Improving Image Retrieval Effectiveness via Multiple Queries

Xiangyu Jin
Department of Computer Science
University of Virginia
Charlottesville, VA 22904
xj3a@cs.virginia.edu

James C. French
Department of Computer Science
University of Virginia
Charlottesville, VA 22904
french@cs.virginia.edu

ABSTRACT

Conventional approaches to image retrieval are based on the assumption that relevant images are physically near the query image in some feature space. This is the basis of the cluster hypothesis. However, semantically related images are often scattered across several visual clusters. Although traditional Content-based Image Retrieval (CBIR) technologies may utilize the information contained in multiple queries (gotten in one step or through a feedback process), this is only a reformulation of the original query. As a result these strategies only get the images in some neighborhood of the original query as the retrieval result. This severely restricts the system performance. Relevance feedback techniques are generally used to mitigate this problem. In this paper, we present a novel approach to relevance feedback which can return semantically related images in different visual clusters by merging the result sets of multiple queries. Further research topics, such as achieving candidate queries' visual diversity, are also discussed. We also provide experimental results to demonstrate the effectiveness of our approach.

Categories and Subject Descriptors

H.2.8 **[Database Management]**: Database Applications – image databases; H.3.1 **[Information Storage and Retrieval]**: Content Analysis and Indexing; H.3.3 **[Information Storage and Retrieval]**: Information Search and Retrieval – search process.

General Terms: Algorithms, Measurement, Performance, Experimentation.

Keywords: Content-based image retrieval, multi-channel CBIR, result merging.

1. INTRODUCTION

Content-base image retrieval (CBIR), which was proposed in the early 1990's [12], has been one of the most active research areas in the past decade. A good survey can be found in [15, 20]. Many visual features, such as color, texture and shape have been explored to describe the content of the images and many systems were built. Most existing CBIR systems search images via a query-by-example (QBE) [7, 12] interface. In QBE, the system presents an initial page of representative image thumbnails to the user. The user then marks one or more images as relevant or irrelevant to his/her intention. These chosen positive/negative examples are employed both to generate a query center and to re-weight the visual space's axes [8]. This feedback process can be repeated continuously until the user is satisfied with the retrieval result, or a predefined threshold is met.

MMDB'03, November 7, 2003, New Orleans, Louisiana, USA.

No matter what feedback strategies, what visual features, and what similarity measurements are used, current CBIR systems unanimously assume all semantically relevant images are clustered in some visual feature space. The function of the feedback process is to move the query center toward positive examples and away from negative examples and also to enhance the importance of some visual features or some dimensions of a specific visual feature. However, images with relevant concepts might be spread out in the entire visual space and can be scattered in several visual clusters rather than one. In this case, the traditional query-center-based feedback approach cannot guarantee an improvement of performance when shifting the query center. Even worse, the query center may get nothing when the relevant images are widely scatted.

Relevance feedback has long been used to enhance retrieval performance in text-based IR systems. More recently this notion has been applied to CBIR systems. A number of techniques are surveyed in [21]. The general idea is to use images identified by the user as relevant to modify the current query so that in the next search iteration the modified query can be used and will be more effective, that is, retrieve more relevant images. The query is modified by query point movement [8,19] or by query expansion [9, 10]. In addition, the distance function might be reweighted [8, 18, 19]. The goal of query point movement is to find the "best" query and generally assumes a convex region of interest [8, 19] but more recently non-convex and disjoint regions [9, 10] have been considered. The notion of disjoint regions is fundamental to our work and our query formulation is similar to disjunctive queries [18] at the feature level while still capturing a single information need at the higher semantic level. Our main contribution to query expansion is our techniques for introducing diversity into the feedback images used in subsequent retrievals.

In this paper, we propose a multi-query retrieval strategy (called a multipoint query in [10]) to solve this problem. Instead of finding a query center for the selected positive examples, our approach issues these positive queries individually and then merges their results later into one synthetic list. This approach can gather a retrieval result that is scattered in the visual space rather than purely in the neighborhood of the query center. Our approach to query reformulation is implicit, that is, we use the feedback images as queries and merge their results rather than explicitly modifying the original query in search of a "best" query. In this way we simultaneously search along several trajectories in the feature space and merge the outcomes for display to the user. Chen et al. [22] have taken this approach of using multiple "seed" queries but they use them to expand the frontier around the "best" query and are still essentially searching a single region of the feature space.

How to achieve visual diversity in the candidate queries is an important research topic for multi-query retrieval. We introduce our multi-representation [6] and multi-system retrieval [5] to help increase visual diversity in the retrieval result, thus the system performance is improved further. These technologies have the ability to improve individual system performance, but they can also achieve extra benefits when combined together. A synthetic retrieval model is proposed and experimental results are also presented to show the effectiveness of these technologies.

