# Relevance Aggregation Projections for Image Retrieval

Wei Liu
Department of Electrical Engineering
Columbia University
New York, NY 10027, USA
wliu@ee.columbia.edu

Wei Jiang
Department of Electrical Engineering
Columbia University
New York, NY 10027, USA
wjiang@ee.columbia.edu

Shih-Fu Chang
Department of Electrical Engineering
Columbia University
New York, NY 10027, USA
sfchang@ee.columbia.edu

## ABSTRACT

To narrow the semantic gap in content-based image retrieval (CBIR), relevance feedback is utilized to explore knowledge about the user's intention in finding a target image or a image category. Users provide feedback by marking images returned in response to a query image as relevant or irrelevant. Existing research explores such feedback to refine querying process, select features, or learn a image classifier. However, the vast amount of unlabeled images is ignored and often substantially limited examples are engaged into learning. In this paper, we address the two issues and propose a novel effective method called Relevance Aggregation Projections (RAP) for learning potent subspace projections in a semi-supervised way. Given relevances and irrelevances specified in the feedback, RAP produces a subspace within which the relevant examples are aggregated into a single point and the irrelevant examples are simultaneously separated by a large margin. Regarding the query plus its feedback samples as labeled data and the remainder as unlabeled data, RAP falls in a special paradigm of imbalanced semi-supervised learning. Through coupling the idea of relevance aggregation with semi-supervised learning, we formulate a constrained quadratic optimization problem to learn the subspace projections which entail semantic mining and therefore make the underlying CBIR system respond to the user's interest accurately and promptly. Experiments conducted over a large generic image database show that our subspace approach outperforms existing subspace methods for CBIR even with few iterations of user feedback.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Relevance feedback*

## General Terms

Algorithms, Experimentation, Measurement, Performance, Theory

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*CIVR'08*, July 7–9, 2008, Niagara Falls, Ontario, Canada.
Copyright 2008 ACM 978-1-60558-070-8/08/07 ...$5.00.

## Keywords

Image Retrieval, Relevance Feedback, Dimensionality Reduction, Subspace Learning, Semi-supervised Learning, Relevance Aggregation Projections.

## 1. INTRODUCTION

Content-based image retrieval (CBIR) [17] has been an active research area in the last decade. In the CBIR paradigm, an image is usually represented by a set of low-level visual features, which often do not have direct connection to high-level semantic concepts. The gap between high-level semantic concepts and low-level visual features remains to be the major obstacle hindering development of CBIR systems. Among several promising efforts, relevance feedback [15][19][10][11][12][13][20][3] has been proposed to narrow the gap. The relevance feedback mechanism is an iterative learning process, which is conventionally treated as supervised learning [15][19][10][12]. During each iteration, the user labels some images to be "relevant" or "irrelevant" according to the query image he provides and the semantic target in mind. The system uses the labeled images as training samples to successively refine the learning model and gives better retrieval quality in the subsequent iteration.

Two key stages known in CBIR are querying and relevance feedback. In the first stage, the user raises a query image as an example of his interested "concept" and the CBIR system later returns the most relevant images with appropriate features (global or local) and a suitable distance metric. In the second stage, the user labels the returned images as positive or negative samples according to their relevances or irrelevances to the query. After that, the CBIR system either refines the distance metric or learns a classification model, which leads to another set of returned images. This process will be iterated until the system's outcome converges to the user's provided concept, gradually narrowing the gap between the user's intention and the response of the system.

In each iteration of relevance feedback, discriminative models such as support vector machines [16] have been used to explore the positive and negative labeled samples to build a decision function which is able to classify any unlabeled samples as relevant or irrelevant. However, traditional CBIR methods [19][10][12] greatly suffer from the scarcity of available labeled images. Characteristics of the vast amount of images remaining in the database that have not been labeled are ignored. As we know, the performance of CBIR depends on the generalization capacity of the adopted models on abundant unlabeled images to be retrieved. All these aspects urge us to draw on the unlabeled samples.

Semi-supervised learning [4] (SSL) is a promising machine learning technique designed for the situations where only few labeled data are available and a large amount of data are unlabeled. The core idea of SSL is to take advantage of both labeled and unlabeled data in optimizing some objective functions considering various criteria such as consistence with known labels, prediction smoothness, and locality preservation. This technique can be readily applied to a wide variety of real-world classification problems in which unlabeled data can be easily obtained, while the acquisition of labeled data is expensive. A family of semi-supervised learning algorithms [22][21][2] have been proposed based on spectral graph theory [6].

In general, the dimensionality of an image space is very large, ranging from several hundreds to thousands. When CBIR systems apply statistical techniques to interactive retrieval tasks, one crucial issue called the "curse of dimensionality" is usually encountered. The high dimensionality makes many methods which are computationally manageable in low dimensional spaces completely intractable. Hence, reducing the dimensionality of the image space is necessary when pursuing computationally manageable techniques for CBIR. Due to the limitations related to the above two issues, current CBIR solutions have not been able to achieve satisfactory performance with adequate generalization capabilities to diverse domains.

Recently, dimensionality reduction has also been a central topic of machine learning research. It assumes that a linear subspace or a nonlinear submanifold be embedded in a high-dimensional ambient space such as an image space. Since 2000, many nonlinear dimensionality reduction methods such as [18][14][1] have emerged in the subarea called "manifold learning" which attempts to discover or learn the low-dimensional submanifold. However, many existing methods are unsupervised and not applicable to new data points. In contrast, linear dimensionality reduction or subspace learning can be naturally extended to novel data points outside the training set and is thus more compatible with the setting described above for CBIR.

