

A Criterion for Optimizing Kernel Parameters in KBDA for Image Retrieval

Lei Wang, Kap Luk Chan, and Ping Xue

Abstract—A criterion is proposed to optimize the kernel parameters in Kernel-based Biased Discriminant Analysis (KBDA) for image retrieval. Kernel parameter optimization is performed by optimizing the kernel space such that the positive images are well clustered while the negative ones are pushed far away from the positives. The proposed criterion measures the goodness of a kernel space, and the optimal kernel parameter set is obtained by maximizing this criterion. Retrieval experiments on two benchmark image databases demonstrate the effectiveness of proposed criterion for KBDA to achieve the best possible performance at the cost of a small fractional computational overhead.

Index Terms—Content-based image retrieval, kernel-based biased discriminant analysis, kernel parameter optimization, relevance feedback.

I. INTRODUCTION

In the last few years, relevance feedback has been used to improve the performance of Content-Based Image Retrieval [1]–[8]. In the relevance feedback, a user labels the retrieved images subjectively and feeds them back to the retrieval system. By treating the labeled image samples as training data, learning is used to explore the high-level concepts encapsulated therein. However, a user is usually unwilling to label too many image samples in the retrieval, and the learning and retrieval performance is often constrained by the scarcity of training data. Different approaches have been proposed to get around this *small sample* problem. Some of them assume the independence among visual feature components or the Gaussianity of the distributions of image classes [9], [10]. Some apply active learning model, such as the Support Vector Machine (SVM)-based active learning [7]. Besides these, the unlabeled images are also incorporated to reduce the adverse effect of the scarcity of labeled image samples [11], [12].

Recently, Kernel-based Biased Discriminant Analysis (KBDA) has also been proposed for image retrieval to deal with the small sample problem [13]. In image retrieval, the positive and negative images are identified by a user based on high-level concepts, and they are often not linearly separable in a low-level visual feature space. Hence, KBDA employs the kernel trick to map the image data to a higher dimensional kernel space. According to Cover's theorem, the data are more likely to be linearly separable in the kernel space [14]. Afterwards, KBDA finds an optimal linear projection from the kernel space to a lower dimensional subspace in which the positive images are well separated from the negative ones. The "biased discriminant analysis" means that different strategies are applied to the two image classes. Comparatively, in the case of a small sample problem, the representation of the negative class is poorer due to its much more complex distribution. Hence, KBDA only requires the labeled positive image samples to be clustered closely in the subspace and does not apply this to the labeled negative ones. Instead, it only requires them to be away, as far as possible, from the positive ones. Based on this idea, an objective function is developed, and the optimal projection is found by maximizing it. After the projection is obtained, all images in a database are projected into a subspace,

and retrieval is performed therein. The Euclidean or Mahalanobis distance from the mean of the labeled positive ones is used as similarity metric. It is experimentally shown in [13] that KBDA achieves better retrieval performance than SVM, KFDA, and BDA, where no kernel trick is used.

However, the good performance of KBDA is based on correct *model selection* [15]. Models that are too complex will over-fit training data, whereas ones that are too simple cannot effectively model the structure therein. Both will result in poor retrieval performance. In KBDA, model selection is mainly to optimize the parameters of the employed kernel function. However, KBDA cannot do this itself. In [13], the kernel parameter Gaussian width σ is empirically set, and the retrieval performance of KBDA may not be maximized for a given training data set. Image retrieval with relevance feedback calls for maximizing retrieval performance for the available training samples because collecting them from a user is not easy. Therefore, the kernel parameters in KBDA should be optimized to approach this target. In addition, considering that the training data set dynamically changes with user feedback, these parameters should not be preset as they should be adjusted with respect to this set. Cross-validation is one of the commonly used methods to perform model selection [15]. However, it is not suitable due to the following considerations: 1) Image retrieval is an application requiring quick response, whereas cross-validation is often time-consuming. Hence, it cannot be used to adjust the kernel parameters in real time; 2) due to the small sample problem, the number of training samples is often so limited that these samples cannot ensure robust estimation and validation. Therefore, to help KBDA achieve the best possible retrieval performance, another way has to be found to perform efficient parameter optimization.

In this paper, a criterion is proposed to do this. It is known that different kernel parameter sets produce different mappings and, then, different kernel spaces. Recall the Cover's theorem, wherein the data in a higher dimensional space are more likely to be linearly separable than those in a lower dimensional space. By applying this principle to KBDA, it follows that if the positive and negative image samples are severely overlapped in the kernel space, it should not be expected that the negative ones can be pushed far away from the positives in a subspace. Hence, selecting the optimal kernel parameter set can be viewed as selecting a good kernel space in which the positive images have been well clustered, whereas the negative ones have been pushed far away from the positives. In other words, this kernel space has been well prepared for the subsequent processing in it. Following this idea, a criterion is developed to measure the goodness of a kernel space, and the optimal kernel parameter set is the one that has this criterion maximized. In addition, considering the quick response required in image retrieval, the computational complexity of the optimization process is discussed. Experimental results of retrieval on two benchmark image databases demonstrate that the proposed criterion can effectively predict the optimal kernel parameter set, and the optimization process only incurs a small fractional computational overhead with respect to the retrieval process. Retrieval with the KBDA using the optimized kernel parameters can obtain the best possible performance with reasonably fast response time. The rest of the paper is organized as follows. In Section II, KBDA is briefly introduced. In Section III, the criterion is proposed, and some issues are discussed. Section IV presents the experimental results and discussions. Finally, concluding remarks are given in Section V.

II. KBDA

Let \mathbf{x}_i ($\mathbf{x}_i \in \mathbb{R}^d$) denote a d -dimensional visual feature vector of the i th image in a database, where \mathbb{R}^d denotes the visual feature space.

Manuscript received August 1, 2003; revised March 18, 2004. This work was supported by A-STAR and NTU, Singapore, under Grants LIT 2002-4 and RGM 14/02. This paper was recommended by Associate Editor Jing Peng.

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798 (e-mail: elwang@ntu.edu.sg; eklchan@ntu.edu.sg; epixue@ntu.edu.sg).

