Building a Compact Relevant Sample Coverage for Relevance Feedback in Content-Based Image Retrieval

Bangpeng Yao[1], Haizhou Ai[1], and Shihong Lao[2]

[1] Computer Science & Technology Department, Tsinghua University, Beijing, China
[2] Sensing & Control Technology Laboratory, Omron Corporation, Kyoto, Japan

Abstract. Conventional approaches to relevance feedback in content-based image retrieval are based on the assumption that relevant images are physically close to the query image, or the query regions can be identified by a set of clustering centers. However, semantically related images are often scattered across the visual space. It is not always reliable that the refined query point or the clustering centers are capable of representing a complex query region.

In this work, we propose a novel relevance feedback approach which directly aims at extracting a set of samples to represent the query region, regardless of its underlying shape. The sample set extracted by our method is competent as well as compact for subsequent retrieval. Moreover, we integrate feature re-weighting in the process to estimate the importance of each image descriptor. Unlike most existing relevance feedback approaches in which all query points share a same feature weight distribution, our method re-weights the feature importance for each relevant image respectively, so that the representative and discriminative ability for all the images can be maximized. Experimental results on two databases show the effectiveness of our approach.

1 Introduction

Recently Content-Based Image Retrieval (CBIR) has been an active research topic. Typical CBIR systems use visual contents for image representation and similarity computation. Good surveys on CBIR can be found in [1] and [2].

One of the most challenging problems in CBIR is the "semantic-gap" problem. It means the low level features used to represent an image do not necessarily represent the human perception of that image. Techniques that were applied to reduce the semantic gap mainly include: (1) using object ontology to define high-level concepts [3]; (2) using machine learning methods to associate low-level features with query concepts [4,5]; (3) using relevance feedback (RF) [6,7,8,9,10,11] to learn users' intention. Compared with object ontology and machine learning, which mainly rely on offline learning, RF is an online approach and has been shown to be effective in boosting image retrieval accuracy [2]. During retrieval with RF, users interact with the system and give feedback scores to the images retrieved by the system. Based on the feedback, the system dynamically updates its query structure so that it can better capture users' semantic concepts.

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part I, LNCS 5302, pp. 697–710, 2008.
© Springer-Verlag Berlin Heidelberg 2008

1.1 Related Work

A typical RF approach is to identify the "ideal" query in user's mind. The classical query-point movement methods [6,7] represent the ideal query as a single point in feature space, and try to move this point toward relevant images as well as away from irrelevant images. Recently, query expansion methods [8,9] become more widely used as they can identify more complex query regions by using multiple queries. In these approaches, the relevant images are divided into many clusters, and the cluster centers are treated as new queries. But there are still two unsolved issues in the clustering based approaches: (1) these methods only used user-labeled relevant images, while neglecting the information contained in irrelevant images; (2) it lacks theoretical support that the modified query points are competent to represent the user's intention.

Another kind of RF approach is to find an appropriate transformation that maps the original feature space into a space that better models the user's high-level concepts. This is usually done by feature re-weighting to dynamically update the similarity metric. Techniques frequently used in re-weighting include Rocchio's formula [6] and machine learning. One problem in most of such approaches is that, only one similarity metric is obtained in each iteration, and this metric will be applied to all the query points. But in some applications different query points may require different feature weight distributions. An example of this situation in face image retrieval is shown in Fig. 1(b-e). In some clustering based RF methods [9] this problem is alleviated by using a specific similarity metric for each cluster center. However, these metrics are not optimal, because on the one hand, they depend on the quality of clustering; on the other hand, when learning these metrics, information in other clusters and irrelevant images are not used. In [12], a "local distance function" method was proposed which can obtain a distance function for each query image. But this approach treats all relevant images as queries. So applying such methods in RF will greatly improve the computational burden, and possibly fail to identify the "ideal" query region.

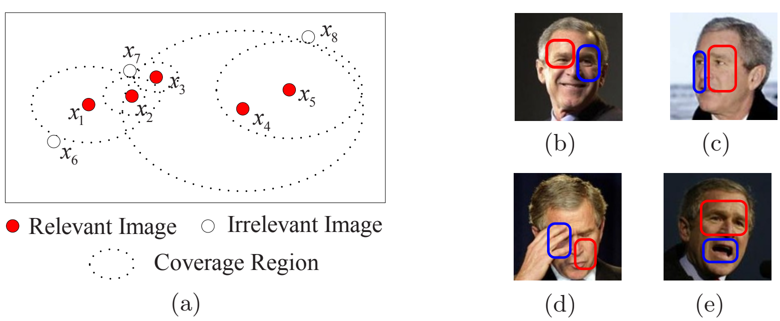


Fig. 1. (a) An illustration of sample coverage. $\{x_1, x_4\}$ is the minimum set that can cover all the other relevant images. (b)-(e) A same facial part may play different roles in similarity measure for face image retrieval. For example, intuitively the red regions should be more significant and discriminative than the blue ones.

