Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval

Yong Rui, Thomas S. Huang, Michael Ortega and Sharad Mehrotra

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 feedback. The experimental results over more than 70,000 images show that the proposed approach greatly reduces the user's effort of composing a query and captures the user's information need more precisely.

Keywords— **Content-Based Image Retrieval, interactive multimedia processing, relevance feedback**

I. Introduction

With 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 one of the most widely used bases 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, the 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, in part by NSF CISE Research Infrastructure Grant CDA-9624396. Yong Rui was also supported in part by a CSE Fellowship, the University of Illinois. Michael Ortega was also supported in part by CONACYT grant 89061.

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 alleviates 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 launched a new work item, MPEG-7 [13], [14], [15], to define a standard Multimedia Content Description Interface. Many special issues from leading journals have been dedicated to CBIR [16], [17], [18], [19] 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 Inquery [20], 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 the 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 needs to also specify the weights for each of the features.
- Based on the selected features and specified weights, the retrieval system tries to find similar images to the user's query.

We refer to such systems as *computer centric* systems. While this 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* [29]. The subjectivity exists at 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.
Among the above three texture images, some may say that (a) and (b) are more similar if they do not care for the intensity contrast, while others may say that (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 perspectives.

In the *computer centric* approach, the "best" features and representations and their corresponding 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, recent research focus in CBIR has moved to an interactive mechanism that involves a human as part of the retrieval process [21], [30], [31], [32]. Examples include *interactive* region segmentation [33]; interactive image database annotation [31], [34]; usage of *supervised* learning before the retrieval [35], [36]; and *interactive* integration of keywords and high level concepts to enhance image retrieval performance [10], [37].

In this paper, to address the difficulties faced by the *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 technique 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 [38], [39], [40]. In the relevance feedback based approach [30], [41], [42], [29], the retrieval process is *interactive* between the computer and human. Under the assumption that high-level concepts can be captured by low-level features, the relevance feedback technique tries to establish the link between high-level concepts and low-level features from the user's feedback. Furthermore, the burden of specifying the weights is removed from the user. The user only needs 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 rest of the paper is organized as follows. Section 2 introduces a multimedia object model 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 the user's information need. Sections 4 and 5 discuss the normalization procedure and dynamic weight updating process, the two bases of the retrieval algorithm. Extensive experimental results over more than 70,000 images for testing both the efficiency and the effectiveness of the retrieval algorithm are given in Section 6. Concluding remarks are given in Section 7.

II. 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 [29]. 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 [43]. 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. Weights exist at various levels. 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 the next section, is to find the appropriate weights to model the user's information need.

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

III. Integrating Relevance Feedback in CBIR

An image object model $O(D, F, R)$, 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 similarity measures, the retrieval process is described below and also illustrated in Figure 2.

(a) (b) (c)

Fig. 1. Subjectivity in perceiving the texture feature

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 is initially of the same importance.

$$W_i = W0_i = \frac{1}{I} \quad (3)$$

$$W_{ij} = W0_{ij} = \frac{1}{J_i} \quad (4)$$

$$W_{ijk} = W0_{ijk} = \frac{1}{K_{ij}} \quad (5)$$

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}:

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

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

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

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

$$S = \sum_i W_i S(f_i) \quad (8)$$

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 (described in Section 5) 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.

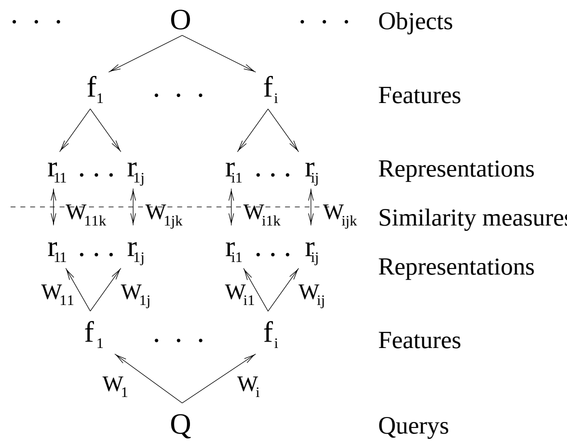

Fig. 2. The retrieval process

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 [39], [38], [40], 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 distinguishes the proposed relevance feedback approach from the *computer centric* approach. In the *computer centric* approach, a 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 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 the weights are proportional to the entities' relative importance [44]. 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[44]. 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. Note that it is not possible to absorb W_{ijk} into W_{ij}, since the calculation of $S(r_{ij})$ can be a non-linear function of W_{ijk}'s, such as Euclidean or Histogram Intersection.

In the next two sections, we will discuss two key components of this retrieval algorithm, i.e. *normalization* and *weight updating*.

IV. Normalization

In the retrieval algorithm described in the previous section, we have assumed that the similarity values of each representation, $S(r_{ij})$'s, are of the same dynamic range, say, from 0 to 1. Otherwise, the linear combination of $S(r_{ij})$'s to form S (Equation (9)) becomes meaningless. One $S(r_{ij})$ may overshadow the others just because its magnitude is large. For the same reason, when calculating $S(r_{ij})$'s, the vector components, r_{ijk}'s, should also be normalized before applying the similarity measure m_{ij}. We refer the normalization of r_{ijk}'s as intra-normalization and the normalization of $S(r_{ij})$'s as inter-normalization [45], [46].

