Retrieval with Knowledge-driven Kernel Design: An Approach to Improving SVM-based CBIR with Relevance Feedback

Lei Wang*, Yan Gao, Kap Luk Chan, Ping Xue
School of EEE, Nanyang Technological University
Singapore
{elwang.assoc, gaoy0004, eklchan, epxue}@ntu.edu.sg

Wei-Yun Yau
Institute for Infocomm Research
Singapore
wyyau@i2r.a-star.edu.sg

Abstract

The performance of SVM-based image retrieval is often constrained by the scarcity of training samples. The total number of image samples labelled by users in a retrieval session is very limited, and these small number of labelled samples cannot effectively represent the true distributions of positive and negative image classes, especially for the negative image class. This paper proposes a novel approach to deal with this problem. Instead of treating it as a problem, the mere existence of the small number of labelled images and their desired distribution in the kernel space is considered as prior knowledge from image retrieval to aid the design of the kernel used by SVMs. This is achieved by maximizing a criterion, such as one based on scatter matrices, through gradient-based search methods, incurring very little computational overhead to real-time retrieval process. Experimental results on two benchmark image databases demonstrate the improved retrieval performance by the dynamically designed kernel and hence the effectiveness of the proposed approach for SVM based image retrieval.

1. Introduction

In recent years, Support Vector Machines (SVMs) have been used in Content-Based Image Retrieval (CBIR) to learn the high-level concepts encapsulated in user feedback [2, 8, 9, 10]. In the process, a user is often asked to provide some "positive" and "negative" image examples via relevance feedback. These labelled image samples are then used as training data to train an SVM classifier to perform a bi-class classification of positive (relevant) and negative (irrelevant) images. Those images having the larger positive decision values are retrieved as positive images. In SVMs, through a kernel, the training data are *implicitly* mapped from a low-level visual feature space to a kernel space, and an optimal separating hyperplane is determined therein. This mapping is often nonlinear, and the dimensionality of the kernel space can be very high or even infinite. The nonlinearity and the high dimensionality help SVMs achieving excellent classification performance, especially for linearly nonseparable patterns. This is very useful for image retrieval in which 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.

However, the number of labelled images is often small because a user cannot bear to label too many images. As a result, this small number of labelled images cannot effectively represent the true distributions of the positive and negative image classes. Especially, in the case of *small sample*, the representation of the negative class is much poorer due to its much more complex distribution. After all, the positive images share a common concept in the user's mind, leading to some degree of aggregation when reasonably good visual features are used. However, negative images have heterogeneous concepts except in that all of them are not positive. The excellent classification performance of SVMs largely depends on sufficient number of representative training samples. Due to the scarcity of training samples in image retrieval, SVMs cannot produce the expected retrieval performance. This is often called the *small sample* problem in image retrieval.

The present literature handles this problem by the following ways. In [9], an active learning mode is used. However, to ensure effective active learning, the user has to label more images before retrieval starts. In [2], Euclidean search is combined to correct SVM's decision instead of improving the performance of SVMs. In our previous work [10], a transductive SVM is used by incorporating unlabelled data. Nevertheless, it is also observed that the retrieval performance is not stable enough. In [7], three techniques of random sampling, random space, and bagging are used to

*This author is now with Research School of Information, Science and Engineering (RSISE), The Australian National University. *Email address*: Lei.Wang@mail.rsise.anu.edu.au.

combine multiple SVMs to address the small sample problem. As seen, different kinds of approaches have been used by collecting more samples from users, incorporating unlabelled samples, or combining other techniques. However, the SVM classifier itself is often ignored there. This is especially true for the kernel function, which plays a crucial role in SVMs. The above work simply takes a commonly used kernel into SVMs, and does not explore its potential for improving the retrieval performance by SVMs.

It has been well recognized in machine learning that, when designing an SVM classifier, incorporating prior knowledge into the employed kernel is important because this often improves its classification performance, especially when training samples are scarce [5]. Also, this is a very efficient way because it needs the minimum change on the SVM classifier itself. In this paper, instead of taking the *small sample* as a problem, we think of it as a prior knowledge from image retrieval with relevance feedback, and incorporate this knowledge into the kernel function. Note that each kernel induces a kernel space. Incorporating prior knowledge into a kernel is to reflect this prior knowledge in its kernel space. In the *small sample* case, the negative sample should not be treated as a class that forms a cluster. Instead, what we can do is to push the negative samples away, as far as possible, from the positive samples while keeping the positive ones well clustered. This is also the motivation behind the Kernel based Biased Discriminant Analysis (KBDA) [12]. We thus develop a criterion based on the above thinking in the kernel space, through which such prior knowledge of *small sample* is reflected therein. Given a set of labelled samples collected through user feedback, the kernel incorporating the prior knowledge is considered as the one whose kernel space has this criterion maximized. Once the labelled sample set is expanded with new feedbacks, this criterion is maximized again over the new sample set, and the kernel is refreshed accordingly. Because this criterion is differentiable, the maximization is accomplished through a gradient-based search method, incurring very little computational overhead to the real-time retrieval process. To our best knowledge, few papers have addressed incorporating prior knowledge from image retrieval to boost the retrieval performance by SVMs[1].

