Which Components are Important for Interactive Image Searching?

Dacheng Tao, *Member, IEEE*, Xiaoou Tang, *Senior Member, IEEE*, and Xuelong Li, *Senior Member, IEEE*

***Abstract*—With many potential industrial applications, content-based image retrieval (CBIR) has recently gained more attention for image management and web searching. As an important tool to capture users' preferences and thus to improve the performance of CBIR systems, a variety of relevance feedback (RF) schemes have been developed in recent years. One key issue in RF is: which features (or feature dimensions) can benefit this human-computer iteration procedure? In this paper, we make theoretical and practical comparisons between principal and complement components of image features in CBIR RF. Most of the previous RF approaches treat the positive and negative feedbacks equivalently although this assumption is not appropriate since the two groups of training feedbacks have very different properties. That is, all positive feedbacks share a homogeneous concept while negative feedbacks do not. We explore solutions to this important problem by proposing an orthogonal complement component analysis. Experimental results are reported on a real-world image collection to demonstrate that the proposed complement components method consistently outperforms the conventional principal components method in both linear and kernel spaces when users want to retrieve images with a homogeneous concept.**

***Index Terms*—Content-based image retrieval (CBIR), kernel machine, orthogonal complement component analysis (OCCA), relevance feedback (RF), support vector machine (SVM).**

I. Introduction

With the explosive growth in image records and the rapid increase of computer power, retrieving images from a large-scale image database has become one of the most active research fields [17]. To give all images text annotations manually is tedious and impractical and to automatically annotate an image is generally beyond current techniques. Moreover, a picture says more than a thousand words.

Content-based image retrieval (CBIR) [24] is a technique to retrieve images, which are semantically relevant to a query image provided by a user, from an image database. It is based on representing images with visual features, which can be automatically extracted from images, such as color, texture, and shape. However, the gap between the low-level visual features and the high-level semantic meanings usually leads to poor performance.

Relevance feedback (RF) [16] is an effective method to bridge this gap and to scale up the performance in CBIR systems. RF focuses on the interactions between the user and the search engine by requiring the user to label semantically positive or negative feedbacks. MARS [15] introduced both the query movement and the re-weighting techniques. MindReader [8] formulated a minimization problem on the parameter estimation process. PicHunter [2] proposed a stochastic comparison search. With the observation that all positive examples are alike and each negative example is negative in its own way, biased discriminant analysis (BDA) [26] and its enhanced version [21] were developed.

Given the user feedback information, the key for a RF scheme is how to construct a suitable classifier. However, RF is much different from the traditional classification problem because users would not like to provide a large number of feedbacks. Among various RF schemes, small sample learning methods, where the number of the training samples is much smaller than the dimension of the descriptive features, are of the most promising.

Support vector machine (SVM) [23] is a popular small sample learning method used in recent years. It obtains top-level performance in different applications [1], [5], [7], [20], [25] because of its good generalization ability. SVM has a very good performance for pattern classification problems by minimizing the Vapnik-Chervonenkis (VC) dimension and achieving a minimal structural risk. Within different RF schemes, SVM-based RF is popular because it outperforms many other classifiers when the size of the training set is small. SVM active learning ($\mathrm{SVM}_{\mathrm{Active}}$) [22] halves the image space each time: 1) retrieved images are selected from the samples, which are farthest from the classifier boundary on the positive side and 2) samples close to the boundary are deemed as the most informative ones for the user to label. Recently, $\mathrm{SVM}_{\mathrm{Active}}$ has been combined with the multimodal concept-dependent process for CBIR [4]. Although $\mathrm{SVM}_{\mathrm{Active}}$-based RF can work better than the conventional SVM-based RF, it requires users to label a lot of training images (about twenty images) in the first round feedback procedure. Guo *et al.* [5] developed a constrained similarity measure (CSM) for image retrieval, in which SVM and AdaBoost are employed as classifier. The CSM also learns a boundary that halves the images in the

Manuscript received October 17, 2006; revised February 15, 2007. The work was supported in part by the Research Grants Council of the Hong Kong SAR under Project AoE/E-01/99, in part by the Internal Competitive Research Grants of the Department of Computing with the Hong Kong Polytechnic University under Project A-PH42, and the in part by the National Natural Science Foundation of China under 60703037.This paper was recommended by Associate Editor L. Guan.

D. Tao is with the Biometrics Research Centre, Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: csdct@comp.polyu.edu.hk; dacheng.tao@gmail.com).

X. Tang was with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. He is now with Microsoft Research Asia, Beijing 100080, China (e-mail: xtang@ie.cuhk.edu.hk).

X. Li is with the School of Computer Science and Information Systems, Birkbeck College, University of London, London WC1E 7HX, U.K. (e-mail: xuelong@dcs.bbk.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.906936

database into two groups, and images inside the boundary are ranked by their Euclidean distance to the query. There are also some other kinds of SVM-based RFs [27]. However, most of SVM-based RFs treat positive and negative feedbacks equivalently. This assumption is not appropriate, since the two groups of training feedbacks have very different properties, that is, all positive feedbacks share a homogeneous concept while negative feedbacks do not.

To explore solutions to the above issue, we propose an orthogonal complement component analysis (OCCA), which captures the invariant subspace of all positive feedbacks or the homogeneous concept shared by all positive feedbacks. The labeled positive feedbacks are mapped to their center. This mapping is realized in a feature subspace, into which other samples are also mapped. Experiments show that OCCA performs better than conventional SVM-based RFs on a real-world image collection. Motivated by the kernel approach successfully used in pattern recognition, OCCA is then generalized as the kernel empirical OCCA (KEOCCA). KEOCCA not only can improve the performance of the conventional SVM-based RF but can also outperform OCCA.

The layout of this paper is as follows. In Section II, the conventional SVM- and the $\mathrm{SVM_{Active}}$-based RF are briefly introduced. Section III proposes the orthogonal complement component analysis (OCCA). Section IV generalizes the OCCA to kernel space as the kernel empirical OCCA (KEOCCA). a large number of experiments are reported in Section V. Conclusions are drawn in Section VI.

II. Support Vector Machine-Based Relevance Feedback

In this section, the conventional SVM- [25] and $\mathrm{SVM_{Active}}$-based RF [22] are briefly introduced.