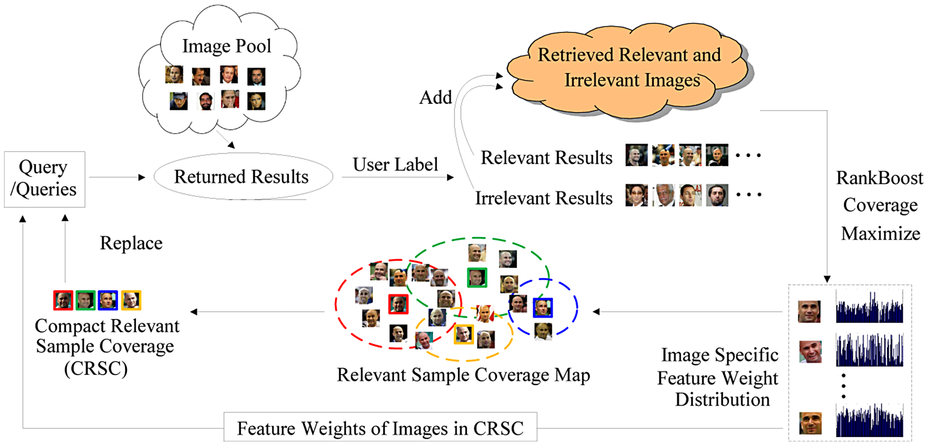


Fig. 2. Overview of our relevance feedback approach

1.2 Outline of Our Approach

In order to address the above issues, in this paper we propose a novel FR approach shown in Fig. 2. Our method directly aims to select a set of images that are competent as well as compact to represent the query region. In one RF iteration, the user specifies whether the retrieved images are relevant to the query, and assigns a relevance score to each relevant image. Based on the user's feedback, we define each relevant image's coverage set, which contains the set of relevant images that can be solved by this image in the nearest neighbor rule. Then, we use a boosting method to obtain a specific feature weight distribution for each relevant image respectively, so that the coverage set of each relevant image can be maximized.

After the coverage maximization stage, our method selects a minimum subset of images that can cover all the other relevant images. This set is called a Compact Relevant Sample Coverage (CRSC). We show that the CRSC extraction problem can be converted to the Minimum Dominating Set (MDS) [13] problem in graph theory. In this work we present an Improved Reverse Heuristic (IRH) method to solve this problem. Images in the CRSC with their feature weight distributions obtained in the coverage maximization stage will be used as new query points in the next retrieval step.

Major contributions in this paper are:

- Using CRSC to represent user's perception, instead of the clustering centers. All the information contained in user's feedback, including irrelevant images, relevant images and their relevance scores, can be used.
- The RankBoost method for simultaneous feature re-weighting and sample coverage maximization.
- The IRH solution for CRSC extraction.
- Each query point has a specific feature weight distribution, so that the representative and discriminative ability for each query can be better utilized.

The rest of this paper is organized as follows. In Sect. 2 we introduce some notations used in the paper and our similarity measure. Section 3 and 4 describe the coverage maximization and CRSC extraction approaches in detail. Experimental results are shown in Sect. 5. Finally in Sect. 6 is the conclusion.

2 Notations and Similarity Measure

In one RF iteration, we have a set of retrieved images $\boldsymbol{X}$. It contains N_{R} relevant images $\boldsymbol{X}_{\mathrm{R}} = \{x_{\mathrm{R},1}, \cdots, x_{\mathrm{R},N_{\mathrm{R}}}\}$ and N_{I} irrelevant images $\boldsymbol{X}_{\mathrm{I}} = \{x_{\mathrm{I},1}, \cdots, x_{\mathrm{I},N_{\mathrm{I}}}\}$. $x_{\mathrm{R},r}$'s relevance score is v_r. The CRSC of $\boldsymbol{X}$ is denoted as $\widetilde{\boldsymbol{X}} = \{\tilde{x}_1, \cdots, \tilde{x}_Q\}$.

Assume that each image is represented as a P dimensional vector. The distance from a relevant image $x_{\mathrm{R},r}$ to another image x is measured by[1]

$$\mathcal{D}(x_{\mathrm{R},r}, x) = \sum_{p=1}^{P} w_{r,p} d_p(x_{\mathrm{R},r}, x) \tag{1}$$

where $d_p(x_{\mathrm{R},r}, x)$ is the distance from $x_{\mathrm{R},r}$ to x measured by feature p, $w_{r,p}$ is feature p's weight when measuring the distance from $x_{\mathrm{R},r}$ to x. The image-specific feature weight distribution for $x_{\mathrm{R},r}$ is denoted as $W_r = \{w_{r,1}, \cdots, w_{r,P}\}$. Note that because we assign different feature weight distributions to different relevant images, the distance from $x_{\mathrm{R},r}$ to $x_{\mathrm{R},j}$ ($\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{R},j})$) is not always equal to the distance from $x_{\mathrm{R},j}$ to $x_{\mathrm{R},r}$ ($\mathcal{D}(x_{\mathrm{R},j}, x_{\mathrm{R},r})$), because $w_{r,p} \neq w_{j,p}$.

In our approach, if $x_{\mathrm{R},r} \in \widetilde{\boldsymbol{X}}$, in the next iteration W_r will be used to measure the distance from $x_{\mathrm{R},r}$ to the images in the candidate pool. The feature weight distribution for $\tilde{x}_q$ is denoted as $\widetilde{W}_q = \{\tilde{w}_{q,1}, \cdots, \tilde{w}_{q,P}\}$. In the retrieval stage, the distance from the set of multiple query points in $\widetilde{\boldsymbol{X}}$ to an image x in the candidate pool is measured by an aggregate function,

$$\mathcal{D}_{agg}^{\tau}(\widetilde{\boldsymbol{X}}, x) = \frac{1}{Q} \sum_{q=1}^{Q} \left(\mathcal{D}(\tilde{x}_q, x)\right)^{\tau} = \frac{1}{Q} \sum_{q=1}^{Q} \left(\sum_{p=1}^{P} \tilde{w}_{q,p} d_p(\tilde{x}_q, x) \right)^{\tau} \tag{2}$$

where a negative value of τ can make the smallest distance have the largest impact on the aggregate distance function [14]. We choose $\tau = -4$.

