# TOWARD INTELLIGENT USE OF SEMANTIC INFORMATION ON SUBSPACE DISCOVERY FOR IMAGE RETRIEVAL

*Jie Yu*
Department of Computer Science
University of Texas at San Antonio

*Qi Tian*
Department of Computer Science
University of Texas at San Antonio

## ABSTRACT

Image retrieval has been widely used in many fields of science and engineering. The semantic concept of user interest is obtained by a learning process. Traditional techniques often assume the images are from certain distribution and all images from the same class are visually similar. Our study shows that those assumptions are inappropriate in many cases. To solve this problem we model the images as lying on non-linear subspaces embedded in the high dimensional space. We also find that a set of low-level feature subspaces may correspond to one high-level semantic concept. Unlike most unsupervised subspace learning techniques, we propose to intelligently use the semantic similarity and dissimilarity information provided by user in discovering the discriminant structure of image subspaces in respect to classification. Theoretical study shows that our methods converge to Linear Discriminant Analysis if certain criteria are met. Extensive experiments are designed to evaluate the performance of our method and compare it to other state-of-the-art techniques. The results show the superior performance of our proposed method.

## 1. INTRODUCTION

With the development of digital imaging technology, more and more information is conveyed in the form of digital images or video clips. The rich context of an image makes the understanding of its semantic meaning very difficult.

Image retrieval aims at automatically retrieving the images of user interest from large databases based on their visual content. The user interest could be summarized by a high-level semantic concept while the visual content of the images could be represented by low-level features such as color, texture and shape. Machine learning techniques are used to bridge the gap between the semantic concept and image features. During this process the user gives semantic information on a few sample images by labeling them. Then a statistical model of the images with same semantic meaning could be estimated based on the labeled training set.

Because the dimensionality of feature space is usually very high, ranging from tens to hundreds, direct model estimation in the high dimensional feature space can fail easily. Dimension reduction is used to map the original space to a low dimensional space. The concept learning is conducted in that projected space. During that process a projection that facilitates the learning is difficult to obtain when few samples are used for training purposes. However, in most image retrieval applications labeling the images requires human effort involved and is computational inefficient. Thus the small training set problem along with the high dimensionality problem become two major challenges for image retrieval.

Traditional techniques, such as Principal Component Analysis [1] and Linear Discriminant Analysis [2], assume image data are from certain distribution model (in most cases a Gaussian Mixture). Recently more and more attention has been drawn on modeling the data as lying on a subspace which is embedded in high dimensional space. The intrinsic structure of the subspace could be discovered and preserved in a low dimensional space by using subspace learning techniques. Because the global structure of the subspaces could be inferred from the local neighborhood information, no assumption on the data distribution is needed.

## 2. SEMANTIC CLASS AND GEOMETRIC SUBCLASS IN IMAGE RETRIEVAL

### 2.1. Semantic Class and Geometric Subclass

As we mentioned in the introduction, in image retrieval the images are often represented by feature vectors, which correspond to visual content such as color, texture and shape. Although the subject of a single image could be complex and ambiguous, a set of images that share certain visual similarity could correspond to a geometric subclass in the feature space. The set of images therefore may correspond to a simple semantic subject, e.g. apple or banana. In that case capturing the structure of that geometric subclass is sufficient to learn the semantic subject of the images.

For a particular image retrieval application, the user's interest, a high-level semantic concept, is the prime target to learn by a classifier. Although in some cases the learning is conducted on the visually similar images, we find the human judgment on the semantic subject of an image may be from non-visual knowledge and consequently the one-to-one relation between geometric subclass and semantic class doesn't exist. For instance, one may be interested in fruit images which contain images of bananas and apples. Visually those images may be quite different and share little geometric correlation between data from the two subclasses. The fact that one semantic class may contain multiple geometric subclasses introduces a more challenging problem. If we are given enough training samples, a novel classifier is expected to learn that the user is interested in images from certain subclasses and, without supervision, generalize the query concept.

### 2.2. Similarity and Dissimilarity

The visual resemblance between images can be defined as similarity and dissimilarity in the geometric space. However for an image query the user's judgment on the semantic similarity and dissimilarity between images may not be solely based the geometric information. Because of the many-to-one relation

1-4244-0367-7/06/$20.00 ©2006 IEEE

between geometric subclass and semantic class, we conclude two guidelines for intelligent use of the semantic information user provides:

1) It is safe to claim two images are visually dissimilar if the user labels them as semantically irrelevant to each other because the user must make the judgment based on some visual difference between the images.

