Convex Experimental Design Using Manifold Structure for Image Retrieval

Lijun Zhang[1]
zljzju@zju.edu.cn

Chun Chen[1]
chenc@zju.edu.cn

Wei Chen[1]
chenw@zju.edu.cn

Jiajun Bu[1]
bjj@zju.edu.cn

Deng Cai[2]
dengcai@cad.zju.edu.cn

Xiaofei He[2]
xiaofeihe@cad.zju.edu.cn

[1]Zhejiang Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, Hangzhou, China
[2]State Key Lab of CAD&CG, College of Computer Science, Zhejiang University, Hangzhou, China

ABSTRACT

Content Based Image Retrieval (CBIR) has become one of the most active research areas in computer science. Relevance feedback is often used in CBIR systems to bridge the semantic gap. Typically, users are asked to make relevance judgements on some query results, and the feedback information is then used to re-rank the images in the database. An effective relevance feedback algorithm must provide the users with the most informative images with respect to the ranking function. In this paper, we propose a novel active learning algorithm, called Convex Laplacian Regularized I-optimal Design (CLapRID), for relevance feedback image retrieval. Our algorithm is based on a regression model which minimizes the least square error on the labeled images and simultaneously preserves the intrinsic geometrical structure of the image space. It selects the most informative images which minimize the average predictive variance. The optimization problem of CLapRID can be cast as a semidefinite programming (SDP) problem, and solved via interior-point methods. Experimental results on COREL database have demonstrate the effectiveness of the proposed algorithm for relevance feedback image retrieval.

Categories and Subject Descriptors

H.3.3 [**Information storage and retrieval**]: Information search and retrieval—*Relevance feedback*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Experimental design*

General Terms

Algorithms, Performance, Theory

Keywords

Image retrieval, active learning, convex optimization, relevance feedback, semidefinite programming

MM'09, October 19–24, 2009, Beijing, China.

1. INTRODUCTION

With the rapid increase in the volume of electronically archived image and video materials, Content Based Image Retrieval (CBIR) has become one of the most active research areas for the last few decades [10, 17]. Query by example (QBE) is the traditional type of query in CBIR. In this environment, users formulate a query by means of giving an example image [21]. CBIR systems use the low level visual features (mostly color, texture and shape) to represent an image's content, and relevant images are retrieved based on the similarity of their visual features. Although CBIR has been extensively studied, the semantic gap between low-level image features and high-level semantic concepts limits its performance largely.

To narrow down the semantic gap, relevance feedback is introduced into CBIR [16]. Typically, users are asked to make relevance judgements on the top images returned by the system, and their preference is used to train a classifier to separate images that match the query concept from those that do not. However, in general the top returned images may not be the most informative ones. In the worst case, all the top images labeled by the user may be positive and thus the standard classification techniques can not be applied due to the lack of negative examples. The key problem then becomes how to select the most informative samples from the image database. In machine learning, this problem is called active learning, which studies the phenomenon of a learner selecting actions or making queries that influence what data are added to its training set [9].

Active learning algorithm is highly correlated with the underlying ranking mechanism. The most popular active learning techniques include Support Vector Machine active learning (SVM_{active}) [19,20] and regression based active learning [2,14,23,24]. SVM_{active} asks the user to label those images which are closest to the SVM boundary. The rationale is that the closer to the SVM boundary an image is, the less reliable its classification is. One of the major disadvantages of SVM_{active} is that the estimated boundary may not be accurate enough, especially when the number of labeled image is small. Moreover, SVM_{active} can not be applied at the first round of the retrieval when there is no feedback images.

In statistics, the problem of selecting samples to label is usually referred to as experimental design. The study of Optimum Experimental Design (OED) [2] is concerned with the design of experiments that are expected to minimize variance of a parameterized model. There are two types of selection criteria of OED. One type is to choose points to minimize the confidence region for the estimated model parameters, which results in D-, A-, and E-optimal Design. The other is to minimize the variance of the pre-

diction value, which results in I- and G-optimal Design. Recently, Yu et al. propose an active learning formulation, called Transductive Experimental Design (TED) [23]. TED selects data points to minimize the average predictive variance of the learned function on some pre-given dataset. It is also formulated into a convex optimization problem (CovTED) [24].

Both SVM$_{active}$ and OED are based on supervised learning algorithms (SVM or linear regression). These algorithms only consider the labeled data while ignoring those unlabel data. In relevance feedback image retrieval, the user is usually not willing to provide too many feedbacks. Thus, we have a large unlabeled image database and a small number of labeled images (feedbacks). In this case, supervised learning algorithms may not be able to get good performance due to the lack of training data. The research in semi-supervised learning [7] shows that unlabeled data, when used in conjunction with a small amount of labeled data, can improve the learning performance greatly. For example, the intrinsic manifold structure in the data space has been shown very useful for improving the learning performance [3, 25].

In this paper, we propose a novel active learning algorithm, called Convex Laplacian Regularized I-optimal Design (CLapRID), for relevance feedback image retrieval. CLapRID is essentially based on a graph-based semi-supervised learning algorithm— Laplacian Regularized Least Squares (LapRLS) [3]. It makes use of both labeled and unlabeled points to discover the intrinsic geometrical structure in the data. By constructing a nearest neighbor graph, the geometrical structure of the image space can be described by the graph Laplacian. CLapRID then selects those images which can minimize the average predictive variance once labeled. The optimization problem of CLapRID can be cast as a semidefinite programming (SDP), and solved exactly via interior-point methods.

The rest of the paper is organized as follows. In Section 2, we provide a brief description of the related work. Our proposed Convex Laplacian Regularized I-optimal Design (CLapRID) is introduced in Section 3. In Section 4, we describe how to perform CLapRID in Reproducing Kernel Hilbert Space (RKHS) which leads to nonlinear extension. In Section 5, we compare our algorithm with the state-of-the-art active learning algorithms and present the experimental results on image retrieval. Finally, we provide some concluding remarks and suggestions for future work in Section 6.

2. RELATED WORK

The generic problem of active learning is the following. Given a set of points $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$ in $\mathbb{R}^n$, find a subset $\mathcal{Z} = \{\mathbf{z}_1, \cdots, \mathbf{z}_k\} \subseteq \mathcal{X}$ which contains the most informative points. That is, if points $\mathbf{z}_i (i = 1, \cdots, k)$ are labeled and used as training points, we can predict the labels of the unlabeled points most precisely. Active learning is usually referred to as experimental design in statistics. Since our approach is motivated by recent progress in experimental design, we begin with a brief description of it. For details, please see [2, 14, 23, 24].