SVM [23] is a very effective binary classification algorithm. Consider a linearly separable binary classification problem

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \quad \text{and} \quad y_i = \{+1, -1\} \tag{1}$$

where $\mathbf{x}_i \in R^M$ and y_i is the label of the class that the vector belongs to. SVM separates the two classes of samples by a hyperplane

$$\mathbf{w}^T\mathbf{x} + b = 0 \tag{2}$$

where $\mathbf{x}$ is an input vector, $\mathbf{w}$ is an adaptive weight vector, and b is a bias. SVM finds the parameters $\mathbf{w}$ and b for the optimal hyperplane to maximize the geometric margin $2/\|\mathbf{w}\|$, subject to

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq +1. \tag{3}$$

The solution can be found through a Wolfe dual problem with the Lagrangian multipliers α_i

$$Q(\alpha) = \sum_{i=1}^{N}\alpha_i - \sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(\mathbf{x}_i \cdot \mathbf{x}_j)\Big/2 \tag{4}$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{N}\alpha_i y_i = 0$.

In the dual format, samples only appear in the inner product. To get a potentially better representation of samples, they are mapped to kernel space and implemented by kernel trick

$$\mathbf{x}_i \cdot \mathbf{x}_j \to \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \tag{5}$$

where $K(.)$ is a kernel function. We then get the kernel version of the Wolfe dual problem

$$Q(\alpha) = \sum_{i=1}^{N}\alpha_i - \sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)\Big/2. \tag{6}$$

Thus for a given kernel function, the SVM classifier is given by

$$F(\mathbf{x}) = \mathrm{sgn}\left(f(\mathbf{x})\right) \tag{7}$$

where $f(\mathbf{x}) = \sum_{i=1}^{l}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$ is the decision function of SVM and l is the number of support vectors.

In general [22], [25], the lower the $|\sum_{i=1}^{l}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b|$ is for a given sample, the closer this sample is to the decision boundary, and the lower the corresponding prediction confidence, and vice versa. RF is used to find an adaptive dissimilarity measure which approaches to the sentiments of the user. Consequently, for the conventional SVM-based RF [25] and the $\mathrm{SVM_{Active}}$-based RF [22], the dissimilarity measure is always given by the decision function of SVM, i.e., $\sum_{i=1}^{l}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$.

$\mathrm{SVM_{Active}}$-based RF [22] has a different feedback procedure from the conventional SVM-based RF [25]. $\mathrm{SVM_{Active}}$-based RF asks the user to label the marginal retrieved images as feedbacks and conventional SVM-based RF asks the user to label images that are farthest from the SVM boundary.

From the statistical learning theory [23], we know that the following inequality holds with probability of at least $1-\delta$ for any $n > h$

$$\begin{cases} R[f] \leq R_{\mathrm{emp}}[f] + G(h, n, \delta) \\ G(h, n, \delta) = \sqrt{\dfrac{h\left(\ln\left(\frac{2n}{h}\right)+1\right) - \ln\left(\frac{\delta}{4}\right)}{n}} \end{cases} \tag{8}$$

where h denotes the VC dimension [23] of the classifier function set, n is the size of the training set, and R_{emp} describes the empirical risk. For all $\delta > 0$ and $f \in F$ the inequality in (8) bounds the risk. The inequality gives us a way to estimate the error on future samples based only on the training error and the VC dimension of the classifier function set. Although the bound is loose, it is a suitable way for us to construct an effective classifier for CBIR RF.

The smaller the risk value $R[f]$ is, the better the performance of the classifier will be. We can see that the risk depends on the empirical risk R_{emp} and $G(h, n, \delta)$. Based on the representation of $G(h, n, \delta)$, we know that $G(h, n, \delta)$ is a strictly monotonically increasing function of h for given n and δ. The VC-dimension h is determined by the support vectors when the number of training samples is smaller than the feature dimension. In addition, the VC dimension h is almost an increasing function of the number of support vectors. Consequently, the performance of an SVM classifier depends mainly on the empirical risk, the number of support vectors, and δ. Since δ cannot be controlled manually, we can restrict R_{emp} and the number of support vectors to achieve a good performance.

III. Orthogonal Complement Components Analysis

With the conventional SVM-based RF [25], the CBIR performance can be improved. However, the conventional SVM-based RF treats the positive and negative feedbacks equivalently although this assumption is not always appropriate, since the two groups of training feedbacks have very different properties. That is, all positive feedbacks share a homogeneous concept while negative feedbacks do not. To explore solutions to this severe limitation of the conventional SVM-based RF, we propose in this Section an OCCA to mainly analyze positive feedbacks. Comparison experimental results are given in Section V.

In CBIR RF, it is not difficult to achieve zero empirical risk R_{emp} by having enough support vectors. However, a large number of support vectors will enlarge the VC dimension of an SVM classifier h. Therefore, we need to restrict both h and R_{emp}. In order to solve the problem, we project all positive feedbacks onto a subspace, in which all positive feedbacks have the same coordinates, and then project all negative feedbacks onto the same subspace. In this subspace, we can decrease the number of support vectors without increasing R_{emp}. Finally, the remaining images in the database are projected onto the same subspace and some similarity or dissimilarity measure is applied to sort images in the database.

For CBIR, the method is reasonable because all positive feedbacks share a homogeneous concept with the query image while negative feedbacks do not. Meanwhile, in the projection step, the optimal hyperplane of an SVM classifier can be transformed by any increasing positive feedbacks and it will not be sensitive to any negative feedbacks. Therefore, more emphasis is put on positive feedbacks and the search engine can find the homogeneous concept shared by all positive feedbacks as the number of positive feedbacks increases. In addition, the resulting SVM classification hyperplane will be simpler around the projection center. Based on this observation, we propose the OCCA to improve SVM.

OCCA can be implemented mainly in three steps: 1) to project all positive feedbacks onto their center and generate a subspace; 2) to project the remaining images including the negative feedbacks onto this subspace; and 3) to construct an SVM classifier in this subspace and resort all images based on new similarities.

