rather than requiring that is known in a priori that overcome the disadvantages of an inappropriate choice of the number of k cluster in k -means and has been widely used in data-processing system.

Note that if we cluster image in each subspace independently, a same image may have different estimate value and assign different label. To avoid this problem, we do clustering from subspace with largest weight to the subspace with smallest weight. Supposed subspaces $\{S^1, S^2, \dots, S^K\}$ are arranged by descent weight. If an unlabeled image have assigned with certain pseudo label in the feature subspace $\{S^n\}$, and then it will keep the label no matter what estimate value obtained in the subspace $\{S^{n+1}\}$.

According to the theory of K-NN algorithm, the close data points tend to have similar class label. We select the nearest unlabeled points for each labeled cluster based on element-element distance. The label (relevant or irrelevant) of each labeled cluster is then propagated to the unlabeled neighbors. As shown in Figure 2, the unlabeled element a surrounded by several unlabeled point in the region R can be used for pseudo-labeling. This is referred to the process of pseudo-labeling in [43].

In our framework, some issues in pseudo-labeling have to be addressed: (1) selection proper images for pseudo-labeling, (2) determination of the pseudo-labels, (3) estimation of

the degree of correct labeling for the pseudo-labeled image. To solve those problem, we use fuzzy logic for correct labeling estimation. The detail is to develop a fuzzy member function $P(I) (P(I) \in [0, 1])$ as the estimation respect to class labels.

Suppose clustering has been performed on positive and negative class separately. In general, the closer an unlabeled image is to the center of a labeled images cluster center, the higher of its degree of the same label. In each cluster, we let \bar{d} represents the average distance from each point to its cluster center. \bar{d}_i in the i th cluster with M points can be represented as

$$\bar{d}_i = \frac{\sum_{k=1}^M d(V_k, V_{si})}{M}, \tag{3}$$

here, $d(V_k, V_{si})$ represents the Euclidean distance from the point V_k to its center V_{si} . Based on the above argument, we construct a fuzzy function as

$$\omega_2(I) = \exp \left\{ -\frac{d(I, V_{si})}{\bar{d}_i} \right\}, i = \underset{i}{\operatorname{argmin}}(I, V_{si}) \tag{4}$$

where I denotes the pseudo-labeled image and $\operatorname{mind}(I, V_{si})$ represents the minimal element-element distance from the pseudo-labeled image to the same labeled cluster center V_{si} .

In addition, another issue must be noticed. If we cluster image in relevance and irrelevance independently, a same unlabeled image may have different estimate value and assign different label according to the above argument. To avoid this problem, a pseudo-labeled image should be far be as far away from its nearest opposite labeled cluster. Therefore, we develop another fuzzy function as

$$\omega_1(I) = \begin{cases} \exp \left\{ -\frac{\operatorname{mind}(I, V_{si})}{\operatorname{mind}(I, V_{oi})} \right\} & \text{if } \frac{\operatorname{mind}(I, V_{si})}{\operatorname{mind}(I, V_{oi})} < 1 \\ 0 & \text{Otherwise} \end{cases} \tag{5}$$

here, $\operatorname{mind}(I, V_{oi})$ represents the minimal distance from the pseudo-labeled image to the opposite labeled cluster center. We combine the above two probability estimation function to obtain the final pseudo-Label estimation function as:

$$P(I) = \omega_1(I) \cdot \omega_2(I). \tag{6}$$

In our work, only those pseudo-labeled images whose probabilistic estimation $P(I)$ satisfies $P(I) > \lambda$ will be used for classifier training in our classifier. Parameter λ is a threshold for pseudo-labeled images selection, and its influence will be discussed in the following section.

2.4 Proposed SVM-based active learning

Active learning studies the strategy for the learner to select the most informative objects to user for feedback and gain the maximal information for decision making. In the process of SVM-based active learning, a better selection of the representative images would make the labeling work of user bring more information and reduce the workload of labeling task. In SVM classification, images which are closest to the separating hyperplane are considered to have low condence to be correctly classied. Thus, the early SVM-based AL methods usually select image as

$$I_i = \underset{i \in U}{\operatorname{argmin}} |f(I_i)|, \tag{7}$$

here, U is the set of unlabeled image and $f(\cdot)$ is decision function in SVM, This active learning strategy has been used in [15, 36]. However, this strategy does not remove the redundancies between these selected images. That is to say, these images may be too similar

to each other in some cases, thus the feature space supported by these images is generally very small. Thus, the learners trained on these instances are usually biased and unstable.

Inspired by works [36, 38], a dynamic clustering method is adopted to realize representative images selection in our work. Let S is the set which include s images closest to the separating hyperplane. The k -means clustering initially applied to S and those images are divided into h clusters, $\{C_1, C_2, \dots, C_h\}$. We assume that features of similar images tend to get together in the feature space. Images will have more diversity in feature space in different clusters.

