

Integrating User Feedback Log into Relevance Feedback by Coupled SVM for Content-Based Image Retrieval

Steven C. H. Hoi and Michael R. Lyu

Dept. of Computer Science & Eng. and Shun Hing Institute of Advanced Eng.
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong S.A.R.
{chhoi,lyu}@cse.cuhk.edu.hk

Rong Jin

Dept. of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824, U.S.A.
rongjin@cse.msu.edu

Abstract

Relevance feedback has been shown as an important tool to boost the retrieval performance in content-based image retrieval. In the past decade, various algorithms have been proposed to formulate relevance feedback in content-based image retrieval. Traditional relevance feedback techniques mainly carry out the learning tasks by focusing low-level visual features of image content with little consideration on log information of user feedback. However, from a long-term learning perspective, the user feedback log is one of the most important resources to bridge the semantic gap problem in image retrieval. In this paper we propose a novel technique to integrate the log information of user feedback into relevance feedback for image retrieval. Our algorithm's construction is based on a coupled support vector machine which learns consistently with the two types of information: the low-level image content and the user feedback log. We present a mathematical formulation of the problem and develop a practical algorithm to solve the problem effectively. Experimental results show that our proposed scheme is effective and promising.

1. Introduction

With the rapid growth of image and video data, visual information retrieval has attracted more and more attention in both the research and the industrial communities. Content-based image retrieval (CBIR) is one of the most popular and challenging problems in visual information retrieval [27].

Although this has attracted extensive research effort over many years, finding a desired image from large multimedia databases is still challenging nowadays. Early studies in CBIR mainly focused on low-level feature extraction and similarity measurement [27]. However, due to the complexity of image understanding, it is almost impossible to discriminate images simply by distance measurements on low-level features. One well-known factor causing difficulty is the semantic gap problem between low-level features and high-level human semantic concepts [13, 25]. To attack the challenging semantic gap problem, one feasible way is to build a textual description index for images. However, making a textual index manually in large image databases is time-consuming and too costly in practice. Although some automatic image annotation techniques have been studied recently [3, 7, 17, 20], really practical auto-annotation techniques are still a long way off. In order to tackle the problem, relevance feedback has been proposed as an alternative technique to narrow the semantic gap by means of learning with user feedback. Many research studies have shown that relevance feedback is a powerful tool for boosting the retrieval performance in CBIR [8, 10, 14, 23, 24, 25].

In the past years various relevance feedback techniques from heuristic methods to all kinds of sophisticated learning techniques have been suggested and studied [5, 11, 15, 19, 29, 33]. Relevance feedback has already considered a key component when designing a CBIR system. In general, a relevance feedback mechanism requires users' relevance judgements on the results returned by the CBIR system in respect to the original query. When a user has made the relevance assessment on the initial retrieval results, relevance

feedback is then engaged as a query refinement method to improve the retrieval results. Since the learning task of relevance feedback is very tough, it is usually necessary to repeat many rounds of feedback in order to achieve satisfactory results. Hence, the learning task of relevance feedback can be a very time-consuming procedure.

Furthermore, the task of providing the relevance marking of images in relevance feedback is a tedious and boring job for the users. Thus, it is advantageous for the retrieval task using relevance feedback in a CBIR system to achieve satisfactory results within as few feedback cycles as possible. Although some research studies have suggested employing active learning techniques to speed up the relevance feedback procedure [30], traditional techniques for relevance feedback may not be able to tackle the problem well when few of the results returned are actually relevant.

However, from a long-term learning perspective, it is reasonable to assume the log information of user feedback is available in a CBIR system as an important supplemental resource for the learning task of relevance feedback. Thus, a challenging problem is to determine how can we integrate the user log information into the learning task of relevance feedback for CBIR, i.e. the log-based relevance feedback problem [12]. In this paper, we suggest treating the problem as a multiple-modality learning task, i.e. learning both on the low-level image content and the log information of user feedback. In order to learn the two types of information consistently, we propose a coupled support vector machine technique to attack the learning task in this paper.

The rest of this paper is organized as follows. We review and discuss the problem of log-based relevance feedback in Section 2. Section 3 provides a brief review of the Support Vector Machine (SVM) technique. Section 4 presents a coupled support vector machine for learning on data with multiple types of information. Section 5 proposes an effective algorithm for log-based relevance feedback using the coupled SVM. Section 6 presents detailed experiment and performance comparison, and addresses some practical problems. Section 7 describes some work related to this paper. Section 8 sets out our conclusion and discusses our future work.

2. Log-based Relevance Feedback Problem

Relevance feedback is an important component for a CBIR system. In general, when a user submits his/her query target, the CBIR system will return a set of similar images to the user. The images returned initially may not be fully relevant to the query target of the user. In order to learn the user’s query concept, relevance feedback is employed as a query refinement method to help the retrieval task. The relevance feedback mechanism solicits the user to judge the relevance of the retrieved images and then refines the results

by learning the feedbacks given by the user. The relevance feedback procedures are repeated again and again until the targets are found. Since the semantic gap problem in CBIR is very challenging, regular techniques normally need a lot of rounds of feedback for achieving satisfactory results. To assist the learning task, we propose engaging the log information of user feedback into relevance feedback for the retrieval tasks from a long-term learning perspective [12].

