This paper is accepted by the 9th International Conf. on Multimedia Modeling, Taiwan, 2003 for publication.

Modeling Data and User Characteristics by Peer Indexing in Content-based Image Retrieval

Jun Yang[1,2] Qing Li[1] Yueting Zhuang[2]

[1]Dept. of Computer Engineering and Information Technology, City University of Hong Kong
83 Tat Chee Avenue, Kowloon, HKSAR, CHINA
yangjun@acm.org itqli@cityu.edu.hk

[2]Department of Computer Science
Zhejiang University,
Hangzhou, CHINA, 310027
yzhuang@cs.zju.edu.cn

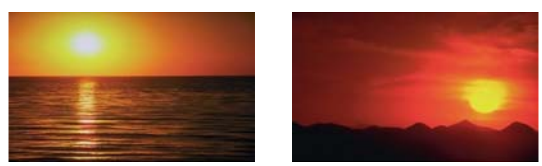
Abstract: Modeling the characteristics of specific images and individual users is a critical issue in content-based image retrieval but insufficiently addressed by the current retrieval approaches. In this paper, we propose a novel approach to data-adaptive and user-adaptive image retrieval based on the idea of *peer indexing*—describing an image through semantically relevant peer images. Specifically, we associate each image with two-level peer index that models the "data characteristics" of the image as well as the "user characteristics" of individual users with respect to this image. Based on two-level image peer indices, retrieval parameters including query vectors and similarity metric can be optimized towards both data and user characteristics by applying the *pseudo feedback strategy*. A cooperative framework is proposed under which peer indices and image visual features are integrated to facilitate data- and user-adaptive image retrieval. Extensive experiments have been conducted on real-world images to verify the effectiveness of our approach.

Keywords: peer indexing, content-based image retrieval, user-adaptive, data-adaptive, pseudo feedback, user modeling, data modeling.

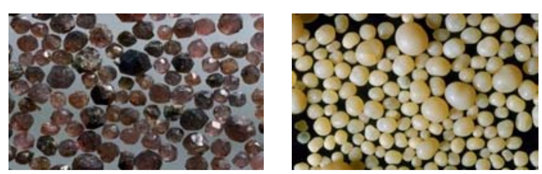
1. Introduction

With the explosive growth of digital images, content-based image retrieval (CBIR) has become one of the most active research areas in the past decade. A generic framework followed by most existing CBIR approaches is that: each image in the database, which is represented by a set of feature vectors of various visual features[1], is matched against a query, which is represented by a set of query vectors, using a similarity metric. Therefore, the retrieval result of a specific query is completely determined by a set of *retrieval parameters*, including query vector(s), visual features, and the similarity metric. Early research on CBIR (Ma et al., 1999) primarily focused on exploring various visual features (e.g., color histogram, Tamura texture, wavelets) as well as various similarity metrics (e.g., Euclidean distance, histogram intersection, Mahalanobis distance), attempting to find the most effective visual feature(s) and similarity metric(s). However, under many circumstances there is no *uniform* optimal solution to these retrieval parameters, due to their dependency on the following two aspects:

- *Data dependency*: The description power of a certain visual feature varies with the types of images it deals with, and different visual features are not equally effective in describing a certain type of images. This can be clearly illustrated by the images in Figure 1 and Figure 2. The feature that best describes the images of "sunset" is dominant color, while most salient feature for the images of "cobble" is texture. Similarly, the effectiveness of various similarity metrics or the same similarity metric with different settings (i.e., the weights for various features) is also dependent on the types of images. The impact of data-dependency on image retrieval is that, for a given query, we need to identify the retrieval parameters that best capture the characteristics of the desired images—*data characteristics*—in order to enhance the retrieval performance.



(a) (b)
Figure 1: Images of "sunset"



(a) (b)
Figure 2: Images of "cobble"

- *User dependency*: Because of the human perception subjectivity (Rui et al, 1998), different persons or the same person under different circumstances may perceive the same image differently. As a manifestation of this subjectivity, different users are probably interested in

[1] In this paper, we regard "visual features" and "low-level features" as the same thing and use them interchangeably.

different features of images. For example, some users may regard image (a) and (b) in Figure 3 to be similar if they care much about the coarseness of the images (specifically, the size of the cobbles), but others may think (b) and (c) are more similar because their overall colors are much closer. Thus, if the visual features and similarity metric that are consistent with the subjectivity of a user—*user characteristics*—are used, the retrieved images can better satisfy the need of this particular user. Note that this user-dependency is not orthogonal with the data dependency; rather, the user's preference on retrieval parameters varies with images, e.g., a user may prefer color features to describe some images and prefer texture features for other images. In this sense, the "user characteristics" itself is data-dependent.

(a) (b) (c)

Figure 3: Illustration of user subjectivity

Due to the data- and user-dependency of retrieval parameters, it is essential in CBIR to figure out the optimal retrieval parameters that adapt to the characteristics of both the desired images of a given query and the particular user who conducts the query, preferably, *without* any user intervention. This offers us a new perspective from which existing works on CBIR can be classified according to the type and degree of the supported adaptation. In most of the early year work, the visual features and their weights in the similarity metric are fixed and therefore *no adaptation* is provided. Other early approaches allow users to select their interested visual feature(s) and specify the feature weights, which is regarded as a form of *manual adaptation*. Recently, relevance feedback techniques have been extensively adopted in CBIR, which improve the retrieval results by updating the various retrieval parameters (mostly, query vectors or feature weights) based on the users' feedbacks in order to better capture the characteristics of the desired images. This can be regarded as a form of *short-term adaptation*, as the adapted parameters take effect only during the current query session and the adaptation needs to start from scratch for the future queries. Some learning approaches based on relevance feedback are devised to "memorize" the adaptations conducted for previous queries and reuse them to benefit future queries, through either sophisticated mathematical models (Minka et al., 1996) or propagation of keywords (Lu et al., 2000). In this regard, they are capable of *long-term adaptation*. All the aforementioned adaptation techniques for CBIR strive to find the best retrieval parameters for specific queries and thus belong to *data-adaptive* approaches. Although the issue of user factors has been discussed in previous works, the emphasis is on the *uncertainty* of the user preferences. To the best of our knowledge, no previous work on CBIR has dedicated to address the *uniqueness* of individual users and the *discrepancy* among various users. Therefore, user-adaptation remains a largely unexplored problem in CBIR.