For a set of positive feedbacks $\{\mathbf{x}_i^+, 1 \le i \le P\}$, where $\mathbf{x}_i^+ \in R^M$ and P is the number of the positive feedbacks. The Karhunen-Leove transformation (KLT) [3] can be used to extract the principal subspace and its orthogonal complement. The principal components describe the variance of the distribution of positive feedbacks while the orthogonal complement components describe the invariance. That is, the orthogonal components correspond to the directions with minimal variances. The basis functions for the KLT are obtained by solving the eigenvalue problem

$$\begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = [\mathbf{\Phi} \quad \mathbf{\Phi}^{\perp}]^T \mathbf{\Xi} [\mathbf{\Phi} \quad \mathbf{\Phi}^{\perp}] \tag{9}$$

where $\mathbf{\Xi}$ is the covariance matrix of positive feedbacks, $\mathbf{\Phi}$ is the principal subspace of $\mathbf{\Xi}$, $\mathbf{\Phi}^{\perp}$ is the orthogonal complement subspace of $\mathbf{\Phi}$ in $\mathbf{\Xi}$, $\mathbf{\Lambda}$ is the corresponding diagonal matrix of eigenvalues of $\mathbf{\Phi}$, and the eigenvalues of $\mathbf{\Phi}^{\perp}$ are 0. The unitary matrix $\mathbf{\Phi}^{\perp}$ defines a coordinate transform, which de-correlates the data, makes explicit the invariant subspace of the matrix operator $\mathbf{\Xi}$, and ensures that all positive feedbacks are mapped to their center. By KLT, we can obtain the orthogonal complement feature vector $\mathbf{y}_i^+ = (\mathbf{\Phi}^{\perp})^T(\mathbf{x}_i^+ - \overline{\mathbf{x}}^+)$, where $\overline{\mathbf{x}}^+ = (1/P)\sum_{i=1}^{P} \mathbf{x}_i^+$ is the center of positive feedbacks, $\mathbf{x}_i^+$ is the ith positive feedback, and $\mathbf{y}_i^+$ is the ith projected positive feedback. We call the transformation as OCCA, which preserves the invariant direction of the data distribution.

After projecting all positive feedbacks onto their center, we can project all negative feedbacks onto the subspace according to $\mathbf{y}_i^- = (\mathbf{\Phi}^{\perp})^T(\mathbf{x}_i^- - \overline{\mathbf{x}}^+)$, where $\mathbf{x}_i^-$ is the ith negative feedback and $\mathbf{y}_i^-$ is the ith projected negative feedback.

Then each image $\mathbf{x}$ in the database is projected onto the subspace through $\mathbf{y} = (\mathbf{\Phi}^{\perp})^T(\mathbf{x} - \overline{\mathbf{x}}^+)$, where $\mathbf{y}$ is the projected datum vector of the original datum vector $\mathbf{x}$.

The standard SVM classification algorithm is executed on $\mathbf{z} = [\mathbf{y}_i^+|_{i=1}^P, \mathbf{y}_j^-|_{j=1}^N]$, where $|\mathbf{z}| = 1 + N$ and N is the number of negative feedbacks. This is because in the projected subspace, all positive samples are merged together. Finally, we can measure the dissimilarity through the output of SVM $\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{z}_i, \mathbf{y}) + b$, where N_S is the number of support vectors. The outline of the proposed algorithm is shown in Table I.

TABLE I
ALGORITHM OF OCCA SVM

1	Calculate the covariance matrix $\mathbf{\Xi}$ of the positive feedbacks.
2	Calculate the orthogonal complement components $\mathbf{\Phi}^{\perp}$ of $\mathbf{\Xi}$ according to $\left[\mathbf{\Phi}^{\perp}\right]^T \mathbf{\Xi}\mathbf{\Phi}^{\perp} = 0$.
3	Project each positive feedback $\mathbf{x}_i^+$ onto their center $\mathbf{y}_i^+$.
4	Project each negative feedback $\mathbf{x}_j^-$ onto the orthogonal complement subspace, $\mathbf{y}_j^- = \left(\mathbf{\Phi}^{\perp}\right)^T \left(\mathbf{x}_j^- - \overline{\mathbf{x}}^+\right)$.
5	Project each remaining image $\mathbf{x}$ in the database onto the orthogonal complement subspace, $\mathbf{y} = \left(\mathbf{\Phi}^{\perp}\right)^T \left(\mathbf{x} - \overline{\mathbf{x}}^+\right)$.
6	Train a standard SVM classifier on $\mathbf{z} = \left[\mathbf{y}_i^+ \mid_{i=1}^{P}, \mathbf{y}_j^- \mid_{j=1}^{N}\right]$.
7	Resort the remaining projected images $\mathbf{y}$ using the output of SVM $\sum_{i=1}^{N_s} \alpha_i y_i K\left(\mathbf{z}_i, \mathbf{y}\right) + b$.

Recently, the locality preserving projections (LPP) [6] has been proposed to discover the nonlinear structure of data, which lie on a low dimensional manifold embedded in a high dimensional space. The difference between LPP and OCCA is the weight matrix. In LPP, the weight matrix is a normalized locality preservation matrix. In OCCA, the weight matrix is the same as the weight matrix in PCA. Both in LPP and OCCA, the eigenvectors associated with smallest eigenvalues are selected for feature representation. However, PCA applies the eigenvectors corresponding to the largest eigenvalues for feature representation.

IV. Kernel Empirical Orthogonal Complement Component Analysis

Herein, we aim to improve the performance of CBIR RF and generalize OCCA in the kernel space [13] and we name the approach as KEOCCA, which can be regarded as an enhanced OCCA. According to the kernel approach, the original input space is first nonlinearly mapped to an arbitrarily high dimensional feature space, in which the distribution of samples is linearized. Then, OCCA is used to obtain a classifier in the kernel feature space.

The direct kernelization of OCCA is to use only positive feedbacks to construct the bases, that is $\tilde{\mathbf{\Phi}}^{\perp} \in \text{span}\{\psi(\mathbf{x}_i^+)|_{i=1}^P\}$. In this paper, we use not only positive feedbacks but also negative feedbacks to construct the bases, that is $\tilde{\mathbf{\Phi}}^{\perp} \in \text{span}\{\psi(\mathbf{x}_i^+)|_{i=1}^P; \psi(\mathbf{x}_j^-)|_{j=1}^N\}$. In fact, the most suitable way to construct the bases is to incorporate all images in the database, because many more kernel features can be generated by this approach. However, the method is practically intractable for CBIR RF. Therefore, we only use all feedbacks to construct the bases and we call the new kernelization of OCCA as KEOCCA, but not the kernel OCCA.