In order to integrate user log information into the learning task of relevance feedback effectively, the first step is to organize the user log information well. In general, when a user launches a query in a CBIR system, he/she may choose to begin a relevance feedback learning procedure if he/she cannot obtain the desired targets from the initial results. To quantify the log information, a typical relevance feedback round can be viewed as a unit of user log session. For each user log session, suppose there are N_l images returned to be judged by users, which are marked as relevant or irrelevant. The relevant and irrelevant images are respectively recorded in the log database as “+1” (positive) and “-1” (negative). To manage the log information well, a relevance matrix is constructed to describe the relevance information. Each user log session is represented as a row in the relevance matrix. The appropriate element in the matrix is represented as “+1” and “-1” for relevant and irrelevant samples while it is given as “0” for unknown status by default.

More formally, let $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$ be the collection of images in image retrieval, where N is the number of images in the image database. Let the first N_l images be the samples labeled by users, and \mathcal{S}_l be the set of labeled images, i.e. $\mathcal{S}_l = \{(\mathbf{z}_i, y_i)\}_{i=1}^{N_l}$, where y_i is the label of image \mathbf{z}_i . Let N' be the number of unlabeled images, here we assume N is the sum of N_l and N' . Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ represent the low-level information of image content. Correspondingly, the log information of user feedback can be organized formally. Let $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ be the relevance representation of user log information, in which each column corresponds to an image in the image database and each row represents a user log session in the log database. Each element $r_{i,j}$ indicates the relevance judgement made about the i -th image during the j -th user log session (“+1” and “-1” for relevant and irrelevant, and “0” for unknown). Based on this representation, each image corresponds to a user log vector \mathbf{r}_i , whose dimension M is the total number of user log sessions collected.

Therefore, the log-based relevance feedback problem is that, given a query \mathbf{q} and initial labeled collection \mathcal{S}_l , how can we integrate the user log information \mathbf{R} into the learning task with the low-level image content \mathbf{X} consistently? This could be considered as a task of learning consistently on the data with two types of information: the low-level image content and the user feedback log. In this paper we pro-

pose a coupled support vector machine technique to tackle this problem. Before going into details, we first briefly introduce some basic background of SVM and then present the formulation of coupled SVM in the subsequent sections.

3. Support Vector Machine

Support vector machine (SVM), a state-of-the-art discriminative learning technique, has already achieved many successes in various empirical pattern recognition applications, drawing on its superior generalization performance. It has a sound theoretical foundation based on Structural Risk Minimization instead of Empirical Risk Minimization [32]. Let us introduce the basic concept of SVM.

Suppose we are given a set of labeled training data $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$ in a binary classification task, where \mathbf{x}_i are the data vectors in some input space $\mathcal{X} \subseteq \mathbb{R}^n$, l is the number of training data instances, and $y_i \in \{+1, -1\}$ are the class labels. In the simplest situation, the learning goal of SVM is to find a separating hyperplane that separates the training data with a maximal margin. The primal form of SVM in a linear kernel setting can be expressed as:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & \forall_{i=1}^l : y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0. \end{aligned}$$

The optimization of SVM is usually solved in a dual form as follows:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to} \quad & \sum_{i=1}^l \alpha_i y_i = 0, 0 \leq \alpha_i \leq C. \end{aligned}$$

In general, we can project the training data from the original data space \mathcal{X} to a higher dimensional feature space \mathcal{F} by a Mercer kernel K . The kernel K , which satisfies a Mercer's condition [32], can be represented as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, where $\Phi(\cdot)$ is a mapping function given by $\Phi : \mathcal{X} \mapsto \mathcal{F}$, where “ \cdot ” denotes an inner product. Therefore, the decision boundary of SVM with a kernel setting can be represented as: $f(x) = \mathbf{w} \cdot \Phi(\mathbf{x})$, where $\mathbf{w} = \sum_{i=1}^l \alpha_i \Phi(\mathbf{x}_i)$.

4. Coupled Support Vector Machine

4.1. Formulation

Without losing generality, we formalize the coupled SVM for learning on data with two types of information. It can be naturally generalized for learning on a

multiple-modality problem. Let us consider a task for learning on data with two types of information: the low-level image content and the user feedback log. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ represent the image content information and $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ be the user log relevance information, as described in Section 2. In a regular SVM based relevance feedback algorithm [30], only the low-level features of image content is considered. Typically, a vector \mathbf{w} is introduced to learn the weights of image features, such that the magnitudes of $\mathbf{w}^T \mathbf{X}$ represent the relevance degrees of images to the given query \mathbf{q} . Formally, learning the optimal solution by SVM can be formulated as follows:

$$\min_{\mathbf{w}, b_w, \xi} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C_w \sum_{k=1}^{N_l} \xi_k$$

$$\text{subject to} \quad \forall_{k=1}^{N_l} : y_k(\mathbf{w}^T \mathbf{x}_k + b_w) \geq 1 - \xi_k, \xi_k \geq 0.$$

Similarly, for the log information, we can also introduce a vector \mathbf{u} as the weights assigned to different user log sessions such that the magnitudes of $\mathbf{u}^T \mathbf{R}$ represent the relevance degrees of images to the given query q . Hence, a maximum margin based approach can also be formulated as follows:

$$\min_{\mathbf{u}, b_u, \eta} \quad \frac{1}{2} \|\mathbf{u}\|^2 + C_u \sum_{k=1}^{N_l} \eta_k$$

$$\text{subject to} \quad \forall_{k=1}^{N_l} : y_k(\mathbf{u}^T \mathbf{r}_k + b_u) \geq 1 - \eta_k, \eta_k \geq 0.$$

The straightforward approach to integrate the user feedback log with the low-level image content is to learn two modalities respectively and then sum up their results. Such an approach is feasible but it may lose some coupling information. In order to construct a unified framework that can combine the two types of information, we can put the two objective functions above together, meanwhile force the relevance prediction based on the two types of information to be consistent. More precisely, this idea can be formulated into the following optimization:

Minimize over $(\mathbf{w}, \mathbf{u}, b_w, b_u, \xi, \xi', \eta, \eta', \mathbf{Y}')$:

$$\begin{aligned} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} \|\mathbf{u}\|^2 + C_w \sum_{i=1}^{N_l} \xi_i + C_u \sum_{i=1}^{N_l} \eta_i \\ + \rho C_w \sum_{j=1}^{N'} \xi'_j + \rho C_u \sum_{j=1}^{N'} \eta'_j \end{aligned} \quad (1)$$

$$\text{subject to} \quad \forall_{i=1}^{N_l} : y_i(\mathbf{w}^T \mathbf{x}_i + b_w) \geq 1 - \xi_i, \xi_i \geq 0$$

$$\forall_{i=1}^{N_l} : y_i(\mathbf{u}^T \mathbf{r}_i + b_u) \geq 1 - \eta_i, \eta_i \geq 0$$

$$\forall_{j=1}^{N'} : y'_j(\mathbf{w}^T \mathbf{x}'_j + b_w) \geq 1 - \xi'_j, \xi'_j \geq 0$$

$$\forall_{j=1}^{N'} : y'_j(\mathbf{u}^T \mathbf{r}'_j + b_u) \geq 1 - \eta'_j, \eta'_j \geq 0$$

where N_l is the number of labeled images given by the user, N' is the number of unlabeled images involved, and

$\mathbf{Y}' \in [-1, 1]^{|S'|}$ is the label vector of unlabeled training data. Here, unlabeled data are engaged in the learning task based on two reasons. First, the unlabeled data are employed to enable the consistent predictions on the two types of information. Meanwhile, the unlabeled data can boost the performance of SVMs according to the transductive learning theory of SVMs [18, 34]. Moreover, to avoid the dominance of unlabeled data in the learning task, a parameter ρ is introduced for a regularization purpose. Finally, the above framework can be extended to nonlinear kernels straightforwardly [32].

4.2. Alternating Optimization

The optimization of the coupled SVM is not an easy task. As solving the optimization of Eq. 1 directly is very difficult, we propose to employ the Alternating Optimization (AO) technique [2] to tackle the problem.

First, we fix the parameters \mathbf{Y}' and try to find the (\mathbf{u}, b_u) and (\mathbf{w}, b_w) that optimize the objective. When fixing \mathbf{Y}' , in accordance with the Eq. 1, we try to solve the following two optimization problems:

$$\min_{\mathbf{w}, b_w, \xi, \xi', \mathbf{Y}'} \frac{1}{2} \|\mathbf{w}\|^2 + C_w \sum_{i=1}^{N_l} \xi_i + \rho C_w \sum_{j=1}^{N'} \xi'_j \quad (2)$$

$$\text{subject to} \quad \forall_{i=1}^{N_l} : y_i(\mathbf{w}^T \mathbf{x}_i + b_w) \geq 1 - \xi_i, \xi_i \geq 0 \\ \forall_{j=1}^{N'} : y'_j(\mathbf{w}^T \mathbf{x}'_j + b_w) \geq 1 - \xi'_j, \xi'_j \geq 0,$$

and

$$\min_{\mathbf{u}, b_u, \eta, \eta', \mathbf{Y}'} \frac{1}{2} \|\mathbf{u}\|^2 + C_u \sum_{i=1}^{N_l} \eta_i + \rho C_u \sum_{j=1}^{N'} \eta'_j \quad (3)$$

$$\text{subject to} \quad \forall_{i=1}^{N_l} : y_i(\mathbf{u}^T \mathbf{r}_i + b_u) \geq 1 - \eta_i, \eta_i \geq 0 \\ \forall_{j=1}^{N'} : y'_j(\mathbf{u}^T \mathbf{r}'_j + b_u) \geq 1 - \eta'_j, \eta'_j \geq 0.$$

To solve these two problems, we can simply apply the technique used in solving regular SVMs. For example, to solve the optimization in Eq. 2, we introduce non-negative Lagrange multipliers $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \dots, \alpha_{N_l+N'})$ to enforce the constraints. For the convenience of discussion, let us denote $\hat{\mathbf{Y}}^T = \{\hat{y}_1, \dots, \hat{y}_{N_l+N'}\}$, where $\hat{y}_i = y_i$ for $i = 1, \dots, N_l$, and $\hat{y}_{N_l+j} = y'_j$ for $j = 1, \dots, N'$. After introducing a Lagrange function in the optimization problem of Eq. 2, it is not difficult to derive the following dual:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \boldsymbol{\alpha}^T \mathbf{1} \\ \text{subject to} \quad \boldsymbol{\alpha}^T \hat{\mathbf{Y}} = 0 \\ \forall_{i=1}^{N_l} : 0 \leq \alpha_i \leq C_w \\ \forall_{i=N_l+1}^{N_l+N'} : 0 \leq \alpha_i \leq \rho C_w$$

where Q is an $(N_l + N')$ by $(N_l + N')$ positive semidefinite matrix. If a kernel is enabled, then $Q_{ij} \equiv \hat{y}_i \hat{y}_j K(\mathbf{x}_i, \mathbf{x}_j)$, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$.