The rest of the paper is organized as follows. The cluster hypothesis and the traditional CBIR search problem are discussed in section 2. Our multi-query retrieval is presented in section 3. Section 3 also deals with achieving visual diversity in the candidate queries and presents our synthetic retrieval model. Afterwards, the analysis of our experimental results is shown in section 4. Finally, we draw conclusions and discuss our future works in section 5.

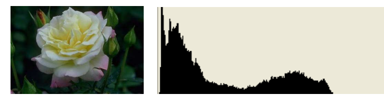
2. THE CBIR SEARCHING PROBLEM

The cluster hypothesis[16, p. 45] states that "closely associated documents tend to be relevant to the same requests." For CBIR we substitute "images" for "documents." This hypothesis is widely assumed and forms the foundation of similarity-based retrieval.

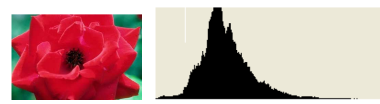
Current CBIR technology assumes that semantically related images are physically clustered in some visual feature space. Therefore retrieval can be performed by getting the images in the neighborhood of the given query in the visual feature space. This clustering phenomenon may hold for many images in image databases, no matter if they are composed of pictures taken from natural scenes, images generated by computer graphics, or paintings drawn by artists. Many images with similar semantics will more or less resemble each other perhaps because they were taken in series at similar time and scene, generated in a series by similar graphic models in computer, or drawn by the same artist in a period time. Take the COREL database as an example. The photos of the flowers (Figure 1) are taken in a series at similar time and place, resulting in quite similar visual features, such as illumination, color composition, etc.

Figure 1. Semantically related images resemble each other in visual features.

However, not all images with the same semantics will be close to each other in the visual space. Semantically related images may be scattered in the visual space. Still considering the flower pictures in COREL database as an example, red flowers and yellow flowers are quite different in their color histogram (Figure 2). Therefore there is not a single cluster for the semantics "flower". Actually there exist several clusters in the visual space. Figure 3 shows such an example. There are yellow, white, and red flowers in the image database, and they form at least three clusters naturally in the visual space. Each cluster contains several images with similar visual features.



(a) Color histogram of a white flower



(b) Color histogram of a red flower

Figure 2. Semantically related images can be quite different in visual features.

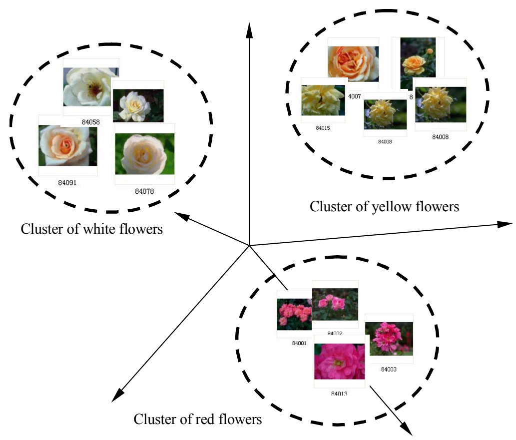
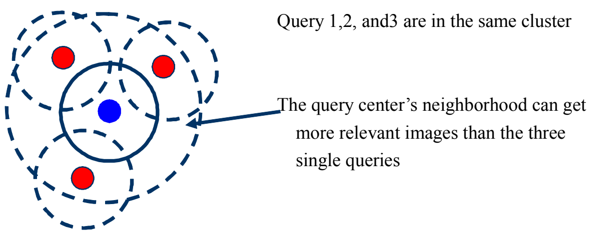
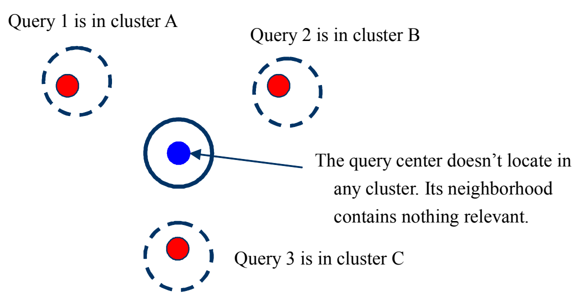


Figure 3. Semantically related images are scattered in several visual clusters.

Traditional CBIR technologies often assume there only exists one cluster. They try to find a "best" representative of the user intention in the visual space and get the images in the neighborhood of this optimistic query. Unfortunately, this approach does not work well unless only one cluster exists. From Figure 3 we can see that no matter how you choose such an optimistic query and how you reshape the query neighborhood, it is unrealistic to cover all the three scattered clusters in a tight region. This results in poor performance of current CBIR technology. Some traditional CBIR feedback approaches try to move the query center by linear combination of the positive feedback images and triy to adjust the neighborhood shape by relevance feedback. Again, this approach may not be effective if more than one cluster exists. Figure 4 shows such an example. In Figure 4 (a), all positive queries happen to be located in one cluster and adjusting the query center may improve retrieval effectiveness. This is the theoretical foundation of traditional relevance feedback technology in CBIR. However, this will not work when the feedback queries are located in several clusters. Figure 4 (b) shows that if the feedback queries are located in several clusters, the query center may not be in any cluster. This results in the neighborhood of the query center containing nothing the user wanted.