3 Coverage Maximization by Feature Re-weighting

3.1 Sample Coverage

Definition 1. *The Coverage Set of an image $x_{\mathrm{R},r}$ is defined as*

$$Cover(x_{\mathrm{R},r}) = \{x_{\mathrm{R},j} | x_{\mathrm{R},j} \in \boldsymbol{X}_{\mathrm{R}}, \mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{R},j}) < D\} \tag{3}$$

where $\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{R},j})$ is the distance from $x_{\mathrm{R},r}$ to $x_{\mathrm{R},j}$, D is the distance from $x_{\mathrm{R},r}$ to its boundary image (the image in $\boldsymbol{X}_{\mathrm{I}}$ which is the nearest to $x_{\mathrm{R},r}$).

[1] In our method, we only need to measure the distance from a relevant image to another image (relevant or irrelevant images, or images in the candidate pool).

The definition is illustrated in Fig. 1(a). In the illustrated situation, x_8 is the boundary image of x_4 and x_5. According to the definition, we have: $Cover(x_1) = \{x_1, x_2\}$, $Cover(x_2) = \{x_2\}$, $Cover(x_3) = \{x_3\}$, $Cover(x_4) = \{x_2, x_3, x_4, x_5\}$, $Cover(x_5) = \{x_4, x_5\}$.

Sample coverage is a well-known concept in case-based reasoning [15]. From its definition, we can see that, the larger the coverage set of a sample, the more significant this sample, because it can correctly solve more relevant samples according to the nearest neighbor rule. In our RF problem, each relevant image has a relevance score. Therefore the coverage competence of a sample $x_{\mathrm{R},r}$ is measured by the sum of relevance scores of the images in its coverage, i.e.

$$\Psi_{Cover(x_{\mathrm{R},r})} = \sum_{x_{\mathrm{R},j} \in Cover(x_{\mathrm{R},r})} v_j . \tag{4}$$

From (3) and (4), we can see that, the definition and measurement of sample coverage makes use of all the information provided by the user. The irrelevant images serve to bound the coverage region, and the relevance scores are used to measure the competence of a coverage set.

3.2 The Image Specific Loss Function

The definition of sample coverage is based on a similarity measure. An image's coverage region can be modified by changing its feature weight distribution. Here, for each $x_{\mathrm{R},r}$, we learn a specific feature weight distribution $W_r = \{w_{r,1}, \cdots, w_{r,P}\}$ to maximize $\Psi_{Cover(x_{\mathrm{R},r})}$, the coverage ability of $x_{\mathrm{R},r}$.

We have two motivations to maximize the coverage ability of each relevant image. First, after the coverage ability of each sample is maximized, we can use a smaller number of images to cover all the relevant images. Second, we can obtain a specific feature weight distribution for each sample, which can be used in the subsequent retrieval stage.

According to Definition 1 and the concepts above, the loss function for obtaining W_r to maximize $x_{\mathrm{R},r}$'s coverage ability can be written as

$$\mathrm{Loss}_{W_r} = \sum_{j=1}^{N_\mathrm{R}} v_j \left\| \sum_{p=1}^{P} w_{r,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{R},j}) > \min_{x_{\mathrm{I},i} \in \boldsymbol{X}_\mathrm{I}} \sum_{p=1}^{P} w_{r,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{I},i}) \right\| \tag{5}$$

where $\|\mathrm{A}\|$ is an indicator function: if A is true $\|\mathrm{A}\| = 1$, otherwise $\|\mathrm{A}\| = 0$.

3.3 Loss Function Minimization Via RankBoost Learning

The algorithm to optimize (5) for all the relevant samples is shown in Fig. 3. Two parameters should be learned in order to minimize Loss_{W_r}: the weight distribution W_r and $x_{\mathrm{R},r}$'s boundary image (the irrelevant image which has the smallest distance to $x_{\mathrm{R},r}$). It is hard to learn the two parameters simultaneously, because the optimal weight distribution varies with respect to the selection of boundary image. Therefore, we treat each irrelevant image $x_{\mathrm{I},i}$ as $x_{\mathrm{R},r}$'s boundary respectively, and obtain a weight distribution $W_{r,i}$. $W_{r,i}$ can maximize the coverage

– **For** each relevant image $x_{\mathrm{R},r}$
- **For** each irrelevant image $x_{\mathrm{I},i}$, treat it as $x_{\mathrm{R},r}$'s boundary sample.
  - ∗ The optimization objective becomes

$$W_{r,i} = \arg\min_{W_{r,i}} \sum_{j=1}^{N_{\mathrm{R}}} v_j \left\| \sum_{p=1}^{P} w_{r,i,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{R},j}) > \sum_{p=1}^{P} w_{r,i,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{I},i}) \right\|. \quad (6)$$

  - ∗ Decompose (6) into N_{R} ranked pairs:

$$\{(x_{\mathrm{R},r}, x_{\mathrm{R},1}), (x_{\mathrm{R},r}, x_{\mathrm{I},i})\}, \cdots, \{(x_{\mathrm{R},r}, x_{\mathrm{R},N_{\mathrm{R}}}), (x_{\mathrm{R},r}, x_{\mathrm{I},i})\}. \quad (7)$$

  - ∗ Assign initial importance value to all the ranked pairs. The importance for $\{(x_{\mathrm{R},r}, x_{\mathrm{R},j}), (x_{\mathrm{R},r}, x_{\mathrm{I},i})\}$ is $\frac{v_j}{\sum_{k=1}^{N_{\mathrm{R}}} v_k}$.
  - ∗ Treat each feature d_p as a weak ranker, run RankBoost P iterations to get a feature weight $w_{r,i,p}$ for each d_p.
  - ∗ Calculate $\ell_{r,i} = \sum_{j=1}^{N_{\mathrm{R}}} v_j \left\| \sum_{p=1}^{P} w_{r,i,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{R},j}) > \sum_{p=1}^{P} w_{r,i,p} d_p(x_{\mathrm{R},r}, x_{\mathrm{I},i}) \right\|$.
- Find $i_r^* = \arg_i \min \ell_{r,i}$, and let $W_r = W_{r,i_r^*}$.