We define the degree of uncertainty of an element x_i as its density in its own cluster. $density(x_i)$ is defined as follow:

$$density(x_i) = \frac{\sum_{j=1}^k D(x_i, x_j)}{k}. \quad (8)$$

Here, k is a constant which represent the k nearest elements around x_i should be considered, $D(\cdot)$ denotes the Euclidean distance between images i and image j in the higher image space. This distance can be estimated by SVM kernel function $K(\cdot)$ as

$$D = \|\varphi(x_1) - \varphi(x_2)\| = \sqrt{K(x_1, x_1) + K(x_2, x_2) - 2K(x_1, x_2)}, \quad (9)$$

where, $\varphi(x)$ is the mapping function in the SVM classification.

After calculating the density of each image in $\{C_i\}$, the image I_i with minimal density is selected as the most representative image, i.e.,

$$I_i = \arg \min_{I_i \in C_i} \{density(I_i)\}. \quad (10)$$

In our work, we use a kernel function based on our proposed feature subspace partition algorithm for SVM training. Its definition is shown as

$$K(I_i, I_j) = \exp \left\{ -\frac{d(I_i, I_j)^2}{\sigma^2} \right\}, \quad (11)$$

where $d(\cdot)$ is the element-element distance and σ is the parameter of RBF kernel which we set as large as in SVM training set.

3 Experiments

We conduct our experiments on two related image processing tasks, including analysis of the influence of parameter and retrieval performance comparison.

3.1 Dataset

The performance of the proposed method is evaluated through Wang' image dataset [29], INRIA Holiday dataset [30] and CIFAR-10 Dataset [14].

Wang database consists of 1,000 of images of various contents ranging from humans to animals to scenery images (shown in Figure 3a), namely, Africans, beaches, buildings, buses, dinosaurs, elephants, owers, horses, mountains, and food. These images have been pre-classied into different categories by domain professionals. Each category has 100 images with resolution of either 256×384 or 384×256 .

The INRIA Holidays dataset is one of a standard bench-marking dataset used in the CBIR to measure the performance of image process technology. This database (shown in

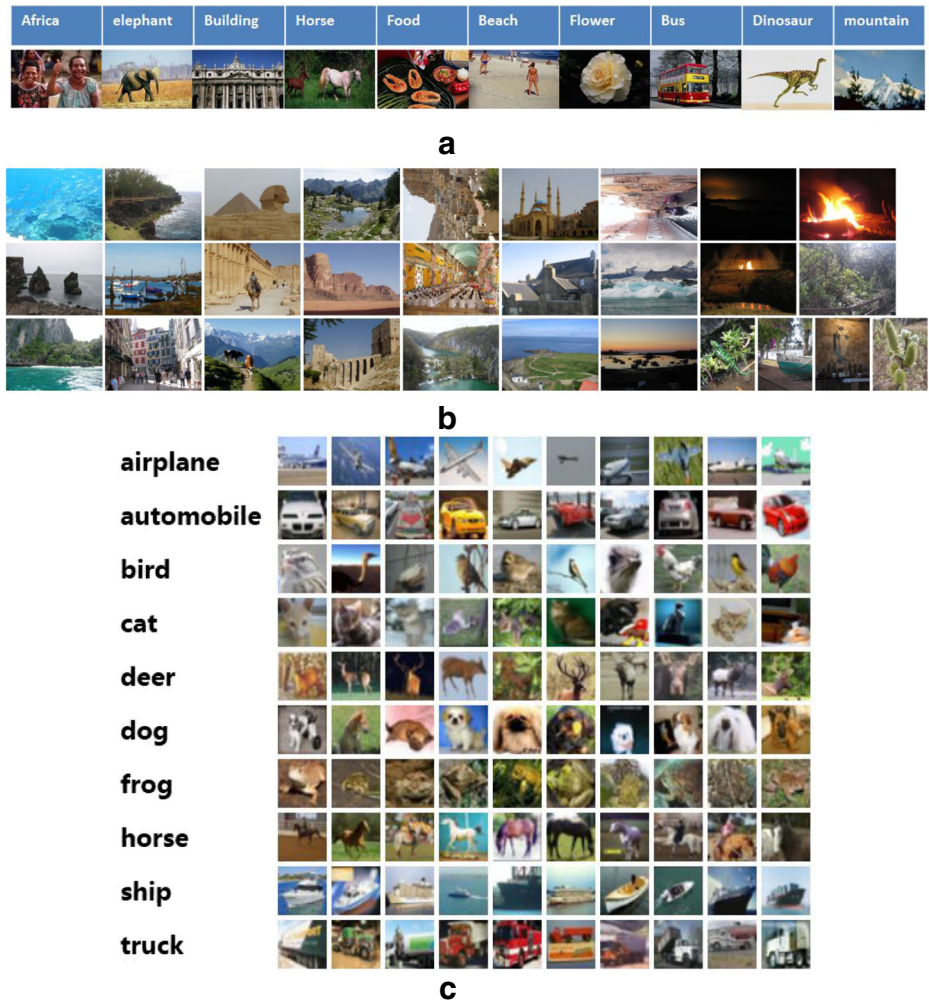


Figure 3 Sample image in experiment datasets: **a** Wang’s dataset, **b** INRIA Holidays dataset, **c** CIFAR-10 Dataset

Figure 3b) consists of 1491 high resolution images with a large variety of scene type, e.g. natural, man-made, water and fire effects, etc.

