Content based image retrieval via a transductive model

Songhao Zhu · Liming Zou · Baojie Fang

Abstract Content based image retrieval plays an important role in the management
of a large image database. However, the results of state-of-the-art image retrieval
approaches are not so satisfactory for the well-known gap between visual features
and semantic concepts. Therefore, a novel transductive learning scheme named
random walk with restart based method (RWRM) is proposed, consisting of three
major components: pre-filtering processing, relevance score calculation, and can-
didate ranking refinement. Firstly, to deal with the problem of large computation
cost involved in a large image database, a pre-filtering processing is utilized to filter
out the most irrelevant images while keeping the most relevant images according
to the results of a manifold ranking algorithm. Secondly, the relevance between a
query image and the remaining images are obtained with respect to the probability
density estimation. Finally, a transductive learning model, namely a random walk
with restart model, is utilized to refine the ranking taking into account both the
pairwise information of unlabeled images and the relevance scores between query
image and unlabeled images. Experiments conducted on a typical Corel dataset
demonstrate the effectiveness of the proposed scheme.

Keywords Content based image retrieval · Semi-supervised learning · Pre-filtering
processing · Relevance score measurement · Candidate ranking refinement

S. Zhu (✉) · L. Zou · B. Fang
School of Automation, Nanjing University of Posts and Telecommunications,
Nanjing, 210046 China
e-mail: play_tree@163.com

L. Zou
e-mail: zoulm@njupt.edu.cn

B. Fang
e-mail: bjfang@njupt.edu.cn

1 Introduction

In the past decades, the size of image collections has grown dramatically, accompanied by the popularity of inexpensive hardware. Faced with the already huge and quickly increasing volume of images, there has been an increasing demand for effective technologies to manage large image databases. Among many technologies, content based image retrieval technology is received more and more attention for its aim is to search in a large image database for those images which are perceptually similar to a given query image.

A common problem faced by researchers in real-world applications is that only a small number of positive samples are given, combined with the need to develop a model to retrieve images with similar visual content to a given query image from a large image database. In this sense, discriminative models, such as support vector machine (SVM), do not work (Natsev et al. 2005). In literature (He et al. 2004), this issue is tackled by selecting several unlabeled images as pseudo-negative samples. Alternatively, this issue is solved by a manifold-ranking algorithm which propagates label information from training samples to testing samples (He et al. 2006). In literature (Cui and Zhang 2007), feedback from users is incorporated into to the system to form an iterative refinement processing. Many other approaches are proposed to deal with content-based image retrieval (Beecks et al. 2011; Su et al. 2011; Akakin and Gurcan 2012; Zhang et al. 2012). In spite of the fact that encouraging performance has been obtained, the content-based image retrieval problem has not been satisfactorily solved due to the large gap between low-level features and high-level semantic concepts, i.e., images of the dissimilar semantic content may share some common low-level features, while images of similar semantic content may be scattered in the feature space.

To narrow the well-known semantic gap, many different methods have been proposed, our scheme consists of following three steps, as illustrated in Fig. 1. First, when a large-scale image database is incorporated, the computation cost will be expensive. A possible solution to this problem is to introduce a pre-filtering process, which can filter out the majority of irrelevant samples while retaining the most relevant ones in accordance with the results of the manifold-ranking algorithm. Then, a probability density estimation algorithm is used to obtain the relevance scores between the input query image and the remaining images in the database, and the relevance scores are stored as the restart vector for a later candidate ranking refinement processing. Finally, to make full use of both the pairwise information of unlabeled images and the relevance scores between input query image and unlabeled images, an image-based semi-supervised learning method, a random walk with restart algorithm, is exploited
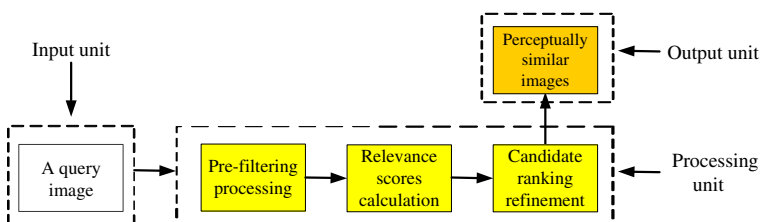

Fig. 1 Proposed content-based image retrieval scheme

to refine candidate ranking. The final retrieval images are those with the highest similarity probability.

The organization of this paper is as follows. Section 2 introduces the related works. Section 3 details the pairwise image similarity and the pre-filtering process. Sections 4 and 5 introduce the method of computing the image relevance scores and the candidate ranking refinement process, respectively. Experimental results are provided in Section 6, followed by concluding remarks and a discussion of future work in Section 7.

2 Related works

Up to now, many different methods have been proposed to deal with the issue of image retrieval, which can be mainly classified into two categories: one is to search for appropriate metrics to estimate perceptual similarity, and the other is to exploit new learning strategies to improve retrieval accuracy.

