International Journal of Innovative Computing, Information and Control
Volume **7**, Number **8**, August **2011**
 ISSN 1349-4198
pp. 4977–4987

UNIFIED LEARNING PARADIGM FOR IMAGE RETRIEVAL

Zheng-Kui Lin, Jun Wu*, Zhi-Bo Xiao, Jing Duan and Ming-Yu Lu

School of Information Science and Technology
Dalian Maritime University
No. 1, Linghai Rd., Dalian 116026, P. R. China
*Corresponding author: wujunas8@gmail.com

Received April 2010; revised August 2010

Abstract. *Dealing with Relevance feedback (RF) using statistical learning has been a key technique to improve the content-based image retrieval (CBIR) performance. However, there is still a big room to further RF performance since the popular RF methods ignore the cooperation among various learning mechanisms. In this paper, we propose a unified learning paradigm (ULP) that integrates the merits of ensemble learning, semi-supervised learning, active learning and long-term learning into a uniform framework. Concretely, unlabeled examples are exploited to facilitate ensemble learning by helping augment the diversity among the base classifiers, and then, a strong ensemble is used to identify the most informative examples for active learning. In particular, the semantic clues are inferred in the long-term learning setting, which serves as the prior knowledge to validate the effectiveness of the unlabeled examples used by ULP. Finally, a bias-weighting strategy is developed to guide the ensemble of classifiers to pay more attention to the positive examples than the negative ones. An empirical study shows that using multiple learning strategies simultaneously in CBIR is beneficial, and that the proposed scheme is significantly more effective than some existing approaches.*
Keywords: Content-based image retrieval, Relevance feedback, Short-term learning, Long-term learning, Unified learning

1. **Introduction.** With the explosive growth of digital images, content-based image retrieval (CBIR) has drawn substantial research attention in the last decade [1]. In general, images are represented with visual features, such as color, texture and shape in CBIR systems. However, the gap between visual features and semantic concepts usually leads to poor performance. To narrow down the semantic gap, a few works focused on designing sophisticated methods to segment out the meaningful objects from an image [2,3]. However, it is impossible to achieve exact segmentation due to the rich content but subjective semantics of an image. Although it is feasible to bridge the semantic gap by building an image index with textual annotation [4,5], fully automatic image annotation is still a long way off. Relevance feedback, as an alternative and more promising way to mitigate the semantic gap issue, has been intensively investigated in recent years [6].

1.1. **Related works in relevance feedback.** Relevance feedback (RF) focuses on the interaction between the user and the search engine by letting the user provide feedback regarding the retrieval results, i.e., by labeling images returned as either positive or negative in terms of whether they are relevant to the query concept or not. From the interaction loop, the search engine is refined and the improved results are returned to the user. In essence, RF can be regarded as a statistical learning problem, and more precisely as a binary classification task between relevant and irrelevant classes. During the past years, many RF techniques based on statistical learning have been proposed, as for instance Bayesian learning, fuzzy sets, support vector machines (SVM) [7-9]. However, in CBIR systems, the labeled images are very limited (typically less than 20), whereas the database often contains thousands of images. Thus, learning algorithms has to handle classification with a few training examples, i.e., *small example issue*. This limitation has been addressed by latter research efforts and we classify them into following groups.

Learning with multi-classifier, or called ensemble learning, aims at mining the complementary information of multiple classifiers to achieve strong generalization performance. Tao et al. [10] trained a set of weak classifiers using different feature subspace and combined them by using Bagging technique. Then, the ensemble classification model was used to estimate the query concept. Wang et al. [11] independently trained and stored an individual classifier at each round of feedback, and then combined them using Adaboost technique for the retrieval task.

Learning with unlabeled data is another feasible way to achieve strong generalization performance. Semi-supervised learning and active learning are two main paradigms for this purpose. Semi-supervised learning focuses on exploiting a few confident unlabeled examples in conjunction with the labeled ones to improve the performance of learning system. For instance, Zhang et al. [7] proposed a stretching Bayesian method which regarded some unlabeled examples near the negative examples in a kernel space as additional negative examples and applied them to improve the estimation of the distribution of the irrelevant semantic class. Active learning aims to actively identify the most informative examples from unlabeled data and then query the user for labels, with the goal of achieving the maximal information gain in decision-making. For example, in SVM active learning [9], the images closest to the classification boundary are deemed as the most informative examples and asked to be labeled by the user, so as to maximally reduce the size of the version space. Moreover, many fusion methods of semi-supervised learning and active learning for CBIR have been reported [12-14], which further prove that exploiting unlabeled data is efficient to improve CBIR performance.

All of above-mentioned methods emphasize the learning within a single query session, so called short-term learning (STL). From a long-term learning (LTL) perspective, the accumulated users' relevance judgments stored in log data could be used as an important resource to aid the retrieval task. *Learning with log data* aims at mining the semantic clues across previous query sessions. He et al. [15] devoted to constructing a semantic space using log data, and a semantic correlation measure is then learned from the semantic space for image retrieval. Hoi et al. [16] utilized the user-specified relevant images as seeds to search through semantic space and obtain more positive examples for training SVM such that the future retrieval precision improves.

