Query Feedback for Interactive Image Retrieval

Azadeh Kushki, *Student Member, IEEE*, Panagiotis Androutsos, *Student Member, IEEE*, Konstantinos N. Plataniotis, *Member, IEEE*, and Anastasios N. Venetsanopoulos, *Fellow, IEEE*

***Abstract*—From a perceptual standpoint, the subjectivity inherent in understanding and interpreting visual content in multimedia indexing and retrieval motivates the need for online interactive learning. Since efficiency and speed are important factors in interactive visual content retrieval, most of the current approaches impose restrictive assumptions on similarity calculation and learning algorithms. Specifically, content-based image retrieval techniques generally assume that perceptually similar images are situated close to each other within a connected region of a given space of visual features. This paper proposes a novel method for interactive image retrieval using query feedback. Query feedback learns the user query as well as the correspondence between high-level user concepts and their low-level machine representation by performing retrievals according to multiple queries supplied by the user during the course of a retrieval session. The results presented in this paper demonstrate that this algorithm provides accurate retrieval results with acceptable interaction speed compared to existing methods.**

***Index Terms*—Feature combination, fuzzy aggregation operators, interactive content-based image retrieval, MPEG-7 visual descriptors, multiple queries, relevance feedback, similarity calculations.**

I. Introduction

Both the consumer and business worlds have witnessed the recent and rapid growth in the generation, processing, and sharing of multimedia data. This trend has resulted in the emergence of numerous multimedia repositories that require efficient methods for storage, sharing, and organization of large volumes of multimedia data. In addressing storage, distribution, and sharing needs of data repositories, data compression, and network management have become relatively mature fields. This has resulted in a subsequent shift of research attention from storage and bandwidth considerations to the *management of information content* in multimedia applications [1]. For instance, content-based image retrieval (CBIR) systems have received much interest as a tool for locating relevant information within image repositories. These systems rely on low-level representations of images in terms of their visual content such as color, shape, and texture in order to compare images. In devising a standard scheme for these low-level features, the MPEG-7 committee has finalized a set of tools for the description of multimedia content [2]. The primary goal of this standard is to provide the means for interoperable searching, indexing, filtering, and access to multimedia content [3]. This interoperability is essential in systems that allow retrieval and searching among distributed repositories such as the Internet.

Manuscript received April 30, 2003; revised August 8, 2003. This work was supported in part by the National Science and Engineering Council (NSERC) of Canada.

The authors are with the Multimedia Laboratory, The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 2G4, Canada (e-mail: azadeh@dsp.toronto.edu; oracle@dsp.toronto.edu; kostas@dsp.toronto.edu; anv@dsp.toronto.edu).

Digital Object Identifier 10.1109/TCSVT.2004.826759

The low-level representation of images used by the MPEG-7 standard as well as most content-based visual data retrieval systems leads to several serious shortcomings when comparing and retrieving images. The difficulty comes from two sources [4]: 1) the semantic gap between low-level image representations and higher level concepts by which humans interpret and understand images and 2) the perceptual subjectivity of the users' similarity judgment.

The semantic gap between the low-level representation of images and the high-level user concepts has many consequences in CBIR systems. First, the ideal query is unknown to the system if it cannot be represented in terms of the low-level features. In addition, specification of a query to the system using an example (a query specification type supported by most CBIR systems) results in many ambiguities in terms of the relevant features as well as importance of each of these features to the user. A second problem is that the mapping between low-level feature spaces and high-level user concepts is not known *a priori* and needs to be determined for each query. The overall effect of this gap is that images that may be similar to a query in terms of the low-level features are deemed similar by the system even if they do not contain the conceptual content intended by the user. In the same light, images that contain the same high-level content may not have similar low-level representations (e.g., concept of flower).

Perceptual subjectivity of similarity judgement comes about as different users interpret the visual contents and the similarity between them differently. The main implication of this subjectivity in a CBIR system is that the measure used to calculate similarity between images must be user and query dependent.

In order to alleviate the problems that come about because of user subjectivity and the semantic gap, *interactive* systems have been proposed that place the user in the loop during retrievals. Such *relevance feedback* approaches aim to learn intended high-level query concepts and adjust for subjectivity in judgement by exploiting user input on successive iterations. Generally, the user provides some quality assessment of the retrieval results to the system by indicating the degree of satisfaction with each of the retrieved results. The system then uses this feedback to adjust its query and/or the similarity measure in order to improve the next set of results.

Since image retrieval is a highly interactive service, efficiency and speed are of paramount importance. To fulfill these requirements, most current approaches make several restrictive assumptions that result in the simplification of learning algorithms. One such assumption is that images that are similar to the user according to some high-level concept also fall close to each other in the low-level feature space. Under this assumption, it is possible to determine images that are similar to each other by employing a distance measure between the low-level feature representations. This presupposition, however, is not generally true as high-level semantic concepts may not be directly mapped to elements in a low-level feature space. For example, the concept of "flower" is not uniquely representable using the attributes such as color or shape. A second presumption is that the ideal conceptual query in the mind of a user can be represented in the low-level space and used to determine the region of this space that corresponds to images relevant to the query. Yet, as the mapping between low-level features and the conceptual space is unknown, it is not possible to represent a high-level query as a single point in the low-level feature space. One way to address this issue is to represent the user query in pieces corresponding to a number of similarity regions in the low-level space in order to obtain a better model of the conceptual intentions of the user.

This work proposes a novel approach for interactive content-based image retrieval that provides user-centered image retrieval by lifting the previously mentioned restrictive assumptions imposed on existing CBIR systems, while maintaining accurate retrieval performance. The proposed system exploits user feedback to generate multiple query images for similarity calculations. Employing multiple low-level realizations of one conceptual query results in a more accurate description of user intentions. The query-feedback system, thus, aims to resolve ambiguities that come about during the query-by-example scheme. The final similarity between a given image and a high-level user concept is then obtained as a fusion of similarity of these images to a set of low-level query representations.

This paper is organized as follows. Section II will provide a brief review of the state of the art in the field of content-based image retrieval and discuss some of the limitations of the current approaches. Section III presents an overview of the proposed system. The details of image representation and similarity calculations will be discussed in Sections IV and V, and the query-feedback algorithm will be presented in Section VI. Finally, some experimental results and their interpretation will be discussed in Sections VII and VIII and we will conclude the paper with Section IX.

II. Prior Work

As a subset of multimedia indexing and retrieval systems, content-based image retrieval (CBIR) systems are considered in this paper. Numerous interactive CBIR designs have been proposed that implement relevance feedback to improve the accuracy of retrievals. In the interest of space, we review representatives of each approach and refer the reader to two surveys papers, [5] and [6], on this subject for a more detailed study.

Traditionally, there have been three main approaches to relevance feedback [7], namely, retrieval based on query shifting, feature re-weighting, and updating the probability distribution of the database images. Each of these approaches are discussed briefly in what follows.

A major problem in CBIR is that the ideal low-level representation of the user query is not known due to the semantic gap previously discussed. Query-shifting systems aim to learn the ideal query by moving the query toward the region of the feature space that contains the relevant images paralleling the approach of text retrieval algorithms. Therefore, in each iteration of relevance feedback, the ideal query vector is determined from some combination of images that are deemed relevant or not relevant to the query by the user. In one version of the multimedia analysis and retreival system (MARS) [8], the ideal query is a weighted combination of the feature vectors of the relevant and nonrelevant images. The system proposed in [9] moves the query so that the relevant k nearest neighbors of the query are in a region with a high probability of finding relevant images. Methods based on query shifting impose a restrictive assumption that there exists one query image which completely conveys the intentions of the user. In other words, it is assumed that low-level feature space already models high-level concepts.

Feature re-weighting approaches exploit user feedback to learn the optimal mapping between low-level features and high-level concepts by adjusting the weights assigned to each feature or by modifying the similarity measure used. The system proposed in [10] updates the feature weights based on the scores assigned by the user to each displayed image. Instead of simply determining the feature-weights, biased discriminant analysis is used in [11] to find the optimum *feature transform* that minimizes the ratio of positive over negative scatter. Thus, a new query and a discriminant transform are learned before a nearest neighbor search is performed such that after the transformation the images of interest to the user will fall near the query image in the low-level feature space. Other classification methods including [12] and [13] aim to determine the distribution of positive images in the feature space using support vector machines (SVM). This is done by determining the boundary that separates the relevant and nonrelevant classes.

