6-2007

Similarity Beyond Distance Measurement

Feng KANG
Michigan State University

Rong JIN
Michigan State University

Steven C. H. HOI
Singapore Management University, chhoi@smu.edu.sg

Similarity Beyond Distance Measurement

Feng Kang kangfeng@msu.edu Michigan State University

Rong Jin rongjin@cse.msu.edu Michigan State University

Steven C. H. Hoi chhoi@cse.cuhk.edu.hk Hong Kong Chinese University

Abstract

One of the keys issues to content-based image retrieval is the similarity measurement of images. Images are represented as points in the space of low-level visual features and most similarity measures are based on certain distance measurement between these features. Given a distance metric, two images with shorter distance are deemed to more similar than images that are far away. The well-known problem with these similarity measures is the semantic gap, namely two images separated by large distance could share the same semantic content. In this paper, we propose a novel similarity measure of images that goes beyond the distance measurement. The key idea is to exploit the clustering structure of images when a large number of images are present. The similarity of two images is determined not only by their Euclidean distance in the space of visual features but also by the likelihood for them to be clustered together, which is further estimated using a marginalized kernel. Our empirical studies with COREL datasets have shown that the proposed similarity measure is effective for traditional content-based image retrieval as well as user relevance feedback.

1 Introduction

Content-based image retrieval (CBIR) has been an active research area for more than a decade (Smeulders *et al.* 2000). One of the keys to CBIR is how to measure the similarity of two images in their low-level visual features. Most similarity measures are related to certain distance measurement: given a distance measure for the space of low-level visual features, two images are deemed to be similar when they are close in the space of low-level visual features. The simplest and probably the most common similarity measure is based on the Euclidean distance. Other distance functions have been developed for similarity measure (Santini & Jain 1999) include L_p norm based distance function and the geodestic distance. In addition, many studies have devoted themselves to the data dependent distance function. Most of these studies learned distance metric by taking advantage of the side information of images. In (Jeon, Lavrenko, & Manmatha 2003; Lavrenko, Manmatha, & Jeon 2003), the authors proposed to learn a distance metric of low-level visual features based on the textual annotations of images. In (He *et al.* 2003) and (Xing *et al.* 2003), the authors presented algorithms that automatically learned distance metrics from the log data of user relevance feedback. In (Yan *et al.* 2004), machine learning algorithms were applied to automatically determine the distance metrics based on the pair-wise constraints.

Despite the extensive study of similarity function in the past, the key assumption of similarity function remains unchanged, i.e., the similarity of two images is inverse to their distance. Evidently, this

assumption may not always hold in reality. It is possible for two images that are far away separated to be more similar in their semantic content than two images that are close to each other. In this paper, we propose a novel similarity measure that goes beyond the distance measurement. In particular, this new similarity measure exploits the clustering structure of a large number of images. The similarity of two images are determined by the following two factors:

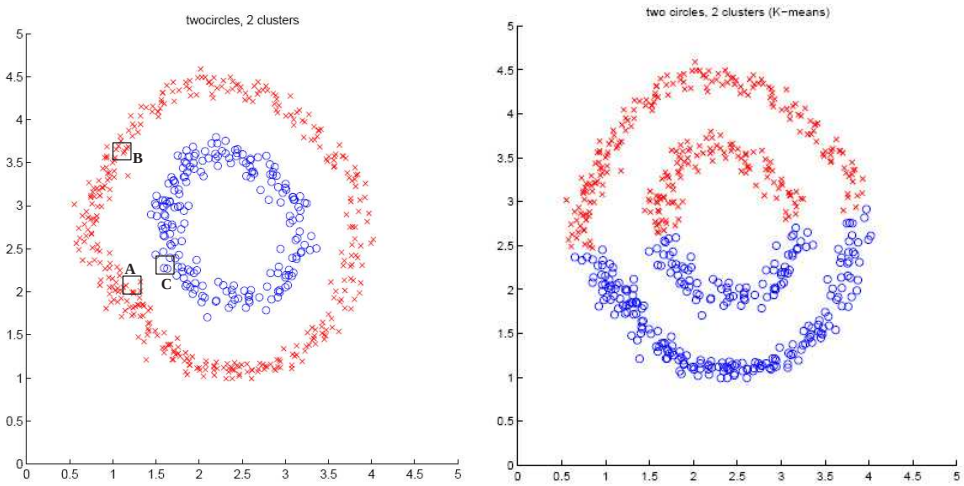
- **Separation Distance**, i.e., the distance of the two images in the space of low-level visual features. The closer the two images are in low-level visual features, the more similar they are.
- **Clustering Likelihood**, i.e., the likelihood for the two images to be clustered together. The more likely the two images are to be clustered together, the more similar they are.