In the first category of the study, several distance functions have been exploited to measure the similarity between a query tag and all the images in the database. In Li et al. (2002), a perceptual distance function based on the Minkowski distance is presented to maximize the similarity between two images. In Jing et al. (2002), the Earth Mover's Distance (Rubner et al. 1998) is utilized as the similarity metrics to implement image retrieval. Another example is the Manhattan distance metrics. In Stricker and Orengo (1995), Pass (1997), Kokare et al. (2003), the authors conclude that Mahalanobis distance is superior to Euclidean distance in handling image retrieval with color information. However, these traditional similarity metrics are based on the pairwise distance, and their effectiveness is not so satisfactory.

In the second category of the study, approaches can be further categorized into inductive and transductive ones according to whether unlabeled data is utilized in the training stage or not. The goal of an inductive method (such as SVM) is to create a classifier with manually pre-labeled samples, and then propagate labels to unlabeled samples by the trained model. However, this learning model has a major drawback. With the limitation of labeled training samples, the distribution character of the entire dataset can not be well represented and the accuracy of retrieval results will not be so high (Jeon et al. 2003). On the other hand, transductive methods aim at accurately predicting the relevance between all unlabeled images and a query image, which can be obtained during the training stage. Graph-based semi-supervised learning methods, a major family of transductive models, have attracted lots of attention recently. Many works on this topic are reported (such as Yuan et al. 2006, Xu et al. 2011, Wang et al. 2012), and some of them have been applied to image or video semantic annotation.

3 Computing pairwise similarity and prefiltering processing

Here, we first introduce the pairwise similarity of samples by taking into account the influence of both distance and neighborhood. Then we analyze the importance and necessity of the pre-filtering processing.

3.1 Pairwise similarity of samples

Graph-based semi-supervised learning actually relies on two basic assumptions as proposed in Zhang and Zhang (2004). The first assumption is the neighborhood assumption: nearby images are likely to have the same label. The second assumption is the structure assumption: images with the same "structure" (typically referred to as a cluster or a manifold) are likely to have the same label. Therefore, the evaluation of perceptual similarity between images plays a crucial role for graph-based methods as it is the basis of label propagation. In many literatures including Su et al. (2011), Akakin and Gurcan (2012), Yuan et al. (2006), Xu et al. (2011), Wang et al. (2012), the first assumption is enforced into pairwise similarity metric between images x_i and x_j with Manhattan distance measure as:

$$S^n(i, j) = \exp\left[-\frac{|x_i - x_j|}{\sigma_n}\right] = \exp\left[-\prod_{l=1}^{k}\frac{\left|x_i^l - x_j^l\right|}{\sigma_n^l}\right] \tag{1}$$

where x_i^l and x_j^l are the lth dimension in the low-level feature vector of images x_i and x_j respectively, k is the dimension number, and σ_n^l is the positive parameter that reflects the scope of different dimension. In our current implementation, σ_n^l is the mean of x_i^l and x_j^l. An intuitive idea of (1) is that perceptual similarity between two images increases as the decrease of their distance in the feature space.

However, the second assumption is not embedded into pairwise similarity metric for many graph-based semi-supervised learning algorithms, including Yuan et al. (2006), Xu et al. (2011), Wang et al. (2012). It means that during the label propagation process, the influence of the structure difference is not taken into account, which leads to an isotropic propagation. To the best of our knowledge and references (Gunnarsson and Jones 1980; Bower and Balogh 2004), it is intuitive that the density difference between different clusters is usually larger than the density difference within a cluster. Therefore, we can assume with confidence that the perceptual similarity between two images, not only increases as the decrease of their distance in the feature space, but also increases as the decrease of their density variation.

The probability density of an image x_i with Laplace kernel is defined as:

$$p_i = \frac{1}{N_i}\sum_{v=1}^{N_i}k(x_i - x_v) = \frac{1}{N_i}\sum_{v=1}^{N_i}\left(\prod_{l=1}^{k}\frac{1}{2\sigma_n^l}\exp\left[-\frac{\left|x_i^l - x_v^l\right|}{\sigma_n^l}\right]\right) \tag{2}$$

where N_i is the number of neighbors to image x_i. Correspondingly, the structure similarity metrics between images x_i and x_j is defined as:

$$S^p(i, j) = \exp\left(-\frac{|p_i - p_j|}{\sigma_p}\right) = \exp\left[-\prod_{l=1}^{k}\frac{\left|p_i^l - p_j^l\right|}{\sigma_p^l}\right] \tag{3}$$

where σ_p^l is the positive parameter that reflects the scope of different dimension. An intuitive idea of (3) is that perceptual similarity between two images increases as the decrease of their density variation.

Combining the neighborhood assumption and structure assumption, a pairwise similarity metric between images x_i and x_j can be formulated as:

$$S(i, j) = S^n(i, j) \bullet S^p(i, j) = \exp\left[-\frac{|x_i - x_j|}{\sigma_n}\right] \bullet \exp\left[-\frac{|p_i - p_j|}{\sigma_p}\right]$$

$$= \exp\left[-\prod_{l=1}^{k}\frac{|x_i^l - x_j^l|}{\sigma_n^l}\right] \bullet \exp\left[-\prod_{l=1}^{k}\frac{|p_i^l - p_j^l|}{\sigma_p^l}\right] \quad (4)$$

where the symbol "$\bullet$" denotes the Hadamard product. The first term in the right side of (4) indicates that the similarity between two images increases as the decrease of their distance in the feature space, and the second one indicates that the similarity increases as the decrease of the corresponding density variation.