In this paper, we demonstrate that both data-adaptation and user-adaptation can be achieved using an elegant and effective approach for image retrieval. Firstly, we propose a new scheme for image indexing, *peer indexing*, which describes images through semantically relevant peer images. In particular, each image is associated with a two-level peer index, which includes a *global* peer index modeling the "data characteristics" of the image and a set of *personal* peer indices modeling the "user characteristics" of individual users with respect to this specific image. Based on two-level image peer indices, the optimal query vectors and similarity metric can be optimized towards a specific query conducted by a particular user by applying *pseudo feedback* strategy. Finally, a *cooperative framework* is proposed under which visual features are integrated with peer indices to support data- and user-adaptive image retrieval. Extensive experiments on real-world images have been conducted to verify the feasibility and effectiveness of the proposed retrieval approach.

The rest of this paper is organized as follows. In Section 2, we describe the presentation, learning algorithm, and similarity metric of the two-level peer index of images. The pseudo feedback strategy for adapting the query vectors and the similarity metric are elaborated in Section 3. The cooperative framework for data- and user-adaptive image retrieval is proposed in Section 4. We present the experimental results in Section 5 and review the related works in Section 6. The conclusion and the future works are given in Section 7.

2. Peer Indexing Scheme

Peer indexing is based on a simple and intuitive idea: describing an image by its semantically relevant images. The underlying assumption is that each image has an intrinsic semantic concept, which becomes emergent through its correlation with other images. This notion is analogous to the idea of estimating the impact factor of a scientific journal based on the citations of its papers by the papers of other journals, or calculating the degree of "authority" of a web page through its hyperlinks with other web pages. In this section, we firstly present the formal representation of the two-level peer index associated with each image. A learning algorithm for the acquisition of peer indices is then proposed, and a similarity metric for peer index is lastly formulated.

2.1. *Two-level peer index*

In peer indexing, each image in the database plays a dual role—either as an image to be indexed or as a "peer image" that is used to index other images. A peer index of an image can be represented as a list of semantically peer images that are semantically relevant to it, with a weight attached to each peer image indicating the degree of relevance. (Specifically, a peer image in peer indices is denoted by its unique identifier, e.g., UID.) In our approach, each image is associated with a two-level per index, which includes a *general peer index* maintaining its relevant peer images from the whole user community point of view, and a set of *personal peer indices* maintaining its relevant images from the perspective of individual users.

The two types of peer index have the same representation but differ in semantics: the general index[2] reflects the perception of the whole user community on the relevance among images, while each personal index captures the perception of a particular user. Specifically, the general peer index of an image I_m is defined as:

$$P_m = \{< p_{m1}, w_{m1} >, \cdots, < p_{mi}, w_{mi} >, \cdots, < p_{mN_m}, w_{mN_m} >\} \quad (1.1)$$

where p_{mi} represents a peer image that is semantically relevant to I_m from a perspective representing the whole user community, with the weight w_{mi} indicating the strength of the relevance. The general index of an image captures its "data characteristics", which is implied by the group of its semantically relevant images.

Similarly, the personal peer index of I_m corresponding to the user k is given by:

$$P_m^k = \{< p_{m1}^k, w_{m1}^k >, \cdots, < p_{mi}^k, w_{mi}^k >, \cdots, < p_{mN_{mk}}^k, w_{mN_{mk}}^k >\} \quad (1.2)$$

where $p_{mi}{}^k$ represents a peer image that is semantically relevant to I_m from the opinion of user k. In contrast to general index, the personal index of an image models the "user characteristics" of a particular user pertaining to the specific image. The relationship between two types of peer index is illustrated Figure 4, and their respective semantics will be made clearer through their learning algorithm and the way they are used in the retrieval process.

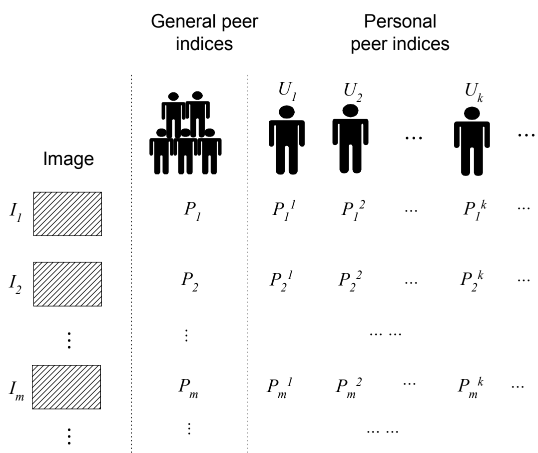


Figure 4: Two-level image peer indices

2.2. *Learning algorithm for index acquisition*

To avoid the significant efforts of building peer indices manually, we suggest a simple machine learning algorithm to derive both types of peer indices progressively from the statistics of user-provided feedback information. This algorithm is embedded in the process of relevance feedback: when a user submits a sample image as the initial query and designates some of the retrieved images as relevant or irrelevant examples, we insert each relevant image and the sample image into each other's peer indices, and remove each irrelevant image and the sample image from each other's peer indices (if exists). Note that the update (insertion and removal) is done to not only the general indices of the involved images (sample image, relevant and irrelevant examples), but also one of their personal indices corresponding to the specific user who conducts the query and the feedback. This algorithm is formally presented in Figure 5.