1.2. **Motivation of our work.** To achieve desirable results, multiple rounds of feedback are generally required for learning. As a result, the RF phase can be extremely time-consuming. Hence, it is necessary for CBIR system to achieve satisfactory results using as few labeled examples as possible. Although a few works have been developed to tackle the small example issue, there is still a big room to elevate the CBIR performance, because few works take the cooperation among various learning techniques into account within a unified framework.

This paper proposes a unified learning paradigm (ULP) that integrates several statistical learning techniques including ensemble learning, semi-supervised learning, active learning and long-term learning in a synergistic way to maximize the generalization capability of a learning system. Furthermore, two smart learning tricks, respectively named Bias-Sampling and Bias-Weighting mechanisms, are developed within ULP framework to facilitate the learning task in CBIR. Our empirical study shows encouraging results in comparison to some existing methods for interactive image retrieval. The rest of this paper is organized as follows: Section 2 provides an overview of the proposed framework for ULP problem, followed by a formal definition and a solution for the problem; Section 3 elaborates the core component of ULP – biased semi-ensemble learning scheme; Section 4 shows experimental evaluations; finally, Section 5 concludes this paper.

2. **Unified Learning Paradigm with Users' Feedback.**

2.1. **Overview of the proposed scheme.** We first give an overview of the proposed scheme that integrates variant learning techniques into a uniform framework with the goal of interacting with the user looking for image concepts in database, so termed as Unified Learning Paradigm, ULP for short. Figure 1 shows the architecture of the proposed ULP that can be roughly divided into four main modules.

(1). *Initialization*: Generally, a retrieval session is initialized using one query image brought by the user. However, only one example can hardly train a reliable classifier. Therefore, the k nearest neighbors (kNN) of the query point in the feature space are identified using Euclidean distances measure, and returned to the user for labeling, which are deemed initial labeled set for learning system.

(2). *Long-term Learning*: Given the labeled set and log data, the semantic correlation between each image and query can be estimated from previous retrieval sessions. It worth noting that semantic correlation is not enough to rank the whole image database since log data cannot cover all target images. In our solution, semantic correlation serves as the prior knowledge which is helpful to select the confident unlabeled examples in the semi-supervised setting.

(3). *Biased Semi-supervised Ensemble*: Ensemble learning and semi-supervised learning are usually employed for enhancing learning system. In this paper, we try to integrate the merits of these two techniques in order to achieve strong learning performance. Concretely, unlabeled data is exploited to facilitate ensemble learning by helping augment the diversity among individual classifiers. In particular, considering the asymmetric distribution between the relevant and irrelevant image classes, ULP processes positive and negative examples in different ways, whose details will be presented in Section 3.

For one thing, exploiting unlabeled data to enlarge positive set is more challenging than doing that for negative set in CBIR, because positive examples make up an extremely small proportion of the unlabeled set. The method presented in Section 3.1 employs the prior knowledge inferred from log data to validate the effectiveness of unlabeled examples before using them to enlarge the positive set. For another, in CBIR context, the user is more interested in positive examples rather than negative examples. The method presented in Section 3.2 can combine the individual classifiers with bias such that the generated ensemble would pay more attention on positive examples than negative examples.

(4). *Active Learning*: ULP does not passively wait for the user to choose image to label. Instead, it actively prepares a pool of images for feedback. To achieve active selection, the retrieval result is separated from the feedback pool. In each round of feedback, the images judged by ULP with low confidence are put into the feedback pool for the user to label. Once the user labeled these examples, the learning system can achieve the maximal information gain in decision making.

2.2. **Problem formulation and preliminaries.** To retrieve the desired images, a user must first pose a query image $\mathbf{q}$. Let $Z=\{\mathbf{z}_1, \cdots, \mathbf{z}_n\}$ denote the identity of target images. Let $\mathbf{X}=\{\mathbf{x}_1 \cdots \mathbf{x}_n\}$ denotes the image database, where each $\mathbf{x}_i$ is a feature vector of image $\mathbf{z}_i$. Let $\mathbf{R}=\{\mathbf{r}_1 \cdots \mathbf{r}_m\}$ denote the log dataset, where each $\mathbf{r}_i$ contains relevance judgments in the i-th log session. From the view of statistical learning, during the relevance feedback process, the image database can be regarded as a collection of labeled and unlabeled