CIFAR-10 dataset, which consists of 60000 32×32 color images in 10 classes as shown in Figure 3c). The dataset is divided into five training batches and one test batch, each with 1000 randomly-selected images from each class.

To evaluate the retrieval performance, we define images within the same category as relevance.

3.2 Image feature extraction

In our work, images are represented by three low-level features: color, texture, and shape.

For color, we selected the global color histogram (GCH) and the color moments [35]. Firstly, we convert the color space from RGB into HSV. Secondly, we quantify the HSV

color space into 64 bins with the (8, 4, 2) quantization scheme and the direction of hue, saturation and value are quantized into 8 bins, 4 bins and 2 bins, respectively. Then, we calculate 2 moments: color mean, color variance in each color channel, respectively. Thus, a 64-dimensional GCH feature and a 6-dimensional color moment is used.

For texture, a Gabor wavelet [48] is performed on the gray images. Gabor wavelet filters spanning five scales: 1, 3, 5, 7, 11 and four orientations: 0 , $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ are applied to the image. The mean and standard deviation of Gabor wavelet coefficients are used to form the texture feature vector. The vector dimension is 40.

For shape, the Edge Direction Histogram (EDH) [43] is used as the shape features. The edge of the images is extracted using the Sobel edge detection algorithm. An image is converted into YCrCb color space and the statistical edges of horizontal, $\frac{\pi}{4}$, vertical, $\frac{3\pi}{4}$, and isotropic edges are calculated on the component of Y. Thus a total of 5 dimensional EDH edge features are extracted.

All of those features are combined into a feature vector, which results in a feature descriptor with 111 values. The distance between pair of images is computed by the Euclidean distance.

3.3 The influence of the parameter

In this section, we investigate the influence of parameters, the threshold λ about pseudo-labeled images selection. By following the experimental setting in the Section 3.2, we evaluated the proposed method using retrieval comparison.

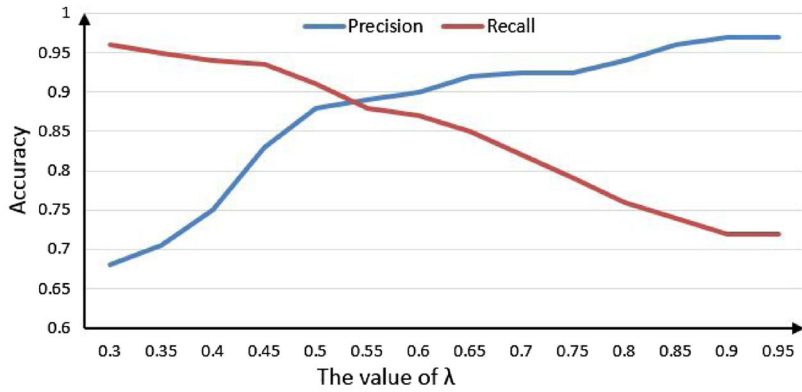
The influence of λ Our proposed method uses the parameter λ to represent a threshold for pseudo-label image selection. Since the unlabeled images in the database are much more than the labeled images and too much uses of pseudo-labeled images would overwhelm the labeled images, an effective system requires a rational selection of the value of λ . To investigate the influence of λ on the retrieval accuracy, we increase the value of λ from 0.3 to 0.9 (the reason we choose to perform our experiment from $\lambda = 0.3$ is that too small value of the threshold would make nonsense in unlabeled image selection) with an interval of 0.05 by performing pseudo-labeling testing on the above three experimental dataset. In our work, we randomly divide each data set into two parts with equal size for training and testing, and calculate the average accuracy of pseudo-labeling in testing set.

We varied the parameter of λ on the three experimental datasets. Figure 4 shows *Precision*/ λ and *Recall*/ λ curve in pseudo-labeling. Clearly, a larger λ will lead to a relative larger pseudo-labeling precision but does not bring a higher recall since a larger λ does reflect a harsh requirement about unlabeled image selection. We can find the pseudo-labeling strategy will get a good performance when $\lambda \in (0.5, 0.65)$ by statistical analyzing (Figure 4).