2) If the user assigns two images to same semantic category, it doesn't suggest that they must be visually similar because that decision may be based on some non-visual prior knowledge.

From the above two observations we can find the semantic dissimilarity information is more important in learning complex concepts while the semantic similarity may not correspond to the visual similarity. It is desirable to use both semantic similarity and dissimilarity information for self-discovery of the geometric subclasses.

### 2.3. Related Work in Subspace Learning

To facilitate data exploration, researchers have been trying to capture the structure of a correlated data group by mapping the original high-dimensional space to a low-dimensional one. Recent advances in subspace learning have drawn more and more attention. Instead of assuming data are from a particular distribution, they could be modeled as lying on a non-linear low-dimensional subspace embedded in the high-dimensional space. The global structure of such a subspace could be inferred by gathering local information of every neighborhood on that subspace. Local Linear Embedding (LLE) [3] assumes the local patch can be approximated by a hyperplane and the linear correlation between local neighbors should be preserved in the projected space. ISOMAP [4] proposes to use geodesic distance to substitute Euclidean distance to reflect the non-linear structure of the subspace. Locality Preserving Projection (LPP) [5] treats the neighborhood as a cluster and tries to find an optimal projection that makes neighbor data close to each other. Those subspace learning techniques are unsupervised while supervised or semi-supervised approaches are further proposed because they are more desirable in image retrieval. Supervised LLE (S-LLE) [6] incorporate user provided information by tuning a parameter $\alpha$ to control the influence of semantic labeling on the geometric structure learnt. Incremental Semi-supervised LPP (I-LPP) [7] tries to use user feedback as a semantic relation to add to or override the geometric neighbor relation. All these techniques try to cluster data but ignoring the relation between semantic concept and geometric subspace as discussed in Section 2.1. Furthermore, none of them use dissimilarity information to make samples from different semantic classes separated in the projected space. Based on the above discussion, we propose a novel subspace discovery technique that could use semantic information intelligently.

## 3. SUBSPACE DISCOVERY FOR CLASSIFICATION

### 3.1. Subspace Discovery for Classification

As we indicate in Section 2, there is a gap between geometric structure and semantic similarity. To bridge this gap, we propose a novel supervised method called subspace-discovery for classification (*SDC*). It could makes use of user provided semantic information and captures the structure of each geometric subspace. A brief introduction of the technique is as follows:

Because the semantic dissimilarity between images must arise from visual differences, those images from different semantic classes should be as separated as possible in the new space. The pair-wise Semantic Dissimilarity information between two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be stored in a matrix *SD* as in equation (1). It uses supervised information to refine the geometric structure discovery.

$$SD_{ij} = \begin{cases} 1 \text{ if } \boldsymbol{x}_{\mathrm{i}}, \boldsymbol{x}_{\mathrm{j}} \notin \text{same semantic class} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

Instead of assuming that each semantic class contains only one subspace, we try to discover the global structure of the data by preserving locality information in a projected space. The local geometric information is first captured by constructing a *K* Nearest Neighbor graph:

$$GeoSim_{ij} = \begin{cases} 1 \text{ if } \boldsymbol{x}_{\mathrm{i}}, \boldsymbol{x}_{\mathrm{j}} \text{within } e\text{ach other's neighbourhood} \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Considering that some visually similar, that is, geometrically close, image data may be from semantically different classes, we incorporate the semantic information into the geometric structure discovered and get a new Geometric-Semantic Similarity Matrix:

$$GSSim_{ij} = \overline{SD_{ij}} \; AND \; GeoSim_{ij} \tag{3}$$

Because each neighborhood may contain a different number of samples, the Geometric-Semantic Similarity Matrix is normalized for each row.

$$GSSim_{ij} = GSSim_{ij} / \sum_i GSSim_{ij} \tag{4}$$

In the above matrix, the semantic dissimilarity is considered for classification purpose and geometric neighborhood information, along with semantic similarity information, is used to discover image subspaces.