A. Intra-Normalization

This normalization process puts equal emphasis on each component, r_{ijk}, within a representation vector r_{ij}. To see the importance of this, note that different components within a vector may be of totally different physical quantities. Their magnitudes can vary drastically, thereby biasing the similarity measure.

To simplify the notations, define $V = r_{ij}$. That is, every representation vector r_{ij} will go through the following normalization procedure.

Assume there are M images in the database and let m be the image id index, then

$$V = V_m = [V_{m,1}, V_{m,2}, ..., V_{m,k}, ..., V_{m,K}] \tag{10}$$

is the representation vector for image m, where K is the length of vector $V = r_{ij}$. If we put all V_m's into a matrix form, we have a $M \times K$ matrix

$$\mathcal{V} = [\mathcal{V}_{m,k}],\ m = 1, ..., M,\ k = 1, ..., K \tag{11}$$

where $\mathcal{V}_{m,k}$ is the k^{th} component in vector V_m. Now, the k^{th} column of matrix $\mathcal{V}$ is a length-M sequence. Denote this sequence as $\mathcal{V}_k$. Our goal is to normalize the entries in each column to the same range so as to ensure that each individual component receives equal emphasis when calculating the similarity between two vectors. One way of normalizing the sequence $\mathcal{V}_k$ is to find the maximum and minimum values of $\mathcal{V}_k$ and normalize the sequence to [0, 1] as follows:

$$\mathcal{V}_{m,k} = \frac{\mathcal{V}_{m,k} - min_k}{max_k - min_k}, \tag{12}$$

where min_k and max_k refer to the smallest and the biggest values in the sequence $\mathcal{V}_k$. Although simple, this is not a desirable normalization procedure. Let's consider a sequence {1.0, 1.1, 1.2, 1.3, 100.0}. If we use Equation (12) to normalize the sequence, most of the [0, 1] range will be taken away by a single entry 100.0, and most of the information in {1.0, 1.1, 1.2, 1.3} is warped into a very narrow range.

A better approach is to use the Gaussian normalization. Assuming the sequence $\mathcal{V}_k$ to be a Gaussian sequence, we compute the mean μ_k and standard deviation σ_k of the sequence. We then normalize the original sequence to a N(0,1) sequence as follows:

$$\mathcal{V}_{m,k} = \frac{\mathcal{V}_{m,k} - \mu_k}{\sigma_k} \tag{13}$$

It is easy to prove that after the normalization according to Equation (13), the probability of an entry's value being in the range of [-1, 1] is 68%. If we use $3\sigma_k$ in the denominator, according to the 3-σ rule, the probability of an entry's value being in the range of [-1, 1] is approximately 99%. In practice, we can consider all the entry values to be within the range of [-1,1] by mapping the out-of-range values to either -1 or 1. The advantage of this normalization process over Equation (12) is that the presence of a few abnormally large or small values, such as the 100.0 entry in the example sequence, *does not* bias the importance of a component r_{ijk} in computing the similarity between vectors.

B. Inter-Normalization

Intra-normalization procedure ensures the equal emphasis of each component r_{ijk} within a representation vector r_{ij}. On the other hand, the inter-normalization procedure ensures equal emphasis of each individual similarity value $S(r_{ij})$ within the overall similarity value S.

Depending on the similarity measure m_{ij} used, the values of $S(r_{ij})$'s can be of quite different dynamic ranges. In order to ensure that no single $S(r_{ij})$ will overshadow the others only because it has a larger magnitude, inter-normalization should be applied. This procedure is summarized as follows:

1. For any pair of images I_m and I_n in the image collection, compute their similarity $S_{m,n}(r_{ij})$:

$$S_{m,n}(r_{ij}) = m_{ij}(r_{ij}, W_{ijk}) \tag{14}$$

$$\begin{aligned} m, n &= 1, ..., M \\ m &\neq n \end{aligned}$$

2. Since there are M images, there are $C_2^M = \frac{M\times(M-1)}{2}$ possible similarity values between any pair of images. Treat them as a data sequence and find the mean μ_{ij} and standard deviation σ_{ij} of the sequence. Store μ_{ij} and σ_{ij} in the database to be used in later normalization.
3. When a query Q is presented, compute the raw (un-normalized) similarity values between Q and the images in the database.

$$S_{m,Q}(r_{ij}) = m_{ij}(r_{ij}, W_{ijk}) \tag{15}$$

4. Normalize the raw similarity values as follows:

$$S'_{m,Q}(r_{ij}) = \frac{S_{m,Q}(r_{ij}) - \mu_{ij}}{3\sigma_{ij}} \tag{16}$$

As explained in the intra-normalization sub-section, this Gaussian normalization procedure ensures that 99% of all the $S'_{m,Q}(r_{ij})$ values will be within the range of [-1,1]. An additional shift will guarantee that 99% of similarity values are within [0,1]:

$$S''_{m,Q}(r_{ij}) = \frac{S'_{m,Q}(r_{ij}) + 1}{2} \tag{17}$$

After this shift, in practice, we can consider all the values to be within the range of [0,1], since an image whose distance from the query is greater than 1 is very dissimilar and can be disregarded without affecting the retrieval results.