Similar to SVM and other kernel machines, we first map a sample $\mathbf{x}$ to $\psi(\mathbf{x})$ in a higher dimensional space, and then the kernel trick $K(\mathbf{x}_i, \mathbf{x}_j) = \psi^T(\mathbf{x}_i)\psi(\mathbf{x}_j)$ is utilized to obtain the solution. We first calculate the covariance matrix of the positive feedbacks in the Hilbert space according to

$$\mathbf{\Xi} = \sum_{i=1}^{P} \left(\psi\left(\mathbf{x}_i^+\right) - \bar{\psi}(\mathbf{x}^+)\right)\left(\psi\left(\mathbf{x}_i^+\right) - \bar{\psi}(\mathbf{x}^+)\right)^T \tag{10}$$

where $\overline{\psi}(\mathbf{x}^+) = (1/P)\sum_{i=1}^{P}\psi(\mathbf{x}_i^+)$ is the center of positive feedbacks in the higher dimensional space. According to the previous analysis of the orthogonal complement components in the higher dimensional space, we know that $\tilde{\mathbf{\Phi}}^{\perp} \in \text{span}\{\psi(\mathbf{x}_i^+)|_{i=1}^P; \psi(\mathbf{x}_j^-)|_{j=1}^N\}$ (this is because we only use all feedbacks to construct the bases for forming the kernelization). Therefore, the basis function for KEOCCA can be solved by an eigenvalue problem

$$\mathbf{0} = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \mathbf{\Xi} \tilde{\mathbf{\Phi}}^{\perp} \tag{11}$$

where $\tilde{\mathbf{\Phi}}^{\perp} = \sum_{i=1}^{P}\xi_i\psi(\mathbf{x}_i^+) + \sum_{i=1+P}^{N+P}\xi_i\psi(\mathbf{x}_{i-P}^-)$.

Through the kernel trick, the eigenvalue problem can be solved by using the kernel Gram matrix $\mathbf{K}$, according to

$$\begin{aligned}
&\left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \mathbf{\Xi} \tilde{\mathbf{\Phi}}^{\perp} \\
&= \sum_{i=1}^{P}\left\{\left(\left\{\sum_{j=1}^{P}\xi_j\psi^T\left(\mathbf{x}_j^+\right) + \sum_{j=P+1}^{N+P}\xi_j\psi\left(\mathbf{x}_{j-P}^-\right)\right\}^T \right.\right. \\
&\qquad \left.\left(\psi\left(\mathbf{x}_i^+\right) - \frac{1}{P}\sum_{k=1}^{P}\psi\left(\mathbf{x}_k^+\right)\right)\right) \\
&\qquad \cdot \left(\left\{\sum_{j=1}^{P}\xi_j\psi^T\left(\mathbf{x}_j^+\right) + \sum_{j=P+1}^{N+P}\xi_j\psi\left(\mathbf{x}_{j-P}^-\right)\right\}^T \right. \\
&\qquad \left.\left.\left(\psi\left(\mathbf{x}_i^+\right) - \frac{1}{P}\sum_{k=1}^{P}\psi\left(\mathbf{x}_k^+\right)\right)\right)^T\right\} \\
&= \boldsymbol{\xi}^T \sum_{i=1}^{P}\left(\mathbf{k}_{(\cdot,i)} - \frac{1}{P}\sum_{k=1}^{P}\mathbf{k}_{(\cdot,k)}\right) \\
&\quad \left(\mathbf{k}_{(\cdot,i)} - \frac{1}{P}\sum_{k=1}^{P}\mathbf{k}_{(\cdot,k)}\right)^T \boldsymbol{\xi}
\end{aligned} \tag{12}$$

where the kernel Gram matrix is defined by (13), shown at the bottom of the page.

Therefore, we can obtain the kernel empirical orthogonal complement component (KEOCC) according to $\boldsymbol{\xi}$, which makes $\mathbf{0} = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \mathbf{\Xi} \tilde{\mathbf{\Phi}}^{\perp}$.

Similar to OCCA combined with the conventional SVM-based RF, we project positive feedbacks, negative feedbacks, and images in the database onto the KEOCC spanned space by $\mathbf{y} = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T(\psi(\mathbf{x}) - \overline{\psi}(\mathbf{x}^+))$. In KEOCCA, positive feedbacks, negative feedbacks, and images in the database are represented by $\mathbf{y}_i^+$, $\mathbf{y}_j^-$, and $\mathbf{y}$, respectively.

Then, using $\mathbf{z} = [\mathbf{y}_i^+|_{i=1}^P, \mathbf{y}_j^-|_{j=1}^N]$, we train the standard SVM classifier. Finally, we can measure the dissimilarity through the output of SVM according to $\sum_{i=1}^{N_s}\alpha_i y_i K(\mathbf{z}_i, \mathbf{y}) + b$, where N_S is the number of the support vectors. The outline of the proposed algorithm is shown in Table II.

V. Experiments

With CBIR [27], the search engine is required to feedback the most semantically relevant images after each previous RF

$$\begin{aligned}
\mathbf{K} &= \begin{bmatrix}\mathbf{k}_{(\cdot,1)} & \cdots & \mathbf{k}_{(\cdot,i)} & \cdots & \mathbf{k}_{(\cdot,P+N)}\end{bmatrix} \\
&= \begin{bmatrix}
K\left(\mathbf{x}_1^+, \mathbf{x}_1^+\right) & \cdots & K\left(\mathbf{x}_1^+, \mathbf{x}_P^+\right) & K\left(\mathbf{x}_1^+, \mathbf{x}_1^-\right) & \cdots & K\left(\mathbf{x}_1^+, \mathbf{x}_N^-\right) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
K\left(\mathbf{x}_P^+, \mathbf{x}_1^+\right) & \cdots & K\left(\mathbf{x}_P^+, \mathbf{x}_P^+\right) & K\left(\mathbf{x}_P^+, \mathbf{x}_1^-\right) & \cdots & K\left(\mathbf{x}_P^+, \mathbf{x}_N^-\right) \\
K\left(\mathbf{x}_1^-, \mathbf{x}_1^+\right) & \cdots & K\left(\mathbf{x}_1^-, \mathbf{x}_P^+\right) & K\left(\mathbf{x}_1^-, \mathbf{x}_1^-\right) & \cdots & K\left(\mathbf{x}_1^-, \mathbf{x}_N^-\right) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
K\left(\mathbf{x}_N^-, \mathbf{x}_1^+\right) & \cdots & K\left(\mathbf{x}_N^-, \mathbf{x}_P^+\right) & K\left(\mathbf{x}_N^-, \mathbf{x}_1^-\right) & \cdots & K\left(\mathbf{x}_N^-, \mathbf{x}_N^-\right)
\end{bmatrix}
\end{aligned} \tag{13}$$