Experimental design [2, 4] considers learning a linear function $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$, from experiments $y_i = \mathbf{w}^T\mathbf{z}_i + \epsilon_i,\ i = 1, \cdots, k$ where ϵ_i are independent Gaussian random variables with zero mean and constant variance σ^2. The most popular estimation method is *least squares*, in which we minimize the residual sum of squares:

$$RSS(\mathbf{w}) = \sum_i^k (y_i - f(\mathbf{z}_i))^2 \tag{1}$$

Let $Z = [z_1, \cdots, z_k]$ and $\mathbf{y} = [y_1, \cdots, y_k]^T$. The optimal solution is

$$\widehat{\mathbf{w}} = (ZZ^T)^{-1}Z\mathbf{y} \tag{2}$$

It can be proved [13] that $\widehat{\mathbf{w}}$ is a unbiased estimation of $\mathbf{w}$ with covariance matrix

$$\mathrm{Cov}(\widehat{\mathbf{w}}) = \sigma^2(ZZ^T)^{-1} \tag{3}$$

The criteria of Optimum Experimental Design (OED) [2] can be classified into two types. The first type is to choose points $\mathbf{z}_i$'s to minimize the confidence region for the estimated parameter $\widehat{\mathbf{w}}$ in some sense [1, 2]. The three most common criteria are:

- D-optimal design: minimizes the determinant of $\mathrm{Cov}(\widehat{\mathbf{w}})$, and thus minimizes the volume of the confidence region.
- A-optimal design: minimizes the trace of $\mathrm{Cov}(\widehat{\mathbf{w}})$, and thus minimizes the dimensions of the enclosing box around the confidence region.
- E-optimal design: minimizes the maximum eigenvalue of $\mathrm{Cov}(\widehat{\mathbf{w}})$, and thus minimizes the size of the major axis of the confidence region.

The other type is to choose points $\mathbf{z}_i$'s to minimize the variance of prediction value over some region of interest $\mathcal{O}$. Given a test point $\mathbf{v} \in \mathcal{O}$, the prediction value is $\widehat{\mathbf{w}}^T\mathbf{v}$ with variance $\mathbf{v}^T\,\mathrm{Cov}(\widehat{\mathbf{w}})\mathbf{v}$. The two most common criteria are:

- I-optimal design: minimizes the average predictive variance: $\int_{\mathbf{v}\in\mathcal{O}} \mathbf{v}^T\,\mathrm{Cov}(\widehat{\mathbf{w}})\mathbf{v}\, d\mu(\mathbf{v})$, where μ is a probability distribution on $\mathcal{O}$.
- G-optimal design: minimizes the maximum predictive variance: $\max_{\mathbf{v}\in\mathcal{O}}\{\mathbf{v}^T\,\mathrm{Cov}(\widehat{\mathbf{w}})\mathbf{v}\}$.

Recently, Yu et. al [23] proposed Transductive Experimental Design (TED) approach which is based on I-optimal design. TED considers the regularized least squares formulation (ridge regression) as follows:

$$\widehat{\mathbf{w}}_{ridge} = \operatorname*{argmin}_{\mathbf{w}} \sum_i^k (y_i - f(\mathbf{z}_i))^2 + \gamma\|\mathbf{w}\|^2 \tag{4}$$

where $\gamma \geq 0$ is the trade-off parameter for the regularizer $\|\mathbf{w}\|^2$, and $\|\cdot\|$ is the vector ℓ_2-norm. The optimal solution is given by

$$\widehat{\mathbf{w}}_{ridge} = (ZZ^T + \gamma I)^{-1}Z\mathbf{y} \tag{5}$$

where I is the identity matrix. The covariance matrix of $\widehat{\mathbf{w}}_{ridge}$ is

$$\begin{aligned}
&\mathrm{Cov}(\widehat{\mathbf{w}}_{ridge})\\
=&(ZZ^T + \gamma I)^{-1}Z\,\mathrm{Cov}(\mathbf{y})Z^T(ZZ^T + \gamma I)^{-1}\\
=&\sigma^2(ZZ^T + \gamma I)^{-1}ZZ^T(ZZ^T + \gamma I)^{-1}\\
=&\sigma^2(ZZ^T + \gamma I)^{-1}(ZZ^T + \gamma I - \gamma I)(ZZ^T + \gamma I)^{-1}\\
=&\sigma^2(ZZ^T + \gamma I)^{-1} - \gamma(ZZ^T + \gamma I)^{-1}(ZZ^T + \gamma I)^{-1}
\end{aligned} \tag{6}$$

Since the regularization parameter γ is usually set to be very small, we have

$$\mathrm{Cov}(\widehat{\mathbf{w}}_{ridge}) \approx \sigma^2(ZZ^T + \gamma I)^{-1} \tag{7}$$

Similar to I-optimal design, TED also selects those samples which can minimize the average predictive variance. Let $X = [\mathbf{x}_1, \cdots, \mathbf{x}_m]$, the average predictive variance is

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^m \mathbf{x}_i^T\,\mathrm{Cov}(\widehat{\mathbf{w}}_{ridge})\mathbf{x}_i \approx& \frac{\sigma^2}{m}\sum_{i=1}^m \mathbf{x}_i^T(ZZ^T + \gamma I)^{-1}\mathbf{x}_i\\
=&\frac{\sigma^2}{m}\,\mathrm{Tr}(X^T(ZZ^T + \gamma I)^{-1}X)
\end{aligned} \tag{8}$$

TED is then formulated as the following optimization problem:

$$\begin{array}{ll}\min & \mathrm{Tr}(X^T(ZZ^T+\gamma I)^{-1}X)\\ \text{s.t.} & \{\mathbf{z}_1,\cdots,\mathbf{z}_k\}\subseteq\mathcal{X}\end{array} \tag{9}$$

with variable $Z=[\mathbf{z}_1,\cdots,\mathbf{z}_k]$. After some mathematical derivation, the above problem can be formulated as an equivalent optimization problem as follows:

$$\begin{array}{ll}\min & \sum_{i=1}^m \|\mathbf{x}_i - Z\boldsymbol{\alpha}_i\|^2+\gamma\|\boldsymbol{\alpha}_i\|^2\\ \text{s.t.} & \{\mathbf{z}_1,\cdots,\mathbf{z}_k\}\subseteq\mathcal{X}\end{array} \tag{10}$$

where the variables are $Z=[\mathbf{z}_1,\cdots,\mathbf{z}_k]$ and $\boldsymbol{\alpha}_i\in\mathbb{R}^k, i=1,\cdots,m$.

The above problem is NP-hard. Yu et al. have proposed a sequential greedy algorithm [23] and a convex relaxation [24] to solve it. The convex relaxation (CovTED) is shown as follows:

$$\begin{array}{ll}\min & \sum\limits_{i=1}^m\left(\|\mathbf{x}_i - X\boldsymbol{\alpha}_i\|^2+\sum\limits_{j=1}^m\frac{\alpha_{i,j}^2}{\beta_j}\right)+\gamma\|\boldsymbol{\beta}\|_1\\ \text{s.t.} & \beta_j\geq 0, j=1,\cdots,m\end{array} \tag{11}$$

where the variables are $\boldsymbol{\beta}\in\mathbb{R}^m$ and $\boldsymbol{\alpha}_i\in\mathbb{R}^m, i=1,\cdots,m$. Here, $\|\boldsymbol{\beta}\|_1$ is the ℓ_1-norm of $\boldsymbol{\beta}$, which is used to enforce some elements of $\boldsymbol{\beta}$ to be zero. An iterative algorithm is proposed to solve it [24].

3. CONVEX LAPLACIAN REGULARIZED I-OPTIMAL DESIGN

Traditional active learning algorithms, such as SVM$_{active}$ and OED, are based on supervised learning algorithms (SVM or linear regression). These approaches only consider the labeled data points while neglecting the large amount of unlabel data points which may play essential rules in selecting informative samples. We introduce in this section a novel active learning algorithm which is based on one semi-supervised learning algorithm. We will first introduce the linear algorithm and then generalize it to the nonlinear case by applying kernel tricks. Our algorithm is fundamentally based on Laplacian Regularized Least Squares (LapRLS) [3], and motivated by recent progress in experimental design [2, 14, 23, 24].