In the above normalization process, the first two steps are done off-line to obtain μ_{ij} and σ_{ij}. The last two steps are done on-line to convert the un-normalized value to normalized ones, by using the pre-calculated statistics μ_{ij} and σ_{ij}.

The above described normalization process assumes that M is large enough, such that μ_{ij} and σ_{ij} calculated based on C_2^M similarity values approximate the true mean and standard deviation of the distribution of all possible $S_{m,n}(r_{ij})$'s by the *Law of Large Number* (LLN) [47]. This assumption is important since it ensures that we can use Equation (16) to normalize the similarity value between an image and a query Q, where the query Q is arbitrary and may not be one of the images in the database.

V. Weight Updating

After the intra- and inter- normalization procedures discussed above, the components r_{ijk} within a vector r_{ij}, as well as $S(r_{ij})$'s within the overall similarity S, are of equal emphasis. This *objective* equality allows us to meaningfully associate *subjectively* unequal intra- and inter- weights for a particular query.

A. Update of W_{ij} (inter-weight)

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:

$$RT = [RT_1, ..., RT_l, ..., RT_{N_{RT}}] \tag{18}$$

Let $Score$ be the set containing the relevance scores fedback by the user for RT_l's (see Section 3):

$$\begin{aligned} & = 3, \quad \text{if highly relevant} && (19)\\ & = 1, \quad \text{if relevant} && (20)\\ Score_l & = 0, \quad \text{if no-opinion} && (21)\\ & = -1, \quad \text{if non-relevant} && (22)\\ & = -3, \quad \text{if highly non-relevant} && (23) \end{aligned}$$

The choice of 3, 1, 0, -1, and -3 as the scores is arbitrary. Experimentally we find that the above scores capture the semantic meaning of *highly relevant*, *relevant*, etc. In Equations (19-23), we provide the user with 5 levels of relevance. Although more levels result in more accurate feedback, it is less convenient for the user to interact with the system. 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 $S(r_{ij})$:

$$RT^{ij} = [RT_1^{ij}, ..., RT_l^{ij}, ..., RT_{N_{RT}}^{ij}] \tag{24}$$

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

$$\begin{aligned} W_{ij} & = W_{ij} + Score_l, \quad \text{if } RT_l^{ij} \text{ is in } RT && (25)\\ & = W_{ij} + 0, \quad \text{if } RT_l^{ij} \text{ is not in } RT && (26)\\ l & = 0, ..., N_{RT} && (27) \end{aligned}$$

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 are then normalized by the total weight to make the sum of the normalized weight equal to 1.

$$W_{ij} = \frac{W_{ij}}{W_{Tij}} \tag{28}$$

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.

B. Update of W_{ijk} (intra-weight)

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 for r_{ijk}'s 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[30]. 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.

$$W_{ijk} = \frac{1}{\sigma_{ijk}} \tag{29}$$

where σ_{ijk} is the standard deviation of the length-M' sequence of r_{ijk}'s. Here we assume that the user will mark at least one image, besides the query image, as relevant or highly relevant, such that σ_{ijk} will not be zero. The assumption is valid since otherwise the user would re-start a new query if nothing relevant is retrieved. Furthermore, just as in Equation (28), we need to normalize W_{ijk}'s in the same way.

$$W_{ijk} = \frac{W_{ijk}}{W_{Tijk}} \tag{30}$$

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

C. Summary

Based on the description of the relevance feedback algorithm in Sections 3, 4, and 5, we briefly summarize the properties of the algorithm.

• *Multi-modality*

The proposed image object model, and therefore the retrieval model, supports multiple features and multiple representations. In contrast to a *computer centric* approach's attempt of finding the single "best" universal feature representation, the proposed approach concentrates on how to organize the multiple feature representations, such that appropriate feature representations are invoked (emphasized) at the right place and right time. The *multi-modality* approach allows the system to better model user's perception subjectivity.

• *Interactivity*

In contrast to a *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 a *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 gets refined.

VI. Experimental Results

To address the challenging research issues involved in CBIR, a Multimedia Analysis and Retrieval System (MARS) project is on going at the University of Illinois [9], [45], [41], [30], [42], [29], [48]. 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, we test our proposed approach over two image collections. The first image collection is provided by the Fowler Museum of Cultural History at the University of California-Los Angeles. It contains 286 ancient African and Peruvian artifacts and is part of the Museum Educational Site Licensing Project (MESL), sponsored by the Getty Information Institute. The second image collection is obtained from Corel Corporation. It contains more than 70,000 images covering a wide range of more than 500 categories. The 120x80 resolution images are available at http://corel.digitalriver.com/commerce/photostudio/catalog.htm.

We have chosen these two test sets because they provide complementary properties to each other. The size of the MESL test set is relatively small but it allows us to explore all the color, texture, and shape features simultaneously in a meaningful way. On the other hand, although the heterogeneity of the Corel test set makes the extraction of some features, such as shape, difficult, it has the advantages of large size and wide coverage. We believe that testing our proposed approach on both sets will provide a fair evaluation of its performance.