3.2 Pre-filtering processing

As mentioned, the computational cost of graph-based semi-supervised learning algorithms is intractable when a large dataset is involved. More specifically, if the number of entire images (including labeled and unlabeled images) is m, then the computation cost will be $O(m^3)$. To tackle this problem, we exploit an efficient pre-filtering processing approach to discard the majority of irrelevant images while preserving the majority of relevant ones.

From the point of view of retrieval results, the process of pre-filtering processing should simultaneously satisfy two criterions: one is low computational cost and the other is high recall rate. Here, a modified nearest neighbor rule is utilized to perform the prefiltering processing. More specifically, images in the database are ranked according to the pairwise similarity obtained using (4): the larger the value of pairwise similarity between an input query image and an image in the database is, the higher the similarity score of the image in the database is. After ranking images in the database according to the similarity scores, a specified percentage of images in the database are discarded and this can significantly reduce the computational cost.

The purpose of such a pre-filtering processing is to ensure enough connection between relevant images while removing weak connection between irrelevant images. Therefore, a graph constructed in this way is a sparsely- connected graph, and will be utilized as a similarity graph to complete the refinement processing in Section 5.

4 Relevance scores

For a query image, a scale should be introduced to describe the correlation between the query image and relevant images in the database. Here, the scale denotes the confidence measure of one image perceptually similar to the query image and is represented by a fuzzy number in the range of [0, 1]. In our current implementation, each relevant image is assigned a real value to indicate the level of the similarity to the query image. The similarity level can be viewed as the membership of the feature value of images relevant to the query, which is computed by hyperbolic tangent function (Kokare et al. 2003). The fuzzy function takes the distance between
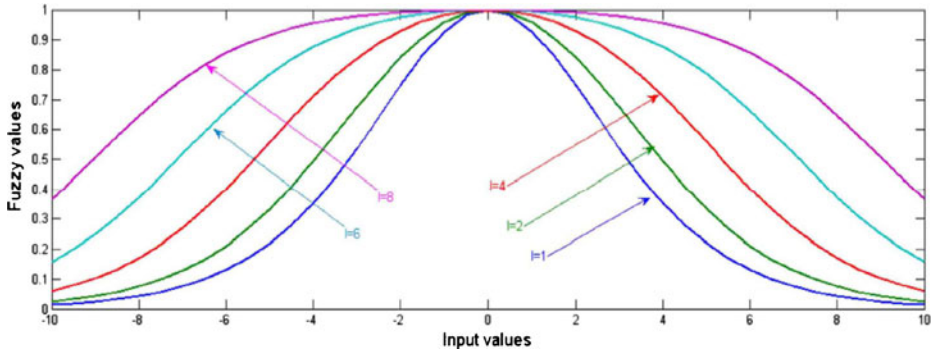
Fig. 2 Plots of hyperbolic tangent function with different shape parameter, l_c, varying from 1.0 to 8.0, where d_c is 3.5 and the mean is 0

a relevant image and a query image as the input and maps it to a value in the range of [0, 1].

Given a query image q with feature vector x_q, the membership $q(i)$ of an image i with feature vector x_i from the same categorization is computed as:

$$q(i) = \frac{1}{2 \times \tanh\left(\frac{l_c}{d_c}\right)} \left[\tanh\left(\frac{|x_q - x_i| + l_c}{d_c}\right) - \tanh\left(\frac{|x_q - x_i| - l_c}{d_c}\right) \right] \quad (5)$$

where $tanh(.)$ is a hyperbolic tangent function, and l_c is the spread of the function that controls the shape. Furthermore, d_c is the average distance between cluster centers of different categories, computed as:

$$d_c = \frac{2}{M(M-1)} \sum_{r=1}^{N} \sum_{t=r+1}^{N} |c_r - c_t| \quad (6)$$

As l_c decreases the shape of the hyperbolic tangent function is more concentrated to the function mean. If users prefer a lose condition, then l_c should be set larger.
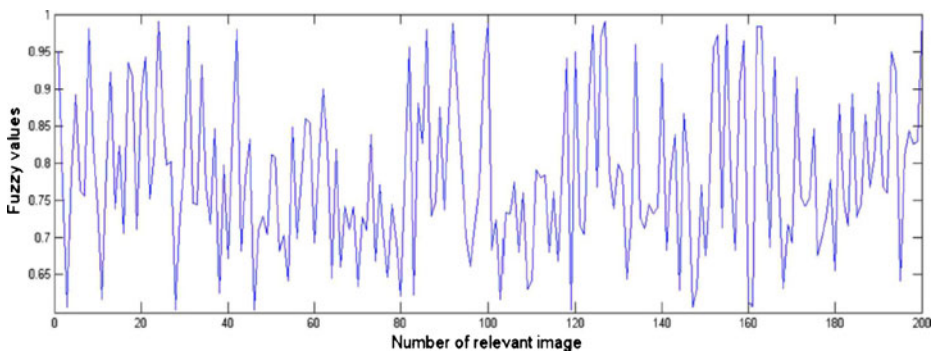

Fig. 3 Fuzzy memberships of a query image. For each query image, the number of relevant images is chosen as 200 after pre-filtering processing

Therefore, one image relatively far from the cluster center can still have a high membership value, as shown in the following figure.