Secondly, when (\mathbf{w}, b_w) and (\mathbf{u}, b_u) are solved from the first step, we can fix them and turn to finding the optimal \mathbf{Y}' that fits the data. After removing constant terms from the objective function in Eq. 1, the original optimization problem can be simplified into the following form:

$$\min_{\xi', \eta', \mathbf{Y}'} \quad C_w \sum_{j=1}^{N'} \xi'_j + C_u \sum_{j=1}^{N'} \eta'_j \\ \text{subject to} \quad \forall_{j=1}^{N'} : y'_j(\mathbf{w}^T \mathbf{x}'_j + b_w) \geq 1 - \xi'_j, \xi'_j \geq 0 \\ \forall_{j=1}^{N'} : y'_j(\mathbf{u}^T \mathbf{r}'_j + b_u) \geq 1 - \eta'_j, \eta'_j \geq 0.$$

If we substitute the slack variables into the objective function, the optimization can be represented into an implicit formulation as follows:

$$\min_{\mathbf{Y}'} \quad \sum_{j=1}^{N'} \left\{ \begin{array}{l} C_w \max(0, 1 - y'_j(\mathbf{w}^T \mathbf{x}'_j + b_w)) + \\ C_u \max(0, 1 - y'_j(\mathbf{u}^T \mathbf{r}'_j + b_u)) \end{array} \right\}$$

Since the label y'_j only takes the value of +1 or -1, it is not difficult to see that the above optimization is a simple integer programming problem that can be solved efficiently.

Based on the above two-step alternating optimization strategy, we can first randomly choose a set of labels for the unlabeled data, and then launch the alternating optimization procedure beginning with a small value of ρ in order to avoid a predominance of unlabeled data; this is similar to the approach in transductive SVM [18]. After each alternating optimization round, we can increase ρ until it achieves a setting threshold.

5. A Practical Algorithm by Coupled SVM

The above coupled SVM can be viewed as a general technique for learning on multiple-modality problems. For our application that learns on data with two types of information, we can apply the coupled SVM technique to develop an algorithm for the log-based relevance feedback task. Although the implementation of the algorithm by the coupled SVM seems straightforward, several practical tricks are useful to develop an effective algorithm.

The first important consideration for the practical algorithm is the strategy for choosing the unlabeled samples. Since a relevance feedback algorithm requires to respond fast, it is impossible to engage all of the unlabeled data in the learning task. One possible strategy is to choose the unlabeled samples closest the decision boundary of SVMs since they are most informative according to the active learning theory [30, 31]. Unfortunately, this kind of approach did not achieve promising improvements in our experiment. The exact answer to this observation in theory

Algorithm LRF-CSVM:

Input:

- \mathbf{q} : a query sample provided by a user
- \mathcal{S}_l : set of N_l labeled samples: $[(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{N_l}, y_{N_l})]$

Parameters:

- N : total number of images in the dataset
- N_l : number of labeled samples provided by users initially
- N' : number of unlabeled samples used in the learning task
- \mathcal{S}' : set of N' unlabeled samples $\{\mathbf{x}'_i, y'_i\}_{i=1}^{N'}$
- C_w, C_u, ρ : parameters in the optimization (1)
- \mathbf{w} : a weight vector assigned to the features of low-level visual content
- \mathbf{u} : a weight vector assigned to the log sessions of user feedback
- Δ : a threshold value to control the degree of error

Output:

- $\{z_1, z_2, \dots, z_{N_r}\}$: a set of N_r images most relevant the query \mathbf{q} .

BEGIN

// 1. Selecting N' unlabeled samples for the learning task $(\mathbf{w}, b_w, \xi) = \text{SOLVE_SVM_QP}([(x_1, y_1), \dots, (x_{N_l}, y_{N_l})], C_w);$ $(\mathbf{u}, b_u, \eta) = \text{SOLVE_SVM_QP}([(r_1, y_1), \dots, (r_{N_l}, y_{N_l})], C_u);$ FOR $i=1$ TO N DO $\text{dist}(z_i) = \text{SVM_Dist}(x_i, \mathbf{w}, b_w) + \text{SVM_Dist}(r_i, \mathbf{u}, b_u);$

END

 $\mathcal{S}' = [];$ $\mathcal{S}' = \text{Add_Unlabeled_Samples_with_Max_Dist}(N'/2, \text{dist}[]);$ $\mathcal{S}' = \text{Add_Unlabeled_Samples_with_Min_Dist}(N'/2, \text{dist}[]);$ *// 2. Training the coupled Support Vector Machine* $\rho^* = 10^{-4}$ WHILE $(\rho^* < \rho)$ DO

BEGIN

 $(\mathbf{w}, b_w, \xi, \xi') = \text{SOLVE_SVM_QP}([(x_1, y_1), \dots, (x_{N_l}, y_{N_l})], [(x'_1, y'_1), \dots, (x'_{N'}, y'_{N'})], C_w, \rho^* C_w);$ $(\mathbf{u}, b_u, \eta, \eta') = \text{SOLVE_SVM_QP}([(r_1, y_1), \dots, (r_{N_l}, y_{N_l})], [(r'_1, y'_1), \dots, (r'_{N'}, y'_{N'})], C_u, \rho^* C_u);$ WHILE $(\exists i: (\xi'_i > 0) \text{ AND } (\eta'_i > 0) \text{ AND } (\xi'_i + \eta'_i > \Delta))$ DO

BEGIN

 FOR $i=1$ TO N' DO IF $((\xi'_i > 0) \text{ AND } (\eta'_i > 0) \text{ AND } (\xi'_i + \eta'_i > \Delta))$ THEN $y'_i = -y'_i;$