In contrast to query-shifting methods, the feature re-weighting techniques assume that similar images are somehow grouped together in a connected region of the feature space centered around the query. Thus, query shifting and feature re-weighting can be considered complementary approaches in the sense that one assumes the user query is known and aims to find the best feature space, whereas the other assumes that the feature space is fixed and aims to find the query point that best represents user intentions. Several techniques have been proposed that combine both of these concepts to improve retrieval accuracy. These include the optimization-based learning algorithms of [14] and [15] that aim to find the optimum query vector and feature transformation with respect to some optimality criterion such as the minimization of the distance of relevant images to the ideal query. A general problem with optimization methods is that they may get trapped in a suboptimum solution and never reach the optimum solution [16]. This situation arises because the optimization techniques cannot differentiate between local (suboptimum solution) and global minima (optimum solution). Optimization-based systems also face the problem of *small sample size* as the parameters in these approaches are learned from the feedback provided by the user using some optimization criteria. Consequently, the performance of the system will be compromised if the number of feedback images or *training samples* is small with respect to the dimensionality of the feature space.

Probability-based approaches compute the probability that any image is associated with a particular semantic *label* or *category*. The PicHunter system [17] considers this label to be the target image and aims to calculate the probability that each image is the target using Bayes rule given the history of user actions and a user model. This approach employs a sigmoidal *user model* that predicts the similarity judgment of the user at each iteration of relevance feedback. A simpler approach is proposed by MUSE [18] whereby database images are semiautomatically clustered into groups such that images within each group exhibit a high degree of similarity to each other. The total probability law is then used to calculate the probability that each particular cluster is the target. Other systems, such as [19] and [20], aim to estimate the posterior probability that a given image belongs to the relevant class. Probability-based approaches suffer from restrictive assumptions regarding the density functions and independence of features. Furthermore, their performance is dependent on the accuracy of the models and parameter estimates.

All the aforementioned approaches based on query shifting, feature re-weighting, and probability updating assume that similar images belong to the same class or that they are grouped together in some feature space. This assumption, however, does not generally hold due to the fact that a direct mapping between the low-level features and the high-level user concepts may not exist. As a result, the high-level concepts will have to be represented using a combination of various regions of the low-level feature space. Recently, systems have been developed to relax this assumption by allowing images that are similar to the query to belong to multiple classes or clusters in the feature space. A common approach is to use a Gaussian mixture density to model the likelihood of a particular image belonging to the class of similar images [21]–[24]. These approaches, however, involve parameter estimation for the probability density models. When offline training is not used for parameter estimation, the assumption of independent and identically distributed (iid) feature vectors as well as that of Gaussianity of the probability density model, are made to simplify the process. Furthermore, for an accurate estimation, a population of labeled images greater than the dimension of the feature space is needed to act as training examples. Yet, during relevance feedback only a few positive and negative examples are provided by the user.

In [25], the importance of each feature is taken into consideration by using modified Gaussian basis function to model the distribution of the relevant class. Furthermore, a method based on a novel neural network is proposed to minimize the user participation in the retrieval process. The Falcon system [26] proposes a combination of the above-mentioned approaches that does not require the use of probability density functions. In this system, the distance between a given image and the query is calculated as the aggregation of distances among this image and the positive examples provided by the user. The complexity of this algorithm, however, is rather high since the pairwise similarity among each candidate image and *all* the positive examples needs to be calculated. Furthermore, many positive images may carry the same information and, therefore, most of the extra complexity may be redundant.

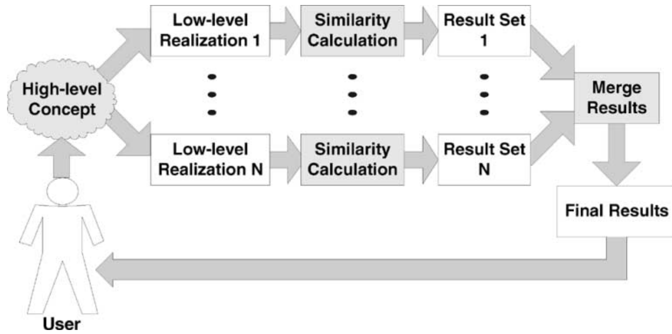


Fig. 1. Conceptual functionality of the query-feedback algorithm.

The query-feedback method proposed in this work allows the similar images to come from various classes. The proposed system selects a subset of the positive images provided by the user as different realization of the high-level user query. A novel aggregation scheme is then proposed to combine the results from these different low-level representations and obtain the overall similarity ranking for the database images.

III. Overview of Method

The semantic gap between the machine representation of images and the human interpretation of visual content means that some high-level concepts cannot be modeled directly using low-level features. In light of this, we propose to consider a high-level query as a combination of various points (exemplars) or queries in the low-level feature space.

In order for a retrieval session to be initiated, the user communicates his/her query intentions to the system by providing an example image. The low-level feature representation of this query image is then compared to the low-level representation of the images in the database using some distance measure. The images that fall closest to the query in the low-level feature space are then returned to the user. The initial example image provided by the user, however, is simply one of the possibly many realizations of the concept in the user's mind. Therefore, the low-level description of the query used by the system may not capture all aspects of the high-level query concept intended by the user. The query-feedback algorithm overcomes this shortcoming by collecting additional low-level realizations of the high-level user concept during iterations of relevance feedback. After the initial retrieval, the user provides feedback on the quality of the retrieved results. More specifically, a retrieved image is marked as a *positive example* if it satisfies the user query and as a *negative example* otherwise. The system will exploit this feedback to determine other images that represent the given user concept. Then, the system retrieves a new set of results by considering this set of example images or realizations of the user concept. Fig. 1 illustrates the conceptual functionality of the proposed system. This method essentially *clusters* the feature space using the user-provided positive examples and combines the similarity results from the various regions to obtain the final results. Therefore, the ideal set of results is computed as a combination of results from a set of different queries that partially reflect or contain the concepts of interest to the user. It must be noted here that the negative examples and the unlabeled images (neither negative nor positive) are also exploited to ensure convergence of the algorithm as discussed in Section VI.

IV. Image Representation

The low-level features selected for image representation are color and texture, as they both provide important visual cues for the delineation and recognition of objects [27]. This work conforms to the MPEG-7 standard color and texture descriptors [28] as the chief objective of the MPEG-7 standard is to offer interoperability among systems and applications aimed at generation, management, distribution, and consumption of multimedia content [29]. This interoperability is essential for interactive services over the Internet and various networks to ensure compatibility among the different system components. The MPEG-7 metadata standard also plays an important role in next generation of multimedia applications geared toward Universal Multimedia Access (UMA) [30]. Furthermore, this compact metadata can be swapped without the need for communicating the actual images.

Three of the MPEG-7 descriptors are used for color: the dominant color descriptor (DCD), the color layout descriptor (CLD), and color structure descriptor (CSD). The MPEG-7 texture descriptors used are the edge histogram descriptor (EHD) and the homogeneous texture descriptor (HTD).

A. Color

The DCD represents the image as a set of N color vectors c_i, together with their percentages p_i. This descriptor is selected since it compactly conveys global information regarding the dominant colors present in the image. The recommended distance measure to be used with the DCD is [31]

$$D_{DCD}(Q,I)=\left(\sum_{i=1}^{N_1} p_{1i}^2+\sum_{j=1}^{N_2} p_{2j}^2-\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} 2a_{1i,2j}p_{1i}p_{2j}\right)^{1/2} \tag{1}$$

where the similarity coefficient $a_{k,l}$ between two RGB color vectors, c_k and c_l, is calculated as

$$a_{k,l} = \begin{cases} 1-\frac{D_{k,l}}{D_{\max}}, & D_{k,l} \le T_d \\ 0, & D_{k,l} > T_d. \end{cases} \tag{2}$$

In the previous expression, $D_{k,l} = \|c_k - c_l\|$ represents the Euclidean distance between two colors vectors, and $T_d = 20$, $\alpha = 1$, and $D_{\max} = \alpha T_d = 20$ are determined experimentally.