The second factor is based on the hypothesis that if two images share similar content, they tend to be clustered together in the space of primitive features. By employing the spectral clustering algorithm, we will allow two far separated images to have a larger similarity than two images that are close by. In order to see this, consider a toy example when all the data points are distributed over two clusters. As illustrated in Figure 1(a) (excerpt from (Ng, Jordan, & Weiss 2001)), these groups of data points form two co-centered circles, highlighted by markers "x" and "o" separately. Assume that data points on the same circle are similar to each other. According to the proposed similarity measure, two data points from different circles will have a small similarity no matter how close they are to each other in the two dimensional space. Hence, data points in the rectangle "A" are more similar to data points in the rectangle "C" than data points in the rectangle "B", even though the rectangle "A" is closer to "B" than "C".

The rest of the paper is arranged as follows: Section 2 describes the research work that is related to this paper. Section 3 introduces the concept of marginalized kernel and its application to the similarity measure. Section 4 presents our empirical studies. Section 5 concludes this work.

2 Related Work of Image Similarity Measures

This work is strongly related to the previous studies that apply machine learning methods to improve the quality of similarity measurement of images. Most of these methods rely on the side information of images, such as textual annotations and user relevance feedback that are logged by CBIR systems. In (Jeon, Lavrenko, & Manmatha 2003; Lavrenko, Manmatha, & Jeon 2003), the authors presented a similarity measure based on automatic image annotation. First the association between low-level visual features and annotated words are automatically learned from a collection of annotated images. Then images in the low level feature space are mapped to the space of annotated words, in which the similarity between two images is measured. In (He, Ma, & Zhang 2004; He *et al.* 2003), the authors presented manifold learning approaches to learn similarity measure from the log data of user relevance feedback. The method assumes that a CBIR system is able to collect and store a large amount of relevance feedback from many users. Using the Laplacian Eigenmap (Belkin & Niyogi 2002), a low dimensional subspace of low-level image features is identified, in which the Euclidean distance measurement of images is consistent with the users' relevance judgments in log data. In (Muller, Pun, & Squire 2004), the authors took another avenue in exploiting log data of user relevance feedback. They extended the weighting schemes such as the TF.IDF methods in text retrieval (Salton & Buckley 1988), and used the log data to estimate the weights for low-level features. Unlike the above approaches that rely on the side information to determine appropriate similarity measure, the proposed approach is based on unsupervised



(a) Spectral clustering (b) K-means

Figure 1. Examples of clustering results

learning methods (i.e., clustering) and therefore does not require any side information when estimating the similarity measurement of images.

Finally, our work is also related to the CLUE image retrieval system (Chen, Wang, & Krvetz 2005). Unlike most image retrieval systems where returned images are presented to users in a ranking list, the CLUE system presented retrieved images in a clustered view. Given a query image, it first identifies a number of images that are similar to the query image based on the predefined similarity measure. Next, it applies a spectral clustering algorithm to group images of related content into a few clusters. Then, the retrieved images are presented to users based on the clustered results. Our work differs from (Chen, Wang, & Krvetz 2005) in that we focus on the design of similarity measure based on marginalized kernel instead of the presentation strategy for retrieved images.

3 Structural Similarity Measure

In this section, we will first describe the concept of Marginalized kernel, and then we will discuss the application of marginalized kernel to computing the structural similarity measure.

3.1 Marginalized Kernel

Marginalized kernel (Tsuda, Kin, & Asai 2002) is used to compute similarity of two objects when parts of the object information are hidden. Let $z = (x, h)$ be the full representation of an object where x stands for the observed part of the representation and h stands for the hidden part of the representation. Given two objects $z = (x, h)$ and $z' = (x', h')$, let $K_z(z, z')$ be a Kernel function that computes the similarity between z and z' based on their full representations. Then, an interesting question is how to compute the similarity of two objects given that only their observed representations x and x' are provided. The main idea of marginalized kernel is to first "guess" the hidden part of the representation h given its observed counterpart x. Then, with the "guessed" hidden representation h, we will apply the kernel function $K_z(z, z')$ to estimate the pair wise similarity. Since we are not sure if the guessed hidden representation h is absolutely correct, we need to take into account the uncertainty in deciding

h given x, which is usually described by a probability distribution $p(h|x)$. As a result, the similarity for two objects based on their observed representations x and x' should be the average of the Kernel function $K_z(z = (x, h), z' = (x', h'))$ over all possible values for the hidden representations h and h'. Formally, the marginalized kernel for two objects with only partial representations x and x' is expressed as follows:

$$K(x, x') = \sum_h \sum_{h'} K_z(z = (x, h), z' = (x', h'))p(h|x)p(h'|x') \quad (1)$$