1. *Collect the sample image I_s, the set of relevant examples I_R, and the set of irrelevant examples I_N from the query and the feedback conducted by user k.*
2. *For each $I_m \in I_R$, if I_m does not exist in P_s, insert it as a peer image into P_s with the initial weight set to 1. Otherwise, increase the weight of I_m in P_s with an increment of 1. Similarly, I_s is also added into P_m or have its weight in P_m increased by 1.*
3. *Update the personal indices of the sample image ($P_s{}^k$) and relevant examples ($P_m{}^k$, for each $I_m \in I_R$) using the similar method as Step 2.*
4. *For each $I_m \in I_N$, if I_m exists in P_s, divide its weight by a factor of 5. If the resulting weight is below 1, remove I_m from P_s. Similarly, I_s is removed from P_m, or has its weight in P_m decreased.*
5. *Update the personal indices of the sample image ($P_s{}^k$) and irrelevant examples ($P_m{}^k$, for each $I_m \in I_N$) using the similar method as Step 4.*

Figure 5: Learning algorithm for two-level peer index

In the algorithm, the weight decrement of an irrelevant example (as a peer image) is much larger than the increment of a relevant example. Therefore, if user behaviors suggest contradictory opinions regarding the relevance between two images (e.g., for a query, some users mark an image as relevant but others mark it as irrelevant), the corresponding peer image(s) will not have a large weight.

The proposed learning algorithm is consistent with the semantics of the two types of peer indices: the general peer index of an image is updated according to the behaviors of all users, while each personal peer index is updated according to the behaviors of the corresponding user. As user queries and feedbacks proceed, both types of peer indices are improved in coverage and quality. Hence, the general index gradually reflects the perception of the whole user community about the relevance among images, while each personal index approximates the personal perception of each particular user.

2.3. *Similarity metric*

From Eq.1.1 and Eq.1.2, one can easily see an analogy between peer index and keyword annotation of images, which is usually represented as a list of weighted keywords. Actually, each peer image in the peer index can be regarded as a "visual keyword", i.e., a visual representation of a semantic concept embodied by the image. Due to this analogy, mature techniques developed for

[2] For simplicity, we use "general index" for "general peer index" and "personal index" for "personal peer index".

text-based information retrieval (IR) can be applied on peer index as well. Among them, *term weighting* is a technique for determining the weights of keywords in a document. A well-known term weighting scheme is the so-called TF*IDF (Salton et al., 1982), which considers two factors: (1) *term frequency* (TF) as the frequency of a keyword in the document, and (2) *inverse document frequency* (IDF) indicating the discriminative power of a keyword by considering the number of documents in which it appears. Peer indices can be weighted using two similar factors. According to the learning algorithm described in Figure 5, the weight w_{mi} (or w_{mi}^k) of a peer image p_{mi} (or p_{mi}^k) in a peer index P_m (or P_m^k) reflects the number of user feedbacks that confirm the relevance of p_{mi} (or p_{mi}^k) with image I_m, and therefore corresponds to the first factor TF. Thus, we need to adjust it to include the discriminability factor using either Eq.2.1 (for general index) or Eq.2.2 (for personal index):

$$r_{mi} = w_{mi}\left(\log\frac{M}{M_{mi}} + 1\right) \tag{2.1}$$

$$r_{mi}^k = w_{mi}^k\left(\log\frac{M}{M_{mi}^k} + 1\right) \tag{2.2}$$

where M is the total number of images in the database, M_{mi} is the number of images whose general peer index has p_{mi} in it, and M_{mi}^k is the number of images whose personal peer index corresponding to user k has p_{mi} in it. Thus, a peer image concentrating on the peer indices of a few images is weighted higher than the one spreading over many images, which are less capable of differentiating among images.

The similarity between the two peer indices (of either type) can be calculated by *cosine similarity,* a similarity function extensively used for text-based IR. That is, we treat a peer index (global or personal) as a vector, with each peer image corresponding to a dimension and the weight (after adjustment by Equ.2.1 or 2.2) of this peer image as the length of the vector along this dimension. The similarity between two peer indices is transformed into cosine value of the angle formed by their corresponding vectors. We define the similarity between image I_m and I_n in terms of their general peer index using Eq.3.1. and in terms of their personal index using Eq.3.2:

$$R_{mn} = \frac{P_m \cdot P_n}{\|P_m\|\|P_n\|} \tag{3.1}$$

$$R_{mn}^k = \frac{P_m^k \cdot P_n^k}{\|P_m^k\|\|P_n^k\|} \tag{3.2}$$

where $\|\ \|$ is the norm of a vector, and $\cdot$ is dot product. It is clear that the similarity as the outcome of Eq.3.1 or Eq.3.2 is within the range of [0,1].

3. Adaptation Strategy by Pseudo Feedback

The previous section has described the two-level peer indices of images, which models the data characteristics of images as well as the user characteristics of individual users. In this section, we suggest a *pseudo feedback* strategy for adapting both queries and similarity metrics towards data and user characteristics.

3.1. *Retrieval model*

Suppose each image I_m can be described by a set of totally K visual features as $[\vec{x}_{m1},...,\vec{x}_{mi},...,\vec{x}_{mK}]$, where each feature can be represented as a vector. We use $\vec{x}_{mi} = [x_{mi1},...,x_{mik},...,x_{miK_i}]$ to denote the i^{th} feature vector of image I_m, where K_i is the length of that feature vector. Similarly, each query Q (usually composed by a sample image) can be represented by a set of query vectors as $[\vec{q}_1,...,\vec{q}_i,...,\vec{q}_K]$, with $\vec{q}_i = [q_{i1},...,q_{ik},...,q_{iK_i}]$ as the query vector for the i^{th} feature.