The CLD complements the DCD by providing information about the spatial color distribution within images. In contrast to the DCD, YC_bC_r space and the discrete cosine transform (DCT) are used in the extraction of the CLD descriptor, making it resolution-invariant and very compact. Since these color descriptors employ different color spaces and may be represented in different domains (spatial domain as opposed to frequency domain of the DCT), the information that cannot be captured in one space/domain may be captured by another. This descriptor is extracted in the YC_bC_r space and the DCT coefficients are used to represent the image. Six coefficients corresponding to the Y component and three for each of the chromatic components are used here, and the matching of images based on the CLD is performed through the calculation of D_{CLD} as shown below [31]

$$\begin{aligned} D_{\text{CLD}}(Q,P) = &\sqrt{\sum_i w_{yi}(Y_{qi}-Y_{pi})^2} \\ &+ \sqrt{\sum_i w_{Cbi}(Cb_{qi}-Cb_{pi})^2} \\ &+ \sqrt{\sum_i w_{Cri}(Cr_{qi}-Cr_{pi})^2}. \end{aligned} \tag{3}$$

In (3), w_i represents the weight associated with coefficient i and is chosen in a manner paralleling the approach of JPEG quantization. The weights decrease in zigzag scanning order to reflect human perceptual characteristics.

Since the number of dominant colors extracted using the DCD is not the same for all images, some existing methods are not able to use the DCD, as explained in Section VII. Therefore, a third color descriptor, the MPEG-7 CSD, is used to carry out parts of the experiments. The CSD provides information regarding color distribution as well as localized spatial color structure in an image. The image is represented by a modified color histogram that incorporates the spatial distribution of each color into the description. The distance between two CSD histograms for images Q and I is calculated using the L_1 norm as follows:

$$D_{\text{CSD}}(Q,I) = \sum_{i=0}^{127} |(H_{Q,i} - H_{I,i})| \tag{4}$$

where $H_{Q,i}$ represents the ith bin of the color structure histogram for image Q and a 128-bin histogram has been.

B. Texture

The rationale behind the use of the EHD and the HTD is similar to the argument of complementarity made for the color descriptors. The two texture descriptors chosen complement each other, since the EHD performs best on large, nonhomogenous areas (such as complicated natural scenery) while the HTD operates on homogeneous texture regions (e.g., portraits).

The EHD captures the edge distribution within an image. It extracts edge information from 16 subimages and categorizes each subimage into five edge classes: horizontal, vertical, 45°, 135°, and nondirectional [2]. The similarity between two histograms is determined by calculating the L_1 norm of the 80-dimensional feature vectors H_Q and H_I [31]

$$D_{\text{EHD}}(Q,I) = \sum_{i=0}^{79} |(H_{Q,i} - H_{I,i})|. \tag{5}$$

The HTD represents the mean and energy deviation of 30 frequency channels modeled using Gabor functions [31]. Since this descriptor is represented in the frequency domain, it captures information that complements the EHD.

In a manner similar to histogram intersection, the distance D_{HTD} between two vectors T_Q and T_I is calculated as follows:

$$D_{\text{HTD}}(Q,I) = \sum_k \left| \frac{(T_{Q,i} - T_{I,i})}{\alpha(k)} \right| \tag{6}$$

where the parameter $\alpha(k) = 1$ was determined experimentally.

V. Similarity Measure

The Unified Framework for Similarity Calculation (UFSC) [32], [33] is used in this work to obtain the similarity between two images based on multiple features and descriptors. In this framework, each low-level descriptor distance, D_{DCD}, D_{CLD}, D_{EHT}, and D_{HTD} [(1)–(6)], is passed through a fuzzy membership function to obtain a decision about the similarity of two images for a particular descriptor. The form of the membership function is highly dependent on the nature of the descriptors and distance measures. For the purposes of this paper, it was experimentally determined that the Cauchy membership function provides reasonable performance. The form of this function is

$$\mu_j(I) = \frac{1}{1 + \left(\frac{|D_j(Q,I)|^\rho}{\lambda}\right)} \tag{7}$$

where $D_j(Q,I)$ is the distance in descriptor j between images Q and I, $\lambda = \text{Median}_i(D(Q,I_i))$ is the median of the distances of all database images to the query, and ρ is a parameter determined experimentally to equal 4.25, -1, -1, 4, and -1 for the DCD, CLD, CSD, EHD, and HTD, respectively.

The descriptor decisions, μ_{DCD}, μ_{CLD}, μ_{CSD}, μ_{EHT}, and μ_{HTD} obtained through (7) for each image I, represent the membership grade of that image to the set of *images similar to the query* based on that particular descriptor. The overall similarity calculation is, therefore, based on decisions obtained from descriptor distances using the membership functions rather than actual feature distances. Since each descriptor or feature similarity comes from its own space and represents a different concept, the raw distances cannot be directly combined. To illustrate this, consider the distance in the dominant colors reflecting the distance between RGB values whereas the color layout difference refers to YC_bC_r DCT coefficients. In such a case, it is not meaningful to directly combine these completely different concepts. In contrast, the UFSC algorithm first maps all of these distances to a decision space by using a membership function and then combining the *decisions*. As a result, aggregation is performed in a decision space where all the values represent the same concept: a decision on the similarity between two images. Fig. 2 shows the overall structure of the UFSC framework.

Similarity calculation between two images is performed as a combination of the feature descriptor decisions using a *compensative* operator as follows:

$$\mu_{S_Q}(I_i) = [\mu_{\text{DCD}} \odot \mu_{\text{CLD}} \odot \mu_{\text{CSD}}] \odot [\mu_{\text{EHT}} \odot \mu_{\text{HTD}}]. \tag{8}$$

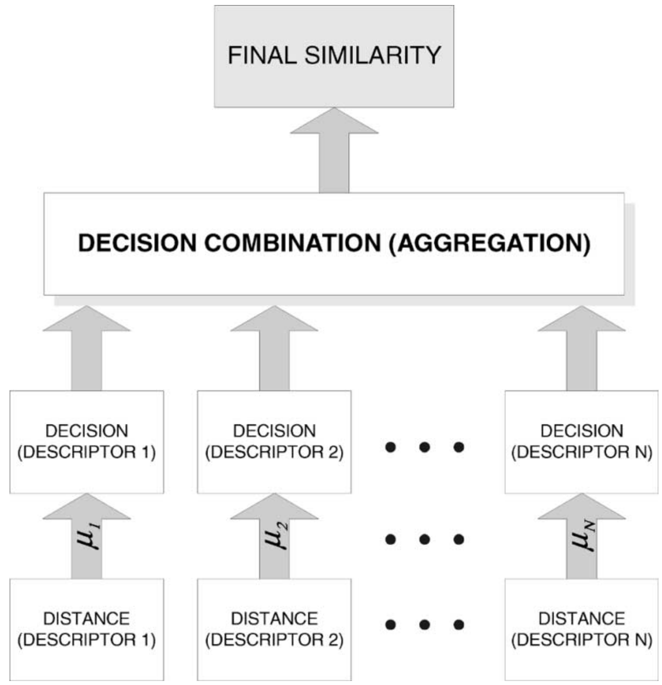


Fig. 2. Similarity calculations are based on aggregation of decisions in the unified framework for similarity calculations.

The aggregation operator $\odot$ is defined as follows:

$$x_1 \odot, \ldots, \odot x_n = \gamma \max(x_1, \ldots, x_n) + (1-\gamma)\min(x_1, \ldots, x_n) \tag{9}$$

where the x_i's are the elements being aggregated and $\gamma \in [0,1]$. Equation (9) represents a family of logical operators ranging between the logical AND and the logical OR. More specifically, the logical behavior of this aggregator can be adjusted based on the compensation parameter γ: for $\gamma = 0$, a conjunctive operator (AND) is obtained, for $\gamma = 1$, a disjunction (OR), and for all other values of this parameter, the operator is a compromise between these two logical extremes. This compensative operator offers many advantages in modeling the structure of conceptual user queries. Based on user needs, the final similarity between two images can be obtained as a conjunction, disjunction, or compromise between the individual descriptor similarities. More details on the rationale behind the choice of this operator, its desirable properties for similarity measurement between images, and adjustment of the parameter γ are discussed in [32].