For the MESL test set, the visual features used are color, texture and shape of the objects in the image. That is,

$$F = \{f_i\} = \{\text{color, texture, shape}\} \tag{31}$$

The representations used are color histogram and color moments [43] for the color feature; Tamura [49], [22] and co-occurrence matrix [50], [51] texture representations for the texture feature, and Fourier descriptor and chamfer shape descriptor [41] for the shape feature.

$$\begin{aligned} R = \{r_{ij}\} &= \{r_1, r_2, r_3, r_4, r_5, r_6\} \\ &= \{\text{color histogram, color moments, Tamura,} \end{aligned}$$

co-occurrence matrix, Fourier descriptor, chamfer shape descriptor}

For the Corel test set, the visual features used are color and texture. That is,

$$F = \{f_i\} = \{\text{color, texture}\} \tag{32}$$

The representations used are color histogram and color moments [43] for color feature; and co-occurrence matrix [50], [51] texture representation for texture feature.

$$\begin{aligned} R = \{r_{ij}\} &= \{r_1, r_2, r_3\} \\ &= \{\text{color histogram, color moments,} \\ &= \{\text{co-occurrence matrix}\} \end{aligned}$$

Our proposed relevance feedback architecture is an *open* retrieval architecture. Other visual features or feature representations can be easily incorporated, if needed. The similarity measures used for the corresponding representations are the following. Color Histogram Intersection [43] is used for the color histogram representation; weighted Euclidean is used for the color moments, Tamura texture, co-occurrence matrix, and Fourier shape descriptor [41] representations; and Chamfer matching [41] is used for the chamfer shape representation.

There are two sets of experiments reported here. The first set of experiments is on the efficiency of the retrieval algorithm, i.e. how fast the retrieval results converge to the true results. The second set of experiments is on the effectiveness of the retrieval algorithm, i.e. how good the retrieval results are subjectively.

A. *Efficiency of the Algorithm*

The ultimate goal of the relevance feedback technique is to help the user retrieve what he or she wants. Because of this, it is very important to verify that the above proposed relevance feedback retrieval algorithm converges to the user's true information need fast.

The only assumption that we make in the experiments is that the user is consistent when doing relevance feedback. That is, the user does no change his or her information need during the feedback process, such that the feedback process can be simulated by a computer.

As we have discussed in Sections 2 and 3, the image object is modeled by the combinations of representations with their corresponding weights. If we fix the representations, then a query can be completely characterized by the set of weights embedded in the query object Q. Let set s_1 be the *highly relevant* set; set s_2 be the *relevant* set; set s_3 be the *no-opinion* set; set s_4 be the *non-relevant* set; and set s_5 be the *highly non-relevant* set. The testing procedure is described as follows:

1. Retrieval results of the ideal case

Let W^* be the set of weights associated with the query object Q. The retrieval results based on W^* are the ideal case and serve as the baseline for comparing other non-ideal case.

(a) Specify a set of weights, W^*, to the query object.
(b) Set $W = [W_{ij}, W_{ijk}]$ to W^*.
(c) Invoke the retrieval algorithm (see Section 3).
(d) Obtain the best N_{RT} returns, RT^*.
(e) From RT^*, find the sizes of sets $s_i, i = 1, ..., 5$, $n_i, i = 1, ..., 5$. s_i's are marked by human for testing purpose.
(f) Calculate the ideal weighted relevant count as:

$$\mathit{count}^* = 3 \times n_1 + 1 \times n_2 \tag{33}$$

Note that 3 and 1 are the scores of the *highly relevant* and *relevant* sets, respectively (see Section 5). Therefore, count^* is the maximal achievable weighted relevant count and serves as the baseline for comparing other non-ideal case.

2. Retrieval results of relevance feedback case

In the real retrieval situation, neither the user nor the computer knows the specified weights W^*. However, the proposed retrieval algorithm will move the initial weights $W0$ to the ideal weights W^* via relevance feedback.

(a) Set $W = W0$.
(b) Set the maximum number of iterations of relevance feedback, P_{fd}.
(c) Initialize the iteration counter, $p_{fd} = 0$.
(d) Invoke the retrieval algorithm and get back the best N_{RT} returns, $RT(p_{fd})$ (see Section 3).
(e) Compute the weighted relevant count for the current iteration:

$$\mathit{count}(p_{fd}) = 3 \times n_1(p_{fd}) + 1 \times n_2(p_{fd}) \tag{34}$$

where $n_1(p_{fd})$ and $n_2(p_{fd})$ are the number of *highly relevant* and *relevant* objects in $RT(p_{fd})$. These two numbers can be determined by comparing $RT(p_{fd})$ against RT^*.
(f) Compute the convergence ratio $CR(p_{fd})$ for the current iteration:

$$CR(p_{fd}) = \frac{\mathit{count}(p_{fd})}{\mathit{count}^*} \times 100\% \tag{35}$$

(g) Set $p_{fd} = p_{fd} + 1$. If $p_{fd} \geq P_{fd}$, quit; otherwise continue.
(h) Feedback the current 5 sets $s_i, i = 1, ..., 5$, to the retrieval system.
(i) Update the weights W according to Equations (25-30). Go to step 2(d).