Digital Object Identifier 10.1109/TSMCB.2005.846660

$k_{\theta}(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ denotes the employed kernel function, where $\langle \cdot, \cdot \rangle$ is the dot product, $\phi(\cdot)$ represents the associated mapping, and θ is the kernel parameter set. \mathbf{K} denotes the kernel matrix, and $\{\mathbf{K}\}_{i,j} = k_{\theta}(\mathbf{x}_i, \mathbf{x}_j)$. Let \mathcal{A} and \mathcal{B} be two data sets, and $\mathbf{K}_{\mathcal{A},\mathcal{B}}$ is the kernel matrix where $\{\mathbf{K}_{\mathcal{A},\mathcal{B}}\}_{i,j} = k_{\theta}(\mathbf{x}_i, \mathbf{x}_j)$ with the constraints of $\mathbf{x}_i \in \mathcal{A}$ and $\mathbf{x}_j \in \mathcal{B}$. The objective function of KBDA is

$$J(\mathbf{W}) = \frac{|\mathbf{W}^{\top} \mathbf{S}_{np}^{\phi} \mathbf{W}|}{|\mathbf{W}^{\top} \mathbf{S}_p^{\phi} \mathbf{W}|} \quad (1)$$

where \mathbf{W} denotes a linear projection from the kernel space \mathcal{F} , with a dimensionality of $d_{\mathcal{F}}$, to a subspace \mathcal{S} , with a dimensionality of $d_{\mathcal{S}}$. $|\cdot|$ denotes the determinant of a matrix. \mathbf{S}_{np}^{ϕ} and \mathbf{S}_p^{ϕ} are two scatter matrices in \mathcal{F} . \mathbf{S}_{np}^{ϕ} describes the scatter of negative image samples with respect to the mean of the positive ones, whereas \mathbf{S}_p^{ϕ} describes the scatter of positive ones with respect to this mean.

$$\begin{aligned} \mathbf{S}_{np}^{\phi} &= \sum_{\mathbf{x}_i \in \mathcal{D}_n} [\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}] [\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}]^{\top} \\ \mathbf{S}_p^{\phi} &= \sum_{\mathbf{x}_j \in \mathcal{D}_p} [\phi(\mathbf{x}_j) - \mathbf{m}_p^{\phi}] [\phi(\mathbf{x}_j) - \mathbf{m}_p^{\phi}]^{\top} \end{aligned} \quad (2)$$

where \mathcal{D}_p and \mathcal{D}_n denote the sets of the labeled positive and negative image samples, respectively. \mathbf{m}_p^{ϕ} denotes the mean vector of the labeled positive samples in \mathcal{F} . Maximizing $J(\mathbf{W})$ is solved as follows. Let $\mathbf{W} = \Phi \mathbf{A}$, where the i th column of Φ is $\phi(\mathbf{x}_i)$ ($i = 1, 2, \dots, |\mathcal{D}|$). Here, $\mathcal{D} = \mathcal{D}_p \cup \mathcal{D}_n$, and $|\mathcal{D}|$ denotes its size. \mathbf{A} is a coefficient matrix of the size $|\mathcal{D}| \times d_{\mathcal{S}}$. In this way, (1) becomes

$$J(\mathbf{A}) = \frac{|\mathbf{A}^{\top} \Phi^{\top} \mathbf{S}_{np}^{\phi} \Phi \mathbf{A}|}{|\mathbf{A}^{\top} \Phi^{\top} \mathbf{S}_p^{\phi} \Phi \mathbf{A}|} \triangleq \frac{|\mathbf{A}^{\top} \mathbf{P} \mathbf{A}|}{|\mathbf{A}^{\top} \mathbf{Q} \mathbf{A}|} \quad (3)$$

where $\mathbf{P} \triangleq \Phi^{\top} \mathbf{S}_{np}^{\phi} \Phi$, and $\mathbf{Q} \triangleq \Phi^{\top} \mathbf{S}_p^{\phi} \Phi$. The matrices of \mathbf{P} and \mathbf{Q} are expressed as

$$\mathbf{P} = \sum_{\mathbf{x}_i \in \mathcal{D}_n} (\mathbf{M}_i - \mathbf{M}_p)(\mathbf{M}_i - \mathbf{M}_p)^{\top} \quad (4)$$

where $\mathbf{M}_i = \Phi^{\top} \phi(\mathbf{x}_i)$, whereas $\mathbf{M}_p = \Phi^{\top} \mathbf{m}_p^{\phi}$.

$$\mathbf{Q} = \mathbf{K}_{\mathcal{D},\mathcal{D}_p} (\mathbf{I} - \mathbf{1}_p) \mathbf{K}_{\mathcal{D},\mathcal{D}_p}^{\top} \quad (5)$$

where \mathbf{I} is the identity matrix of the size $|\mathcal{D}_p| \times |\mathcal{D}_p|$, and $\mathbf{1}_p$ is the matrix in which all elements are $1/|\mathcal{D}_p|$. The dimensionality of the kernel space $d_{\mathcal{F}}$ is often high. Hence, \mathbf{S}_p^{ϕ} and, then, \mathbf{Q} are often singular. Numerical instability may occur when maximizing $J(\mathbf{A})$. To avoid this, \mathbf{Q} is regularized as $\mathbf{Q}_{\text{Reg}} = \mathbf{Q} + \mu \mathbf{K}_{\mathcal{D},\mathcal{D}}$, where μ is a small positive real number that is empirically set beforehand. The \mathbf{A} maximizing $J(\mathbf{A})$ is obtained by filling its columns with the eigenvectors of $\mathbf{Q}_{\text{Reg}}^{-1} \mathbf{P}$ corresponding to the larger $d_{\mathcal{S}}$ eigenvalues. $d_{\mathcal{S}}$ is the size of the set $\Lambda = \{\lambda_i | \lambda_i \geq \tau \lambda_{\max}\}$, where λ_{\max} denotes the largest eigenvalue, and τ is a small positive real number, e.g., 0.01. Thus, the projection of \mathbf{x} into the subspace is

$$\mathbf{y} = \mathbf{W}^{\top} \phi(\mathbf{x}) = \mathbf{A}^{\top} \mathbf{K}_{\mathcal{D},\{\mathbf{x}\}}. \quad (6)$$

In [13], the popular Gaussian radial basis function (RBF) kernel is used. It is

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (7)$$

where the kernel parameter is the Gaussian width σ , which controls the complexity of model [16].