(a) When the queries are located in the same cluster, the query center can help improve performance



(b) When the queries are located in different clusters, the query center can degrade performance

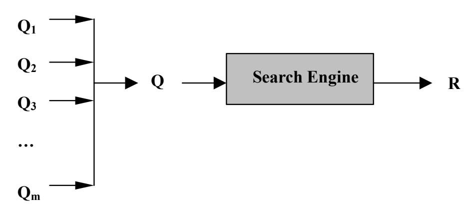
Figure 4. Retrieval by query center of multiple queries may achieve different effects when the queries are located in one or more than one cluster.

This CBIR search problem occurs because many CBIR approaches, no matter what feedback strategies, what visual features, and what distance formulations are used, all tend to find a single query center to perform retrieval task. This single query center does not sufficiently utilize the information contained in several feedback queries. In order to solve this problem, approaches using multiple queries have been developed, e.g. multipoint queries [9,10], or support for disjoint regions (disjunctive queries) [18]. We propose a novel multi-query retrieval technique, which can return semantically related images in different visual clusters by merging the result sets of multiple queries.

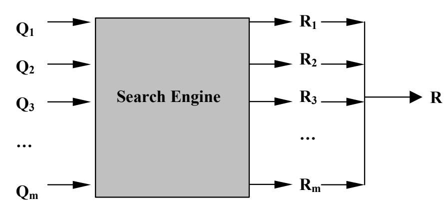
For a given retrieval task, the user may pick different queries, which are all semantically related to the images the user desires. These queries will generate different retrieval results by the same CBIR system. These different result lists (referred to as *channels* in this paper) can be thought of as different viewpoints [3] regarding the retrieval task in user's mind. The retrieval results in these two viewpoints have the potential to be merged into a better synthetic channel.

Merging the channels of multiple queries has two advantages. First, merging the channels of queries in one cluster, called intra-cluster merging, can help filter some irrelevant images in the list, because the likelihood that an irrelevant image is similar to all the given queries will tend to be very low, although it may be quite similar to some query. This effect can improve retrieval precision. Second, merging the channels of queries in different clusters, called inter-cluster merging, can help to insert images from different clusters into the synthetic channel. Therefore images in different clusters will get the chance to appear in the synthetic list fairly. This effect can improve retrieval recall. Consequently, both retrieval precision and recall can be improved if we merge the channels of multiple queries.

In earlier work we conducted a preliminary investigation of the effects of query diversification in CBIR systems [4]. The work reported here extends that work by applying query diversification to the query process. This idea is analogous to the work in text IR on combination-of-evidence strategies and dates back to the early 90's. Two approaches have generally been used to capture an information need more precisely by a diversity of queries [1, 2]. The first approach combines the queries Q_1 to Q_m before searching, and then submits the synthetic query to the search engine to the get the result R (Figure 6(a)). The second approach issues search Q_1 through Q_m individually and then merges the results of each query R_1 to R_m afterwards into a synthetic list R (Figure 6(b)). Query point movement approaches take the first strategy, while our approach adopts the second one.



(a) Combine queries before search



(b) Combine results after search

Figure 6. Two strategies for multiple query retrieval.

Our previous work [5, 6] has already shown that merging channels of different representations and different systems can improve retrieval effectiveness. In this paper we extend the work to multiple queries. This can greatly improve retrieval effectiveness. And such improvement will be even greater when combined with our multi-representation and multi-system retrieval technologies.

3. ACHIEVING VISUAL DIVERSITY OF THE CANDIDATE QUERIES

According to the cluster hypothesis, it is easy to see that the more clusters the queries cover, the higher the potential of performance improvement. In a QBE user interface, multiple relevant images are selected by the user as feedback queries after the original query is performed. However, in a single channel CBIR system [6], the candidate queries provided by the user interface are usually in one cluster, because they are quite similar to the original query in visual space. For example, if the user only chooses candidate queries from the cluster of red flowers (see

Figure 3), how can the system have the chance to retrieve yellow and white flowers? Therefore, diversity of visual clusters in the images provided to the user is required. Our multi-channel retrieval approaches, such as multi-representation and multi-system, provide such an opportunity.

3.1 Multi-representation Retrieval

Representations are the processing results of certain transformations. Unlike extracted features, these representations will still keep the original media format and can be directly processed by the retrieval system. For example, a color histogram of a given image is not a representation as we define it, for it is not a viewable image anymore and cannot be processed by the image retrieval system directly. A grayscale image is a representation of the original color image. Figure 7 shows four representations for a yellow flower, 84060.jpg, in COREL database. We use the original color image (C+) together with the grayscale image (B+) and both the color negative (C-) and the grayscale negative (B-) to form four different representations in this paper. These representations form the channels of our augmented CBIR system.