There are 3 parameters that affect the behavior of the retrieval algorithm, i.e. number of feedbacks P_{fd}, number of returns N_{RT}, and specified query weights W^*. For P_{fd}, the more relevance feedback iterations, the better the retrieval performance. However, we can not expect the user to do relevance feedback forever. In the experiments reported here, we set $P_{fd} = 3$ to study the convergence behavior of the first 3 iterations. The experiments show that the greatest CR increase occurs in the first iteration of feedback, which is a very desirable property.

In all the experiments reported here, for both the MESL and the Corel test sets, 100 randomly selected images are used as the query images and the values of CR listed in the tables are the averages of the 100 cases.

6.1.1 CR as a function of W^*

In the experiments here, we will concentrate on the effect of W_{ij}. The effect of W_{ijk} has been studied in our previous research in [30]. Specifically, only W_{ij} is specified for W^* in the experiments. In the MESL test set, there are 6 r_{ij}'s as described at the beginning of this section. Therefore, both W^* and $W0$ have 6 components. In addition,

$$W0 = [\frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ ] \tag{36}$$

where each entry in the vector $W0$ is the weight for its corresponding representation.

In the Corel test set, there are 3 r_{ij}'s as described at the beginning of this section. Therefore, both W^* and $W0$ have 3 components. In addition,

$$W0 = [\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}\ ] \tag{37}$$

where each entry in the vector $W0$ is the weight for its corresponding representation.

Obviously, the retrieval performance is affected by the offset of the specified weights W^* from the initial weights $W0$. We classify W^* into two categories, i.e. moderate offset, and significant offset, by considering how far away they are from the initial weights $W0$.

For the MESL test set, the six moderate offset testing weights are:

$$\begin{aligned}
W_1^* &= [0.5\ 0.1\ 0.1\ 0.1\ 0.1\ 0.1] \\
W_2^* &= [0.1\ 0.5\ 0.1\ 0.1\ 0.1\ 0.1] \\
W_3^* &= [0.1\ 0.1\ 0.5\ 0.1\ 0.1\ 0.1] \\
W_4^* &= [0.1\ 0.1\ 0.1\ 0.5\ 0.1\ 0.1] \\
W_5^* &= [0.1\ 0.1\ 0.1\ 0.1\ 0.5\ 0.1] \\
W_6^* &= [0.1\ 0.1\ 0.1\ 0.1\ 0.1\ 0.5]
\end{aligned}$$

The six significant offset testing weights are:

$$\begin{aligned}
W_7^* &= [0.75\ 0.05\ 0.05\ 0.05\ 0.05\ 0.05] \\
W_8^* &= [0.05\ 0.75\ 0.05\ 0.05\ 0.05\ 0.05] \\
W_9^* &= [0.05\ 0.05\ 0.75\ 0.05\ 0.05\ 0.05] \\
W_{10}^* &= [0.05\ 0.05\ 0.05\ 0.75\ 0.05\ 0.05] \\
W_{11}^* &= [0.05\ 0.05\ 0.05\ 0.05\ 0.75\ 0.05] \\
W_{12}^* &= [0.05\ 0.05\ 0.05\ 0.05\ 0.05\ 0.75]
\end{aligned}$$

For the Corel test set, the three moderate offset testing weights are:

$$\begin{aligned}
W_1^* &= [0.6\ 0.2\ 0.2] \\
W_2^* &= [0.2\ 0.6\ 0.2] \\
W_3^* &= [0.2\ 0.2\ 0.6]
\end{aligned}$$

The three significant offset testing weights are:

$$\begin{aligned}
W_4^* &= [0.8\ 0.1\ 0.1] \\
W_5^* &= [0.1\ 0.8\ 0.1] \\
W_6^* &= [0.1\ 0.1\ 0.8]
\end{aligned}$$

TABLE I

MESL Moderate Offset Convergence Ratio with $N_{RT} = 12$.

weights	0 feedback	1 feedback	2 feedbacks	3 feedbacks
W_1^*	88.4	99.7	99.7	99.8
W_2^*	64.0	98.4	98.5	98.0
W_3^*	80.7	95.4	93.2	95.8
W_4^*	56.5	94.9	94.9	94.0
W_5^*	66.9	87.9	88.1	88.0
W_6^*	83.5	95.4	97.4	97.5
Avg	73.3	95.3	95.3	95.5

TABLE II

MESL Significant Offset Convergence Ratio with $N_{RT} = 12$.

weights	0 feedback	1 feedback	2 feedbacks	3 feedbacks
W_7^*	40.8	89.6	97.4	98.5
W_8^*	38.9	95.7	98.8	99.0
W_9^*	39.0	77.7	74.2	77.3
W_{10}^*	34.8	91.9	94.6	94.1
W_{11}^*	62.9	85.7	87.1	87.1
W_{12}^*	64.1	87.6	94.4	95.3
Avg	46.8	88.0	91.1	91.9

The experimental results for these cases are summarized in Tables I - IV.

To better represent the process of convergence, we redraw the average Convergence Ratio of the MESL test set and the Corel test set in Figure 3.