Compared with the existing reported ones, this approach has the following advantages: (1) It is simple and efficient. It improves the retrieval performance of SVM and needs the minimum modification. It does not involve combination with other methods or collection of more labelled data, which may increase the computation load or inconvenience to users. Furthermore, since the retrieval performance of SVM is improved, a better result can be expected when it is combined with other techniques or when more data are available; (2) This approach can be conveniently realized. As shown in the experiments later, a commonly used Gaussian RBF kernel is converted to the one incorporating the prior knowledge by simply tuning its parameter according to the proposed criterion; (3) Kernel parameter optimization of SVMs is seldom considered in practical retrieval although it is very important. This is because image retrieval requires quick response while the commonly used techniques such as leave-one-out cross-validation [3] are often time-consuming, and cannot be used to dynamically optimize the kernel parameters in real time. This approach solves this problem. By incorporating prior knowledge, it can even achieve better performance than the SVM using a kernel whose parameter is optimized by cross-validation. Moreover, this approach incurs very little computational overhead and does not significantly affect the retrieval speed; (4) This approach leads to an open framework for incorporating the retrieval prior knowledge into kernel-based classifiers. It provides better scope to boost the performance of image retrieval. (5) This approach is also a contribution to the method of KBDA. It can be shown that, by using the kernel designed with the proposed approach, KBDA can achieve its best possible retrieval performance. That is, this approach solves the kernel selection problem in KBDA. It will be seen in the experiments later that, with this designed kernel, SVM and KBDA show high retrieval performance among the compared methods. However, SVM with this kernel has the advantage in that it achieves this good performance readily rather than specially developing a new algorithm. Furthermore, SVMs have been widely used and have shown many elegant properties. These advantages are preserved by using the proposed approach. Experimental results of retrieval on two benchmark image databases demonstrate the effectiveness of the proposed approach.

2. Support Vector Machines and The Kernel

Let $\mathcal{D}$ be a training data set and $\mathcal{D} = \{(\mathbf{x}, y)\} \in (\mathbb{R}^n \times \mathcal{Y})^{|\mathcal{D}|}$, where $\mathbb{R}^n$ denotes an n-dimensional feature space, $\mathcal{Y} = \{\pm 1\}$ denotes the label set of $\mathbf{x}$, and $|\mathcal{D}|$ is the size of $\mathcal{D}$. Given $\mathcal{D}$, SVM finds an optimal separating hyperplane which classifies the two classes by the minimum expected test error. Let $\langle \mathbf{w}^*, \mathbf{x} \rangle + b^* = 0$ denote this hyperplane, where $\mathbf{w}^*$ and b^* are normal vector and bias, respectively. $\mathbf{w}^*$ and b^* can be found by minimizing

$$\begin{array}{rl} & \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{|\mathcal{D}|} \xi_i^p \\ \textit{subject to}: & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) \geq 1 - \xi_i, \;\; i = 1, \cdots, |\mathcal{D}| \end{array} \tag{1}$$

where ξ_i ($\xi_i \geq 0$) is the i-th slack variable and C is the regularization parameter controlling the trade-off between function complexity and training error. $p = 1$ or 2 corresponds to the case of L1 or L2-norm based soft margin,

[1]Our recent paper [11] has discussed this problem but the approach is not based on kernel design.

respectively. The solution can be given as

$$\begin{cases} \mathbf{w}^* &= \sum_{i=1}^{|\mathcal{D}|} \alpha_i y_i \mathbf{x}_i \\ b^* &= 1 - \langle \mathbf{w}^*, \mathbf{x}^s \rangle \end{cases} \tag{2}$$

where α_i is a non-negative coefficient of $\mathbf{x}_i$, and $\mathbf{x}^s$ denotes a support vector. In this way, the optimal separating hyperplane can be expressed as

$$\langle \mathbf{w}^*, \mathbf{x} \rangle + b^* = \sum_{i=1}^{|\mathcal{D}|} \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b^* = 0 \tag{3}$$

In the classification stage, $f(\mathbf{x}) = \langle \mathbf{w}^*, \mathbf{x} \rangle + b^*$ is used as the decision function, and a test sample is labelled as $\mathrm{sgn}\,[f(\mathbf{x})]$, where $\mathrm{sgn}(\cdot)$ denotes the sign function. The *kernel* trick can be conveniently embedded into SVMs to handle the linearly nonseparable patterns. A kernel, k, is defined to be $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$, where $\phi(\cdot)$ is the associated mapping from a feature space, $\mathbb{R}^n$, to a kernel space, $\mathcal{F}$. This mapping is often nonlinear, and the dimensionality of $\mathcal{F}$ can be of high or even infinite dimensions. The linearly nonseparable patterns in $\mathbb{R}^n$ can become linearly separable in $\mathcal{F}$ with high probability. Hence, the optimal separating hyperplane is sought in $\mathcal{F}$ instead of $\mathbb{R}^n$, and it is

$$f(\mathbf{x}) = \sum_{i=1}^{|\mathcal{D}|} \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle + b^* = \sum_{i=1}^{|\mathcal{D}|} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b^* \tag{4}$$

The commonly used kernel functions consist of Gaussian RBF kernel, polynomial kernel, and sigmoid kernel.