TABLE II
ALGORITHM OF KEOCCA SVM

1	Calculate the kernel matrix $\mathbf{K}$.
2	Calculate the kernel empirical orthogonal complement components $\tilde{\mathbf{\Phi}}^{\perp}$ of the kernel covariance matrix $\mathbf{\Xi}$ of the positive feedbacks by $\mathbf{0} = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \mathbf{\Xi} \tilde{\mathbf{\Phi}}^{\perp}$.
3	Project all positive feedbacks $\mathbf{x}_i^+$ onto their center $\mathbf{y}_i^+$ according to $\mathbf{y}_i^+ = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \left(\psi\left(\mathbf{x}_i^+\right) - \bar{\psi}\left(\mathbf{x}^+\right)\right)$.
4	Project all negative feedbacks $\mathbf{x}_i^-$ onto the empirical kernel orthogonal complement subspace according to $\mathbf{y}_i^- = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \left(\psi\left(\mathbf{x}_i^-\right) - \bar{\psi}\left(\mathbf{x}^+\right)\right)$.
5	Project each remaining image $\mathbf{x}$ in the database onto the subspace according to $\mathbf{y} = \left(\tilde{\mathbf{\Phi}}^{\perp}\right)^T \left(\psi\left(\mathbf{x}\right) - \bar{\psi}\left(\mathbf{x}^+\right)\right)$.
6	Train a standard SVM classifier on $\mathbf{z} = \left[\mathbf{y}_i^+ \,|_{i=1}^{P}, \mathbf{y}_j^- \,|_{j=1}^{N}\right]$.
7	Resort the projected remaining images $\mathbf{y}$ using the output of SVM $\sum_{i=1}^{N_s} \alpha_i y_i K\left(\mathbf{z}_i, \mathbf{y}\right) + b$.

iteration. The user will not label many images for each iteration and will usually only do a few iterations. Thus, the following CBIR framework is used, into which any RF algorithm can be embedded.

When a query image is inputted, its low-level visual features are extracted. Then, all images in the database are sorted based on a similarity metric, e.g., Euclidean distance. If the user is satisfied with the result, the retrieval process is ended. If, however, the user is not satisfied, s/he can label some top query relevant images as positive feedbacks and/or some query irrelevant images as negative feedbacks. Using these feedbacks, the system is trained based on a learning machine (an embedded RF algorithm). Then, all the images are re-sorted based on the new similarity metric. If the user is still not content with the result, s/he repeats the process.

In this Section, we report the results of a large number of experiments. The experiments have two parts: statistical experiments (part B) and real-world experiments (part C). Prior to describing experiments, we introduce the database and visual features in part A.

A. Groundtruth and Feature

Groundtruth: For the experiments we used part of the Corel Photo Gallery [24], comprising 10 800 images. In the original Corel Photo Gallery, each folder has a name and includes 100 images. However, in the original database, names of most folders are not suitable as conceptual classes, because many images with similar concepts are not in the same folder and some images whose semantic contents are quite different are in the same folder. The existing folders in the Corel Photo Gallery were therefore ignored and all 10 800 images were manually divided into 80 concept groups, such as castle, aviation, bonsai, ship, steam-engine, train, dog, stalactite, autumn, cloud, iceberg, waterfall, elephant, primates, tiger, etc. These concept groups were only used in the evaluation of the results of our experiments. This large-scale groundtruth is used in both part B and part C.

Feature: Generally in a CBIR RF system images are represented by three main features: color [9], [14], [18], texture [10], [11], [19], and shape [9], [12]. Color information [18] is the most informative feature because of its robustness with respect to scaling, rotation, perspective, and occlusion. Texture information [11] can be another important feature and previous studies have shown that texture structure and orientation fit well the model of human perception, similarly with shape information [9].

For color [18], we selected hue, saturation, and value histogram. Hue and saturation were both quantized into 8 bins and value into 4 bins. A 128 dimensional Color coherence vector (CCV) [14] in Lab color space and a 9 dimensional color moment feature [10] in Luv color space were both employed.

For texture, a pyramidal wavelet transform (PWT) was extracted from the Y component in the YCrCb space. PWT results in a feature vector of 24 values. We also extracted the tree-structured wavelet transform (TWT) in the form of a 104 dimensional feature.

For shape, the edge direction histogram [12] was calculated from the Y component in YCrCb space. Edges were grouped into five classes, namely horizontal, 45° diagonal, vertical, 135° diagonal, and edges curving back on themselves.

Each of these features has its own power to characterize a type of image content. We combined color, texture, and shape features into a feature vector.

B. Statistical Experimental Results

Precision is widely used to evaluate retrieval performance. It is the ratio of the number of relevant images retrieved in the top N retrieved images. The error-bar is also given in the paper to evaluate the robustness of an evaluated algorithm. In our experiments, comparisons are made of the performances of the $\mathrm{SVM}_{\mathrm{Active}}$, PCA [3] with SVM, kernel PCA (KPCA) [13] with SVM, OCCA with SVM, and KEOCCA with SVM. We do not give out the comparison experimental results between the SVM and the proposed algorithms, because they have already been given in [20].