To compute the distance between the query Q and an image I_m, we firstly calculate their distance in each feature space, which is defined by the *generalized Euclidean metric*. Specifically, the distance between Q and I_m on the i^{th} feature is formulated as:

$$d_{mi} = (\vec{q}_i - \vec{x}_{mi})^T W_i (\vec{q}_i - \vec{x}_{mi}) \tag{4}$$

where W_i is a $K_i \times K_i$ symmetric full matrix for the i^{th} feature space. The diagonal elements of W_i model the importance of the components x_{mik} of the i^{th} feature vector, and the non-diagonal elements of W_i model the correlations between different components of the i^{th} feature vector.

The overall distance between query Q and image I_m is defined as a linear combination of their distances on individual feature spaces, given by:

$$d_m = \vec{u}^T \vec{d}_m \tag{5}$$

where $\vec{d}_m = [d_{m1},...,d_{mi},...,d_{mK}]$ is a vector consisting of the distances between Q and I_m on individual features, and $\vec{u} = [u_1,...,u_i,...,u_K]$ is a vector of length K with each component u_i being the weight for the distance on the i^{th} feature space.

The overall distance d_m defines the final similarity of I_m to the query, which is fully determined by the query vectors $\vec{q}_i$, matrices W_i for various features, and the vector $\vec{u}$ specifying the weights of the distances on individual features. Among them, the last two parameters can be regarded as components of the similarity metric. Therefore, our objective is to figure out the optimal values of the three parameters (denoted as $\vec{q}_i^{\,*}$, W_i^*, and $\vec{u}^{\,*}$) so that images retrieved under the optimized parameters can best satisfy the requirements of a particular user.

3.2. *Pseudo feedback strategy*

Rui et al. (2000) has proposed a learning approach to optimizing the query vectors and the similarity metric based on user feedbacks. The inputs of their approach includes query vectors $\vec{q}_i$ for various features, a set of N images as training samples (which are the relevant examples labeled by the user in feedbacks), and a vector $\vec{\pi} = [\pi_1,...,\pi_n,...,\pi_N]$ denoting the degree of relevance of each training sample to the query. Each training sample has a set of feature vectors $\vec{x}_{ni}$ corresponding to various features. By formulating a minimizing problem on the sum of distances between individual training samples and the query, Rui et al. (2000) provides the optimal solutions to the query vectors and the similarity metric. Specifically, the optimal query vector of the i^{th} feature is given by:

$$\vec{q}_i^{T*} = \frac{\vec{\pi}^T X_i}{\sum_{n=1}^{N} \pi_n} \tag{6}$$

where X_i is a $N \times K_i$ training sample matrix for the i^{th} feature constituted by stacking the feature vector $\vec{x}_{ni}$ of each training sample. The optimal matrix W_i^* for the i^{th} feature is defined as:

$$W_i^* = (\det(C_i))^{\frac{1}{K_i}} C_i^{-1} \tag{7}$$

where C_i is the $K_i \times K_i$ weighted covariance matrix of X_i, given as:

$$C_{i_{rs}} = \frac{\sum_{n=1}^{N} \pi_n (x_{nir} - q_{ir})(x_{nis} - q_{is})}{\sum_{n=1}^{N} \pi_n} \quad r, s = 1, ..., K_i \tag{8}$$

The optimal weight for the i^{th} feature in the vector $\vec{u}^*$ is given by:

$$u_i^* = \sum_{j=1}^{K} \sqrt{\frac{f_j}{f_i}} \tag{9}$$

where $f_i = \sum_{n=1}^{N} \pi_n d_{ni}$ with d_{ni} as the distance between the n^{th} training sample and the query on the i^{th} feature.

In the approach proposed by Rui et al. (2000), the training samples and their relevance scores $\vec{\pi}$ are explicitly designated by users in the feedback process. In our approach, we eliminate the need of user feedbacks by obtaining the training samples and $\vec{\pi}$ based on the two-level image peer indices immediately after a sample image is submitted as a query from a user. In particular, we use both the general peer index and the personal peer index (of the current user) of the sample image to match with the general and personal index of each image in the database, and treat the matched images as training samples. To compute the degree of relevance π_n of each matched image, a similarity metric integrating both general and personal index is formulated as:

$$\pi_n = \max(\varepsilon \cdot R_{nq}, R_{nq}^k) \tag{10}$$

where R_{nq} is the similarity between the general peer indices of an image I_n and the sample image I_q calculated by Eq.3.1, while $R_{nq}{}^k$ is the similarity between their respective personal indices corresponding to user k calculated by Eq.3.2. The variable ε is set to a value between [0, 1] in order to give higher priority to the personal index. Therefore, an image matched by a personal index (corresponding to the particular user conducting the query) is given a higher degree of relevance than an image matched by the general index with the same resultant similarity.

Whenever a user submits a query as a sample image, we match it against all the images in the database using Eq.10, and the images with non-zero π_n are regarded as the training samples, with π_n as their degree of relevance. The training samples, together with the query vectors $\vec{q}_i$ extracted from the sample image, are fed into Eq.6, Eq.7 and Eq.9 to calculate the optimal $\vec{q}_i^{\,*}$, W_i^*, and $\vec{u}^*$. This strategy is named as "pseudo feedback" as it executes the same adaptation process as the genuine feedback (Rui et al., 2000) using the training samples implicitly obtained based on peer indices rather than explicitly supplied by users. (As an exception, if the number of matched images is less than two, we cannot optimize the three retrieval parameters, which remain as their initial values. That is, $\vec{q}_i^{\,*}$ is the same as $\vec{q}_i$, W_i^* is an identity matrix, and $\vec{u}^*$ is a vector with all elements set to 1.) As the training samples are obtained based on general and personal indices, which model the data and the user characteristics respectively, the optimal parameters are essentially adapted to the characteristics of the current query and the user who conducts the query. Naturally, the images retrieved based on these parameters (see Section 4) will "inherit" such adaptation effect.