ENDIF

 $(\mathbf{w}, b_w, \xi, \xi') = \text{SOLVE_SVM_QP}([(x_1, y_1), \dots, (x_{N_l}, y_{N_l})], [(x'_1, y'_1), \dots, (x'_{N'}, y'_{N'})], C_w, \rho^* C_w);$ $(\mathbf{u}, b_u, \eta, \eta') = \text{SOLVE_SVM_QP}([(r_1, y_1), \dots, (r_{N_l}, y_{N_l})], [(r'_1, y'_1), \dots, (r'_{N'}, y'_{N'})], C_u, \rho^* C_u);$

END

 $\rho^* = \min(2 * \rho^*, \rho);$

END

*// 3. Retrieving the results by the coupled SVM*FOR $i=1$ TO N DO $\text{dist}(z_i) = \text{CSVM_Dist}(x_i, r_i, \mathbf{w}, b_w, \mathbf{u}, b_u);$

END

 $\{z_1, z_2, \dots, z_{N_r}\} = \text{Select_Samples_with_Max_CSVM_Dist}(N_r, \text{dist}[]);$ RETURN $\{z_1, z_2, \dots, z_{N_r}\};$

END

Figure 1. Algorithm for Log-based Relevance Feedback by Coupled Support Vector Machine

is unclear, but a reasonable explanation is that the learning machine may take too much effort on learning the unlabeled data.

In order to avoid the problem of putting overlarge effort on learning the label information, we suggest choosing the unlabeled data most similar to the labeled data in order to provide more informative labels initially. The way for choosing the unlabeled data can be assisted by both the low-level visual information of image content and the log information of user feedback [12]. The idea is that we learn two SVM classifiers on the two types of information respectively. Then, we choose the unlabeled samples based on the sum of SVM distances on the two types of information. Fig. 1 shows more details of the algorithm for the log-based relevance feedback problem by the coupled SVM (LRF-CSVM) to integrate the log information of user feedback into the relevance feedback. In the algorithm, there are three main part, i.e. choosing the unlabeled data, training the coupled SVM, and retrieving the results by the coupled SVM. In the training procedure, a parameter Δ is introduced for controlling the error degree of label correction to avoid overlarge change in the label set.

6. Experimental Results

In our experiment, we perform detailed performance comparison to evaluate the effectiveness of our proposed technique. We want to answer the following questions through empirical studies in our experiment. The first question is whether the log-based relevance feedback techniques can achieve better retrieval performance than the regular relevance feedback technique. If they can, how much can they perform better? Furthermore, we want to know whether the log-based relevance feedback algorithm using the coupled SVM will perform better than the heuristic approach by combining two SVMs linearly. Meanwhile, we are interested to know the performance of log-based relevance feedback technique working on two datasets with different number of categories. We present the details of our experiment as follows.

6.1. Datasets

To perform empirical evaluation of our proposed algorithm, we choose real-world images from the COREL image CDs. There are two sets of data collected in our experiment: 20-Category and 50-Category. The 20-Category dataset contains 20 categories and the 50-Category one contains 50 categories. Each category in the datasets consists exactly 100 images selected from the COREL image CDs. The categories represent different semantic meanings, such as *antique*, *antelope*, *aviation*, *balloon*, *botany*, *butterfly*,

car, *cat*, *dog*, *firework*, *horse* and *lizard*, etc. Fig. 2 shows some images used in our experiment.



Figure 2. Some images selected from COREL image CDs in our experiment

The motivation for selecting the semantic categories are twofold. First, it enables us to evaluate whether the approach can retrieve the images that are not only visually relevant but also have similar semantic meaning. Second, the approach can help us evaluate the performance automatically, which can reduce the subjective errors arising from manual evaluations by different people.

6.2. Image Representation

Image representation is an important step in the implementation of relevance feedback algorithms in CBIR. Three different features are chosen in our experiment to represent the images: color, edge and texture.

The color feature is widely adopted in CBIR for its simplicity and effectiveness. The color feature engaged in our experiment is color moment since it is naturally closer to human perception, and many previous research studies have showed the effectiveness of color moment applied in CBIR. For the employed color moment, we extract 3 moments: color mean, color variance and color skewness in each color channel (H, S, and V), respectively. Thus, 9-dimensional color moment is adopted as the color feature in our experiment.

The edge feature can be very effective in CBIR when the contour lines of images are evident. The edge feature in our experiment is the edge direction histogram [16]. The images in the datasets are first translated to gray images. Then a Canny edge detector is applied to obtain the edge images. From the edge images, the edge direction histogram can then be computed. The edge direction histogram is quantized into 18 bins of 20 degrees each; hence an 18-dimensional edge direction histogram is employed to represent the edge feature.

The texture feature is known to be an important cue for image retrieval. A variety of texture analysis methods have been studied in past years. In our experiment, we employ the wavelet-based texture technique [21, 28]. The original color images are transformed to gray images. Then we perform the Discrete Wavelet Transformation (DWT) on the gray images employing a Daubechies-4 wavelet filter [28]. Each wavelet decomposition on a gray 2D-image results in four subimages with a $0.5 * 0.5$ scaled-down image of the input image and the wavelets in three orientations: horizontal, vertical and diagonal. The scaled-down image is fed into the DWT operation to produce the next four subimages. In total, we perform 3-level decompositions and obtain 10 subimages in different scales and orientations. One of the 10 subimages is a subsampled average image of the original image; this is discarded since it contains less useful texture information. For the other 9 subimages, we compute the entropy of each subimage respectively. Therefore, we obtain a 9-dimensional wavelet-based texture feature to describe the texture information for each image.

