Relevance Feedback Techniques in Interactive Content-Based Image Retrieval *

Yong Rui, Thomas S. Huang, and Sharad Mehrotra

Beckman Institute and Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
E-mail:{yrui, huang}@ifp.uiuc.edu, sharad@cs.uiuc.edu

Abstract

Content-Based Image Retrieval (CBIR) has become one of the most active research areas in the past few years. Many visual feature representations have been explored and many systems built. While these research efforts establish the basis of CBIR, the usefulness of the proposed approaches is limited. Specifically, these efforts have relatively ignored two distinct characteristics of CBIR systems: (1) the gap between high level concepts and low level features; (2) subjectivity of human perception of visual content.

This paper proposes a relevance feedback based *interactive* retrieval approach, which effectively takes into account the above two characteristics in CBIR. During the retrieval process, the user's high level query and perception subjectivity are captured by dynamically updated weights based on the user's relevance feedback. The experimental results show that the proposed approach greatly reduces the user's effort of composing a query and captures the user's information need more precisely.

1 Introduction

With the advances in the computer technologies and the advent of the World-Wide Web, there has been an explosion in the amount and complexity of digital data being generated, stored, transmitted, analyzed, and accessed. Much of this information is multimedia in nature, including digital images, video, audio, graphics, and text data. In order to make use of this vast amount of data, efficient and effective techniques to retrieve multimedia information based on its content need to be developed. Among the various media types, images are of prime importance. Not only it is the most widely used media type besides text, but it is also the basis for representing and retrieving videos and other multimedia information. This paper deals with the retrieval of images based on their contents, even though the approach is readily generalizable to other media types.

Keyword annotation is the traditional image retrieval paradigm. In this approach, the images are first annotated manually by keywords. They can then be retrieved by their corresponding annotations. However, there are three main difficulties with this approach, i.e. the large amount of manual effort required in developing the annotations, the differences in interpretation of image contents, and inconsistency of the keyword assignments among different indexers [1, 2, 3]. As the size of image repositories increases, keyword annotation approach becomes infeasible.

*This work was supported in part by NSF/DARPA/NASA DLI Program under Cooperative Agreement 94-11318, in part by ARL Cooperative Agreement No. DAAL01-96-2-0003, and in part by a CSE Fellowship, UIUC.

To overcome the difficulties of the annotation based approach, an alternative mechanism, Content-Based Image Retrieval (CBIR), has been proposed in the early 1990's. Besides using human-assigned keywords, CBIR systems use the visual content of the images, such as color, texture, and shape features, as the image index. This greatly alleviate the difficulties of the pure annotation based approach, since the feature extraction process can be made automatic and the image's own content is always consistent. Since its advent, CBIR has attracted great research attention, ranging from government [4, 5], industry [2, 6, 7], to universities [8, 9, 10, 11, 12]. Even ISO/IEC has recently launched a new work item, MPEG-7 [13, 14, 15, 16], to define a standard Multimedia Content Description Interface. Many special issues from leading journals have been dedicated to CBIR [17, 18, 19, 20] and many CBIR systems, both commercial [1, 2, 3, 6, 7] and academic [8, 9, 10, 11, 12], have been developed recently.

Despite the extensive research effort, the retrieval techniques used in CBIR systems lag behind the corresponding techniques in today's best text search engines, such as Yahoo, Alta Vista, Lycos, etc. At the early stage of CBIR, research primarily focused on exploring various feature representations, hoping to find a "best" representation for each feature. For example, for texture feature alone, almost a dozen representations have been proposed [21], including Tamura [22], MSAR [23], Word decomposition [24], Fractal [25], Gabor Filter [26, 11], and Wavelets [27, 28, 12], etc. The corresponding system design strategy for early CBIR systems is to first find the "best" representations for the visual features. Then,

- During the retrieval process, the user selects the visual feature(s) that he or she is interested in. In the case of multiple features, the user has to also specify the weights for the representations.
- Based on the selected features and specified weights, the retrieval system tries to find similar images to user's query.