3.1 Laplacian Regularized Least Squares

Laplacian Regularized Least Squares (LapRLS) [3] makes use of both labeled and unlabeled points to discover the intrinsic geometrical structure in the data. It assumes that if two points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close then their measurements $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ are close as well. Specifically, LapRLS adds a new locality preserving regularizer into the loss function of ridge regression (Eq. 4). Let W be a similarity matrix, the new loss function is defined as follows:

$$J_L(\mathbf{w})=\sum_i^k(f(\mathbf{z}_i)-y_i)^2+\frac{\alpha}{2}\sum_{i,j=1}^m(f(\mathbf{x}_i)-f(\mathbf{x}_j))^2W_{ij}+\beta\|\mathbf{w}\|^2 \tag{12}$$

where $\alpha\geq 0$ and $\beta\geq 0$ are the regularization parameters. The second term of the right-hand side in the cost function is the locality preserving regularizer, which incurs a heavy penalty if neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ are mapped far apart.

There are many choices of similarity matrix W. A simple definition is as follows:

$$W_{ij}=\begin{cases}1, & \text{if } \mathbf{x}_i \text{ is among the } p \text{ nearest neighbors of } \mathbf{x}_j,\\ & \text{or } \mathbf{x}_j \text{ is among the } p \text{ nearest neighbors of } \mathbf{x}_i;\\ 0, & \text{otherwise.}\end{cases} \tag{13}$$

Let D be a diagonal matrix with $D_{ii}=\sum_{j=1}^m W_{ij}$, and $L=D-W$. The matrix L is called *Graph Laplacian* in spectral graph theory [8]. The solution to minimize equation (12) is given as follows:

$$\widehat{\mathbf{w}}_L=(ZZ^T+\alpha XLX^T+\beta I)^{-1}Z\mathbf{y} \tag{14}$$

Let $H=ZZ^T+\alpha XLX^T+\beta I$, the covariance matrix of $\widehat{\mathbf{w}}_L$ is

$$\begin{aligned}\mathrm{Cov}(\widehat{\mathbf{w}}_L)=&H^{-1}Z\,\mathrm{Cov}(\mathbf{y})Z^TH^{-1}\\ =&\sigma^2H^{-1}ZZ^TH^{-1}\\ =&\sigma^2H^{-1}(H-\alpha XLX^T+\beta I)H^{-1}\\ =&\sigma^2H^{-1}-\sigma^2H^{-1}(\alpha XLX^T+\beta I)H^{-1}\end{aligned} \tag{15}$$

Since the regularization parameters (α and β) are usually set to be very small, we have

$$\mathrm{Cov}(\widehat{\mathbf{w}}_L)\approx\sigma^2H^{-1}=\sigma^2(ZZ^T+\alpha XLX^T+\beta I)^{-1} \tag{16}$$

3.2 Convex Laplacian Regularized I-optimal Design

Through making use of both labeled and unlabeled data, LapRLS estimates a linear fitting function $f(\mathbf{x})=\widehat{\mathbf{w}}_L^T\mathbf{x}$ that respects the intrinsic geometrical structure in the data space. An ideal design would choose a subset $\mathcal{Z}\subseteq\mathcal{X}$ which simultaneously minimizes the confidence region for $\widehat{\mathbf{w}}_L^T$ and the predictive variance of $f(\mathbf{x})$. However, usually a choice has to be made between these desiderata [2]. In image retrieval, we aim at learning a regression function which can distinguish the relevant images from irrelevant ones. It is natural to require that the predictions of the learned function on the image database are as stable as possible. Thus, we use the I-optimal design criterion to select those images which can minimize the average predictive variance of learned regression function.

Here, we consider a set $\mathcal{V}=\{\mathbf{v}_1,\cdots,\mathbf{v}_l\}$ of test data points besides candidates in $\mathcal{X}=\{\mathbf{x}_1,\cdots,\mathbf{x}_m\}$. In special cases, $\mathcal{V}$ and $\mathcal{X}$ can be the same set. Given a test point $\mathbf{v}$, its prediction value is $f(\mathbf{v})=\widehat{\mathbf{w}}_L^T\mathbf{v}$ with variance $\mathrm{Var}(f(\mathbf{v}))=\mathbf{v}^T\,\mathrm{Cov}(\widehat{\mathbf{w}}_L)\mathbf{v}$. Let $V=[\mathbf{v}_1,\cdots,\mathbf{v}_l]$, the average predictive variance on $\mathcal{V}$ is

$$\begin{aligned}&\frac{1}{l}\sum_{i=1}^l\mathbf{v}_i^T\,\mathrm{Cov}(\widehat{\mathbf{w}}_L)\mathbf{v}_i\\ \approx&\frac{\sigma^2}{l}\sum_{i=1}^l\mathbf{v}_i^T(ZZ^T+\alpha XLX^T+\beta I)^{-1}\mathbf{v}_i\\ =&\frac{\sigma^2}{l}\,\mathrm{Tr}(V^T(ZZ^T+\alpha XLX^T+\beta I)^{-1}V)\end{aligned} \tag{17}$$

Then, our problem is to find a subset $\mathcal{Z}\subseteq\mathcal{X}$ to minimize equation (17). A simple sequential greedy approach was suggested to select $\mathbf{z}_i$'s one after another in [14].

By introducing m indicator variables $\{\lambda_i\}_{i=1}^m\in\{0,1\}$ where λ_i indicates whether or not point $\mathbf{x}_i$ is chosen, finding a subset $\mathcal{Z}$ to minimize equation (17) is equivalent to the following optimization problem:

$$\begin{array}{ll}\min & \mathrm{Tr}(V^T(\sum_{i=1}^m\lambda_i\mathbf{x}_i\mathbf{x}_i^T+\alpha XLX^T+\beta I)^{-1}V)\\ \text{s.t.} & \{\lambda_i\}_{i=1}^m\in\{0,1\},\ \sum_{i=1}^m\lambda_i=k\end{array} \tag{18}$$

where the variables are $\{\lambda_i\}_{i=1}^m$ and k is the number of data points to be chosen. To simplify our presentation, we use vector $\boldsymbol{\lambda}=[\lambda_1,\cdots,\lambda_m]$ to denote all the m variables. The variable vector $\boldsymbol{\lambda}$ is sparse and has only k non-zero entries.

Due to its combinatorial nature, the above optimization is NP-hard. In order to solve the above optimization problem efficiently, we relax the integer constraints on λ_i's and allow them to take real

nonnegative values. Then, the value of λ_i indicates how significantly $\mathbf{x}_i$ contributes to the minimization in problem (18). The sparseness of $\boldsymbol{\lambda}$ can be controlled through minimizing the ℓ_1-norm of $\boldsymbol{\lambda}$, which is a very popular technique in regression [4, 13].