6.3. Log Data Collection of User Feedback

Log data collection of user feedback is an important step toward performance evaluation of a log-based relevance feedback algorithm for CBIR. Instead of producing simulated log data by computers, we collect the feedback logs from real-world users. The main reason is that the user feedback log data collected from real-world users typically contain more or less noise that is difficult to be simulated. In order to collect the log data, we have developed a CBIR system powered with a relevance feedback mechanism [10, 11]. In our CBIR system, users can judge the relevance of images simply by ticking out the relevant images.

In our experiment, the log data of relevance feedback are collected from users on both the 20-Category and 50-Category dataset. The reason to evaluate the performance on the two datasets is that we want to evaluate algorithms on the datasets with different diversity. The 50-Category dataset is more diverse in visual content than the 20-Category dataset since the number of categories is larger than the 20-Category one. Hence, the log information may be less helpful for the 50-Category dataset which is more diverse than the 20-Category one. It is interesting for us to observe it empirically.

More specifically, the way to organize and collect the log data from users is articulated as follows. For each participant user, he or she first specifies a query example and submits it to the CBIR system. The CBIR system returns 20 initial similar images to the user according the measurement of low-level visual features of image content. The user then employs the relevance feedback tool to improve the re-

trieval performance. For the given 20 images, he/she marks positive (relevant) or negative (irrelevant) labels on the images according to his/her query target. When a relevance feedback round is finished, the information of user feedback will be logged into a log database. Each relevance feedback round corresponds to a log session unit of user feedback in the log database. Since different people may have different subjectivity, a certain amount of noise is inevitable to appear in the collected log data. The noise problem is not further discussed in this paper, although it may also be a critical factor for the performance evaluation of the log-based relevance feedback algorithms.

In total, we respectively collect 150 log sessions for each of the two datasets from users in the experiment. Although the number of log sessions is not very large, they are enough to evaluate the effectiveness of our algorithm. In reality, many more log sessions can be collected in a real-world CBIR application from a long-term learning perspective; however, we hope to demonstrate that our proposed algorithm can work well even with limited log sessions.

6.4. Performance Evaluation

We have developed the log-based relevance feedback algorithm by coupled SVM (LRF-CSVM) in our experiment. We implement the coupled SVM algorithm by modifying the LIBSVM library [4]. In order to evaluate our method's effectiveness, we compare the performance with regular relevance feedback algorithm by SVM (RF-SVM) and the straightforward log-based relevance feedback approach by simply combining two SVMs for the two types of information (LRF-2SVMs). In the experiment, the same experimental settings are adopted in the schemes compared. The kernel function for all schemes is based on the Gaussian RBF kernel [4]. The performance metric used in the experiment is *Average Precision*, which is defined as the number of relevant samples in the returned images divided by the total number of returned images.

For an objective performance comparison, 200 queries are generated randomly. For each query, we first simply calculate the distances between the images in the database based on the low-level visual features and return top 20 similar results for evaluation. The procedure of relevance evaluation is automatic: we simulate the relevance judgements that would have been made by users. Based on a query q and 20 labeled images, we try the three different relevance feedback schemes and compare their improvement on the retrieval performance. The experimental results are obtained by taking an average over the 200 queries.

Fig. 3 and Fig. 4 illustrate the visual comparison of the experimental results on the two datasets. In the figures, the curve of Euclidean is given as a reference, which is obtained based on the Euclidean distance measure on the low-level

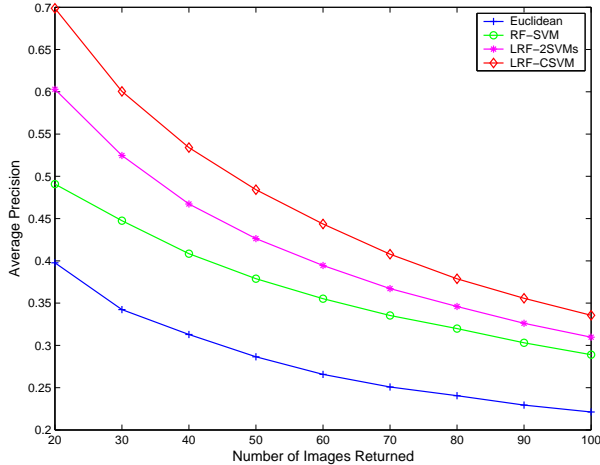


Figure 3. Performance comparison on the 20-Category dataset

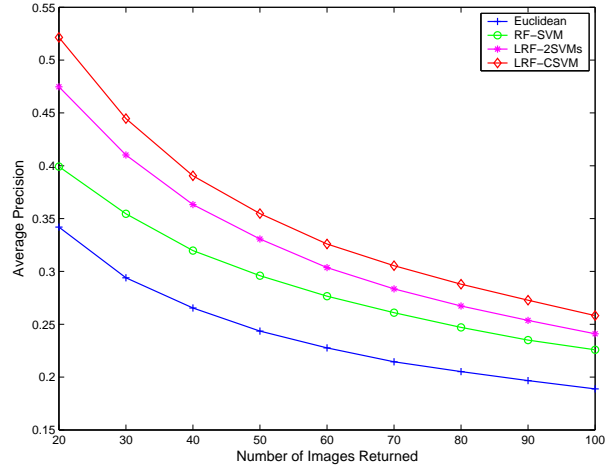


Figure 4. Performance comparison on the 50-Category dataset

image features. The curve of RF-SVM is the baseline for performance comparison. The both figures evidently show the answer of our question, i.e. the log-based relevance feedback techniques can improve the retrieval performance substantially compared with the regular relevance feedback scheme. Moreover, from both figures, we can observe the log-based relevance feedback algorithm using the coupled SVM shows promising improvement on the retrieval performance compared with the log-based relevance feedback by a simple combination of two SVMs.