Fig. 3. Algorithm of sample coverage maximization and feature re-weighting

ability of $x_{\mathrm{R},r}$ when $x_{\mathrm{I},i}$ is the boundary, taking no account of the distance from $x_{\mathrm{R},r}$ to the other irrelevant images. The loss resulted from $W_{r,i}$ is $\ell_{r,i}$. After all the irrelevant images are considered, the minimum loss value ℓ_{r,i_r^*} is selected, and its associated W_{r,i_r^*} and x_{I,i_r^*} are the optimal feature weight distribution and boundary image respectively.

When $x_{\mathrm{I},i}$ is set as $x_{\mathrm{R},r}$'s boundary image, the optimization objective becomes (6). As shown in Fig. 3, here (6) is solved by RankBoost [16]. We start by decomposing (6) into a set of ranked pairs as shown in (7). We assume that the image features are P weak rankers, where $d_p(x_{\mathrm{R},r}, x_{\mathrm{I},i}) > d_p(x_{\mathrm{R},r}, x_{\mathrm{R},j})$ means $(x_{\mathrm{R},r}, x_{\mathrm{I},i})$ is ranked higher than $(x_{\mathrm{R},r}, x_{\mathrm{R},j})$ by the pth feature. Our goal is to find a strong ranker $\mathcal{D}$, which is a linear combination of $\{d_1, \cdots, d_P\}$ using a set of weights $\{w_{r,i,1}, \cdots, w_{r,i,P}\}$, so that $\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{I},i})$ can be ranked higher than $\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{R},j})$ for all the $j = 1, \cdots, N_{\mathrm{R}}$. The mis-ranking between $\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{I},i})$ and $\mathcal{D}(x_{\mathrm{R},r}, x_{\mathrm{R},j})$ is penalized with v_j, the relevance score of $x_{\mathrm{R},j}$. This learning objective is exactly consistent with the formal ranking problem defined in [16], and thus the optimal $W_{r,i}$ can be found with RankBoost learning.

Like other boosting methods, RankBoost [16] operates iteratively and in each iteration, selects a "best" weak ranker and determines its importance. In the tth iteration, the best weak ranker h_t and its weight α_t is selected according to

$$h_t = \arg\max_{d_p} \sum_{j=1}^{N_{\mathrm{R}}} \rho_{t,j} \left(d_p(x_{\mathrm{R},r}, x_{\mathrm{I},i}) - d_p(x_{\mathrm{R},r}, x_{\mathrm{R},j}) \right) \quad (8)$$

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 + \sum_{j=1}^{N_{\mathrm{R}}} \rho_{t,j} (h_t(x_{\mathrm{R},r}, x_{\mathrm{I},i}) - h_t(x_{\mathrm{R},r}, x_{\mathrm{R},j}))}{1 - \sum_{j=1}^{N_{\mathrm{R}}} \rho_{t,j} (h_t(x_{\mathrm{R},r}, x_{\mathrm{I},i}) - h_t(x_{\mathrm{R},r}, x_{\mathrm{R},j}))} \right) \quad (9)$$

where $\rho_{t,j}$ is the importance of distance pair $\{(x_{\mathrm{R},r}, x_{\mathrm{R},j}), (x_{\mathrm{R},r}, x_{\mathrm{I},i})\}$ in the tth iteration. As shown in Fig. 3, $\{(x_{\mathrm{R},r}, x_{\mathrm{R},j}), (x_{\mathrm{R},r}, x_{\mathrm{I},i})\}$'s initial importance is $\rho_{1,j} = \frac{v_j}{\sum_{k=1}^{N_{\mathrm{R}}} v_k}$. After h_t is selected, the distance pair's importance value is updated by

$$\rho_{t+1,j} = \frac{\rho_{t,j} \exp(\alpha_t (h_t(x_{\mathrm{R},r}, x_{\mathrm{R},j}) - h_t(x_{\mathrm{R},r}, x_{\mathrm{I},i})))}{Z_t} \tag{10}$$

where Z_t is a normalization factor so that $\rho_{t+1,j}$ is a distribution.

In our method, once a feature has been chosen, it cannot be selected again. For each learning task we implement RankBoost P iterations, and thus each feature can have a weight value.

4 An Improved Reverse Heuristic Solution for CRSC Extraction

After all the relevant images' coverage sets are maximized, we shall extract the CRSC from $\boldsymbol{X}_{\mathrm{R}}$. Here we show that the CRSC extraction problem can be converted to the Minimum Dominating Set (MDS) problem [13] in graph theory. We propose an Improved Reverse Heuristic (IRH) method to solve this problem.

4.1 Convert CRSC Extraction to the MDS Problem

Definition 2. *The Dominating Set (DS) of a graph G is defined as,*

$\boldsymbol{S}$ *is a subset of the vertex set* $\boldsymbol{U}(G)$. $\boldsymbol{N}_G[\boldsymbol{S}]$ *is the set of vertices in G which are in* $\boldsymbol{S}$ *or adjacent to a vertex in* $\boldsymbol{S}$. *If* $\boldsymbol{N}_G[\boldsymbol{S}] = \boldsymbol{U}(G)$, *then* $\boldsymbol{S}$ *is said to be a dominating set (of vertices in G).*

If there does not exist another dominating set $\boldsymbol{S}'$ *whose* $|\boldsymbol{S}'| < |\boldsymbol{S}|$, *then* $\boldsymbol{S}$ *is the Minimum DS (MDS) of G. ($|\boldsymbol{S}|$ is the number of vertices in* $\boldsymbol{S}$*).*

Proposition 1. *CRSC extraction can be converted to the MDS problem.*

Proof. Given a set of user-labeled images $\boldsymbol{X}$, $\widetilde{\boldsymbol{X}} = \{\tilde{x}_1, \cdots, \tilde{x}_Q\}$ is its CRSC.