We refer such systems as *computer centric* systems, since they do not involve human input in the retrieval process. While the *computer centric* approach establishes the basis of CBIR, the performance is not satisfactory due to the following two reasons:

- *The gap between high level concepts and low level features*
  The assumption that the *computer centric* approach makes is that the high level concepts to low level features mapping is easy for the user to do. While in some cases the assumption is true, e.g. mapping a high level concept (fresh apple) to low level features (color and shape), in other cases, this may not be true. One example is to map an ancient vase with sophisticated design to an equivalent representation using low level features. The gap exists between the two levels.
- *The subjectivity of human perception*
  Different persons, or the same person under different circumstances, may perceive the same visual content differently. This is called *human perception subjectivity.* The subjectivity exists in various levels. For example, one person may be more interested in an image's color feature while another may be more interested in the texture feature. Even if both people are interested in texture, the way how they perceive the similarity of texture may be quite different. This is illustrated in Figure 1.

(a) (b) (c)

Figure 1: Subjectivity in perceiving texture feature

Among the above three texture images, some may say (a) and (b) are more similar if they do not care for the intensity contrast, while others may say (a) and (c) are more similar if they ignore the local property on the seeds. No single texture representation can capture everything. Different representations capture the visual feature from different angles.

In the *computer centric* approach, the "best" representations and weights are fixed, which can not effectively model high level concepts and user's perception subjectivity. Furthermore, specification of weights imposes a big burden on the user, as it requires the user to have a comprehensive knowledge of the low level feature representations used in the retrieval system, which is normally not the case.

Motivated by the limitations of the *computer centric* approach, recently research focus in CBIR has moved to interactive mechanism that involves human as part of the retrieval process [21, 29, 30]. Examples include *interactive* region segmentation [31]; interactive image annotation [29, 32]; usage of *supervised* learning before the retrieval [33, 34]; and *interactive* integration of keywords and high level concepts to enhance image retrieval performance [10, 35].

In this paper, to overcome the difficulties faced by *computer centric* approach, we present a *Relevance Feedback* based approach to CBIR, in which human and computer interact to refine high level queries to representations based on low level features. Relevance feedback is a powerful concept used in traditional text-based Information Retrieval systems. It is the process of automatically adjusting an existing query using the information fed-back by the user about the relevance of previously retrieved objects such that the adjusted query is a better approximation to the user's information need [36, 37, 38]. In the relevance feedback based approach [39, 40, 41], the retrieval process is *interactive* between the computer and human. The burden of concept mapping and specification of weights are removed from the user. The user only need to mark which images he or she thinks are relevant to the query. The weights embedded in the query object are *dynamically* updated to model the high level concepts and perception subjectivity.

The remaining of the paper is organized as follows. In section 2, a multimedia object model is introduced, which supports multiple features, multiple representations, and their corresponding weights. The weights are essential in modeling high level concepts and perception subjectivity. Section 3 discusses how the weights are *dynamically* updated based on the relevance feedback to track user's information need. Experimental results of the effectiveness of the retrieval algorithm are given in Section 4. Concluding remarks are given in Section 5.

2 The Multimedia Object Model

Before we describe how the relevance feedback technique can be used for CBIR, we first need to formalize how an image object is modeled. An image object O is represented as:

$$O = O(D, F, R) \tag{1}$$

- D is the raw image data, e.g. a JPEG image.
- $F = \{f_i\}$ is a set of low-level visual features associated with the image object, such as color, texture, and shape.
- $R = \{r_{ij}\}$ is a set of representations for a given feature f_i , e.g. both color histogram and color moments are representations for the color feature [42]. Note that, each representation r_{ij} itself may be a vector consisting of multiple components, i.e.

$$r_{ij} = [r_{ij1}, ..., r_{ijk}, ..., r_{ijK}] \tag{2}$$

where K is the length of the vector.

In contrast to the *computer centric* approach's single representation and fixed weights, the proposed object model supports multiple representations with dynamically updated weights to accommodate the rich content in the image objects. Different weights, W_i, W_{ij}, and W_{ijk}, are associated with features f_i, representations r_{ij}, and components r_{ijk}, respectively. The goal of relevance feedback, described in next section, is to find the appropriate weights to model user's information need.

Further notice that a query Q has the same model as that of the image objects, since it is also an image object in nature.

3 Integrating Relevance Feedbacks in CBIR

An image object model O, together with a set of similarity measures $M = \{m_{ij}\}$, specifies a CBIR model (D, F, R, M). The similarity measures are used to determine how similar or dissimilar two objects are. Different similarity measures may be used for different feature representations. For example, Euclidean is used for comparing vector-based representations while Histogram Intersection is used for comparing color histogram representations.