To examine the quantitative amount of improvement, let us look into more detailed experimental results from Table 1 and Table 2. The two tables show the results of average precision on the top returned images and the mean average precision (MAP) of different compared schemes. For example, on the 20-Category dataset, by evaluating the relevance on the top 20 returned images, the log-based relevance feedback algorithm by the coupled SVM achieves 42.4% improvement compared with the regular relevance feedback approach by SVM, which greatly outperforms 22.9% improvement of the log-based relevance feedback approach by the combination of two SVMs (LRF-2SVMs). On average, the coupled SVM based approach achieves 25.9% improvement on MAP compared with the regular relevance feedback algorithm by SVM on the 20-Category dataset, while the LRF-2SVMs approach only obtains 12.3% improvement on MAP compared with the regular approach. Similarly, on the 50-Category dataset, the coupled SVM approach achieves 20.0% improvement on MAP compared with the regular relevance feedback technique by SVM on average, while the approach of combination of two SVMs only obtains 11.2% improvement. We also find that the amount of improvement on the 50-Category dataset is less

Table 1. Quantitative evaluation for different approaches on the 20-Category dataset

#TOP	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
20	0.398	0.491	0.603 (+22.9%)	0.699 (+42.4%)
30	0.342	0.448	0.525 (+17.2%)	0.600 (+34.2%)
40	0.313	0.409	0.467 (+14.4%)	0.534 (+30.7%)
50	0.287	0.379	0.426 (+12.5%)	0.484 (+27.8%)
60	0.266	0.355	0.394 (+11.0%)	0.444 (+24.9%)
70	0.251	0.335	0.367 (+9.5%)	0.408 (+21.6%)
80	0.241	0.320	0.346 (+8.2%)	0.379 (+18.4%)
90	0.229	0.303	0.326 (+7.6%)	0.356 (+17.4%)
100	0.221	0.289	0.310 (+7.2%)	0.336 (+16.1%)
MAP	0.283	0.370	0.418 (+12.3%)	0.471 (+25.9%)

than that on the 20-Category dataset since it is more diverse for more categories. However, the improvements by the log-based relevance feedback algorithms are still very promising on average.

6.5. Discussions

Although we have demonstrated the effectiveness of our proposed algorithm from the above experimental results, several empirical findings are worth discussion. First, the selection strategy of unlabeled data for the coupled SVM is important in an image retrieval environment. A good strategy is to choose unlabeled images closest to the positive labeled images for half the samples, and those closest to the negative labeled images for the other half. The reason for the success of this strategy is not yet clear in theory exactly, but it can be explained by noting that the samples closest to the positive samples can provide more precise label information, reducing the effort in learning the labels by the

transductive inference approach. Further, the choice of parameter ρ is also important for the scheme. Whether existing an optimal parameter for the scheme is still an open question.

Moreover, it is worth making some comments on the coupled SVM for learning on multiple-modality problems. Instead of two types of information, our model can be easily generalized to learn the data with multiple types of information. However, there are several open problems to be solved. First, the current approach for the optimization of the coupled SVM is based on the Alternating Optimization technique, which may not be able to guarantee the optimal solution globally. It is interesting to seek other optimization techniques for tackling the problem. Moreover, whether existing a better formulation of the coupled SVM is worth discussing both theoretically and empirically.

7. Related Work

Relevance feedback originally comes from traditional text-based information retrieval community [22, 26]. It is a bit surprising that relevance feedback has been received much more research attention from the image retrieval community in the past decade [25, 15]. Most of the past research studies focused mainly on studying various algorithms and theories for traditional relevance feedback scheme. However, due to the difficulty of the learning task, it is almost impossible to bridge the semantic gap between low-level visual features and high-level semantic concepts by learning low-level information of image content only.

Hence, exploiting the log resource of user feedback has become a promising direction by which to attack the challenge [12]. Although there are some research work for studying user logs in traditional text information retrieval [1, 6], there is little research attention paid to image retrieval. Some related work in this area has been recently reported by Zhou and Zhang et al. [35], He and King et al. [8], Hoi and Lyu [12], He and Ma, et al. [9], amongst others. Different from previous work, our work in this paper is based on a novel coupled support vector machine which can integrate the log information of user feedback into traditional relevance feedback with learning on the low-level visual features of image content.

8. Conclusion and Future Work

In this paper we study a log-based relevance feedback scheme by integrating the log information of user feedback with low-level image content for content-based image retrieval. In order to combine the user log information with low-level image content consistently, we propose a unified learning framework, i.e. a coupled support vector machine,

Table 2. Quantitative evaluation for different approaches on the 50-Category dataset

#TOP	Euclidean	RF-SVM	LRF-2SVMs	LRF-CSVM
20	0.342	0.399	0.475 (+18.9%)	0.522 (+30.6%)
30	0.294	0.355	0.410 (+15.7%)	0.445 (+25.4%)
40	0.265	0.320	0.363 (+13.6%)	0.391 (+22.1%)
50	0.244	0.296	0.331 (+11.7%)	0.355 (+19.8%)
60	0.228	0.277	0.304 (+9.8%)	0.326 (+17.9%)
70	0.215	0.261	0.283 (+8.6%)	0.305 (+17.1%)
80	0.205	0.247	0.267 (+8.2%)	0.288 (+16.5%)
90	0.197	0.235	0.254 (+7.9%)	0.273 (+16.1%)
100	0.189	0.226	0.241 (+6.7%)	0.258 (+14.4%)
MAP	0.242	0.291	0.325 (+11.2%)	0.351 (+20.0%)

for learning the two types of information. The suggested coupled SVM technique is generic and may also be applicable for other multi-modality learning tasks. To apply the coupled SVM for the log-based relevance feedback problem effectively, we develop a practical algorithm that can tackle the problem well. Experimental results show that our proposed algorithm is effective and promising.