Following the convention in the field of optimization, we use $\boldsymbol{\lambda} \succeq 0$ to denote that all the elements in $\boldsymbol{\lambda}$ should be nonnegative. And because all the elements of $\boldsymbol{\lambda}$ are nonnegative, $\|\boldsymbol{\lambda}\|_1$ is equal to $\mathbf{1}^T\boldsymbol{\lambda}$, where $\mathbf{1}$ is a column vector containing all ones. Finally, our optimization problem becomes:

Definition 1. Convex Laplacian Regularized I-optimal Design (CLapRID):

$$\begin{array}{ll} \min & \mathrm{Tr}(V^T(\sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I)^{-1}V) + \gamma \mathbf{1}^T\boldsymbol{\lambda} \\ \text{s.t.} & \boldsymbol{\lambda} \succeq 0 \end{array} \tag{19}$$

where the variable is $\boldsymbol{\lambda} \in \mathbb{R}^m$, and $\gamma \geq 0$ is the trade-off parameter for sparsity.

THEOREM 1. *Problem (19) is a convex optimization problem with variable* $\boldsymbol{\lambda} \in \mathbb{R}^m$.

PROOF. Let $g(X) = \mathrm{Tr}(V^TX^{-1}V) = \sum_{j=1}^{l} \mathbf{v}_j^T X^{-1}\mathbf{v}_j$ and $h(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I$. We know that matrix fractional function $f_1(X) = \mathbf{v}^TX^{-1}\mathbf{v}$ is a convex function of X [4]. Since nonnegative weighted sum preserves convexity, $g(X)$ is also a convex function of X. We define

$$g \circ h(\boldsymbol{\lambda}) = \mathrm{Tr}(V^T(\sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I)^{-1}V)$$

Because $h(\boldsymbol{\lambda})$ is an affine function of $\boldsymbol{\lambda}$ and composition with an affine function preserves convexity, $g \circ h$ is a convex function of $\boldsymbol{\lambda}$.

Since $f_2(\boldsymbol{\lambda}) = \gamma \mathbf{1}^T\boldsymbol{\lambda}$ is a convex function of λ, the objective function of problem (19) ($g \circ h(\boldsymbol{\lambda})$+ $f_2(\boldsymbol{\lambda})$) is also convex.

Because the objective function is convex, the inequality constraint function $(-\boldsymbol{\lambda})$ is convex, problem (19) is a convex optimization problem with variable $\boldsymbol{\lambda} \in \mathbb{R}^m$ [4]. □

3.3 Optimization Scheme

The success of Semidefinite programming (SDP) in various applications motivates us to formulate and solve CLapRID as an SDP problem. Semidefinite programming has been the most exciting mathematical development in mathematical programming. It has applications in traditional convex constrained optimization, as well as in such diverse domains as control theory and combinatorial optimization [12]. Moreover, the powerful interior-point methods for linear programming have been extended to SDP [11].

By introducing a new variable $P \in \mathbb{R}^{l\times l}$, optimization problem (19) can be equivalently rewrote as:

$$\begin{array}{ll} \min & \mathrm{Tr}(P) + \gamma \mathbf{1}^T\boldsymbol{\lambda} \\ \text{s.t.} & P \succeq_{\mathbb{S}_l^+} V^T(\sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I)^{-1}V \\ & \boldsymbol{\lambda} \succeq 0 \end{array} \tag{20}$$

with variables $P \in \mathbb{R}^{l\times l}$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$. Here, $\mathbb{S}_l^+$ denotes the set of symmetric positive semidefinite $l \times l$ matrices, which is called positive semidefinite cone in the field of optimization. The associated generalized inequality $\succeq_{\mathbb{S}_l^+}$ is the usual matrix inequality: $A \succeq_{\mathbb{S}_l^+} B$ means $A - B$ is a positive semidefinite $l \times l$ matrix [4].

THEOREM 2. *Problem (19) is equivalent to problem (20).*

PROOF. Let $\boldsymbol{\lambda}_a^*$ be the optimal solution of problem (19), and (P^*, $\boldsymbol{\lambda}_b^*$) be the optimal solutions of problem (20). Then, $\boldsymbol{\lambda}_a^*$= $\boldsymbol{\lambda}_b^*$ is a sufficient condition for Theorem 2. Let $f(\boldsymbol{\lambda}) = T^T(\sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I)^{-1}T$.

Assume $\boldsymbol{\lambda}_a^* \neq \boldsymbol{\lambda}_b^*$. Since $\boldsymbol{\lambda}_a^*$ minimizes problem (19), we must have $\mathrm{Tr}\, f(\boldsymbol{\lambda}_a^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_a^* < \mathrm{Tr}\, f(\boldsymbol{\lambda}_b^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_b^*$. Because ($P^*$, $\boldsymbol{\lambda}_b^*$) satisfies the constraints in problem (20), we have

$$\begin{aligned} P^* \succeq_{\mathbb{S}_l^+} f(\boldsymbol{\lambda}_b^*) &\Leftrightarrow P^* - f(\boldsymbol{\lambda}_b^*) \in \mathbb{S}_l^+ \\ &\Rightarrow \mathrm{Tr}(P^* - f(\boldsymbol{\lambda}_b^*)) \geq 0 \\ &\Rightarrow \mathrm{Tr}(P^*) \geq \mathrm{Tr}\, f(\boldsymbol{\lambda}_b^*) \\ &\Rightarrow \mathrm{Tr}(P^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_b^* \geq \mathrm{Tr}\, f(\boldsymbol{\lambda}_b^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_b^* \\ &\Rightarrow \mathrm{Tr}(P^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_b^* > \mathrm{Tr}\, f(\boldsymbol{\lambda}_a^*) + \gamma \mathbf{1}^T\boldsymbol{\lambda}_a^* \end{aligned}$$

It is clear that $(f(\boldsymbol{\lambda}_a^*), \boldsymbol{\lambda}_a^*)$ satisfies the constraints in problem (20). Then, for problem (20), $(f(\boldsymbol{\lambda}_a^*), \boldsymbol{\lambda}_a^*)$ is more optimal than (P^*, $\boldsymbol{\lambda}_b^*$), which contradicts our assumptions. So, we mush have $\boldsymbol{\lambda}_a^* = \boldsymbol{\lambda}_b^*$. □

Problem (20) can be cast as an SDP using the Schur complement theorem [4]. Given a symmetric matrix X partitioned as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

If A is invertible, the matrix $S = C - B^TA^{-1}B$ is called the Schur complement of A in X. Schur complement theorem states that, if A is positive definite, then X is positive semidefinite if and only if S is positive semidefinite.

According to this theorem, problem (20) is equivalent to the following semidefinite programming (SDP):

$$\begin{array}{ll} \min & \mathrm{Tr}(P) + \gamma \mathbf{1}^T\boldsymbol{\lambda} \\ \text{s.t.} & \begin{bmatrix} \sum_{i=1}^{m} \lambda_i \mathbf{x}_i\mathbf{x}_i^T + \alpha XLX^T + \beta I & V \\ V^T & P \end{bmatrix} \succeq_{\mathbb{S}_{n+l}^+} 0 \\ & \boldsymbol{\lambda} \succeq 0 \end{array} \tag{21}$$

with variables $P \in \mathbb{R}^{l\times l}$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$. As explained previously, $A \succeq_{\mathbb{S}_{n+l}^+} 0$ means A is a positive semidefinite $(n+l)\times(n+l)$ matrix. We can solve this problem exactly via interior-point methods [4]. After obtaining the optimal solution $\boldsymbol{\lambda}^*$, we select k points with the largest significant indicators (λ_i^*'s) for user to label.