Based on the tables and figures, some observations can be made:

- In all the cases, CR increases the most in the first iteration. Later iterations only result in minor increase in CR. This is a very desirable property, which ensures that the user gets reasonable results after only one iteration of feedback. No further feedbacks are needed, if time is a concern.
- CR is affected by the degree of offset. The less the offset, the higher the final absolute CR. However, the more the offset, the higher the relative increase of CR.
- Although the finial absolute CR is higher for the MESL test set than that for the Corel test set, the final relative

TABLE III

Corel Moderate Offset Convergence Ratio with $N_{RT} = 1000$.

weights	0 feedback	1 feedback	2 feedbacks	3 feedbacks
W_1^*	57.1	71.7	74.7	75.2
W_2^*	55.2	54.7	54.6	54.6
W_3^*	59.1	71.4	74.0	74.5
Avg	57.1	65.9	67.8	68.1

TABLE IV

Corel Significant Offset Convergence Ratio with $N_{RT} = 1000$.

weights	0 feedback	1 feedback	2 feedbacks	3 feedbacks
W_4^*	40.2	63.4	68.8	70.0
W_5^*	31.1	34.4	35.1	35.3
W_6^*	41.8	62.0	66.7	67.6
Avg	37.7	53.3	56.9	57.6

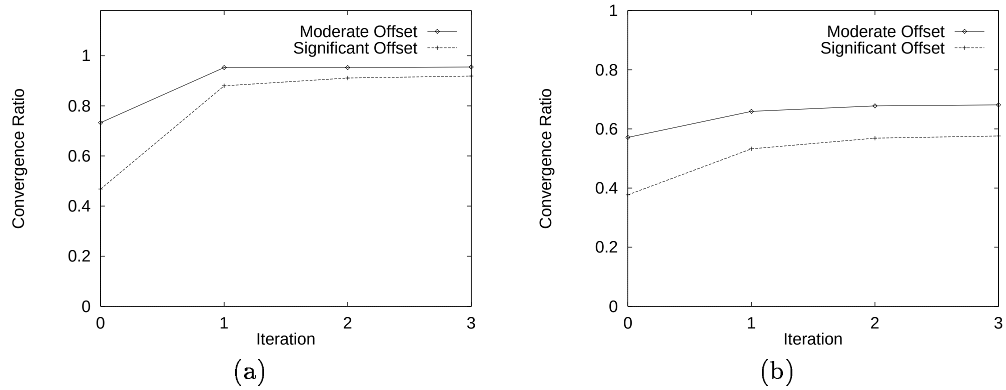

Fig. 3. Convergence Ratio curves. (a) MESL test set, (b) Corel test set

TABLE V
Convergence Ratio for MESL test set with W_7^*.

N_{NT}	0 feedback	1 feedback	2 feedbacks	3 feedbacks
10	40.4	89.7	97.4	98.6
12	40.9	89.6	97.4	98.5
14	40.8	90.2	97.9	98.4
16	40.2	91.3	98.0	98.2
18	40.4	93.2	97.8	98.0
20	40.5	95.9	97.8	97.9

TABLE VI
Convergence Ratio for Corel test set with W_4^*.

N_{NT}	0 feedback	1 feedback	2 feedbacks	3 feedbacks
850	40.6	63.4	68.5	69.6
900	40.9	63.8	69.0	70.1
950	41.2	64.2	69.6	70.6
1000	41.4	64.6	70.0	71.1
1050	41.6	65.1	70.6	71.7
1100	41.9	65.5	70.9	72.0

increase of CR is comparable for both test sets (around 10-20%). The convergence process is more challenging for the Corel test set, because of its much bigger size and less feature representations.

6.1.2 CR as a function of N_{RT}

N_{RT} is related to the size of the test data set. Normally only 2% - 5% of the whole data set is needed. For the MESL test set, we test $N_{RT} = 10, ..., 20$. For the Corel test set, we test $N_{RT} = 850, ..., 1100$ The experimental results are listed in Tables V and VI.

Some observations can be made based on the experiments:

- The first iteration's CR increases the most when N_{RT} is large. This is because, the larger the number of returns, the more the fed-back information and thus better retrieval performance.
- In the second and third iterations, CR is almost independent of different N_{RT}'s. This is because, after first iteration's feedback, most of the desired objects have been found and later performance is almost independent of N_{RT}.

B. *Effectiveness of the Algorithm*

Previous sub-section's experiments focus on the convergence of the algorithm. This sub-section will focus on how good the returns are *subjectively*. The only way of performing subjective tests is to ask the user to evaluate the retrieval system subjectively. Extensive experiments have been carried out. 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 than the *computer-centric* approach in terms of capturing their perception subjectivity and information need.

A typical retrieval process on the MESL test set is given in Figures 4 and 5.

Fig. 4. 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 4, 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 others are similar to the query image in terms of color or texture feature.

Fig. 5. The retrieval results after the relevance feedback

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 map 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 5. Note that our shape representations are invariant to translation, rotation, and scaling. Therefore, images 164 and 96 are relevant to the query image.

Similarly, a typical retrieval process over the Corel test set is given in Figures 6 and 7.

In Figure 6, the top left image is the query image (a glacier). Before any feedback, several human constructed structures appear in the retrieval results. After the user feeds back his interests in ice/water related images, in the next iteration (Figure 7), much more ice/water related images are returned. Note that the retrieval of such images is based on both the color and the texture features.