In the UFSC framework, the final similarity scores are represented using a fuzzy set which indicates the grade of membership of each image to the query's similarity set. The similarity set of a query image Q, denoted as $\mathcal{S}_Q$, is a conceptual fuzzy set that contains images similar to this query. Using this definition, the similarity set becomes

$$\mathcal{S}_Q = \{\mu_{S_Q}(I_1)/I_1, \ldots, \mu_{S_Q}(I_N)/I_N\} \tag{10}$$

where N is the number of images in the database, I_i is the ith database image, and $\mu_{S_Q}(I_i)$ is the decision made on similarity of images Q and I_i using (8). The higher the grade of membership of a database image I_i to this set, the more similar it is to the query. Further details on interpretation of the similarity set can be found in [32].

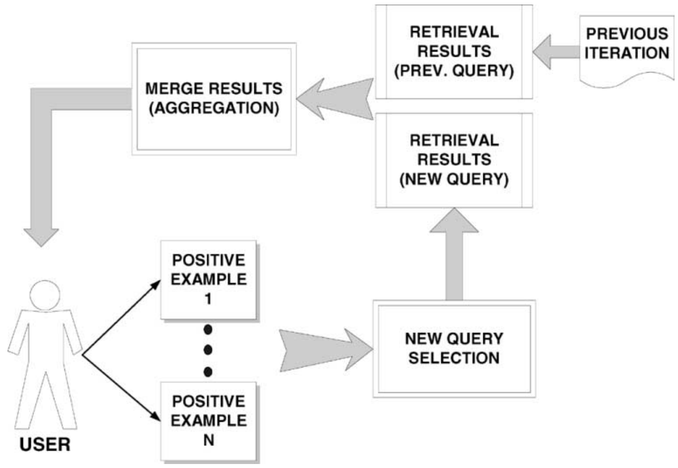


Fig. 3. Overall structure of query-feedback system.

Input:

$0 < \alpha, \beta < 1$,

set of positive examples $\mathcal{P}$ and set of negative examples $\mathcal{N}$.

Output:

A set of rankings for the images in the database.

Query feedback algorithm:

Step 1. Modify the previous similarity set based on positive and negative examples. from the user:

$\forall I_i \in \mathcal{P}, \mu_{\mathcal{S}'_{Q_t}}(I_i) = \mu_{\mathcal{S}_{Q_t}}(I_i)$,

$\forall I_i \in \mathcal{N}, \mu_{\mathcal{S}'_{Q_t}}(I_i) = 0$,

$\forall I_i \notin \mathcal{P} \vee \mathcal{N}, \mu_{\mathcal{S}'_{Q_t}}(I_i) = \alpha \mu_{\mathcal{S}_{Q_t}}(I_i)$,

Step 1. Choose the new query, I_{new}, from the positive examples provided by the user (using Query Selection algorithm).

Step 2. Perform the rankings based on the new query to obtain $\mathcal{S}_{new}$.

Step 3. Calculate final ranks:

$S_{Q_{t+1}} = \min(\mathcal{S}_{new}, \mathcal{S}'_{Q_t})^{\beta} \max(\mathcal{S}_{new}, \mathcal{S}'_{Q_t})^{(1-\beta)}$.

Fig. 4. Outline of the query-feedback algorithm.

VI. Query Feedback

In calculating the similarity scores using (8), it was assumed that there exists exactly one ideal low-level representation of the query. Similarity measurements were then performed by comparing each image to this ideal image. Unfortunately, such a presupposition is not justified because of the subjective interpretation of images. The example image provided upon initiation of the query is believed to contain the concepts of interest as perceived by the user. This example image, however, is simply a starting point and may not be a complete representation of the user needs. For example, the user may be interested in retrieving all images containing cars. To this end, he provides a picture of a yellow Ferrari as the example. This, however, does not mean that other makes and colors of cars will not be acceptable to this user. It is, therefore, possible for a high-level concept to have different low-level manifestations. Considering this fact, the query-feedback algorithm performs the retrieval based on multiple low-level queries. The objective of relevance feedback is, therefore, to exploit user feedback to locate various partitions of the decision space that correspond to intended high-level concepts. The conceptual functionality of the query-feedback algorithm is depicted in Fig. 3.

A. Query-Feedback Algorithm

Since the query image provided by the user may not completely convey the intentions of the user to the system, the query-feedback algorithm aims to improve the retrieval results by performing the retrieval based on several query images. To this end, a subset of images marked as positive examples by the user during relevance feedback will be selected to serve as new query images. The subsequent retrieval is then performed based on a set of images rather than on a single one. The query-feedback algorithm presented in this section is summarized in Fig. 4.

The new query images should be chosen in a systematic way to maximize the amount of information the system gains. For example, it is likely that the top two images are close to each other in the low-level feature space. Adding the second image to the set of examples, therefore, will not bring much new knowledge to the system because of the redundancy that exists between the low-level representations. In addition to the positive examples, the negative and unlabeled images may be used for determination of the new query points as well.

A number of ways to choose the set of new queries from the positive examples provided by the user exist. The most intuitive approach to pick the next query image is to select the one positive image that maximizes a given distance function to the query. This ensures that the clusters of similarity are as far as possible in the decision space and, thus, provide complementary information. In its simplest form, this method translates to choosing the positive image ranked last during similarity measurement to the query. The distance measure in this case will be the L_1 norm between the query and the final similarity decisions in $\mathcal{S}_Q$ obtained through (8). To be precise

$$Q_{\text{new}} = \arg\max_{\text{I}_\text{i}} |1 - \mu_{\text{S}_{\text{Q}_\text{old}}}(\text{I}_\text{i})|. \tag{11}$$

There is, however, no indication that this distance is a meaningful measure in the decision space. Thus, a novel *set distance* measure is introduced here for choosing the next query image to be considered.

As previously mentioned, the query-feedback algorithm performs the retrievals based on a set of query images rather than just one. This is achieved by selecting a subset of the positive examples provided by the user as the new query images. For simplicity, we only add one extra query image during each iteration of relevance feedback as opposed to using multiple images. Clearly, there is a direct relationship between the number of exemplars chosen and the computational complexity of the algorithm. Furthermore, it is not clear that increasing the number of exemplars will improve the performance of the system. This is due to the fact that the user only marks the positive examples among the set of images that are presented to him/her during each retrieval. This set generally contains a small set of images that are deemed to be similar to the query in terms of their low-level feature representation. As a result, the set of positive images chosen by the user are not very far away from each other during the initial iterations of the feedback algorithm. In light of this, a small number of exemplars should be sufficient to characterize the set of positive examples. The effect of the number of new queries on the accuracy of the results is further investigated in Section VII. In addition, the set distance algorithm used to choose the most informative representative will be discussed in Section VI-B.

Once the new query I_{new} is chosen by the query-feedback algorithm, the similarity set $\mathcal{S}_{\text{new}}$ containing the rankings of the database images is generated using (8). The next step is to merge these results with the results based on the previous query image at iteration t of the algorithm $\mathcal{S}_{Q_t}$. In order to guarantee the convergence of the query-feedback algorithm, modifications are made to the previous set of rankings that generate a new set ${\mathcal{S}'}_{Q_t}$ before merging of the two sets is performed. Specifically, the ranks of the images that are unlabeled or not visited by the user are first multiplied by a forgetting factor $\alpha \in (0,1)$. The rank of an unlabeled image thus decreases every time the user does not mark it as a positive image. Thus, if an image is not labeled in many consecutive iterations, its rank will approach zero and it will be treated a negative example. The forgetting factor α determines the strictness in outlier filtering. The outliers are images that are conceptually dissimilar to the query but have some low-level features in common with it. The ranks of the images that are not deemed relevant (i.e., are not ranked within the top N relevant images to a query) are multiplied by this forgetting factor. Thus, at iteration k, the rank of an image that is not similar to any of the previous k queries is decreased by α^k. This method has the effect of implicitly filtering the outlier data and is a result of the assumption that the truly relevant images are likely to be similar to more than one query (or positive example). If an image is similar to several positive examples, it has a higher likelihood of being retrieved during initial iterations since it will receive a high rank using any of such positive examples as the query. In Section VII, the effects of this parameter on the accuracy of retrievals are further investigated.