4. CONVEX KERNEL LAPLACIAN REGULARIZED I-OPTIMAL DESIGN

Traditional experimental design only considers linear functions. When the data is highly nonlinear distributed, the linear function might not be able to fit the data well. In this Section, we extend CLapRID to handle nonlinear cases by performing experimental design in the Reproducing Kernel Hilbert Space (RKHS) [3]. We begin with a brief description of Kernel Laplacian Regularized Least Squares [3].

4.1 Kernel Laplacian Regularized Least Squares

Let K be a positive definite mercer kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, and $\mathcal{H}_K$ be the corresponding Reproducing Kernel Hilbert Space (RKSH). Consider the optimization problem (12) in RKHS. Then, we seek a function $f \in \mathcal{H}_K$ such that the following objective function is minimized:

$$J_L(f) = \sum_{i=1}^{k} (y_i - f(\mathbf{z}_i))^2 + \frac{\alpha}{2} \sum_{i,j=1}^{m} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 W_{ij} + \beta \|f\|_{\mathcal{H}_K}^2 \tag{22}$$

The Representer Theorem [3] can be used to show that the solution is an expansion of kernel functions over both the labeled and the unlabeled data:

$$\widehat{f}(\mathbf{x}) = \sum_{i=1}^{m} \widehat{\alpha}_i K(\mathbf{x}, \mathbf{x}_i) \tag{23}$$

Let $\widehat{\boldsymbol{\alpha}} = [\widehat{\alpha}_1, \cdots, \widehat{\alpha}_m]^T$. The optimal solution is given by

$$\widehat{\boldsymbol{\alpha}} = (K_{XZ}K_{ZX} + \alpha K_{XX}LK_{XX} + \beta K_{XX})^{-1}K_{XZ}\mathbf{y} \tag{24}$$

with covariance

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) \approx \sigma^2(K_{XZ}K_{ZX} + \alpha K_{XX}LK_{XX} + \beta K_{XX})^{-1} \tag{25}$$

where K_{XZ} is a $m \times k$ matrix with $K_{XZ,ij} = K(\mathbf{x}_i, \mathbf{z}_j)$, K_{XX} is a $m \times m$ matrix with $K_{XX,ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, and $K_{ZX} = K_{XZ}^T$.

4.2 Convex Kernel Laplacian Regularized I-optimal Design

The estimated function of Kernel Laplacian Regularized Least Squares is $\widehat{f}(\mathbf{x}) = \sum_{i=1}^{m} \widehat{\alpha}_i K(\mathbf{x}, \mathbf{x}_i)$. Given a test point $\mathbf{v}$, its prediction value is $\widehat{\boldsymbol{\alpha}}^T K_{\mathbf{v}}$ with variance $K_{\mathbf{v}}^T \mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) K_{\mathbf{v}}$, where $K_{\mathbf{v}} = [K(\mathbf{x}_1, \mathbf{v}), \cdots, K(\mathbf{x}_m, \mathbf{v})]^T$. Let $M = \alpha K_{XX}LK_{XX} + \beta K_{XX}$, the average predictive variance of $\widehat{f}(\mathbf{x})$ on test set $\mathcal{V} = \{\mathbf{v}_1, \cdots, \mathbf{v}_l\}$ is

$$\begin{aligned} &= \frac{1}{l}\sum_{i=1}^{l} K_{\mathbf{v}_i}^T \mathrm{Cov}(\widehat{\boldsymbol{\alpha}}) K_{\mathbf{v}_i} \\ &\approx \frac{\sigma^2}{l}\sum_{i=1}^{l} K_{\mathbf{v}_i}^T (K_{XZ}K_{ZX} + M)^{-1} K_{\mathbf{v}_i} \\ &= \frac{\sigma^2}{l}\mathrm{Tr}(K_{VX}(K_{XZ}K_{ZX} + M)^{-1}K_{XV}) \end{aligned} \tag{26}$$

where $K_{\mathbf{v}_i} = [K(\mathbf{x}_1, \mathbf{v}_i), \cdots, K(\mathbf{x}_m, \mathbf{v}_i)]^T$, K_{VX} is a $l \times m$ matrix with $K_{VX,ij} = K(\mathbf{v}_i, \mathbf{x}_j)$, and $K_{XV} = K_{VX}^T$. Then, our problem is to find a subset $\mathcal{Z} \subseteq \mathcal{X}$ to minimize equation (26). Following the steps in Section 3.2, this combinatorial problem can be relaxed as:

Definition 2. Convex Kernel Laplacian Regularized I-optimal Design (CKerLapRID):

$$\begin{aligned} &\min\ \mathrm{Tr}(K_{VX}(\textstyle\sum_{i=1}^{m} \lambda_i K_{\mathbf{x}_i} K_{\mathbf{x}_i}^T + M))^{-1}K_{XV} + \gamma \mathbf{1}^T \boldsymbol{\lambda} \\ &\text{s.t.}\ \ \boldsymbol{\lambda} \succeq 0 \end{aligned} \tag{27}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the variable, and $K_{\mathbf{x}_i} = [K(\mathbf{x}_1, \mathbf{x}_i), \cdots, K(\mathbf{x}_m, \mathbf{x}_i)]^T$.

Using the optimization scheme in Section 3.3, problem (27) can also be cast as an SDP:

$$\begin{aligned} &\min\ \mathrm{Tr}(P) + \gamma \mathbf{1}^T \boldsymbol{\lambda} \\ &\text{s.t.}\ \ \begin{bmatrix} \sum_{i=1}^{m} \lambda_i K_{\mathbf{x}_i} K_{\mathbf{x}_i}^T + M & K_{XV} \\ K_{VX} & P \end{bmatrix} \succeq_{\mathbb{S}_{m+l}^+} 0 \\ &\quad \boldsymbol{\lambda} \succeq 0 \end{aligned} \tag{28}$$

with variables $P \in \mathbb{R}^{l \times l}$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$. After solving it, the most informative images are defined as those with the largest significant indicators (λ_i^*'s).

5. CONTENT BASED IMAGE RETRIEVAL USING CLAPRID

To demonstrate the effectiveness of our proposed CLapRID algorithm, we compare it with Laplacian Regularized Least Squares (LapRLS) [3], Support Vector Machine (SVM) [5], Support Vector Machine Active Learning (SVM$_{active}$) [19,20], and Convex Transductive Experimental Design (CovTED) [24].

The SVM$_{active}$, CovTED, and CLapRID are active learning algorithms, while LapRLS and SVM are standard classification algorithms. SVM only makes use of the labeled images, while LapRLS is a semi-supervised learning algorithm which makes use of both labeled and unlabeled images. For SVM$_{active}$, CovTED, and CLapRID, the training images are selected by the algorithms themselves. While for LapRLS and SVM, we use the top images as training data. It would be important to note that SVM$_{active}$ is based on the ordinary SVM, CovTED is based on ridge regression, and CLapRID is based on LapRLS.

5.1 A Toy Example

A toy example is given in Figure 1. The data set contains two circles with random noise added. There are twenty points on the big circle, and ten points on the small circle. We let CovTED and CLapRID select 6 most informative points respectively. We use the same gaussian kernel for the two algorithms. As can be seen, all the points selected by CovTED are from the small circle, while CLapRID selects 4 points from the big circle and 2 from the small circle. The numbers beside the selected points denote their orders to be selected. Clearly, the points selected by our CLapRID algorithm can better represent the original data set. We did not compare our algorithm with SVM$_{active}$, because SVM$_{active}$ can not be applied in this case due to the lack of the labeled points.