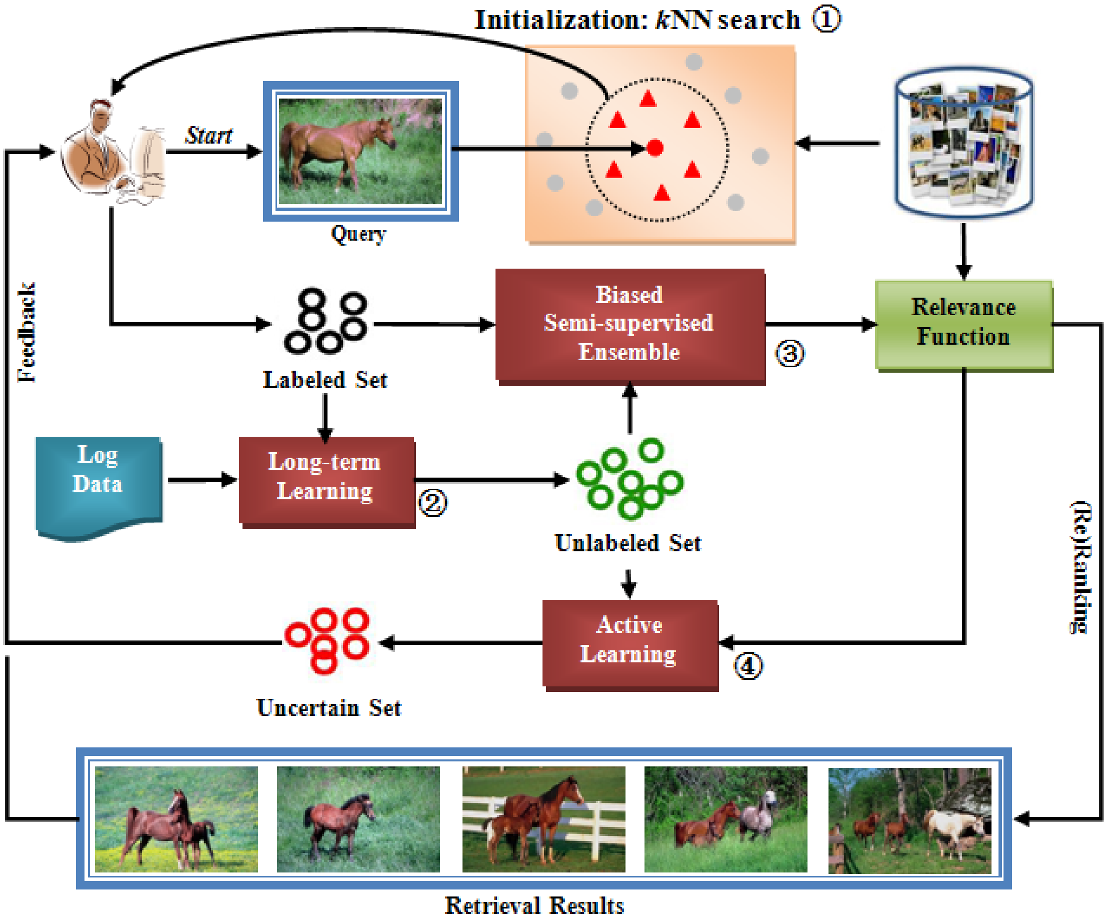

FIGURE 1. Framework of unified learning paradigm

images. Let $L=\{\langle \mathbf{z}_i, y_i\rangle \,|y_i \in \{-1,+1\}\}_{i=1}^{nl}$ and $U=\{\langle \mathbf{z}_i, y_i\rangle \,|y_i = 0\}_{i=1}^{nu}$ $(nl + nu = n)$ respectively denote the labeled and unlabeled image example set, where $y_i = +1$ denotes that $\mathbf{z}_i$ is labeled as positive, $y_i = -1$ denotes negative and $y_i = 0$ denotes unlabeled. Based on the above formulation, we now define the ULP problem as follows.

Definition 2.1. *ULP for interactive image retrieval is to look for a relevance function $f_{\mathbf{q}}$ that maps each image $\mathbf{z}_i$ to a real value of relevance degree within -1 and $+1$,*

$$f_{\mathbf{q}} : Z \mapsto [-1, +1]$$

based on the feature representation of images $\mathbf{X}$, the log dataset $\mathbf{R}$, and the training examples, including L and U, acquired from online feedback.

Let $\mathrm{Abs}(a)$ denote the function used to produce the absolute value of a real number a. In general, when $\mathrm{Abs}(f_{\mathbf{q}})$ for a target image $\mathbf{z}_i$ is high, the corresponding prediction confidence will be high. Similarly, a low $\mathrm{Abs}(f_{\mathbf{q}})$ of a target image means that the image is close to the decision boundary and its corresponding prediction confidence will be low. Thus, in our solution, images judged with highly positive confidence $(f_{\mathbf{q}}(\mathbf{z}_i) \approx +1)$ are regarded as the retrieved results, while the unlabeled images judged with low confidence $(\mathrm{Abs}(f_{\mathbf{q}}(\mathbf{z}_i)) \approx 0)$ are regarded as the informative examples which are prepared for the user to label.

2.3. **Solution to the problem.** Given L, a few semantic clues can be inferred from $\mathbf{R}$ in the LTL setting; meanwhile, a visual similarity between each target image and query concept can be learned from $\mathbf{X}$ in the STL setting. A good learning algorithm should achieve satisfactory results within only a few feedback steps, i.e., requiring a small number of labeled examples collected from the online interaction. ULP tries to employ both LTL and STL, with the goal of learning a valid relevance function by compensating the visual similarity using the semantic clues. In the following, we will describe how to achieve LTL and STL separately.