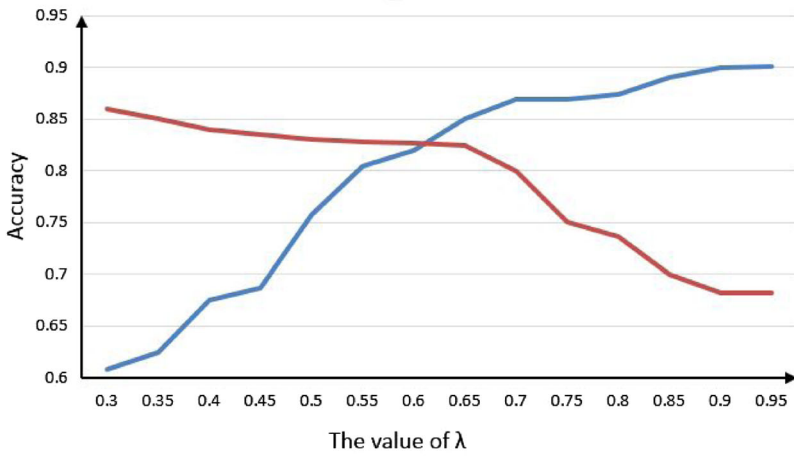
3.4 Comparative performance evaluation

To evaluate the performance of the proposed method, a set of CBIR experiments have been conducted by comparing the proposed algorithm to several state-of-the-art SVM feedback methods [3, 18, 36] that have been used in CBIR system. Here,

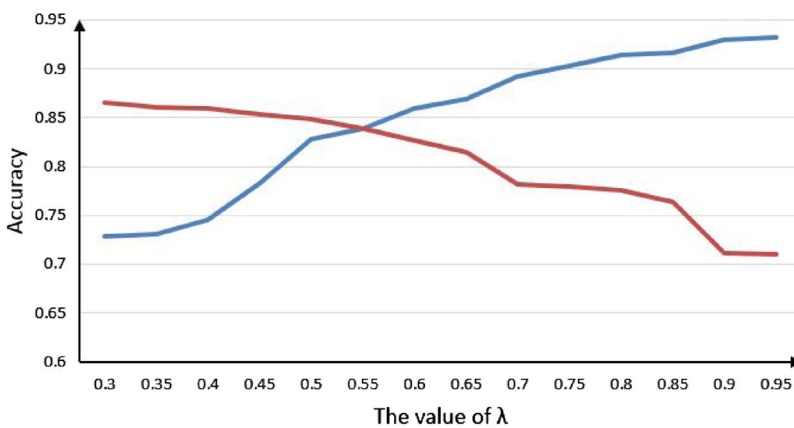
- (1) RF method proposed by S. Tong et al. [36]. This is a SVM-based AL algorithm which simply choose a batch size of examples closet to the current decision boundary for feedback;



a



b



c

Figure 4 Effect of varying various threshold of pseudo-labeled image selection in the three datasets: **a** Wang's dataset, **b** INRIA Holiday dataset, **c** CIFAR-10 dataset



Figure 5 Image retrieved results after the fifth feedback iteration for the query image “Africa” in Wang database: **a** the RF scheme in Ref. [36], **b** the RF scheme in Ref. [3], **c** the RF scheme in Ref. [18] and **d** is the proposed method

- (2) RF methods proposed by B. Demir et al. [3]. This method proposed a new triple criteria method for representative image selection for feedback in active learning;
- (3) RF method proposed by R.J. Liu et al. [18]. This method proposed a SVM-based active feedback scheme using clustering algorithm and unlabeled data;

In addition, an image retrieval method based on former experimental settings without any process of RF is used as the baseline.

In our work, we use the average precision (AP) which defined by NISTTREC video (TRECVID) to evaluate image retrieval performance. The precision in each query is defined as

$$Precision@k = \frac{\sum_{i=1}^k Rel(i)}{k}, \quad (12)$$

where $Rel(i) \in \{0, 1\}$ is a binary function which denotes the ground truth relevance between a query I_q and the i -th ranked images with 1 as relevance, 0 otherwise.

$$AP = \frac{\sum_{M=1}^{m=1} Precision_m}{M}, \quad (13)$$

here, M is the number of queries, $Precision_m$ is the retrieval precision in the m -th query.

At the beginning of image retrieval, the top k images which ranked according to their similarity measure to the query are returned to users for feedback. In our work, only the non-relevant images are marked by user and all other images are automatically marked as relevance by the system. According to the analysis of influence by the parameter in the above section, we do a trade-off selection on the values of parameter for real-world experiment - for our experimental setup, we use $\lambda = 0.5$.