Two of the most popular subspace learning techniques are PCA and LDA [7] which have been extensively used in numerous computer vision and pattern recognition applications. PCA finds a maximum metric subspace for data representation and reconstruction. LDA seeks a discriminative subspace for classification. PCA is unsupervised while LDA is fully supervised. Locality Preserving Projections (LPP) [8] aim to find a set of projections on which data are projected with local geometric structures preserved. Following LPP, Local Discriminant Embedding (LDE) [5] generalizes standard LDA from the point of view of locality discrimination. LDE is essentially a supervised version of LPP.

Subspace learning is broadly exploited due to its ease of implementation. A good subspace $A$ readily sets up a distance metric as $(\mathbf{x}_i - \mathbf{x}_j)^T AA^T(\mathbf{x}_i - \mathbf{x}_j)$ between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the subspace, and will alleviate the high-dimensionality problem discussed above. To exploit the potential of unlabeled data, subspace learning should be redesigned to work under a semi-supervised setting. With both unlabeled images and labeled images collected from relevance feedback, image representation within semi-supervised subspaces can better reveal the semantic structure of the image data and meanwhile improve the generalization capabilities of the underlying CBIR systems.

Interesting semi-supervised subspace learning algorithms for CBIR have been developed recently, including Augmented Relation Embedding (ARE) [13], Semantic Subspace Projection (SSP) [20] and Spectral Regression [3]. All of these algorithms are based on a *manifold assumption* that images reside in or close to a submanifold embedded in the ambient space. Analogous to the principle used in Laplacian Eigenmaps [1], they employ the graph Laplacian [6] to push nearby images close to each other in the desired subspaces and maintain certain discriminating properties from the labeled images. These approaches have been shown useful for CBIR in several experiments. Nonetheless, information is lacking about the intrinsic dimensions of the subspaces learned by these algorithms although they hold on the same manifold assumption. Moreover, all of the prior work ignores the intrinsic asymmetry in CBIR that requires to treat the "relevant" and "irrelevant" sets unequally with more emphasis needed for the "relevant" set. The relevant images are more important for semi-supervised subspace learning because they (along with the query image) jointly define the underlying semantic target.

In this paper, we develop a novel framework for semi-supervised subspace leaning in the context of CBIR with relevance feedback. SSL has been applied to solve many computer graphics and vision problems ranging from interactive image colorization, interactive image segmentation, object categorization, and object tracking. The proposed learning framework successfully marries relevance feedback and SSL, addresses the two fundamental problems mentioned earlier, and is thus expected to reduce the gap between low-level features and high-level semantics. Using the framework, we propose a new algorithm Relevance Aggregation Projection (RAP) to learn a set of semantically meaningful projections which span the image subspace for retrieval. Our algorithm relies on the simple geometric intuition that a good subspace is one within which the relevant examples including the query are aggregated into a single point and the irrelevant examples are simultaneously separated by a large margin. We construct a constrained quadratic optimization problem whose solution generates such a subspace via QR factorization and Laplacian regularization.

The rest of this paper is organized as: Section 2 reviews the related work in subspace learning applied to CBIR. Section 3 describes our subspace learning method based on relevance aggregation. Section 4 presents an experimental study on a large database composed of generic images. Conclusions are drawn in Section 5.

## 2. RELATED WORK

We first describe a general framework for subspace learning, which unifies almost all existing subspace learning algorithms from the viewpoint of optimization. We state that learning a subspace $A \in \mathbb{R}^{d\times r}$ $(r \le d)$ for a special intent may be formulated as maximizing the generalized Rayleigh quotient

$$\max_A \mathcal{R}(S_1, S_2, A) = \frac{tr(A^T S_1 A)}{tr(A^T S_2 A)}, \tag{1}$$

where $S_1$ and $S_2$ are $d \times d$ real symmetric matrices. The optimal $A$ is solved such that its columns are the eigenvectors corresponding to the maximum eigenvalues of the generalized eigen-problem:

$$S_1\mathbf{a} = \lambda S_2\mathbf{a}. \tag{2}$$

The general framework eq. (1) shows that the two matrices $S_1$ and $S_2$ play essential roles in designing subspace approaches. The choices of $S_1$ and $S_2$ can be very flexible, and with different choices this framework will lead to many popular subspace algorithms, e.g., PCA, LDA, LPP and LDE. Let us denote the total scatter matrix, the within-class scatter matrix, and the between-class scatter matrix as $S_t$, $S_w$, and $S_b$, respectively ($S_t = S_w + S_b$). When $S_1 = S_t$ and $S_2 = I$, the framework reduces to PCA. When $S_1 = S_b$ and $S_2 = S_w$, the framework becomes LDA.

To this end, we do not access the manifold assumption to which diverse kinds of machine learning techniques such as nonlinear dimensionality reduction and spectral clustering conform. The common characteristic of manifold-based learning techniques is to resort to graphs and their Laplacians to approximate the manifold structure of the observed data. Specifically, we construct an undirected weighted graph $G(\mathbf{V}, \mathbf{E}, W)$ over all $n$ data points $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^d$, including labeled and unlabeled, of which each point $\mathbf{x}_i$ corresponds to a node $v_i \in \mathbf{V}$ in the graph. An edge $(v_i, v_j) \in \mathbf{E}$ is established between two nodes $v_i$ and $v_j$ if the corresponding two points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close, i.e., they are among $k$ nearest neighbors of each other. Afterwards, we may weight the edge $(v_i, v_j)$ as $W_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma^2)$. This forms the weight matrix $W \in \mathbb{R}^{n\times n}$ as well as the diagonal degree matrix $D \in \mathbb{R}^{n\times n}$ where $D_{ii} = \sum_j W_{ij}$ (notice $W_{ii} = 0$). It is noticeable that $G$ and $W$ can be redefined to characterize certain statistical or geometric properties of the data set. The $k$-NN graph and the weighting function of the RBF kernel are the most intuitive ones. What's more, we can construct graphs supervisedly in contrast to traditional unsupervised graph construction schemes.