Let us first consider the LTL. Formally, the semantic space is represented by a relevance matrix $\mathbf{R}$ (say of size $m \times n$), where each column of such a matrix represents a target image, and each row represents a log session in past interactions. $r_{i,j}= +1$ denotes that an image is judged as relevant in a log session, similarly, $r_{i,j}=-1$ denotes that it is judged as irrelevant, and $r_{i,j}=0$ denotes that it is not judged. Given $L = L_+ \cup L_-$, for each target image $\mathbf{z}_i$, its semantic similarity to current query concept can be inferred from $\mathbf{R}$. We apply the simplified method [16] to compute the semantic similarity:

$$f_{\mathbf{R}}(\mathbf{z}_i) = \max_{k \in L_+} \{c_{k,i}\} - \max_{k \in L_-} \{c_{k,i}\} \tag{1}$$

where k are identities of images that belong to L_+ or L_-. The semantic correlation of each image $\mathbf{z}_i$ to the current query concept, $f_{\mathbf{R}}(\mathbf{z}_i)$, is determined using its overall semantic correlations to both positive and negative images. In the semantic space $\mathbf{R}$, the correlation between two images $\mathbf{z}_u$ and $\mathbf{z}_v$ can be estimated by:

$$c_{u,v} = \sum_{t=1}^{m} \delta_{t,u,v} \cdot r_{t,u} \cdot r_{t,v} \tag{2}$$

where $\delta_{t,u,v} = \begin{cases} 1, & r_{t,u} + r_{t,v} \geq 0 \\ 0, & r_{t,u} + r_{t,v} < 0 \end{cases}$ that is used to remove $(-1,-1)$ pairs among $(r_{t,u}, r_{t,v})$ in the computation of similarity because nothing can be inferred when two images were both marked as irrelevant. Based on above discussion, it is easy to conclude that a target image $\mathbf{z}_i$ is semantically relevant to current query concept when $f_{\mathbf{R}}(\mathbf{z}_i)$ is positive, similarly, irrelevant when $f_{\mathbf{R}}(\mathbf{z}_i)$ is negative.

After obtaining the semantic clues, we can use them in learning the visual similarity. Learning the visual similarity is a standard STL problem in CBIR. Dozens of algorithms are available [6]. Among various approaches, support vector machine (SVM) is one of the most effective techniques in practice. SVM select the optimal separating hyperplance which has the largest margin and, hence, the lowest vicinal risk. Also, it applies kernel trick to process the non-linear problem. As a state-of-the-art classification technique, SVM has shown superior performance in many applications. However, its performance will deteriorate significantly for its application to CBIR due to the limited training examples. To generate strong learning performance, ULP focuses on dealing with SVMs ensemble in the semi-supervised setting. Furthermore, two biased learning strategies are developed to process the asymmetric distribution between the relevant and irrelevant semantic classes. In the following section, we first introduce the semi-supervised ensemble framework and then elaborate the two biased learning mechanisms.

3. **Training Biased SVMs Ensemble in the Semi-supervised Setting.** To generate strong learning systems, ensemble learning tries to mine the complementary information of multiple classifiers, while semi-supervised learning aims to benefit from the unlabeled data. Indicated by Zhou [17], however, ensemble learning and semi-supervised learning are actually mutually beneficial, a key element is that exploiting unlabeled data in ensemble is helpful to augment the diversity among the individual classifiers. Based on this viewpoint, we introduce Semi-supervised Boosting (SemiBoost) technique [18-20] into our solution and modify it for CBIR purpose.

Similar to boosting, the main idea of SemiBoost is to train an ensemble classifier iteratively. At each round of iterations, the pseudo-labels of the unlabeled examples are predicted using existing ensemble and the pairwise similarity between examples, and then a few confident pseudo-labeled examples in conjunction with all labeled ones are used to train a new classifier. Finally, all of the learned classifiers will be combined to form the final ensemble. An outline of the SemiBoost algorithm is presented in Figure 2.

Input: labeled examples, unlabeled examples, supervised learning method.
Parameter: T – the number of individual classifiers used in ensemble.
Output: ensemble

- Start with an empty ensemble;
- **for** $t = 1 \ldots T$
  - Compute the pseudo-label and its confidence for each unlabeled example (using existing ensemble and the pairwise similarity);
  - Sample the most confident pseudo-labeled examples; combine them with the labeled examples and train an individual classifier using a supervised learning method;
  - Update the ensemble by including the individual classifier with an appropriate weight.
- **end for**

FIGURE 2. An outline of the SemiBoost algorithm

Let $\mathbf{S} = [S_{i,j}]^{n\times n}$ denote the symmetric similarity matrix, where $S_{i,j} \geq 0$ represents the similarity between example $\mathbf{x}_i$ and $\mathbf{x}_j$. Let $\mathbf{y} = [\mathbf{y}_l; \mathbf{y}_\mathrm{u}]$ denote the labels of entire image database. Let $h^{(\mathrm{t})}(\mathbf{x}) : X \to \{-1, +1\}$ denote the individual classifier learned at the t-th iteration. Let $H(\mathbf{x}) : X \to R$ denote the combined classification model learned after the T iterations. It is computed as a linear combination of the T individual classifiers, i.e., $H(x) = \sum_{t=1}^{T} \alpha_t h^{(t)}(x)$ where α_t is the combination weight. At the $(T+1)$-st iteration, SemiBoost aims to find a new component classifier $h(\mathbf{x})$ and the combination weight α by solving the following optimization problem:

$$\begin{aligned}
&\underset{h(\mathbf{x}),\alpha}{\mathbf{argmin}} \left(\mathbf{F}(\mathbf{y}, \mathbf{S}) = \mathbf{F_l}\left(y, \mathbf{S}^{lu}\right) + \mathbf{CF_u}(\mathbf{y_u}, \mathbf{S}^{uu})\right) \\
&\Leftrightarrow \underset{h(\mathbf{x}),\alpha}{\mathbf{argmin}} \left(\sum_{i=1}^{nl} \sum_{j=1}^{nu} S_{i,j}^{lu} \exp\left(-2y_i^l (H_j + \alpha h_j)\right) \right. \\
&\quad \left. + C \sum_{i,j=1}^{nu} S_{i,j}^{uu} \exp(H_i - H_j) \exp(\alpha (h_i - h_j)) \right) \\
&\quad \text{s.t.} \quad h_i = y_i^l, \quad i = 1 \cdots nl
\end{aligned} \tag{3}$$

where $H_i \equiv H(\mathbf{x}_i)$, $h_i \equiv h(\mathbf{x}_i)$ and the constant $C = nl/nu$ is introduced to weight the importance between the labeled and the unlabeled data.

To simplify the computation, the above optimization problem can be transformed to a simple format. In detail, by substituting $H_i \leftarrow H_i + \alpha h_i$ into F and regrouping the terms, an equivalent and simplified optimization problem can be obtained.

$$\mathrm{F}_1 = \sum_{i=1}^{nu} \exp(-2\alpha h_i) p_i + \exp(2\alpha h_i) q_i \tag{4}$$

$$p_i = \sum_{j=1}^{nl} S_{i,j}^{ul} \exp\left(-2H_i\right) \delta\left(y_j, 1\right) + \frac{C}{2} \sum_{j=1}^{nu} S_{i,j}^{uu} \exp\left(H_j - H_i\right) \tag{5}$$

$$q_i = \sum_{j=1}^{nl} S_{i,j}^{ul} \exp\left(2H_i\right) \delta\left(y_j, -1\right) + \frac{C}{2} \sum_{j=1}^{nu} S_{i,j}^{uu} \exp\left(H_i - H_j\right) \tag{6}$$

where $\delta(x, y) = 1$ when $x = y$ and 0 otherwise. The quantities p_i and q_i can be interpreted as the confidence in classifying the unlabeled example $\mathbf{x}_i$ into the positive class and the negative class, respectively. Since F_1 is difficult to optimize, its upper bound F_2 is then constructed. More details can be found in [19].

$$\mathrm{F}_1 \leq \mathrm{F}_2 = \sum_{i=1}^{nu} \left(p_i + q_i\right) \left(\exp\left(2\alpha\right) + \exp\left(-2\alpha\right) - 1\right) - \sum_{i=1}^{nu} 2\alpha h_i \left(p_i - q_i\right) \tag{7}$$

Obviously, F_2 is linear in $h_i(p_i - q_i)$ and is minimized when $h_i = \mathrm{sign}(p_i - q_i)$, for maximum values of $\mathrm{Abs}(p_i - q_i)$. Therefore, to minimize F_2, the optimal pseudo-label c_i for the example $\mathbf{x}_i$ is $c_i = \mathrm{sign}(p_i - q_i)$ and its corresponding prediction confidence is $\mathrm{Abs}(p_i - q_i)$. Also, by differentiating F_2 with regard to α and setting it to 0, the optimal α that minimizes the objective function is

$$\alpha = \frac{1}{4} \ln \frac{\sum_{i=1}^{nu} p_i \delta\left(h_i, 1\right) + \sum_{i=1}^{nu} q_i \delta\left(h_i, -1\right)}{\sum_{i=1}^{nu} p_i \delta\left(h_i, -1\right) + \sum_{i=1}^{nu} q_i \delta\left(h_i, 1\right)}. \tag{8}$$

Input: L – labeled example set, U – unlabeled example set, $\mathbf{R}$ – log data.
Parameter: σ – sampling scale used in boosting
Output: L^* – sampled example set.

- $L^* = \emptyset$;
- **while** $|L^*| < m = [\sigma \cdot |U|]$
  - Compute p_i and q_i for every unlabeled example using Equations (5) and (6)
  - Compute the pseudo-label $c_i = \mathrm{sign}(p_i - q_i)$ for each example
  - Single out example $\mathbf{x}_i$ from U by the weight $\mathrm{Abs}(p_i - q_i)$
  - **if** $c_i = +1$ **then**
    - * compute the semantic correlation $f_{\mathbf{R}}(z_i)$ using Equation (1)
    - * **if** $f_{\mathbf{R}}(z_i) \geq 0$ **then** $L^* \leftarrow L^* \cup \{\langle \mathbf{x}_i, c_i \rangle\}$ **end if**
  - **else** $L^* \leftarrow L^* \cup \{\langle \mathbf{x}_i, c_i \rangle\}$ **end if**
- **end while**

FIGURE 3. Bias-sampling mechanism

3.1. **Bias-sampling the unlabeled examples.** Given any query concept, the relevant target images are much less than the irrelevant ones. Besides, the classifiers learned during the online RF process are not strong, especially at the beginning of retrieval. As a result, singling out the positively confident examples from unlabeled data set only depending upon the classifiers' prediction is unreliable. Almost none of existing semi-supervised techniques can deal with this problem [21]. In contrast, selecting negatively confident examples is relatively easy since irrelevant images make up an extremely large proportion of the existing database. Considering this, we propose a Bias-Sampling strategy used by ULP for retrieval purpose, which is summarized in Figure 3.