3. Incorporating Prior Knowledge in Kernel Design for Image Retrieval

3.1. The basic idea

Image retrieval with relevance feedback is known to suffer from the *small sample* problem. Treating it as a prior knowledge, we know that the positive images should be clustered while the negative images should be pushed away from the positive ones. We explore to incorporate this information into a retrieval algorithm to improve its performance. With no exception, this also applies to the design of an SVM classifier for image retrieval. Better is that SVMs really provides a basis to conveniently accommodate the prior knowledge of a learning task, that is, the kernel function (or kernel matrix in general). Kernel is the soul of SVMs. Based on the same training data set, very different classification results can be obtained by simply changing the employed kernel function or its parameters. As mentioned in Section 2, SVMs learn classification rules solely based on training samples and the kernel, k. It will result in poor performance if training samples are insufficient. However, the situation can be remedied by using a well-designed kernel. The proposed approach can be extended as a general framework, and other prior knowledge and domain theories can also be incorporated once suitable criteria are found to represent them. The following presents a novel way of incorporating the prior knowledge of the *small sample* in SVM-based image retrieval.

3.2. The proposed approach

Figure 1 illustrates the idea we discussed above. The knowledge of the *small sample* problem is reflected in a kernel space by requiring that the positive samples be tightly clustered while the negative samples be pushed far away from the positive ones. We design a criterion in the kernel space to measure whether this requirement is well met.

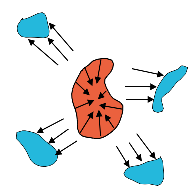


Figure 1. This figure shows how the positive samples and negative samples should be distributed in the kernel space. The center cluster represents the positive class, and the four surrounding clusters represent the negative class

Let $\mathbf{x}_i$ ($\mathbf{x}_i \in \mathbb{R}^d$) denote a d-dimensional visual feature vector of an image i (similarly, $\mathbf{x}_j$ for an image j), where $\mathbb{R}^d$ denotes the visual feature space. $k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ denotes the employed kernel function, where $\langle \cdot, \cdot \rangle$ is the dot product, $\phi(\cdot)$ represents the associated mapping, and $\boldsymbol{\theta}$ is the kernel parameter set. $\mathbf{K}$ denotes the kernel matrix and $\{\mathbf{K}\}_{i,j} = k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$. Let $\mathcal{A}$ and $\mathcal{B}$ be two data sets, and $\mathbf{K}_{\mathcal{A},\mathcal{B}}$ is the kernel matrix where $\{\mathbf{K}_{\mathcal{A},\mathcal{B}}\}_{i,j} = k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$ with the constraints of $\mathbf{x}_i \in \mathcal{A}$ and $\mathbf{x}_j \in \mathcal{B}$.

Let's define a criterion in the kernel space, $\mathcal{F}$, to design a kernel incorporating the prior knowledge. $\mathbf{S}_{np}^{\phi}$ and $\mathbf{S}_{p}^{\phi}$ are two scatter matrices in $\mathcal{F}$. $\mathbf{S}_{np}^{\phi}$ describes the scatter of negative image samples with respect to the mean of the positive ones while $\mathbf{S}_{p}^{\phi}$ describes the scatter of positive ones with respect to this mean.

$$\begin{aligned} \mathbf{S}_{np}^{\phi} &= \textstyle\sum_{\mathbf{x}_i \in \mathcal{D}_n} \left[\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}\right] \left[\phi(\mathbf{x}_i) - \mathbf{m}_p^{\phi}\right]^{\top} \\ \mathbf{S}_{p}^{\phi} &= \textstyle\sum_{\mathbf{x}_j \in \mathcal{D}_p} \left[\phi(\mathbf{x}_j) - \mathbf{m}_p^{\phi}\right] \left[\phi(\mathbf{x}_j) - \mathbf{m}_p^{\phi}\right]^{\top} \end{aligned} \tag{5}$$

