A Scalable Graph-Based Semi-Supervised Ranking System for Content-Based Image Retrieval

Xiaojun Qi, Department of Computer Science, Utah State University, Logan, UT, USA

Ran Chang, Department of Computer Science, Utah State University, Logan, UT, USA

ABSTRACT

The authors propose a scalable graph-based semi-supervised ranking system for image retrieval. This system exploits the synergism between relevance feedback based transductive short-term learning and semantic feature-based long-term learning to improve retrieval performance. Active learning is applied to build a dynamic feedback log to extract semantic features of images. Two-layer manifold graphs are then built in both low-level visual and high-level semantic spaces. One graph is constructed at the first layer using anchor images obtained from the feedback log. Several graphs are constructed at the second layer using images in their respective cluster formed around each anchor image. An asymmetric relevance vector is created for each second layer graph by propagating initial scores from the first layer. These vectors are fused to propagate relevance scores of labeled images to unlabeled images. The authors' extensive experiments demonstrate the proposed system outperforms four manifold-based and five state-of-the-art long-term-based image retrieval systems.

Keywords: *Anchor Images, Content-Based Image Retrieval, Graph-Based Semi-Supervised Ranking System, Long-Term Learning, Semantic Features, Short-Term Learning*

1. INTRODUCTION

With the rapidly growing number of digital images found on the Internet and housed in digital libraries, the need for effective and efficient tools to manage large image databases has grown dramatically. Specifically, the development of efficient image retrieval systems to find images of interest in this haystack of data has become an active research area in recent years (Thomee, 2010).

Content-based image retrieval (CBIR) techniques (Datta *et al.*, 2008; Lew *et al.*, 2006) are viable solutions to find desired images from multimedia databases and have evolved significantly since the early 1990s. They make use of low-level visual image features (e.g., color, texture, shape, etc.) instead of keywords to represent images, where each feature can be automatically and consistently extracted without human intervention. Consequently, they overcome the limitations entailed by text-based

DOI: 10.4018/ijmdem.2013100102

image retrieval, which include the large amount of manual labor required to annotate each image in the database, and the inconsistency among different annotators in perceiving the same image. However, as the ranking of retrievals is calculated based on selected image features, the retrieval accuracy may be unsatisfactory due to the semantic gap between low-level visual features and high-level semantic concepts. This semantic gap exists because objects of the same type do not have the same visual representation. For example, images of similar semantic content may be scattered far away from each other in the feature space, while images of dissimilar semantic content may share similar low-level features. To bridge the semantic gap, a great deal of research work has been focused on developing effective relevance feedback (RF) techniques (Liu *et al.*, 2007; Zhou & Huang, 2003), which utilize users' interaction to learn better representation of images as well as the query concept. RF, as an interactive search technique, has been used in CBIR systems to repeatedly modify the query descriptive information (feature, matching models, metrics or any meta knowledge) as response to the users' feedback on retrieved results. Therefore, it learns the query close to its optimal and returns more user-desired images (i.e., improves the retrieval precision) after each round.

Most existing RF techniques use short-term learning or intra-query learning to find out images that are relevant to the user's query in a retrieval session. Representative short-term learning techniques include query updating (e.g., query reweighting, query shifting, and query expansion) and statistical learning techniques (e.g., inductive learning and transductive learning) (Qi *et al.*, 2011). However, query updating methods (Muneesawang & Guan, 2004) do not fully utilize the information embedded in feedback images and therefore cannot achieve satisfactory retrieval results. Inductive learning methods (Tong & Chang, 2001; Wu & Yap, 2006) yield degraded retrieval results when the chosen classifier is trained with insufficient labeled training samples. Moreover, these two categories of techniques ignore the manifold structure of image features. Therefore, the latest trend has been moving towards RF-based transductive learning, which explores the relationship of all database images in the feature space and propagates ranking scores of labeled images to unlabeled images via a weighted graph. To this end, He *et al.* (2006) propose a generalized manifold-ranking-based image retrieval (gMRBIR) algorithm, which works well for any query image inside or outside the database. Their proposed algorithm represents images and their relationships as a graph and propagates labeled information among images via the graph structure. It further exploits the distribution of unlabeled images to improve the retrieval performance. Wan (2007) proposes to apply the MRBIR algorithm to non-overlapping equal-sized blocks of each database image and fuse the ranking scores of all blocks in the image as the final retrieval score of each image. Cai *et al.* (2007) incorporate a locality preserving regularizer into the manifold structure to learn a classification function in the image manifold. They then apply the user's RFs to update the manifold structure for better classification. Bian and Tao (2010) combine the biased discriminative Euclidean embedding with the manifold regularization-based item to discover a more accurate manifold structure for better classification. Geng *et al.* (2012) propose an ensemble manifold regularization framework to implicitly estimate hyperparameters involved in the regularization to better explore the intrinsic manifold structure of the image database. All above transductive methods achieve better retrieval precision in each iterative step. However, they do not apply users' accumulated historical RF information to improve the manifold graph. They also cannot run on a computer when the number of images in the database reaches a certain level due to the use of several large square matrices. Furthermore, all these short-term learning techniques cannot capture the semantic meaning of an image and therefore cannot achieve satisfactory retrieval results. They also cannot remember users' historical feedback and therefore cannot utilize it in future retrievals.

Recently, long-term learning or inter-query learning extends short-term learning by utilizing the information gathered from the past retrieval sessions to improve the retrieval results in future retrieval sessions. Specifically, these long-term learning techniques first store the accumulated feedback history collected from multiple query sessions in a feedback log. They then aggregate the information in the feedback log into a semantic matrix, relevance matrix, or affinity matrix, which can be further used to discover extra knowledge. Finally, they infer relationships between images by analyzing the transformed matrix and estimate the semantic relevance level of a database image to the current query. Representative long-term learning techniques include retrieval pattern-based and feature vector model-based learning (Fechser *et al.*, 2010). For example, Hoi *et al*. (2006) apply the statistical correlation on the feedback log to analyze the relationship among the current and past retrieval sessions. Han *et al*. (2005) use the memory learning technique to form a knowledge memory model to store the semantic information and learn semantic relations. Yin *et al*. (2008) design a virtual-features-based technique to explore the historical feedback to estimate the semantic relationship between images. Qi *et al.* (2011) enhance the retrieval performance by developing a short-term block-based fuzzy SVM and long-term dynamic semantic clustering technique to adaptively learn and update semantic categories. Xiao *et al.* (2012) propose a short-term and long-term combined learning strategy by applying users' RF-related semantic knowledge to create dynamic semantic features for database images. Hence, the high-level semantic similarity can benefit from these updated semantic features to boost the retrieval accuracy. Although these long-term RF learning techniques achieve impressive retrieval performance, they usually require a large matrix to store historical RF information. Furthermore, these long-term RF learning techniques use the piecewise distance calculation and oversimplify the relationship among all images in the database. In other words, they do not explore the relationship of all database images in the feature space. Therefore, their effectiveness is limited.

To address the aforementioned shortcomings of the short-term learning and long-term learning techniques, Chang and Qi (2011) and Chang *et al.* (2012) create semantic clusters based on users' historical RF to group semantically similar images. They then construct a weighted semantic clusters-based manifold structure to represent image relationship in low-level visual feature space (Chang & Qi, 2011) and both low-level visual and high-level semantic feature spaces (Chang *et al.*, 2012) for better retrieval performance. However, these two learning techniques cannot be directly applied to a large scale CBIR system due to the use of several large square matrices whose size equals to the square of the number of database images. To make the system scalable, Chang and Qi (2013) propose a novel hierarchical manifold ranking system which constructs a two-layer intrinsic weighted structure using the visual space at the first layer and the visual and semantic spaces at the second layer. The relevance scores of labeled images are propagated to unlabeled images via this hierarchical manifold. However, a relatively large matrix is used to store semantic features of each database image. The size of this matrix equals to the number of database images multiplying by the number of training queries (e.g., 10% of the number of database images). As the size of the image database grows, the size of this matrix grows too. Eventually, computer may not have enough memory space to store this matrix. So the scalability issue still presents in all existing long-term manifold-based CBIR systems.