Several plots of hyperbolic tangent function with different l_c is shown in Fig. 2, where d_c is 3.5 and the mean is 0. For each relevant image in the database, its membership value to a query is computed and utilized as the basic of the process of ranking refinement. One example is shown in Fig. 3, from which two interesting phenomenon can be seen: one is the selected hyperbolic tangent function can better demonstrate the similarity of relevant images in the database to the query image, and the other is the pre-filtering processing well preserve these images relevant to a query image.

5 Basic algorithm

From the point of view of pattern recognition, the process of ranking refinement can be considered as a process of generating more accurate relevance score of database images for a query. With the membership values of a query image and the correlations between database images, a transductive method, namely random walk with restart method, is here exploited to deal with the issue of ranking refinement. For random walk with restart method, there are mainly two steps: construction of a similarity graph and refinement of candidate ranking.

5.1 Construction of a similarity graph

A similarity graph is constructed based on the following two main phases: the choice of vertexes and the connection of vertexes. The construction of the similarity graph helps to fully describe the relationship between a query image and these relevant images in the database.

Step 1: Choosing vertexes. For a query image q, let $\Gamma(=\{1, 2,\ldots,N\})$ be a set of N images generated by the nearest neighborhood rule. Then, a similarity graph G is constructed with the query image q and all the images in A:

$$G = (V, E) \tag{7}$$

where V is the image set Γ, and all vertexes are connected to form the weighted edge set E.

Step 2: Connecting vertexes. All vertexes in the similarity graph G is connected by a $N \times N$ affinity matrix W, where $w_{i,j}$ is 0 and $w_{i,j}$ when $i \neq j$ is:

$$w_{i,j} = S(i, j) = S^n(i, j) \bullet S^p(i, j) \tag{8}$$

5.2 Refinement of candidate ranking

The process of ranking refinement consists of the following several phases: the setting of restart vector, the construction of a normalized matrix, and the completion of a random walk with restart algorithm. Such a processing helps to improve the original ranking results of image retrieval.

Step 1: Setting Restart Vector. In our current implementation, membership values of a query image q, $qm = \{q(1), q(2), \ldots, q(N)\}$ are chosen as the restart

vector e, and e is normalized to ensure the sum of all components in e is one.

Step 2: Constructing a normalized matrix. In our current implementation, the affinity matrix W is normalized to ensure the sum of each column is one.

Step 3: Completing a random walk with restart algorithm.

Step 3.1: Segmenting the similarity graph G into k partitions according to the idea in Karypis and Kumar (1999).

Step 3.2: Decomposing the normalized matrix W into two diagonal matrices:

$$W = W_1 + W_2 \tag{9}$$

where W_1 contains all within-partition links, and W_2 contains all cross-partition links which is the basic of the ranking.

Step 3.3: Denoting diagonal matrix W_1 as:

$$W_1 = \begin{pmatrix} W_{1,1} & 0 & ... & 0 \\ 0 & W_{1,2} & ... & 0 \\ ... & ... & ... & ... \\ ... & ... & 0 & W_{1,n} \end{pmatrix} \tag{10}$$

where $W_{1,i}$ is the ith component of diagonal matrix W_1.

Step 3.4: For each component $W_{1,i}$, the corresponding component $\sigma_{1,i}^{-1}$ is:

$$\sigma_{1,i}^{-1} = \left(I - \alpha^* W_{1,i}\right)^{-1} \tag{11}$$

where I is an identity matrix, and α is a parameter that control the probability of returning to the vertex i.

Step 3.5: Performing low-rank approximation of W_2 to achieve a vertex-concept matrix U and a concept-concept matrix S:

$$W_2 = U^* S^* U^T \tag{12}$$

where S is such a diagonal matrix whose diagonal elements are the eigen-values of W_2 ranked in decreasing order, and U is such a matrix where each column is the eigen-vector of W_2 and the order of each column is similar to the corresponding eigen-vector. Furthermore, the superscripts 'T' and '$*$' denote the matrix transpose operation and the matrix multiplication operation respectively.

Step 3.6: Constructing a diagonal matrix σ_1^{-1} as:

$$\sigma_1^{-1} = \begin{pmatrix} \sigma_{1,1}^{-1} & 0 & ... & 0 \\ 0 & \sigma_{1,2}^{-1} & ... & 0 \\ ... & ... & ... & ... \\ ... & ... & 0 & \sigma_{1,n}^{-1} \end{pmatrix} \tag{13}$$

Step 3.7: Achieving optimum membership values of all relevant database images to the query image q as:

$$\gamma = (1 - \alpha)^* \left(\sigma_1^{-1} + \alpha^* \sigma_1^{-1*} U^* \Lambda^* U^{T*} \sigma_1^{-1}\right)^* e \tag{14}$$

where the superscripts '*' denotes the matrix multiplication operation and the formula of Λ is:

$$\Lambda = \left(S^{-1} - \alpha^* U^{T*} \sigma_1^{-1*} U\right)^{-1} \tag{15}$$

The ith element of $n \times 1$ vector γ is the optimum membership value for image i. Top t images with the highest probability are selected as retrieved images for each query image q, and these t images are termed as *top@t*.