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

where $\mathcal{D}_p$ and $\mathcal{D}_n$ denote the sets of the labelled positive and negative image samples, respectively. $\mathbf{m}_p^\phi$ denotes the mean vector of the labelled positive samples in $\mathcal{F}$. Since the high dimensionality of $\mathcal{F}$ often makes both $\mathbf{S}_{np}^\phi$ and $\mathbf{S}_p^\phi$ singular, their determinants cannot be used here. Hence, the proposed criterion is developed based on the traces of $\mathbf{S}_{np}^\phi$ and $\mathbf{S}_p^\phi$ instead. Let $Sum(\cdot)$ denote the summation of all elements in a matrix therein. The traces are derived as

$$
\begin{aligned}
& tr(\mathbf{S}_{np}^\phi) \\
= \; & tr\left[\sum_{\mathbf{x}_i \in \mathcal{D}_n} (\phi(\mathbf{x}_i) - \mathbf{m}_p^\phi)(\phi(\mathbf{x}_i) - \mathbf{m}_p^\phi)^\top\right] \\
= \; & \sum_{\mathbf{x}_i \in \mathcal{D}_n} \left[\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_i) - 2\phi(\mathbf{x}_i)^\top \mathbf{m}_p^\phi + {\mathbf{m}_p^\phi}^\top \mathbf{m}_p^\phi\right] \\
= \; & \sum_{\mathbf{x}_i \in \mathcal{D}_n} k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_i) - \tfrac{2}{|\mathcal{D}_p|} Sum(\mathbf{K}_{\mathcal{D}_n, \mathcal{D}_p}) \\
& + \tfrac{|\mathcal{D}_n|}{|\mathcal{D}_p|^2} Sum(\mathbf{K}_{\mathcal{D}_p, \mathcal{D}_p})
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
& tr(\mathbf{S}_p^\phi) \\
= \; & tr\left[\sum_{\mathbf{x}_i \in \mathcal{D}_p} (\phi(\mathbf{x}_i) - \mathbf{m}_p^\phi)(\phi(\mathbf{x}_i) - \mathbf{m}_p^\phi)^\top\right] \\
= \; & \sum_{\mathbf{x}_i \in \mathcal{D}_p} \left[\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_i) - 2\phi(\mathbf{x}_i)^\top \mathbf{m}_p^\phi + {\mathbf{m}_p^\phi}^\top \mathbf{m}_p^\phi\right] \\
= \; & \sum_{\mathbf{x}_i \in \mathcal{D}_p} k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_i) - \tfrac{1}{|\mathcal{D}_p|} Sum(\mathbf{K}_{\mathcal{D}_p, \mathcal{D}_p})
\end{aligned}
\tag{7}
$$

The criterion is then defined as

$$
\mathcal{J}(k, \boldsymbol{\theta}) = \frac{tr(\mathbf{S}_{np}^\phi)}{tr(\mathbf{S}_p^\phi)} \tag{8}
$$

It measures the ratio of the scatter of negative images to that of 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 in $\mathcal{F}$. The optimal kernel function, k^*, or the optimal parameter set, $\boldsymbol{\theta}^*$, for a pre-selected kernel function can be expressed as

$$
(k^*, \boldsymbol{\theta}^*) = \arg \max_{k \in \mathcal{K}, \boldsymbol{\theta} \in \boldsymbol{\Theta}} \left[\frac{tr(\mathbf{S}_{np}^\phi)}{tr(\mathbf{S}_p^\phi)}\right] \tag{9}
$$

where $\mathcal{K}$ denotes the set of possible kernel functions while $\boldsymbol{\Theta}$ denotes the parameter space for a pre-selected kernel function. The SVMs with L2-norm based soft margin is used. This helps the criterion $\mathcal{J}$ avoiding the numerical instability in maximization.

The computational load of kernel design should not be heavy because image retrieval task requires quick response. For a given kernel function, the criterion $\mathcal{J}$ has continuous first and second derivatives with respect to the kernel parameters as long as the employed kernel function has. The maximization of $\mathcal{J}$ (or the minimization of $-\mathcal{J}$) is solved by applying a nonlinear optimization technique. The BFGS Quasi-Newton method [4] is often favored because of its less number of iterations for convergence. The computational load in each iteration is largely due to evaluating $\mathcal{J}$, 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 vector. 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 number of images searched. Hence, it can be expected that the optimization procedure will not take much time, and it will not significantly slow down the response required. This can be seen from the experimental results later.

Finally, 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, for example, through the simple Euclidean search around this query. Afterwards, the labelled image samples are collected via relevance feedback. After each round of feedback, treating the labelled samples as training data, maximize the criterion $\mathcal{J}$ to carry out kernel design, and then perform retrieval by the SVM with the dynamically designed kernel. Those images having the larger positive decision values are retrieved as positive images. It is expected that this knowledge-based kernel helps SVMs achieving improved retrieval performance.

4. Experimental Results

4.1. Image databases and visual features

In this experiment, two benchmark image databases are used. One is selected from a subset of $20,000$ *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 4,800 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 intra-class variations and inter-class overlap in the visual feature space. A perceptually uniform color space, $CIE-Lab$, is used, and a feature vector of color moments is defined for each image [6]. 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 UCSB [1]. The original database includes 40 large aerial photos. Each of them is divided into 128×128 sub-images and 40 image classes are formed. In this experiment, 100 sub-images are taken from each class and a database of 4,000 images is constructed. The Gabor based texture feature described in [1] is used as the visual feature for retrieval here. The two pre-defined image databases are the ground truth for performance evaluation. The commonly used *Recall* is used to measure the performance against the rounds of feedback. *Recall* of top k for the r-th round is the percentage of true positive images retrieved so far among the total positive ones in an image database when the top k images are retrieved at the r-th round.