Over a well-defined graph, dimensionality reduction is described as mapping each node of the graph to a low dimensional data vector which is expected to maintain connections between adjacent nodes. Obviously, the connections are measured by the edge strengths, i.e. weights. This mapping process is called as *graph embedding* in [9]. Typically, LPP introduces a local scatter matrix $S_{local}$ to implement linear graph embedding:

$$S_{local} = \frac{1}{2}\sum_{i,j=1}^{n}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T W_{ij} = XLX^T, \qquad (3)$$

where $X = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ is the sample matrix and $L = D - W \in \mathbb{R}^{n\times n}$ is the graph Laplacian matrix. As a key ingredient of spectral graph theory [6], graph Laplacians have received increasing attention in a lot of fields including dimensionality reduction, clustering, and semi-supervised learning. We find that LPP still falls in the general framework eq. (1) since its objective is

$$A^{LPP} = \arg\max_A \frac{tr(A^TXWX^TA)}{tr(A^TXLX^TA)}. \qquad (4)$$

For the sake of CBIR, we will briefly review three state-of-the-art subspace learning algorithms: Augmented Relation Embedding (ARE) [13], Semantic Subspace Projection (SSP) [20], and Spectral Regression (SR) [3]. All these algorithms also conform to the general framework. Particularly, the disadvantages of these algorithms will be pointed out within this framework.

## 2.1 Augmented Relation Embedding (ARE)

To embody the semantic relations, ARE uses an extra relational graph $G^{ARE}$ to encode the label information determined in the user's relevance feedbacks. Suppose $F^+$ denotes the index set of relevant images including the query in one iteration of feedback, and $F^-$ denotes the index set of irrelevant images. In a supervised way, ARE builds a relational graph as:

$$W_{ij}^{ARE} = \begin{cases} -\alpha, & i \in F^+ \wedge j \in F^+ \wedge i \neq j \\ 1, & (i \in F^+ \wedge j \in F^-) \vee (i \in F^- \wedge j \in F^+) \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

where $\alpha > 0$ is a parameter in charge of the possibility of unbalanced feedback. Note that $W_{ii}^{ARE} = 0$.

ARE acquires the optimal subspace by:

$$A^{ARE} = \arg\max_A \frac{tr(A^TXL^{ARE}X^TA)}{tr(A^TXLX^TA)}, \qquad (6)$$

where $L^{ARE} = D^{ARE} - W^{ARE}$ is the graph Laplacian of the new graph $G^{ARE}$. Clearly, ARE and LPP share and use the same denominator of the general framework.

Now we analyze the intrinsic dimension $r(< d)$ of the subspace $A^{ARE}$ learned by ARE. In accordance with eq. (2), $A^{ARE}$ is composed of the eigenvectors associated with the $r$ largest eigenvalues of the eigen-system $eig(XL^{ARE}X^T, XLX^T)$. Because $L^{ARE}$, $L$, $XL^{ARE}X^T$ and $XLX^T$ are all positive semidefinite, the considered eigenvalues must not include zeros in order to exclude trivial eigenvectors in $A^{ARE}$. As shown in [6], $L$ has at least one zero eigenvalue while $L^{ARE}$ has exactly $n - |F^+ \cup F^-| + 1$ zero eigenvalues since the maximal connected subgraph $G^{ARE}(\{v_i\}_{i\in F^+\cup F^-})$ is single-connected. Consequently, we have

$$\begin{aligned} r = rank(A^{ARE}) &\leq \min\{rank(XL^{ARE}X^T), rank(XLX^T)\} \\ &\leq \min\{rank(L^{ARE}), rank(L)\} \\ &\leq \min\{|F^+ \cup F^-| - 1, n - 1\} \\ &= |F^+ \cup F^-| - 1, \end{aligned} \qquad (7)$$

which reveals that ARE can provide at most $|F^+ \cup F^-| - 1$ nontrivial projections to span the desired subspace $A^{ARE}$.

## 2.2 Semantic Subspace Projection (SSP)

Following ARE, SSP also defines a relational graph $G^{SSP}$ to encode the label information provided by the user's relevance feedbacks, that is

$$W_{ij}^{SSP} = \begin{cases} 1, & (i \in F^+ \wedge j \in F^-) \vee (i \in F^- \wedge j \in F^+) \\ 0, & \text{otherwise} \end{cases} \qquad (8)$$

SSP looks for the optimal subspace by:

$$A^{SSP} = \arg\max_A \frac{tr(A^TX\overline{W}^TL^{SSP}\overline{W}X^TA)}{tr(A^TX\widetilde{L}X^TA)}, \qquad (9)$$

where $L^{SSP}$ is the graph Laplacian of $G^{SSP}$ and $\overline{W} = D^{-1}W$. Let us define a diagonal matrix $\widetilde{D} \in \mathbb{R}^{n\times n}$ with the entries being the row sums of $\overline{W} + \overline{W}^T$. Then we compute $\widetilde{L} = \widetilde{D} - \overline{W} - \overline{W}^T$ that behaves as a new Laplacian matrix corresponding to the weight matrix $\overline{W} + \overline{W}^T$.

Similar to $G^{ARE}$, the maximal connected subgraph $G^{SSP}(\{v_i\}_{i\in F^+\cup F^-})$ is also single-connected, which implies

$rank(L^{SSP}) = |F^+ \cup F^-| - 1$. We herewith derive follows

$$\begin{aligned} r = rank(A^{SSP}) &\leq \min\{rank(\overline{X}L^{SSP}\overline{X}^T), rank(X\widetilde{L}X^T)\} \\ &\leq \min\{rank(L^{SSP}), rank(\widetilde{L})\} \\ &\leq \min\{|F^+ \cup F^-| - 1, n - 1\} \\ &= |F^+ \cup F^-| - 1, \qquad (10) \end{aligned}$$

where $\overline{X} = X\overline{W}^T$. So far, it turns out that SSP is also able to learn at most $|F^+ \cup F^-| - 1$ nontrivial projection vectors like ARE.