In this paper, we propose a novel scalable graph-based semi-supervised ranking system for CBIR. This proposed system extends the short-term learning by utilizing the RF information gathered from the past retrieval sessions to hierarchically explore the relationship of all database images in both low-level visual feature space and high-level semantic feature space. It treats both labeled and unlabeled im-

ages as vertices in their respective graph and builds pairwise edges between these vertices, which are weighted by both visual and semantic affinities between the corresponding image pairs. The small portion of vertices carrying seed labels (e.g., the users' RF information) are then harnessed via information propagation to predict the labels of the unlabeled vertices (images). Positively predicted images are finally returned as the retrieval results. Specifically, the proposed system first learns semantic features of each database by using the users' historical RF. It then builds a two-layer manifold graph ranking system which models the intrinsic structure for the image space in several manageable small scales. The first layer manifold graph ranking system is constructed using both low-level visual similarity and high-level semantic similarity of the anchor images in the database. These anchor images are chosen based on the users' RF. They normally contain key semantic concepts of the image database. The number of anchor images approximately corresponds to the number of semantic concepts contained in all images in a database. The second layer manifold graph ranking system is constructed based on the clusters formed around anchor images. For each cluster, both low-level visual and high-level semantic similarities of the images in the cluster are integrated to construct its manifold graph to achieve a more meaningful structure in the image space. The size of these graphs is significantly smaller compared to the size of the traditional manifold graph, which makes the proposed system scalable. Finally, an asymmetric relevance vector is created for each second layer graph by assigning initial scores from the first layer graph. This vector then propagates the relevance scores of labeled images to unlabeled images via the hierarchical graph-based structure. In the proposed RF-based CBIR system, the training and retrieval processes have the following advantages over other common RF-based CBIR systems:

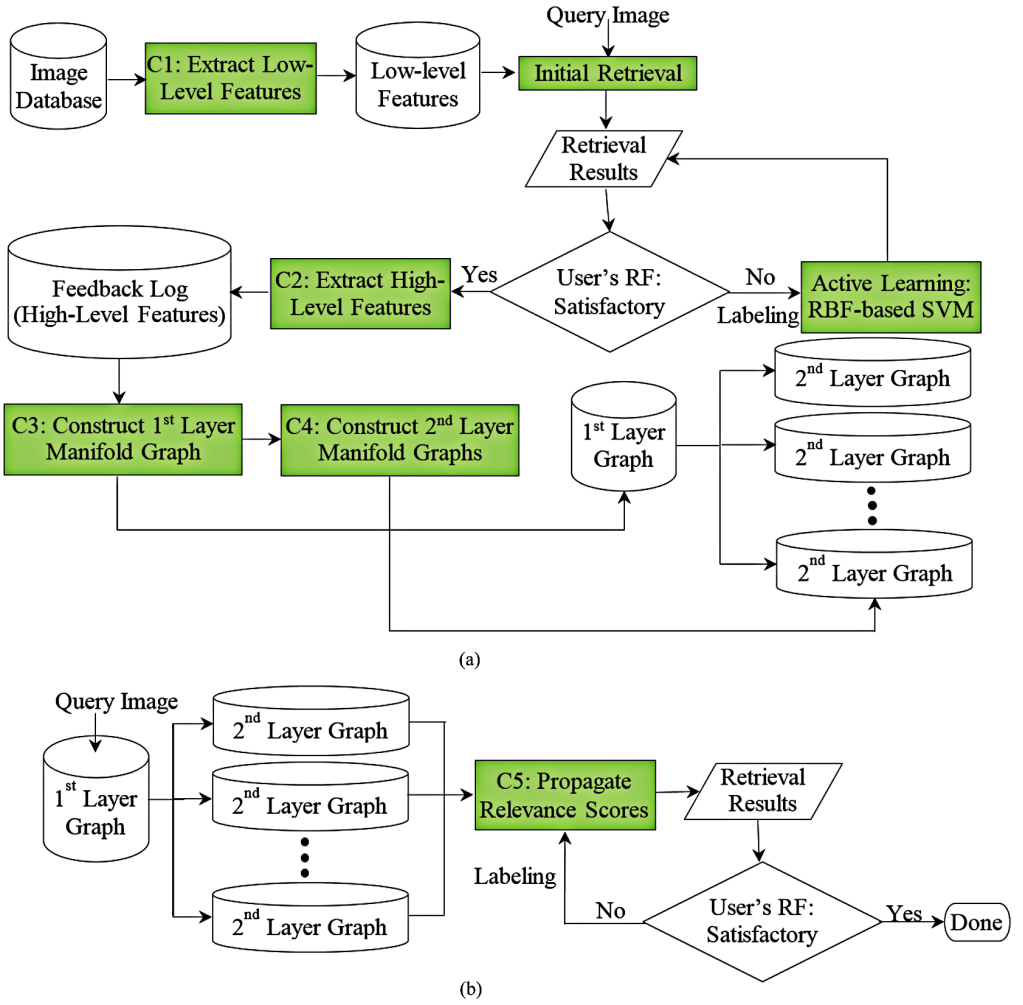
- Quick construction of a compact dynamic feedback log to store unique retrieval patterns (i.e., the similarity of relevant and irrelevant images) of historical query sessions;
- Efficient merging of similar retrieval patterns to maintain a reasonable number of meaningful semantic concepts to represent all images in a database;
- Effective composition of low-level visual and high-level semantic similarity measure to build the manifold graph, which explores the intrinsic structure of images in both low-level visual feature space and high-level semantic feature space;
- Effective layered design of the relevance vector to propagate the relevance scores from anchor images to the second layer graphs and further propagate the relevance scores of labeled images to unlabeled image via the hierarchical graph-based structure.

The rest of the paper is organized as follows: Section 2 presents the proposed scalable graph-based semi-supervised ranking system. Section 3 compares the proposed CBIR system and its variant systems with four manifold-based CBIR systems and five representative long-term-based CBIR systems on five databases. Section 4 draws conclusions and presents future directions.

2. PROPOSED SCALABLE GRAPH-BASED SEMI-SUPERVISED RANKING SYSTEM

Our proposed scalable graph-based semi-supervised ranking approach consists of offline training and online retrieval phases, which are demonstrated in Figure 1. One of the aims of offline training is to collect users' historical RF to learn semantic features of each database image. Specifically, SVM active learning is first applied to select the most informative unlabeled positive images based on the decision boundary learned from user's positively and negatively labeled images. The relevancy information for each retrieved image in each query session is stored in a dynamic feedback log, which is

Figure 1. Block diagram of the proposed system: (a) offline training phase; (b) online retrieval phase



updated after each query session. To this end, the relevancy information of the current query may be iteratively merged with the relevancy information of past query sessions if they contain sufficient overlapping information. The final merged relevancy information is then used to update the feedback log. On the other hand, the relevancy information of the current query may contain unique information which is not present in the past query sessions. This new relevancy information is then appended to the feedback log. After all query sessions have been performed, this dynamic feedback log holds the semantic information of each database image. Another aim of offline training is to build a scalable graph-based semi-supervised ranking system for future retrievals. To this end, anchor images are first located based on the feedback log. A cluster is formed around each anchor image and any image outside of any cluster is assigned to an appropriate cluster using the minimum-distance-based strategy. In this way,

the image database is divided into several clusters (categories). Finally, the first layer manifold graph is constructed by incorporating low-level visual and high-level semantic features of all anchor images. Several second layer manifold graphs are also constructed by using low-level visual and high-level semantic features of all images in their respective clusters. The online retrieval phase focuses on designing a strategy to asymmetrically propagate the relevance scores of labeled positive and negative images through the hierarchical manifold graphs. Specifically, the first layer graph is capable of quickly identifying the potential clusters that a query image belong to and propagating its relevance scores to the second layer graphs. The final relevance scores can then be propagated to unlabeled images via the hierarchical manifold structure. In the following, we explain the major components of each phase in detail.

2.1. Extraction of Low-Level Visual Features

All three important features, e.g., color, edge, and texture features, are utilized to represent each image in the database. The proposed system uses a 100-dimensional vector to represent low-level features of an image. These global features were proven to be effective in (Qi & Chang, 2007) and they are easy to compute and complementary to each other. Specifically, the 100-dimensional vector includes a 64-bin (8×2×4) HSV-based color histogram and a 36-dimensional complementary feature vector (Hoi *et al.*, 2006), which contains 9 color, 18 edge, and 9 texture components. To this end, it computes the first three moments in each HSV color channel to represent color features; computes the 18-bin edge direction histogram to represent the edge features in the grayscale image; and computes the entropy of each of nine detail subbands of a 3-level wavelet transform to represent texture features in the grayscale image.