4. A Cooperative Framework for Data- and User-adaptive Image Retrieval

Undoubtedly, images matched by peer indices are mostly relevant to the sample image, since peer indices are essentially the memorizations of the user-perceived relevance among images. Nevertheless, as peer indices need to be gradually accumulated from user feedbacks, their contribution to the retrieval task is limited when they are unavailable or insufficient for either the candidate images (i.e., images in the database) or the sample image. To reach its full capacity, we use peer index in conjunction with image visual features such that they can benefit each other to yield better retrieval results. A cooperative framework is proposed under which they are seamlessly integrated to support data- and user-adaptive image retrieval.

4.1. *The retrieval process*

For each image in the database, we extract three types of visual features, including 256-d HSV color histogram, 64-d Lab color coherence, and 32-d Tamura directionality. Sample images used to compose user queries can be either new images submitted by the user or existing images selected from the database. The similarity between two images in terms of visual features is defined as the inverse of their distance calculated by Eq.5.

The retrieval process is performed in a multi-pass process. Firstly, the relevance of each candidate image to the sample image is calculated based on two-level peer index using Eq.10. The images matched in the first step, together with their relevance score, are used to calculate the optimal query vectors and the optimal similarity metric using Eq.6, Eq.7 and Eq.9 in the second step. The first two steps together correspond to the pseudo feedback strategy described in Section 3.2. In the third step, the optimized query and similarity metric are used to compute the similarity of each image to the sample image in terms of visual features by Eq.5. Finally, overall similarity of each candidate image is combined from the similarity on peer index and that on visual features, given by:

$$G_m = (1+\pi_m)s_m \tag{11}$$

where G_m is the overall similarity of image I_m to the query, π_m denotes the relevance between I_m and the sample image based on two-level peer index (calculated in the first step), and $s_m = 1/d_m$ is the similarity between I_m and the optimized query vectors in terms of visual features (calculated in the third step).

If the user is not satisfied with the retrieved images, he can provide feedback information by marking the retrieved images as relevant or irrelevant examples. On acceptance of the feedback information, the learning algorithm described in Figure 5 is firstly executed to update the two-level peer indices of the feedback examples as well as of the sample image. After that, we go through a similar multi-pass retrieval process as that for the first round of retrieval: images are matched with the sample image based on peer indices in the first step, and the matched images are fed into the pseudo feedback strategy to calculate the optimal query vector and similarity metric. The difference between the feedback process and the first round of retrieval lies in that, instead of calculating similarity of each image by Eq.11, we formulate a more comprehensive similarity function by considering both the relevant and irrelevant examples, given by:

$$G_m = (1+\pi_m)s_m + \frac{\beta}{N_R}\sum_{k\in N_R}\left[(1+\pi_{mk})s_{mk}\right] - \frac{\gamma}{N_N}\sum_{k\in N_N}\left[(1+\pi_{mk})s_{mk}\right] \tag{12}$$

Similar to Eq.11, π_m is the relevance between image I_m and the sample image based on the updated two-level peer index calculated by Eq.10, and s_m is the similarity between I_m and the optimal query vectors calculated using Eq.5. N_R and N_N are the number of relevant and irrelevant examples respectively. π_{mk} is relevance between I_m and the k^{th} relevant (or irrelevant) example based on the updated two-level peer index calculated by Eq.10, and s_{mk} is their visual feature similarity calculated by Eq.5. β and γ are parameters adjusting the contribution of relevant and irrelevant examples to the similarity function. Please note that if there is no feedback example (i.e., both N_R and N_N are zero), Eq.12 can be reduced to Eq.11. Therefore, we can use Eq.12 as a uniform similarity function even for the first round of retrieval. The whole retrieval process is summarized in Figure 6.

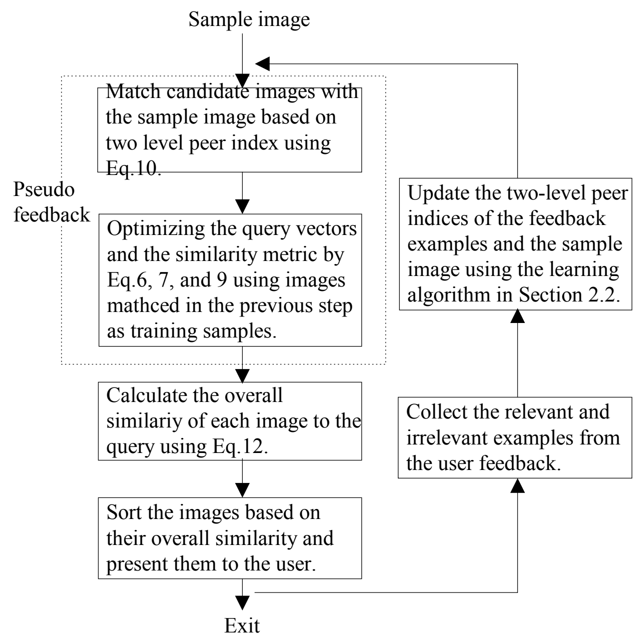


Figure 6: Cooperative framework for data- and user-adaptive image retrieval

4.2. *Discussion*

In the retrieval approach described above, the adaptation towards the data and user characteristics is reflected on the following aspects:

(1) In both Eq.11 and Eq.12, the relevance π_m between each image and the sample image calculated based on peer indices contributes to the overall similarity. Therefore, images matched by either type of the peer index are favored (i.e., given a larger overall similarity) and are likely to be ranked higher than the other images. As the general peer index reflects the common knowledge on the relevance among images, the images matched by it are those generally regarded as relevant to the semantics of the sample image (as the "data characteristics"). On the other hand, since personal peer index corresponds to a particular user's perception on the relevance among images, the images matched by it are consistent with such personal perception (as the "user characteristics"). This form of adaptation is rather basic in the sense that it simply retrieves the images "memorized" by the peer indices.