## 2.3 Spectral Regression (SR)

To utilize the label information, SR constructs a labeled graph $G^{SR}$ as

$$W_{ij}^{SR} = \begin{cases} 1/|F^+|, & i \in F^+ \wedge j \in F^+ \\ 1/|F^-|, & i \in F^- \wedge j \in F^- \\ 0, & \text{otherwise} \end{cases} \qquad (11)$$

in which two "classes" $F^+, F^-$ are explicitly formed. Notice $W_{ii}^{SR} \neq 0$ for $i \in F^+ \cup F^-$. Define a diagonal matrix $D^{SR} \in \mathbb{R}^{n\times n}$ where $D_{ii}^{SR} = \sum_{j=1}^{n} W_{ij}^{SR}$.

The optimal subspace of SR is obtained by

$$A^{SR} = \arg\max_{A} \frac{tr(A^T X W^{SR} X^T A)}{tr\,(A^T X(D^{SR} + L)X^T A)}, \qquad (12)$$

which invokes to solve the dense eigen-system $eig(XW^{SR}X^T, X(D^{SR} + L)X^T)$. When the image dimension $d$ is rather large, the eigen-solver takes high computational cost. Hence, SR instead solves the sparse eigen-system:

$$\mathbf{y}^{SR} = \arg\max_{\mathbf{y}} \frac{\mathbf{y}^T W^{SR}\mathbf{y}}{\mathbf{y}^T(D^{SR} + L)\mathbf{y}}, \qquad (13)$$

where $W^{SR}, D^{SR}, L$ are all sparse due to the sparse construction schemes of graphs $G^{SR}, G$. According to each learned eigenvector $\mathbf{y}^{SR}$, SR finds the optimal projection vector using the following ridge regression

$$\mathbf{a}^{SR} = \arg\min_{\mathbf{a}} \|X^T\mathbf{a} - \mathbf{y}^{SR}\|^2 + \beta\|\mathbf{a}\|^2, \qquad (14)$$

where $\beta > 0$ is the regularization parameter to control the shrinkage of the above regularized least square problem.

We also derive the dimension of the SR subspace by

$$\begin{aligned} r &= rank(A^{SR}) = |\{\mathbf{y}^{SR}\}| \\ &= \min\{rank(W^{SR}), rank(D^{SR} + L)\} \\ &= \min\{2, rank(D^{SR} + L)\} = 2. \qquad (15) \end{aligned}$$

Thus, the dimension of the SR subspace is the number of classes formed in $W^{SR}$.

# 3. RELEVANCE AGGREGATION PROJECTIONS (RAP)

In all CBIR systems, the involved learning process must tackle a fundamental problem: what features are more representative for explaining the current query concept than the others. This refers to the problem of feature extraction. Projecting the original vector-formed samples into a known subspace is linear feature extraction, which is the issue we mainly address in this paper. Compared with other machine learning problems, subspace learning for CBIR has two challenges. 1) Small size of the labeled set: the labeled images, specified by the user during each query session, are very few compared with the image dimension and the size of the database. Therefore, the unlabeled samples should be utilized to prevent overfitting the few labeled samples. 2) Intrinsic asymmetry: the images labeled to be "relevant" during a query session share some common semantic cues, while the "irrelevant" images are different from the "relevant" ones in different ways. Thus, the relevant images are more important for the system to grasp the query concept. This asymmetry requirement makes it necessary to treat the relevant and irrelevant sets unequally with an emphasis on the relevant one. Through reviewing previous work in Section 2, we find that both SSP and SR fail to engage the asymmetry inherent in relevance feedback. SSP only emphasizes the irrelevant set, while SR treats the relevant and irrelevant sets equally, i.e., two classes. Our method will highlight the difference between the relevant and irrelevant sets.

The subspace dimension is a key parameter for a great number of projection-related methods: if the dimension is too small, important features might concentrate on the same projection direction, and if the dimension is too large, the projections undergo noise and, in some cases, are unstable. Since the user labels a very small fraction of image samples, $|F^+ \cup F^-| \ll n$ holds. The conclusions in eq. (7)(10)(15) tell that any of ARE, SSP and SR produces a very low-dimensional subspace. Especially, SR produces a 2D subspace. In this section, we will seek a subspace of higher dimensions.

## 3.1 Learning Prototype

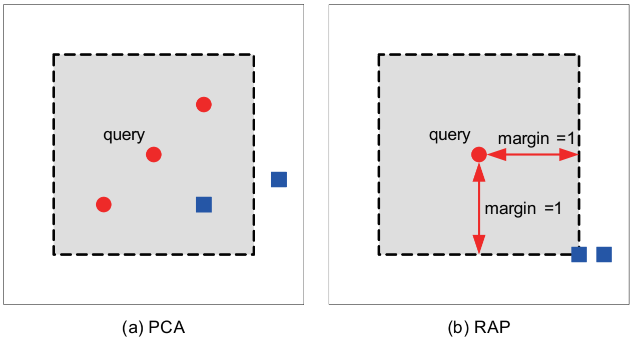

**Figure 1: Schematic illustration of relevance aggregation. For the query point, two points with the same color and shape share the same label, and the others take different labels. (a) 2D projections by PCA; (b) 2D projections by RAP which optimizes projections such that: (i) the positive points are aggregated into a single point; (ii) the negative points lie outside the unit square centered on the query, with a margin of at least one unit distance at each projection direction.**

Regarding the query plus its feedback images as labeled data and the remainder as unlabeled data, the feedback scheme actually provides an imbalanced paradigm of semi-supervised learning. Let us consider the relevant images (including the query) in $F^+$ as positive labeled samples, and the irrelevant images in $F^-$ as negative labeled samples. Also let $l^+ = |F^+|$ and $l^- = |F^-|$. We form the sample matrix $X = [\mathbf{x}_1, \cdots, \mathbf{x}_l, \mathbf{x}_{l+1}, \cdots, \mathbf{x}_n]$ such that the first $l = l^+ + l^-$ columns correspond to the labeled samples.