In our experiments, we use the Gaussian kernel

$$K(\mathbf{x}, \mathbf{y}) = e^{-\rho \|\mathbf{x}-\mathbf{y}\|^2} \tag{14}$$

in SVM, $\mathrm{SVM}_{\mathrm{Active}}$, KPCA, and KEOCCA. We chose the kernel parameters from a series of values according to the retrieval performance. For SVM and $\mathrm{SVM}_{\mathrm{Active}}$, $\rho = 1$. For KPCA and KEOCCA, $\rho = 9$. The retrieval performance is sensitive to the kernel parameters. We need to tune the kernel parameter and kernel type for different databases. Furthermore, we can also achieve much better performance by tuning the kernel parameter for different queries according to current kernel machine techniques.

The problem of mislabeling feedbacks is an open issue in small sample learning. The number of labeled samples is small so, when the number of the mislabeled samples is much less than the correct labeled samples, the learning machine can still obtain a correct model for the retrieval process by ignoring the

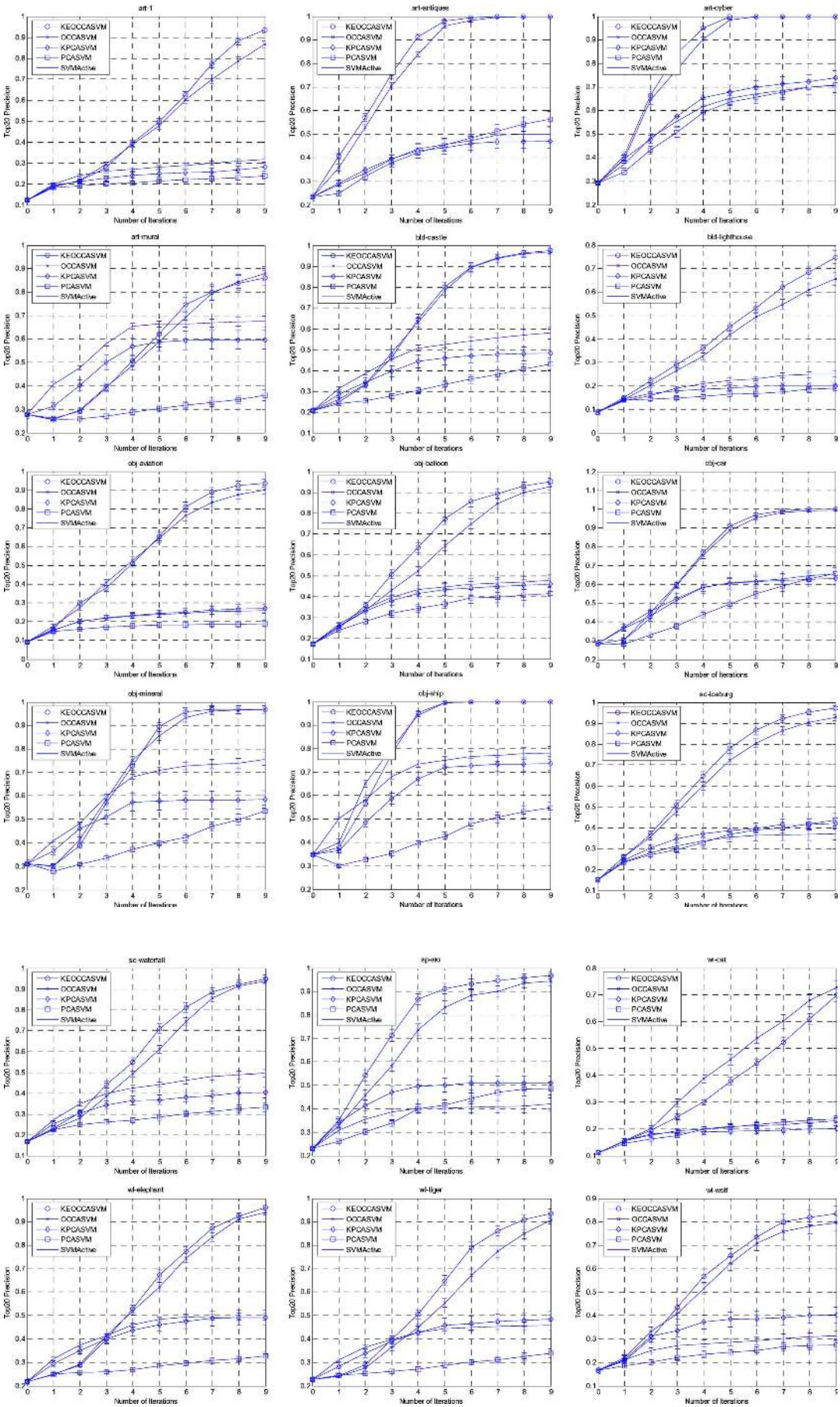

Fig. 1. Statistical retrieval results.

