SEMANTIC CLUSTERS BASED MANIFOLD RANKING FOR IMAGE RETRIEVAL

Ran Chang and Xiaojun Qi

ran.c@aggiemail.usu.edu and Xiaojun.Qi@usu.edu
Computer Science Department, Utah State University, Logan, UT 84322-4205

ABSTRACT

We propose a novel weighted manifold-ranking based image retrieval method to improve the effectiveness of traditional manifold methods. Specifically, we apply the SVM-based relevance feedback technique to create semantic clusters for computing the reliability score of each database image. We then incorporate the reliability scores into the affinity matrix to construct a weighted manifold structure. We finally create an asymmetric relevance vector to store users' positively and negatively labeled information. Our system ensures to propagate the labels in the relevance vector to the images with high reliability scores and discriminately spread the ranking scores of positive and negative images via the weighted manifold structure. Extensive experiments demonstrate our system outperforms the other manifold systems and SVM-based systems in the context of both correct and erroneous feedback.

Index Terms— Content based image retrieval (CBIR), backbone image, semantic clusters, weighted manifold

1. INTRODUCTION

A successful CBIR system should bridge the semantic gap between low-level visual features and high-level semantic concepts. Relevance feedback (RF) techniques [1] are online learning techniques that have been widely used in CBIR to refine the user query through interactive sessions. They are generally considered as promising approaches to bridge the semantic gap and improve the retrieval performance. However, most research focused on the following: 1) Apply learning schemes to tune query using a limited number of labeled images in a single retrieval session. 2) Apply distance functions [2] to measure the perceptual similarity between images. As the emergence of large-scale databases, it is paramount to efficiently use the information of the entire database instead of the labeled images to facilitate learning. This can be achieved by transductive learning to explore the relationships of all database images and propagate the ranking scores of labeled images to unlabeled images via a weighted graph. Here, we review representative transductive learning in CBIR.

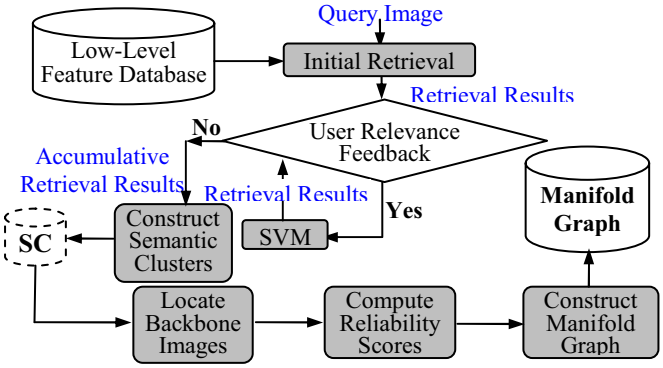
He *et al.* [3] propose the manifold ranking based image retrieval (MRBIR) algorithm to represent images and their relationships as a graph and propagate labeled information among images according to the graph structure. They exploit the distribution of unlabeled images to enhance the ranking and improve the retrieval accuracy. Wan [4] apply the MRBIR algorithm on equal-sized blocks of an image. The retrieval score of each image is a fusion of the ranking scores of all blocks in the image. Cai *et al.* [5] incorporate a locality preserving regularizer into the manifold structure to learn a classification function in the image manifold. They then apply the user's RFs to update the manifold structure for better classification. He *et al.* [6] propose a generalized MRBIR (gMRBIR) algorithm by propagating the scores of a neighborhood-based pseudo seed vector to all unlabeled database images. This gMRBIR algorithm allows the user to submit any query image, inside or outside the database. He *et al.* [7] use the geodesic distances on manifold to measure similarities between images. They then use the radial basis function (RBF) neural network to map low-level features to high-level semantics for inferring the semantics of a new image. Wang *et al.* [8] apply the affinity propagation clustering (APC) algorithm to reduce the manifold graph and preserve its manifold structure. This reduced graph damps the effect of noisy images while emphasizing the effect of reliable images. However, the retrieval performance may be degraded when clusters do not resemble the semantic concepts. All these transductive learning techniques improve the retrieval performance after each iterative feedback step. However, they are sensitive to users' erroneous feedback due to the possible propagation of wrong labels.