As a basis of imbalanced semi-supervised subspace learning, the learning prototype is depicted in follows

$$\begin{aligned}
&\min_{A\in\mathbb{R}^{d\times r}} \; tr(A^TXLX^TA) \\
&s.t. \quad A^T\mathbf{x}_i = \sum_{j\in F^+} A^T\mathbf{x}_j/l^+, \forall\, i \in F^+ \\
&\qquad \left\| A^T(\mathbf{x}_i - \sum_{j\in F^+}\mathbf{x}_j/l^+)\right\|^2 \geq r, \forall\, i \in F^-
\end{aligned} \tag{16}$$

Like LPP, this prototype also minimizes the local scatter of $tr(A^TXLX^TA)$ so that nearby samples are pushed together in the target subspace $A$. What's more, this prototype enforces two constraints in eq. (16) to successfully endow the subspace with a good property of margin maximization.

From the viewpoint of projection vectors $\{\mathbf{a}_i\}_{i=1}^r$ spanning $A$, we parsimoniously transform eq. (16) to follows in terms of each projection vector $\mathbf{a} \in \mathbb{R}^d$

$$\begin{aligned}
&\min_{\mathbf{a}} \; \mathbf{a}^TXLX^T\mathbf{a} \\
&s.t. \quad \mathbf{a}^T\mathbf{x}_i = \mathbf{a}^T\mathbf{c}^+, \forall\, i \in F^+ \\
&\qquad \left\|\mathbf{a}^T(\mathbf{x}_i - \mathbf{c}^+)\right\|^2 \geq 1, \forall\, i \in F^-
\end{aligned} \tag{17}$$

in which $\mathbf{c}^+ = \sum_{j\in F^+}\mathbf{x}_j/l^+$ is the positive center. Figure 1 schematically illustrates the geometrical intuition behind eq. (17) that on each projection direction, (i) the positive points collapse into a single point; (ii) the negative points and the aggregated positive points maintain a large margin of at least one unit distance. Naturally, we call the projections optimized by eq. (17) as **Relevance Aggregation Projections** since the relevant images, i.e. the positive samples, are indeed aggregated.

## 3.2 Solution

Eq. (17) formulates a quadratically constrained quadratic optimization problem, but the quadratic constraint is not convex. So it is very hard to solve directly. Here we adopt a heuristic trick to explore the solution.

### *3.2.1 Initialization with PCA*

We intend to obtain initial projections using PCA. Without loss of generality, we assume that $\{\mathbf{x}_i\}_{i=1}^n$ be zero-centered. This can be simply achieved by subtracting the mean vector from all $\mathbf{x}_i$s. Let $U$ consist of the $r < \min\{d, n\}$ principle eigenvectors of $XX^T$, i.e., $U = [\mathbf{u}_1, \cdots, \mathbf{u}_r]$, corresponding to the eigenvalues $\lambda_1, \cdots, \lambda_r$ in a decreasing order. Then we define the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_r)$ and have

$$U^TXX^TU = \Lambda. \tag{18}$$

We calculate the whitened eigenvectors $V \in \mathbb{R}^{d\times r}$ by

$$V = U\Lambda^{-1/2}, \tag{19}$$

such that

$$V^TXX^TV = \Lambda^{-1/2}U^TXX^TU\Lambda^{-1/2} = I. \tag{20}$$

For each column vector $\mathbf{v} \in \mathbb{R}^d$ in $V$, eq. (20) leads to

$$\begin{aligned}
& \mathbf{v}^TXX^T\mathbf{v} = 1 \\
\Longrightarrow \quad & \sum_{i=1}^n \left(\mathbf{v}^T\mathbf{x}_i\right)^2 = 1 \\
\Longrightarrow \quad & |\mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{x}_j| < 2, \; i,j = 1,\cdots,n.
\end{aligned} \tag{21}$$

### *3.2.2 Relevance Aggregation with QR Factorization*

If the two constraints were removed, eq. (17) would be easy to solve via routine optimization procedures. Prompted by eq. (21), we may amend the 1D projections $\{\mathbf{v}^T\mathbf{x}_i\}_{i=1}^l$ of the labeled points as

$$y_i = \begin{cases} \mathbf{v}^T\mathbf{c}^+, & i \in F^+ \\ \mathbf{v}^T\mathbf{x}_i, & i \in F^- \wedge |\mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{c}^+| \geq 1 \\ \mathbf{v}^T\mathbf{c}^+ + 1, & i \in F^- \wedge 0 \leq \mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{c}^+ < 1 \\ \mathbf{v}^T\mathbf{c}^+ - 1, & i \in F^- \wedge -1 < \mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{c}^+ < 0 \end{cases} \tag{22}$$

Suppose $X_l = [\mathbf{x}_1, \cdots, \mathbf{x}_l] \in \mathbb{R}^{d\times l}$ and $\mathbf{y} = [y_1, \cdots, y_l]^T$. We impose

$$X_l^T\mathbf{a} = \mathbf{y}, \tag{23}$$

which entirely satisfies the two constraints of eq. (17). Due to $l \ll d$ in this paper, QR factorization on $X_l$ results in

$$X_l = [Q_1 \; Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1R, \tag{24}$$

where $[Q_1\, Q_2] \in \mathbb{R}^{d\times d}$ is a unitary matrix, forming a set of bases in $\mathbb{R}^d$. Furthermore, $Q_1 \in \mathbb{R}^{d\times l}$, $Q_2 \in \mathbb{R}^{d\times(d-l)}$, $Q_1^TQ_2 = 0$, and $R^{l\times l}$ is an invertable matrix.