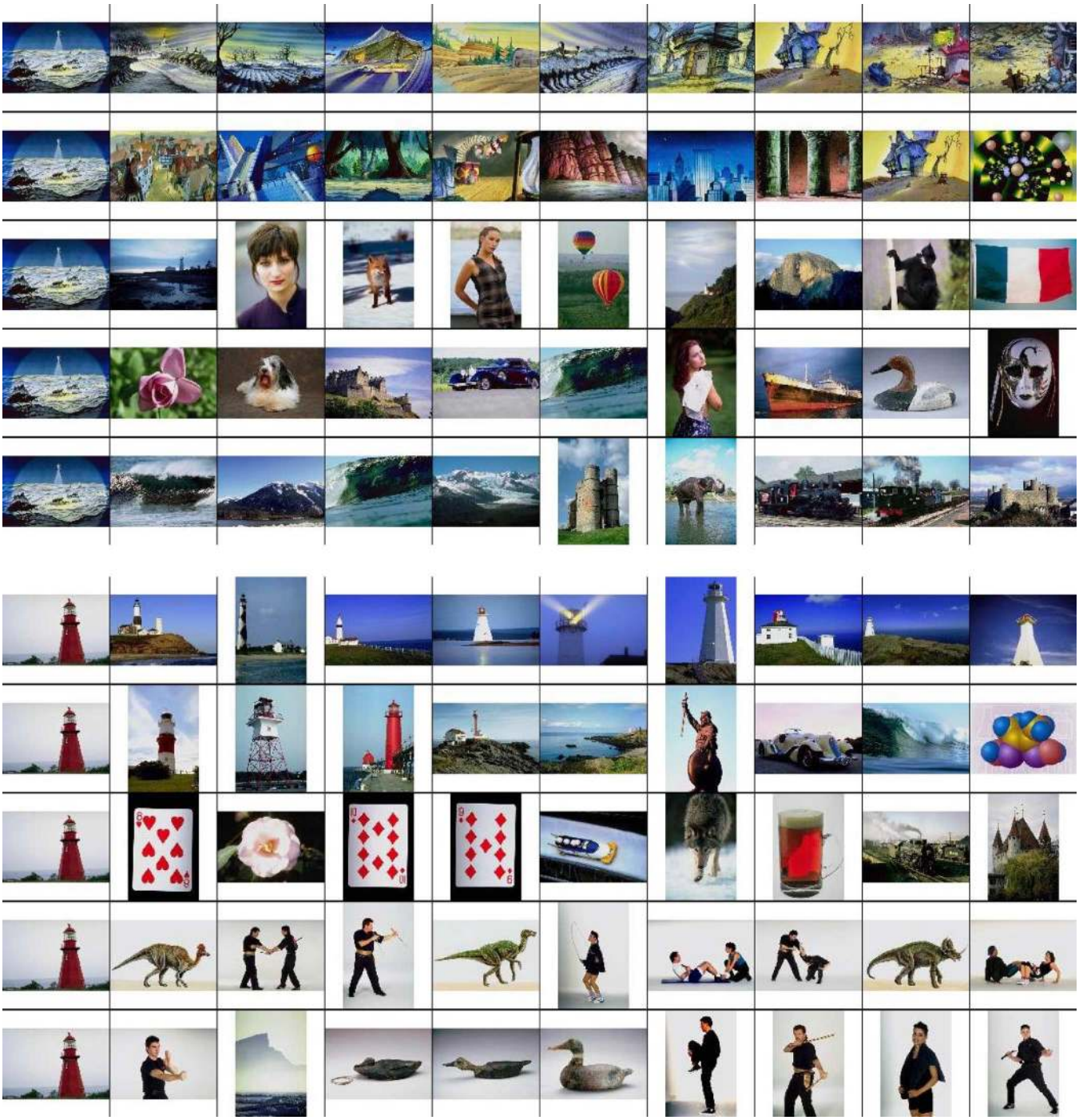

Fig. 2. Real-world retrieval results (top 12–20 results).

minor mistake. However, if a user mislabels too many images during the relevance feedback, the learning will be misled to an incorrect retrieval model. Thereafter, the retrieval system cannot give a satisfactory performance.

To avoid this kind of mistake, in our experiments, the computer does the relevance feedback iterations automatically using the conceptual classes described in Part A.

We conducted statistical experiments separately according to 18 concepts, which are art_pic, antiques, cyber, mural, castle, lighthouse, aviation, balloon, car, mineral, ship, iceberg, waterfall, ski, wildcat, elephant, tiger, and wolf. Each of these concepts is homogeneous. For each concept, 30 percent of images were selected as queries for statistical experiments. In an experiment there were none iterations. For each iteration the top 20 images resulting from the resort were examined serially from the top and each image was marked as correct or incorrect. The first five correct images and the first five incorrect images were then used as feedbacks unless fewer such images were found among the top 20 in which case the fewer number found was used as feedback.

As can be seen from Fig. 1, the proposed KEOCCA and OCCA combined with SVM-based RF algorithms consistently

outperform the KPCA and PCA combined with SVM-based RF algorithms. Furthermore, the new algorithms can also perform better than $\mathrm{SVM}_{\mathrm{Active}}$-based RF. Finally, the kernel based algorithms (KEOCCA and KPCA) combined with SVM-based RFs can consistently outperform their linear versions (OCCA and PCA) combined with SVM-based RFs, respectively. Each subfigure in Fig. 1 shows the precision with error-bar in the top 20 retrieval results for the 30% selected images as queries in each concept. Furthermore, in our experiments, we found that the numbers of support vectors of KEOCCA and OCCA combined with the SVM-based RFs were much less than those of the KPCA and PCA combined with the SVM-based RFs and that of $\mathrm{SVM}_{\mathrm{Active}}$-based RF. Previously, we found that the numbers of support vectors of the KEOCCA and OCCA combined with the SVM-based RFs were much less than that of the conventional SVM-based RF. Finally, the training errors in all RFs were zero in the feedback procedure.

In the proposed CBIR system and its RF algorithms, to fairly compare the experimental results, no indexing technique was used to improve either the speed or the precision. There are two major indexing styles and either of them has its intrinsic advantage. 1) Classification based indexing focuses on the improvement of retrieval precision. In this method, each image is assigned one or more distinct labels which are supported by the majority of people. Then, based on these labels, the indexing can be constructed through semantic classifications. Thereafter, the search results will cater to most of the users. 2) Low-level visual feature based indexing is employed to speed up the retrieval. There are many feature-based indexing approaches such as a variety of tree-based indexing structures for high-dimensional databases and VQ and VA methods etc. The promising way should be the combination of both feature and classification information for indexing structure, so that both speed and precision are enhanced.

C. Real-World Experimental Results

Based on the same groundtruth, we performed some real-world experiments. We randomly selected some images as the queries. For each query, we did RF iteration five times. For each RF iteration, we randomly selected some query concept relevant and irrelevant images as positive and negative feedbacks from the first screen shot, respectively. The number of the positive (negative) feedbacks is about 5. Meanwhile, they may not be the top retrieved images. We chose them according to whether the images share the same concept with the query or not. Fig. 2 shows the experimental results. The first image of each row in each subfigure is the query. Because the top 1 to top 11 retrieved results are usually query relevant, we only show the top 12 to top 20 results. The rows of each subfigure are the retrieval results given by KEOCCA SVM, OCCA SVM, KPCA SVM, PCA SVM, and $\mathrm{SVM}_{\mathrm{Active}}$, respectively. From this experiment, we can see that the proposed KEOCCA algorithm can work well practically.