4.2. Experimental setting and procedure

In the experiment, the commonly used Gaussian RBF kernel is optimized to incorporate the prior knowledge of *small sample*. Gaussian RBF kernel is defined as $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right)$, where σ is the Gaussian width. The incorporation is achieved through optimizing its parameter, σ, by maximizing the proposed criterion, $\mathcal{J}$. Thus, the problem of kernel design is now simplified to the one of finding an optimal kernel parameter. The proposed approach, called SVM with the designed kernel, is compared with the following five retrieval methods: (1) SVM with LOO. There, in each round of relevance feedback, a brutal search of σ on the training data set is performed through Leave-One-Out (LOO) cross validation technique. The σ corresponding to the minimum LOO test error is selected and used in the Gaussian RBF kernel to train an SVM classifier for retrieval; (2) Kernel based Biased Discriminant Analysis (KBDA) [12] with the designed kernel; (3) SVM active learning [9] with LOO; (4) Query refinement with Euclidean search, in which the mean of labelled positive samples is taken as the query; (5) Euclidean search around the initial query only. 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 twenty retrieved images are labelled and excluded from the database for the next retrieval. Five feedbacks are performed in total. To achieve robust statistics, these methods will not be used until at least five positive and five negative image samples are labelled. In this place, the Euclidean search around the initial query is used instead. The BFGS Quasi-Newton method is used to find the σ for the designed kernel. The initial value of σ is set to $\sigma_0 = 1.0$, and the stopping 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 corresponding similarity distances; (4) According to the ground truth, label the top twenty images to simulate user feedback. After that, these labelled 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, use the proposed method to find the optimal σ w.r.t. the criterion $\mathcal{J}$. Also, run the LOO cross validation to find the σ corresponding to the minimum LOO test error; (6) Perform all the compared methods, and evaluate retrieval performance; (7) Redo steps 3 to 6 five times to simulate five user feedbacks; (8) To accumulate statistics, redo steps 2 to 7 thirty times. The obtained retrieval performance and the criterion values are averaged, respectively; (9) Redo steps 1 to 8 for each image class in the database, and the obtained retrieval performance is averaged.

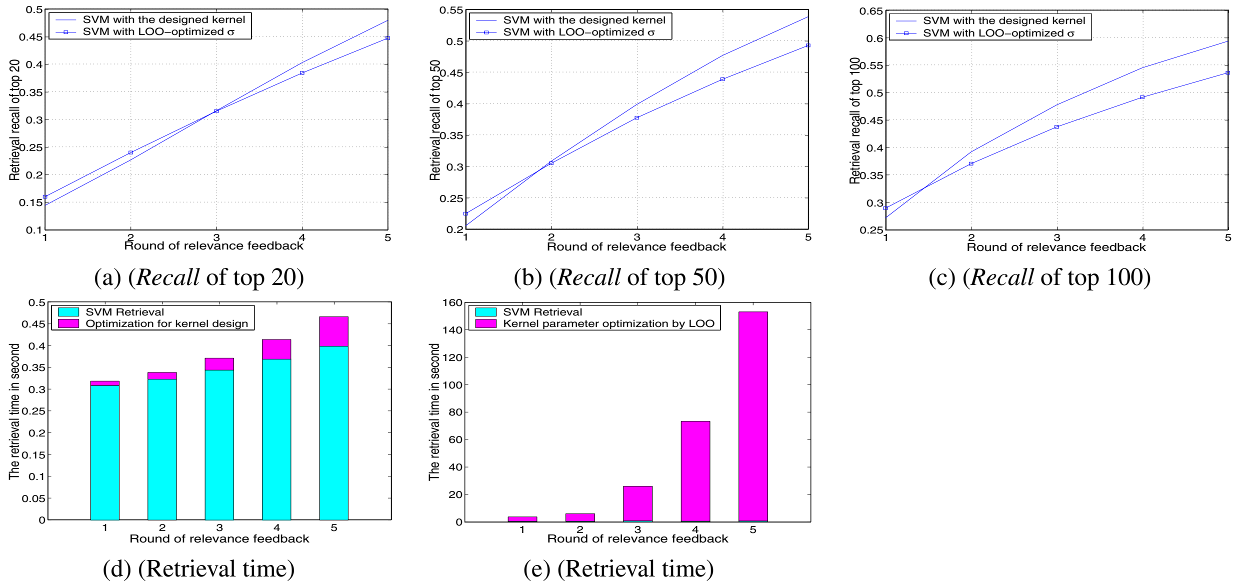
4.3. Results and discussions

Figure 2(a) - (c) show the *Recall* of top 20, 50 and 100 against the rounds of feedback, respectively. The *Corel* photo image database is used. It can be seen that SVM with the designed kernel achieves the performance comparable to that obtained by a brutal search of σ at the first 1 to 2 rounds of feedback, and increasingly outperforms the latter with more rounds of feedback. This result shows the better performance of the proposed method. As the designed kernel incorporates prior knowledge, which directs the mapping into the kernel space where the negative image samples are pushed far away from the positive ones, it gives rise to the performance better than the best one obtainable via a brutal search of σ on the training data, which does not include prior knowledge of how the two classes should be separated and treats them equally as default.