Suppose a projection *W* map any sample $\boldsymbol{x}_i$ in original space to a corresponding sample $\boldsymbol{y}_i$ in a lower dimension space.

$$\boldsymbol{y}_i = W \cdot \boldsymbol{x}_i \tag{5}$$

Obviously in the local neighborhood, the mean can be estimated as:

$$m_i = \sum_j \boldsymbol{x}_j GSSim_{ij} \tag{6}$$

And the projected mean can be calculated as follows:

$$m_i^{(y)} = \sum_j \boldsymbol{y}_j GSSim_{ij} \tag{7}$$

Intuitively we propose a new cost function for finding the optimal projection:

$$\begin{aligned} &\max \frac{|\sum (m_i^{(y)} - m_j^{(y)})(m_i^{(y)} - m_j^{(y)})^T SD_{ij}|}{|\sum (y_i - y_j)(y_i - y_j)^T GSSim_{ij}|} \\ &= \max \frac{|\sum W(m_i - m_j)(m_i - m_j)^T SD_{ij} W^T|}{|\sum W(x_i - x_j)(x_i - x_j)^T GSSim_{ij} W^T|} \end{aligned} \tag{8}$$

In the above cost function the numerator corresponds to the separation between neighbors from different semantic classes. The different semantic classes are more far away from each other when the numerator gets larger. Note that we don't use distances between class means for numerator as in LDA because there may be more than one subclass clusters in one semantic class. The denominator describes the preserving of global structure by clustering the samples within the same neighbor, which is important for data modeling and classification in the projected space. Note that by using *GSSim*, only samples that are both semantic and geometric similar are clustered together unlike in traditional approaches where all the samples from same semantic classes are forced together. That may result in better discovery of the global class structure because all subclasses' structure is preserved. By maximizing the ratio of the numerator and denominator we can find an optimal projection that makes the subspaces from different semantic classes more separate from each other and the samples within same neighborhood clustered together.

We denote semantic Dissimilarity Scatter Matrix as follows:

$$S_{Diss} = \sum_{i,j} (m_i - m_j)(m_i - m_j)^T SD_{ij} \tag{9}$$

The Geometric-Semantic Similarity Scatter Matrix can be defined as:

$$S_{GS-Sim} = \sum_{i,j} (x_i - x_j)(x_i - x_j)^T GSSim_{ij} \tag{10}$$

Consequently the optimal projection can be obtained by solving the following optimization problem:

$$W = \arg\max_W \frac{|WS_{Diss}W^T|}{|WS_{GS-Sim}W^T|}$$
$$\Rightarrow W = \underset{\max}{eig}(S_{Diss}S_{GS-Sim}^{-1}) \tag{11}$$

The optimal projection $W$ consists of the eigenvectors corresponding to the largest eigenvalues of $S_{Diss}S_{GS-Sim}^{-1}$.

Compared to other subspace learning techniques such as LLE, LPP and their supervised version, our method is novel in that: i) we consider not only preserving the subspaces' structure after the projection but also separating samples from different classes. For a classification task the latter obviously can't be neglected. ii) Our method uses the semantic similarity and dissimilarity intelligently to refine the local geometric structure while the full supervised S-LLE and I-LPP tries to use semantic similarity information to substitute geometric structure. As we analyze in Section 2 when one semantic class corresponds to multiple feature subspaces, the semantic similarity can't guarantee similarity in low-level features. Thus clustering all samples from the same semantic class is unnecessary and easy to fail. However that problem doesn't exist for our method because our method tries to capture the structure of all subspaces from all classes. This is accomplished by clustering only samples within a geometric local neighborhood from the same class and consequently no effort will be put to cluster samples from different subspaces.

### 3.2. Relation to LDA

Linear discriminant analysis has been widely used in image retrieval applications. It assumes that one subspace from a Gaussian distribution corresponds to one semantic concept. That can be generalized as a special case of our proposed method when the neighborhood is defined as large enough to cover all images from the same semantic class. If we have totally $N$ training samples, that condition can be satisfied when $K \geq N$.