5.2 Relevance Feedback Image Retrieval

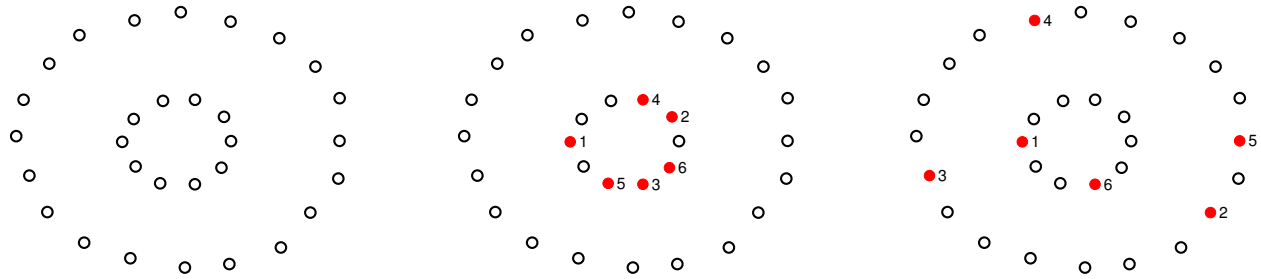
The generic problem of image retrieval is the following. Given a query image and an image database, return the most relevant images to the user. Relevance feedback is one of the most important techniques to narrow down the gap between low level visual features and high level semantic concepts [16]. A typical scenario for relevance feedback in CBIR is as follows [26].

- Initially, the user submits a query image example to the CBIR system. The system ranks the images in database according to some pre-defined distance metric and presents to the user the top ranked images.
- Then, the user is asked to provide relevance judgements on some chosen images. The system uses the user's provided information to re-rank the images in database and returns to the user the top images. This process continues until the user is satisfied.

Active learning algorithms are applied to select the most informative images to label. Take our CLapRID algorithm as an example. Firstly, we use CLapRID to select the most informative images to label. After we get user's feedback, relevant images are treated as having a real value measurement 1, and irrelevant images as -1. A linear function $f(\mathbf{x}) = \widehat{\mathbf{w}}_L^T \mathbf{x}$ can be learned based on LapRLS as discussed in Section 3.1. Then, all the images in database are re-ranked according to their function values.

5.3 Experimental Design

The image database we used consists of 7,700 images of 77 semantic categories, from COREL data set. It is a large and heterogeneous image set. Figure 2 shows some sample images. In this work, we combine 64-dimensional color histogram and 64-dimensional Color Texture Moment (CTM) [22] to represent the images. The color histogram is calculated using $4 \times 4 \times 4$ bins in HSV space. The Color Texture Moment is proposed by Yu et al. [22]. Then, each image is represented as a 128-dimensional vector.



(a) Data set (b) Data points chosen by CovTED (c) Data points chosen by CLapRID

Figure 1: Data selection by different active learning algorithms. The numbers beside the selected points denote their orders to be selected.

(a) Bird (b) Dish (c) Surfing

Figure 2: Sample images from category Bird, Dish, and Surfing.

To exhibit the advantages of using our algorithm, we need a reliable way of evaluating the retrieval performance and the comparisons with other algorithms. We list different aspects of the experimental design below.

5.3.1 Evaluation Metrics

We use *precision-scope curve* and *precision rate* [15] to evaluate the effectiveness of the image retrieval algorithms. The scope is specified by the number (N) of top-ranked images presented to the user. The precision is the ratio of the number of relevant images presented to the user to the scope N. The precision-scope curve describes the precision with various scopes and thus gives an overall performance evaluation of the algorithms. On the other hand, the precision rate emphasizes the precision at a particular value of scope. In general, it is appropriate to present 20 images on a screen. Putting more images on a screen may affect the quality of the presented images. Therefore, the precision at top 20 (N = 20) is especially important.

In real world image retrieval systems, the query image is usually not in the image database. To simulate such environment, we randomly select 20 images per category as query images, and the other images are used as the database for retrieval. The precision-scope curve and precision rate are computed by averaging the results over the 1540 (20×77) queries.

5.3.2 Automatic Relevance Feedback Scheme

We designed an automatic feedback scheme to model the retrieval process. For each submitted query, our system retrieves and ranks the images in the database. At the beginning of retrieval, the Euclidean distances in the original 128-dimensional space are used to rank the images in database. To reduce the computational cost, we select the most informative images from the top 300 images. For active learning algorithms (SVM$_{active}$, CovTED, and CLapRID), 5 images are selected from the database for user labeling. For passive learning algorithms (LapRLS and SVM) the top 5 images are selected for labeling. The label information is used by the system for re-ranking. Images which have been selected at previous iterations are excluded from later selections. For each query, the automatic relevance feedback mechanism is performed for four iterations.

5.4 Experimental Results

For CLapRID and LapRLS algorithms, we use the same graph structure (Eq. 13) and set the value of p (number of nearest neighbors) to be 5. The parameters α and β in our CLapRID algorithm are empirically set to be $1e-3$ and $1e-6$. The parameter γ which controls sparsity is set to be $1e9$. We use Gaussian Kernel for all the five algorithms. SVM and SVM$_{active}$ are implemented based on LIBSVM [6], and CLapRID is implemented based on SDPT3 [18].

Figure 3 shows the precision-scope curves of the five algorithms for the four feedback iterations. It is important to note that, during the retrieval process of each query, the user-labeled images at the previous iterations are excluded from future retrieval. Because SVM$_{active}$ can only be applied when there is a initial classifier available. It can not be applied at the first round and we use the standard SVM to build the initial classifier. As can be seen, our CLapRID algorithm outperforms the other four algorithms on the entire scope.

By iteratively adding the user's feedbacks, the corresponding precision results (at top 10, top 20, and top 30) of the five algo-

Table 1: Precision at top 20 returns of the five algorithms after the second feedback iteration. The highest precision is in bold for each category.