Based on the image object model and the set of the similarity measures, the retrieval process is described below and also illustrated in Figure 2.

1. Initialize the weights $W = [W_i, W_{ij}, W_{ijk}]$ to $W0$, which is a set of no-bias weights. That is, every entity are initially of the same importance.

$$W_i = W0_i = \frac{1}{I} \tag{3}$$

$$W_{ij} = W0_{ij} = \frac{1}{J_i} \tag{4}$$

$$W_{ijk} = W0_{ijk} = \frac{1}{K_{ij}} \tag{5}$$

where I is the number of features in set F; J_i is the number of representations for feature f_i; K_{ij} is the length of the presentation vector r_{ij}.

2. The user's information need, represented by the query object Q, is distributed among different features f_i, according to their corresponding weights W_i.
3. Within each feature f_i, the information need is further distributed among different feature representations r_{ij}, according to the weights W_{ij}.
4. The objects' similarity to the query, in terms of r_{ij}, is calculated according to the corresponding similarity measure m_{ij} and the weights W_{ijk}:

$$S(r_{ij}) = m_{ij}(r_{ij}, W_{ijk}) \tag{6}$$

5. Each representation's similarity values are then combined into a feature's similarity value:

$$S(f_i) = \sum_j W_{ij} S(r_{ij}) \tag{7}$$

6. The overall similarity S is obtained by combining individual $S(f_i)$'s:

$$S = \sum_i W_i S(f_i) \tag{8}$$

7. The objects in the database are ordered by their overall similarity to Q. The N_{RT} most similar ones are returned to the user, where N_{RT} is the number of objects the user wants to retrieve.
8. For each of the retrieved objects, the user marks it as *highly relevant*, *relevant*, *no-opinion*, *non-relevant*, or *highly non-relevant*, according to his information need and perception subjectivity.
9. The system updates the weights according to the user's feedback such that the adjusted Q is a better approximation to the user's information need.
10. Go to Step 2 with the adjusted Q and start a new iteration of retrieval.

In Figure 2, the information need embedded in Q flows up while the content of O's flows down. They meet at the dashed line, where the similarity measures m_{ij} are applied to calculate the similarity values $S(r_{ij})$'s between Q and O's.

Following the Information Retrieval theories [37, 36, 38], the objects stored in the database are considered *objective* and their weights are fixed. Whether the query is considered *objective* or *subjective* and whether its weights can be updated distinguish the proposed relevance feedback approach from the *computer centric* approach. In the *computer centric* approach, query is considered *objective*, the same as the objects stored in the database, and its weights are fixed. Because of the fixed weights, this approach can not effectively model high level concepts and human perception subjectivity. It requires the user to specify a precise set of weights at the query stage, which is normally not possible. On the

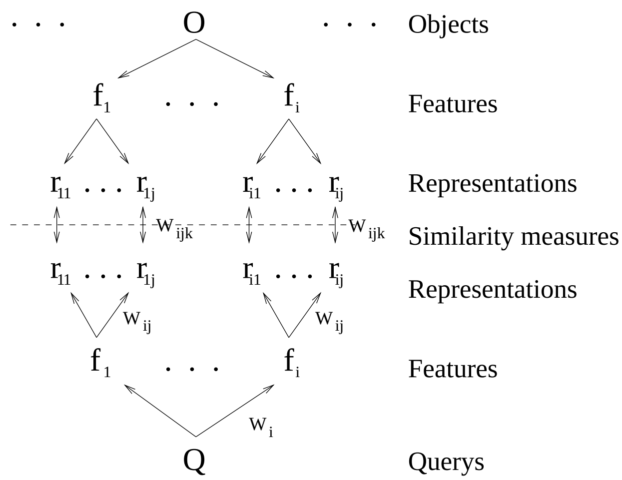

Figure 2: The retrieval process

other hand, queries in the proposed approach are considered as *subjective*. That is, during the retrieval process, the weights associated with the query can be *dynamically* updated via relevance feedback to reflect the user's information need. The burden of specifying the weights is removed from the user.