Actually, ULP pays more attention on the positively pseudo-labeled examples than the negatively pseudo-labeled examples during the sampling process. Roughly speaking, it selects unlabeled examples for relevant class depending upon both classifiers' prediction and semantic clues inferred from log data, while do that for irrelevant class only using classifiers' prediction. This Bias-Sampling strategy aims at validating the effectiveness of the positively pseudo-labeled examples by means of the semantic clues, in order to compensate the weak predictions of learned classifiers.

3.2. **Bias-weighting the individual classifiers.** As mentioned before, relevant images are much less than irrelevant images in the database and thus positive examples will be also less than negative examples in the sampled example set. However, SemiBoost fails to take this class-imbalance problem into account, and thus it tends to be overwhelmed by the majority (irrelevant) class and ignore the minority (relevant) class.

In CBIR context, the user is more interested in positive images rather than negative images. Hence, individual classifiers with high *true positive rate* should be emphasized. Considering this, we propose a Bias-Weighting mechanism used by ULP for retrieval purpose.

$$\begin{aligned} \alpha &= \frac{\eta}{4}\ln\frac{\sum_{i=1}^{nu} p_i\delta\left(h_i,1\right)+\sum_{i=1}^{nu} q_i\delta\left(h_i,-1\right)}{\sum_{i=1}^{nu} p_i\delta\left(h_i,-1\right)+\sum_{i=1}^{nu} q_i\delta\left(h_i,1\right)} + (1-\eta)\exp\left(tpr\right) \\ tpr &= \Pr\left[h\left(\mathbf{x}_j\right)=y_j \ \&\ y_j=+1\right], \quad j=1,\cdots,nl+m \end{aligned} \tag{9}$$

where tpr denotes the true positive rate of classifier $h(\mathbf{x})$ learned from a mixture of nl labeled examples and m pseudo-labeled examples. $\exp(tpr)$ is used to augment the relative contribution of tpr. Under the influence of Bias-Weighting, the ensemble pays more attention on positive images than negative ones. $\eta \in (0, 1]$ is used to control the relative contribution of each component.

4. **Experiments and Discussions.** For the evaluation purpose, we pick 5000 real-word images from COREL collection with 50 semantic categories. Two kinds of low-level features are used to describe images. (1) **Color**: The color features are derived using $4\times4\times4$ bins histogram in HSV space. (2) **Texture**: the texture features are derived using 3-level pyramidal wavelet transform (PWT) from the Y component in YCbCr space, and then the mean and variance calculating in each of 9 high-frequent sub-bands is used to form a 18-dimension vector.

To demonstrate the effectiveness of the proposed ULP, we compare it with three other well-known RF approaches: SVM Active Learning (SVM-AL) [9], Boost SVM Active Learning (BSVM-AL) [11] and Transductive SVM Active Learning (TSVM-AL) [12]. Furthermore, in order to study whether the Bias-Sampling and Bias-Weighting strategies used by our algorithm are useful, two degenerated variant of ULP, i.e., ULP-d1 and ULP-d2, are evaluated in the comparison. Unlike ULP, the ULP-d1 directly uses the original weighting strategy used by SemiBoost. In detail, ULP aims to emphasize the importance of positive examples by setting $\eta = 0.3$ in Equation (9) while ULP-d1 regards the positive and negative examples equally, i.e., $\eta = 1$. Based on ULP-d1, the ULP-d2 can be further obtained by omitting the Bias-Sampling process, i.e., ULP-d2 samples the unlabeled examples in Boosting iterations only depending upon the classifiers' prediction.

Initially 10 images are presented to the user for labeling. After obtaining the labeled images, four learning algorithms are employed to rerank the image database separately. For each scheme, five rounds of feedback are conducted and 10 images are labeled by the user in each round. We adopted the *experimental design* technique to select the optimal values of parameters T and σ in Figures 2 and 3. The feasible values of them are set to $\{5, 10, 15, 20\}$ and $\{5\%, 10\%, 15\%, 20\%\}$, respectively. In experiment, we found that, with the growing of T and σ, the performance of ULP improves slowly while the computational time increases quickly. Considering the tradeoff between the effectiveness and the computational complexity, T and σ are set to 5 and 5%, respectively.