Category	SVM	LapRLS	SVM$_{active}$	CovTED	CLapRID	Category	SVM	LapRLS	SVM$_{active}$	CovTED	CLapRID
Antelope	0.12	0.13	0.15	0.17	**0.21**	Horse	0.71	0.86	0.72	**0.92**	0.90
Antique	0.35	0.48	0.37	0.42	**0.62**	Indoor decorate	0.19	0.50	0.20	0.39	**0.57**
Aquarelle	0.14	0.13	0.16	0.17	**0.18**	Jewelry	0.06	0.07	0.07	**0.10**	0.10
Balloon	0.24	0.25	0.27	0.40	**0.41**	Kungfu	0.88	0.79	0.89	0.89	**0.89**
Beach	0.13	0.13	0.13	0.14	**0.15**	Leopard	0.25	0.22	0.23	0.24	**0.32**
Bead	0.16	0.12	0.16	**0.23**	0.20	Lighthouse	0.10	0.06	0.10	**0.11**	0.10
Bird	0.05	0.05	0.06	**0.07**	0.06	Lion	0.28	0.28	0.26	0.29	**0.31**
Bobsled	0.22	0.25	0.25	0.31	**0.34**	Lizard	0.18	0.17	0.14	0.22	**0.29**
Bonsai	0.22	0.38	0.25	0.38	**0.46**	Marble	0.29	0.28	0.28	**0.35**	0.29
Building	0.08	0.13	0.08	0.12	**0.13**	Mask	0.32	0.44	0.32	0.43	**0.57**
Bus	0.35	0.40	0.33	0.35	**0.44**	Men	0.12	0.10	**0.12**	0.11	0.10
Butterfly	0.31	0.34	0.33	0.31	**0.45**	Model	0.11	0.13	0.12	0.15	**0.16**
Cactus	0.12	0.10	0.14	**0.15**	**0.15**	Mosaic	0.61	0.62	0.57	**0.72**	0.69
Canvas	0.31	0.27	0.27	**0.39**	0.35	Mountain	0.29	0.28	0.25	0.31	**0.40**
Cards	0.88	**0.94**	0.89	0.90	0.93	Old Car	0.41	0.39	0.38	0.41	**0.45**
Castle	**0.17**	0.17	0.16	0.09	0.15	Orbit	0.19	0.20	0.25	**0.28**	0.27
Cat	0.17	0.19	0.20	0.26	**0.30**	Owl	0.63	0.64	0.65	0.67	**0.68**
Cave	0.16	0.19	0.17	0.15	**0.26**	Penguin	0.20	**0.22**	0.18	0.17	0.20
Cell	0.36	0.41	0.44	**0.54**	0.51	Plane	0.12	0.09	0.11	**0.13**	0.11
Cougar	0.10	0.14	0.11	**0.15**	0.11	Postcard	0.74	0.83	0.75	0.87	**0.91**
Couples	0.08	0.08	0.08	0.10	**0.11**	Pyramid	0.46	0.49	0.49	0.59	**0.66**
Cuisine	0.44	0.51	0.40	**0.61**	0.57	Race Car	0.39	0.45	0.40	0.42	**0.49**
Dinosaur	0.96	0.96	0.97	0.92	**1**	Road Sign	0.16	0.18	0.20	**0.26**	0.23
Dish	0.65	0.80	0.67	0.76	**0.84**	Rodeo	0.46	0.48	0.45	0.52	**0.58**
Dog	0.10	**0.14**	0.11	0.12	0.12	Shell	0.14	0.16	0.13	**0.16**	**0.16**
Doll	0.53	0.53	0.53	**0.75**	0.74	Ship	0.31	0.37	0.31	0.35	**0.42**
Drink	0.31	0.41	0.35	0.43	**0.49**	Ski	0.21	0.20	0.22	**0.24**	0.21
Eagle	0.30	0.31	0.31	**0.35**	0.31	Stamp	0.43	0.53	0.40	0.55	**0.61**
Elephant	0.21	0.35	0.29	0.37	**0.40**	Sunset	0.46	0.46	0.45	**0.63**	0.62
Firework	0.44	0.45	0.44	**0.60**	0.48	Surfing	0.45	0.39	0.39	0.45	**0.47**
Fitness	0.75	0.80	0.76	0.69	**0.94**	Tiger	0.24	0.24	0.23	**0.26**	0.24
Flag	0.67	0.72	0.66	0.91	**0.96**	Tools	0.46	0.50	0.48	0.44	**0.55**
flower	0.39	0.38	0.43	0.55	**0.57**	Train	0.22	0.22	**0.23**	0.22	0.21
Forest	0.10	0.12	0.09	0.11	**0.14**	Tropical fish	0.29	0.35	0.30	0.37	**0.38**
Fox	0.07	0.06	0.07	**0.10**	**0.10**	Volcano	0.13	0.14	0.16	**0.27**	0.23
Fruit	0.44	0.50	0.41	0.51	**0.62**	Water fall	0.12	0.18	0.17	0.19	**0.20**
Fungus	0.09	0.12	0.10	**0.13**	0.11	Waves	0.29	0.29	0.29	0.32	**0.37**
Goat	0.14	0.13	0.14	0.12	**0.15**	Wolf	0.12	**0.14**	0.13	0.14	0.13
Gun	0.27	0.36	0.27	0.38	**0.50**						

rithms are respectively shown in Figure 4. Again, CLapRID algorithm performs the best in all the cases. Table 1 shows the precision at top 20 returns after the second feedback iteration for all the 77 categories. As can be seen, the retrieval performance of all the algorithms varies with the different categories. Among all the 77 categories, our CLapRID algorithm performs the best on 48 categories, CovTED performs the best on 25 categories, LapRLS performs the best on 4 categories, SVM$_{active}$ performs the best on 2 categories, and SVM performs the best on 1 category.

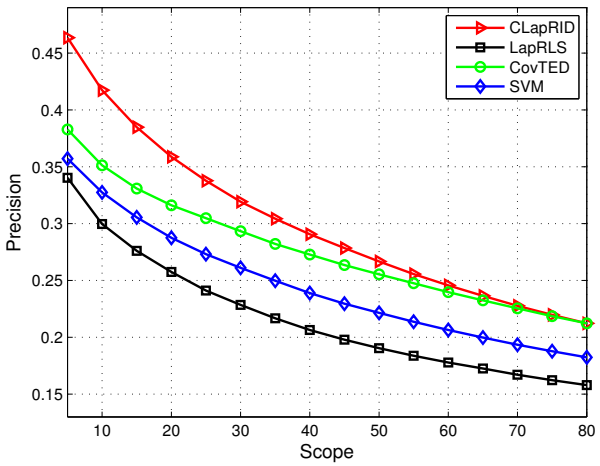
We summarize some important points as follows:

1. CLapRID outperforms the other four algorithms on the entire scope and in every iteration.

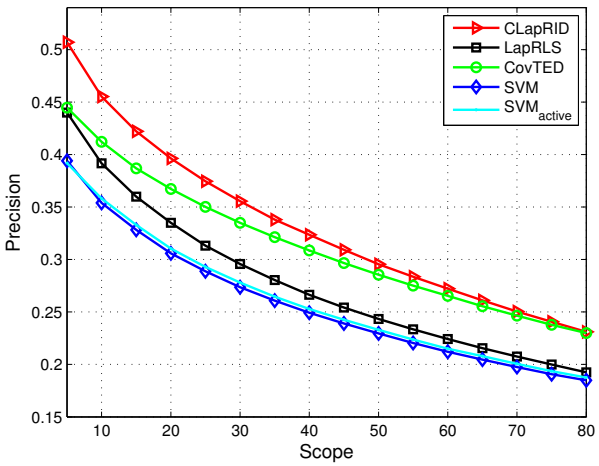
2. Both CovTED and CLapRID are regression based algorithms, and both select points to minimize the average variance of prediction value. The difference is that CovTED is based on a supervised learning algorithm (ridge regression) and CLapRID is based on a semi-supervised learning algorithm (LapRLS). The advantage of CLapRID compared with CovTED shows that exploiting the manifold structure discovered from both labeled and unlabeled data can improve the precision significantly.

3. Both CovTED and CLapRID outperform SVM$_{active}$. This indicates that, for image retrieval, active learning based on regression is more effective than SVM$_{active}$. Compared with SVM, SVM$_{active}$ gained little improvement. The main reason is that in our experiments, we excluded labeled images at the previous iterations from future retrieval, which is the key difference from previous studies [19, 20].