For initial retrieval, the Euclidean distance is used on the normalized low-level features to measure the similarity between query and database images.

2.2. Extraction of High-Level Semantic Features

Each image is also represented by semantic features, which are learned from the users' historical RF. Since semantic features are closely related to the high-level semantics of an image, we also call semantic features as high-level semantic features. The more images in the database, the more possible semantic concepts are. So we fix the maximal length of high-level semantic features (e.g., maximal semantic concepts) to be linked with the total number of images in the database. In our system, we initially confine the maximal length of high-level semantic features for a database image to be 10% of the total number of images in the database, which is a reasonable and conservative estimate for the maximal number of semantic concepts contained in all images in a database. These high-level semantic features are directly constructed from the dynamic feedback log $\boldsymbol{R}$. This $\boldsymbol{R}$ is a 2-D matrix whose row number equals to the total number of images in the database (e.g., N) and column number starts with 0 and is updated with each training query. It should be noted that the number of training queries equals to the confined maximal length of high-level semantic features (i.e., $10\% \times N$) so sufficient learning can be achieved to learn the semantic features of each image.

The system randomly selects 10% of database images as query images to perform the training task. For each query, the system initially calculates its Euclidean distance to all database images and returns top v images that have the smallest distances to the query. The user marks the retrieved images that are similar to the query as relevant (positive) images. The unmarked images are automatically considered by the system as irrelevant (negative) images. In the remaining RF iterations, the system applies active learning, a Gaussian radial basis func-

tion (RBF) SVM classifier, on the accumulated positive and negative images to find a better classification boundary to discriminate positive images from negative images. It then returns top v images that have the largest positive distances to the classification boundary for another round of user's feedback. This process continues for a few iterations till the maximum iteration is reached. A query session is completed at this time as well. In order to collect a sufficient number of unique positive and negative images, the system ensures that a retrieved image is not returned in later iterations. It also specifies v as 25 so all retrieved images can be easily fit into one screen for users to provide their feedback information. The number of iterations is set to be 4 to keep the users' labeling time in a query session to be the minimum.

After each query session, the system creates a candidate column with all 0's. It then marks the cells corresponding to the rows of positive images as 1's and marks the cells corresponding to the rows of negative images as -1's. A merging technique is then carried out to iteratively combine this candidate column with other similar columns in $\boldsymbol{R}$. If no columns in $\boldsymbol{R}$ are similar to this candidate column, the merging operation does not take place and the candidate column is added as the last column in $\boldsymbol{R}$. The basic idea of this merging strategy is as follows:

1. The candidate column is sequentially compared with each column in $\boldsymbol{R}$;
2. If the candidate column is similar to an existing column in $\boldsymbol{R}$, these two columns are combined to form a new candidate column by performing an addition operation;
3. This newly merged candidate column is continuously examined against the remaining columns in $\boldsymbol{R}$ until there is no merging operation occurs.

The algorithmic view of comparing the similarity of two columns c_1 and c_2 and performing the merging operation is as follows:

1. Put the IDs of cells of 1's in c_1 into a set A;
2. Put the IDs of cells of 1's in c_2 into a set B;
3. If |intersect (A, B)| > 0.5×min(|A|, |B|)
   Merge c_1 and c_2 by $c_1 + c_2$
   Endif

All the columns that have been iteratively merged with the candidate column are deleted from $\boldsymbol{R}$. The final merged candidate column is then added as the last column in $\boldsymbol{R}$.

After performing the query session for all the training images, $\boldsymbol{R}$ holds the possible semantic information for each database image. The column number of $\boldsymbol{R}$ equals to the number of learned semantic concepts (e.g., foreground objects or background implicitly marked by the users as a set of relevant images in the RF step). High-level semantic features of an image correspond to the respective row in $\boldsymbol{R}$. Each value in $\boldsymbol{R}$ represents the relationship between a database image and the semantic concept corresponding to the respective queries encapsulated in the corresponding column. For example, the first row in $\boldsymbol{R}$ represents semantic features of the first database image. If the first column in $\boldsymbol{R}$ represents the semantic concept (e.g., sky and mountain) of the first (merged) query, the value at (1, 1) in $\boldsymbol{R}$ means the relevance of the first database image to the sky and mountain concepts. A larger positive value indicates the database image likely to possess the corresponding semantic concept. A smaller negative value indicates the database image unlikely to possess the corresponding semantic concept.

The high-level semantic relevance relation $S_{i,j}$ between images i and j is computed by the semantic-correlation-based distance:

$$S_{i,j} = HSF_i \cdot HSF_j = \sum_{k=1}^{p} HSF_i(k) \times HSF_j(k) \quad (1)$$

where HSF_i and HSF_j respectively represent semantic features of images i and j, $HSF_i(k)$ and $HSF_j(k)$ respectively are the k^{th} element of semantic features of images i and j, p is the

dimensionality of semantic features of each image (i.e., the number of columns in $\boldsymbol{R}$), and the × operation is defined as shown in Box 1.

This operation yields positive results when $HSF_i(k)$ and $HSF_j(k)$ are positive values (i.e., both images have the k^{th} semantic meaning represented in the k^{th} column of $\boldsymbol{R}$), yields negative results when $HSF_i(k)$ and $HSF_j(k)$ have different signs (i.e., one image has the k^{th} semantic meaning and the other image does not have the k^{th} semantic meaning), and yields 0's otherwise (i.e., both images do not have the k^{th} semantic meaning or no semantic meaning is learned for either of the two images or both).

2.3. Construction of the First Layer Manifold Graph

The first layer manifold graph is constructed from $\boldsymbol{R}$. Suppose that there are p columns in $\boldsymbol{R}$ after performing the query session for all training images, the proposed system sequentially investigates each of p columns to find its anchor image. To this end, it first records the IDs of the images that have positive values in the corresponding column. It then computes the centroid of these images (i.e., the average of their low-level visual features) and finds the image in the recorded set that has the closest distance to this centroid. The found image is considered as the anchor image for the respective column (i.e., the representative image for the respective semantic concepts). The other images in the recorded set are considered as the members for the respective column. They share similar semantic meanings as their anchor image. In total, there are p anchor images. These anchor images contain key semantic concepts of the image database, which are learned from the users' historical RF.

The system constructs the first layer manifold graph using p anchor images. It builds a $p \times p$ affinity matrix, in which each element represents the relationship between each pair of anchor images. The constructed first layer graph is capable of spreading relevance scores of the query to all anchor images. The algorithmic view of constructing the first layer manifold graph is summarized in Figure 2.

In step 2, two popular Minkowski distances, e.g., the Manhattan (L_1) distance and the Euclidean (L_2) distance, can be used to compute each element in $GW_{i,j}$. If the L_1 distance is

Box 1.

$$HSF_i(k) \times HSF_j(k) = \begin{cases} HSF_i(k) \times HSF_j(k) & \text{if } HSF_i(k) > 0, HSF_j(k) > 0 \text{ or } HSF_i(k) \times HSF_j(k) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Figure 2. The algorithmic view of constructing the first layer manifold graph

1. Initialize the first layer manifold graph FMG and three intermediate graphs (e.g., GW, GD, and GN) as all 0's. The sizes of these four graphs are all $p \times p$.
2. For each pair of anchor images A_i and A_j, $1 \le i \le p$, $1 \le j \le p$, $i \ne j$, compute their distance by using the respective low-level visual features and high-level semantic features. The computed distance is stored in the i^{th} row and j^{th} column of GW (e.g., $GW_{i,j}$).
3. Update the diagonal element in GD as the sum of all elements in its corresponding row in GW. That is, $GD_{i,i} = \sum_{k=1}^{p} GW_{i,k}$
4. Update each element in GN by symmetrically normalizing GW. That is, $GN = GD^{-1/2} \times GW \times GD^{-1/2}$.
5. Update each element in FMG by computing $(1-\alpha \times GN)^{-1}$, where α is a parameter in [0, 1).

employed, $GW_{i,j}$ is computed by the Laplacian kernel-based 100-dimensional low-level visual and high-level semantic features:

$$GW_{ij} = \prod_{l=1}^{100} \exp\left(-\frac{\left|lvf_{il} - lvf_{jl}\right|}{\sigma_L}\right) \times \exp\left(-\frac{1 - NS_{i,j}}{\sigma_H}\right) \quad (3)$$

where lvf_i and lvf_j are respectively normalized low-level visual features of two anchor images A_i and A_j, lvf_{il} and lvf_{jl} are respectively the l^{th} element of normalized low-level visual features lvf_i and lvf_j, σ_L is a positive parameter reflecting the standard deviation of the low-level visual similarity, $NS_{i,j}$ is the normalized high-level semantic relevance relation between A_i and A_j, and σ_H is a positive parameter reflecting the standard deviation of the high-level semantic similarity. If the L_2 distance is employed, $GW_{i,j}$ is computed as shown in Box 2.