Note that in the proposed retrieval algorithm, both S and $S(f_i)$ are linear combinations of their corresponding lower level similarities. The basis of the linear combination is that weights are proportional to the entities' relative importance [43]. For example, if a user cares twice as much about one feature (color) as he does about another feature (shape), the overall similarity would be a linear combination of the two individual similarities with the weights being 2/3 and 1/3 respectively. Furthermore, because of the nature of linearity, these two levels can be combined into one, i.e.:

$$S = \sum_i \sum_j W_{ij} S(r_{ij}) \tag{9}$$

where W_{ij}'s are now re-defined to be the weights by which the information need in Q is distributed directly into r_{ij}'s.

3.1 Update of W_{ij}

The W_{ij}'s associated with the r_{ij}'s reflect the user's different emphasis of a representation in the overall similarity. The support of different weights enables the user to specify his or her information need more precisely. We will next discuss how to update W_{ij}'s according to user's relevance feedback.

Let RT be the set of the most similar N_{RT} objects according to the overall similarity value S:

$$RT = [RT_1, ..., RT_l, ..., RT_{N_{RT}}] \tag{10}$$

Let $Score$ be the set containing the relevance scores fed-back by the user for RT_l's:

$$\begin{aligned} & = 3, \quad \text{if highly relevant} && (11)\\ & = 1, \quad \text{if relevant} && (12)\\ Score_l & = 0, \quad \text{if no-opinion} && (13)\\ & = -1, \quad \text{if non-relevant} && (14)\\ & = -3, \quad \text{if highly non-relevant} && (15) \end{aligned}$$

The choice of 3, 1, 0, -1, and -3 as the scores is arbitrary. Experimentally we find the above scores capture the semantic meaning of *highly relevant*, *relevant*, etc. In Equations (11-15), we provide the user with 5 levels of relevance. Although the more the levels, the more accurate is the feedback, it is less convenient for the user to interact with the systems. Experimentally we find that 5 levels is a good trade-off between convenience and accuracy.

For each r_{ij}, let RT^{ij} be the set containing the most similar N_{RT} objects to the query Q, according to the similarity values $S(r_{ij})$:

$$RT^{ij} = [RT_1^{ij}, ..., RT_l^{ij}, ..., RT_{N_{RT}}^{ij}] \tag{16}$$

To calculate the weight for r_{ij}, first initialize $W_{ij} = 0$, and then use the following procedure:

$$\begin{aligned} W_{ij} &= W_{ij} + Score_l, \quad \text{if } RT_l^{ij} \text{ is in } RT & (17) \\ &= W_{ij} + 0, \quad \text{if } RT_l^{ij} \text{ is not in } RT & (18) \\ l &= 0, ..., N_{RT} & (19) \end{aligned}$$

Here, we consider all the images outside RT as marked with *no-opinion* and have the score of 0. After this procedure, if $W_{ij} < 0$, set it to 0. Let $W_{Tij} = \sum W_{ij}$ be the total weights. The raw weights obtained by the above procedure is then normalized by the total weight to make the sum of the normalized weight equal to 1.

$$W_{ij} = \frac{W_{ij}}{W_{Tij}} \tag{20}$$

As we can see, the more the overlap of relevant objects between RT and RT^{ij}, the larger the weight of W_{ij}. That is, if a representation r_{ij} reflects the user's information need, it receives more emphasis.

3.2 Update of W_{ijk}

The W_{ijk}'s associated with r_{ijk}'s reflect the different contributions of the components to the representation vector r_{ij}. For example, in the wavelet texture representation, we know that the mean of a sub-band may be corrupted by the lighting condition, while the standard deviation of a sub-band is independent of the lighting condition. Therefore more weight should be given to the standard deviation component, and less weight to the mean component. The support of different weights enables the system to have more reliable feature representation and thus better retrieval performance.