The second modification made to the previous similarity set before merging involves changing the ranks of negative images to zero. As a direct consequence of the merging scheme used here, these negative examples will not show up in the top results in subsequent iterations of the algorithm. Finally, the set of results from the previous iteration ${\mathcal{S}'}_{Q_t}$ and the set of results from the current iteration $\mathcal{S}_{\text{new}}$ are merged to generate the new rankings for the images based on user feedback. A compensatory operator is used to perform the merging, as follows:

$$\mathcal{S}_{Q_{t+1}} = \min({\mathcal{S}'}_{Q_t}, \mathcal{S}_{\text{new}})^{\beta} \max({\mathcal{S}'}_{Q_t}, \mathcal{S}_{\text{new}})^{(1-\beta)}. \quad (12)$$

The operator in (12) has been chosen since it offers several desirable and essential properties. This operator's logical behavior ranges between the logical AND and OR depending on the parameter β. For $\beta = 1$, a logical AND (or intersection) of the two result sets is obtained and $\beta = 0$ results in a logical OR (or union) of the two resulting sets. For all other values of β, a compromise between the AND and OR (or a soft intersection or union) is obtained. This compensation parameter β can be set based on the distance between new query images generated. For instance, if the two exemplars are too close, this parameter can be set closer to 1 to result in an intersection of the two result sets. This is based on the assumption that the two exemplars should provide more or less similar results and, therefore, anything outside their intersection can be considered an outlier or irrelevant. The effects of the compensation parameter on the retrieval results will be investigated in Section VII.

Another advantage of the operator of (12) is that it provides a null element that can be used to completely determine the decision regardless of the other elements. If an image is chosen as negative even once, it never again shows up in the top results regardless of its similarity to any of the exemplars.

An important consideration in designing relevance-feedback algorithms is their convergence. This ensures that there is an improvement in the retrieval results from one iteration to another. In light of this, a theorem on the convergence of the query-feedback algorithm is presented next. The proof can be found in [32].

Theorem 1: Let $\mathcal{P}$, $\mathcal{U}$, and $\mathcal{N}$ denote the set of images marked by the user as positive, undecided, and negative, respectively, during the tth iteration of the algorithm. Then, $\exists \alpha, \beta, \mu$ such that $\mu_{S_{Q_{t+1}}}(\forall I_i \in \mathcal{P}) \geq \mu_{S_{Q_{t+1}}}(\forall I_i \in \mathcal{U}) \geq \mu_{S_{Q_{t+1}}}(\forall I_i \in \mathcal{N})$.

Theorem 1 states that there exists a value of the system parameters such that images labeled by the user as positive will always receive a higher ranking than the unlabeled and negative images. This property ensures that these images will not be presented to the user before the positive labeled images. Furthermore, this theorem shows that by properly choosing α, the number of positive images can never decrease in successive iterations of the relevance-feedback algorithm. Thus, the accuracy of the results never decreases during successive iterations, meaning that the query feedback can never degrade the quality of the results.

The next section discusses the choice of new query images during relevance feedback.

B. *Query-Point Selection*

Clearly, many choices for the query-selection algorithm exist ranging from heuristic methods such as choosing the last positive image to more involved methods. This section proposes a theoretical method for selecting the new query image based on set distances. Other heuristic methods will be discussed in the experiments section.

In order to minimize the probably of overlap between two clusters, it is important to ensure that the chosen cluster seeds are the farthest possible from each other. It is not theoretically clear, however, whether this heuristic method is sound for measuring distances in a low-level decision space. Therefore, a new measure is introduced for quantifying the amount of overlap between the information conveyed by each the cluster center.

In the proposed scheme, each image in the database I_p will be associated with a fuzzy subset $\mathcal{S}_{I_p}$ by utilizing (8) and considering the low-level representation of image I_p to be the low-level query image. This fuzzy subset represents the grade of similarity of each of the images in the database to image I_p. These fuzzy subsets are then used to measure the degree of redundancy between the information carried by two images, I_p and I_r. If S_{I_p} and S_{I_r} are close, the database images will have similar rankings using either I_p or I_r as the query. Therefore, the two images more or less convey the same information. On the other hand, if S_{I_p} and S_{I_r} are dissimilar, it can be concluded that the two

Input:

Set of positive images $\mathcal{P} = \{I_1, I_2, \dots, I_N\}$.

Output:

New query image.

Set-Distance Algorithm:

Step 1. Calculate membership of database images to similarity set of images in *I*, $S_{I_1}, \dots, S_{I_N}$.

Step 2. Calculate set overlaps $\mathcal{OV}_{I_1}, \mathcal{OV}_{I_2}, \dots, \mathcal{OV}_{I_N}$ and the query set: (Equation 13).

Step 3. $I_{new} = \arg\min_i \mathcal{OV}_{I_i}$.

Fig. 5. Outline of the set-distance algorithm.

images carry different information and as a result, new information may be obtained by combining the two. Various distance measures can be used to find the difference between two fuzzy sets. For simplicity, the absolute distance is employed here. The amount of overlap between two images, $\mathcal{OV}$, is, therefore, defined as follows:

$$\mathcal{OV}(I_p, I_r) \equiv S_{I_p} - S_{I_r} \equiv \sum_{i=1}^{N} |\mu_{S_{I_p}}(I_i) - \mu_{S_{I_r}}(I_i)|. \quad (13)$$

A summary of the query-point selection algorithm is shown in Fig. 5.

VII. Experiments

Various experiments were conducted to evaluate the retrieval performance of the proposed relevance-feedback algorithm with respect to accuracy and convergence speed. Furthermore, the effects of system parameters on the results are examined.

A. Dataset

These experiments were carried out on a database of 2850 general images containing a wide variety of subjects such as flowers, sunsets, ancient ruins, people, and animals. This type of database is chosen because large variations exist between the low-level representation of images in the same semantic class. For example, images of natural scenery may have varying color composition or texture depending on the type of image. Our experiments show that despite this semantic gap, the proposed algorithm performs well in learning the mapping between the low-level features and high-level user concepts. Furthermore, this general database is selected to reflect the wide range of CBIR user needs in a distributed network. In these types of environments, the retrieval may be performed using various databases containing different high-level concepts and, hence, the retrieval and relevance-feedback engines should be able to handle such repository generality.

It must be emphasized here that since this work does not focus on optimization of search speed, the distribution of images within the test database is the important factor to consider rather than the the database size. For example, consider a case where the distribution of feature descriptors is similar for a given class (in-class) and very distinct from the remaining classes (between-class). Since a one-to-one mapping exists between the low-level representation and high-level concepts, the retrieval is greatly simplified regardless of the size of the database. The test database used in this paper has been manually selected such that different degrees of in-class and between-class descriptor variations are exhibited for the various semantic topics.

TABLE I
Overall Improvement in Recall After Ten Iterations of Relevance Feedback

Method	Improvement in $\overline{Recall}$
Query Feedback	26%
Integrated Probability Function	3%
Bayesian estimation	6%
Falcon	3%

B. Performance Measures

The performance of the proposed algorithm in terms of accuracy is evaluated using the Recall measure [34]

$$\text{Recall}(i) = \frac{N_{\text{relevant}}(i)}{N_{\text{total}}} \quad (14)$$

where $N_{\text{relevant}}(i)$ represents the number of retrieved images relevant to the user query after i retrievals and N_{total} is the total number relevant images to the query as determined *a priori* by a human subject (ground truth). The Recall measure is an indication of the percentage of the ground truth retrieved at each iteration of the algorithm.