While the Semantic Dissimilarity matrix *Diss* is the same, geometric neighborhood information is omitted in that

$$GeoSim_{ij} = 1 \text{ for any pair of images} \tag{12}$$

Since $GSSim_{ij} = \overline{Diss_{ij}}\ AND\ GeoSim_{ij}$, we have

$$GSSim_{ij} = \begin{cases} \dfrac{1}{n_l} \text{if } \boldsymbol{x}_\text{i}, \boldsymbol{x}_\text{j} \in \text{semantic class l} \\ 0 \text{ otherwise} \end{cases} \tag{13}$$

Instead of using semantic information together with geometric information for subspace discovery, here the semantic information overrides the geometric information. Consequently for each sample, its neighborhood mean $m_i$ becomes its class mean $m_{L(i)}^c$.

$$m_i = \sum_j \boldsymbol{x}_j GSSim_{ij} = \sum_{L(j)=L(i)} \boldsymbol{x}_j / n_{L(i)} = m_{L(i)}^C \tag{14}$$

We denote the label of sample $\boldsymbol{x}_i$ as $L(i)$ and the number of samples in class *l* as $n_l$.

Thus the Dissimilarity Scatter Matrix converges to Between-Class Scatter Matrix. It can be proven that (note: we have to omit the proof here due to the limited space).

$$S_{Diss} = 2N \cdot S_{Diss} \tag{15}$$

Similarly we can prove $S_{GS-Sim} = \frac{1}{2}S_W$ (for the same reason, we omit the proof here), when the geometric neighborhood is overridden by semantic class.

Thus we have

$$W_{LDA} = \arg\max_W \frac{|WS_{Diss}W^T|}{|WS_{GS-Sim}W^T|} = \arg\max_W \frac{|WS_BW^T|}{|WS_WW^T|} \tag{16}$$

From the above equation we can conclude that, LDA could be a special case of a subspace discovery method where each class is assumed to contain only one subspace. According to our discussion in Section 2, that assumption is inappropriate in many cases. Compared to LDA, our method could handle more complex situation where multiple subclasses exist in one semantic class. Besides, our SDC can capture the structure of the subspaces better because geometric neighborhoods can be preserved while only semantic similarity is considered in LDA. Thus our proposed method could model the real scenario more accurately.

## 4. EXPERIMENTS AND ANALYSIS

### 4.1. Test on Benchmark Datasets

The first experiment is designed to evaluate the performance of our proposed method on some benchmark datasets. Different setting of neighborhood size *(K)* is tested to find the optimal setting. The benchmark datasets tested are the heart and breast-cancer (B.C.) datasets from UCI repository. In all the experiments we run the program 20 times to get average performance.

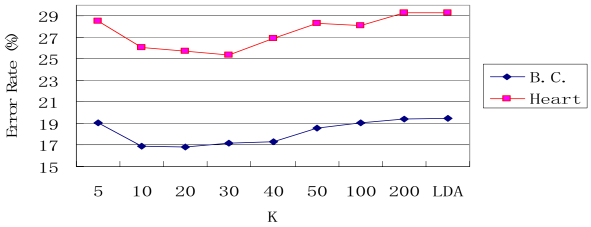

**Figure 1. Test Neighborhood size on Benchmark Dataset**

From the result in above table we can find the size of neighborhood does have an influence on the classification performance. However we find it is stable between 10 to 30 and the smallest error rates lie in that range. This may be used as guideline for the future settings. It is clear that our proposed method performs better than LDA in most settings and will converge to LDA when the neighborhood size gets larger and contains all the training samples.

### 4.2. Comparison to State-of-the-Art

In this experiment, we test the performance of our proposed method and compare it to the state-of-the-art techniques discussed in Section 2.3: LDA, LLE, S-LLE, LPP and I-LPP. ISOMAP is not tested because it's computational expensive to calculate the global geodesic distance. *K* is set to 20 for our proposed method according to the previous experiment. Best performance for S-LLE is listed by searching for optimal α from 0 to 1 with step size 0.01. For simplicity, nearest neighbor classifier is used in the projected space.

We first apply those methods on face identification on three popular databases: Harvard, ATT and UMIST facial image databases [8]. Harvard Face Image Database consists of grayscale images of 10 persons. Each person has totally 66 images which were classified into 10 sets. The ATT Face Image Database consists of 400 images for 10 persons. The UMIST Face Database consists of 564 images of 20 people. In all three experiments 67% of the images are randomly picked up as the training sample and the rest is used for testing.