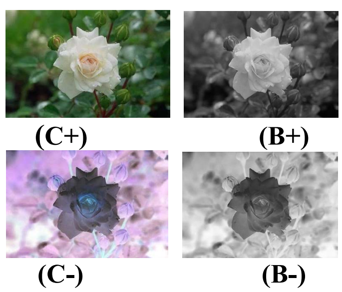


Figure 7. Multiple representations of the same image.

Extra information can be acquired through other channels. By doing so, the system can gain wider bandwidth through different representations. Our previous work [6] has shown that merging the retrieval results of multi-representations introduces some diversity of images in visual space. We claim this characteristic will help achieving diversity of candidate queries in our multi-query retrieval.

3.2 Multi-system Retrieval

Different retrieval systems will generate different viewpoints and retrieve different results for the same retrieval task. On one hand, two systems may have different opinions on some images of whether they should be returned to the user. On the other hand, both of the systems may agree on some images that should be returned to the user. This partial duplication is extremely valuable for it shows that merging the retrieval results of multi-systems has the potential to gain extra benefits. It can help improve retrieval effectiveness, including precision and recall. Our previous work [5] has already shown this fact. It also introduces some diversity of the relevant images in visual space.

3.3 A Synthetic Retrieval Model

Putting multi-representation, multi-system, and multi-query technologies together, we get a synthetic[1] CBIR model (Figure 10). This synthetic retrieval model is composed of different channels at different levels. Each channel is a viewpoint of the user intended images. These channels will be merged into one and provided to the user in an integrated user interface. Relevance feedback is also employed to let the user pick up relevant images from the synthetic channel. Weights of different channels can be adjusted. The integrated retrieval system gains wider bandwidth through different representations and systems, and improves performance further by different queries, thus better performance is achieved than with any individual approach.

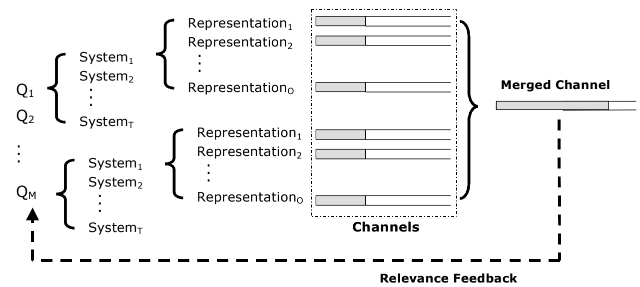


Figure 10. The framework of the Synthetic Retrieval Model.

4. EXPERIMENTAL RESULTS

In this section, we report on experiments that evaluate the effects of combining multi-representation, multi-system retrieval, and multi-query retrieval together.

4.1 Test Environment

We selected 34 categories from COREL, each with 100 images, to form our test set with 3,400 images. The selection principle was whether the images contained one salient foreground object. The purpose of this is so that the image set can work well under segmentation-based image retrieval systems.

We use the COREL category as a ground truth, i.e., we regard all the images inside the same COREL category to be relevant. Four image representations are employed in this experiment, which are color, color negative, grayscale, and grayscale-negative. Two typical CBIR systems are employed in our evaluation. One is global-feature-based and the other is region-based. The global-feature-based system is referred to as the Basic CBIR system in this paper. It is built similar to MiAlbum system [11] with seven visual features, including three color features and four texture features. All these features are compared respectively to get a similarity and then linearly combined with equal weights. The region-based system employed in this experiment is SIMPLIcity [17]. This system performs image segmentation and then uses an integrated-region-matching (IRM) approach in distance calculation.

4.2 Simulated Relevance Feedback

In order to mimic proper user behavior, we perform simulated relevance feedback in our test bed. First, an original query is used

[1] We call this a *synthetic* model because we merge a variety of channels to *synthesize* new channels.

to generate an initial retrieval result. Then we simulate the user interaction by selecting up to *k* relevant images from the retrieval result according to the ground truth. These selected relevant images are issued individually and then their retrieval results are merged together to generate a synthetic channel. Relevance feedback is performed according to the following strategies:

- Random relevant in all

  k relevant images (including the original query image) are randomly selected from the relevant images. This feedback strategy has nothing to do with what the system returns for the original query. It is just like randomly sorting all the 3,400 images in the image set in a list, and then asking the user to pick *k* relevant images from it, no matter how many images he/she finally looked at. Unfortunately, this optimistic case will never happen in a real application, for the task of finding relevant images from a random list is completely impractical. However, this strategy is a good one to use for performance upper bound analysis, because it will provide queries that are widely distributed in the visual space and let us know the potential of how well multi-query retrieval can do.