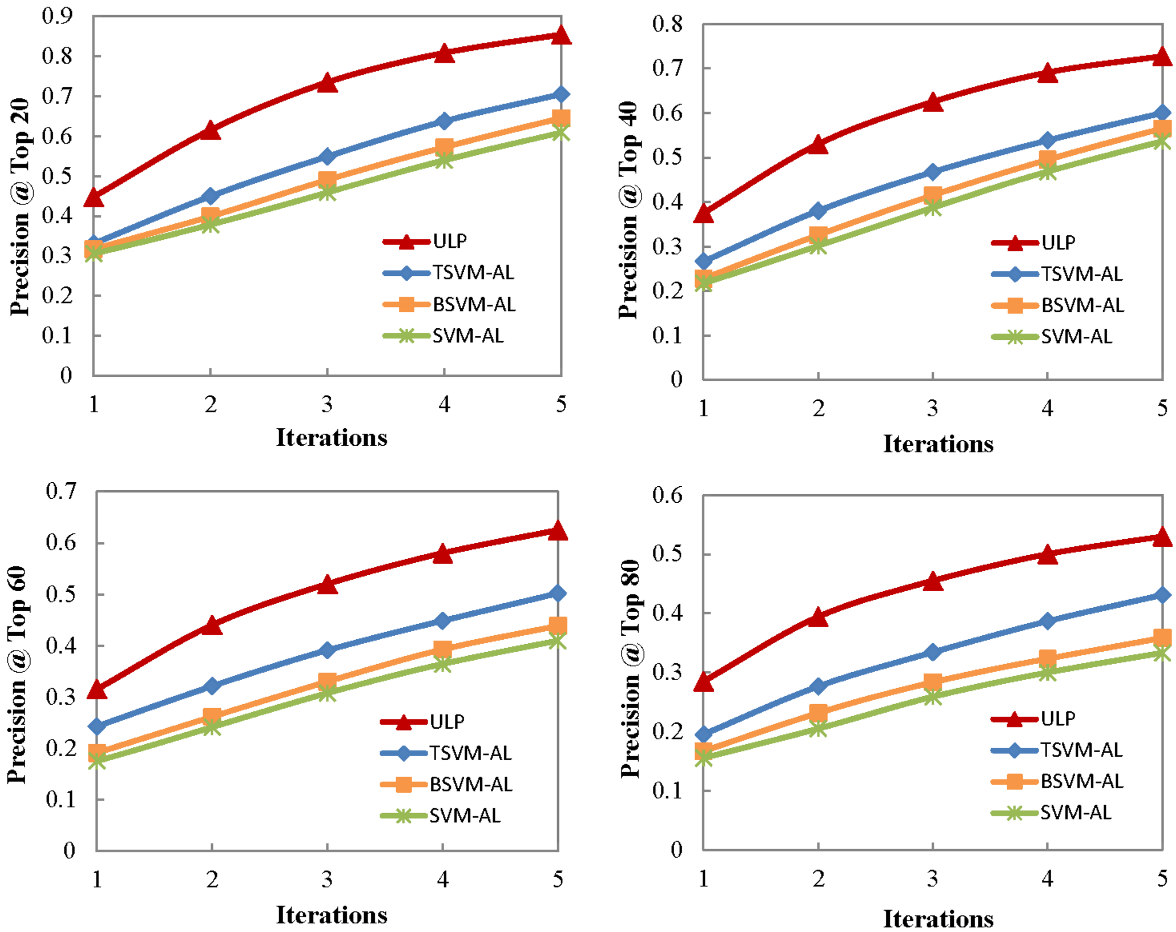

FIGURE 4. Performance of the proposed algorithm compared with some existing algorithms

At first, the performance of ULP, SVM-AL, BSVM-AL and TSVM-AL are compared. Figure 4 shows the precision curves of the different methods at the top 20, top 40, top 60 and top 80 retrieval results. Several observations can be drawn from the experimental results. First, by examining the results with all methods, we found that all methods outperform the baseline SVM-AL, which demonstrates that both ensemble learning and semi-supervised learning are beneficial to improve the retrieval performance of a CBIR system. Second, however, the ensemble solution BSVM-AL is only marginally better than the baseline method SVM-AL. The main reason is that the AdaBoost may degenerates to a single strong classifier due to the over-fitting caused by the limited number of labeled examples. Finally, by comparing the two semi-supervised solutions, it is impressive that the performance of ULP is always the best, which shows that integrating the merits of various leaning techniques is more effective than only using a single solution to deal with the small example issue.

Furthermore, in order to study whether the two biased learning strategies are helpful or not, the performance of ULP, ULP-d1 and ULP-d2 are compared. Figure 5 shows the comparison results (Precision @ Top 50) of the three algorithms. As can been seen, the performance of ULP and ULP-d1 are much better than ULP-d2, which illustrates that using the semantic clues is helpful to enhance the performance of short-term learning. What's more, ULP increasingly outperforms ULP-d1 with the rounds of feedback growing. It is conjectured that positive and negative examples are nearly balance at the early rounds of feedback. By gradually adding the user's feedbacks, the positive and negative examples become imbalance and thus Bias-Weighting strategy contributes more.

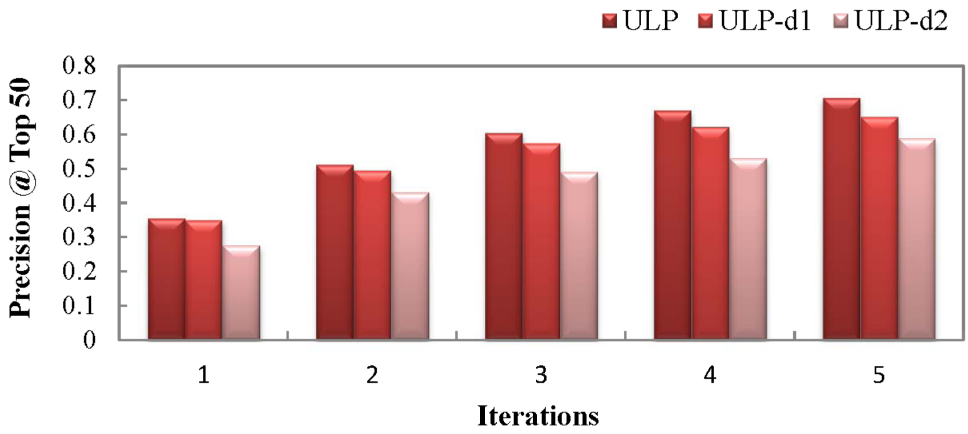

Figure 5. Performance the proposed algorithm compared with its degenerated variants