The equation $d(lvf_i, lvf_j)$ represents the Euclidean distance between normalized low-level features of A_i and A_j, σ is a positive parameter reflecting the standard deviation of the low-level visual and high-level semantic similarity, and w_h is the contribution factor of high-level semantic features.

2.4. Construction of the Second Layer Manifold Graphs

The second layer manifold graphs are constructed from the clusters around anchor images. For each of p anchor images, the system forms a cluster around it and constructs a second layer manifold graph. As a result, there are p second layer manifold graphs in total.

Each anchor image and its associated positively labeled images form the initial cluster. Other database images that are not retrieved from the system or are negatively labeled in all query sessions are assigned to their appropriate cluster using the minimum-distance-based strategy. We denote the set of these other database images as *UnassignedSet* and the set of images in p clusters as *AssignedSet*. For each image *Imx* in *UnassignedSet*, the system computes its distances to all images in *AssignedSet* and finds the image *Imy* in *AssignedSet* that has the closest distance to *Imx*. The system then assigns *Imx* to the same cluster as *Imy*. In this way, all images in *UnassignedSet* are assigned to exactly one cluster. Each of p clusters contains positively labeled images and some images in *UnassignedSet*. Each image has its own anchor image, which represents the characteristic semantic concepts of the cluster. In this way, the images with the same anchor image are considered to be in the same cluster since they assume to share similar semantic concepts as the anchor image. The system then uses a vector *AnchorVec* to store the ID of the anchor image for each database image so that the cluster related information can be easily acquired. The number of clusters equals to the number of anchor images or the number of columns in R, where each column is obtained by the merging strategy explained in Section 2.2.

Suppose a cluster k contains n_k images including positively labeled images and some images in *UnassignedSet*. The system builds an $n_k \times n_k$ affinity matrix as the second layer manifold graph SMG_k for cluster k. The con-

Box 2.

$$GW_{ij} = \exp\left(-\frac{\left[\left(1 - w_h\right) \times d\left(lvf_i, lvf_j\right) + w_h \times \left(1 - NS_{i,j}\right)\right]^2}{2\sigma^2}\right) \quad (4)$$

struction of each second layer manifold graph SMG_k, $1 \le k \le p$, follows the same five steps as summarized in Figure 2 with two exceptions: 1) The size of SMG_k is $n_k \times n_k$. 2) Each element in graph represents the relationship between each pair of n_k images in cluster k.

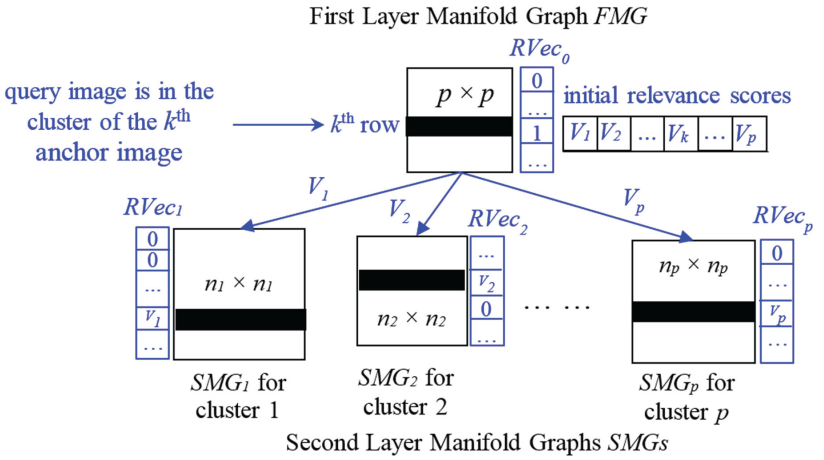
Figure 3 shows the structure of the proposed scalable graph-based semi-supervised ranking system generated at the end of the training phase. In the first layer, one manifold graph *FMG* is constructed using p anchor images, where p (i.e., $p << N$) is the number of columns in $\boldsymbol{R}$ and is also the number of clusters. In the second layer, there are p manifold graphs SMG_k, $1 \le k \le p$. Each graph is constructed using all member images in its respective cluster. For example, if there are n_1 images in the first cluster, the size of the corresponding graph SMG_1 is $n_1 \times n_1$. The sum of all images in p manifold graphs at the second layer is N. Here, $n_i << N$ $(1 \le i \le p)$. It is clear that the size of each of $p + 1$ graphs at both the first layer and the second layer is significantly smaller than the size of the traditional manifold graph, which equals to $N \times N$. As a result, the need for a computer to allocate several large consecutive $N \times N$ memory spaces to store the graph is eliminated. It should be noted that a computer runs out of memory or swap space to satisfy such a need and therefore the proposed scalable manifold graphs can be employed for a large scale image database as long as each graph does not exceed the memory capability of the running machine.

2.5. Propagation of the Relevance Scores

Propagating relevance scores of positive and negative images collected in iterative RF steps to unlabeled images via scalable graphs is the major focus of the online retrieval phase. These propagated relevance scores are also used as similarity scores between query and database images.

Since one *FMG* is constructed to represent the relationship between anchor images and p *SMGs* are constructed to represent the relationship between images in their corresponding clusters, p+1 relevance vectors, i.e., $RVec_i$, $0 \le i \le p$, are used to propagate the relevance scores among images in their respective graphs. Here, $RVec_0$ denotes the relevance vector for *FMG* and $RVec_i$ $(1 \le i \le p)$ denotes the relevance vector for SMG_i. Initially, the system sets all relevance vectors as all 0's. That is, $RVec_i = 0$ for $0 \le i \le p$.

Figure 3. The structure of the proposed scalable graph-based semi-supervised ranking system and illustration of the layered design of the relevance vectors together with their initialization



For each submitted query image, the system first locates its anchor image from *AnchorVec*. If the index of the query's anchor image in *FMG* is k, the system then sets the k^{th} element of $RVec_0$ as 1's. That is, $RVec_{0,k} = 1$. Next, the system propagates $RVec_0$ through FMG (i.e., $\left[FMG\right]_{p\times p}\times\left[RVec_0\right]_{p\times 1}$) to obtain the relevance score of each anchor image to the query. These relevance scores correspond to the values in the k^{th} row (or the k^{th} column) of *FMG*. The system then propagates each value in the k^{th} row (e.g., $V_i (1 \le i \le p)$) as the initial relevance score for its corresponding second layer manifold graph $SMG_i (1 \le i \le p)$. Specifically, the system sets $RVec_{i,m}$ as V_i, where $1 \le i \le p$ and m is the index of the anchor image in its respective graph SMG_i. For example, V_1 is put at the row of the anchor image of cluster 1 in $RVec_1$ and V_2 is put at the row of the anchor image of cluster 2 in $RVec_2$, etc. Finally, the system performs one more operation if the query image is a positive image in a cluster k ($1 \le k \le p$). To this end, the system finds respective rows of all positive images in cluster k and then puts V_k at these same rows in $RVec_k$. This layered design of the relevance vectors together with their initialization is also demonstrated in Figure 3 in blue color.