- Blind feedback

  The top *k* images (including the original query itself), no matter relevant or irrelevant, in the synthetic channel of the original query, are used as the feedback queries. If blind feedback itself can improve system performance, we can improve system performance without any human intervention.

- Top relevant in truncated

  Top *k* relevant images (including the original query itself) are selected from the truncated synthetic channel of the original query. Obviously, the worst case is there is nothing relevant except the original query. And the best case is there are more than *k-1* relevant images in addition to the original query. Consequently, the feedback query number can vary from 1 to *k* (including the original query). This strategy is used to simulate a practical user interface. Only part of the synthetic channel will be returned to the user, and the user will only look for relevant images in this much smaller image set. We assume the user tends to choose the first *k* relevant images from the image list.

- Random relevant in truncated

  This strategy is actually a mix of the first and the third strategies. First we randomly sort the truncated synthetic channel, and then let the user choose up to *k* relevant images from this list. We still assume the user tends to choose the first *k* relevant images seen from the image list. Hence top *k* relevant images (including the original query itself) will be chosen from the resorted synthetic channel. Similar to the third strategy, the feedback query number can vary from 1 to *k*, including the original query. Obviously, this is also a practical user interface. We do not introduce more work for the user to do than in the previous strategy.

4.3 Merging Technology

There are many varieties of merging technologies. In this paper, we use an intuitive but effective merge strategy called mid-rank merge [6]. In this merge strategy, all channels will be treated as having the same importance, i.e., they have same weights. This strategy is basically a rank sum that is adjusted for ties. Although we could use a more complex merge strategy, which may improve the performance further, we want to show that even a simple merging strategy results in greatly improved retrieval performance.

Each query is presented to two different systems having 4 representations (channels) each. The resulting 8 channels are merged to get a synthetic channel for the query. Then these independent synthetic channels are merged further to generate a single synthetic channel. This channel is presented to the user as the retrieval result.

4.4 Evaluation Metrics

All the 3,400 images are issued individually as queries. Average precision[2] at 50 (P50) and average precision at 100 (P100) are used as evaluation metrics. We claim P50 and P100 to be suitable metrics for image retrieval, for each screen can hold about 50 thumbnails and the user seldom clicks more than twice to turn the pages to view more images. Moreover, we argue that including the rank of image in the evaluation is meaningless because the content of all the images in one page can be viewed at a glance and does not require a great deal of work to be checked one by one (this is different from text retrieval). Finally, we do not use average non-interpolated precision as the evaluation metric because the merged list will be longer in length, thus resulting in unfair comparisons.

4.5 Performance Evaluation

The system performance is affected by many factors, such as the number of feedback queries, feedback iterations, truncation lengths of the channels, feedback strategies, the composition of the synthetic channel, etc. In this section, we mainly analyze three factors' affect on the system performance:

- The number of feedback queries

  Six different settings are used for comparison, single query (just the original query without any feedback), and 4, 8, 12, 16, 20 queries. We expect multi-query retrieval to be better than single query retrieval. The more queries employed, the better the performance. Therefore 20 queries should have the highest performance and a single query should have the lowest.

- The composition of synthetic channel

  Five different settings are used for comparison, including single system with single representation, single system with multi-representations, and multi-systems with multi-representations. The 5 different configurations are: Basic C+, SIMPLIcity C+, Basic 4 (with 4 representations), SIMPLIcity 4 (with 4 representations), and all 8 (2 systems with 4 representations each). We expect multi-representation

[2] Precision is the ratio of the number of relevant images in the result list to the number of returned images.

and multi-system retrieval technologies to help to improve the performance further due to diversity of the queries in visual space. Therefore, all 8 channels should have the highest performance and Basic C+ or SIMPLIcity C+ should have the lowest.

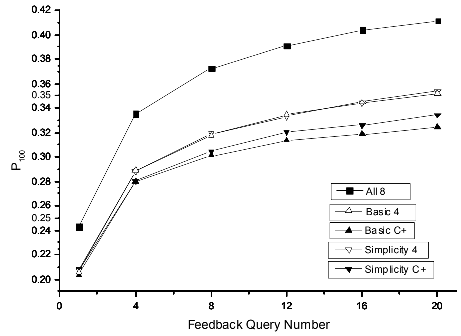
- Feedback strategies

Four feedback strategies are used for comparison. We expect ***random relevant in all*** have the highest performance, because this strategy can introduce more diversity of queries. We don't expect ***blind feedback*** to improve performance, because the feedback queries have high possibility of being in one visual cluster. And we expect ***random relevant in truncated*** to have higher performance than ***top relevant in truncated*** and lower performance than ***random relevant in all***. Because it should cover more visual clusters than the ***top relevant in truncated*** strategy, but fewer visual clusters than ***random relevant in all*** strategy.