The target projection vector $\mathbf{a}$ satisfying eq. (23) must be expressed in

$$\mathbf{a} = Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2, \tag{25}$$

and then we deduce

$$\begin{aligned}
& X_l^T\mathbf{a} = R^TQ_1^T(Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2) = R^T\mathbf{b}_1 = \mathbf{y} \\
& \Longrightarrow \mathbf{b}_1 = (R^T)^{-1}\mathbf{y}.
\end{aligned} \tag{26}$$

With solved $\mathbf{b}_1$ and arbitrary $\mathbf{b}_2$, $\mathbf{a}$ in the expression of eq. (25) implements relevance aggregation as well as margin maximization.

### *3.2.3 Semi-supervised Learning with Regularization*

Most semi-supervised learning approaches [22][21][2] try to minimize a transductive energy function containing 1) a fidelity term ensuring the consistency of the labels of graph transduction and the prior labels provided by the user; and 2) a regularization term ensuring that neighboring data points are likely to have the same labels. Again, we develop a novel transductive framework to optimize projections instead of labels, and the fidelity term is designed as $\|\mathbf{a} - \mathbf{v}\|^2$. Incorporating the objective function of eq. (17) into the regularization term, our framework is formulated by

$$f(\mathbf{a}) = \|\mathbf{a} - \mathbf{v}\|^2 + \gamma\mathbf{a}^TXLX^T\mathbf{a}, \tag{27}$$

where $\gamma > 0$ is the regularization parameter that controls the trade-off between PCA initialization and Laplacian-driven transduction. Plugging $\mathbf{a} = Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2$ into the above equation, we have

$$\begin{aligned}
\min_{\mathbf{b}_2} \; f(\mathbf{b}_2) \;\; &= \;\; \mathbf{b}_2^T(\mathbf{b}_2 - 2Q_2^T\mathbf{v}) \\
&+ \;\; \gamma\mathbf{b}_2^TQ_2^TXLX^T(Q_2\mathbf{b}_2 + 2Q_1\mathbf{b}_1).
\end{aligned} \tag{28}$$

The derivatives of eq. (28) with respect to $\mathbf{b}_2$ will vanish at the minimizer $\mathbf{b}_2^*$:

$$\mathbf{b}_2^* = \left(I + \gamma Q_2^T XLX^T Q_2\right)^{-1} \left(Q_2^T \mathbf{v} - \gamma Q_2^T XLX^T Q_1 \mathbf{b}_1\right). \quad (29)$$

So far, we grasp a close-form solution $\mathbf{a}^* = Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2^*$ to the original optimization problem eq. (17).

## 3.3 Algorithm

The proposed prototype for imbalanced semi-supervised subspace learning and its solution lead to the Relevance Aggregation Projection (RAP) algorithm, which is summarized in below. It is appreciable that the dimensionality $r$ of the RAP subspace is independent on the size of the labeled set and can stretch until $r = min\{d, n-1\}$.

```
1. Construct a $k$-NN graph: Construct an undirected, weighted graph $G$ upon all $n$ input samples $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$[1]:

   $W_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right), \ \mathbf{x}_i \in \mathcal{N}^k(\mathbf{x}_j) \vee \mathbf{x}_j \in \mathcal{N}^k(\mathbf{x}_i) \\ 0, \ \text{otherwise} \end{cases}$   (30)

   in which $\mathcal{N}^k(\mathbf{x}_i)$ denotes the set consisting of $k$-nearest neighbors of $\mathbf{x}_i$. Calculate the graph Laplacian $L = D - W$ where $D_{ii} = \sum_{j=1}^n W_{ij}$, and the local scatter matrix $S = XLX^T$.
2. PCA initialization: Run PCA on $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ to get the matrix $V = [\mathbf{v}_1, \cdots, \mathbf{v}_r] \in \mathbb{R}^{d \times r}$ $(r < d)$ such that $V^T XX^T V = I$.
3. QR factorization: Perform QR factorization on the labeled data matrix $X_l \in \mathbb{R}^{d \times l}$, resulting in $Q_1, Q_2, R$ according to eq. (24).
4. Transductive Regularization:
     For $j = 1$ to $r$
         given $\{\mathbf{v}_j^T \mathbf{x}_i\}_{i=1}^l$, use eq. (22) to form $\mathbf{y}$;
         $\mathbf{b}_1 \longleftarrow (R^T)^{-1}\mathbf{y}$;
         $\mathbf{b}_2 \longleftarrow \left(I + \gamma Q_2^T S Q_2\right)^{-1} \left(Q_2^T \mathbf{v}_j - \gamma Q_2^T S Q_1 \mathbf{b}_1\right)$;
         $\mathbf{a}_j \longleftarrow Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2$;
     End.
5. Projecting: Form the matrix $A^{RAP} = [\mathbf{a}_1, \cdots, \mathbf{a}_r]$, and then project any sample $\mathbf{x} \in \mathbb{R}^d$ into an $r$-dimensional Euclidean space with the new vector $(A^{RAP})^T\mathbf{x}$.
```

# 4. EXPERIMENTS

Interactive image search or relevance feedback is the process which helps a user highlight his search intention and target difficult concepts. This process often consists of partially labeling a very small fraction of an image database and iteratively refining some learning model using both labeled and unlabeled data. Training this kind of learning models is referred to as semi-supervised learning. In this section, we evaluate several semi-supervised subspace learning models with relevance feedback.