Figure 5 shows image retrieval results after five feedback iteration by four different image retrieval algorithm in Wang's dataset. Figure 5d shows as the retrieval results by using the proposed method after five feedback iteration for the query of an "Africa image" in Wang database. The "black" color for the image boarder is used to distinguish images with relevance and non-relevance. For comparison, the retrieval results by several RF method in Ref. [2, 13, 14] are shown in Figure 5a–c. Figure 6 shows the accuracy comparison about the four algorithms in top-40 results returned by system and a varied number of feedback iteration from 10 categories in Wang's database.

We varied both the number of results returned by system and retrieval feedback iteration of the experiments in the three experiment datasets. Figure 7a–b show the effect of varying model parameters (the number of results returned by system and the number of retrieval iteration is varying from 20–60 and 1–10 respectively) in the four RF algorithm. Clearly a small number of returned images or a large number of feedback iteration are beneficial to retrieval accuracy. Additionally, a larger number of results returned by the system means the number of feedback information given by the user would increase, and a stable classifier could be obtained in a relative short feedback iterations. Clearly, our proposed method is competitive from experimental results under the same terms.

To examine the performance of the proposed method for large image retrieval, we conduct experiments with performance comparison with two state-of-arts methods which are based on supervised hashing and deep CNN [17, 44].

In order to meet the demand of the work of deep learning, we randomly select 1K images (500 images per class) as train set, and the rest images as test query. In those works of deep learning framework, all images are represented by 128 bits binary code and hamming distance is used as similarity measure. Additionally, same train set is used in our proposed

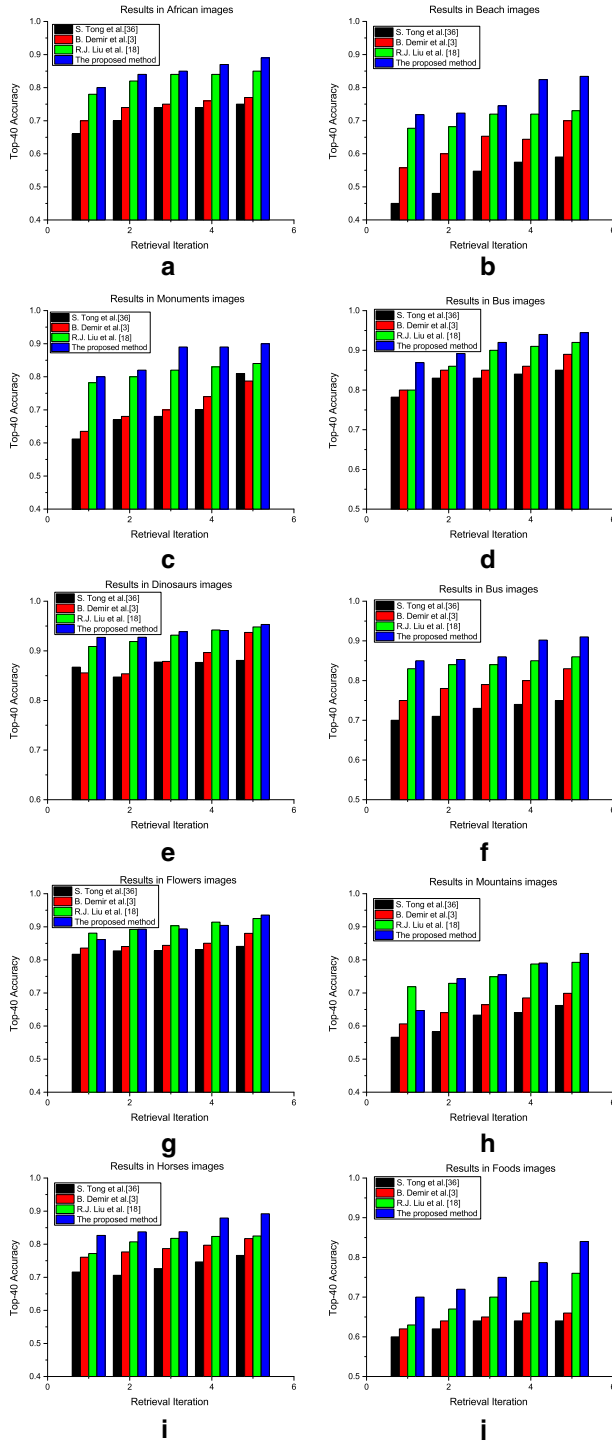
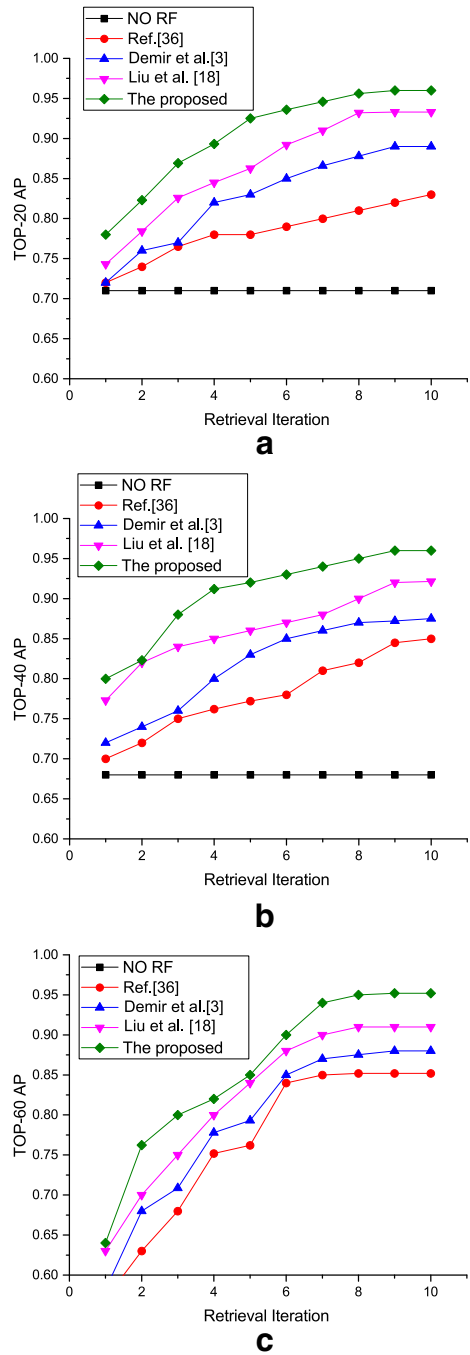