6 Implementation issues

Response time is an issue that should be seriously considered when designing an image retrieval system. It is intolerable for users to wait several minutes before retrieval system returns satisfactory results. For random walk with restart algorithm, there exists the inversion and multiplication of large scale matrices as shown in (14), which are difficult to be implemented due to the limitation of calculation ability and memory quantity. For an image dataset with 5,000 images, it will need 5,000 *5,000 *32 bytes memory to represent the similarity matrix. To tackle this problem, pre-filtering processing is here utilized to construct a sparse affine matrix by only connecting neighborhood points. With such processing, the memory quantity and computation cost can be greatly reduced.

Another issue is the construction of sparse affine matrix W. In our current implementation, the sparse affine matrix W is constructed in such a way: first, the relevance score of each image in the database is computed off-line to construct the original sparse affine matrix W; second, M nearest neighbors of each query image are selected, and the relevance of M nearest neighbors to each query image are calculated using (4); finally, one row and column are added to the sparse affine matrix W. All the other operations will be performed using this enlarged matrix W.

7 Experimental results

7.1 Experiment design

To test the proposed method and compare it with other methods described in the literature, Corel dataset is here used as the ground truth datasets.

Corel dataset As a general-purpose image database, Corel consisting of 5,000 image is manually sub-divided into 50 semantic categories and there are 100 images on the same topic in each category. Table 1 lists some exemplary categories. Each image in Corel database is used as a query, and the performance over all the 5,000 images is used as the evaluation metric.

Table 1 Examples of some categories in Corel dataset

No	1	2	3	4	5	6	7	8	9	10
Category	Beach	Flower	Horse	Tiger	Sunset	Benthos	Plane	Pyramid	People	Mountain

Feature selection   Feature selection is an open issue and have a great impact on the performance of image retrieval. In our current implementation, low-level feature vector consists of the following aspects: color histogram, color moment and wavelet texture.

- 128-D color histogram in HSV color space, where 8 bins for Hue, 4 bins for Saturation and 4 bins for Value;
- 225-D block-wise color moments in LAB color space: extracted over 5×5 grid partitions, where each grid partition is described by a 9-D feature;
- 36-D Pyramid Wavelet textures: extracted over 6-level Haar Wavelet transformation, where each level is described by a 6-D feature: mean and variance of coefficients in high/high, high/low, and low/high bands.

Parameters choice   There are six parameters needed to be set in the proposed algorithm: $N, l_c, \alpha, \sigma_n^l, \sigma_p^l$ and iteration steps. In our current implementation, K-NN is exploited to find appropriate nearest neighbors of each query image to construct the sparse matrix W. Based on the experimental results introduced in Section 7.2, N is chosen 200. In our experiments, l_c and α set to be are 1.2 and 0.4 respectively through cross validation. σ_n^l and σ_p^l are set as the average in the lth dimension respectively. The number of iteration steps is set to be 50 due to no improvement in performance with more iteration.

Evaluation metric   Similar to previous literatures on image retrieval, precision, recall, and F_1 are used to evaluate the performance of various approaches. For each query image q, the number of detected images is denoted as $|q_{\text{detected}}|$, the number of relevant images among detected images is denoted as $|q_{\text{correct}}|$, and the number of relevant images from the same category is denoted as $|q_{\text{ground}}|$. Then corresponding formulations are:

$$
\begin{cases}
precision = \dfrac{|q_{\text{correct}}|}{|q_{\text{detected}}|}, & recall = \dfrac{|q_{\text{correct}}|}{|q_{\text{ground}}|} \\
F_1 = \dfrac{2 \times precision \times recall}{precision + recall} = \dfrac{2 \times |q_{\text{correct}}|}{|q_{\text{detected}}| + |q_{\text{ground}}|}
\end{cases}
\tag{16}
$$

7.2 Comparison experiment

In the section, three kinds of calculation methods are considered and compared: the proposed random walk with restart based method (RWRM) which is a transductive method, the Support Vector Machine based method (SVMM) in Natsev et al. (2005) which is an inductive model based method, and the manifold ranking method (MRM) in He et al. (2006) which is another transductive model based method. Figures 4, 5 and 6 reports the comparison experiment results in terms of average precision, average recall and average F1-measure respectively.