Figure 2(d) and (e) plot the retrieval time. A Linux system with Pentium Xeon 2.8GHz and 2G memory is used. Compared with the time taken by the retrieval process, the time taken by the process of kernel design is insignificant. In contrast, if a brutal search of σ is performed, the time used can be longer than the time taken by the retrieval process by hundreds of times. Such a long search time confirms the unsuitability of the brutal search approach for practical SVM-based image retrieval. Moreover, the time taken by the kernel design process is mainly affected by the number of labelled image samples rather than the size of the pool of images. Hence, even if a larger pool of images is used, it can be expected that this time still holds.

Figure 3 plots the experimental results on the aerial photo image database. As shown in the sub-figures (a) - (c), the SVM with the designed kernel achieves better performance than a brutal search against σ at more rounds of feedback. The sub-figures (d) and (e) illustrates the retrieval time. Again, it is confirmed that the time taken by the proposed method for parameter optimization is much less than that taken by the SVM retrieval process while the time taken by a brutal search based optimization is still much longer.

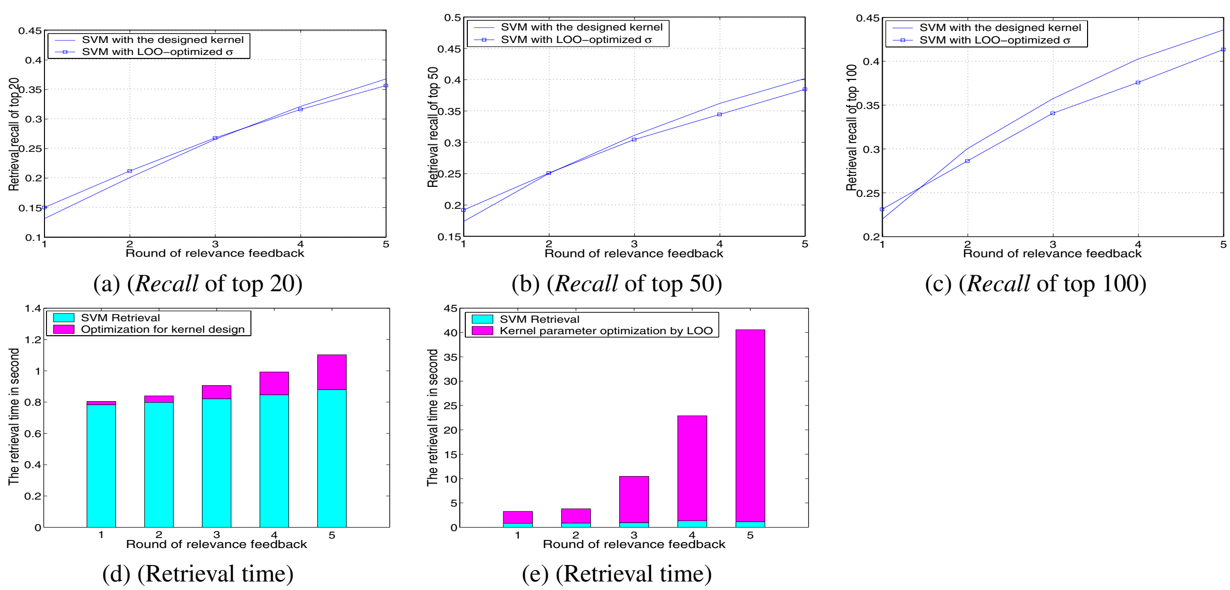
Beside these, the proposed method is also compared with another four retrieval methods, namely, KBDA, SVM active learning, Query refinement with Euclidean search, and Euclidean search only. The original KBDA method also lacks a method to choose the kernel parameters, which again can be solved by our proposed method. Thus, in the experiment, the σ optimized by our proposed method is also used in KBDA (We will show that the σ obtained through this approach can help KBDA attaining the best possible retrieval performance in Figure 6 and present more discussion on it later). All the methods take the same retrieval proce-



(a) (*Recall* of top 20) (b) (*Recall* of top 50) (c) (*Recall* of top 100)

(d) (Retrieval time) (e) (Retrieval time)

Figure 2. Retrieval performance and retrieval time (*Corel* photo image database)



(a) (*Recall* of top 20) (b) (*Recall* of top 50) (c) (*Recall* of top 100)

(d) (Retrieval time) (e) (Retrieval time)

Figure 3. Retrieval performance and the retrieval time (Aerial photo image database)

dure as described in Section 4.2. For SVM active learning, a leave-one-out cross validation for the optimal σ is also carried out to get its best possible performance for comparison. Note that σ cannot be well tuned like this in practice because this procedure is very time-consuming. Instead, it is often heuristically or empirically set. All the six retrieval methods are compared and the retrieval *Recall* of top 20, 50 and 100 against 4 rounds of feedback are plotted in Figure 4 for the *Corel* photo image database and in Figure 5 for the aerial photo image database. Here, from each class, ten images are randomly selected one by one to launch ten retrieval sessions. The retrieval performance values are av-

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

eraged over all these sessions. It can be seen that the designed kernel makes SVM and KBDA leading among all the methods for the *Corel* image database. In aerial photo image database, SVM with the designed kernel leads all the others, while the performance of KBDA with the designed kernel is slightly worse than SVM with LOO optimization. A browse through the aerial database shows very large overlapping between different classes, which may explain the overall lower retrieval performance on the aerial database and the different ordering of the performance of these methods on this database.