III. OPTIMIZING KERNEL PARAMETERS FOR KBDA

Let us find a criterion to measure the goodness of a kernel space \mathcal{F} . Recall that \mathbf{S}_{np}^{ϕ} and \mathbf{S}_p^{ϕ} are the matrices characterizing the scatter of the positive and negative training samples in \mathcal{F} . Because the high dimensionality of \mathcal{F} often makes both \mathbf{S}_{np}^{ϕ} and \mathbf{S}_p^{ϕ} singular, their determinant information cannot be used here (the determinant-based criterion can be used in KBDA because the scatter is measured in a lower dimensional subspace). Hence, the proposed criterion is developed based on the traces of \mathbf{S}_{np}^{ϕ} and \mathbf{S}_p^{ϕ} instead. Let $\text{Sum}(\cdot)$ denote the summation of all elements in the matrix therein. The traces can be derived as follows:

$$\begin{aligned} \text{tr}(\mathbf{S}_{np}^{\phi}) &= \text{tr} \left[\sum_{\mathbf{x}_i \in \mathcal{D}_n} (\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}) (\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi})^{\top} \right] \\ &= \sum_{\mathbf{x}_i \in \mathcal{D}_n} [\phi(\mathbf{x}_i)^{\top} \phi(\mathbf{x}_i) - 2\phi(\mathbf{x}_i)^{\top} \mathbf{m}_p^{\phi} + \mathbf{m}_p^{\phi \top} \mathbf{m}_p^{\phi}] \\ &= \sum_{\mathbf{x}_i \in \mathcal{D}_n} k_{\theta}(\mathbf{x}_i, \mathbf{x}_i) - \frac{2}{|\mathcal{D}_p|} \text{Sum}(\mathbf{K}_{\mathcal{D}_n, \mathcal{D}_p}) \\ &\quad + \frac{|\mathcal{D}_n|}{|\mathcal{D}_p|^2} \text{Sum}(\mathbf{K}_{\mathcal{D}_p, \mathcal{D}_p}) \end{aligned} \quad (8)$$

$$\begin{aligned} \text{tr}(\mathbf{S}_p^{\phi}) &= \text{tr} \left[\sum_{\mathbf{x}_i \in \mathcal{D}_p} (\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}) (\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi})^{\top} \right] \\ &= \sum_{\mathbf{x}_i \in \mathcal{D}_p} [\phi(\mathbf{x}_i)^{\top} \phi(\mathbf{x}_i) - 2\phi(\mathbf{x}_i)^{\top} \mathbf{m}_p^{\phi} + \mathbf{m}_p^{\phi \top} \mathbf{m}_p^{\phi}] \\ &= \sum_{\mathbf{x}_i \in \mathcal{D}_p} k_{\theta}(\mathbf{x}_i, \mathbf{x}_i) - \frac{1}{|\mathcal{D}_p|} \text{Sum}(\mathbf{K}_{\mathcal{D}_p, \mathcal{D}_p}). \end{aligned} \quad (9)$$

The criterion is then given as

$$J(\theta) = \frac{\text{tr}(\mathbf{S}_{np}^{\phi})}{\text{tr}(\mathbf{S}_p^{\phi})}. \quad (10)$$

It measures the ratio of the scatter of the negative images to that of the positive ones. This criterion is used in this paper to reflect whether the positive images have been well clustered and the negative ones have been pushed away from the positive ones as far as possible.

Like KBDA, the proposed criterion also encounters numerical instability when $\text{tr}(\mathbf{S}_p^{\phi})$ approaches zero. Enlightened from the L2-norm-based soft margin in SVM, the following modified kernel is used:

$$\tilde{k}_{\theta}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} k_{\theta}(\mathbf{x}_i, \mathbf{x}_j), & \text{when } i \neq j \\ k_{\theta}(\mathbf{x}_i, \mathbf{x}_j) + \mu', & \text{when } i = j. \end{cases} \quad (11)$$

where μ' is a regularization parameter. As a result, a regularization item is appended to $\text{tr}(\mathbf{S}_p^{\phi})$, and the numerical instability is avoided. In this way, the optimal kernel parameter set θ^* can be expressed as

$$\theta^* = \arg \max_{\theta \in \Theta} [J(\theta)] = \arg \max_{\theta \in \Theta} \left[\frac{\text{tr}(\mathbf{S}_{np}^{\phi})}{\text{tr}_{\text{Reg}}(\mathbf{S}_p^{\phi})} \right] \quad (12)$$

where Θ denotes the parameter space, and $\text{tr}_{\text{Reg}}(\mathbf{S}_p^{\phi})$ denotes the regularized version.

Now, we discuss some of the issues on this criterion.

- 1) Note that the kernel parameters cannot be optimized by maximizing the KBDA objective function in (1), although it is also a function of them. This is explained as follows. Kernel parameter optimization performs model selection for a kernel-based learning method. As pointed out by the principle of Structural Risk Minimization (SRM) [14], the optimal model commonly does not correspond to the one that maximizes or minimizes the objective function

on a given training data set. This is because the learning machines in this case often overfit the training data with over complicated models rather than correctly generalizing the nature of data. Consequently, the learning machines cannot effectively predict the attributes of test data, such as the class labels. This applies to KBDA. It is observed that when the Gaussian RBF kernel is used, the KBDA objective function value monotonically increases with decreasing value of the Gaussian width σ . A smaller σ often means a more complicated mapping function. Therefore, kernel parameter optimization has to be done via other criteria, e.g., the proposed one.

- 2) The computational load of kernel parameter optimization should not be heavy because image retrieval requires quick response. Here, the maximization of $J(\theta)$ (or the minimization of $-J(\theta)$) is solved by applying a nonlinear optimization technique. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) Quasi-Newton method [17] is often favored because it has fewer iterations before stopping. The computational load in each iteration is largely due to evaluating $J(\theta)$, which involves calculating $\mathbf{K}_{\mathcal{D}_p, \mathcal{D}_p}$ and $\mathbf{K}_{\mathcal{D}_n, \mathcal{D}_p}$, and the complexity is $O(\max\{|\mathcal{D}_p|^2, |\mathcal{D}_p| \cdot |\mathcal{D}_n|\})$ for a given visual feature. In image retrieval, both $|\mathcal{D}_p|$ and $|\mathcal{D}_n|$ often remain small, even after several rounds of feedback. In addition, they are independent of the size of the image database. Hence, it can be expected that the optimization process will not take much time, and it will not significantly slow down the response required. This can be seen from the experimental results.
- 3) It is found that the proposed criterion cannot effectively optimize the value of μ' in (11) because it is a monotonic function of μ' (the proof is omitted here). To avoid introducing a new parameter, μ' is empirically set to the value of μ given in KBDA beforehand. This is based on the consideration that both μ' and μ are regularization parameters and that they are both small positive real numbers. The experimental results will show that this empirical setting will not significantly compromise the effectiveness of the proposed criterion. (Ideally, regularization parameters should be optimized by using cross-validation. However, they are often empirically set in practice due to the little adverse effect on the algorithm performance but significant saving on computation.)
- 4) Finally, the relationship between kernel parameter optimization and KBDA is highlighted as follows: 1) KBDA cannot optimize the kernel parameters automatically, and the above criterion is proposed to do this. In other words, kernel parameter optimization serves KBDA; 2) Kernel parameter optimization finds the optimal mapping from the visual feature space to a higher dimensional kernel space. Differently, KBDA finds the optimal projection from the kernel space to a lower dimensional subspace. Hence, their goals are different. Kernel parameter optimization does not conflict with KBDA, and it cannot replace KBDA for image retrieval.