## 4.1 Features

The experiments were conducted on a large database with 10,000 generic images from the Corel gallery [12]. These images were pre-grouped to 100 categories by high-level semantics (defined by a large group of human observers as standard groundtruths), such as autumn, balloon, bird, dog, eagle, sunset, tiger, etc. Each category contains 100 images and represents a semantic concept. Some images sampled from category 1, 16, 33, and 49 are shown in Figure 2. Two types of color features and two types of texture features are used in the experiments, which are: the nine-dimensional color moments in LUV color space (the first three-order moments) and the 64-dimensional color histogram in HSV color space; the ten-dimensional coarseness vector and the eight-dimensional directionality. Therefore, we obtain a 91-dimensional data vector for each image.

## 4.2 Relevance Feedback Scheme

We illustrate an automatic feedback scheme to simulate a CBIR system. During each query session, the user looks for images carrying the same concept with the query image. Provided each submitted query, the CBIR system uses some distance metric to rank the images in the database and conducts five feedback iterations. At each feedback iteration, the top-10 ranked images among the returned ones, which have not been labeled in previous feedback iterations, are labeled as the feedback images. Their label information, relevant or irrelevant to the query in semantics, is employed for re-ranking. Note that the images which have been selected at previous iterations are excluded in later iterations.

## 4.3 Image Retrieval Performance

The statistical average top-$N$ precision is used for performance measurement. $N$ is referred to as the scope of which top-ranked images will be returned to the user. The precision is the ratio of the number of relevant images presented to the user to the scope $N$. We use the precision-scope curves [12] to testify the effectiveness of subspace-based image retrieval approaches. This kind of curves capture the precision with various scopes and hence take on an overall performance evaluation. We also use the precision-iteration curves to describe the precision dynamics with feedback iteration enumerated from 1 to 5.

To start relevance feedback, the Euclidean distance metric in the original 91-dimensional space is used to rank images for the first time. After the user labels the top-10 returned images, the first feedback iteration begins with applying RAP, SR, ARE, SSP, and PCA, respectively. Form the distance metrics $AA^T$ with subspaces $A$ learned by the five algorithms, and then apply them to re-rank images.

To sustain fair comparisons, we construct the same $k$-NN graph $G$ for RAP, SR, ARE, and SSP where $k$ is fixed to 6. Set the parameters $\alpha = 2$, $\beta = 10^{-6}$, and $\gamma = 0.01$ in ARE, SR, and RAP, respectively. The dimensions of ARE and SSP subspaces change with the size of the labeled set, while that of SR subspace is always 2. Both PCA and RAP use 76-dimensional subspaces since PCA contributes the initial solution to RAP.

Figure 3 shows the precision-iteration curves as well as the precision-scope curves for five subspace algorithms. Particularly, figure 4 shows the precision-scope curves corresponding to four concepts. We find that the proposed RAP algorithm consistently outperforms the other four algorithms on the entire scope and at all feedback iterations, and that ARE exhibits second best at the 5th feedback iteration. In summary, these quantitative comparisons show a clear and consistent gain of our subspace learning algorithm with respect to the state-of-the-art subspace learning algorithms in CBIR, especially over a small scope.

[1] Let $\{\mathbf{x}_i\}_{i=1}^n$ be zero-centered, i.e., $\sum_{i=1}^n \mathbf{x}_i = 0$.

**Figure 2: Corel image examples. (a) Concept 1: Antelope; (b) Concept 16: Car; (c) Concept 33: EasterEgg; (d) Concept 49: Jewelry.**

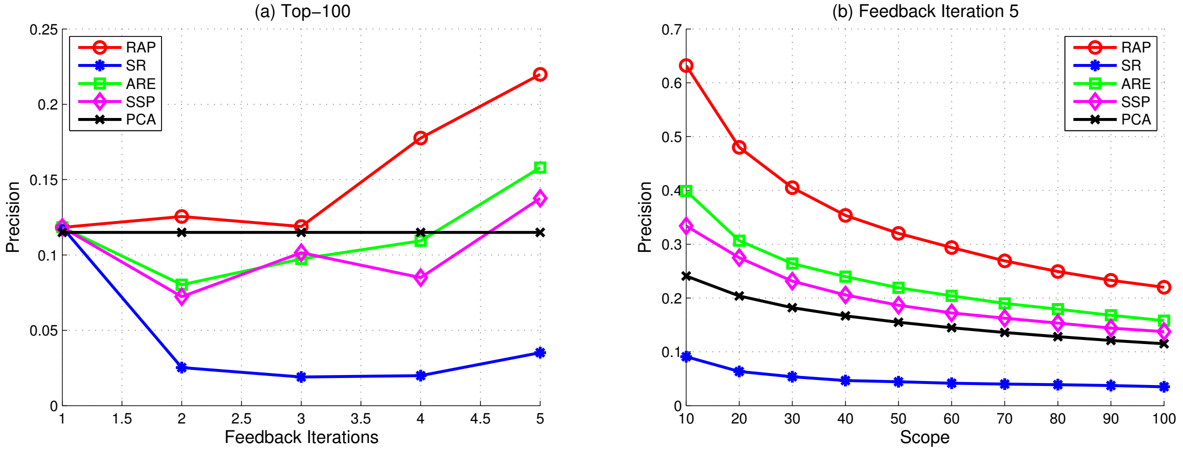

**Figure 3: Comparisons of the retrieval performance of five subspace methods. (a) The precision-iteration curves of RAP, SR, ARE, SSP, and PCA over top-100 retrieved images (scope=100); (b) the precision-scope curves of RAP, SR, ARE, SSP, and PCA at feedback iteration 5.**

## 5. CONCLUSIONS

In this paper, a new subspace learning technique, Relevance Aggregation Projections (RAP), for content-based image retrieval is proposed. It exploits the labeled images provided in the user's relevance feedback to learn a semantic subspace within which the positive (relevant) samples collapse to a single point while the negative (irrelevant) samples are pushed outward with a large margin. Our technique falls in the general category of imbalanced semi-supervised learning, and can be conveniently implemented in interactive image search systems. Experimental results on the COREL image database demonstrate the proposed method can achieve a significantly higher precision for image retrieval than the stat-of-the-art subspace learning algorithms even with few feedback iterations.