3.2 Similarity Measure Incorporating Structural Information

As already described in Section 1, our hypothesis is that if two images are clustered into the same group, they are more likely to share similar content than images that are clustered into different groups. The cluster membership of images, to certain degree, indicates the semantic relation in image content. Hence, we will modify the distance-based similarity of two images by incorporating the likelihood for them to be clustered together. To this end, we introduce a hidden variable for each image to represent its cluster membership. Then, each image $\mathcal{I} = (\vec{f}, c)$ is represented by its low-level features $\vec{f}$ and cluster membership c. Here low-level feature vector $\vec{f}$ corresponds to the observed representation of images and cluster membership c corresponds to the hidden representation of images. Given two images $\mathcal{I} = (\vec{f}, c)$ and $\mathcal{I}' = (\vec{f'}, c')$, the original kernel definition is $K_{\mathcal{I}}(\mathcal{I}, \mathcal{I}') = \exp(-\frac{|\vec{f}-\vec{f'}|_2^2}{\sigma^2})$ where σ^2 is the variance of image features, and $|\vec{f} - \vec{f'}|_2^2$ is the square of Euclidean distance between vector $\vec{f}$ and $\vec{f'}$. After we apply marginalized kernel, the similarity between the two images become(1), i.e.,

$$\begin{aligned} & K(\vec{f}, \vec{f'}) \\ &= \sum_c \sum_{c'} K_{\mathcal{I}}(\mathcal{I} = (\vec{f}, c), \mathcal{I}' = (\vec{f'}, c'))p(c|\vec{f})p(c'|\vec{f'}) \\ &= \exp(-\tfrac{|\vec{f}-\vec{f'}|_2^2}{\sigma^2})(\sum_c \sum_{c'} \delta(c, c')p(c|\vec{f})p(c'|\vec{f'})) \\ &= \exp(-\tfrac{|\vec{f}-\vec{f'}|_2^2}{\sigma^2})(\sum_c p(c|\vec{f})p(c|\vec{f'})) \end{aligned} \quad (2)$$

The above expression is a product of two terms: the first term is based on the Euclidean distance measurement; and the second term can be interpreted as the likelihood for two images $\mathcal{I}$ and $\mathcal{I}'$ to be clustered together under all possible assignments of cluster membership to $\mathcal{I}$ and $\mathcal{I}'$. In the extreme case, when using hard cluster membership, each image is assigned to an unique cluster and the related probability $p(c|\vec{f})$ becomes a binary value. As a result, $\sum_c p(c|\vec{f})p(c|\vec{f'})$ degenerates to a $\delta(c, c')$ where c and c' are the cluster memberships assigned to images $\mathcal{I}$ and $\mathcal{I}'$, respectively.

Note that there are two types of uncertainties when assigning each image to a cluster:

- *Partitioning uncertainty*. This refers to the case when there are multiple ways of dividing a collection of images into a number of clusters and each way of partitioning is almost equally good. This happens because most clustering algorithms can only find solutions of local maximum. For example, in the EM clustering algorithm, a different initialization of parameters usually results in a different clustering structure.

- *Membership uncertainty*. Many clustering algorithms, such as EM clustering algorithm and fuzzy clustering algorithm (Bezdek 1981), employ soft membership. In particular, images close to the cluster boundaries will have large uncertainties in determining their cluster memberships and vice versa. Hence, given a determined partition of image collections, we still have uncertainties in assigning cluster membership to each cluster.

To explicitly express the two types of uncertainties, we expand the similarity function in (2) as follows:

$$\begin{aligned} & K(\vec{f}, \vec{f'}) \\ & = \sum_t \sum_{c,c'} p(t) K_{\mathcal{I}}(\mathcal{I} = (\vec{f}, c), \mathcal{I}' = (\vec{f'}, c')) p(c|\vec{f}; t) p(c'|\vec{f'}; t) \\ & = \exp(-\frac{|\vec{f} - \vec{f'}|_2^2}{\sigma^2}) (\sum_p \sum_c p(t) p(c|\vec{f}; t) p(c|\vec{f'}; t)) \end{aligned} \tag{3}$$

In the above formulism, random variable t is introduced to represent the partitioning uncertainty. Probability $p(t)$ stands for the likelihood of applying partitioning t to an image collection. $p(c|\vec{f}; t)$ is the probability of assigning image vector $\vec{f}$ to cluster c given the partitioning t. Slightly different from the interpretation for (2), the second term in (3) is interpreted as the likelihood for two images to be clustered together under any partitioning of image collections as well as any assignment of cluster memberships. In practice, to create multiple partitioning for a given image collection, we will run a clustering algorithm a number of times and each time starts with a different initialization of parameters. Furthermore, we set probability $p(t)$ to be a constant for all different runs of the clustering algorithm. This is based on the assumption that each run will result in a clustering structure of similar quality.