The Recall measure was averaged over 20 different classes to obtain $\overline{\text{Recall}}$

$$\overline{\text{Recall}}(i) = \sum_{j=1}^{20} \text{Recall}_j(i). \quad (15)$$

In order to compare the speed of convergence for the various methods examined, the *Recall Rate of Change* (RRC) measure was used. This measure is defined as the derivative of $\overline{\text{Recall}}$:

$$\text{RRC} = \frac{d\overline{\text{Recall}}}{di}. \quad (16)$$

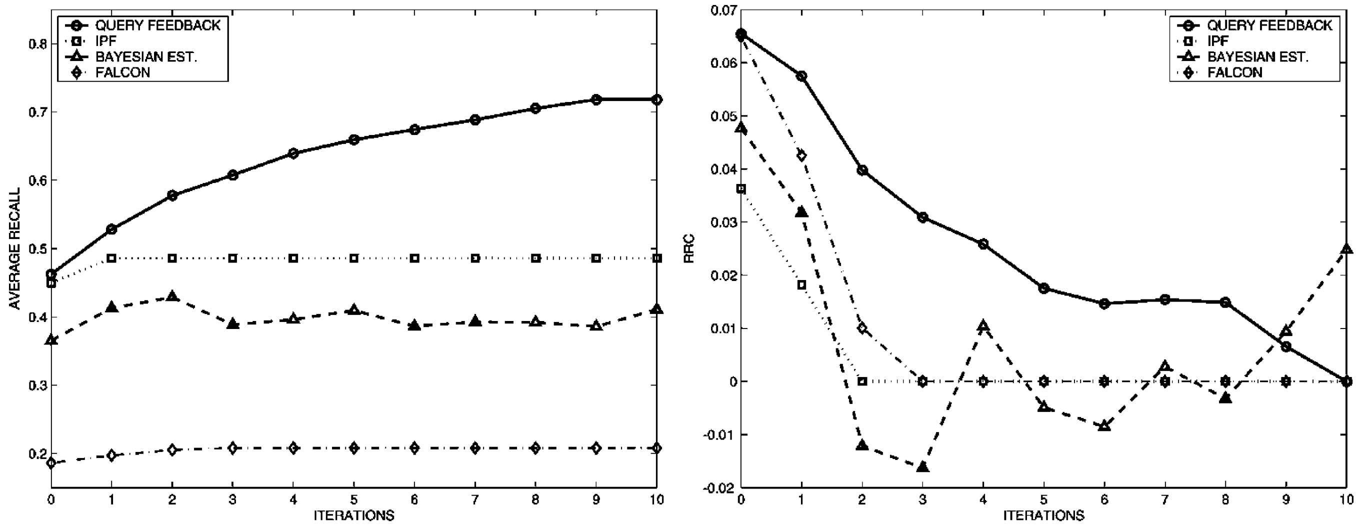
The RRC is a measure of the rate of change in $\overline{\text{Recall}}$ at each iteration of the algorithm. It is essentially an indication of the speed of convergence to the final recall value.

C. Comparison to Other Methods

Performance of the proposed system is compared to representatives of some of the existing methods in the literature, namely, the integrated probability function (IPF) [19], Bayesian estimation (BE) [20], and the aggregated dissimilarity measure of Falcon (AD) [26].

It must be noted that the Euclidean distance approach employed by each of these three methods requires that the descriptor vectors are of the same length and type. This restriction, therefore, prevents the usage of the MPEG-7 DCD since the number of dominant colors extracted from various images may not be the same.

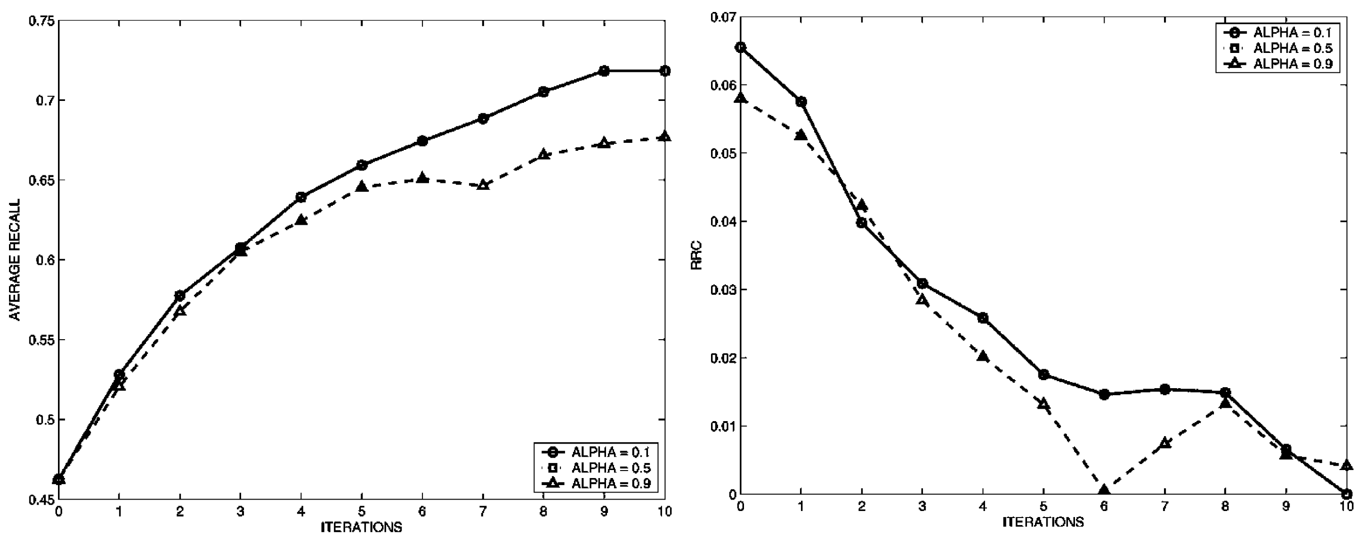
A quantitative performance comparison between query feedback, IPF, BE, and AD is provided in Fig. 6. Specifically, a plot of $\overline{\text{Recall}}$ versus iterations of the relevance-feedback algorithms



(a) $\overline{Recall}$ versus. iterations of relevance feedback.

(b) RRC versus iterations of relevance feedback

Fig. 6. Comparison between query feedback, integrated probability function, Bayesian estimation, and Falcon.



(a) Effect of α on retrieval performance.

(b) Effect of α on convergence rate.

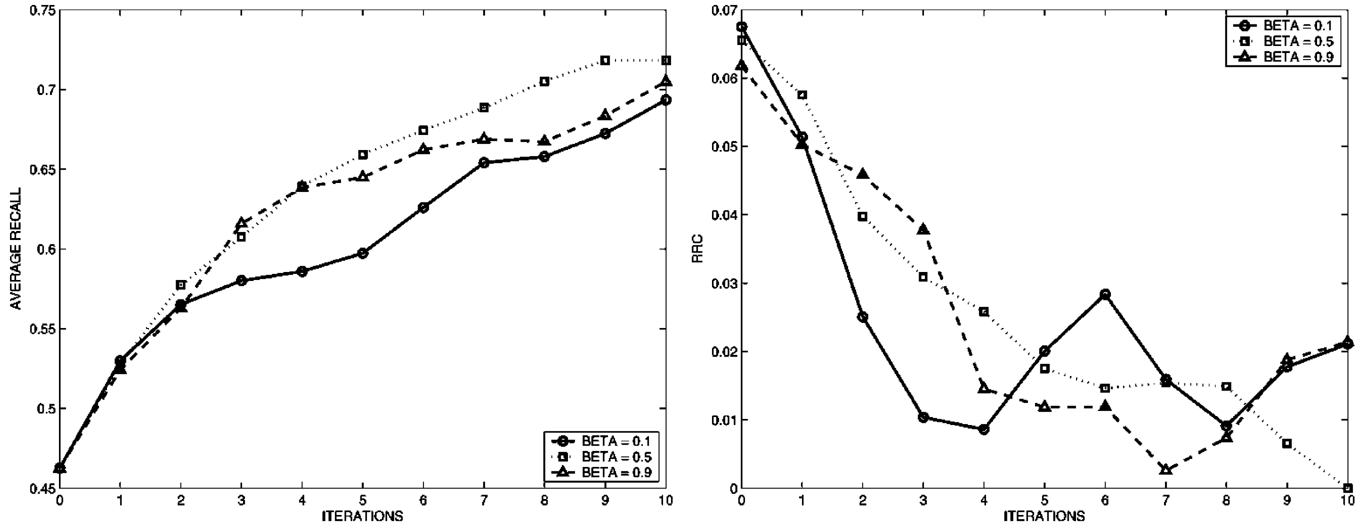
Fig. 7. Effects of forgetting factor α system performance.