Build a directed graph G with N_{R} vertices, where N_{R} is the number of relevant images in $\boldsymbol{X}$. The vertex u_r corresponds to the relevant image $x_{\mathrm{R},r}$ in $\boldsymbol{X}$. In G, there is an edge from u_r to u_j iff $x_{\mathrm{R},r} \in \widetilde{\boldsymbol{X}}$ and $x_{\mathrm{R},j} \in Cover(x_{\mathrm{R},r})$.

According to G's construction process and Definition 2, $\boldsymbol{S} = \{s_1, \cdots, s_Q\}$ is the DS of G, where s_q is the point corresponds to $\tilde{x}_q$.

If G has another DS whose sample size is smaller than Q, then the corresponding images of this set is also a CRSC of $\boldsymbol{X}$. This contradicts the pre-condition that $\widetilde{\boldsymbol{X}}$ is the CRSC of $\boldsymbol{X}$. Therefore, $\boldsymbol{S}$ is the MDS of G.

Thus, if we find $\boldsymbol{S}$ from G, we find $\widetilde{\boldsymbol{X}}$ from $\boldsymbol{X}$. Therefore, extracting the CRSC from $\boldsymbol{X}$ can be converted to a MDS problem. □

The corresponding graph of the situation in Fig. 1(a) is shown in Fig. 4.

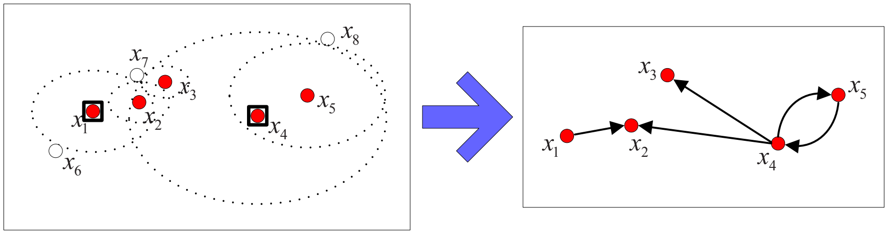


Fig. 4. The left figure is a sample coverage figure (the same as Fig. 1(a)), and the right is its corresponding graph. Extracting a Compact Relevant Sample Coverage of the left figure can be converted to finding the right figure's MSD, which is $\{x_1, x_4\}$.

4.2 The Improved Reverse Heuristic Solution

Based on the above observation, a method for MDS can be directly used to extract CRSC. However, MDS is a well-known NPC problem in graph theory. It is rather time-consuming to find a globally optimal solution. For approximate optimal solutions there are two well-known approaches, *Tabu Search (TS)* and *Reverse Heuristic Algorithm (RH)*. TS works by modifying an initial solution iteratively according to some searching criterions. Although in many situations TS can alleviate the local minima problem that exists in many other approaches, it relies too much on the initial solution, and its convergence speed might be very slow. RH is another algorithm to find approximate solutions for NP problems. In the MDS application, in each iteration of RH, the vertex that has the largest coverage will be selected, and this vertex and those that are connected with it will be removed from the graph. This approach can find an approximate dominating set rapidly. The drawback of RH is the local minima problem.

One reason for the local minima problem in heuristic based algorithms is that, the heuristic rule is not good enough. In the original RH algorithm, only sample coverage is considered. That is, the larger the coverage set, the more significant the sample. However, besides sample coverage, another concept, sample reachability is also important for measuring a sample's competence.

Definition 3. *The Reachability Set of a sample $x_{\mathrm{R},r}$ is defined as*

$$Reach(x_{\mathrm{R},r}) = \{x_{\mathrm{R},j} | x_{\mathrm{R},j} \in \boldsymbol{X}_{\mathrm{R}}, x_{\mathrm{R},r} \in Cover(x_{\mathrm{R},j})\}. \tag{11}$$

For a sample, the larger its reachability set, the less important this sample, because it can be covered by many other samples. Since both sample coverage and sample reachability reflect the importance of a sample, they should be combined to result in a more reliable measure. In our work a sample $x_{\mathrm{R},r}$'s competence for RH is measured as

$$\begin{aligned} Comp(x_{\mathrm{R},r}) &= \Psi_{Cover(x_{\mathrm{R},r})} - \Psi_{Reach(x_{\mathrm{R},r})} \\ &= \sum_{x_{\mathrm{R},j} \in Cover(x_{\mathrm{R},r})} v_j - \sum_{x_{\mathrm{R},k} \in Reach(x_{\mathrm{R},r})} v_k \end{aligned} \tag{12}$$

– **Input**: A set of relevant images $\boldsymbol{X}_{\mathrm{R}}$ and irrelevant images $\boldsymbol{X}_{\mathrm{I}}$.
– **Initialize**: $\widetilde{\boldsymbol{X}} = NULL$.
– **While** $\boldsymbol{X}_{\mathrm{R}} \neq NULL$
  - For each $x_{\mathrm{R},r} \in \boldsymbol{X}_{\mathrm{R}}$, calculate its $Cover(x_{\mathrm{R},r})$ and $Reach(x_{\mathrm{R},r})$;
  - Get $x^*_{\mathrm{R}} \in \boldsymbol{X}_{\mathrm{R}}$ so that, $x^*_{\mathrm{R}} = \arg \max_{x_{\mathrm{R},r} \in \boldsymbol{X}_{\mathrm{R}}} Comp(x_{\mathrm{R},r})$; Ties are broken by selecting the sample with larger coverage set.
  - Append x^*_{R} to $\widetilde{\boldsymbol{X}}$;
  - Remove $Cover(x^*_{\mathrm{R}})$ from $\boldsymbol{X}_{\mathrm{R}}$.
– **Output**: $\widetilde{\boldsymbol{X}}$.