3.3 Image Retrieval based on Structural Similarity

The key to the proposed similarity measure is how to accurately compute the likelihood of two images to be clustered together. We will discuss two involved issues: the choice of clustering algorithms and the determination of the numbers of clusters. In this study, we choose the spectral clustering algorithm for proposed similarity measure. This is because spectral clustering usually significantly outperform other clustering algorithms. This has been demonstrated in several application domains, including image segmentation (Shi & Malik 2000), text categorization (Ding *et al.* 2002), and data mining (Ng, Jordan, & Weiss 2001). Another more important reason is because spectral clustering captures not only the compactness of clusters but also the connectivity of clusters. This property of spectral clustering has been demonstrated in Figure 1(a). Hence, by using the spectral clustering algorithm, the structural similarity measure is able to go beyond the traditional distance-based similarity measure in that two images can have a small similarity even though they share similar low-level visual features.

Another important issue related to clustering algorithms is how to determine the number of clusters, which can have significant impact on the proposed similarity measure. Overestimate may lead to splitting of same category into different parts and thus have small similarity based on structural similarity, while underestimate will group different semantic categories into the same group. In this study, we use the algorithm in (Sugar & James 2003) to automatically determine the number of clusters. The main idea of this algorithm is to compare the change of average Mahalonobis distance between successive change of the number of clusters. It is based on the observation that when the number of clusters is close to the optimal one, the change in the size of clusters will become less noticeable because some of the partitions involves arbitrary splitting of true clusters and therefore will not be able to shrink the cluster size substantially.

4 Experimental results

In this experiment, we will examine the effectiveness of the proposed structural similarity measure. In particular, we will address the following research questions:

1. *Will the proposed similarity measure be effective for content-based image retrieval*? To address this question, we will evaluate the proposed similarity measure for two retrieval tasks: the standard content-based image retrieval, and the regular relevance feedback for image retrieval. We will compare the performance of the proposed similarity measure to the performance of other well-known similarity measures.

2. *How sensitive is the proposed similarity measure to the number of clusters*? We will vary the number of clusters and see how it effects the retrieval accuracy of the proposed similarity measure. We will also examine the presented algorithm for automatically determining the number of clusters.

4.1 Datasets

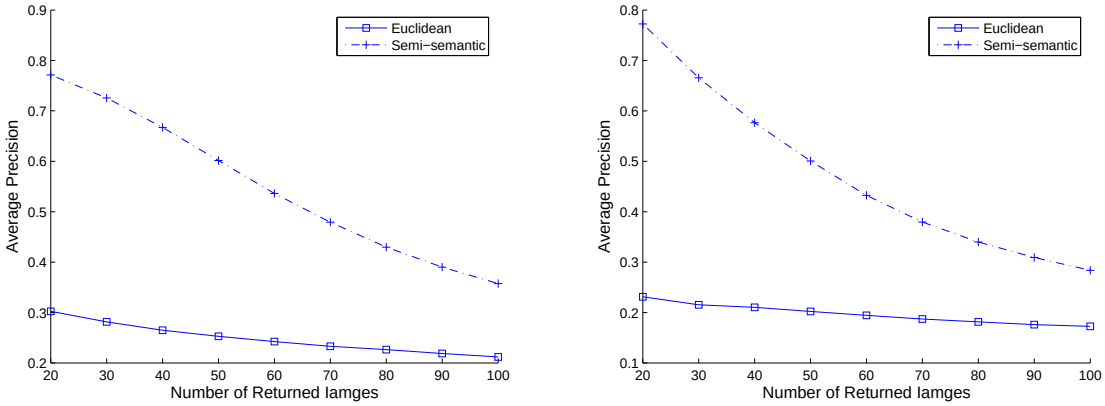
In this experiment, we use the data provided in (Hoi & Lyu 2004). It consists of two testbeds: the "20-category" testbed comprises images from 20 semantic categories, and the "50-category" comprises images from 50 semantic categories. For each semantic category, 100 images are randomly selected from COREL database and are included in the testbed. Hence, there are totally 2000 images in the 20-category testbed and 5000 images in the 50-category testbed. Every image in a testbed is used as a query image for the testbed. Hence, there are totally 2000 query images for the 20-category testbed and 5000 query images for the 50-category testbed. For a given query image, an image in the testbed is deemed relevant if it shares the same category as the query image. Otherwise, it is marked as irrelevant. The average precision of the top retrieved images is used to measure the quality of retrieved results. Despite that such a definition of relevance judgments may not accurately reflect the characteristics of relevance judgments by real-world users, it is able to avoid the subjectiveness in manual relevance judgments. Furthermore, it automates the process of evaluation and allows different approaches to be compared based on the same ground truth. In practice, this evaluation methodology has been adopted by many studies of image retrieval, such as (He *et al.* 2004; He, Ma, & Zhang 2004; Tong & Chang 2001; He *et al.* 2003; Hoi & Lyu 2004).

