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 opti-

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 combin ch748y

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].

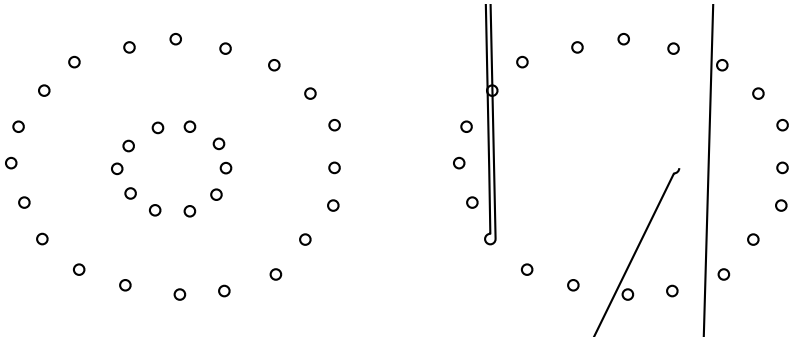
4.1 Kernel Laplacian Regularized Least Squares

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:

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



(a) Data set

precision-scope curve

e experi-mental design below.

ithms. The numbers beside the selected points denote their o

ompar-isons with other algorithms. We

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									

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

[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.