Fig. 5. Improved Reverse Heuristic Algorithm for Minimum Dominating Set Detection

Using (12), we propose an improved reverse heuristic solution for CRSC, shown in Fig. 5. The only difference between our method and the original RH is the heuristic rule.

5 Experiment

5.1 Databases and Evaluation Settings

In this section, we compare our method with some state-of-the-art methods, including Query Movement (Mindreader [7]), Query Expansion (Qcluster [9]), and pure Machine Learning (SVM with triangular kernel [5]).

Experiments are conducted on two publicly available databases: Labeled Faces in the Wild (LFW) [17] and Caltech 101 [18]. LFW is a database of face photographs designed for studying the problem of unconstrained face recognition. It contains 13,233 face images of 5,749 people collected from web, among which 57 people have more than 20 images. The LFW database contains large variations of head pose, illumination, and expression. The Caltech 101 database is a collection of object pictures belonging to 101 categories, such as airplane, panda, etc. Each category has 40 to 800 images. Most categories have about 50 images. The size of each image is roughly 300×200 pixels.

We use the two databases because face/object recognition and retrieval are both active research topics. Moreover both LFW and Caltech 101 are suitable for CBIR performance evaluation, because of their large size, great homogeneous and heterogeneous variations, and human annotated ground truth available.

5.2 Performance Measure and Initial Queries

We use two performance measures: Recall (Re) and Rank (Ra) to evaluate the effectiveness of these RF methods. Recall is defined as the number of retrieved relevant images over the total number of relevant images in the candidate pool. For two RF algorithms A and B, if $Re_A > Re_B$, then A is better than B in terms of Re, because A retrieves more relevant images than B. Rank is the average rank of the retrieved relevant images returned by the system. Obviously, if $Ra_A < Ra_B$, then A is better than B in terms of Ra.

Scope (Sc) is the number of images returned to the user in each RF iteration. It may also affect the performance of a RF approach. Here we measure the effectiveness of these RF methods when Sc is 40 and 80 respectively. The iteration number of each experiment is 4. Note that we only label "relevant" ($v_r = 1$) and "irrelevant" to the returned images, without providing different relevance scores. This is because in the databases human annotated ground truth is available.

It is known that the quality of initial query is important for CBIR. Having a frontal neutral face as the initial query usually achieves better retrieval results than that obtained by using a profile image. Thus, a system that performs well on selected queries does not necessarily work well on not-selected images. In this work, the initial queries are selected randomly. On LFW, we randomly selected 100 images from the 57 people with more than 20 images as initial queries. On Caltech 101, one random image per class was selected. On the two databases, average retrieval results are reported for performance evaluation.

5.3 Visual Features

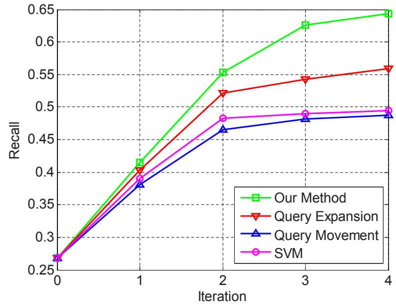
Early CBIR systems mainly rely on low-level features such as color and edge. With the advances of computer vision research in these years, many novel and domain-specific features are available. Here we use Local Binary Pattern (LBP) [19] for face similarity measure, and use Bag of Words (BoW) [18] to measure generic object similarity. Both LBP and BoW are recently developed features and have been widely used in face and object recognition respectively.

In the LBP representation, each face image is normalized to the size of 64×64 pixels according to two-eye centers. The normalized image is divided into 8×8 sub-regions. LBP histogram in each region is treated as a feature. The histogram similarity is measured using Chi-square distance, as in [19]. BoW is also a histogram based feature. We generate the texture codebook using 200 images that are randomly selected from the Caltech 101 database. The codebook histogram size is 10×10, and has 128 bins. In the BoW representation each histogram bin is treated as a feature.

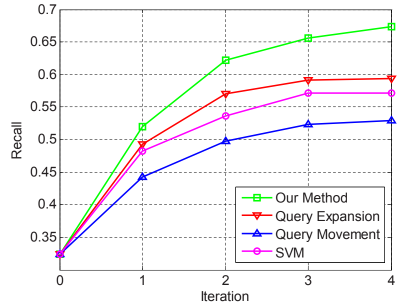
It should be noted that the selection of visual features in our experiment is not optimal. For example, we did not use any color information in extracting LBP and BoW features. It is possible that using other features can obtain better results than that reported in this paper. The focus of this paper is the RF scheme. Which feature is the best for CBIR is beyond the content of this paper.

5.4 Results and Analysis

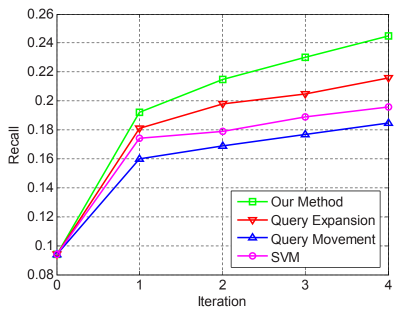
Experiment results of the four approaches are shown in Fig. 6 and Table 1. Figure 6 illustrates that, our method consistently outperforms the other approaches by a large margin. What is more, the margin increases with the iteration number. This is important for a RF system, because the images retrieved after 2 or 3 iterations are usually distinct from the initial query in terms of feature representation. However, these images are closely related to the query in human's perception. So we can say our method better reduces the "semantic gap" between feature representation and human perception.