From Figs. 4–6, it can be seen that the proposed RWRM shows better performance than SVM-based method and MR-based method, which can be analyzed from the following aspects. For instructive model based method, the generalization ability of a trained classifier is seriously influenced by the insufficiency of labeled training samples. Correspondingly, the proposed transductive model based method can better
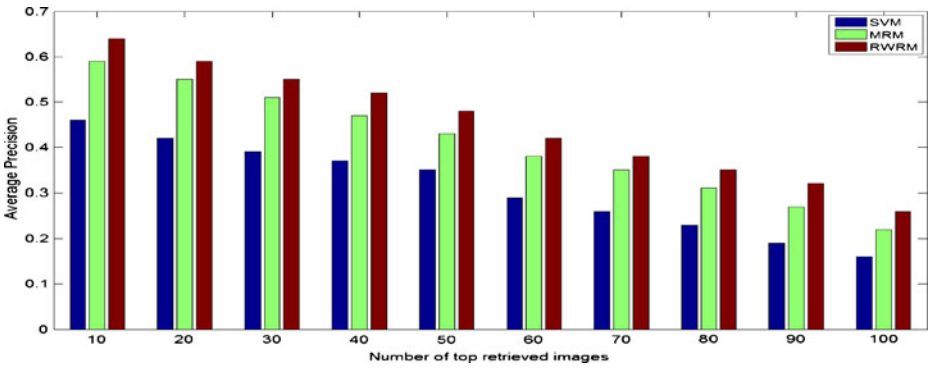
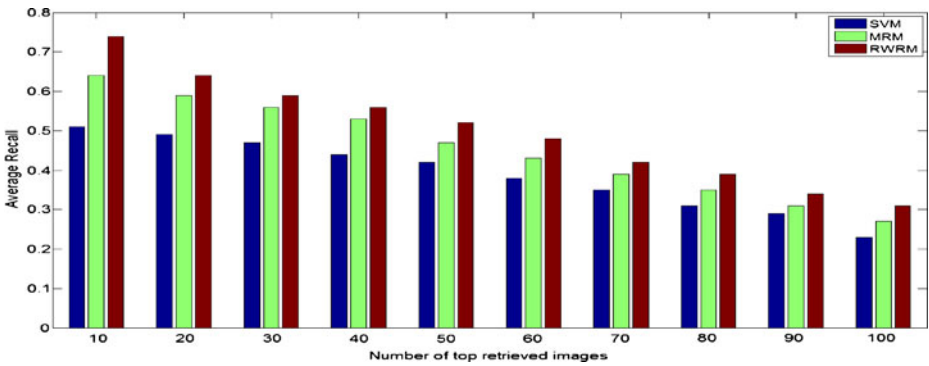**Fig. 4** Average precision with different top returned images for three methods

Fig. 5 Average recall with different top returned images for three methods
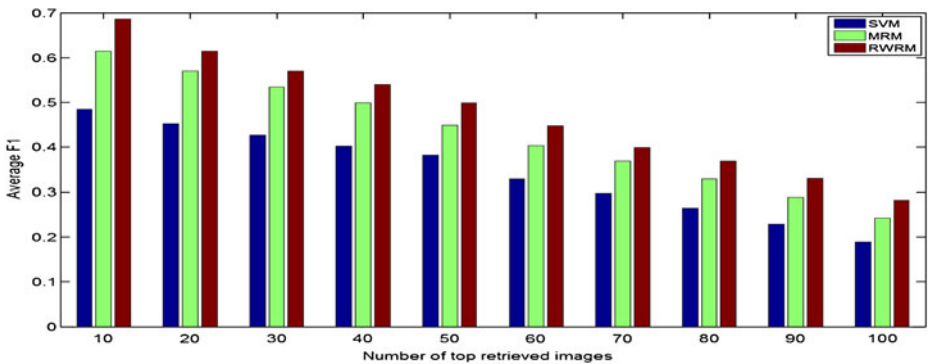