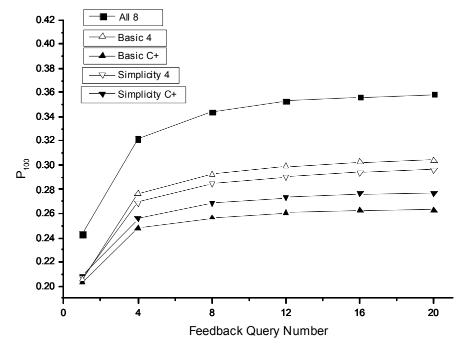
As we described before, there are 6 different settings of feedback query number (1, 4, 8, 12, 16, 20), 5 different settings of synthetic channel composition (Basic C+, SIMPLIcity C+, Basic 4, SIMPLIcity 4, all 8), and 4 different settings of feedback strategies (random relevant in all, blind feedback, top relevant in truncated, random relevant in truncated). So there are 120 different configurations altogether. In our experiments, we run all the 3,400 queries under these 120 configurations to get the average P_{50} and P_{100}. All merging is performed on the truncated channels of length 150. And the synthetic channel is also truncated to be 150. Therefore, the user can only look for relevant images from the top 150 images in the synthetic channel. This is based on the assumption that the user seldom looks for images after three full screens, each with 50 thumbnails. Moreover, only one time feedback is employed in our current experiment. Due to space limitations, we only make our analysis based on P_{100}, but from P_{50} we can draw quite similar conclusions. The experimental results are shown in Figure 11. The horizontal axis shows the feedback query number, which can be 1, 4 … and 20. The vertical axis shows the precision at 100. Different settings of synthetic channel compositions are shown by 5 curves. And the 4 feedback strategies are shown in four sub-figures, respectively.

We can draw the following conclusions from Figure 11. Multi-query retrieval does help to improve system performance. From Figure 11(a)(b)(c), the system precision will rise with the increment of horizontal axis. The more queries are employed in the retrieval, the higher the system performance. We also can see, the improvement is greater when the query number increased from 1 to 4 than increased from 4 to 8. For instance, in *random relevant in all* (Figure 11(a)), the performance of Basic C+ relatively improved 37.7% when merging 4 queries, and improved 10.5% again when merging another 4 queries (thus getting 8 queries). Although merging multiple queries will help improve the performance, this improvement will become smaller and smaller as more queries are added, until all clusters are covered.

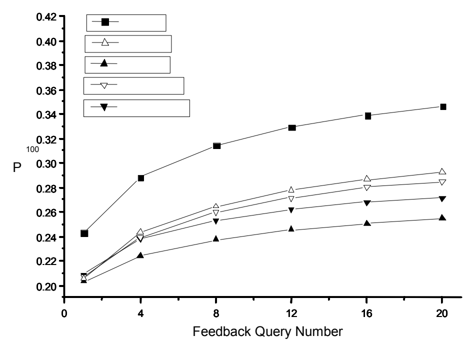
Blind feedback does not improve performance at all because its feedback queries are usually located in the same visual cluster. Figure 11(d) shows the 5 curves all remain flat with the increase in number of queries. Even worse, some curves indeed drop a little, this is because the feedback images maybe not semantically relevant to the user intention.



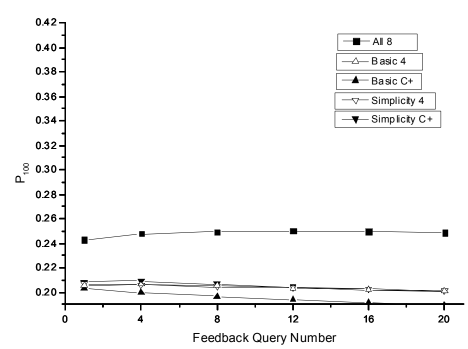
(a) random relevant in all



(b) random relevant in truncated



(c) top relevant in truncated



(d) blind feedback

Figure 11. Performance evaluation.

Random relevant in all has the highest performance due to its high coverage to visual clusters. An unusual phenomenon here is the performance of *random relevant truncated* is better than the performance of *top relevant in truncated*, especially when the query number is small. This is contrary to a widely accepted belief. For many years in the CBIR world almost all feedback strategies tend to pick up the top relevant images from the result list. But our result shows that this is not necessarily the best strategy. When we randomly reshuffle the truncated result list and present it to the user, we might move some bad examples forward and put some good examples backward. But actually, this strategy works much better under our multi-query retrieval environment because it introduces more chances for the user to select positive examples from different visual clusters. That is, this strategy tends to create more query diversity in the top *k* relevant images.