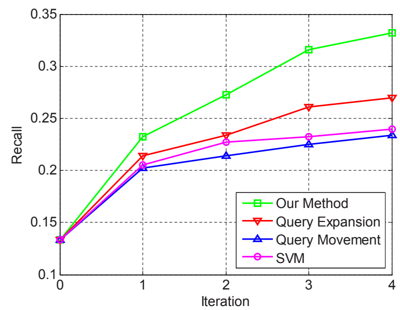
(a) Comparison of recall in LFW database when $Sc = 40$.



(b) Comparison of recall in LFW database when $Sc = 80$.



(c) Comparison of recall in Caltech 101 database when $Sc = 40$.



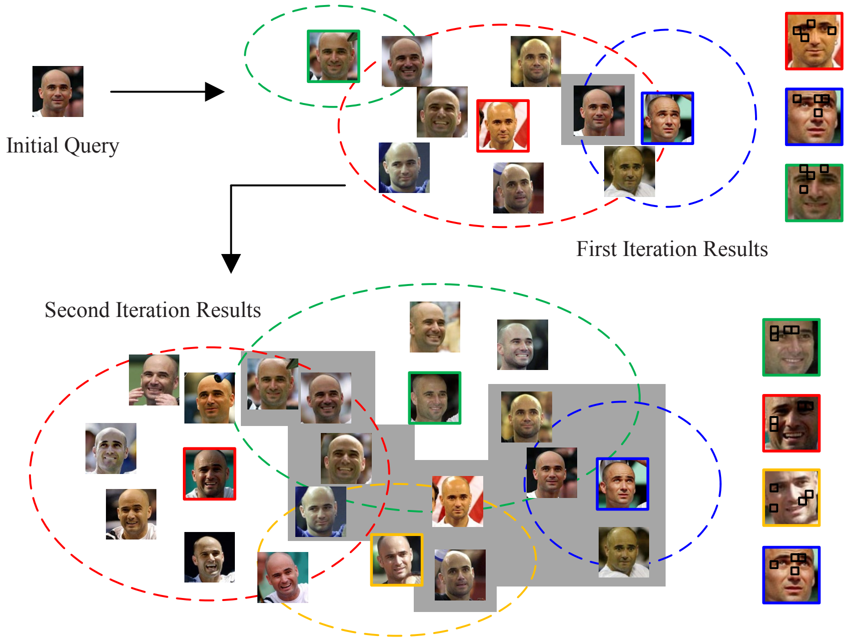
(d) Comparison of recall in Caltech 101 database when $Sc = 80$.

Fig. 6. Comparison of the four RF approaches on LFW and Caltech 101 databases

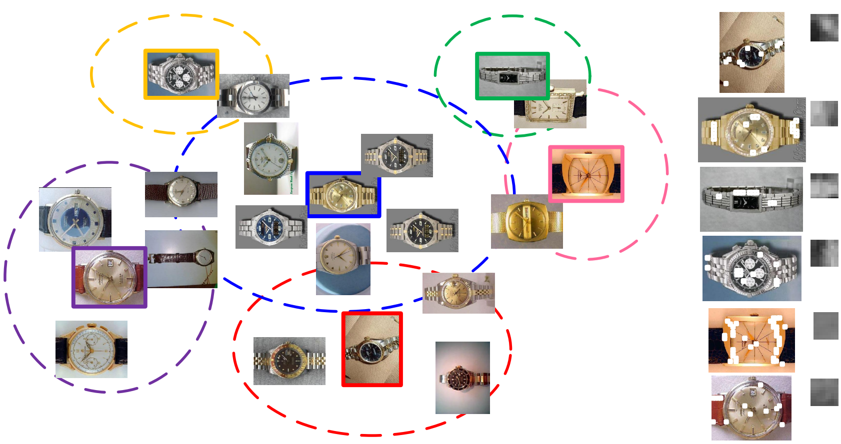
Table 1. Using Rank to evaluate different RF approaches' performance on LFW and Caltech 101 databases. The smallest rank in each experiment are in bold.

		$Sc = 40$					$Sc = 80$				
		It.=0	It.=1	It.=2	It.=3	It.=4	It.=0	It.=1	It.=2	It.=3	It.=4
LFW	Our Method	11.0	**10.3**	**11.6**	11.2	**16.1**	20.3	**18.5**	**18.8**	**14.4**	20.2
	QCluster	11.0	13.3	13.8	12.4	17.0	20.3	19.8	19.6	17.2	**15.9**
	Mindreader	11.0	11.7	12.6	11.9	20.3	20.3	20.6	19.2	18.1	18.4
	SVM	11.0	12.1	11.9	**10.7**	16.9	20.3	20.7	19.1	16.9	19.3
Caltech101	Our Method	13.2	**13.0**	14.2	**13.7**	16.3	30.6	28.8	**28.1**	**25.6**	**25.0**
	QCluster	13.2	15.3	14.9	15.2	**15.9**	30.6	**28.6**	29.0	29.1	27.7
	Mindreader	13.2	13.9	**13.3**	14.1	16.4	30.6	29.1	28.8	27.3	27.8
	SVM	13.2	14.7	15.0	15.3	16.9	30.6	30.3	29.8	28.9	27.4

One reason for our method's better performance in higher iterations is that, it uses the information in irrelevant images. With the increasing of iteration number, more and more irrelevant images are available as boundary image candidates. Thus the CRSC built by our method are more tight and more consistent with the human's perception. In our experiment, we find that when the iteration



(a) Visualization of face sample coverage. In the right part, the first four regions selected by RankBoost for the images in the CRSC are illustrated. We showed the retrieval process with 2 RF iterations, with one image of Agassi as the initial query. The images with shadow background means it has been retrieved in previous iterations.



