# CONTENT BASED IMAGE RETRIEVAL USING SEMANTIC GROUPING AND LINEAR DISCRIMINANT ANALYSIS

Xiaojun Qi, Robert Qualls
NSF REU at Utah State University

**ABSTRACT**

The biggest problem in content-based image retrieval (CBIR) is bridging the gap between low-level features and high-level semantics. Relevance feedback is a popular method for incorporating semantic information in CBIR. This paper discusses an improvement upon a relevance feedback approach that utilizes semantic grouping and linear discriminant analysis to close the gap between low-level features and high-level semantics. Specifically, the prior system is improved by incorporating the images in the same group as the query image in the collection of retrieved images. The experimental results show that this significantly improves the performance of the second and later retrievals in each session.

***Index Terms—*** content-based image retrieval, computer vision, linear discriminant analysis, semantic grouping

## 1. INTRODUCTION

Recent years have seen a vast increase in the amount of visual information available to personal computer users. Like other information, it is logical and preferable to organize this information so that it is categorized or sorted for personal or commercial use. One example of such organization is the ability to find images in a data set that are similar to a query. While it is technically possible to have users contribute tag information to images, as the number of images in the database increases, it becomes much more of a burden on the users. Ultimately, it makes more sense to have an automated system that can find and return similar images based on the content of a query image, freeing users to do other things.

Low-level features have been proposed as a means of determining the content of images. These include descriptors such as color, shape, and texture [2, 3, 4]. In practice, however, low-level features alone do not supply enough information to describe the high-level semantic concepts of an image. This is in part caused by how differently a user and a computer perceive an image. For example, a user could more easily describe the structures, animals, or atmosphere of an image rather than color, shape, or texture.

Various methods have been utilized to bridge the semantic gap. One of the more popular groups of methods is relevance feedback. Relevance feedback involves getting feedback from the user as to whether an image returned by the system is relevant to the query image. This feedback is then fed into an algorithm which utilizes the user information to improve later retrieval results. Depending on how long user preferences stay in the system, a relevance feedback approach may be called short-term or long-term.

The relevance feedback method discussed in this paper uses a semantic grouping algorithm to group similar images in semantic groups and apply a classification method called linear discriminant analysis (LDA) to improve retrievals. Semantic grouping uses user feedback to place relevant images together in groups while making sure that these groups do not contain irrelevant images. The LDA classification technique maximizes the ratio of between-group variance to within-group variance. In this approach, LDA is used to modify the descriptor matrix of the data set such that semantically similar images have smaller Euclidean distances than semantically dissimilar images.

The shortcoming of the prior relevance feedback system is that it does not consider the images in the query's group as candidates. In theory, since the images are in the same group as the query, they are semantically similar to the query based on the users' past decisions. In this paper, we propose to use information gathered from semantic groups to facilitate retrieval results. The remainder of this paper is organized as follows: Section 2 presents and overview of relevance feedback. Section 3 discusses semantic grouping. Section 4 describes linear discriminant analysis. Section 5 provides the Null-Space method for LDA. Section 6 covers the experimental process. Section 7 shows the results obtained from the project. Section 8 discusses the conclusions made. Last, section 9 lists works cited.

## 2. RELATED WORKS

Zhang et al. discussed another relevance feedback system that used a memorization technique [1]. Bing et al. described a system that used LDA for face retrieval [6]. Yoshizawa and Schweitzer developed the system this paper is based on [4].

## 3. PROPOSED APPROACH

### 3.1. Relevance Feedback

When a query image is submitted to the CBIR system, the system returns a specified number of images to the user. The user then marks each returned image as relevant or irrelevant based on his or her semantic understanding of the similarity between the query image and returned image. For this approach, it is assumed that all images in the entire data set are represented in terms of a fixed set of low level features. These low-level features are stored in a vector of real values for each image. The relevance feedback process occurs in the following steps [4].

1. Obtain the low-level feature vector for the query image.
2. Compute the Euclidean distance between the query image feature vector and the rest of the image feature vectors in the data set.
3. If the query image is not already assigned to a semantic group, take the top *k* smallest distances and use them as the candidates for the current retrieval. If the query image is already in a group, include the rest of the images in the group in the candidates. Then use as many of the remaining smallest distance images until there are *k* candidates. The latter was not taken into consideration in the original approach.

### 3.2. Semantic Grouping

Once the candidates for the current retrieval are generated, the user can mark each one as relevant or irrelevant (*positive* or *negative*). This is the user's relevance feedback. Next, the candidates are assigned to groups in the history through the semantic grouping algorithm. The algorithm can be interpreted as actions that occur in four possible cases [4].

Case 1: If any of the negative images has not been assigned to a group, assign it to a new group.

Case 2: If *all* of the positive images have not been assigned to any groups, add all of the positives together to a new semantic group.

Case 3: If some of the positive images are assigned to some of the groups, and no negative images are in those groups, assign all of the positives and the images in their groups to a new semantic group.

Case 4: If some of the positive images are assigned to some of the groups, and some negative images are in those groups, assign all of the positives to a new semantic group, and remove the positives from the groups containing negatives.

### 3.3. Generating a New Metric With Multi-Class LDA

Once grouping information has been obtained, the system can start generating a new distance metric via LDA. There are two approaches to LDA: a two-class approach and a multi-class approach.

In the two-class approach, all positive images are considered as one class, and all negative images are considered as another class. The goal is to find a distance metric which provides good separation between the two classes. The downside to this approach is that the negative images may contain many dissimilar semantic concepts, so it does not make any sense to consider them all as one class, and it reduces the effectiveness of LDA.