The retrieval procedure is briefly described as follows. At the beginning, an initial retrieval result is obtained based on a query given by a user. Afterwards, the labeled image samples are collected via relevance feedback. After each round of feedback, by treating the labeled samples as training data, optimize the kernel parameters, and perform retrieval by the KBDA with these parameters. It is expected that this dynamic kernel parameter optimization helps KBDA achieve the best possible retrieval performance.

IV. EXPERIMENTAL RESULTS

A. Image Databases and Visual Features

In this experiment, two benchmark image databases are used. One is selected from a subset of *Corel* Stock Photos. This selection removes the image classes with very abstract concepts like “Thailand” or “Autumn” because they cannot be learned by the retrieval algorithms at the present stage. After this selection, a database including 4800 general color images is constructed. These images form 48 classes, and each class has 100 image samples. The ground truth is based on the labels of CDs by *Corel*. Even with generally agreeable semantics, these image categories still exhibit sufficient intraclass variations and interclass overlap in the visual feature space. A perceptually uniform color space *CIE-Lab* is used, and a feature of color moments is defined for each image [18]. It consists of the mean, variance, and skewness of the pixel values in an image along L , a , and b axes, respectively. The other database is selected from the aerial photo image database provided by the Vision Research Lab of the University of California at Santa Barbara (UCSB) [19]. The original database includes 40 large aerial photos. Each of them is divided into the subimages of size 128×128 , and 40 image classes are formed. In this experiment, 100 subimages are taken from each class, and a database of 4000 images is constructed. The Gabor-based texture feature provided in [19] is also used here. The two predefined image databases are the ground truth for performance evaluation. *Precision* and *Recall* are used to measure the performance of a retrieval. In this experiment, *Precision* describes the performance for a given round of feedback, and *Precision* of top k is the percentage of true positive images among the top k retrieved at that round. *Recall* is used to describe the performance against the rounds of feedback. *Recall* of top k for the i th round is the percentage of true positive images retrieved so far among the total positive ones in an image database.

B. Experiment Setting and Procedure

In the experiment, the Gaussian RBF kernel used in [13] is selected for consistency, and the Gaussian width σ is optimized. Two regularization parameters are set as $\mu' = \mu = 0.1$. The commonly used “Query by example” retrieval model is adopted, and the initial retrieval result is obtained by an Euclidean search around the given query example. In each round of relevance feedback, the top 20 retrieved images are labeled and excluded from the database for the next retrieval. Four rounds of feedback are performed in total. To achieve robust statistics, KBDA and kernel parameter optimization will not be performed until at least five positive and five negative image samples are labeled. In this period, the Euclidean search around the given query example is used instead. The BFGS Quasi-Newton method is used to find the σ that maximizes the proposed criterion. The initial value of σ is set to $\sigma_0 = 1.0$, and the stop criterion is $|J(\sigma_{i+1}) - J(\sigma_i)| \leq 10^{-6} J(\sigma_i)$.

The procedure of the experiment is described as follows.

- 1) Treat the i th image class as the positive and the remaining as the negative.
- 2) Select a sample from the positive class as a query to launch retrieval, and perform the Euclidean search.
- 3) Sort the images in the database according to the ascending order of the corresponding similarity distances.
- 4) According to the ground truth, label the top 20 images to simulate user feedback. After that, these labeled images are added into the current training data set and removed from the database for the next retrieval.
- 5) Based on the available training data, maximize the proposed criterion to find the optimal σ . In addition, calculate the criterion value under different values of σ .