Comparing Figure 11(a), (b), and (c), we can see that the performance curve increases differently. In Figure 11(a) and (c), the performance curve increases sharply from a single query to 8 queries, and then increases gradually from 8 queries to 20 queries. But the curve in Figure 11(b) is quite different. It increases sharply from single query to 8 queries, and then remains almost unchanged when more queries added. In the *random relevant in all* strategy, when more queries are added, more clusters are covered, hence the performance is continually increased. But in the *random relevant in truncated* strategy, the images provided for selection is a subset of all images, hence it only covers some visual clusters. After adding 8 queries, there is almost no chance of adding more queries to cover more clusters. This is why the curve almost does not increase after 8 queries added. However, in the *top relevant in truncated* strategy, the situation changes again. This strategy will return images, which are most visually similar to the original query. Therefore the selected images have high possibility to reside in some restricted clusters. Consequently, under this strategy, adding more queries will no doubt help the query set to cover more clusters.

Another interesting proof of our cluster hypothesis is the different behavior between the two CBIR systems, Basic and SIMPLIcity. We see that SIMPLIcity C+ improves its performance much better than Basic C+. This is because our basic CBIR system is merely based on global visual features, and it has a high tendency to have its retrieval results reside in the same visual cluster. Therefore, relevant images from the channel of Basic C+ do not help as much to achieve visual diversity. But in the *random relevant in all* strategy, relevant images are selected widely from the whole image set, this can compensate for this vulnerability.

The synthetic channel helps to improve performance further. Figure 11(a)(b)(c) shows this phenomenon. As we described before, multi-representation retrieval does not improve the performance directly, but introduces some visual diversity. This can enhance performance improvement under multi-query retrieval. Take the basic CBIR system as an example, Basic C+ (0.2032) and Basic 4 (0.2056) have quite similar precision when performing single query retrieval. When performing 8-query retrieval, both settings increase their performance. The precision of Basic 4 is improved to 0.2925 and performance of Basic C+ is improved to 0.2564 under *random relevant in truncated* feedback (Figure 11(b)). Although both settings increase their performance, the synthetic channel no doubt gains more advantage from multi-query retrieval technology. Another example, the synthetic channel of all 8 channels has a precision of 0.2428, which is higher than a Basic C+ channel, which has a precision of 0.2032. But when merging multiple queries, this difference has been increased. Under *random relevant in truncated* feedback (Figure 11(b)), we can see Basic C+ channel improves its performance to 0.2564, relatively improving 26.2%, while the synthetic channel of all 8 channel improved its performance to 0.3440, relatively improving 41.7%. Therefore, multi-representation and multi-system retrieval cannot only improve the system performance directly, but also can achieve a latent benefit, which will be released by the multi-query retrieval technology.

From the analysis above, we conclude that combing multi-representation, multi-system, and multi-query retrieval technologies together would greatly improve system performance. Our *random relevant in truncated* feedback strategy can perform better than merely picking up the top *k* relevant images. On the other hand, although more queries will provide more information and guarantee a better performance, it also introduces more computation. It is also unlikely that a user will choose so many relevant images in the user interface. In our system, it seems that 8 queries is a good compromise between performance and computational expenses.

When using the *random relevant in truncated* feedback strategy (feedback 8 queries), the synthetic 8 channels will achieve a precision of 0.3440. If using the traditional CBIR technology as a baseline, Basic CBIR at 0.2032 or SIMPLIcity at 0.2087, our technology improves the performance relatively about 67.0%. This is a great improvement of the performance of image retrieval. On the other hand, we claim our random shuffle will not increase the workload much for the user. The synthetic channel has an initial precision at 0.2428 when performing the original single query. This means that on average there are 24 relevant images ranked in top 100 returned by the synthetic channel. Even if no more relevant images appear in the next 50 images, these 24 images already guarantee when we randomly distribute them in the top 150 images, that on average 8 relevant images will appear in the top 50, i.e., the first page. Consequently, in most cases, a user can find enough relevant images (8 in this paper) already in the first page of the retrieval results.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel multi-query retrieval technology to improve retrieval effectiveness in CBIR. And also introduce our multi-representation and multi-system retrieval technologies to improve the visual diversity in the candidate queries, thus the system performance is improved further. Different feedback strategies are also compared. Our experimental results show that the more visual clusters are covered in the candidate queries, the higher the potential of performance improvement when we merge these queries together.

Tahaghoghi et al. [14] have demonstrated improved retrieval performance when using two images but question the usefulness of using more than two claiming that “further improvements in using more than two examples may not justify the added processing required.” [14, p. 138] Our results show that there is added benefit when more than two queries are used. The apparent discrepancy is due, we believe, to the fact that our techniques increase visual diversity and allow us to find related images from other parts of the visual space.

Achieving the diversity in the candidate queries is an important research issue of our future work. There is a tradeoff between the precision of the retrieval result and the visual diversity in the images presented to the user. If we want to keep a relatively high precision in the results presented to the user, we'd better show the top rank images to the user. But this strategy will result in the user selected images tending to be co-clustered. On the contrary, if we introduce some randomness in the retrieval result, such as randomly sorting part of them, we can achieve more visual diversity, but perhaps with a lower precision in the retrieval result. Therefore, we should present relevant images in different visual clusters to the user as earlier as possible. Proper user interface study and clustering technologies are our further research issues.