Figure 6 Accuracy comparison of four algorithms in Wang database

Figure 7 Relationship between AP and the number of feedback iterations: **a** top-20 returned images, **b** top-40 returned images and **c** top-60 returned images



work to train the SVM classifier in advance for the sake of fairness. As shown in Figure 8, our proposed method can achieve almost the same retrieval accuracy by certain number of feedback iterations. In fact, our proposed method have its advantage by taking consideration

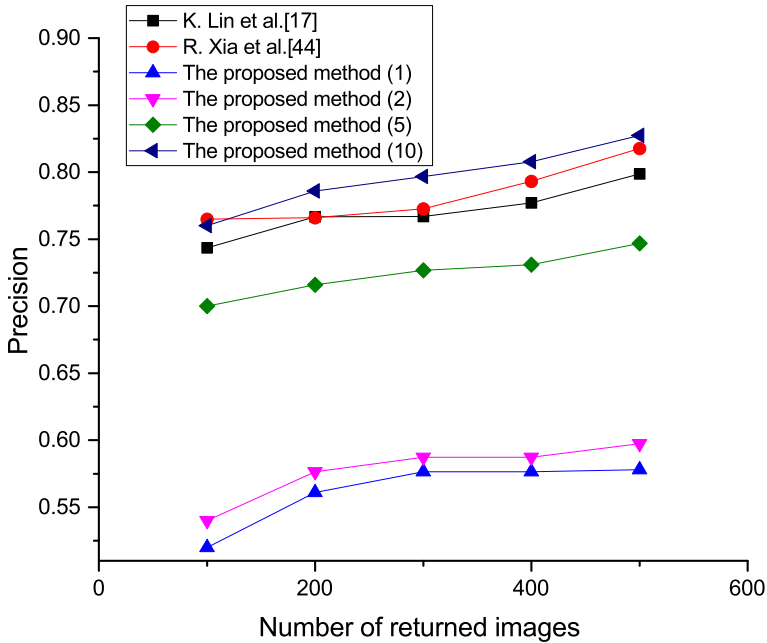


Figure 8 the result on CIFAR-10, the proposed method (k) presents a work with k feedback iteration

of some extra condition which deep learning based technologies must requisite, such as hardware support, vast train set etc.

4 Conclusion

In this paper we have demonstrate the application of CBIR with SVM-based RF technology. In particular, we have mainly focused on solving a drawback that SVM model is sensitive to the training samples, especially when the size of training set is small. Despite this difficulty we have shown is possible to solve with many feedback iterations, the proposed method can build a stable and efficient classifier in a short span of feedback iterations and reduce the effort from user's annotation.

While we have proved the efficacy of the proposed method by qualitative experimental results, such techniques can just reach a certain level of accuracy without further processing. Future work might include contextual information annotation, such as image foreground annotation or scene recognition.

Acknowledgments This work was supported by the Fundamental Research Funds for the Central Universities of China(No.A03013023001050,No.ZYGX2016J095).the National Natural Science Foundation of Sichuan(No.2017JY0229).