(2) The visual feature based similarity s_m of each image is calculated based on the optimal retrieval parameters (namely, $\vec{q}_i^{\,*}$, W_i^*, and $\vec{u}^{\,*}$), which are adjusted based on the training samples matched by peer index using the pseudo feedback strategy. Since the training samples are data- and user-adapted, these optimal retrieval parameters and in turn the similarity s_m are intrinsically adapted to the data and user characteristics. Compared with the adaptation in discussed in (1), this form of adaptation is more sophisticated and desirable since it generalizes the

knowledge in peer indices to facilitate adaptive retrieval of the images that are not "memorized" by peer indices.

(3) According to the definition of Eq.10, the images matched by personal indices are given higher priority over the images matched by general indices. Therefore, in both forms of adaptation mentioned in (1) and (2), the "user characteristics" is considered in preference to "data characteristics". (In fact, since the "user characteristics" itself is image-specific, the "data characteristics" is addressed anyway.) If the personal indices of the images are unavailable or insufficient, their general indices may dominate. In this case, only the adaptation to "data characteristics" is supported.

In conclusion, personal peer index provides a means of reusing and generalizing the information embedded in a particular user's past behaviors to adapt image retrieval (results) towards his personal need. In this regard, it enables "learning from one's past behaviors". In comparison, general peer index allows a user to reuse and generalize the information embedded in the past behaviors of the whole user community for retrieving images consistent with common knowledge. Therefore, it supports "learning from other users".

5. Experimental Results

To demonstrate the effectiveness of our retrieval approach, experiments based on real-world images are conducted and the results are presented in this section.

5.1. *Experiment setup*

We have implemented a prototypical image retrieval system based on the retrieval approach presented in this paper and conducted experiments on real-world images. The test data consists of 8,000 images from Corel Image Gallery. The test images have been already classified into 80 topical categories by domain experts, with exactly 100 images in each category. Note that this classification is based on high-level concepts rather than visual features. In order to study the system performance under different situations, the test data set is chosen such that the images in some categories (e.g., "horse", "dawn") have very similar visual aspects, while images in other categories (e.g., "insect", "city") look largely different. The classification is used as the *ground truth* in our experiments. That is, if the query is composed by a sample image selected from a certain category, the rest images in the same category are relevant to the query, and none of the other images is relevant even if it is visually similar to the sample. Using this setting, we examine the performance of data-adaptive retrieval by using general peer index only, as well as the performance of data- and user-adaptive retrieval by using the two-level peer index.

5.2. *Performance of data-adaptive retrieval*

The test on the performance of data-adaptive retrieval investigates the effect of both short-term adaptation and long-term adaptation. In this experiment, we ignore personal indices and use general indices only in the retrieval process, in order to eliminate the influence of user factors. This adjustment does not require any change to the retrieval procedure shown in Figure 6 except that the similarity calculated based on personal indices is assumed to be zero and no update needs to be done on personal indices.

The experiment is conducted in a fully automated way which does not need any user intervention. Specifically, each query is composed by a sample image randomly selected from the test data set by the system. For each query, the system returns 100 images that are ranked top by our retrieval approach as the results. User feedbacks are automatically generated by the system among the 100 retrieved images according to the ground truth, i.e., images that belong to the same category as the sample image are labeled as relevant and the rest are labeled as irrelevant. Based on these feedback examples, the system refines the retrieval results by following the procedure shown in Figure 6. The feedback can last for more than one round. Since the number of retrieved images is equal to the number of relevant images, the value of precision and recall are the same and we use "retrieval accuracy" to refer to both of them.

The short-term adaptation effect can be examined from the change of retrieval accuracy in the course of relevance feedbacks. For this purpose, we generated totally 320 random queries (4 queries for each category) and conducted 15 rounds of feedback for each query. Before each query is conducted, the general peer indices of all images are cleared. The average retrieval accuracy achieved at each round of feedback is shown in Figure 7. For comparison purpose, we run the same 320 queries using the CBIR relevance feedback approach suggested by Rui et al. (2000), the same approach used by our pseudo feedback strategy. The performance of this "comparison experiment" is plot together with that of our approach in Figure 7.

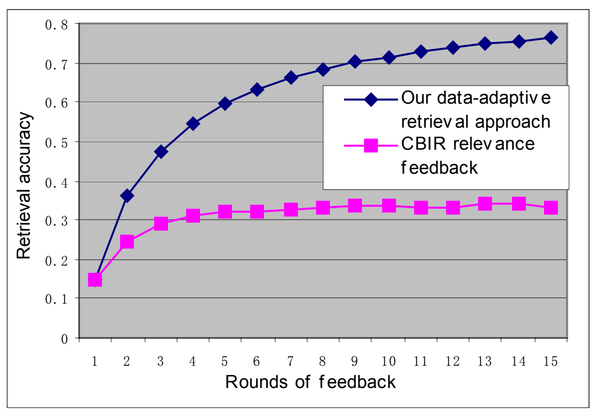


Figure 7: Performance comparison on short-term adaptation

As shown in Figure 7, our approach is very effective in short-term adaptation in terms of enhancing the retrieval accuracy. Initially, the two approaches are at the same performance level, because without initial peer index our approach is reduced to the CBIR approach (referred as the "comparison approach" subsequently). As the feedback proceeds, our approach outperforms significantly the comparison approach, achieving accuracy as high as 76.6% after 15 rounds of feedback. In contract, the accuracy of the comparison approach hovers around 33% after 6 feedbacks and does not grow further even with more feedbacks. To test the robustness of the two retrieval approaches, we also calculate the standard deviation of the final accuracy (after 15 feedbacks) in different image categories, which is 10.3% for our approach, compared with 15.5% for the comparison approach. Looking into the final accuracy on each image category, we find that the performance of the comparison approach fluctuates greatly over categories, achieving 83.7% for a certain category (which has visually similar images in it) and only 16% for another category (which has visually heterogeneous images). In comparison, our approach performs more steadily over images of different categories, with the lowest accuracy being 52.3%.