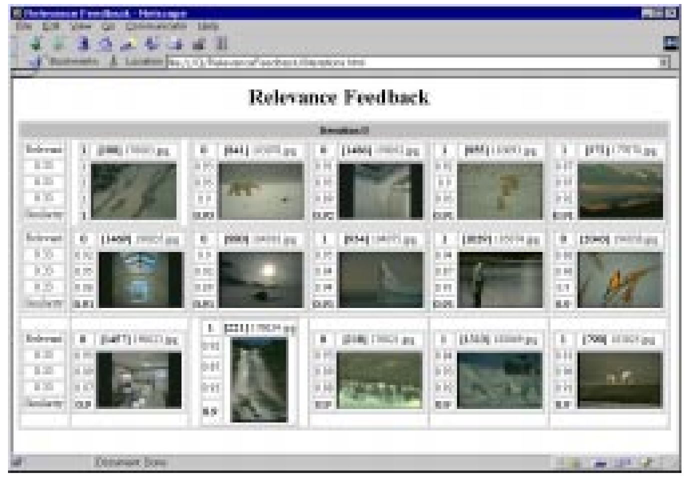

Fig. 6. The retrieval results before the relevance feedback

Fig. 7. The retrieval results after the relevance feedback

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.

VII. 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 a 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 [52]. 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.

One of the future research directions of this approach is to explore optimal or sub-optimal weight updating strategies. Currently, the weight updating strategy is heuristic-based and may not be the best solution. Techniques, such as Expectation Maximization (EM), are promising techniques worth exploring.

VIII. Acknowledgments

The authors would like to thank their colleagues, Porkaew Kriengkrai and Hu Yu, for their help in implementing the system. The authors also would like to thank the anonymous reviewers for their valuable comments in improving the quality of this paper. The example images used in this article are from two sources. One set of images are used with permission from the Fowler Museum of Cultural History at the University of California–Los Angeles. These images were part of an image database delivery project called the Museum Educational Site Licensing Project (MESL), sponsored by the Getty Information Institute. This goal of the two-year MESL project was to test the licensing and delivery of digital images and meta data from seven U. S. museums to seven U. S. universities. The University of Illinois was selected as a participant in the MESL project. The second set of images were obtained from the Corel collection and used in accordance with their copyright statement.

References

[1] W. Niblack, R. Barber, and et al., "The QBIC project: Querying images by content using color, texture and shape," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, Feb 1994.

[2] M. Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Q. Huang, Byron Dom, Monika Gorkani, Jim Hafine, Denis Lee, Dragutin Petkovic, David Steele, and Peter 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 Storage and Retrieval for Image and Video Databases*, 1993.

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

[6] Jeffrey R. Bach, Charles Fuller, Amarnath Gupta, Arun Hampapur, Bradley Horowitz, Rich Humphrey, Ramesh Jain, and Chiao fe Shu, "The Virage image search engine: An open framework for image management," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*.

[7] James Dowe, "Content-based retrieval in multimedia imaging," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

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

[9] Thomas S. Huang, Sharad Mehrotra, and Kannan 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, "Visually searching the web for content," *IEEE Multimedia Magazine*, vol. 4, no. 3, pp. 12–20, Summer 1997, also Columbia U. CU/CTR Technical Report 459-96-25.

[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, no. 3, pp. 557–565, Aug 1996.

[13] "MPEG-7: Context and objectives (v.5)," *ISO/IEC JTC1/SC29/WG11 N1920, MPEG97*, Oct 1997.

[14] "Third draft of MPEG-7 requirements," *ISO/IEC JTC1/SC29/WG11 N1921, MPEG97*, Oct 1997.

[15] "MPEG-7 applications document," *ISO/IEC JTC1/SC29/WG11 N1922, MPEG97*, Oct 1997.

[16] Venkat N. Gudivada and Jijay V. Raghavan, "Special issue on content-based image retrieval systems," *IEEE Computer Magzine*, vol. 28, no. 9, 1995.

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

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

[19] Bruce Schatz and Hsinchun Chen, "Building large-scale digital libraries," *Computer*, 1996.

[20] James P. Callan, W. Bruce Croft, and Stephen M. Harding, "The inquery retrieval system," in *Proc of 3rd Int Conf on Database and Expert System Application*, Sept 1992.

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

[22] Will Equitz and Wayne 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] Fang Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Patt. Recog. and Mach. Intell.*, vol. 18, no. 7, July 1996.

[25] Biao 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] John R. Smith and Shih-Fu Chang, "Transform features for texture classification and discrimination in large image databases," in *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[28] Tianhorng Chang and C.-C. Jay Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Proc.*, vol. 2, no. 4, pp. 429–441, October 1993.

[29] Yong Rui, Thomas S. Huang, and Sharad Mehrotra, "Relevance feedback techniques in interactive content-based image retrieval," in *Proc. of IS&T SPIE Storage and Retrieval of Images/Video Databases VI, EI'98*, 1998.

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

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

[32] Stan Sclaroff, Leonid Taycher, and Marco La 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.