In this paper, we propose a novel weighted manifold approach that effectively builds the manifold structure using RF-based semantic clusters (SCs). First, we apply the SVM-based RF technique to create SCs, where all the images within each cluster are semantically similar from the user's perspective. These RF-based SCs correctly divide the database images into meaningful semantic categories to facilitate future learning. Second, we locate the backbone image for each SC and compute the reliability score of each image based on the distance to its backbone image. The higher reliability score an image has, the more semantic information we know about an image, and the more propagation power an image processes. As a result, the backbone images have the highest reliability scores and can then be used to suppress the decayed effects of erroneous

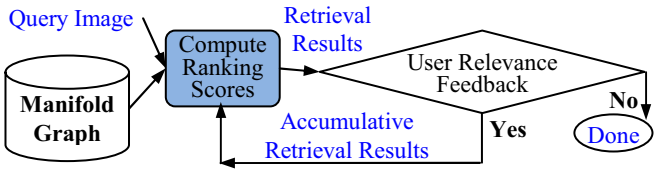
feedback. Third, we incorporate the reliability scores into the affinity matrix to construct the SCs-based weighted manifold graph, which significantly suppresses the noise propagation among the images and is therefore more robust than the traditional manifold graph. Fourth, we asymmetrically construct the relevance vector based on the user's RF and propagate the ranking scores of labeled images in the relevance vector to unlabeled images via the weighted manifold graph. This asymmetrical assignment ensures the propagation on the positive images is dominated and helps unlabeled images to obtain more proper ranking scores than the traditional manifold approach. The rest of the paper is organized as follows: Section 2 presents our proposed SCs-based weighted manifold learning approach. Section 3 compares our system with various MRBIR systems and SVM-based systems. Section 4 draws conclusions and presents future directions.

2. PROPOSED WEIGHTED MANIFOLD APPROACH

The block diagram of our method is shown in Fig. 1. The offline training process of constructing the SCs-based manifold graph is shown in Fig. 1(a). The online retrieval process of propagating the ranking scores of labeled images to unlabeled images is shown in Fig. 1(b). The following subsections explain each component in detail.



(a) Offline training process



(b) Online retrieval process

Fig. 1: The block diagram of the proposed system.

2.1. Offline Process: Initial Retrieval

We use 64-bin HSV color histogram and 36 features to represent each database image. The 36 features consist of 9 color, 18 edge, and 9 texture components. Color features are the first three moments in HSV color space. Edge features correspond to an 18-bin edge direction histogram of the converted grayscale image. Texture features are the entropy of each of 9 detail wavelet subbands of the grayscale image. The Euclidean distance is initially used to measure the similarity between query and database images.

2.2. Offline Process: Construct Semantic Clusters

The aim of constructing SCs is to divide the database images into meaningful semantic categories using users' RFs. Three intuitive observations guide this construction: 1) The semantic relationship among images is complicated and therefore it is difficult to effectively group images using low-level features. 2) An image mainly contains a few interesting objects and therefore belongs to a few important semantic categories. 3) Humans tend to classify objects into semantic categories and remember how well each object belongs to each category [9].

In our proposed system, we randomly select 10% of the database images as training images to perform the query task. For each query, the system first performs initial retrieval to return top n images. The user then selects relevant (i.e., positive) images from the returned pool while treating non-selected images as irrelevant (i.e., negative) images. The system next applies radial basis function (RBF) kernel-based SVM on the user's accumulative feedback to find a better classification boundary to discriminate positive images from negative images in the database. This process continues for a few feedback iterations until the query session finishes. In order to speed up the initial learning and maximize the amount of the semantic relationship information that could be learned on the training set, we ensure that a retrieved image would not be returned in the following iterations. To limit the number of the user's interactions, we return 25 images at each of the four iterations since these images can be easily fit into one screen for users to provide their feedback information.