It is also observed that SVM active learning does not show any advantage in both databases over SVMs, at least in the top four rounds of feedback. As discussed in [9], in order to ensure effective SVM active learning, instead of performing Euclidean search, the user is asked to label twenty images randomly sampled from the database to expand the training set. In our experiment, the most common retrieval setting is used, in which Euclidean search is applied at the initial retrieval. As expected, SVM active learning cannot produce good performance.

As mentioned earlier, our method can also be used in kernel parameter optimization for KBDA. Through the kernel trick, KBDA maps the training samples from a feature space to a kernel space. Afterwards, it finds an optimal linear projection from this kernel space to a lower-dimensional subspace such that the positive images are well separated from the negative ones therein. However, it cannot find the optimal kernel parameters itself. We compare the retrieval performance using our proposed criterion with that using a brutal search against possible values of σ. As shown in Figure 6, our method achieves the performance comparable to the best one obtainable by a brutal search against σ, on both *Corel* photo image database and aerial photo image database.

In summary, the above experimental results demonstrate the effectiveness of the designed kernel and the excellent retrieval performance of the SVM with it. The procedure 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 SVM with the designed kernel achieves better or at least comparable retrieval performance, especially as the number of feedback increases.

5. Conclusion

This work aims to improve the retrieval performance of SVMs by incorporating retrieval prior knowledge into the employed kernel, and a criterion is developed to achieve this. This criterion measures the goodness of a kernel space, and the optimal kernel is obtained by maximizing this criterion. The best kernel space is judged to be one that tightly clusters the positive images and pushes away the negative images from the positive ones therein. The computational load incurred by kernel design is light and the fast response requirement in image retrieval is maintained. Experimental results on two benchmark image databases demonstrate the effectiveness of the proposed criterion in optimizing the kernel parameters. The comparison with the state-of-the-art retrieval methods shows the advantage of our proposed approach.

6. Acknowledgements

This work is partially supported by the grants LIT 2002-4 of A-STAR and RGM 14/02 of NTU, Singapore.

References

[1] *http://vision.ece.ucsb.edu/datasets/*.

[2] G. Guo, A. K. Jain, W.-Y. Ma, and H. Zhang. Learning similarity measure for natural image retrieval with relevance feedback. *IEEE Transactions on Neural Networks*, 13(4):811–820, 2002.

[3] S. Haykin. Neural Networks A Comprehensive Foundation (Second Edition). *Prentice Hall International, Inc*, 1999.

[4] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. Numerical Recipes in C: The Art of Scientific Computing. *Cambridge University Press*, 1988.

[5] B. Schölkopf, P. Simard, A. Smola, and V. Vapnik. Prior knowledge in support vector kernels. *Proceedings of the 1997 conference on Advances in neural information processing systems 10, Denver*, 1997.

[6] M. Stricker and M. Orengo. Similarity of color images. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, 2420:381–392, 1995.

[7] D. Tao and X. Tang. Random sampling based SVM for relevance feedback image retrieval. *Proceedings of IEEE International Conference on Computer Visiona and Pattern Recognition (CVPR)*, 2:647–652, 2004.

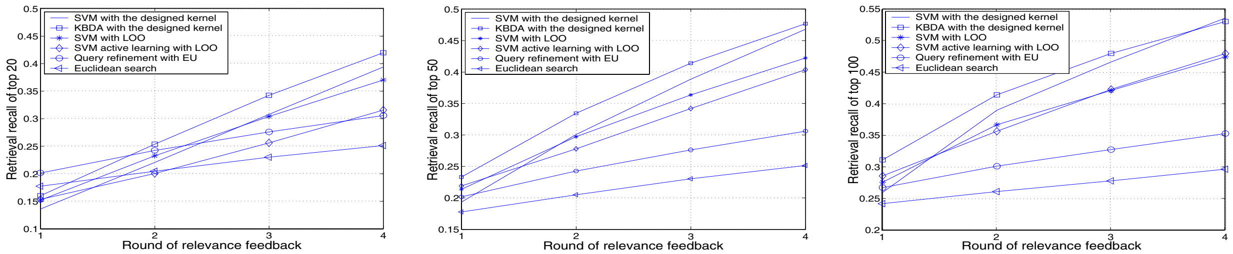
[8] Q. Tian, P. Hong, and T. S. Huang. Update Relevant Image Weights for Content-Based Image Retrieval using Support Vector Machines. *IEEE International Conference on Multimedia and Expo (II)*, pages 1199–1202, 2000.