Three types of visual features are used to represent images: 9 for color, 18 for edges detected by canny filter, 9 for texture detected by Daubechies-4 wavelet filter (Smith & Chang 1996). Overall, each image is represented by 36 features. More details of image features can be found in (Hoi & Lyu 2004).

Unless specified, a variant of normalized cut (Yu & Shi 2003) is used as the spectral clustering algorithm for computing structural similarity measure throughout all the experiments.

4.2 Experiment (I): Content-based Image Retrieval

In this experiment, we test the effectiveness of the proposed structural similarity measure for content-based image retrieval. The number of clusters used for both testbeds are automatically determined by the algorithm that is already described in Section 3.3. The resulting number is 11 clusters for the 20-category testbed and 41 clusters for the 50-category testbed. Note that the two numbers of clusters are not identical to the number of semantic categories. Using the proposed structural similarity measure, images will be ranked in the descending order of their similarities to query images and the most similar images are presented to users. Figure 2(a) and 2(b) show the average precision of the structural similarity measure for the two testbeds, respectively. For the purpose of comparison, in Figure 2, we also plot the average precision of CBIR based on the Euclidean distance metric.



(a) 20-Category (b) 50-Category

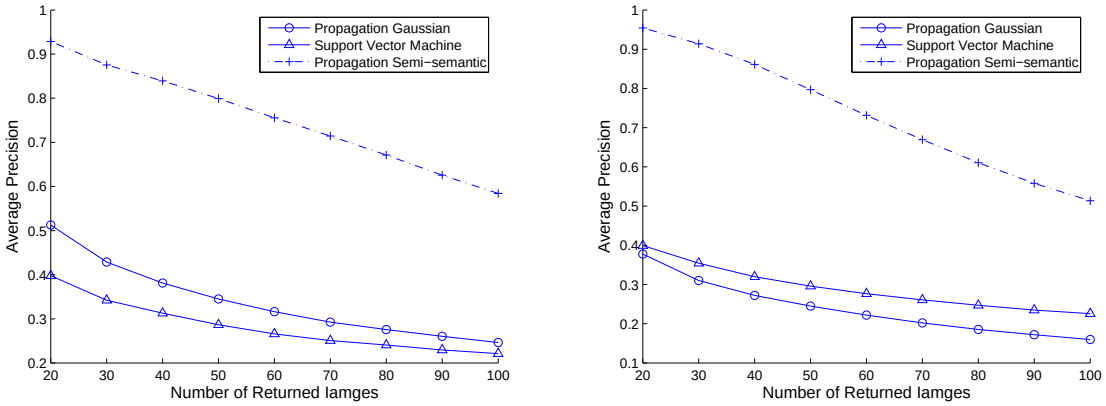
Figure 2. Average precision for traditional content-based image retrieval using the Euclidean distance and the structural similarity measure

According to Figure 2, we observe that the structural similarity measure significantly outperforms the Euclidean distance metric for all evaluation points. For both testbeds, the average precision of the structural similarity measure for the top 20 images is over 78%. Compared to the average precision of the Euclidean distance metric for the same evaluation point, which only 30.2% for the 20-category testbed and 23.1% for the 50-category testbed, the improvement by the structural similarity measure is dramatic given that no side information is provided. Even more surprisingly, for both testbeds, the precision of the Euclidean distance metric for the top 20 returned images is even lower than the precision of the structural similarity for the top 100 returned images, which is 38.4% for the 20-category testbed and 27.8% for the 50-category testbed.

4.3 Experiment (II): Relevance Feedback for Image Retrieval

In this experiment, we test the effectiveness of the proposed similarity measure for relevance feedback of image retrieval. In the past, there have been a large number of studies on applying relevance feedback techniques to image retrieval (Chang *et al.* 2003; Cox *et al.* 1998; Hoi & Lyu 2004; Huang & Zhou 2001; Tieu & Viola 2000; Vasconcelos & Lippman 2000; Laaksonen, Koskela, & Oja 1999), ranging from heuristic methods to sophisticated learning methods. In general, any relevance feedback mechanism requires users' relevance judgements for the results returned by a CBIR system in response to a user query. Given the relevance judgments for the retrieved results, relevance feedback is then engaged as a query refinement method to improve the retrieval accuracy of the CBIR system.