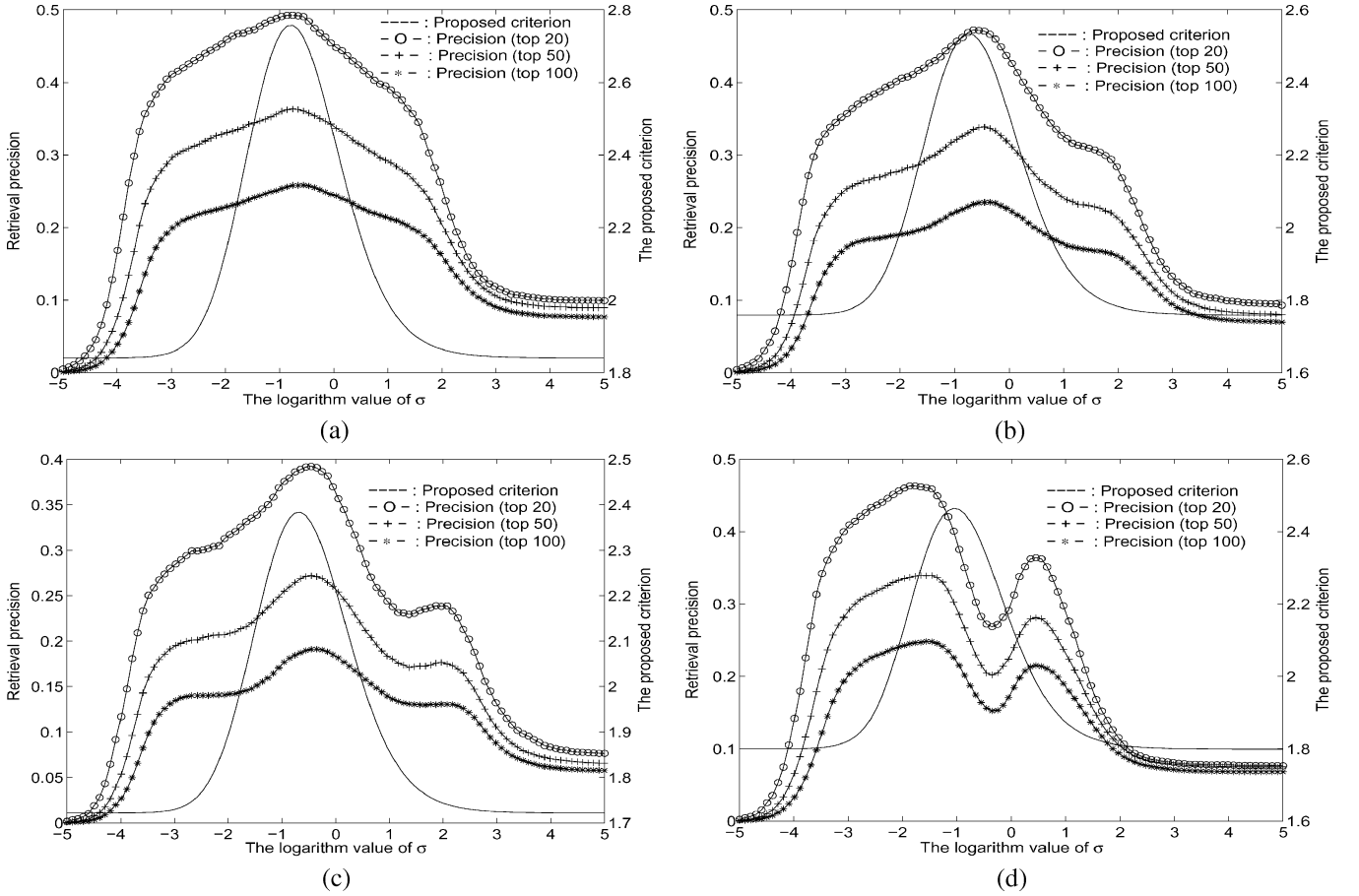


Fig. 1. Proposed criterion $J(\sigma)$ and the retrieval performance of KBDA against σ (Corel image database). (a) Second feedback. (b) Third feedback. (c) Fourth feedback. (d) First feedback.

- 6) Perform KBDA for image retrieval by using the optimized σ , and evaluate retrieval performance. In addition, evaluate the retrieval performance of the KBDA using different values of σ .
- 7) Redo steps 3 to 6 four times to simulate four rounds of user feedback.
- 8) To accumulate statistics, redo steps 2 to 7 30 times. The retrieval sessions where KBDA is performed in all four rounds of feedback are selected, and the obtained retrieval performance and the criterion values are averaged, respectively.
- 9) Redo steps 1 to 8 until each class has been treated as the positive at least once, and the obtained retrieval performance and criterion values are averaged.

C. Results and Discussions

Figs. 1–4 show the experimental results on the Corel image database. Fig. 1 plots the values of the proposed criterion and the retrieval performance of KBDA against σ in four rounds of feedback. Fig. 1(a) is for the second round. (The one for the first round will be discussed later.) As shown, each *Precision* curve fluctuates significantly against σ , and a maximum can be found. By cross-referencing these curves to that of the proposed criterion, it can be seen that they show similar profiles, and the σ values corresponding to their maxima lie within a close range. This indicates that the retrieval performance of KBDA varies with σ , and the σ corresponding to the best retrieval performance can be predicted by the proposed criterion well. Similar results can be seen from Fig. 1(b) and (c). In addition, note that *Precision* degrades slightly from the second to fourth feedback. This is because the images labeled

by the user in each round of feedback are excluded from the database used in the next retrieval, and they often include some positive images.

It is also observed that this criterion cannot predict the optimal σ in the first round of feedback. This result can be explained as follows. In the first round, the labeled images are retrieved by an Euclidean search around the initial query. Due to the Euclidean search, they lie in a small hypersphere in the visual feature space. Consequently, they only represent a very localized distribution of image data. In this case, both proposed criterion and KBDA cannot function well. Therefore, the maxima of these curves do not necessarily correspond to each other.

Fig. 2 compares the proposed criterion and the KBDA objective function against σ . It can be seen that the objective function value monotonically increases with decreasing value of σ rather than showing a maximum as the proposed criterion. By referring to Fig. 1, the effect of “over-fitting” on training data can be observed. That is, the retrieval performance becomes poorer, although the objective function value becomes larger. This result demonstrates that maximizing the objective function will not give rise to the optimal σ .

To test the performance of the proposed criterion in practical applications, the BFGS Quasi-Newton method is used to find the optimal σ in real time. Afterwards, KBDA with the optimized σ is performed for retrieval. Also, the best performance of KBDA with a brutal search of σ is listed for comparison. In addition, the retrieval performance of using kernel parameter optimization only is plotted. That is, with the optimized σ , the Euclidean search around the mean of the labeled positive images is applied in the induced kernel space instead of the subspace obtained by KBDA. This is to demonstrate that KBDA cannot be replaced by kernel parameter optimization. Fig. 3(a)–(c) show the *Recall*