In the multi-class approach, each negative image is considered a class by itself. As in the two-class approach, LDA will attempt to maximize the distance between all of the classes. The downside to this approach is that some negative images may actually belong to the same semantic category; however, they are grouped into the same semantic group based on the user's relevance feedback history, so considering them as different classes reduces the effectiveness of LDA. The two-class approach was not tested for this paper, but the authors of the original paper claimed that the multi-class approach provided better performance.

### 3.4. Null-Space LDA

The LDA computation approach used in the paper was the Null Space approach. "The idea is to project the between-class scatter into the null space of the within-class scatter, and then find the eigenvectors corresponding to the largest eigenvalues of the transferred between-class scatter." [4]. The algorithm for the Null Space method is described below [4].

1. Compute the eigenvector factorization $S_w = \tilde{H}\Lambda\tilde{H}'$
2. Discard the eigenvectors corresponding to the positive eigenvalues in $\Lambda$. The remaining eigenvectors span the null-space of $S_w$. The matrix of remaining eigenvalues is denoted by $H$. It satisfies $\| H'S_w H \| = 0$
3. Project the between-class scatter into the within-class scatter: $\tilde{S}_b = H'S_b H$ .

4. Compute the eigenvector factorization: $\tilde{S}_b = \tilde{U}\Sigma\tilde{U}$
5. Select the k eigenvectors in $\tilde{U}$ corresponding to the largest eigenvalues in $\Sigma$. Their matrix is denoted by $U$.
6. Set $W = HU$.

## 3. EXPERIMENTAL PROCESS

The same experimental parameters in the original paper were used for this project. The data set consisted of 2000 flickr images. For each image, two retrievals were performed for each session (query image), and each retrieval returned 30 images. In addition, another experimental trial was performed in which there were three retrievals per session. The experimental process is described below [4].

1. A query image is randomly selected from the database
2. If it is the first retrieval of the session, the original distance metric is used. If it is the second or later retrieval, the distance metric computed in the previous retrieval is used.
3. Using the distance metric, the 30 smallest distance images are returned to the user.
4. The user tells the system which images are positive or negative.
5. The system updates the semantic group history
6. Using the updated history, the system computes a new distance metric via LDA.

## 4. EXPERIMENTAL RESULTS

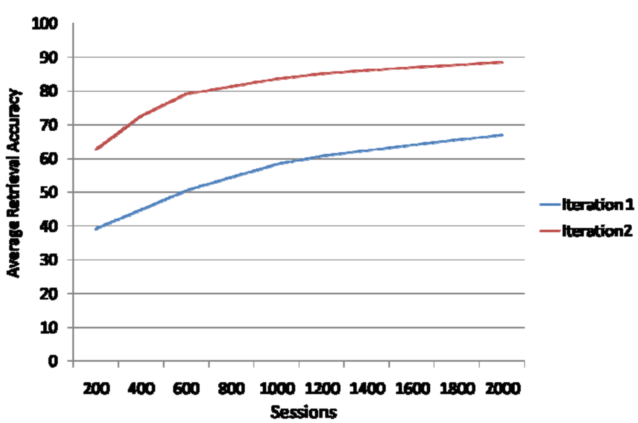


Figure 1: Average retrieval accuracy for two retrievals through 2000 sessions.

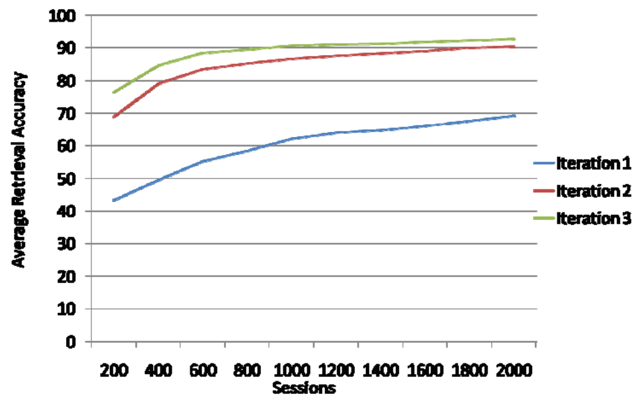


Figure 2: Average retrieval accuracy for three retrievals through 2000 sessions.

## 4. CONCLUSIONS AND FUTURE WORK

The modified approach has better performing retrievals after the first retrieval each session. In the original approach, the difference between the first and second retrieval was negligible. In the future, the focus should be on increasing the accuracy of the first retrieval as it still seems to have room for improvement.

## 5. REFERENCES


[1] Lei Zhang, Fang Qian, Mingiing Li, Hong-Jiang Zhang. An e±cient memorization scheme for relevance feedback in image retrieval. Multimedia and Expo, International Conference on, 2003.

[2] Smith, J. R. and Chang S-F. Transform features for texture classi¯cation and discrimination in large image databases. In Proceedings of IEEE Intl. Conf. on Image Processing, 1994.

[3] Stricker, M. and Orengo, M. Similarity of color images. In Proceedings of SPIE, 1995.

[4] T. Yoshizawa and H. Schweitzer. Long-term learning of semantic grouping from relevance-feedback. In Proc of the ACM SIGMM Intl workshop on Multimedia Information Retrieval, pages 165{172, New York, NY, USA, 2004. ACM.

[5] Zhou, X. S. and Huang, T. S. Edge-based structural feature for content-base image retrieval. Pattern Recognition Letters, Special issue on Image and Video Indexing, 2000.

[6] Yu Bing, Jin Lianfu, Chen Ping. A new lda-based method for face recognition. Pattern Recognition, 2002.