A standard deviation based weight updating approach has been proposed in our previous work[39]. Out of the N_{RT} returned objects, for those objects that are marked with *highly relevant* or *relevant* by the user, stack their representation vector r_{ij}'s to form a $M' \times K$ matrix, where M' is the number of objects marked with *highly relevant* or *relevant*. In this way, each column of the matrix is a length-M' sequence of r_{ijk}'s. Intuitively, if all the relevant objects have similar values for the component r_{ijk}, it means that the component r_{ijk} is a good indicator of the user's information need. On the other hand, if the values for the component r_{ijk} are very different among the relevant objects, then r_{ijk} is not a good indicator. Based on this analysis, the inverse of the standard deviation of the r_{ijk} sequence is a good estimation of the weight W_{ijk} for component r_{ijk}. That is, the smaller the variance, the larger the weight and vice versa.

$$W_{ijk} = \frac{1}{\sigma_{ijk}} \tag{21}$$

where σ_{ijk} is the standard deviation of the length-M' sequence of r_{ijk}'s. Again, just as in Equation (20), we need to normalize W_{ijk}'s in the same way.

$$W_{ijk} = \frac{W_{ijk}}{W_{Tijk}} \tag{22}$$

where $W_{Tijk} = \sum W_{ijk}$.

3.3 Summary

Based on the above description of the relevance feedback algorithm, we briefly summarize its properties here.

- *Multi-modality*
  The proposed the image object model, and therefore the retrieval model, supports multiple features and multiple representations. In contrast to *computer centric* approach's attempt of finding the single best universal feature representation, the proposed approach concentrates on the way of organizing multiple feature representations, such that appropriate feature representations are invoked (emphasized) at the right time and place. The *multi-modality* approach allows the system to better model user's perception subjectivity.
- *Interactivity*
  In contrast to *computer-centric* approach's *automated* system, the proposed approach is *interactive* in nature. The interactivity allows the system to make use of the ability both from computer and from human.
- *Dynamic*
  In contrast to *computer-centric* approach's fixed query weights, the proposed approach *dynamically* updates the query weights via relevance feedback. The advantages are twofold:
  - *Remove burden from the user*
    The user is no longer required to specify a precise set of weights at the query formulation stage. Instead, the user interacts with the system, indicating which returns he or she thinks are relevant. Based on the user's feedback, query weights are dynamically updated.
  - *Remove burden from the computer*
    The computer is no longer required to understand the high level concept. Based on user's feedback, the high level concept embedded in the query weights automatically get refined.

4 Experimental Results

To address the challenging research issues involved in CBIR, a Multimedia Analysis and Retrieval System (MARS) project has been started at University of Illinois[9, 44, 45, 46, 40, 39, 41, 47]. MARS-1 is accessible via internet at http:// jadzia.ifp.uiuc.edu:8000. The relevance feedback architecture proposed in this paper is currently being integrated into MARS-2.

In the experiments reported here, the image database is provided by Fowler Museum of Cultural History at the University of California-Los Angeles. The image database is part of the Museum Educational Site Licensing Project (MESL), sponsored by the Getty Information Institute.

The test database contains 286 images. The reason that we have chosen this test set is that it allows us to explore all the color, texture, and shape features simultaneously in a meaningful way. Even though there exist larger test sets, they do not provide the ability to test multiple visual features simultaneously. In our current retrieval system, the visual features used include color, texture and shape of the objects in the image. That is,

$$F = \{f_i\} = \{\text{color, texture, shape}\} \tag{23}$$

To validate the proposed approach, multiple representations are used for each feature, e.g. color histogram and color moments [42] are used for color feature; Tamura [48, 22] and co-occurrence matrix [49, 50] texture representations are used for texture feature; Fourier descriptor and chamfer shape descriptor [40] are used for shape feature.

$$\begin{aligned} R = \{r_{ij}\} &= \{r_1, r_2, r_3, r_4, r_5, r_6\} \\ &= \{\text{color histogram, color moments, Tamura, co-occurrence matrix,} \\ &\quad \text{Fourier descriptor, chamfer shape descriptor}\} \end{aligned}$$

The proposed relevance feedback architecture is an *open* retrieval architecture. Other visual features or feature representations can be easily incorporated, if needed.

Extensive experiments have been carried out to evaluate the system's performance. Users from various disciplines, such as Computer Vision, Art, Library Science, etc., as well as users from industry, have been invited to compare the retrieval performance between the proposed *interactive* approach and the *computer centric* approach. All the users rated the proposed approach much higher in terms of capturing their perception subjectivity and information need. A typical retrieval process is given in Figures 3 and 4.