(b) Visualization of object sample coverage. In the right part, the first texture code selected by RankBoost, and the position of these codes appear in the CRSC images are illustrated. Due to space limitation, we only show the final results with 2 RF iterations.

Fig. 7. Visualization of sample coverage. Images with colored frames form the Compact Relevant Sample Coverage in each figure. The dotted ellipses are coverage regions.

number is large, the number of images in the CRSC is usually larger than the number of clusters in the query expansion method. That is to say, the cluster centers are not competent to represent the whole query region.

Although SVM also uses irrelevant images, in the training procedure, it is likely that many relevant images are very similar to the initial query, and thus the classification boundary learned by SVM will be specialized to the initial query rather than human perception. This problem will not happen in our method, because the similar images are likely to be in a same coverage region, and only the images in CRSC will be used as new queries in the next iteration.

Furthermore, in most situations, images retrieved by our method have the smallest rank, as shown in Table 1. Actually, rank and recall are two contrary measures, because the more images are retrieved, the harder to make all these images rank high. Our method can result in better performance in terms of both recall and rank shows that, it not only retrieves the most relevant images, but also all those retrieved images are closer to the top than the other approaches.

In addition, from the aspect of time-cost, our RF mechanism implements faster than SVM, but slower than the query movement approach. Compared with query expansion, our method has the similar time-cost. That is to say, our approach can obtain much better performance without much extra time cost.

5.5 Visualization of Sample Coverage and Feature Re-weighting

Figure 7 shows a 2D visualization of sample coverage and feature re-weighting. For explicitness, only the coverage sets of images in the CRSC are illustrated. Irrelevant images which are used for coverage boundary are also not shown. These figures are obtained from realistic experimental results. The sample coverage are drawn as follows. First, we apply PCA to the image histograms. The projection values on the first two components are taken to get an initial image. Then, we manually move some of the images to make sure that, on the 2D planar, each image in the CRSC can cover all the images in its coverage set.

6 Conclusion

In this paper, we presented a novel RF scheme for CBIR. Our method explicitly aims at selecting a minimum subset of images to cover all the relevant images returned by the system. RankBoost learning is used to maximize each relevant image's coverage set, as well as obtaining a image-specific feature weight distribution. Future research will focus on two directions. One is to build an experiment scenario where user can give detailed relevance score to each relevant image. The other is to explore better image feature representation.

Acknowledgement

This work is supported in part by National Science Foundation of China under grant No. 60673107, National Basic Research Program of China under grant No. 2006CB303100, and a grant from Omron Corporation.

References

1. Rui, Y., Huang, T., Chang, S.: Image retrieval: Current techniques, promising directions, and open issues. J. Vis. Commun. Image R 10(4), 39–62 (1999)
2. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high level semantics. Pattern Recogn. 40, 262–282 (2007)
3. Mezaris, V., Kompatsiaris, I., Strintzis, M.G.: An ontology approach to object-based image retrieval. In: ICIP, vol. 2, pp. 511–514 (2003)
4. Jin, W., Shi, R., Chua, T.S.: A semi-naïve bayesian method incorporating clustering with pair-wise constraints for auto image annotation. In: ACMMM (2004)
5. Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: ACMMM, pp. 107–118 (2001)
6. Rui, Y., Huang, T., Mehrotra, S.: Content-based image retrieval with relevance feedback in mars. In: ICIP, vol. 2, pp. 815–818 (1997)
7. Ishikawa, Y., Subramanya, R., Faloutsos, C.: Mindreader: Querying databases through multiple examples. In: VLDB, pp. 208–217 (1998)
8. Porkaew, K., Chakrabarti, K.: Query refinement for multimedia similarity retrieval in mars. In: ACMMM, pp. 235–238 (1999)
9. Kim, D.H., Chung, C.W.: Qcluster: Relevance feedback using adaptive clustering for content-based image retrieval. In: SIGMOD, pp. 599–610 (2003)
10. Su, Z., Zhang, H., Li, S., Ma, S.: Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. IEEE T. Image Process 12(8), 924–937 (2003)
11. Sahbi, H., Audibert, J.Y., Keriven, R.: Graph-cut transducers for relevance feedback in content-based image retrieval. In: ICCV (2007)
12. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globally-consistent local distance functions for shape-based image retrieval and classification. In: ICCV (2007)
13. Bray, N.: Dominating set. In: Weisstein, E.W. (ed.) From MathWorld - A Wolfram Web Resource, `http://mathworld.wolfram.com/DominatingSet.html`
14. Salton, G., Fox, E.A., Wu, H.: Extended boolean information retrieval. Commun. ACM 26(11), 1022–1036 (1983)
15. Racine, K., Yang, Q.: Maintaining unstructured case bases. In: Leake, D.B., Plaza, E. (eds.) ICCBR 1997. LNCS, vol. 1266, pp. 553–564. Springer, Heidelberg (1997)
16. Freund, Y., Iyer, R., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. J. Mach. Learn. Res. 4, 933–969 (2003)
17. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report 07-49 (October 2007)
18. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from new training examples: An incremental bayesian approach tested on 101 object categories. In: WGMBV (2004)
19. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 469–481. Springer, Heidelberg (2004)