The long-term adaptation effect is studied by examining the retrieval accuracy across different retrieval sessions. (A retrieval session consists of a query and the subsequent queries.) The experiment is designed as follows: For each category, we applied a succession of retrieval sessions, with each session consisting of a random query followed by a single round of feedback. Since the feedback in each session causes the general peer indices of the images updated, which will be exploited in the subsequent sessions, the retrieval accuracy is expected to increase across sessions. We conduct this experiment on all the image categories and show the change of average retrieval accuracy in Table.1. As we can see, the accuracy improves substantially over sessions, reaching 42.0% after 12 sessions. Given that only a single round of feedback is conducted in each session, our approach is very effective in long-term adaptation.

Table.1: Performance of long-term adaptation

session	1	2	3	4	5	6
accuracy(%)	14.1	25.7	30.1	33.5	35.7	37.3
session	7	8	9	10	11	12
accuracy(%)	38.5	39.5	40.3	40.9	41.6	42.0

5.3. *Performance of user-adaptive retrieval*

The experiment on the performance of user adaptation investigates the ability of our approach to address the *uniqueness* of user's information need, specifically, the situation that different users prefer different images from the same query. The experiment is conducted by simulating the information needs and the behaviors of real users. Firstly, the test data set and the ground truth used in the previous experiments need to be adjusted. In particular, for each image category, we take one random image out and apply the k-means unsupervised clustering algorithm (Hartigan et al. 1979) to cluster the rest 99 images into three subcategories based on their visual features. Since the sizes of subcategories are not fixed and some of them may be very small, we discard all the image categories having at least one subcategory with less than 15 images in it. We number the remaining 26 categories by digital order (from 1 to 26), and number the sub-categories by the order of their category plus an alphabet (from *A* to *C*), e.g., *8B* represents the second subcategory of the category *8*. The random image picked out from each category serves as the sample image of that category. In addition, we suppose there are three "simulated" users as *A*, *B*, and *C*, who, if using the sample image of a category as the query, regard the images in the subcategory with the same order as correct results (i.e., different users have different ground truth). For example, user *C* regards the images in subcategory *8C* as the correct results of a query composed by the sample image of category *8*, while images in *8A* and *8B* are irrelevant results for him.

Similar to the previous experiments, queries and feedbacks in this experiment are also conducted automatically by the three "simulated" users, except that the query must be composed by the designated sample image of each category. The number of images to retrieve for a specific query is equal to the size of the subcategory preferred by the user who conducts the query. (Therefore, the value of prevision and recall are still the same.) Automatic feedbacks are conducted based on the retrieved images, among which the images preferred by the current user are labeled as relevant and the rest as irrelevant. For example, if we assume that user *A* conducts a query using the sample image of category *8* as the query, we return the retrieved images with the number equal to the size of subcategory *8A*, and among them, the images belonging to *8A* are marked as relevant examples.

For each image category, a succession of retrieval sessions is conducted by the three simulated users in an interleaved manner (i.e., *A*, *B*, *C, A, B, C…)*, with each session consisting of a single round of feedback. The consecutive three sessions (i.e., a round of '*A, B, C*') are collectively called a "batch", and the retrieval accuracy of a batch is defined as the average of that of the three sessions. We conduct such experiment on all image categories and track the change of average retrieval accuracy across different batches, under the following two conditions: (1) using general peer index only, and (2) using two-level peer index. The performance achieved under these two conditions is shown in Figure 8.

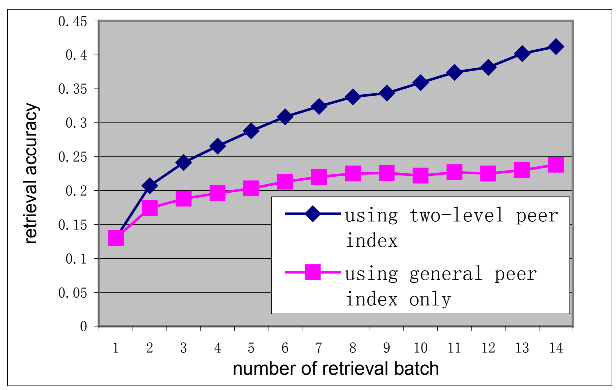


Figure 8: Performance comparison on user-adaptation

As can be seen from Figure 8, the retrieval accuracy increases steadily over retrieval batches (41.1% after 14 batches) by using the two-level peer index. In comparison, if only the general index is used, the performance improvement is not very substantial and the retrieval accuracy hovers around 23% after about 7 batches. The performance difference can be explained as follows: Since different users stick to different ground truth, their feedback behaviors "pull" the general indices towards different directions. To each user, only part of the general indices is useful to retrieve correct results, while other parts are misleading. Therefore, the retrieval accuracy achieved by general indices alone cannot be considerably improved even if a substantial amount of feedbacks has been conducted. In comparison, since personal indices only record the feedback behaviors of a particular user, and they are consulted in advance to general indices in the retrieval process, the retrieval accuracy achieved by two-level peer index is rather high as long as the personal indices have been accumulated from user feedbacks.