In this experiment, we apply the semi-supervised learning approach in (He *et al.* 2004) to relevance feedback. This relevance feedback method is based on the idea of label propagation, which was originally introduced in (Zhou *et al.* 2003). It first constructs a similarity matrix $\mathbf{W} = [w_{i,j}]_{n\times n}$ and each element $w_{i,j}$ represents the similarity between the i-th image and the j-th image. Next, it calculates the normalized similarity matrix $\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, where matrix $\mathbf{D}$ is a diagonal matrix and the i-th diagonal element $\mathbf{D}_i = \sum_{j\neq i} w_{i,j}$. Then, it encodes the user relevance judgments into a vector $\vec{y} = [y_i]_{n\times 1}$



(a) 20-Category (b) 50-Category

Figure 3. Average precision for image retrieval using different relevance feedback methods

as follows:

$$y_i = \begin{cases} 1 & \text{if the } i\text{-th image is judged as relevant} \\ -1 & \text{if the } i\text{-th image is judged as irrelevant} \\ 0 & \text{if the } i\text{-th image is not judged} \end{cases}$$

The final relevance score of each image is determined by the propagation of vector $\vec{y}$ through the normalized similarity matrix. If $\vec{r} = [r_i]_{n\times 1}$ represents the relevance score for all images, then

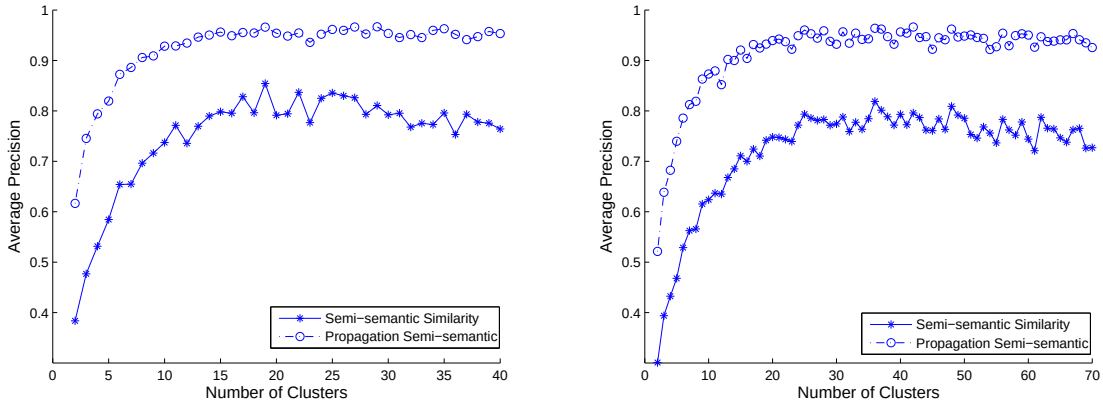
$$\vec{r} = (1 - \alpha \mathbf{S})^{-1}\vec{y}$$

More details of this algorithm can be found in (He *et al.* 2004). We refer to this method as the "*Propagation Method*".

One of the keys to the propagation method is the similarity measurement. In (He *et al.* 2004), a guassian function is used to compute the pair wise similarity of images, i.e., $w_{i,j} = \exp(-\frac{|\vec{f}_i - \vec{f}_j|_2^2}{\sigma^2})$ where σ is the overall variance of image features. We can also use the structural similarity measure to construct $\mathbf{W}$. We refer to the former method as "*Propagation Gaussian*" and the later one as "*Propagation Structural*". For the parameter α in the propagation method, we followed (He *et al.* 2004) and set $\alpha = 0.99$. Furthermore, each user feedback provides relevance judgments for the top 20 returned images.

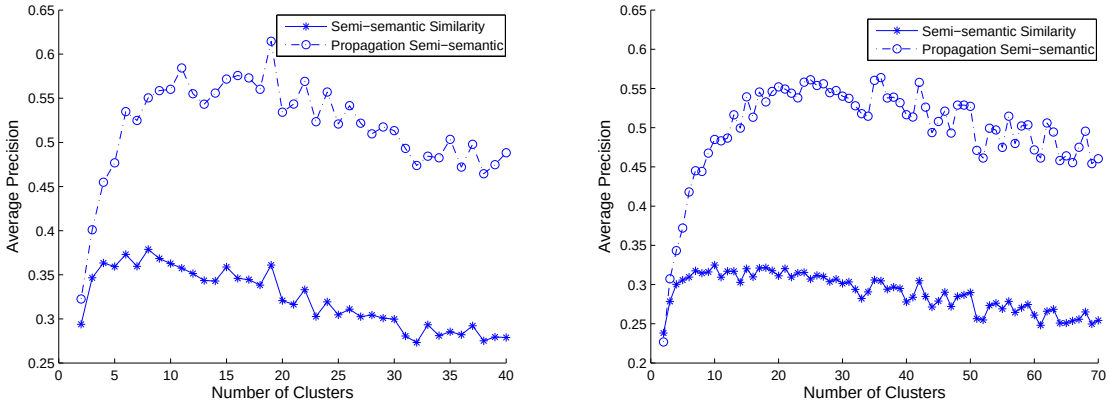
The average precision for the propagation method using different similarity measure for the two testbeds is plotted in Figure 3(a) and 3(b), respectively. For the completeness of comparison, we also include the results for relevance feedback using support vector machines (SVM), which has been a popular method for relevance feedback in CBIR (Tong & Chang 2001).