References

1. Bezdek, J.C.: Convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**(1), 1–8 (2009)

2. Cao, X., Zhang, C., Fu, H., Guo, X.J.: Saliency-aware nonparametric foreground annotation based on weakly labeled data. *IEEE Trans. Neural Netw. Learn. Syst.* **27**(6), 1253–1265 (2016)
3. Demir, B., Bruzzone, L.: A novel active learning method in relevance feedback for content-based remote sensing image retrieval. *IEEE Trans. Geosci. Remote Sens.* **53**(5), 2323–2334 (2015)
4. Elhamifar, E., Sapiro, G., Yang, A., Satri, S.S.: A convex optimization framework for active learning. In: *IEEE International Conference on Computer Vision (ICCV)*. IEEE Computer Society, pp. 209–216 (2013)
5. Feng, J., Wei, Y., Tao, L., Zhang, C., Sun, J.: Salient object detection by composition. In: *Proc IEEE Int. Conf. Comput. Vis.*, pp. 1028–1035 (2011)
6. Gao, X., Xiao, B., Tao, D., Li, X.: Image categorization: Graph edit distance+edge direction histogram. *Pattern Recogn.* **41**(10), 3179–3191 (2008)
7. Hoi, S.C.H., Jin, R., Zhu, J., Lyu, M.R.: Semi-supervised SVM Batch mode active learning with applications to image retrieval. *Acm Trans. Inf. Syst.* **27**(3), 16–32 (2009)
8. Hu, R.Y., Zhu, X.F., Cheng, D.B., He, W., Yan, Y., Song, J.K., Zhang, S.C.: Graph self-representation method for unsupervised feature selection. *Neurocomputing* **220**, 130–137 (2017)
9. Hu, M.Q., Yang, Y., Shen, F.M., Zhang, L.M., Shen, H.T., Li X.L.: Robust Web Image Annotation via Exploring Multi-facet and Structural Knowledge. *IEEE Transactions on Image Processing* **26**(10), 4871–4884 (2017)
10. Hu, M.Q., Yang, Y., Shen, F., Xie, N., Shen, H.T.: Hashing with Angular Reconstructive Embeddings. *IEEE Transactions on Image Processing* **27**(2), 545–555 (2018)
11. Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. *Euro. Conf. Comput. Vis.* **5302**, 304–317 (2008)
12. Jones, S., Shao, L., Zhang, J., Liu, Y.: Relevance feedback for real-world human action retrieval. *Pattern Recogn. Lett.* **33**(4), 446–452 (2012)
13. Kim, D.W., Lee, K.Y., Lee, D.: kernel-based subtractive clustering method. *A Pattern. Recogn. Lett.* **26**(7), 879–891 (2005)
14. Krizhevsky, A.: Learning multiple layers of features from tiny images (2009)
15. Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9), 1075–1088 (2003)
16. Li, J., Allinson, N.M.: Relevance feedback in content-based image retrieval: a survey. In: *Handbook on Neural Information Processing*, Springer, Berlin, Heidelberg, pp. 433–469 (2013)
17. Lin, K., Yang, H.F., Hsiao, J.H., Chen, C.S.: Deep learning of binary hash codes for fast image retrieval, *Computer Vision and Pattern Recognition Workshops* (2015)
18. Liu, R.J., Wang, Y.H., Baba, T., Masumoto, D., Nagata, S.: SVM-Based active feedback in image retrieval using clustering and unlabeled Data. *Pattern Recogn.* **41**(8), 2645–2655 (2008)
19. Liu, H., Wang, R., Shan, S., Chen, X.: Deep supervised hashing for fast image retrieval. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 2064–2072 (2016)
20. Luo, Y.D., Yang, Y., Shen, F.M., Huang, Z., Zhou, P., Shen, H.T.: Robust discrete code modeling for supervised hashing. *Pattern Recognition* **75**, 128–135 (2018)
21. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 837–842 (1996)
22. Markus, S., Markus, O.: Similarity of color images, *SPIE Storage and Retrieval for Image and Video Databases* (1995)
23. Min, R., Cheng, H.D.: Effective image retrieval using dominant color descriptor and fuzzy support machine. *Pattern Recogn.* **4**(12), 147–157 (2009)
24. Niblack, W., Barber, R., Equitz, W., Fickner, M., Glasman, E., Petkovic, D., Yanker, P.: The QBIC project: Querying images by content using color, texture and shape. In: *Proceedings of the SPIE Conference on Geometric Methods in Computer Vision II*, San Diego, California, USA (1993)
25. Niu, B., Cheng, J., Bai, X., Lu, H.Q.: Asymmetric propagation based batch mode active learning for image retrieval. *Signal Process.* **93**(6), 1639–1650 (2013)
26. Qin, T., Zhang, X.D., Liu, T.Y., Wang, D.S., Ma, W.Y., Zhang, H.J.: An active feedback framework for image retrieval. *Pattern Recogn. Lett.* **29**(5), 637–646 (2008)
27. Rahman, M.M., Antani, S.K., Thoma, G.R.: A learning-based similarity fusion and filtering approach for biomedical image retrieval Using SVM classification and relevance feedback. *IEEE Trans. Inf. Technol. Biomed.* **15**(4), 640–646 (2011)
28. Rocchio, J.: Relevance feedback in information retrieval, *Computer Science* (2000)
29. Russell, B., Torralba, A., Liu, C., Fergus, R., Freeman, W.T.: Object recognition by scene alignment. In: *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 1241–1248 (2007)
30. Samadi, A., Lillicrap, T.P., Tweed, D.B.: Deep learning with dynamic spiking neurons and fixed feedback weights. *Neural Comput.* **29**(3), 578 (2017)