is shown in Fig. 6(a) and the convergence speeds as defined by the RRC are shown in Fig. 6(b). The overall improvement achieved in recall after ten iterations of relevance feedback for all methods is summarized in Table I. These results show that query feedback outperforms the other three methods by providing 26% improvement compared to 3%, 6%, and 3% by IPF, BE, and AD, respectively. Furthermore, the proposed method achieves the highest RRC during the initial iterations of the algorithm, indicating that most of the improvement is achieved in these stages. This is an advantageous characteristic as users generally desire to achieve accurate results with only a small number of iterations.

D. Effect of System Parameters

Fig. 7(a) shows the $\overline{\text{Recall}}$ plots for various values of α versus the number of iterations of the relevance-feedback algorithm. It can be seen that values of this parameter close to unity cause instability in the algorithm in terms of convergence. This instability manifests itself as a sudden drop in RRC shown in Fig. 7(b). As discussed previously, this forgetting factor causes the filtering of outliers. The higher this value, the less strict the filtering process, thus, for α values close to unity the outliers will not be ignored, causing a performance degradation as a result of the false positives.

The compensation parameter β is used to merge the results of the various queries to obtain a final set of rankings. The higher this value, the more this combination will resemble a union or a logical OR (more optimism in believing that the similarity set for each individual query provides us with a true ranking of images based on the user intentions). Similarly, for smaller values of the compensation parameter, the merging operation will behave closer to a logical AND with more pessimism. The choice of this parameter is dependent on the system, the database, and on the user needs. The $\overline{\text{Recall}}$ and RRC graphs for various values of β are shown in Fig. 8(a) and (b) indicating that a value of 0.5 for this parameter results in the highest accuracy in terms of $\overline{\text{Recall}}$ and an acceptable RRC.



(a) Effect of the β on retrieval performance.

(b) Effect of the β on convergence rate.

Fig. 8. Effects of compensation parameter β on system performance.

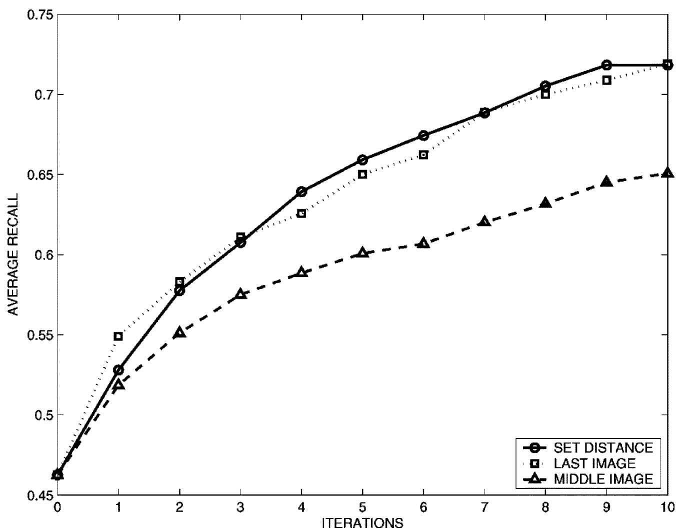


Fig. 9. $\overline{\text{Recall}}$ versus iterations of the relevance-feedback algorithm using various query-selection methods.

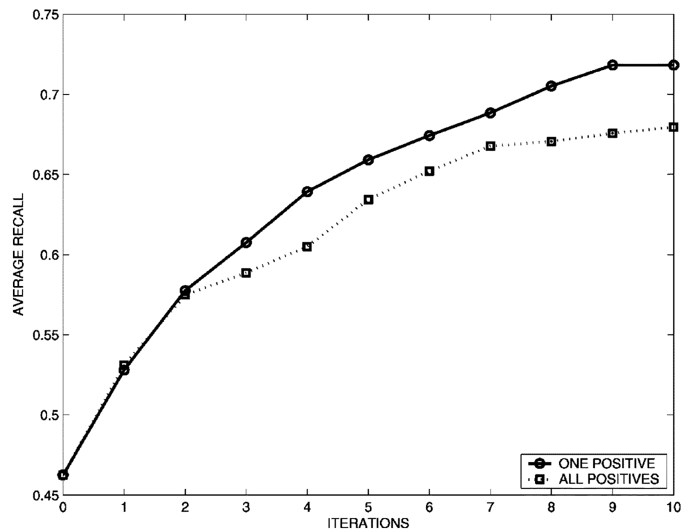


Fig. 10. $\overline{\text{Recall}}$ versus iterations for different number of query images.

E. *Effect of New Query Image*

In this section, the query-selection algorithm based on set distances is compared to possible heuristic methods. Fig. 9 depicts $\overline{\text{Recall}}$ for three different methods using the proposed method, choosing the last and middle positive examples selected by the user as the new query. It can be seen that although all methods have similar performance, the proposed method performs slightly better than the heuristic ones. From these results, it can be concluded that the complexity of the system can be greatly reduced by employing the heuristic method based on choosing the last positive image as the new query.

Fig. 10 shows the retrieval performance using different number of new queries. Specifically, it is observed that using all positive examples as representations of the high-level query degrades the retrieval performance. Since most of the positive images lie close to the original query in the feature space, employing multiple queries selected from the positive set may introduce some bias toward the initial query. This bias causes the system to only consider the region of the feature space close to the original query and this results in a performance degradation.

It must be emphasized that the choice of the query-selection algorithm is a design issue that should take into account the specific system requirements in terms of the tradeoff between complexity and accuracy.

VIII. Discussion

A. *Effectiveness*

The query-feedback method proposed in this paper removes several of the restrictive assumptions made by existing relevance-feedback techniques and yet achieves high accuracy in retrieval results. The main advantage of the proposed method is its flexibility and ability to bridge the gap between the machine representation of visual content and its semantic meaning to humans. This ability comes about because the query-feedback model does not assume a one-to-one mapping between a low-level feature space and high-level user concepts. Instead, this method allows retrievals based on multiple low-level representations. Therefore, after a few iterations of relevance feedback, the system identifies regions of the low-level space that correspond to different manifestations of the concepts of interest. The overall results are then obtained as a union of these low-level concepts. Furthermore, it must be noted that the query-feedback algorithm also takes advantage of the negative and unlabeled images to ensure convergence of the algorithm.

Query feedback incorporates the history of user actions into its design (as opposed to methods such as Falcon). This is important to ensure that once an image is marked as nonrelevant to the query, it will not be shown to the user again. Furthermore, if an image is chosen as relevant during multiple iterations of the relevance feedback, its rank increases relative to other images.

B. Generality

In addition to its effectiveness in retrieval and flexibility in closing the semantic gap, the query-feedback system does not require any assumptions of user models, probability density functions, or the nature of the feature space. It must be noted that the performance of this algorithm can even be further optimized using methods such the one proposed in Garlic [35].

The design of the query-feedback algorithm is general enough to accommodate various databases, users, and environments. The parameters, such as the query-selection algorithm component, forgetting factor, and compensation parameter, can be adjusted independently of the actual algorithm in order to satisfy the specific system requirements.