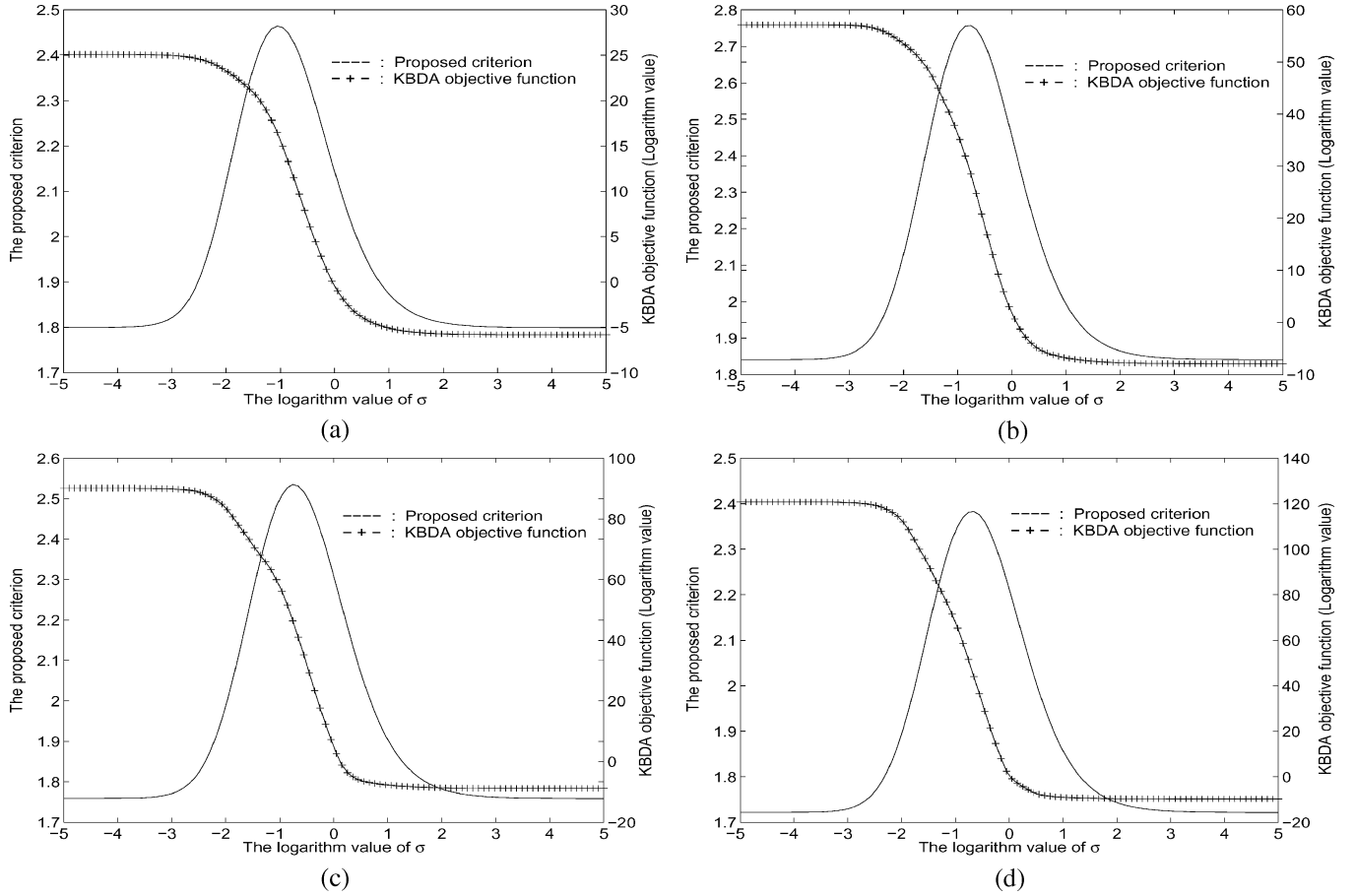


Fig. 2. Proposed criterion $J(\sigma)$ and the KBDA objective function against σ (Corel image database). (a) First feedback. (b) Second feedback. (c) Third feedback. (d) Forth feedback.

of top 20, 50, and 100 against the times of feedback, respectively. It can be seen that the KBDA with the optimized σ achieves the performance comparable with the best one obtained by a brutal search of σ . Again, this result shows the effectiveness of the proposed criterion. It can also be seen that the retrieval performance of using kernel parameter optimization only is much lower than that using the KBDA. Fig. 3(d) plots the retrieval time. A Linux system with Pentium Xeon 2.8 GHz and 2 G memory is used. Compared with the time taken by KBDA, the time taken by kernel parameter optimization is insignificant. Moreover, this time is mainly affected by the number of labeled image samples rather than the size of the image database. Hence, even if a larger database is used, it can be expected that this time still holds.

Finally, to justify the empirical setting of $\mu' = \mu$, Fig. 4 plots the retrieval performance for $\mu' = \mu = 10^{-2}$ and 10^{-3} (Fig. 3 shows the case for $\mu' = \mu = 10^{-1}$). For simplicity, *Recall* of top 20 is plotted only. It is found that the retrieval performance of the KBDA with the optimized σ drops slightly when μ' and μ decrease. This may be due to the fact that μ' is too small to achieve effective regularization. However, this retrieval performance is still close to the best one obtained in a brutal search against σ . These results indicate that the setting of $\mu' = \mu$ does not bring much negative effect on the effectiveness of the proposed criterion.

Fig. 5 plots the experimental results on the aerial photo image database. As shown in Fig. 5(a), KBDA with the optimized σ still achieves the performance comparable to the best one obtained by a brutal search against σ . Fig. 5(b) illustrates the retrieval time. Due to the higher dimension of Gabor texture features, the time is a little longer than that on the general color image database. However, the time taken by kernel

parameter optimization is still much less than that taken by the KBDA. The results similar to those shown in Figs. 1 and 2 are also obtained from the aerial photo image database. They are not listed here to avoid redundancy.

Beside these, the KBDA with the optimized σ is also compared with two of the state-of-the-art retrieval methods: SVM and SVM active learning [6], [7]. This demonstrates that it is worth performing kernel parameter optimization for KBDA. All of the three methods are kernel based. They all treat the image samples labeled in relevance feedback as training data and those in the database as test data. Here, the Gaussian RBF kernel is also used in SVM and SVM active learning. To achieve a fair comparison, the Gaussian width σ and the regularization parameter C in each SVM classifier are also optimized in each round of feedback by using Leave-One-Out (LOO). Note that they cannot be well tuned like this in practice because this procedure is very time-consuming. Often, they are heuristically or empirically set instead. The three retrieval methods are compared under two procedures. Procedure A is just the one used for the previous experiments, which is described in Section IV-B. Procedure B is different from A at the initial stage of retrieval. Instead of the Euclidean search, random sampling is used to select 20 images from the database for the user to label. This procedure is proposed by [7] to ensure effective SVM active learning. The database of 4800 general color images is used. From each class, ten images are randomly selected, and each of them is used as a query once. The obtained retrieval performance values are averaged over these 480 retrieval sessions.

Fig. 6 shows the comparison on the two procedures. As shown in Fig. 6(a), the *Recall* value for each method increases with the times of