6. REFERENCES

[1] Belkin N., Cool C., Croft W., and Callan J. The Effect of Multiple Query Representations on Information Retrieval System Performance, *Proc. of ACM SIGIR*, pp339-346, Pittsburgh, PA, 1993.

[2] Belkin N., Kantor P., Cool C., and Quatrain R. Combining Evidence for Information Retrieval, *Proc. of TREC-2,* pp35-44, Gaithersburg, MD, March 1994.

[3] French J.C., Chapin A.C., and Martin W.N. Multiple Viewpoints as an Approach to Digital Library Interfaces, *Workshop on Document Search Interface Design and Intelligent Access in Large-scale Collections,* Portland, OR, July, 2002.

[4] French J.C., Martin W.N., and Watson J.V.S. A Qualitative Examination of Content-Based Image Retrieval Behavior using Systematically Modified Test Images, *45th IEEE International Midwest Symposium on Circuits and Systems,* Tulsa, OK, August, 2002.

[5] French J.C., Watson J.V.S., Jin X., and Martin W.N. Integrating Multiple Multi-channel CBIR Systems, *International Work Shop on Multimedia Information System (MIS'03),* Ischia, Italy, May, 2003.

[6] French J.C., Watson J.V.S., Jin X., and Martin W.N. Using Multiple Image Representations to Improve the Quality of Content-based Image Retrieval, Tech. Report CS-2003-10, Dept. of Computer Science, Univ. of Virginia, March, 2003.

[7] Hirata K. and Kato T. Query by visual example-content based image retrieval, *Proc. of 3rd international conference on extending database technology*, vol. 580 of LNCS, pp56-71, Vienna, Austria, March 1992.

[8] Ishikawa Y., Subramanya R., and Faloutsos C., MindReader: Querying database through multiple examples, *24th International Conference on Very Large Databases (VLDB'98*), New York, USA, August, 1998.

[9] Kim D. and Chung C. Qcluster: Relevance Feedback Using Adaptive Clustering for Content-based Image Retrieval, *Proc. of SIGMOD'03*, San Diego, CA, June, 2003.

[10] Porkaew, K., Chakrabarti, K. and Mehrotra, S. Query refinement for content-based multimedia retrieval in MARS. *Proc. of the 7th. ACM Int. Conf. on Multimedia*, pp235-238, 1999.

[11] Liu W., Dumais S., Sun Y., Zhang, H. Czerwinski M., and Field B. Semi-Automatic Image Annotation, *Proc. of Human-Computer Interaction-Interact*, pp326-333, 2001.

[12] Niblack W., Barber R., Equitz W., and et al. The QBIC project: Querying images by content using color, texture, and shape, Proc. of SPIE Electronic Imaging: Science and Technology, San Jose, CA, 1993.

[13] Salton G. and McGill M.J. Introduction to Modern Information Retrieval, McGraw-Hill, 1989.

[14] Tahaghoghi S.M.M., Thom, J.A., and Williams, H.E.. Are Two Pictures Better Than One? *Proc. of the 12th Australasian Database Conference (ADC'01)*, Gold Coast, Australia, pp. 138-144, January, 2001.

[15] Huang, T.S. and Rui Y. Image Retrieval: Past Present and Future, in *Proc. of Int. Symposium on Multimedia Information Processing*, Taiwan, Dec. 1997.

[16] van Rijsbergern C.J. Information Retrieval (2nd Edition), Butterworths, London, 1979.

[17] Wang J.Z. and Du Y. Scalable Integrated Region-based Image Retrieval using IRM and Statistical Clustering, *Proc. of ACM/IEEE Joint Conference on Digital Libraries (JCDL'01)*, Roanoke, VA, June, 2001.

[18] Wu L., Faloutsos C., Sycara K., and Payne T. Falcon: Feedback adaptive loop for content-based retrieval, Proc. of 26th International conference on very large data bases (VLDB), Cairo, Egypt, 2000.

[19] Rui, Y., Huang, T.S., Ortega, M. and Mehrotra, S. Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval , *IEEE Tran on Circuits and Systems for Video Technology,* Special Issue on Segmentation, Description, and Retrieval of Video Content, pp. 644-655, Vol. 8, No. 5, Sept, 1998.

[20] Yoshitaka A. and Ichikawa T. A Survey on Content-based Retrieval for Multimedia Database, *IEEE Trans. on Knowledge and Data Engineering*, Vol.11, No.1, pp81-93.

[21] Zhang, H., Chen, Z., Li, M. and Su, Z. Relevance Feedback and Learning in Content-Based Image Search, *World Wide Web*, Vol. 6, No. 2, pp. 131-155, 2003.

[22] Chen, Y., Wang, J.Z. and Krovetz, R. An Unsupervised Learning Approach to Content-based Image Retrieval, *Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Paris, July, 2003.