VI. Conclusion

Recently, SVM has been widely applied in RF, which plays an essential role in improving the performance of CBIR. The main advantage of SVM is that it can generalize better than many other classifiers. To improve the conventional SVM based RF we propose the OCCA. OCCA can be implemented mainly in three steps: 1) to project all positive feedbacks onto their center and generate a subspace to represent the homogeneous concept shared by all positive feedbacks and the query image; 2) to project all the remaining images including the negative feedbacks onto this subspace; and 3) to construct an SVM classifier in this subspace and resort all the images based on new similarities. We then generalize the OCCA to the Hilbert space. The direct kernelization of OCCA is to use only positive feedbacks to construct the bases and in order to achieve additional kernel representation all images in the database could be used to construct the kernel bases. Due to the inefficient reason, we use the positive and negative feedbacks to construct the kernel bases. Using these bases, we define the kernel empirical OCCA (KEOCCA). Through experiments on a subset of Corel Photo Gallery with 10 800 images, we show that our new method can improve the conventional SVM-based RF consistently.

Acknowledgment

The authors would like to thank the editors and the three anonymous reviewers for their constructive comments, which helped improve the paper significantly

References

[1] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. IEEE Int. Conf. Image Process.*, Thessaloniki, Greece, 2001, pp. 815–818.

[2] I. J. Cox, L. Miller, P. Minka, V. Papthomas, and P. Yianilos, "The bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 20–37, Jan. 2000.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Boston: Academic Press, 1990.

[4] K. Goh, E. Chang, and W. Lai, "Concept-dependent multimodal active learning for image retrieval," in *Proc. ACM Int. Conf. Multimedia*, New York, 2004, pp. 564–571.

[5] G. Guo, A. K. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 811–820, Apr. 2002.

[6] X. He and P. Niyogi, "Locality preserving projections," *Adv. Neural Inf. Process. Syst.*, vol. 16, 2003.

[7] P. Hong, Q. Tian, and T. S. Huang, "Incorporate support vector machines to content-based image retrieval with relevant feedback," in *Proc. IEEE Int. Conf. Image Process.*, 2000, pp. 750–753.

[8] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Querying databases through multiple examples," in *Proc. Int. Conf. Very Large Data Bases*, New York, 1998, pp. 218–227.

[9] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233–1244, 1996.

[10] W. Y. Ma and H. J. Zhang, "Content-based image indexing and retrieval," in *Handbook of Multimedia Computing*, B. Furht, Ed. Boca Raton, FL: CRC Press, 1998.

[11] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[12] B. S. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Oct. 2001.

[13] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, 2001.

[14] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. ACM Int. Conf. Multimedia*, 1996, pp. 65–73.

[15] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Process.*, 1997, vol. 2, pp. 815–818.

[16] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool in interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Jul. 1998.
[17] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 1349–1380, Jan. 2000.
[18] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comp. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.
[19] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.s*, vol. SMC-8, no. 6, pp. 460–473, Jun. 1978.
[20] D. Tao and X. Tang, "Orthogonal complement component analysis for positive samples in SVM based relevance feedback image retrieval," in *Proc. IEEE Int. Conf. Comp. Vis. Pattern Recognit.*, 2004, vol. 2, pp. 586–591.
[21] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716–727, Apr. 2006.
[22] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2001, pp. 107–118.
[23] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[24] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 9, pp. 947–963, Sep. 2001.
[25] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *Proc. IEEE Int. Conf. Image Process.*, Thessaloniki, Greece, 2001, pp. 721–724.
[26] X. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using biasmap," in *Proc. IEEE Int. Conf. Comp. Vis. Pattern Recognit.*, 2001, vol. 1, pp. 11–17.
[27] X. Zhou and T. S. Huang, "Relevance feedback for image retrieval: A comprehensive review," *ACM Multimedia Syst. J.*, vol. 8, no. 6, pp. 536–544, 2003.

Dacheng Tao (M'04) received the B.Eng. degree from the University of Science and Technology of China (USTC), Hefei, China, the M.Phil. degree from the Chinese University of Hong Kong (CUHK), Hong Kong, and the Ph.D. degree from the University of London (UoL), London, U.K.

He is currently an Assistant Professor at the Department of Computing, Hong Kong Polytechnic University, Hong Kong. His research interests include artificial intelligence, biometrics, computer vision, data mining, machine learning, and visual surveillance. He has published extensively at IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON MULTIMEDIA (TMM), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE International Conference on Data Mining (ICDM), ACM International Conference on Multimedia (MM), ACM International Conference on Knowledge Discovery and Data Mining (KDD), etc., with Best Paper Awards and nominations. He is an editor of two books.

Dr. Tao received several Meritorious Awards from the Internationall Interdisciplinary Contest in Modeling, which is the highest level mathematical modeling contest in the world, organized by COMAP. He is an Associate Editor of *Neurocomputing* and a Co-Guest Editor of six special issues.

Xiaoou Tang (S'93–M'96–SM'02) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He was a Professor and the Director of Multimedia Laboratory in the Department of Information Engineering, the Chinese University of Hong Kong until 2005. Currently, he is the Group Manager of the Visual Computing Group at the Microsoft Research Asia. His research interests include computer vision, pattern recognition, and video processing.

Dr. Tang is the local chair of the IEEE International Conference on Computer Vision (ICCV) 2005, the area chair of ICCV'07, the program co-chair of ICCV'09, the general co-chair of the IEEE ICCV International Workshop on Analysis and Modeling of Faces and Gestures 2005. He is a Guest Editor of the Special Issue on Underwater Image and Video Processing for IEEE JOURNAL OF OCEANIC ENGINEERING and the Special Issue on Image- and Video-based Biometrics for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He is an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI).

Xuelong Li (M'02–SM'07) holds a permanent academic post at the University of London, London, U.K., as well as a Visiting Professor post at Tianjin University, Tianjin. China.

His research activities are partly sponsored by EPSRC, British Council, Royal Society, and Chinese Academy of Sciences. He has published around eighty scientific papers with several Best Paper Award and nominations. He is an editor of two books.

He is an Editor of 13 international journals, including IEEE Transactions on Circuits and Systems for Video TECHNOLOGY, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, and the IEEETRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, and a Guest Co-Editor of seven special issues. He serves as various chairs at dozens of conferences and a Programme Committee Member for around 70 conferences. He is a Reviewer for over 100 journals and conferences, including 11 IEEE TRANSACTIONS. He is a member of the academic committee of China Society of Image and Graphics, and a senior member several IEEE technical committees, including member-elected, IEEE Signal Processing Society Technical Committee on Machine Learning for Signal Processing.