We construct a new SC after each query session, which includes all the positive images labeled by the user during the current query session. We then iteratively compare this new SC with each of the existing SCs, which are created by prior query sessions, to determine their similarity level and whether they should be merged. Specifically, we find the number of images coexisting in both clusters. If this number is at least a half of the total number of images in either SC, we will merge these two SCs by including all the images in both clusters. This newly merged cluster is then compared with all existing SCs to see whether it should be merged again using the same criterion. The merge process is terminated until there is no significant overlapping among all SCs. Our system respectively creates 39, 106, and 213 SCs for the 2000-, 6000-, and 8000-image DBs, which are close to the actual semantic categories in the database.

2.3. Offline Process: Locate Backbone Images

After constructing SCs using 10% of the database images, we represent each SC by its backbone image. This

backbone image is an image that is in the corresponding SC and has the closest Euclidean distance to the center of the SC. All the remaining images in the SC are considered to be similar to the backbone image. Therefore, their backbone image is the backbone image of their corresponding SC. Since an image may belong to multiple SCs, we assign its backbone image as the one corresponding to its largest SC. We then treat the images outside of all the SCs as outliers whose semantic information has not been learned in the training process. So, we do not assign a backbone image for those outliers.

2.4. Offline Process: Compute Reliability Scores

We compute the reliability score of each image based on the two observations: 1) An image inside a SC contains more reliable semantic information than an image outside the SCs (i.e., an outlier image). 2) The closer an image is to its backbone image, the more reliable its semantic information is. Specifically, if an image is an outlier, we assign its reliability score as a small value (e.g., 0.05). Otherwise, we compute its reliability score r_i by:

$$r_i = r(x_i) = \exp\left(1 - \frac{d(i, B(i))}{\frac{1}{N} \times \sum_{k=1}^{N} d(k, B(k))}\right) \quad (1)$$

where $B(i)$ denotes the backbone image of the i-th database image x_i, $d(i, B(i))$ is the Euclidean distance between x_i and its backbone image, k denotes the k-th database image x_k, and N is the total number of images in the database. This computation ensures that the backbone images have the maximum reliability score and the outlier images have small reliability score. That is, the images with higher reliability scores are believed to be much more representative than those with lower scores.

2.5. Offline Process: Construct the Manifold Graph

The steps for constructing the manifold graph are:

1. Construct an affinity matrix $W = [w_{ij}]_{N \times N}$ where each element w_{ij} represents the similarity between the i-th image and the j-th image in the database and N is the total number of images in the database.
2. Compute the symmetrically normalized affinity matrix S by $D^{-1/2}WD^{-1/2}$, where D is a diagonal matrix with the i-th diagonal element $D(i, i)$ being the sum of the i-th row of W.
3. Compute the final manifold graph M as $(1-\alpha S)^{-1}$, where α is set to be 0.99 in our system.

A suitable affinity matrix W is paramount for achieving good retrieval performance. We integrate the reliability score of each image into the Laplacian kernel based distance, known as L_1 distance, to compute w_{ij} by:

$$w_{ij} = r_i \times r_j \times \prod_{l=1}^{m} \exp\left(-\frac{|x_{il} - x_{jl}|}{\sigma_l}\right) \quad (2)$$

Here, r_i is the reliability score of image x_i; r_j is the reliability score of image x_j; $|x_{il} - x_{jl}|$ is the absolute distance between image x_i and x_j in the l-th dimension, where l ranges from 1 to m (i.e., the feature dimension), and σ is a hyperparameter controlling variance for each dimension. We also designed a variant system by integrating the reliability score of each image into the Gaussian kernel based distance, known as L_2 distance, to compute w_{ij} by:

$$w_{ij} = r_i \times r_j \times \exp\left(-\frac{[d(x_i, x_j)]^2}{\sigma^2}\right) \quad (3)$$

where $d(x_i, x_j)$ is the Euclidean distance between image x_i and image x_j, and σ is the overall variance of image features.