[9] S. Tong and E. Chang. Support Vector Machine Active Learning for Image Retrieval. *Proceedings of ACM International Conference on Multimedia*, pages 107–118, Ottawa, Canada, Oct, 2001.

[10] L. Wang, K. L. Chan, and Z. Zhang. Bootstrapping SVM Active Learning by Incorporating Unlabelled Images for Image Retrieval. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

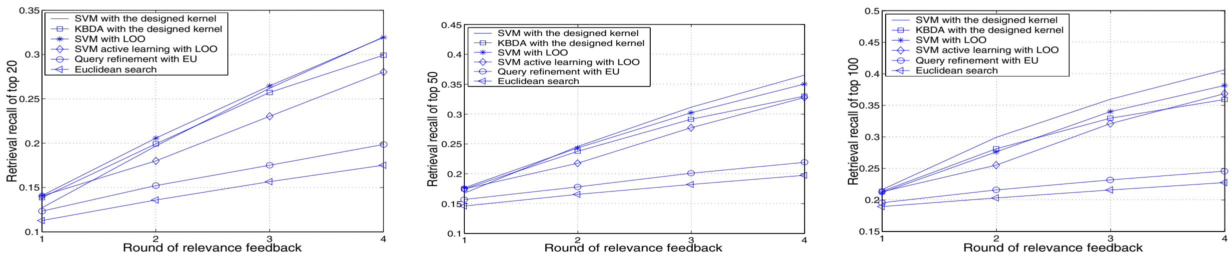
[11] L. Wang, P. Xue, and K. L. Chan. Incorporating Prior Knowledge into SVM for Image Retrieval. *Proceedings of International Conference on Pattern Recognition (ICPR), UK*, 2004.

[12] X. S. Zhou and T. S. Huang. Small Sample Learning during Multimedia Retrieval using BiasMap. *Proceedings of IEEE International Conference on Computer Visiona and Pattern Recognition (CVPR), Hawaii*, Dec, 2001.



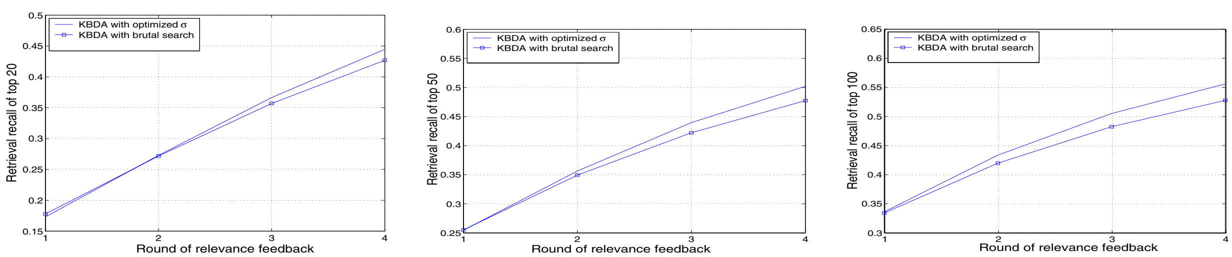
(a) (*Recall* of top 20) (b) (*Recall* of top 50) (c) (*Recall* of top 100)

Figure 4. Comparison of six retrieval methods (*Corel* photo image database)

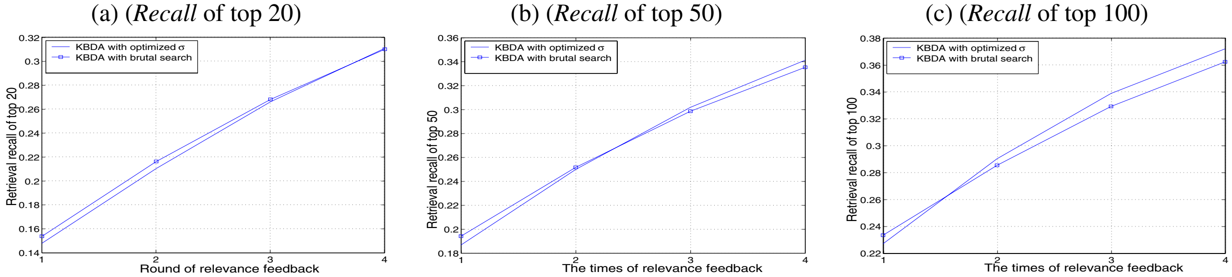


(a) (*Recall* of top 20) (b) (*Recall* of top 50) (c) (*Recall* of top 100)

Figure 5. Comparison of six retrieval methods (Aerial photo image database)



(a) (*Recall* of top 20) (b) (*Recall* of top 50) (c) (*Recall* of top 100)



(d) (*Recall* of top 20) (e) (*Recall* of top 50) (f) (*Recall* of top 100)

Figure 6. Retrieval performance by KBDA using the designed kernel (1st row: *Corel* image database; 2nd row: Aerial photo image database)