6. Related Work

Content-based image retrieval (CBIR) has received extensive study during the last decade (Smeulders et al., 2000). In this section, we specifically review the previous work on adaptive image retrieval techniques, which target at the same problem as our approach.

As discussed in Section 1, there are two classes of techniques that provide non-manual adaptations in image retrieval: relevance feedback techniques and learning approaches based on relevance feedback. Relevance feedback is a powerful technique that improves the retrieval results by adjusting the original query or the similarity metric based on the information fed-back by users about the relevance of the previously retrieved results. For example, MARS system (Rui et al, 1997) has implemented two independent methods that are used to adjust the original query and the visual feature weights respectively, such that both of them can better describe the characteristics of the user-desired images. The feedback method proposed by Ishikawa et al. (1998) for the MindReader system formulates a global optimization problem, the solution to which includes both the optimal similarity metric and the optimal query vector. In most existing relevance feedback approaches for image retrieval, however, the query and/or the similarity metric are adapted only for a specific retrieval session and discarded when the session is finished. Therefore, these adapted parameters cannot be utilized in the subsequent retrieval sessions, where the adaptation has to start from scratch. The long-term retrieval performance remains unchanged even though a large number of feedbacks (adaptations) have been conducted. In this regard, most existing relevance feedback techniques only provide *short-term adaptation.*

A few learning mechanisms have been proposed to build "memories" into relevance feedback such that future queries can benefit from the adaptations achieved in past feedback processes. Hence, these mechanisms are capable of *long-term adaptation.* For example, the image retrieval system proposed by Minka et al. (1996) precomputes many possible groupings of images based on "a society of models" and learns the "bias" towards these groupings from relevant/irrelevant examples to facilitate future queries. Lee et al. (1998) proposed a method to capture the semantic correlations between images from feedbacks and embed them into the system by splitting/merging image clusters, based on which image retrieval is conducted. Both approaches are based on rather complicated mathematical models. The *iFind* system (Lu et al., 2000) adopts a simple keyword propagation mechanism that learns the keyword annotation of images from user feedbacks. To make it work, however, the query must be composed using keywords.

In comparison, our approach achieves long-term adaptation through the learning and exploration of image peer indices, which has advantages over the existing learning approaches on several aspects. First, the representation, update, and similarity calculation of peer index are much simpler and computationally more efficient than the models used by Minka et al. (1996) and Lee et al. (1998). Besides, keyword involvement is not required in our adaptation process. Furthermore, our approach achieves user-adaptation simultaneously with data-adaptation, while existing approaches only support data-adaptation.

7. Conclusions

This paper has presented an elegant and effective approach to model data and user characteristics in content-based image retrieval based on the idea of peer indexing. Specifically, each image is described by a two-level peer index that captures both the "data characteristics" of the image as well as the "user characteristics" of individual users with respect to the specific image. Based on two-level image peer indices, retrieval parameters including query vectors and similarity metric can be adapted towards both data and user characteristics by applying pseudo feedback strategy. A cooperative framework is proposed under which peer indices and visual features are integrated to support data- and user-adaptive image retrieval. Extensive experiments have been conducted to demonstrate the effectiveness of our approach.

In our future work, we plan to investigate two research issues to extend the capability of our approach. First, the current experiment studying the performance of user adaptation is conducted by simulating the behaviors of real users, which is not convincing enough from a practical point of view. In our future work, we will test the user-adaptation performance of our approach using human subjects. Second, personal peer index is a rather primitive and low-level representation of user features in that it simply memorizes the past feedbacks conducted by a user. A challenging future work is to extract high-level user characteristics (e.g., preference, interest) from their low-level features recorded in personal indices. The use of such high-level user characteristics for the retrieval purpose will be investigated. In addition, users can be grouped based on their high-level characteristics such that better collaboration among users can hopefully be achieved.

Reference

1. Hartigan, J. A. and Wong, M. A. 1979. A K-*Means* Clustering Algorithm. Applied Statistics 28:100--108.

2. Ishikawa, Y., Subramanya, R., and Faloutsos, C. 1998. Mindreader: Query databases through multiple examples. In Proc. 24th Conf. Very Large Database, New York, pp. 218-227.

3. Lee, C.S., Ma, W.Y., Zhang, H.J. 1998. Information Embedding Based on User's Relevance Feedback for Image Retrieval. Technical Report, HP Labs.

4. Lu, Y., Hu, C.H., Zhu, X.Q., Zhang, H.J., and Yang, Q. 2000. A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems. In Proc. ACM Int. Conf. Multimedia, pp. 31-38.

5. Ma, W.Y. and Zhang, H.J. 1999. Content-based image indexing and retrieval. The Handbook of Multimedia Computing, Furht, B. (eds.) CRC Press LLC.

6. Minka, T.P. and Picard, R.W. 1996. Interactive learning with a 'society of models'. In Proc. of IEEE Conf. Computer Vision and Pattern Recognition, pp. 447-452.

7. Rui, Y., Huang, T.S., and Mehrotra, S. 1997. Content-based image retrieval with relevance feedback in MARS. In Proc. IEEE Int. Conf. on Image Processing, pp. 815-818.

8. Rui, Y., Huang, T.S., Ortega, M., and Mehrotra, S. 1998. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. IEEE Trans. on Circuits and Systems for Video Technology, 8(5): 644-655.

9. Rui, Y. and Huang, T.S. 2000. Optimizing Learning in Image Retrieval. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 236-243.

10. Salton, G. and Buckley, C. (eds.). 1982. Introduction to Modern Information Retrieval. McGraw-Hill Book Company, New York.

11. Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(12): 1349–1379.

12. Tamura, H. and Yokoya, N. 1984. Image database systems: A survey. Pattern Recognition, 17 (1): 29-43.