After initializing all p relevance vectors, the relevance score of each image is determined by propagating $RVec_i$ through each SMG_i. A relevance score vector T_i for SMG_i is computed by:

$$T_i = [t_{ij}]_{n_i\times 1} = \left[SMG_i\right]_{n_i\times n_i}\times\left[RVec_i\right]_{n_i\times 1} \quad (5)$$

where SMG_i is the i^{th} second layer manifold graph whose size is $n_i \times n_i$ and $RVec_i$ is its initialized relevance vector. The system finally concatenates all relevance score vectors T_i's computed from the second layer manifold graphs into a long relevance score vector T with a length of N. It then returns v images with the highest relevance scores in T.

Based on the user's RF information on v returned images, the system first finds the anchor images for all labeled images and their respective clusters. It then updates the relevance vectors of these pertinent clusters using the following strategies:

1. For positive images, set the corresponding cells in their relevance vector as 1's;
2. For negative images, set the corresponding cells in their relevance vector as -0.25.

This assignment is empirically determined to be optimal and ensures that the propagation on the negative images is not dominated since negative images do not provide sufficient information as the positive images. The system continues to use updated relevance vectors to compute the relevance scores in T to propagate these relevance scores to unlabeled images and return top v images for the user to label. This process iterates several times until the user is satisfied with retrieval results.

It should be mentioned that the following rules should be employed to update a value in the relevance vector: 1) The cells corresponding to positively labeled images are assigned positive values; 2) The cells corresponding to negatively labeled images are assigned negative values; 3) The magnitude of the values assigned to positively labeled images should be larger than the magnitude of the values assigned to negatively labeled images. Our experimental results show that setting positive image cells as 1's and negative image cells as -0.25's achieves the optimal retrieval performance with the minimal computational cost.

The error resulted from the first layer manifold graph usually can be corrected based on users' RF information. Since this kind of error comes from the possibly inaccurate cluster assignment of the query image, the user's correct RF makes the system have a higher chance to select potentially correct clusters. In addition, when the query image is assigned to

the appropriate cluster, the initial score is only assigned to the query itself in the relevance vector. Therefore, the possible error will not affect other images in the same cluster and the error propagation is prohibited. In other words, clusters with more positively labeled images are likely to be returned in the next iteration based on the asymmetric propagation of positively and negatively labeled images.

3. EXPERIMENTAL RESULTS

We conduct a set of carefully designed experiments to evaluate the performance of the proposed scalable graph-based semi-supervised ranking system on five image databases. In Section 3.1, we explain these five image databases. In Section 3.2, we evaluate the effectiveness of the proposed CBIR system by comparing with seven variant systems on the benchmark database. In Section 3.3, we evaluate the performance of the proposed CBIR system together with four manifold-based ranking systems, five state-of-the-art long-term-based CBIR systems, and several representative variant systems on five image databases. In section 3.4, we evaluate the complexity and the storage effectiveness of the compared long-term-based CBIR systems.

3.1. Five Image Databases

To simplify the retrieval process and reduce the burden of soliciting user's labeling, we manually organize database images into several semantic classes. As a result, the image relevance can be automatically determined by checking whether returned images belong to the same manually defined class as the query. It should be noted that this ground truth is exclusively used to evaluate the retrieval performance during each iterative RF process and is not assumed to provide additional class-related information for the proposed system. Thus, the proposed technique can be directly applied in any new unorganized database. We collect the following images to evaluate retrieval performance:

- **6000 COREL images:** We select 60 distinct categories from the COREL database. Each category contains 100 images covering various real-world scenes;
- **2000 Flickr images:** We download a large collection of images from the social photography site http://www.flickr.com. Flickr's API is used to download top 200 images (based on relevance) for each of the chosen 20 categories. We then manually choose 100 images that best represent the category;
- **4000 online images:** We download another set of images from http://images.google.com and http://picasa.google.com through their APIs. Similarly to Flickr images, we download top 200 images for each of 40 distinct keywords and manually pick the most appropriate 100 images for each keyword;
- **22000 NUS-WIDE images:** We download a set of real-world web images from National University of Singapore (Chua *et al.*, 2009). We randomly choose 100 images from each of 81 concepts, which are used for annotation evaluation. We then choose 100 images from each of additional 139 concepts, which contain a sufficient number of images.

Three graduate students are asked to check the appropriateness for each image in its semantic class based on the majority of the agreement. The inappropriate images are replaced by appropriate images approved by at least two graduate students. We then build five image databases as follows: 1) the 6000-image database containing COREL images; 2) the 2000-image database containing Flickr images; 3) the 8000-image database containing 6000 COREL images and 2000 Flickr images; 4) the 12000-image database containing 6000 COREL, 2000 Flickr, and 4000 online images; 5) the 22000-image database containing NUS-WIDE images. Each image in the database is represented by a 100-dimensional low-level

visual feature vector and a high-level semantic feature vector, whose dimensionality is known after the training phase.

3.2. Effectiveness Evaluation

To simulate the practical retrieval process of online users, we randomly generate a sequence of query images to conduct various experiments. At each query session, the proposed CBIR system refines its retrievals by taking advantages of both RF-based transductive short-term learning and semantic feature-based long-term learning techniques and exploiting the synergism between them for several iterations. We use the retrieval precision (RP), which is defined as the ratio of the number of relevant images retrieved to the total number of irrelevant and relevant images retrieved, as the performance measure. We then compute the mean RP (MRP) of the chosen sequence of query images as the final performance measure to evaluate the overall retrieval performance for a large set of query images. The MRP is defined as the total of RP of all query images divided by the total number of queries. In each experiment, we perform four iterations of RF with the top 25 images returned in each iterative step.

In the proposed system, we incorporate L_2-based low-level visual similarity and high-level semantic similarity into both the first layer manifold graph and the second layer manifold graphs to build the scalable graph-based semi-supervised ranking system. The positive parameter σ_L and σ_H in Eq. (3) are respectively set to be 0.05, the positive parameter σ and w_h in Eq. (4) are respectively set to be 0.05 and 0.5, the convergence rate α of the affinity matrix is set to be 0.99, and the parameter γ in the RBF kernel is set to be 0.5. These values are empirically chosen to achieve the optimal retrieval performance.

To evaluate the effectiveness of the proposed system, we implement its three L_2-based variants:

Variant 1: The CBIR system that incorporates L_2-based low-level visual and high-level semantic similarities into the first layer graph and L_2-based low-level visual similarity into the second layer graphs;

Variant 2: The CBIR system that incorporates L_2-based low-level visual similarity into the first layer graph and L_2-based low-level visual and high-level semantic similarities into the second layer graphs;

Variant 3: The CBIR system that incorporates L_2-based low-level visual similarity into both first layer and second layer graphs.

Similarly, we implement four L_1-based counterpart variant systems:

Variant 4: The CBIR system that incorporates L_1-based low-level visual and high-level semantic similarities into both first layer and second layer graphs;

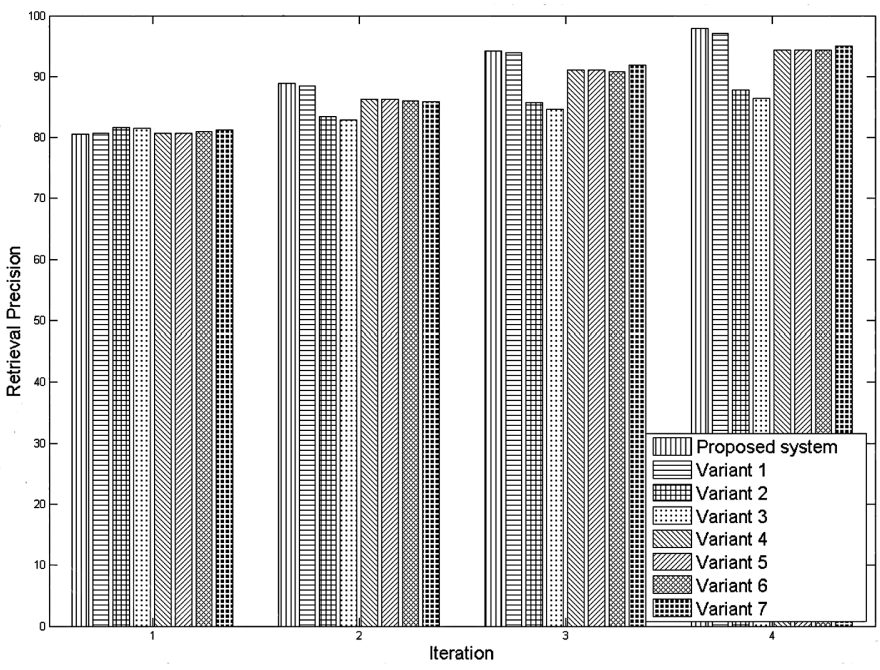
Variant 5: The CBIR system that incorporates L_1-based low-level visual and high-level semantic similarities into the first layer graph and L_1-based low-level visual similarity into the second layer graphs;