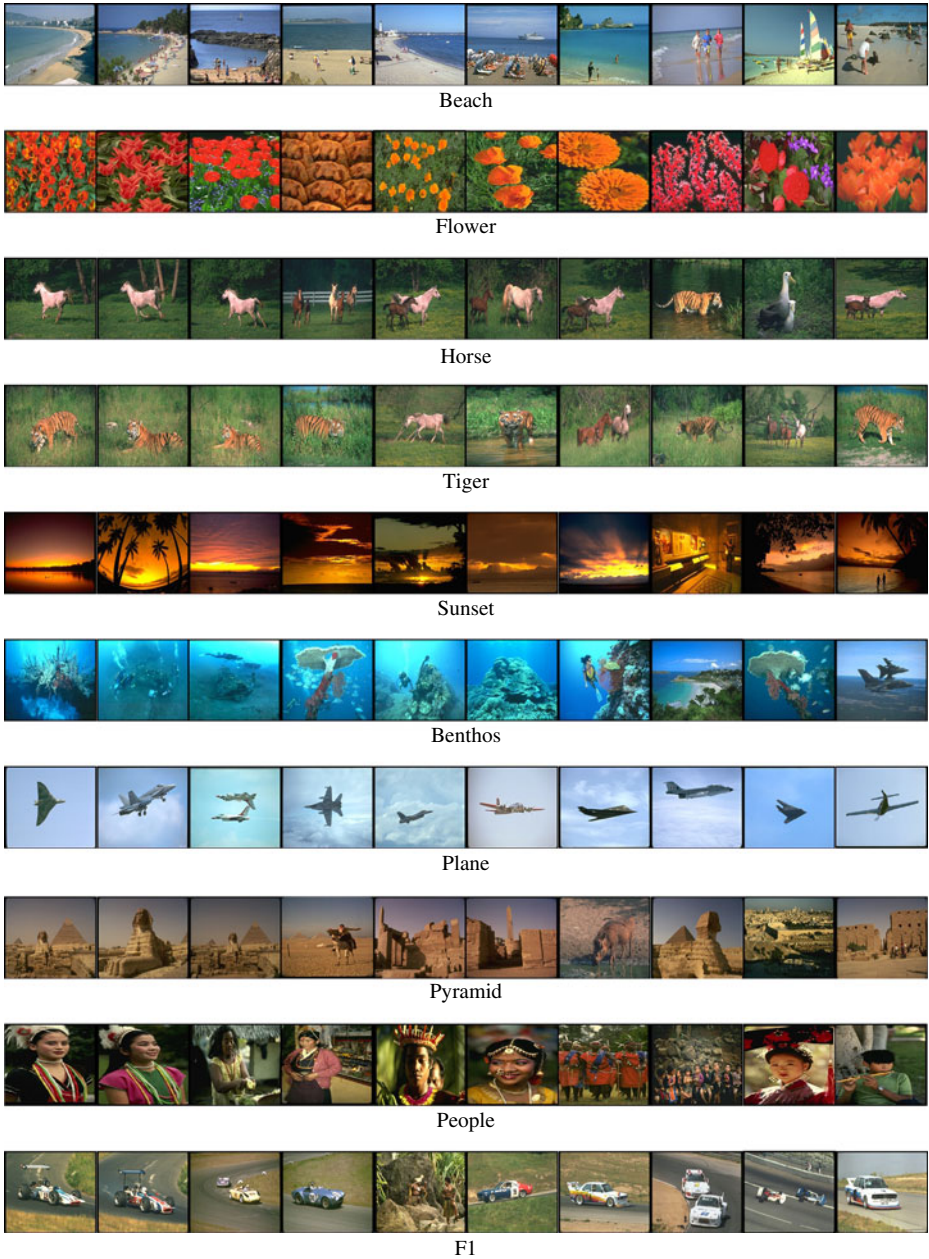**Fig. 6** Average F1-measure with different top returned images for three methods

Fig. 7 Examples of the top ten retrieved images for ten categories. In each group, the first image is the query image, and also the top retrieved image

evaluate the perceptual similarity among a query image and relevant dataset images by exploring the relevance of all data points (including both labeled and unlabeled ones), which helps to achieve better performance on the label propagation.
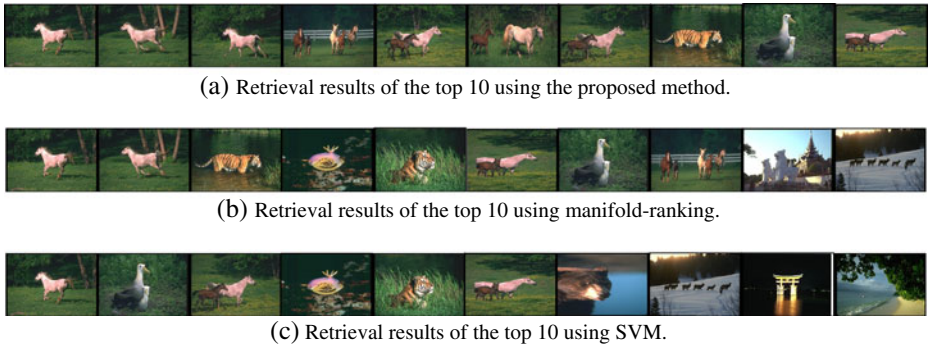
(a) Retrieval results of the top 10 using the proposed method.

(b) Retrieval results of the top 10 using manifold-ranking.

(c) Retrieval results of the top 10 using SVM.

Fig. 8 Examples of the top ten retrieved images, where the first image in each group is the query image and also the top retrieved image. **a** Retrieval results of the top ten using the proposed method. **b** Retrieval results of the top ten using manifold-ranking. **c** Retrieval results of the top ten using SVM

Compared with MRM, our RWRM gains improvement on recall and precision, which happens due to different construction forms of similarity matrix between a query image and corresponding relevant images. More specifically, the pairwise similarity based on neighborhood assumption can not sufficiently reflect the contribution from one image to another one in MR-based model. Correspondingly, our method implements the "anisotropic" contribution by simultaneously taking into account both neighborhood assumption and structure assumption when designing a pairwise similarity matrix, which further improves the performance of normal transductive model based method.

Figure 7 shows the top ten matched images for ten categories of all the 50 categories using the proposed approach. In each group, the first image is the query image, which is also the top matched image.

Comparison of the top ten returned images responsive to an exemplar query image for different methods is shown in Fig. 8.