According to Figure 3, we observe that the propagation method using the Gaussian function as the similarity measure *does not* consistently outperforms the SVM method. It performs noticeably worse than the SVM method on the 50-category testbed. In contrast, the propagation method using the proposed similarity measure performs significantly better than the SVM method for both testbeds. In particular, it achieves very high precision when the number of returned images is small. For instance, for both testbeds, the retrieval accuracy of the top 20 images returned by the propagation method using the structural similarity measure is over 90%. Furthermore, similar to standard content-based image retrieval,



(a) 20-Category (b) 50-Category

Figure 4. Average precision for the top 20 returned images using different number of clusters



(a) 20-Category (b) 50-Category

Figure 5. Average precision for the top 100 returned images using different number of clusters

for both testbeds, the retrieval accuracy of "Propagation structural similarity" for the top 100 returned images is even better than that of "Propagation Gaussian" for the top 20 returned images. Finally, comparing Figure 2 with Figure 3, we could see that for both testbeds, even the structural similarity measure without any relevance feedback significantly outperforms the two relevance feedback approaches that does not use the structural similarity measure. For example, for "Propagation Gaussian", its average precision for the top 20 returned images is 51.3% for the 20-category testbed and 37.8% for the 50-category testbed. These two numbers are significantly lower than the average precision of the structural similarity measure for the same evaluation point, which is above 78% for both testbeds. These empirical results indicate that the structural similarity measure is effective not only for standard CBIR and but also for relevance feedback of image retrieval.

4.4 Experiment (III): Number of Clusters

In this section, we will show how different numbers of clusters will influence the retrieval accuracy of CBIR using the proposed similarity measure. To this end, for each number of clusters, we apply the structural similarity measure to image retrieval and compute the average precision of both standard CBIR and relevance feedback for the top 20 and 100 returned images. The results for the first 20 and 100 returned images are plotted in Figure 4 and Figure 5, respectively.

First, we examine the change in the retrieval accuracy for the top 20 images in response to different numbers of clusters. According to Figure 4, we see that when the number of clusters is small, the retrieval accuracy of both standard CBIR and relevance feedback is usually improved by increasing the number of clusters. When the number of clusters becomes sufficiently large (i.e., 15 for 20-category and 25 for 50-category), the performance of both image retrieval tasks will remain almost unchanged regardless of the number of clusters that is set for structural similarity measure. This fact indicates that if we only concern with the precision for a small number of returned images, the number of clusters will have little impact on the retrieval accuracy of the structural similarity measure as long as it is reasonably large.

Second, we examine the change in the retrieval accuracy for the top 100 images in response to different number of clusters. Similar to the above discussion, the retrieval accuracy is first improved when the number of clusters increases from a small number. However, unlike the performance for the top 20 returned images, the retrieval accuracy for the top 100 returned images begins to decline after the number of clusters reaches its optimum value. As the number of cluster increases, the number of images in the clusters become less yet the images become more compact. Thus if we only return top 20 retrieved images, the cluster is sufficient to provide 20 images with high accuracy. However, as the number of cluster increases, the images in the cluster get less and less and might not be sufficient enough to provide 100 images. As we know, the marginalized kernel will set the similarity low as the two images do not belong to the same cluster. Thus, to acquire the high precision for a large number of returned images, it is important to find a good estimation of the number of clusters. We need to balance the tradeoff between the accuracy of small number of returned images and large number of returned images. According to Figure 5, for the two estimated numbers of clusters (i.e., 11 for 20-category and 41 for 50-category), we see that they are reasonably close to the optimal number of clusters.

5 Conclusion

In this paper, we presented a novel similarity measure for images retrieval based on unsupervised learning method. Thus, we can adjust the similarity of two images according to how likely for them to be clustered together. Based on this idea, we proposed the "structural similarity measure" that combines the tradition distance-based similarity measure with the likelihood for two images to be clustered together. A marginalized kernel approach is used to estimate the similarity of two images in terms of their cluster memberships. Compared to the distance-based similarity measure, this new similarity measure encompasses both low level similarity and the certain semantic similarity. By exploiting the clustering structure of a large number of images, this new similarity is able to express the semantic similarity of two images to a certain degree, thus reducing the semantic gap in content-based image retrieval.

Empirical studies with both traditional content-based image retrieval and regular relevance feedback have indicated that the proposed similarity measure can dramatically improve the performance of image

retrieval tasks. Meanwhile, the proposed similarity measure appears to be robust to the number of clusters as long as it is reasonably large.