Although our experimental results show that our algorithm is effective in practice, we may need to evaluate our algorithm further on larger databases and other different environments. Moreover, although the formulation of the proposed coupled SVM is sound, some theoretical problems, such as the convergence issue, the selection strategy for unlabeled samples, and the noise problem need to be further considered in our future work. Finally, we need to study the computation cost problem when applying the algorithm to large scale applications.

Acknowledgment

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4182/03E), and partially supported by Shun Hing Institute of Advanced Engineering (SHIAE).

References

- [1] P. Anick. Using terminological feedback for web search refinement: a log-based study. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 88–95, 2003.
- [2] J. C. Bezdek and R. J. Hathaway. Convergence of alternating optimization. *Neural, Parallel Sci. Comput.*, 11(4):351–368, 2003.
- [3] D. Blei and M. I. Jordan. Modeling annotated data. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 127–134, 2003.

- [4] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [5] I. J. Cox, M. Miller, T. Minka, and P. Yianilos. An optimized interaction strategy for bayesian relevance feedback. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 553–558, Santa Barbara, CA, USA, 1998.
- [6] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma. Probabilistic query expansion using query logs. In *Proc. of the eleventh international conference on World Wide Web (WWW'02)*, 2002.
- [7] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proc. the 7th European Conf. on Computer Vision*, pages 97–112, 2002.
- [8] X. He, O. King, W.-Y. Ma, M. Li, and H. J. Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):39–48, Jan. 2003.
- [9] X. He, W.-Y. Ma, and H.-J. Zhang. Learning an image manifold for retrieval. In *Proc. ACM Multimedia*, pages 17–23, New York, US, 2004.
- [10] C. H. Hoi and M. R. Lyu. Biased support vector machine for relevance feedback in image retrieval. In *Proceedings International Joint Conference on Neural Networks (IJCNN'04)*, pages 3189–3194, Budapest, Hungary, 2004.
- [11] C. H. Hoi and M. R. Lyu. Group-based relevance feedback with support vector machine ensembles. In *Proceedings 17th International Conference on Pattern Recognition (ICPR'04)*, pages 874–877, Cambridge, UK, 2004.
- [12] C. H. Hoi and M. R. Lyu. A novel log-based relevance feedback technique in content-based image retrieval. In *Proc. ACM Multimedia (ACM-MM'04)*, pages 24–31, New York, USA, 2004.
- [13] C. H. Hoi and M. R. Lyu. Web image learning for searching semantic concepts in image databases. In *Poster Proc. 13th Int. World Wide Web Conference (WWW 2004)*, New York, US, 2004.
- [14] P. Hong, Q. Tian, and T. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In *Proc. IEEE International Conference on Image Processing (ICIP'00)*, Vancouver, BC, Canada, 2000.
- [15] T. S. Huang and X. S. Zhou. Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods. In *Proceedings of IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, Oct. 2001.
- [16] A. K. Jain and A. Vailaya. Shape-based retrieval: a case study with trademark image database. *Pattern Recognition*, (9):1369–1390, 1998.
- [17] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03)*, pages 119–126, 2003.
- [18] T. Joachims. Transductive inference for text classification using support vector machines. In *Proc. 16th Int. Conf. on Machine Learning (ICML'99)*.
- [19] J. Laaksonen, M. Koskela, and E. Oja. Picsom: Self-organizing maps for content-based image retrieval. In *Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington, DC, USA, 1999.
- [20] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In *Advances in Neural Information Processing Systems (NIPS'03)*, 2003.
- [21] B. Manjunath, P. Wu, S. Newsam, and H. Shin. A texture descriptor for browsing and similarity retrieval. *Signal Processing Image Communication*, 2001.
- [22] J. Rocchio. Relevance feedback information retrieval. In G. Salton, editor, *The Smart retrieval system: experiments in automatic document processing*, 1971.
- [23] Y. Rui and T. Huang. Optimizing learning in image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000.
- [24] Y. Rui and T. S. Huang. A novel relevance feedback technique in image retrieval. In *Proc. ACM Multimedia (MM'99)*, pages 67–70, Orlando, Florida, USA, 1999.
- [25] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, Sept. 1998.
- [26] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 44(4):288–287, 1990.
- [27] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
- [28] J. Smith and S.-F. Chang. Automated image retrieval using color and texture. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Nov. 1996.
- [29] K. Tieu and P. Viola. Boosting image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, South Carolina, USA, 2000.
- [30] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia (MM'01)*, pages 107–118, 2001.
- [31] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, 2001.
- [32] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [33] N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries (CVPR'00)*, South Carolina, USA, 2000.
- [34] L. Wang, K. L. Chan, and Z. Zhang. Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*, volume 1, pages 629–634, 2003.
- [35] X.-D. Zhou, L. Zhang, L. Liu, Q. Zhang, and B.-L. Shi. A relevance feedback method in image retrieval by analyzing feedback log file. In *Proc. International Conference on Machine Learning and Cybernetics*, volume 3, pages 1641–1646, Beijing, 2002.