## 6. ACKNOWLEDGMENTS

This work has been supported by the U.S. Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government.

## 7. REFERENCES

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

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

[3] D. Cai, X. He, and J. Han. Spectral regression: A unified subspace learning framework for content-based image retrieval. In Proc. *ACM Conference on Multimedia*, 2007.

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

[5] H.-T. Chen, H.-W. Chang, and T.-L. Liu. Local discriminant embedding and its variants. In Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[6] F. Chung. Spectral graph theory. In *CBMS Regional Conference Series in Mathematics, American Mathematical Society*, number 92, 1997.

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

[8] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA, 2004.

[9] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[10] C. H. Hoi, M. R. Lyu, and R. Jin. A unified log-based

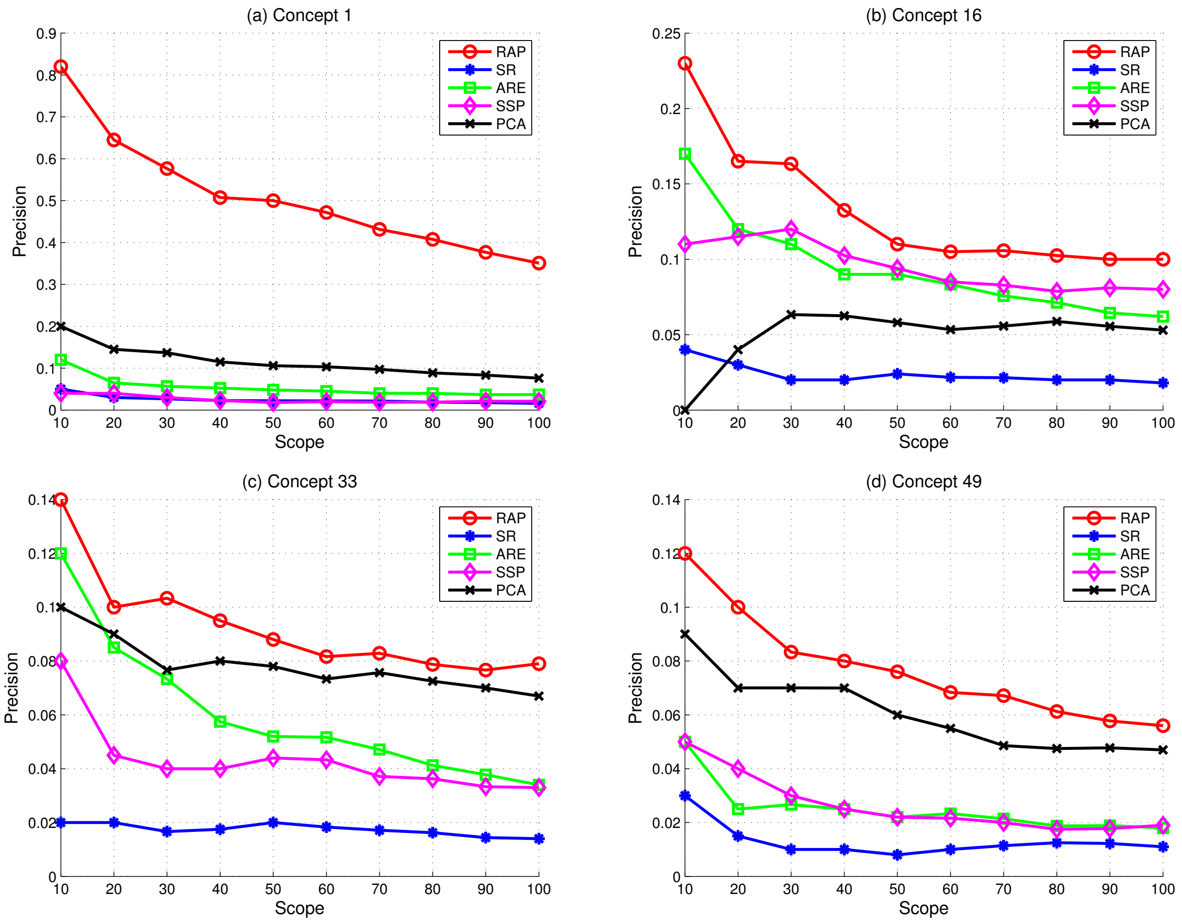

**Figure 4: The precision-scope curves corresponding to four concepts at feedback iteration 5. (a) Concept 1; (b) Concept 16; (c) Concept 33; (d) Concept 49.**

relevance feedback scheme for image retrieval. *IEEE Trans. on Knowledge and Data Engineering*, 18(4):509–524, 2006.

[11] C. H. Hoi, W. Liu, M. R. Lyu, and W. Y. Ma. Learning distance metrics with contextual constraints for image retrieval. In Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[12] W. Jiang, G. Er, Q. Dai, and J. Gu. Similarity-based online feature selection in content-based image retrieval. *IEEE Trans. on Image Processing*, 15(3):702–711, 2006.

[13] Y.-Y. Lin, T.-L. Liu, and H.-T. Chen. Semantic manifold learning for image retrieval. In Proc. *ACM Conference on Multimedia*, 2005.

[14] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[15] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A powerful tool in interactive content-based image retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[16] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[17] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the ealy years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[18] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[19] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In Proc. *ACM Conference on Multimedia*, 2001.

[20] J. Yu and Q. Tian. Learning image manifolds by semantic subspace projection. In Proc. *ACM Conference on Multimedia*, 2006.

[21] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, MIT Press, Cambridge, MA, 2004.

[22] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In Proc. *International Conference on Machine Learning*, 2003.