8 Conclusions

In this paper, we present a novel transductive learning framework to deal with the issue of content-based image retrieval, inspired by a recently developed random walk with restart algorithm. To decrease computation cost, a pre-filtering process is first introduced into the proposed scheme. The relevance score of each database images to a query is then represented as a membership value calculated by hyperbolic tangent function. To consider both the membership value of a query image and the correlations between images in the database, a random walk with restart algorithm is utilized to solve the difficult case where only a small number of positive training images are available. Experiments conducted on a general-purpose image database demonstrate that the proposed framework can effectively improve the performance of image retrieval.

Though random walk with restart based algorithm achieved better performance, semantic modeling for image retrieval is still a challenging problem due to the difficulty of describing image content, as well as the lack of training data. Hence,

there is still huge room for improvement, such as the issue of choosing which similarity metric and learning strategy. Firstly, region-based features may be incorporated into the pairwise similarity metric to further improve the accuracy of image retrieval. Secondly, most of existing approaches use the same pairwise similarity metrics for different genres. However, it is observed that low-level features may have different influences for different types of images. To learn an optimal similarity measure for each genre, i.e., the optimal choice of discriminative features, distance metrics, and structure differences, is promising to improve system performance. Finally, there exists certain correlation between image content for images from the same genre. Therefore, modeling multiple genres simultaneously may have better performance than modeling them separately.

References

Akakin, H., & Gurcan, M. (2012). Content-based microscopic image retrieval system for multi-image queries. *IEEE Transactions on Information Technology in Biomedicine, 16*(4), 758–769.

Beecks, C., Assent, I., Seidl, T. (2011). Content-based multimedia retrieval in the presence of unknown user preferences. In *Proceeding of IEEE conference on multimedia modeling* (pp. 140–150).

Bower, R., & Balogh, M. (2004). The difference between clusters and groups: a journey from cluster cores to their outskirts and beyond. In *Carnegie observatories astrophysics series* (pp. 1–20).

Cui, J., & Zhang, C. (2007). Combining stroke-based and selection-based relevance feedback for content-based image retrieval. In *Proceedings of the ACM conference on multimedia* (pp. 329–332).

Gunnarsson, O., & Jones, R. (1980). Density functional calculations for atoms, molecules and clusters. *Physica Scripta, 21*(3–4), 394–401.

He, J., Li, M., Zhang, H., Tong, H., Zhang, C. (2004). Manifold-ranking based image retrieval. In *Proceedings of the ACM conference on multimedia* (pp. 9–16).

He, J., Li, M., Zhang, H., Tong, H., Zhang, C. (2006). Generalized manifold-ranking based image retrieval. *IEEE Transactions on Image Processing, 15*(10), 3170–3177.

Jeon, J., Lavrenko, V., Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proc. Proceedings of the ACM conference on multimedia information retrieval* (pp. 119–126).

Jing, F., Li, M., Zhang, H., Zhang, B. (2002). An effective region-based image retrieval framework. In *Proceedings of the ACM conference on multimedia* (pp. 456–465).

Karypis, G., & Kumar, V. (1999). Parallel multilevel k-way partitioning for irregular graphs. *SIAM Review, 41*(2), 278–300.

Kokare, M., Chatterji, B., Biswas, P. (2003). Comparison of similarity metrics for texture image retrieval. In *Proceedings of the IEEE conference on convergent technologies for Asia-Pacific region* (pp. 571–575).

Li, B., Chang, E., Wu, C. (2002). DPF-a perceptual distance function for image retrieval. In *Proceedings of the IEEE conference on image processing* (pp. 597–600).

Natsev, Naphade, M., Tesic, J. (2005). Learning the semantics of multimedia queries and concepts from a small number of examples. In *Proceedings of the ACM conference on multimedia* (pp. 598–607).

Pass, G. (1997). Comparing images using color coherence vectors. In *Proceedings of the ACM conference on multimedia* (pp. 65–73).

Rubner, Y., Tomasi, C., Guibas, L. (1998). A metric for distributions with applications to image databases. In *Proceedings of the IEEE conference on computer vision* (pp. 59–66).

Stricker, M., & Orengo, M. (1995). Similarity of color images. In *Proceedings of the IEEE conference on storage and retrieval for image and video databases* (pp. 381–392).

Su, J., Huang, W., Yu, P., Tseng, V. (2011). Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE Transactions on Knowledge and Data Engineering, 23*(3), 360–372.

Wang, B., Pan, F., Hu, K., Paul, J. (2012). Manifold-ranking based retrieval using k-regular nearest neighbor graph. *Pattern Recognition, 45*(4), 1569–1577.

Xu, B., Bu, J., Chen, C. (2011). Efficient manifold ranking for image retrieval. In *Proceedings of the ACM SIGIR conference on research and development in information retrieval* (pp. 525–534).

Yuan, X., Hua, X., Wang, M., Wu, X. (2006). Manifold-ranking based video concept detection on large database and feature pool. In *Proceedings of the ACM conference on multimedia* (pp. 623–628).

Zhang, R., & Zhang, Z. (2004). Hidden semantic concept discovery in region based image retrieval. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 996–1001).

Zhang, L., Wang, L., Lin, W. (2012). Generalized biased discriminant analysis for content-based image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, 42*(1), 282–290.