[33] Dirk Daneels, D. Campenhout, Wayne Niblack, Will Equitz, Ron Barber, Erwin Bellon, and Freddy Fierens, "Interactive outlining: An improved approach using active contours," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

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

[35] 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.

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

[37] John R. Smith and Shih-Fu Chang, "An image and video search engine for the world-wide web," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1997.

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

[39] Gerard Salton and Michael J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company, 1983.

[40] Chris Buckley and Gerard Salton, "Optimization of relevance feedback weights," in *Proc. of SIGIR'95*, 1995.

[41] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega, "Automatic matching tool selection using relevance feedback in MARS," in *Proc. of 2nd Int. Conf. on Visual Information Systems*, 1997.

[42] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael 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.

[43] Michael Swain and Dana Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, 1991.

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

[45] Michael Ortega, Yong Rui, Kaushik Chakrabarti, Sharad Mehrotra, and Thomas S. Huang, "Supporting similarity queries in MARS," in *Proc. of ACM Conf. on Multimedia*, 1997.

[46] Sharad Mehrotra, Kaushik Chakrabarti, Mike Ortega, Yong Rui, and Thomas S. Huang, "Multimedia analysis and retrieval system," in *Proc. of The 3rd Int. Workshop on Information Retrieval Systems*, 1997.

[47] Henry Stark and John W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Pretice-Hall, 1986.

[48] Yong Rui, Thomas S. Huang, and Sharad Mehrotra, "Exploring video structures beyond the shots," in *Proc. of IEEE conf. Multimedia Computing and Systems*, 1998.

[49] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. on Sys, Man, and Cyb*, vol. SMC-8, no. 6, 1978.

[50] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein, "Texture features for image classification," *IEEE Trans. on Sys, Man, and Cyb*, vol. SMC-3, no. 6, 1973.

[51] Philippe P. Ohanian and Richard C. Dubes, "Performance evaluation for four classes of texture features," *Pattern Recognition*, vol. 25, no. 8, pp. 819–833, 1992.

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

Yong Rui received the B.S. degree from Southeast University, P. R. China in 1991 and the M.S. degree from Tsinghua University, P. R. China in 1994, both in Electrical Engineering. He is currently finishing the Ph.D. degree in Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. He was a Graduate Research Assistant at the Department of Automation, Tsinghua University from 1991 to 1994. He joined KangTuo Microcomputer Corp., Beijing in 1994 as a Research Engineer. Since 1995 he has been a Graduate Research Assistant at the Image Formation and Processing Group of the Beckman Institute for Advance Science and Technology at UIUC. During the summer of 1998 he is a summer Research Intern at the Vision Technology Group of Microsoft Research, in Redmond, WA.

His research interests include multimedia information retrieval, multimedia signal processing, computer vision and artificial intelligence. He has published over 20 technical papers in the above areas.

He is a Huitong University Fellowship recipient 1989-1990, a Guanghua University Fellowship recipient 1992-1993, and a CSE Engineering College Fellowship recipient 1996-1998.

Thomas S. Huang received his B.S. Degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, China; and his M.S. and Sc.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973; and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology.

During his sabbatical leaves, Dr. Huang has worked at the MIT Lincoln Laboratory, the IBM Thomas J. Watson Research Center, and the Rheinishes Landes Museum in Bonn, West Germany, and held visiting Professor positions at the Swiss Institutes of Technology in Zurich and Lausanne, University of Hannover in West Germany, INRS-Telecommunications of the University of Quebec in Montreal, Canada and University of Tokyo, Japan. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad.

Dr. Huang's professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 12 books, and over 300 papers in Network Theory, Digital Filtering, Image Processing, and Computer Vision. He is a Fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of American; and has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Acoustics, Speech, and Signal Processing Society's Technical Achievement Award in 1987, and the Society Award in 1991. He is a Founding Editor of the International Journal Computer Vision, Graphics, and Image Processing; and Editor of the Springer Series in Information Sciences, published by Springer Verlag.

Michael Ortega Received his B.E. degree with honors from the Mexican Autonomous Institute of Technology in Aug. 1994 with a SEP fellowship for the duration of the studies. Currently he is pursuing his graduate studies at the University of Illinois at Urbana Champaign. Michael Ortega received a Fulbright/CONACYT/García Robles scholarship to pursue graduate studies as well as the Mavis Award at the University of Illinois and is a member of the Phi Kappa Phi honor society, the IEEE computer society and member of the ACM. His research interests include multimedia databases, database optimization for uncertainty support and content based multimedia information retrieval.

Sharad Mehrotra received his M.S. and PhD at the University of Texas at Austin in 1990 and 1993 respectively, both in Computer Science. Subsequently he worked at MITL, Princeton as a scientist from 1993-1994. He is an assistant professor in the Computer Science department at the University of Illinois at Urbana-Champaign since 1994. He specializes in the areas of database management, distributed systems, and information retrieval. His current research projects are on multimedia analysis, content-based retrieval of multimedia objects, multidimensional indexing, uncertainty management in databases, and concurrency and transaction management. Dr. Mehrotra is an author of over 50 research publications in these areas. Dr. Mehrotra is the recipient of the NSF Career Award and the Bill Gear Outstanding junior faculty award in 1997.