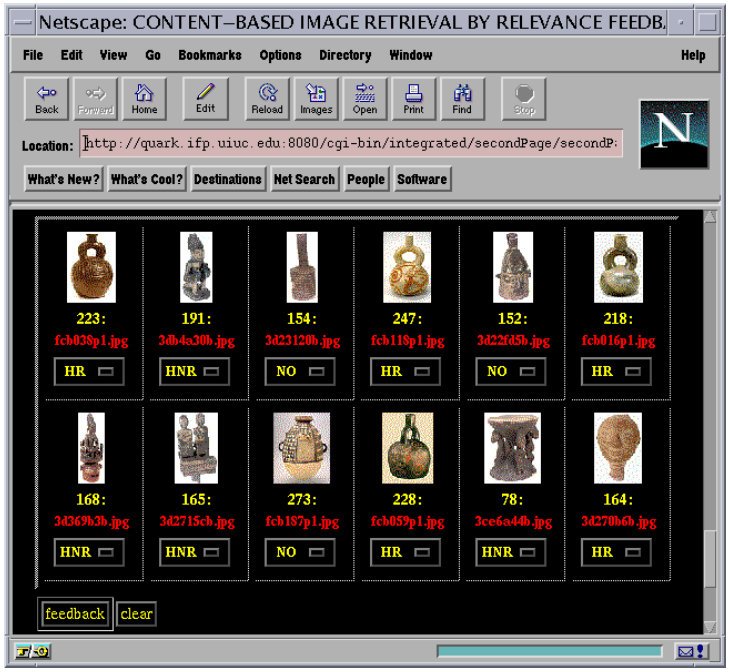

Figure 3: The initial retrieval results

The user can browse through the image database. Once he or she finds an image of interest, that image is submitted as a query. Alternating to this query-by-example mode, the user can also submit images outside the database as queries. In Figure 3, the query image is displayed at the upper-left corner and the best 11 retrieved images, with $W = W0$, are displayed in the order from top to bottom and from left to right. The retrieved results are obtained based on their overall similarities to the query image, which are computed from all the features and all the representations. Some retrieved images are similar to the query image in terms of shape feature while other are similar to the query image in terms of color or texture feature.

Assume the user's true information need is to "retrieve similar images based on their shapes". In the proposed retrieval approach, the user is no longer required to explicitly maps his information need to low-level features, but rather he or she can express his intended information need by marking the relevance scores of the returned images. In this example, images 247, 218, 228 and 164 are marked *highly relevant*. Images 191, 168, 165, and 78 are marked *highly non-relevant*. Images 154, 152, and 273 are marked *no-opinion*.

Based on the information fed-back by the user, the system *dynamically* adjusts the weights, putting more emphasis on the *shape feature*, possibly even more emphasis to one of the two shape representations which matches user's perception subjectivity of shape. The improved retrieval results are displayed in Figure 4. Note that our shape representations are invariant to translation, rotation, and scaling. Therefore, images 164 and 96 are relevant to the query image.

Unlike the *computer centric* approach, where the user has to precisely decompose his information need into different features and representations and precisely specify all the weights associated with them, the proposed *interactive* approach allows the user to submit a coarse initial query and continuously refine his information need via relevance feedback. This approach greatly reduces the user's effort of composing a query and captures the user's information need more precisely.