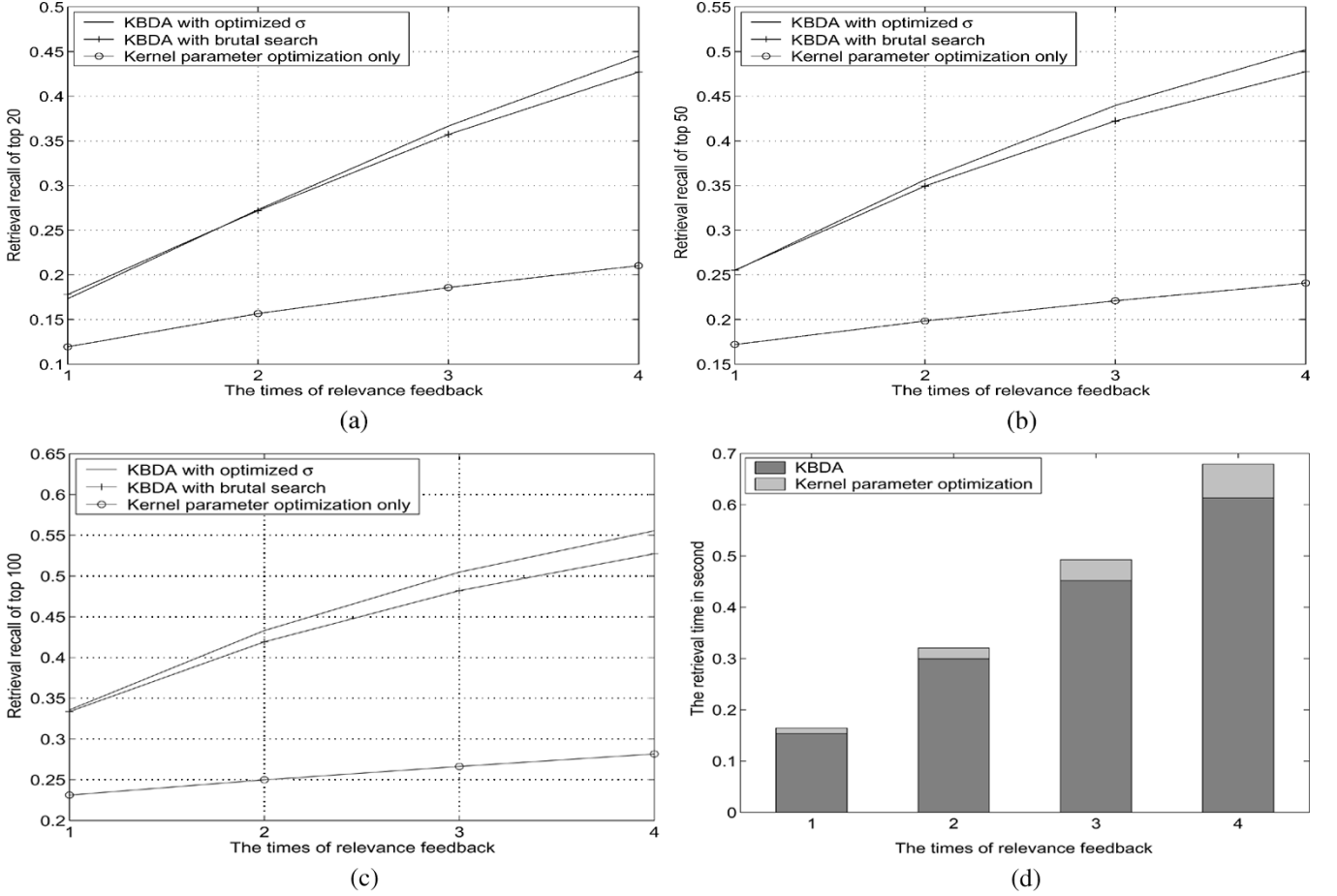


Fig. 3. Retrieval performance & the retrieval time (*Corel* image database). (a) Recall of top 20. (b) Recall of top 50. (c) Recall of top 100. (d) Retrieval time.

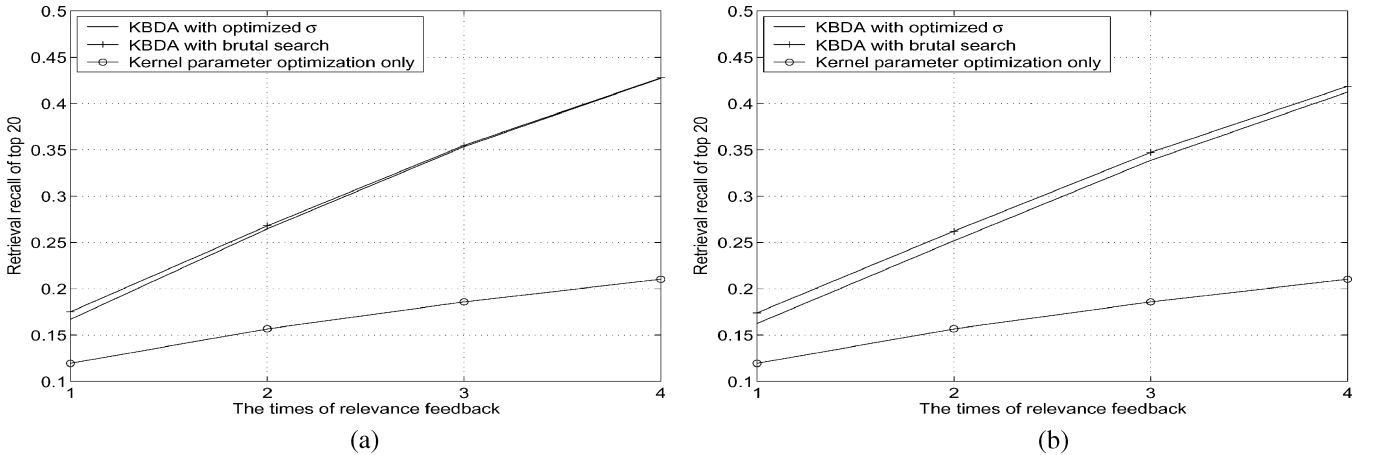


Fig. 4. Retrieval performance when different values of μ' and μ are used (*Corel* image database). (a) $\mu' = \mu = 10^{-2}$. (b) $\mu' = \mu = 10^{-3}$.

feedback, and the KBDA with the optimized σ always achieves the highest value. This is because KBDA applies different strategies to the two image classes. In addition, the kernel parameter optimization makes a large contribution. In the Fig. 6(b), the three methods have similar performance in general. However, KBDA still shows a minor advantage. In addition, SVM active learning shows no advantage over SVM in both procedures, at least at the top four rounds of feedback, although it often has slightly lower test error. This may imply that better classification does not necessarily give rise to more accurate ranking of

images. Note that these conclusions can also be drawn from the results of the Recall of the top 50 and 100, although they are not listed.

In summary, the above experimental results demonstrate the effectiveness of the proposed criterion and the good performance of the KBDA with it. The process of kernel parameter optimization incurs a small fractional computational overhead, and it does not significantly increase the response time. Compared with the well-tuned SVM-based retrieval methods, the KBDA with the optimized σ achieves better or at least comparable retrieval performance.