2.6. Online Process: Compute Ranking Scores

For each query, we propagate the ranking scores of labeled images, which are collected during RF iterations, to unlabeled images via the SC-based weighted manifold graph. The propagated ranking scores are then used as the similarity scores between query and database images.

Initially, we encode a relevance vector $Y=[y_i]_{N\times 1}$ by setting the row corresponding to the query image as 1's and setting the remaining elements as 0's. The relevance score of each image is determined by the propagation of vector Y through the manifold graph M. Let $P=[p_i]_{N\times 1}$ represent the relevance score for all images. We compute P by $M\times Y$. Here, the higher score corresponds to more similarity to the query image. As a result, we return n images based on the top relevance scores. The user then labels the returned images as relevant or irrelevant to the query. These labeled images are then incorporated into $Y=[y_i]_{N\times 1}$ by:

$$y_i = \begin{cases} 1 & \text{if the ith image is judged as relevant} \\ -0.25 & \text{if the ith image is judged as irrelevant} \end{cases} \quad (4)$$

The manifold graph M is then multiplied with this updated Y to compute the relevance scores for the next round. This process continues for a few iterations or until the user is satisfied with the retrieval results.

Since the negative images do not provide sufficient information as the positive images, we assign more weight (e.g., 1) to the positive images and less weight (e.g., -0.25) to the negative images in Y. This assignment ensures the propagation on the negatives will not be dominated.

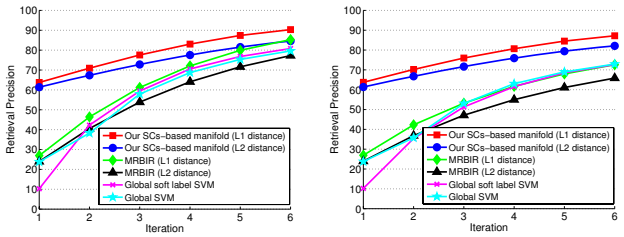
3. EXPERIMENTAL RESULTS

To date, we have tested our proposed system on three data sets: the 2000-Flickr DB, the 6000-COREL DB, and the combined 2000-Flickr and 6000-COREL DB. The COREL DB contains 60 distinct semantic categories with 100 images per category. The Flicker DB contains 20 distinct categories with 100 images per category. The images for the 20 categories were obtained by searching for category related keywords using Flickr's API. We downloaded the top 150 images for each category, and manually picked the most representative 100 images based on semantic contents.

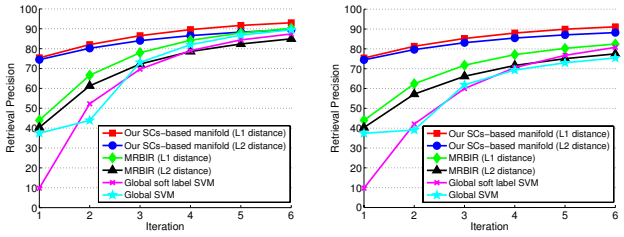
To facilitate the evaluation process, we designed an automatic feedback scheme to construct the SCs-based manifold graph by performing query sessions using 10% unique, randomly chosen database images. For each query session, our system performed four iterative RF retrieval processes and returned top 25 images for each iteration. A retrieved image is considered to be relevant if it belongs to the same category as the query image. The retrieval accuracy is computed as the ratio of the relevant images to the total returned images. We compared the proposed L_1–based manifold system with three manifold systems (our L_2–based manifold, traditional L_1–based manifold [3], and traditional L_2–based manifold [3]) and two SVM-based systems (global soft label SVM and global SVM) on three databases. Fig. 2 compares the average retrieval precision of these systems in the context of having no erroneous feedback and having a level of 5% erroneous feedback. To introduce the noise, we let the simulated "user" misclassify some relevant images as irrelevant and some irrelevant images as relevant. The remaining 90% of the database images (i.e., the images not used in the training process) are used as queries for all the experiments. It clearly shows that our proposed SCs-based manifold systems, both L_1–based and L_2–based manifold systems, outperform the other systems in the context of correct and erroneous feedback. The L_1–based manifold system always performs better than the corresponding L_2–based manifold system. Our SCs-based manifold systems are also resilient to the erroneous feedback since the other systems significantly drop the retrieval precision in all iterations. Specifically, at the last iteration, our proposed L_1–based manifold system respectively achieves the average retrieval accuracy of 90.32%, 92.99%, and 84.74% on the 2000 DB, 6000 DB, and 8000 DB when correct feedback is involved; it respectively achieves the average retrieval accuracy of 87.18%, 91.11%, and 82.86% on 2000 DB, 6000 DB, and 8000 DB when erroneous feedback is involved. This noise resilience feature mainly results from the robust, meaningful SCs and their reliable backbone images learned in the training process.