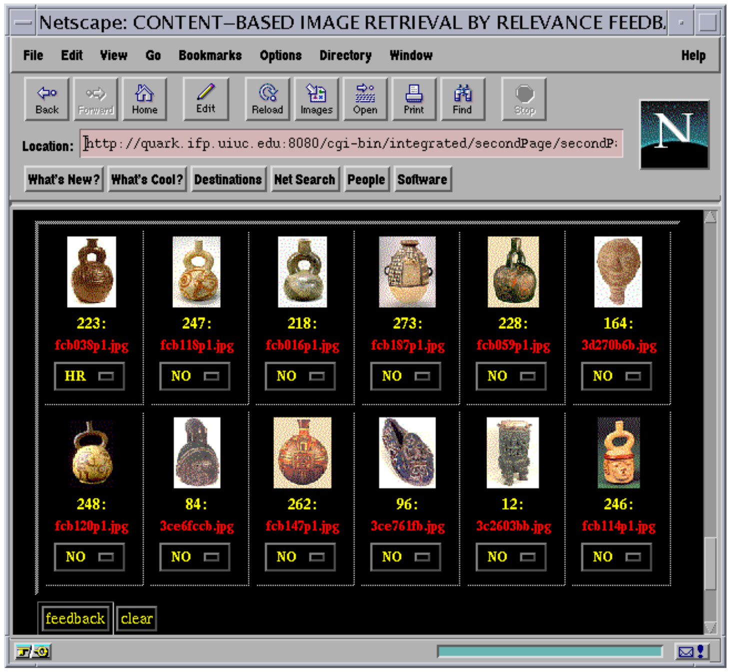

Figure 4: The retrieval results after the relevance feedback

5 Conclusions

CBIR has emerged as one of the most active research areas in the past few years. Most of the early research effort focused on finding the "best" image feature representations. Retrieval was performed as summation of similarities of individual feature representation with fixed weights. While this *computer centric* approach establishes the basis of CBIR, the usefulness of such systems was limited due to the difficulty in representing high level concepts using low level features and human perception subjectivity.

In this paper, we introduce a Human-Computer Interaction approach to CBIR based on relevance feedback. Unlike the *computer centric* approach, where the user has to precisely decompose his information need into different feature representations and precisely specify all the weights associated with them, the proposed *interactive* approach allows the user to submit a coarse initial query and continuously refine his information need via relevance feedback. This approach greatly reduces the user's effort of composing a query and captures the user's information need more precisely. Furthermore, the efficiency and effectiveness of the proposed approach have been validated by large amount of experiments.

Although the proposed retrieval model is for CBIR, it can be easily expanded to handle other media types, such as video and audio. The proposed model also has a close relationship to MPEG-7, as discussed in our previous MPEG-7 proposal [16]. Furthermore, the proposed model provides a natural way of combining keyword features with visual features. We envision the importance of supporting keywords with visual features and are currently expanding our system to handle this.

References

[1] W. Niblack, R. Barber, and et al., "The QBIC project: Querying images by content using color, texture and shape," in *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*, Feb 1994.

[2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafine, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Computer*, 1995.

[3] C. Faloutsos, M. Flickner, W. Niblack, D. Petkovic, W. Equitz, and R. Barber, "Efficient and effective querying by image content," tech. rep., IBM Research Report, 1993.

[4] R. Jain, "Workshop report: NSF workshop on visual information management systems," in *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*, 1993.

[5] R. Jain, A. Pentland, and D. Petkovic, "NSF-ARPA workshop on visual information management systems," (Cambridge, MA), June 1995.

[6] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, and C. fe Shu, "The Virage image search engine: An open framework for image management," in *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*

[7] J. Dowe, "Content-based retrieval in multimedia imaging," in *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*, 1993.

[8] A. Pentland, R. Picard, and S.Sclaroff, "Photobook: Content-based manipulation of image databases," *International Journal of Computer Vision*, 1996.

[9] T. S. Huang, S. Mehrotra, and K. Ramchandran, "Multimedia analysis and retrieval system (MARS) project," in *Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.

[10] J. R. Smith and S.-F. Chang, "Searching for images and video on the world-wide web," Tech. Rep. 459-96-25, Columbia Univ., 1996.

[11] W.Y.Ma and B.S.Manjunath, "Netra: A toolbox for navigating large image databases," in *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[12] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Transactions on Consumer Electronics*, vol. 42, pp. 557–565, Aug 1996.

[13] "MPEG-7: Context and objectives (v.4 - Stockholm)," *ISO/IEC JTC1/SC29/WG11 N1733, MPEG97*, July 1997.

[14] "MPEG-7 applications document," *ISO/IEC JTC1/SC29/WG11 N1734, MPEG97*, July 1997.

[15] "Second draft of MPEG-7 requirements," *ISO/IEC JTC1/SC29/WG11 N1735, MPEG97*, July 1997.

[16] Y. Rui, T. S. Huang, and S. Mehrotra, "Mars and its applications to MPEG-7," *ISO/IEC JTC1/SC29/WG11 M2290, MPEG97*, July 1997.

[17] V. N. Gudivada and J. V. Raghavan, "Introduction: Content-based image retrieval systems," *Computer*, 1995.

[18] A. Pentland and R. Picard, "Special issue on digital libraries," *IEEE Trans. Patt. Recog. and Mach. Intell.*, 1996.

[19] A. D. Narasimhalu, "Special section on content-based retrieval," *Multimedia Systems*, 1995.

[20] B. Schatz and H. Chen, "Building large-scale digital libraries," *Computer*, 1996.