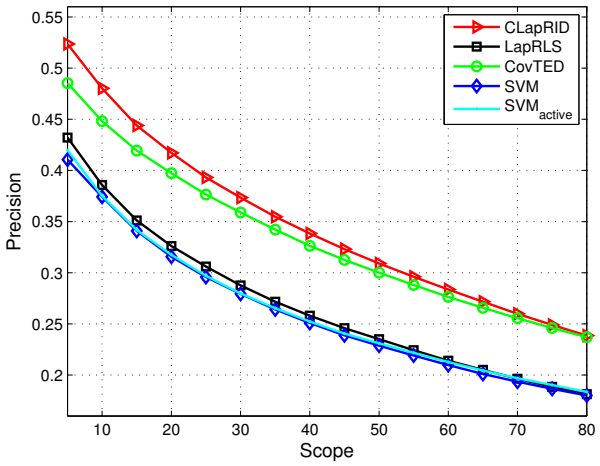
4. The precision rate of CovTED and CLapRID keep increasing as more feedback is added. But for SVM and SVM$_{active}$, the increase rate is very slow. For LapRLS, the precision



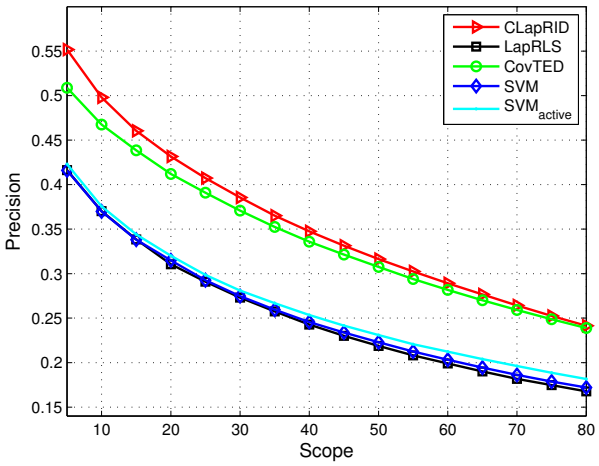
(a) First Iteration



(b) Second Iteration

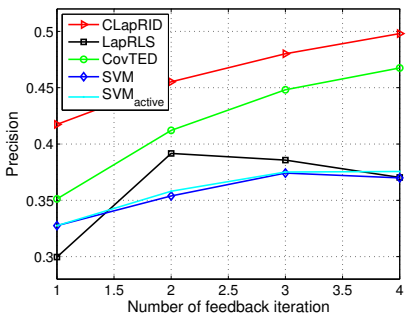


(c) Third Iteration

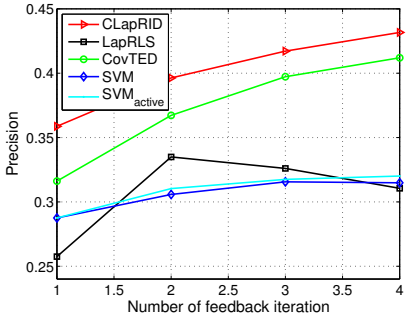


(d) Fourth Iteration

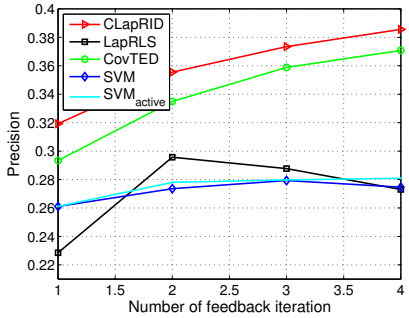
Figure 3: The average precision-scope curves of different algorithms for the four feedback iterations. The CLapRID algorithm performs the best on the entire scope. Note that, at the first round of feedback, the SVM$_{active}$ algorithm can not be applied.



(a) Precision at Top 10



(b) Precision at Top 20



(c) Precision at Top 30

Figure 4: Precision rate of the five learning algorithms at Top 10, 20 and 30. As can be seen, CLapRID consistently outperforms the other four algorithms.

rate even decreases after the second feedback iteration. This phenomenon validates that the top images may not be the most informative ones.

6. CONCLUSIONS

In this paper, we propose a novel active learning algorithm, called Convex Laplacian Regularized I-optimal Design (CLapRID), for relevance feedback image retrieval. Our algorithm is fundamentally based on Laplacian Regularized Least Squares (LapRLS), and motivated by many recent advances in experimental design [2, 14, 23, 24]. CLapRID makes use of both labeled and unlabeled points to discover the intrinsic geometrical structure in the data. It selects images to minimize average variance of prediction value, and can be solved via semidefinite programming. Experimental results on COREL database show that the proposed approach outperforms Support Vector Machines [5], Laplacian Regularized Least Squares [3], Support Vector Machine Active Learning [19,20], Convex Transductive Experimental Design [24].

In this paper we use I-optimal design criterion. However, other classic optimal criteria, such as D-, A-, E-, and G-optimal designs, can also be reformulated under this framework to reflect the underling geometrical structure.

7. ACKNOWLEDGMENTS

This work was supported by China National Key Technology R&D Program (2008BAH26B00 & 2007BAH11B06).

8. REFERENCES

[1] S. P. Asprey and S. Macchietto. Designing robust optimal dynamic experiments. *Journal of Process Control*, 12(4):545–556, 2002.

[2] A. Atkinson, A. Donev, and R. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press, USA, 2007.

[3] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 7:2399–2434, 2006.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[5] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[6] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[7] O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006.

[8] F. R. K. Chung. *Spectral Graph Theory*, volume 92 of *Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[9] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.

[10] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[11] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, volume 65 of *Applied Optimization*. Kluwer Academic Publishers, Dordrecht, 2002.

[12] R. M. Freund and S. Mizuno. Interior point methods: Current status and future directions. In *High Performance Optimization*, pages 441–466. H. Frenk et al. (eds.), Kluwer Academic Publishers, 2000.

[13] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

[14] X. He, W. Min, D. Cai, and K. Zhou. Laplacian optimal design for image retrieval. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 119–126, New York, NY, USA, 2007. ACM.

[15] D. P. Huijsmans and N. Sebe. How to complete performance graphs in content-based image retrieval: Add generality and normalize scope. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):245–251, 2005.

[16] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, Sep 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 Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[18] K. C. Toh, M. J. Todd, and R. H. Tütüncü. Sdpt3 — a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11(1):545–581, 1999.

[19] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, pages 107–118, New York, NY, USA, 2001. ACM.

[20] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, 2002.

[21] K. Vu, K. A. Hua, and N. Jiang. Improving image retrieval effectiveness in query-by-example environment. In *SAC '03: Proceedings of the 2003 ACM symposium on Applied computing*, pages 774–781, New York, NY, USA, 2003. ACM.

[22] H. Yu, M. Li, H. jiang Zhang, and J. Feng. Color texture moments for content-based image retrieval. In *International Conference on Image Processing*, pages 929–932, 2002.

[23] K. Yu, J. Bi, and V. Tresp. Active learning via transductive experimental design. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 1081–1088, New York, NY, USA, 2006. ACM.

[24] K. Yu, S. Zhu, W. Xu, and Y. Gong. Non-greedy active learning for text categorization using convex transductive experimental design. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 635–642, New York, NY, USA, 2008. ACM.

[25] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. S. Olkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328. MIT Press, 2004.

[26] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, April 2003.