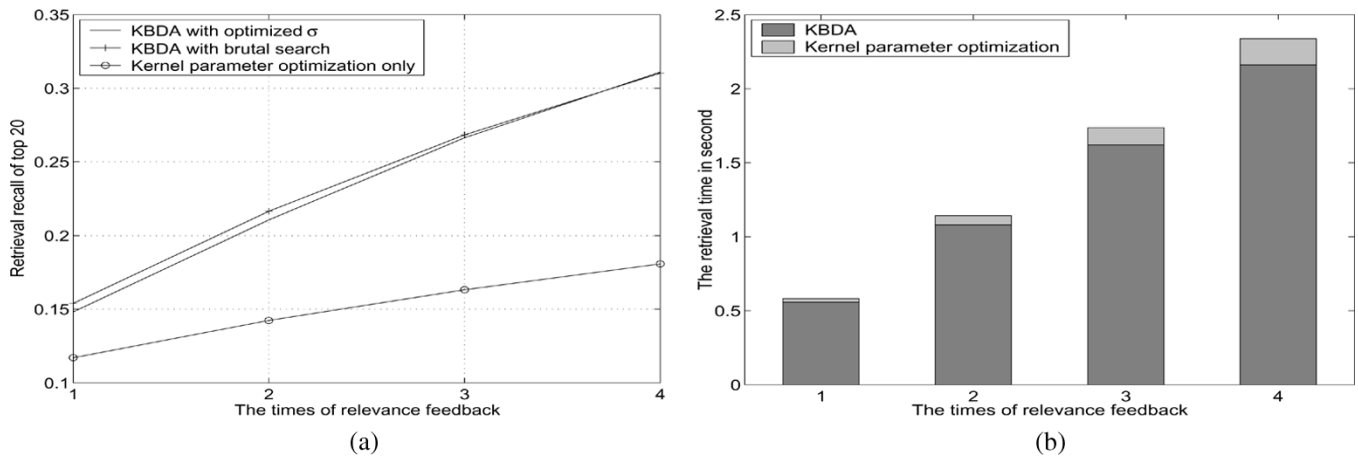


Fig. 5. Retrieval performance and the retrieval time (Aerial photo image database). (a) Recall of top 20. (b) Retrieval time.

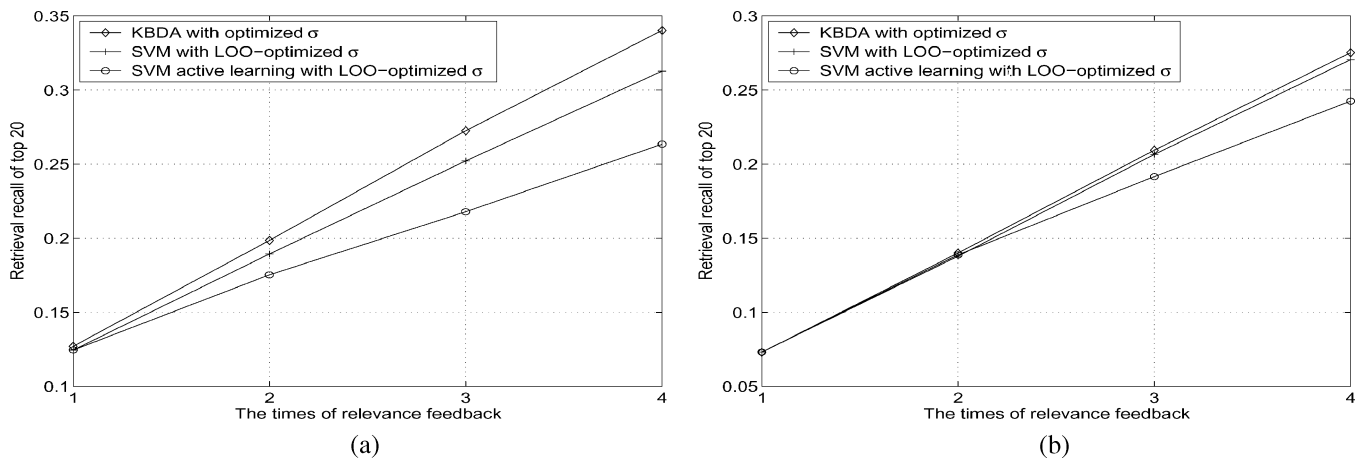


Fig. 6. Comparison of KBDA, SVM, and SVM active learning. (a) Recall of top 20, Procedure A. (b) Recall of top 20, Procedure B.

V. CONCLUSION

This work aims to maximize the retrieval performance of KBDA by optimizing the kernel parameters therein, and a criterion is developed to achieve this target. This criterion measures the goodness of the kernel space, and the optimal parameter set is obtained by maximizing this criterion. The computational load incurred by kernel parameter optimization is light, and the fast response required in image retrieval is maintained. Experimental results on two benchmark image databases demonstrate the effectiveness of the proposed criterion in optimizing the kernel parameters. Comparison with state-of-the-art retrieval methods shows the advantage of the KBDA using this criterion.

REFERENCES

- [1] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1998.
- [2] N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval systems," in *Proc. Neural Inf. Process. Syst.*, Denver, CO, 1999.
- [3] T. P. Minka and R. W. Picard, "Interactive learning using a 'society of models'," Media Laboratory Perceptual Computing Section, Mass. Inst. Technol., Cambridge, MA, Tech. Rep. 349, 1995.
- [4] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *Comput. Vision Image Understanding*, vol. 75, no. 1, pp. 150–164, 1999.
- [5] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Self-organizing maps as a relevance feedback technique in content-based image retrieval," *Pattern Anal. Applicat.*, vol. 4, no. 2–3, pp. 140–152, 2001.
- [6] G. Guo, H. Zhang, and S. Z. Li, "Distance-from-boundary as a metric for texture image retrieval," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2001.
- [7] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. 9th ACM Int. Conf. Multimedia*, Ottawa, ON, Canada, 2001, pp. 107–118.
- [8] X. Huang, S.-C. Chen, and M.-L. Shyu, "Incorporating real-valued multiple instance learning into relevance feedback for image retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 1, Baltimore, MD, 2003, pp. 321–324.
- [9] C. Meilhac and C. Naster, "Relevance feedback and category search in image retrieval," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, vol. 1, 1999, pp. 512–517.
- [10] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Process.*, Santa Barbara, CA, 1997.
- [11] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant-EM algorithm with application to image retrieval," in *Proc. Int. Conf. Computer Vision Pattern Recogn.*, vol. 1, Hilton Head Island, SC, 2000.
- [12] L. Wang and K. L. Chan, "Bootstrapping SVM active learning by incorporating unlabeled images for image retrieval," in *Proc. Int. Conf. Comput. Vision Pattern Recogn.*, vol. 1, 2003, pp. 629–634.
- [13] X. S. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using BiasMap," in *Proc. Int. Conf. Comput. Vision Pattern Recogn.*, vol. 1, 2001, pp. 11–17.
- [14] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995.
- [15] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. New York: Wiley, 2002.
- [16] N. Cristianini, C. Campbell, and J. Shawe-Taylor, "Dynamically adapting kernels in support vector machines," *Adv. Neural Inform. Process. Syst.*, vol. 11, pp. 204–210, 1998.
- [17] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1988.
- [18] M. Stricker and M. Orengo, "Similarity of color images," in *Proc. SPIE Storage Retrieval Image Video Databases*, vol. 2420, 1995, pp. 381–392.
- [19] <http://vision.ece.ucsb.edu/datasets/> [Online]