[21] T. S. Huang and Y. Rui, "Image retrieval: Past, present, and future," in *Proc. of Int. Symposium on Multimedia Information Processing*, Dec 1997.

[22] W. Equitz and W. Niblack, "Retrieving images from a database using texture – algorithms from the QBIC system," Tech. Rep. RJ 9805, Computer Science, IBM Research Report, May 1994.

[23] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognition*, vol. 25, no. 2, pp. 173–188, 1992.

[24] F. Liu and R. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 18, July 1996.

[25] B. Cheng, "Approaches to image retrieval based on compressed data for multimedia database systems," *PhD Thesis, University of New York at Buffalo*, 1996.

[26] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE T-PAMI special issue on Digital Libraries*, Nov. 1996.

[27] J. R. Smith and S.-F. Chang, "Transform features for texture classification and discrimination in large image databases," in *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[28] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Proc.*, vol. 2, pp. 429–441, October 1993.

[29] R. Picard and T.P.Minka, "Vision texture for annotation," *Multimedia Systems: Special Issue on Content-based retrieval*.

[30] S. Sclaroff, L. Taycher, and M. L. Cascia, "Imagerover: A content-based image browser for the world wide web," in *Proc IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[31] D. Daneels, D. Campenhout, W. Niblack, W. Equitz, R. Barber, E. Bellon, and F. Fierens, "Interactive outlining: An improved approach using active contours," in *Proc. SPIE Conf. on Vis. Commun. and Image Proc.*, 1993.

[32] T.P.Minka and R.W.Picard, "Interactive learning using a "society of models"," in *Proc IEEE CVPR*, pp. 447–452, 1996.

[33] B.S.Manjunath and W.Y.Ma, "Image indexing using a texture dictionary," in *Proceedings of SPIE conference on Image Storage and Archiving System*, vol. 2606.

[34] W.Y.Ma and B.S.Manjunath, "Texture features and learning similarity," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 425–430, 1996.

[35] J. R. Smith and S.-F. Chang, "An image and video search engine for the world-wide web," tech. rep., Columbia Univ.

[36] W. M. Shaw, "Term-relevance computations and perfect retrieval performance," *Information Processing and Management*.

[37] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, 1983.

[38] C. Buckley and G. Salton, "Optimization of relevance feedback weights," in *Proc. of SIGIR'95*, 1995.

[39] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[40] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega, "Automatic matching tool selection using relevance feedback in MARS," in *Proc. of 2nd Int. Conf. on Visual Information Systems*, 1997.

[41] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega, "A relevance feedback architecture in content-based multimedia information retrieval systems," in *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97*, 1997.

[42] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, 1991.

[43] R. Fagin and E. L. Wimmers, "Incorporating user preferences in multimedia queries," in *Proc of Int. Conf. on Database Theory*, 1997.

[44] S. Mehrotra, Y. Rui, K. Chakrabarti, M. Ortega, and T. S. Huang, "Multimedia analysis and retrieval system," in *Proc. of The 3rd Int. Workshop on Information Retrieval Systems*, 1997.

[45] S. Mehrotra, Y. Rui, M. Ortega-B., and T. S. Huang, "Supporting content-based queries over images in MARS," in *Proc. of IEEE Int. Conf. on Multimedia Computing and Systems*, 1997.

[46] M. Ortega, Y. Rui, K. Chakrabarti, , S. Mehrotra, and T. S. Huang, "Supporting similarity queries in MARS," in *Proc. of ACM Conf. on Multimedia*, 1997.

[47] Y. Rui, A. C. She, and T. S. Huang, "Modified fourier descriptors for shape representation – a practical approach," in *Proc of First International Workshop on Image Databases and Multi Media Search*, 1996.

[48] H. Tamura, S. Mori, and T. Yamawaki, “Texture features corresponding to visual perception,” *IEEE Trans. on Sys, Man, and Cyb*, vol. SMC-8, no. 6, 1978.

[49] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Texture features for image classification,” *IEEE Trans. on Sys, Man, and Cyb*, vol. SMC-3, no. 6, 1973.

[50] P. P. Ohanian and R. C. Dubes, “Performance evaluation for four classes of texture features,” *Pattern Recognition*, vol. 25, no. 8, pp. 819–833, 1992.