Variant 6: The CBIR system that incorporates L_1-based low-level visual similarity into the first layer graph and L_1-based low-level visual and high-level semantic similarities into the second layer graphs;

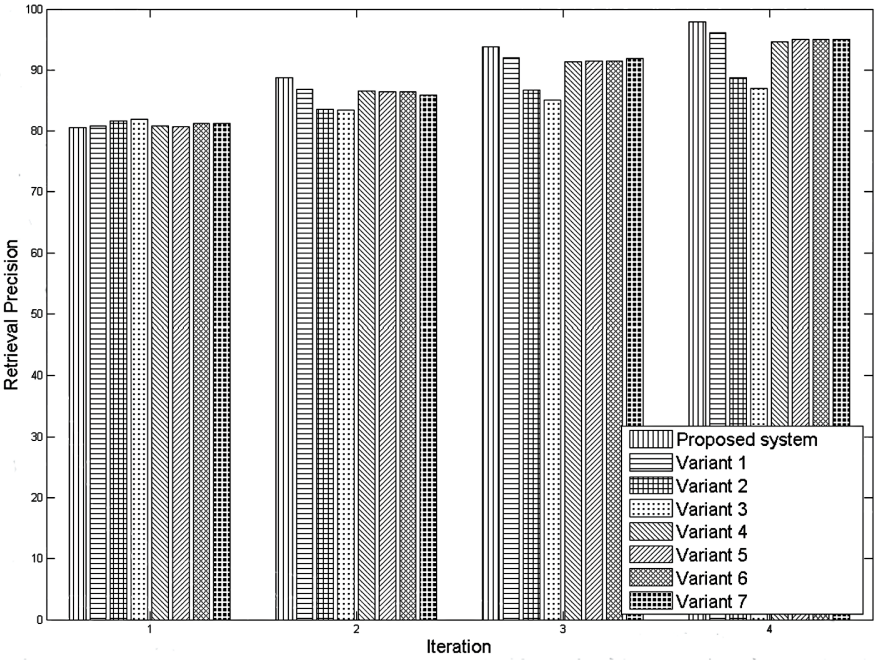
Variant 7: The CBIR system that incorporates L_1-based low-level visual similarity into both first layer and second layer graphs.

Figure 4(a) compares the retrieval performance of the proposed system and its seven variant systems on the COREL benchmark database. It clearly shows that the proposed system achieves the best MRP of 88.83% at iteration 2, 94.23% at iteration 3, and 97.89% at iteration 4. At the last iteration, the proposed system improves the second-best system (variant 1) by 0.9%, the third-best system (variant 7) by 3.04%, and the worst system (variant 3) by 13.35%. It clearly demonstrates the effectiveness to include both L_2-based visual and semantic similarities in the first layer graph since the proposed system and its variant 1

Figure 4. Comparison of the proposed system and its seven variant systems, which are built from (a) the compact feedback log (b) the full feedback log



(a)



(b)

achieve the best MRP. However, four L_1-based variant systems interestingly achieve comparable retrieval performance regardless of the incorporation of the semantic similarity. These four L_1-based variant systems outperform variant 2 and variant 3. These L_1-based results are consistent with the experimental results of Pass (1997) and Chang and Qi (2011), where L_1 distance outperforms other distances on color images. On the other hand, these L_1-based results indicate that incorporating semantic similarity does not significantly improve the retrieval performance since the semantic similarity is computed by the correlation measure, which cannot be evaluated on a dimensionality basis. In other words, the semantic similarity is evaluated by one value and the visual similarity is evaluated by multiple values (e.g., 100 values) as shown in Eq. (3), which significantly reduces the effect of the semantic similarity.

To further prove the effectiveness of the construction of the compact feedback log to extract high-level semantic features, we implement eight respective systems which are built on the full feedback log without applying any merging operations. We call these eight systems as full feedback log based systems. Figure 4(b) compares the retrieval performance of these eight full feedback log based systems on the COREL benchmark database. It clearly shows that these eight systems demonstrate the same retrieval performance as their counterpart systems built from the compact feedback log. In addition, the proposed system and its seven variant systems achieve similar average MRP as their eight counterpart systems built on the full feedback log. As a result, we claim that the simple merging strategy works well to reformulate the users' historical feedback in a compact feedback log and extract representative semantic concepts of the image database.

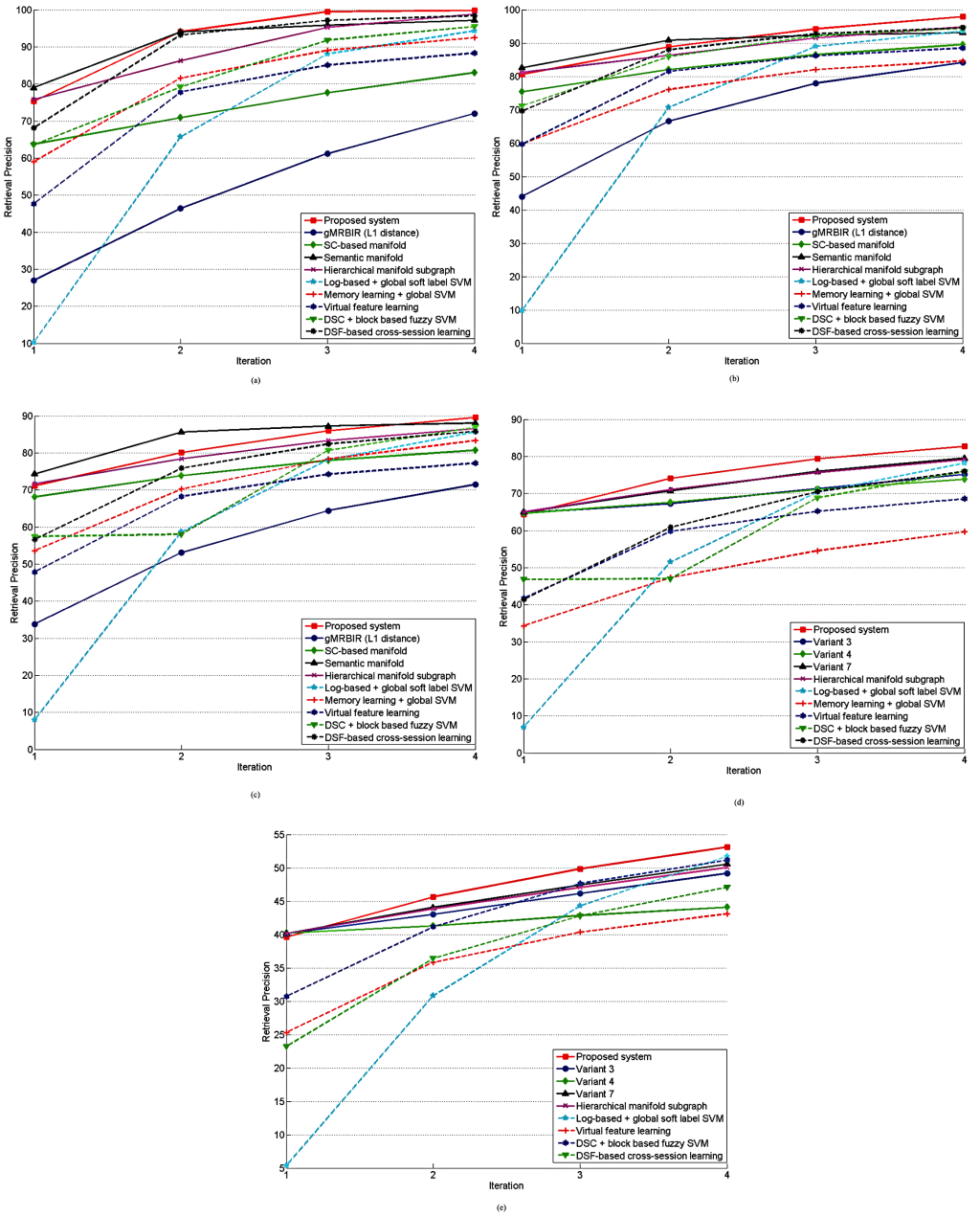
3.3. Performance Evaluation

We compare the proposed system and its representative variants with nine state-of-the-art long-term-based CBIR systems on five image databases. These compared systems can be categorized into two groups:

- **Manifold-based long-term learning systems:** L_1-distance based gMRBIR (He *et al.*, 2006), semantic clusters based manifold ranking system (i.e., SC-based manifold) (Chang & Qi, 2011), weighted semantic manifold ranking system (i.e., Semantic manifold) (Chang *et al.*, 2012), and hierarchical manifold subgraphs ranking system (i.e., Hierarchical manifold subgraph) (Chang & Qi, 2013);
- **Other long-term learning systems:** Log-based system (i.e., Log-based + global soft label SVM) (Hoi *et al.*, 2006), memory learning system (i.e., Memory learning + global SVM) (Han *et al.*, 2005), virtual feature-based system (i.e., Virtual feature learning) (Yin *et al.*, 2008), dynamic semantic clustering system (i.e., DSC + block based fuzzy SVM) (Qi *et al.*, 2011), dynamic semantic feature-based long-term cross-session learning system (i.e., DSF-based cross-session learning) (Xiao *et al.*, 2012).

Figure 5 shows the retrieval performance of the compared systems at each of four iterations on five image databases, where manifold-based systems are shown in solid lines and other long-term systems are shown in dashed lines. For the three smaller databases, the proposed system and all the aforementioned systems are included in the comparison. For the 12000-image database, gMRBIR, SC-based manifold, and semantic manifold systems cannot run on a computer due to its requirement of several matrices of 12000×12000. For the 22000-image database, the same three manifold-based systems and memory learning system cannot run on a computer due to its requirement of several matrices of 24000×24000. As a result, these systems are not included in the comparison for either 12000-image or 22000-image or both databases. Instead, we include our variant

Figure 5. Comparison of ten state-of-the-art long-term-based CBIR systems. Retrieval results on (a) 2000-Flickr database; (b) 6000-COREL database; (c) 8000-image database; (d) 12000-image database; and (e) 22000-image database.



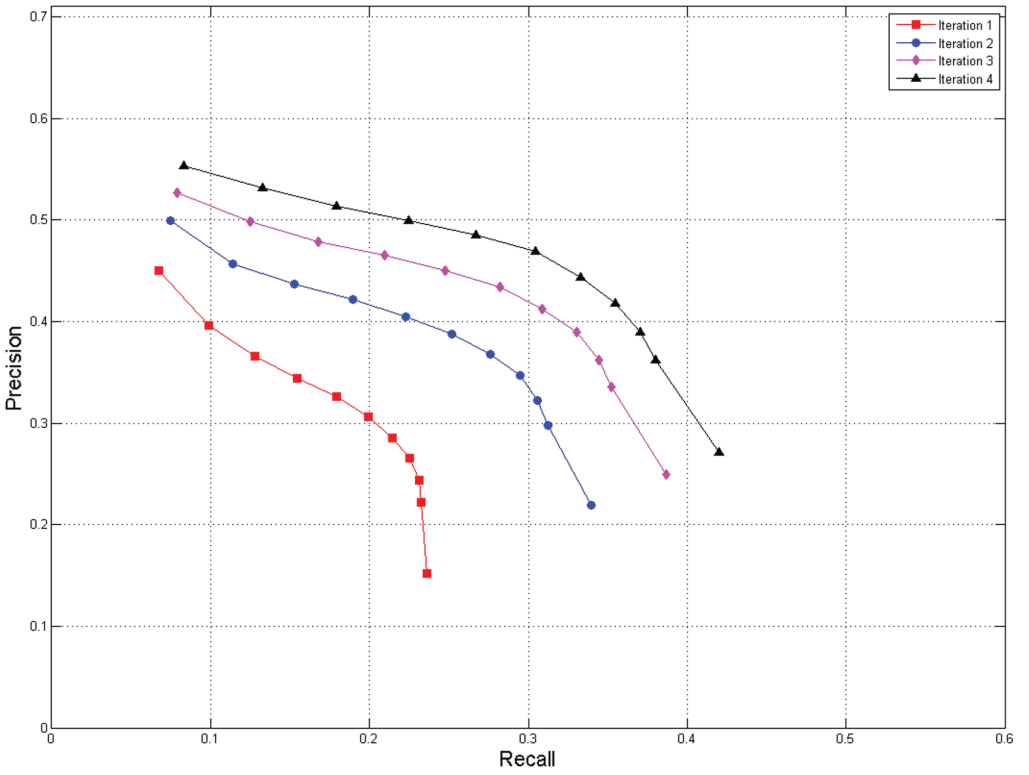
4, variant 5, and variant 7 systems in the comparison for these two larger databases. Figure 5 clearly shows that the proposed system consistently outperforms the other systems for all five databases. Specifically, at the last iteration, it achieves the best MRP of 99.84%, 97.89%,

89.57%, 82.69%, and 53.11% for 2000-image, 6000-image, 8000-image, 12,000-image, and 22,000-image databases, respectively. For the 2000-image database, it improves the second-best system (i.e., hierarchical manifold subgraph) by 1.03%. For the 6000-image database, it improves the second-best system (i.e., hierarchical manifold subgraph) by 3.29%. For the 8000-image database, it improves the second best system (i.e., semantic manifold) by 1.73%. For the 12000-image database, it improves the second best system (i.e., variant 7) by 4.0%. For the 22000-image database, it improves the second best system (i.e., Log-based + global soft label SVM) by 2.7%.

Figure 6 plots the precision and recall curves of the proposed system on the 22000-image database when a different number of images (e.g., 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, and 155) are returned at each of four iterations. Here, precision represents MRP. Recall represents the mean recall that is computed as the total of recall values of all query images divided by the total number of queries, where recall is defined as the ratio of the number of relevant images retrieved to the total number of relevant images (e.g., 100) in the database. It clearly shows recall increases when the number of retrieved images increases for each iteration. However, precision for each iteration drops along with the increasing number of retrieved images. Figure 6 also shows the proposed system is effective in returning above 40% of relevant images at the last iteration (i.e., MRP is above 40%) when the number of images returned is less than 95.

Finally, we summarize the MRPs for several representative categories of the 22000-image database. Specifically, two NUS-WIDE categories (e.g., flags with different backgrounds and water/water drops with different backgrounds)

Figure 6. Precision and recall curve of the proposed system on the 22000-image database



achieve the worst MRP of 7.33% at the last iteration. Five COREL categories (e.g., dinosaurs with pure backgrounds, elephants, masks with pure backgrounds, mineral samples with pure backgrounds, and molecular diagram) and the skyscraper category from 4000 online images achieve the best MRP of 100% at the last three iterations. The pills category of the COREL database achieves the average MRP of 53.33% at the last iteration. The portraits category of the COREL database achieves the median MRP of 54.44% at the last iteration. This clearly shows the effectiveness of the proposed retrieval process on a majority of semantic categories (classes).

3.4. Comparative Complexity and Storage Evaluation

Finally, we compare the above ten CBIR systems from the perspectives of the storage and computational complexity. The proposed CBIR system requires $O(N\times p)$ space to store historical RF information in a compact feedback log for extracting semantic knowledge, where N denotes the total number of images in the database and p denotes the number of columns in the feedback log. Based on our experiments, p is 27, 89, 192, 361, and 1188 for the 2000-image, 6000-image, 8000-image, 12000-image, and 22000-image databases, respectively. Dynamic semantic clustering system requires $O(N\times NumC)$ space, where $NumC$ is the number of learned clusters and is approximately 21, 68, 98, 139, and 326 for the 2000-image, 6000-image, 8000-image, 12000-image, and 22000-image databases, respectively. All the other long-term-based CBIR systems require $O(c\times N\times N)$ space. The c's in hierarchical manifold subgraph, virtual feature learning, and DSF-based cross-session learning systems are a fractional number (e.g., 0.1). The c's in log-based and memory learning systems are 1 and 3, respectively. The c's in gMRBIR, SC-based manifold, and semantic manifold systems are all equal to 4. It clearly shows that the proposed CBIR system requires a little more storage space than dynamic semantic clustering system and a small fraction of storage space as required by the other eight long-term-based CBIR systems. This efficient storage is necessary for real-world situations with databases of millions of images.