31. Schmid, C.: Weakly supervised learning of visual models and its application to content-based retrieval. *Int. J. Comput. Vis.* **56**(1), 7–16 (2004)
32. Schohn, G., Cohn, D.: Less is more: active learning with Support Vector Machines. In: Proc 17th International Conf. Mach. Learn., Stanford, CA, USA, pp. 839–846 (2000)
33. Scott, G., Klaric, M., Davis, C., Shyu, C.R.: Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. *IEEE Trans. Geosci. Remote Sens.* **49**(5), 1603–1616 (2011)
34. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(11), 2298–2304 (2017)
35. Tao, D.C., Tang, X.O., Li, X.L., Wu, X.D.: Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(7), 1088–1099 (2006)
36. Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: Proceedings of the ACM Multimedia, pp. 107–118 (2001)
37. Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y.D., Li, J.T.: Deep learning for content-based image retrieval: a comprehensive study. In: ACM International Conference on Multimedia. ACM, pp. 157–166 (2014)
38. Wang, X.Y., Chen, J.W., Yang, H.Y.: A new integrated SVM classifiers for relevance feedback content-based image retrieval using EM parameter estimation. *Appl. Soft Comput.* **4**(11), 2787–2804 (2011)
39. Wang, X., Zhang, B., Yang, H.: Active SVM-based relevance feedback using multiple classifiers ensemble and features re-weighting. *Eng. Appl. Artif. Intell.* **26**, 368–381 (2013)
40. Wang, X.Y., Li, Y.W., Yang, H.Y., Chen, J.W.: An image retrieval scheme with relevance feedback using feature reconstruction and SVM reclassification. *Neurocomputing* **127**, 214–230 (2014)
41. Wang, X.Y., Liang, L.L., Li, W.Y., Li, D.M., Yang, H.Y.: A new SVM-based relevance feedback image retrieval using probabilistic feature and weighted kernel function. *J. Vis. Commun. Image Represent.* **38**, 256–275 (2016)
42. Weir, J.P.: Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength Cond. Res.* **19**(1), 231 (2005)
43. Wu, K., Yap, K.H.: Fuzzy SVM for content-based image retrieval. *IEEE Comput. Intell. Mag.* **1**(2), 10–16 (2005)
44. Xia, R., Pan, Y., Lai, H., Liu, C., Yan, S.: Supervised hashing for image retrieval via image representation learning. In: Proceedings of the Twenty-Eighth AAAI Conf. on Artificial Intelligence, vol. 3, pp. 2156–2162 (2014)
45. Yang, Y., Shen, F.M., Shen, H.T., Li, H.X., Li, X.L.: Robust discrete spectral hashing for large-scale image semantic indexing. *IEEE Transactions on Big Data* **1**(4), 162–171 (2015)
46. Yang, Y., Ma, Z.G., Yang, Y., Nie, F.P., Shen, H.T.: Multitask spectral clustering by exploring intertask correlation. **45**(5), 1083–1094 (2015)
47. Yang, Y., Shen, F.M., Huang, Z., Shen, H.T., Li, X.L.: Discrete nonnegative spectral clustering. *IEEE Transactions on Knowledge and Data Engineering* **29**(9), 1834–1845 (2017)
48. Zhang, Z., Ji, R.R., Yao, H.X., Xu, P.F., Wang, J.C.: Random sampling SVM based soft query expansion for image retrieval. In: IEEE International Conference on Image and Graphics, Chengdu, Sichuan, China, pp. 805–809 (2007)
49. Zhu, X.F., Zhang, L., Huang, Z.: A sparse embedding and least variance encoding approach to hashing. *IEEE transactions on image processing* **23**(9), 3737–3750 (2014)
50. Zhu, X.F., Li, X.L., Zhang, S.C.: Block-row sparse multiview multilabel learning for image classification. *IEEE transactions on cybernetics* **46**(2), 450–461 (2016)
51. Zhu, X.F., Li, X.L., Zhang, S.C., Ju, C.H., Wu, X.D.: Robust joint graph sparse coding for unsupervised spectral feature selection. **27**(2), 545–555 (2018)