5. Conclusions.

In this paper, we proposed a novel RF scheme that integrates the merits of ensemble learning, semi-supervised learning, active learning and long-term learning to address the small example problem. In particular, two biased learning strategies are used within our framework to deal with the asymmetric distribution between the relevant and irrelevant classes. The empirical results showed the advantages of the proposed solution compared with some existing methods. In future, we will study more efficient solution to reduce the redundancy among the informative examples by using clustering technique.

Acknowledgment. This research was supported in part by Natural Science Foundation of China No. 60973067 and No. 61073133 and the Liaoning Provincial National Natural Science Foundation (No. 20092145).

REFERENCES

[1] R. Datta, D. Joshi, J. Li and J. Z. Wang, Image retrieval: Ideas, influences, and trends of the new age, *ACM Computing Surveys*, vol.40, no.2, pp.1-60, 2008.
[2] O. Marques, L. Mayron, G. Borba et al., Using visual attention to extract regions of interest in the context of image retrieval, *Proc. of ACM Int. Conf. Multimedia*, pp.638-643, 2006.
[3] F. Jing, M.-J. Zhang, H.-J. Zhang and B. Zhang, An efficient and effective region-based image retrieval framework, *IEEE Trans. Image Processing*, vol.35, no.5, pp.699-709, 2004.
[4] J. Li and J. Z. Wang, Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, no.9, pp.1075-1088, 2003.
[5] R. Ji, X. Lang, H. Yao and Z. Zhang, Semantic sensitive region retrieval using keyword-integrated Bayesian reasoning, *International Journal Innovative Computing, Information and Control*, vol.3, no.6(B), pp.1645-1656, 2007.
[6] X. Zhou and T. S. Huang, Relevance feedback in image retrieval: A comprehensive review, *ACM Multimedia Syst. J.*, vol.8, pp.536-544, 2003.
[7] R. Zhang and Z. Zhang, BALAS: Empirical Bayesian learning in the relevance feedback for image retrieval, *Image and Vision Computing*, vol.24, pp.211-223, 2006.
[8] H. Zhang and Y. Zhang, Fuzzy set based image retrieval by relationship of objects, *ICIC Express Letters*, vol.3, no.3(B), pp.733-738, 2009.

[9] S. Tong and E. Chang, Support vector machine active learning for image retrieval, *Proc. of ACM Int. Conf. Multimedia*, pp.107-118, 2001.
[10] D. Tao, X. Tang, X. Li and X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, no.7, pp.1088-1099, 2006.
[11] W. Jiang, G. Er and Q. Dai, Boost SVM active learning for content-based image retrieval, *Proc. of Asilomar Conf. Signals, Systems and Computers*, pp.1585-1589, 2003.
[12] L. Wang, K. L. Chan and Z. Zhang, Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp.629-634, 2003.
[13] Z. H. Zhou, K. J. Chen and Y. Jiang, Exploiting unlabeled data in content-based image retrieval, *Proc. of European Conf. Machine Learning, LNAI*, vol.3201, pp.219-244, 2004.
[14] J. Wu and M. Y. Lu, Asymmetric semi-supervise boosting scheme for interactive image retrieval, *ETRI Journal*, vol.32, no.5, pp.766-773, 2010.
[15] X.-F. He, O. King, W.-Y. Ma, O. King, M. Li and H. J. Zhang, Learning a semantic space from user's relevance feedback for image retrieval, *IEEE Trans. Circuits and Systems for Video Technology*, vol.13, no.1, pp.39-48, 2003.
[16] S. C. H. Hoi, M. R. Lyu and J. Rong, A unified log-based relevance feedback scheme for image retrieval, *IEEE Trans. Knowledge and Data Engineering*, vol.18, no.4, pp.509-524, 2006.
[17] Z.-H. Zhou, When semi-supervised learning meets ensemble learning, *Proc. of Int. Workshop Multiple Classifier System, LNCS*, vol.5519, pp.529-538, 2009.
[18] K. Chen and S. Wang, Regularized boost for semi-supervised learning, *Advances in Neural Information Processing Systems*, vol.20, pp.281-288, 2007.
[19] P. K. Mallapragada, R. Jin, A. K. Jain and Y. Liu, SemiBoost: Boosting for semi-supervised learning, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.11, pp.2000-2014, 2009.
[20] L. Zheng, S. J. Wang, Y. Liu and C.-H. Lee, Information theoretic regularization for semi-supervised boosting, *Proc. of ACM Int. Conf. Knowledge Discovery and Data Mining*, pp.1017-1025, 2009.
[21] J. Wu, M.-Y. Lu and C.-L. Wang, Collaborative learning between visual content and hidden semantic for image retrieval, *Proc. of IEEE Int. Conf. on Data Mining*, Sydney, Australia, pp.1144-1138, 2010.