The complexity of our proposed retrieval algorithm is $O(N\times p)$. The complexity of dynamic semantic clustering system is $O(N\times NumC+NumC\times NumC)$. The complexity of the other eight long-term-based CBIR systems is $O(c\times N\times N)$. It clearly shows that our proposed is computationally efficient.

4. CONCLUSION AND FUTURE WORK

We propose a novel scalable graph-based semi-supervised ranking system for image retrieval. It takes the advantages of both RF based transductive short-term learning and semantic feature-based long-term learning techniques to improve retrieval performance. Our major contributions are:

- Quickly constructing a compact dynamic feedback log to store retrieval patterns of each past query session;
- Efficiently merging similar semantic concepts to maintain a reasonable number of representative semantics for all images in a database;
- Effectively combining low-level visual and high-level semantic similarity measure to build a scalable manifold graph, which explores the intrinsic structure of images in both low-level visual and high-level semantic feature spaces;
- Effectively designing a layered relevance vector to propagate the relevance scores from anchor images to the second layer graphs and further propagate relevance scores of labeled images to unlabeled image via the hierarchical graph-based structure.

We plan to test the proposed technique for its effectiveness and scalability on a larger database by comparing with emerged state-of-the-art systems. We will first investigate the

usefulness of incorporating other sophisticated and distinguishable features such as histograms of oriented gradients (HOG) to extract low-level visual features. We will then incorporate erroneous feedback, which is resulted from the inherent subjectivity of judging relevance, user laziness, or maliciousness, into the current system to evaluate its resilience to the noisy feedback. Next, we plan to obtain a sufficient number of human subject tests to simulate the user's query log information and investigate how the proposed system would do with real human feedback. Finally, we may explore the potential of applying the proposed technique in the image annotation task.

REFERENCES

Bian, W., & Tao, D. (2010). Biased discriminant Euclidean embedding for content-based image retrieval. *IEEE Transactions on Image Processing*, *19*(2), 545–554. doi:10.1109/TIP.2009.2035223

Cai, D., He, X., & Han, J. (2007). Regularized regression on image manifold for retrieval. In *Proceedings of the International Workshop on Multimedia Information Retrieval* (pp. 11-20).

Chang, R., & Qi, X. (2011). Semantic clusters based manifold ranking for image retrieval. In *Proceedings of the International Conference on Image Processing* (pp. 2473-2476).

Chang, R., & Qi, X. (2013). A hierarchical manifold subgraph ranking system for content-based image retrieval. In *Proceedings of the International Conference on Multimedia and Expo* (Article No. 236).

Chang, R., Xiao, Z., Wong, K. S., & Qi, X. (2012). Learning a weighted semantic manifold for content-based image retrieval. In *Proceedings of the International Conference on Image Processing* (pp. 2402-2404).

Chua, T., Tang, J., Hong, R., Li, H., Luo, Z., & Zheng, Y. (2009). NUS-WIDE: A real-world web image database from National University of Singapore. In *Proceedings of the ACM International Conference on Image and Video Retrieval* (Article No. 48).

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, trends of the new age. *ACM Computing Surveys*, *40*(2), 1–60. doi:10.1145/1348246.1348248

Fechser, S., Chang, R., & Qi, X. (2010). Inter-query semantic learning approach to image retrieval. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (pp. 1246-1249).

Geng, B., Tao, D., Xu, C., Yang, L., & Hua, X. S. (2012). Ensemble manifold regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(6), 1227–1233. doi:10.1109/TPAMI.2012.57

Han, J., Ngan, K. N., Li, M., & Zhang, H. J. (2005). A memory learning framework for effective image retrieval. *IEEE Transactions on Image Processing*, *14*(4), 511–524. doi:10.1109/TIP.2004.841205

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. doi:10.1109/TIP.2006.877491

Hoi, S., Lyu, M. R., & Jin, R. (2006). A unified log-based relevance feedback scheme for image retrieval. *IEEE Transactions on Knowledge and Data Engineering*, *18*(4), 509–524. doi:10.1109/TKDE.2006.1599389

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *2*(1), 1–19. doi:10.1145/1126004.1126005

Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*(1), 262–282. doi:10.1016/j.patcog.2006.04.045

Muneesawang, P., & Guan, I. (2004). An interactive approach for CBIR using a network of radial basis functions. *IEEE Transactions on Multimedia*, *6*(5), 703–716. doi:10.1109/TMM.2004.834866

Pass, G. (1997). Comparing images using color coherence vectors. In *Proceedings of the 4th ACM International Conference on Multimedia* (pp. 65-73).

Qi, X., Barrett, S., & Chang, R. (2011). A noise-resilient collaborative learning approach to content-based image retrieval. *International Journal of Intelligent Systems*, *26*(12), 1153–1175. doi:10.1002/int.20503

Qi, X., & Chang, R. (2007). Image retrieval using transaction-based and SVM-based learning in relevance feedback sessions. In *Proceedings of the 4th International Conference on Image Analysis and Recognition* (pp. 637-649).

Thomee, B. (2010). *A picture is worth a thousand words, content-based image retrieval techniques*. Ph.D. Dissertation, Leiden University, Nethelands.

Tong, S., & Chang, E. (2001). Support vector machine active learning for image retrieval. In *Proceedings of the 9th ACM International Conference on Multimedia* (pp. 107-118).

Wan, X. (2007). Content based image retrieval using manifold-ranking of blocks. In *Proceedings of the International Conference on Multimedia and Expo* (pp. 2182-2185).

Wu, K., & Yap, K. H. (2006). Fuzzy SVM for content-based image retrieval. *IEEE Computational Intelligence Magazine*, *1*(2), 10–16. doi:10.1109/MCI.2006.1626490

Xiao, Z., Clark, M. J., Wong, K., & Qi, X. (2012). Dynamic semantic feature-based long-term cross-session learning approach to content-based image retrieval. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (pp. 1033-1036).

Yin, P. Y., Bhanu, B., Chang, K. C., & Dong, A. (2008). Long-term cross-session relevance feedback using virtual features. *IEEE Transactions on Knowledge and Data Engineering*, *20*(3), 352–368. doi:10.1109/TKDE.2007.190697

Zhou, X. S., & Huang, T. S. (2003). Relevance feedback in image retrieval: A comprehensive review. *ACM Multimedia Systems*, *8*(6), 536–544. doi:10.1007/s00530-002-0070-3

Xiaojun Qi received the BS degree in Computer Science from Donghua University, Shanghai, China, in 1993, the MS degree in Pattern Recognition and Intelligent System from Shenyang Institute of Automation, the Chinese Academy of Sciences, Shenyang, China, in 1996, and the PhD degree in Computer Science from Louisiana State University, Baton Rouge, USA in 2001. In Fall 2002, she joined the Department of Computer Science at Utah State University as a tenure-track assistant professor. In April 2008, she received tenure and was promoted to Associate Professor. She has authored and co-authored more than 60 peer-reviewed journal and conference publications. She has been served as technical program committees for 14 international conferences. She is a senior IEEE member since 2010. Her research interests include content-based image retrieval, digital image watermarking, pattern recognition, and computer vision.

Ran Chang received the BS degree in Electrical Engineering from University of Electronics and Science of China, Chengdu, China, in 2001, the MS degree in Electrical Engineering from University of Electronics and Science of China, Chengdu, China, in 2004, the MS degree in Electrical Engineering from Utah State University, Logan, USA in 2006. In Fall 2006, he joined the Department of Computer Science at Utah State University as a PhD student and became a Ph.D candidate in 2008. He has authored and co-authored about 15 peer-reviewed journal and conference publications. In 2011, he joined the Energy Dynamics Lab in Utah State University as a research assistant and participated in the research of moving object tracking and inductive electricity power transfer for electrical vehicles. His research interests include content-based image retrieval, moving object tracking, pattern recognition, computer vision, and smart vehicle control.