In the future, we plan to investigate the effectiveness of the structural similarity measure for image classification. For example, we can introduce the proposed similarity measure into support vector machines through a kernel function and apply the kernelized support vector machines to image classification.

6 Acknowledgements

This research was supported by National Science Foundation (IIP-0610784 and IIS-0643494) and National Institute of Health (1R01GM079688-01).

References

Belkin, M., and Niyogi, P. 2002. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*.

Bezdek, J. C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.

Chang, E. Y.; Li, B.; Wu, G.; and Goh, K.-S. 2003. Statistical learning for effective visual information retrieval. 609–612.

Chen, Y.; Wang, J. Z.; and Krvetz, R. 2005. Clue:cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing* 14(8).

Cox, I. J.; Miller, M.; Minka, T.; and Yianilos, P. 1998. An optimized interaction strategy for bayesian relevance feedback. In *Procedings of IEEE CVPR'98*.

Ding, C.; He, X.; Zha, H.; Gu, M.; and Simon, H. 2002. Spectral Min-Max Cut for Graph Partitioning and Data Clustering. In *Proceedings of 1st IEEE International Conference on Data Mining*, 107–114.

He, X.; King, O.; Ma, W.-Y.; Li, M.; and Zhang, H. J. 2003. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* 13(1):39–48.

He, J.; Li, M.; Zhang, H.-J.; Tong, H.; and Zhang, C. 2004. Manifold-ranking based image retrieval. In *Proceedings of ACM MM 2004*. New York, NY, USA: ACM Press.

He, X.; Ma, W.-Y.; and Zhang, H.-J. 2004. Learning an image manifold for retrieval. In *Proceedings of ACM MM 2004*.

Hoi, C.-H., and Lyu, M. R. 2004. A novel log-based relevance feedback technique in content-based image retrieval. In *Proceedings of ACM MM 2004*. New York, NY, USA: ACM Press.

Huang, T. S., and Zhou, X. S. 2001. Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods. In *Proceedings of IEEE International Conference on Image Processing (ICIP'01)*.

Jeon, J.; Lavrenko, V.; and Manmatha, R. 2003. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. In *Proceedings of the 26th annual international ACM SIGIR*.

Laaksonen, J.; Koskela, M.; and Oja, E. 1999. Picsom: Self-organizing maps for content-based image retrieval. In *Proc. International Joint Conference on Neural Networks (IJCNN'99)*.

Lavrenko, V.; Manmatha, R.; and Jeon, J. 2003. A Model for Learning the Semantics of Pictures. In *Proceedings of Advance in Neutral Information Processing*.

Muller, H.; Pun, T.; and Squire, D. 2004. Learning from user behavior in image retrieval: Application of market basket analysis. *Int. J. Comput. Vision* 56(1-2):65–77.

Ng, A.; Jordan, M.; and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. In *NIPS*.

Salton, G., and Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing and Management: an International Journal* 24(5):513–523.

Santini, S., and Jain, R. C. 1999. Measures of image similarity. *IEEE Transactions on pattern Analysis and Machine Intelligence* 21(9):871–883.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8):888–905.

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

Smith, J., and Chang, S.-F. 1996. Automated image retrieval using color and texture. *IEEE Transaction on Pattern Analysis and Machine Intelligence*.

Sugar, C. A., and James, G. M. 2003. Finding the number of clusters in a dataset: an information-theoretic approach. *Journal of the American Statistical Association*.

Tieu, K., and Viola, P. 2000. Boosting image retrieval. In *Proceedings of IEEE CVPR'00*.

Tong, S., and Chang, E. 2001. Support vector machine active learning for image retrieval. In *Proceedings of ACM MM 2001*. ACM Press.

Tsuda, K.; Kin, T.; and Asai, K. 2002. Marginalized kernels for biological sequences. *Bioinformatics* 18(90001):S268–S275.

Vasconcelos, N., and Lippman, A. 2000. Bayesian relevance feedback for content-based image retrieval. In *Proceedings of IEEE CVPR'00*.

Xing, E. P.; Ng, A. Y.; Jordan, M. I.; and Russell, S. 2003. Distance metric learning with application to clustering with side-information. In S. Becker, S. T., and Obermayer, K., eds., *Advances in Neural Information Processing Systems 15*, 505–512. Cambridge, MA: MIT Press.

Yan, R.; Zhang, J.; Yang, J.; and Hauptmann, A. G. 2004. A discriminative learning framework with pairwise constraints for video object classification. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR)*.

Yu, S. X., and Shi, J. 2003. Multiclass spectral clustering. In *Proceedings of Ninth IEEE International Conference on Computer Vision*.

Zhou, D.; Weston, J.; Gretton, A.; Bousquet, O.; and Schlkopf, B. 2003. Ranking on data manifolds. In *NIPS*.