4. CONCLUSIONS

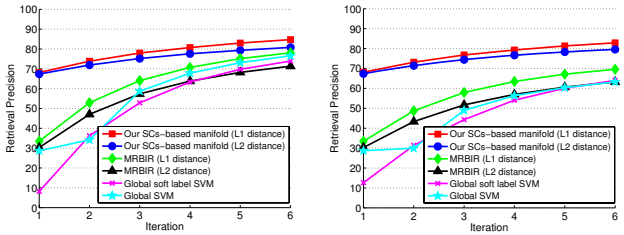
We propose a novel SCs-based manifold ranking system for image retrieval. Major contributions consist of: 1) Using SVM-based RF technique to create SCs to compute the reliability score of each database image. 2) Incorporating the reliability score into the affinity matrix to construct a weighted manifold structure. 3) Creating a relevant vector to ensure the label information can be propagated to high reliable images. Extensive experiments demonstrate our system outperforms the other manifold systems and SVM-based systems. The SCs will be investigated for reducing the manifold graph in our future research.



(a) Retrieval results on the 2000-Flickr DB



(b) Retrieval results on the 6000-COREL DB



(c) Retrieval results on the 8000-image DB

Fig. 2: Comparison of six CBIR systems on three DBs with correct (left) and 5% erroneous (right) relevance feedback.

5. REFERENCES

[1] X. S. Zhou and T. S. Hung, "Relevance Feedback in Image Retrieval: a Comprehensive Review," *ACM Multimedia Systems Journal*, Vol. 8, pp. 536-544, 2003.

[2] M. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Trans. Multimedia, Computing, Communications, and Applications*, Vol. 2, pp. 1-19, 2006.

[3] J. He, M. Li, H. Zhang, H. Tong, and C. Zhang, "Manifold-Ranking Based Image Retrieval," *Proc. ACM Multimedia*, pp. 9-16, 2004.

[4] X. Wan, "Content Based Image Retrieval Using Manifold-Ranking of Blocks," *Proc. of ICME*, pp. 2182-2185, 2007.

[5] D. Cai, X. He, and J. Han, "Regularized Regression on Image Manifold for Retrieval," *Proc. of Int. Workshop on Multimedia Information Retrieval*, pp. 11-20, 2007.

[6] J. He, M. Li, H. Zhang, H. Tong, and C. Zhang, "Generalized Manifold-Ranking-Based Image Retrieval," *IEEE Trans. Image Processing*, Vol. 15, No. 10, pp. 3170-3177, 2006.

[7] X. He, W. Ma, and H. Zhang, "Learning an Image Manifold for Retrieval," *Proc. of ACM Int. Conf. on Multimedia*, pp. 17-23, 2004.

[8] F. Wang, G. Er, and Q. Dai, "Inequivalent Manifold Ranking for Content-Based Image Retrieval," *Proc. of Int. Conf. on Image Processing*, pp. 173-176, 2008.

[9] S. Pinker, How the Mind Works, W. W. Norton & Company, New York, New York, 1997.