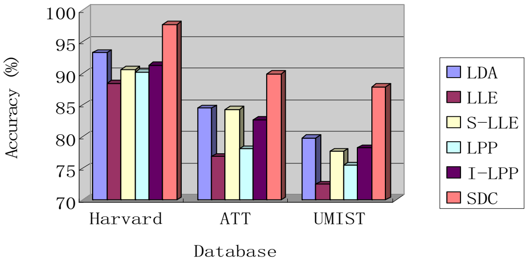

**Figure 2. Comparison to the State-of-the-Art**

From the results in Figure 2 we find that our proposed method performs best on all three databases. Because the facial image of one person could be well modeled by a subspace, this problem of face identification can be considered as a learning semantic concept from multiple image subspaces.

### 4.3. Test on Image Classification

In the final experiment, we test our proposed method and the state-of-the-art techniques on a real image classification application. The dataset used is the COREL image database. It contains color images which are roughly categorized into 10 classes. For simplicity we randomly pick up two classes of images to conduct the training and testing. Only supervised techniques are compared since they are more suitable for classification. The performance of these techniques is illustrated in *Precision-Recall* graph.

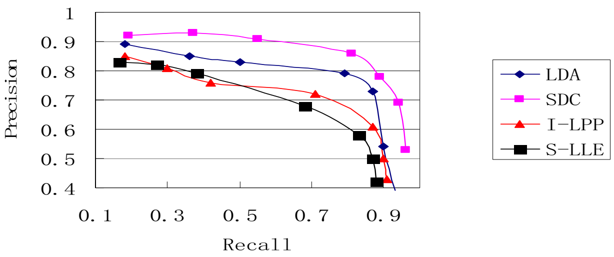

**Figure 3. Precision-Recall graph for test on COREL database**

From Figure 3 we can conclude that the new technique outperforms other state-of-the-art techniques in this experiment. Considering the rich background of images in COREL database, it shows that our method not only discovers the subspaces within semantic class but also capture the most discriminant features of the images to facilitate classification.

## 5. CONCLUSIONS AND FUTURE WORK

The mapping between semantic concept and geometric subclasses is the main object of the learning process in image retrieval. Traditional approaches assume image data are from certain distribution model and correspond to one semantic class. Due to the rich visual content of images, it is more appropriate to model the images as lying on a non-linear subspace embedded in the feature space. Besides, we find that user's semantic interest may arise from non-visual knowledge and consequently one semantic class may contain multiple image subspaces which have different geometric structures. In that sense dissimilarity relation between images is more important for classification purpose while the geometric structure of the subspaces can only be captured by self-discovery method. Based on above analysis we propose a new technique that infers the global structure of subspaces from neighborhood information and intelligently use semantic information to find the optimal projection that facilitates classification. Theoretical analysis shows that our proposed method converges to well-known LDA when one semantic class corresponds to one geometric subspace. Experiments are conducted on benchmark datasets and two image retrieval applications. Our proposed method outperforms other state-of-the-art techniques in these tests.

This research work will be continued in two directions: 1) selecting unlabeled data that is useful for classification based on subspace structure and 2) connecting local neighborhood into meaningful cluster to avoid tuning parameter *K*.

**Acknowledgement**: This work was supported in part by the Army Research Office (ARO) grant under W911NF-05-1-0404, and by the Center of Infrastructure Assurance and Security (CIAS), the University of Texas at San Antonio. We thank Mark Doderer for polishing the paper.

## 6. REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*. 2nd edition, New-York: Springer-Verlag, 2002.

[2] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., 2001.

[3] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, v.290, 2000

[4] J. B. Tenenbaum, V. de Silva and J. C. Langford. "A global geometric framework for nonlinear dimensionality reduction," *Science*, 2000.

[5] X. He and P. Niyogi, "Locality Preserving Projections," *Neural Information Processing Systems 16 (NIPS'2003)*.

[6] D. de Ritter *et al*. "Supervised locally linear embedding," *Proc. Of ICANN/ICONIP*, 2003

[7] X. He, "Incremental semi-supervised subspace learning for image retrieval," *Proc. of ACM Multimedia*, 2004

[8] M. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. PAMI*, vol. 24, 2002