C. Complexity

Each iteration of the query-feedback algorithm involves the selection of a new query image, retrieval based on this new query, and merging of the results. Since the new query image is simply the positive image with the lowest ranking during the previous iteration, this operation does not incur any extra computational cost. Since the query-feedback scheme allows for any descriptor distance measure to be used, the complexity of the new retrieval is variable and design dependent. The merging operation is performed for all database images and is linear in the size of the database. It must be noted that the $\min$ and $\max$ operators are applied to two elements only, namely, previous rank and current rank. The complexity of the combination operator, therefore, remains constant.

IX. Conclusion

Most content-based image retrieval systems assume that the high-level user concepts can be modeled as single points in the low-level feature space. This presumption, however, is an incorrect one due to the semantic gap that exists between high-level user concepts and low-level features. Query feedback, an interactive, user-centered retrieval technique, has been proposed in this paper to address these shortcomings by performing retrievals based on multiple query examples rather than just one. These query examples are determined during relevance feedback to obtain an accurate representation of high-level user concepts as a combination of various low-level features regions.

The query-feedback algorithm uses a subset of images marked as positive by the user to perform the retrieval. Although the negative and unlabeled are implicitly used to ensure the convergence of the algorithm, they are not directly used in similarity calculations. Future work will focus on incorporating these images into the method. In addition, further work will be done to consider different weights indicating the relative importance of each positive image during similarity combination.

References

[1] R. J. Lopes, A. T. Lindsay, and D. Hutchison, "The utility of MPEG-7 systems in audio-visual applications with multiple streams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 16–25, Jan. 2003.

[2] "Multimedia content description Interface – Part 3: Visual," NISO/IEC, Tech. Rep. 15 938-3:2001, 1st ed., 2001.

[3] P. Salembier and J. R. Smith, "MPEG-7 multimedia description schemes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 748–759, June 2001.

[4] C.-Y. Chiu, H.-C. Lin, and S.-N. Yang, "Learning user preference in a personalized CBIR system," in *Proc. 16th Int. Conf. Pattern Recognition*, vol. 2, 2002, pp. 532–535.

[5] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, Dec. 2000.

[6] S. Antani, R. Kasturi, and R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing, and retrieval of images and video," *Pattern Recognit.*, vol. 35, no. 4, pp. 945–965, 2002.

[7] X. He, O. King, W.-Y. Ma, M. Li, and H.-J. Zhang, "Learning a semantic space from user's relevance feedback for image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 39–48, Jan. 2003.

[8] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. Int. Conf. Image Processing*, vol. 2, 1997, pp. 815–818.

[9] G. Giacinto, F. Roli, and G. Fumera, "Comparison and combination of adaptive query shifting and feature relevance learning for content-based image retrieval," in *Proc. 11th Int. Conf. Image Analysis and Processing*, 2001, pp. 422–427.

[10] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 644–655, Sept. 1998.

[11] T. S. Huang, X. S. Zhou, M. Nakazato, Y. Yu, and I. Cohen, "Learning in content-based image retrieval," in *Proc. 2nd Int. Conf. Development and Learning (ICDL'02)*, 2002, pp. 155–162.

[12] Y. Chen, X. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. Int. Conf. Image Processing*, vol. 1, 2001, pp. 34–37.

[13] G.-D. Guo, A. K. Jain, W.-Y. Ma, and H.-J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Networks*, vol. 13, pp. 811–820, July 2002.

[14] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "MindReader: Querying databases through multiple examples," in *Proc. 24th Int. Conf. Very Large Data Bases*, 1998, pp. 218–227.

[15] Y. Rui and T. Huang, "Optimizing learning in image retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 236–243.

[16] Y. Wu and A. Zhang, "A feature re-weighting approach for relevance feedback in image retrieval," in *Proc. Int. Conf. Image Processing*, 2002, pp. 581–584.

[17] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychological experiments," *IEEE Trans. Image Processing*, vol. 9, pp. 20–37, Jan. 2000.

[18] M. Oge and F. Borko, "MUSE: A content-based image search and retrieval system using relevance feedback," *Multimedia Tools Appl.*, vol. 17, pp. 21–50, 2002.

[19] I. King and Z. Jin, "Integrated probability function and its application to content-based image retrieval by relevance feedback," *Pattern Recognit.*, vol. 36, no. 9, pp. 2177–2186, 2003.

[20] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Trans. Image Processing*, vol. 12, pp. 924–937, 2003.

[21] J.-L. Tao and Y.-P. Hung, "A Bayesian method for content-based image retrieval by use of relevance feedback," in *Proc. 5th Int. Conf. Recent Advances in Visual Information Systems, (VISUAL)*, vol. 2314, Lecture Notes in Computer Science, 2002.

[22] N. Vasconcelos, "Decision-theoretic image retrieval with multi-resolution mixtures," Cambridge Research Lab., Cambridge, MA, Tech. Rep. CRL 2002/04, 2002.

[23] F. Qian, M. Li, L. Zhang, H.-J. Zhang, and B. Zhang, "Gaussian mixture model for relevance feedback in image retrieval," in *Proc. Int. Conf. Multimedia and Expo.*, vol. 1, 2002, pp. 229–232.

[24] M. R. Naphade, I. V. Kozintsev, and T. S. Huang, "Factor graph framework for semantic video indexing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 40–52, Jan. 2002.

[25] P. Muneesawang and L. Guan, "Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture," *IEEE Trans. Neural Networks*, vol. 13, pp. 821–834, July 2002.

[26] L. Wu, C. Faloutsos, K. P. Sycara, and T. R. Payne, "FALCON: Feedback adaptive loop for content-based retrieval," in *Proc. 26th Int. Conf. Very Large Data Bases*, 2000, pp. 297–306.

[27] H. Geerts, J.-M. Geusebroek, A. Smeulders, and R. Van Den Boomgaard, "Color invariance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1338–1350, Dec. 2001.

[28] B. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 703–715, June 2001.

[29] S.-F. Chang, T. Sikora, and A. Purl, "Overview of the MPEG-7 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 688–695, June 2001.

[30] P. V. Beek, J. R. Smith, T. Ebrahimi, T. Suzuki, and J. Askelof, "Metadata-driven multimedia access," *IEEE Signal Processing Mag.*, vol. 20, pp. 40–52, Mar. 2003.

[31] B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. New York: Wiley, 2002.

[32] A. Kushki, "An interactive framework for content-based image retrieval," Master's thesis, Univ. Toronto, Toronto, ON, Canada, 2003.

[33] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Retrieval from artistic repositories using a decision fusion framework," *IEEE Trans. Image Processing*, vol. 13, pp. 277–292, Mar. 2004.

[34] J. Smith, "Quantitative assessment of image retrieval effectiveness," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 52, no. 11, pp. 969–979, 2001.

[35] R. Fagin, "Combining fuzzy information from multiple systems," *J. Comput. Syst. Sci.*, vol. 58, no. 1, pp. 83–99, 1999.

Azadeh Kushki (S'02) received the B.A.Sc. and M.A.Sc. degrees in 2002 and 2003, respectively, from the University of Toronto, Toronto, ON, Canada, where she is currently working toward the Ph.D. degree.

Her research interests include content-based image retrieval, pattern recognition, and interactive learning in multimedia systems.

Panagiotis Androutsos (S'03) received the B.A.Sc, and M.A.Sc degrees in 1997 and 1999, respectively, from the University of Toronto, Toronto, ON, Canada, where he is currently working toward the Ph.D. degree.

His research interests include content-based image retrieval, MPEG-7, distributed indexing, relevance-feedback techniques, image analysis and enhancement, and perceptual encoding.

Konstantinos N. Plataniotis (S'90–M'92) received the B. Eng. degree in computer engineering from the Department of Computer Engineering and Informatics, University of Patras, Patras, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology, Melbourne, FL, in 1992 and 1994, respectively.

He was with the Digital Signal Image Processing Laboratory, Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Cananda, from 1995 to 1997. From August 1997 to June 1999, he was an Assistant Professor with the School of Computer Science at Ryerson University, Toronto. While at Ryerson, he served as a Lecturer in 12 courses in industry and continuing education programs. Since 1999 he has been with the University of Toronto. He is currently an Assistant Professor at the Edward S. Rogers Sr. Department of Electrical Computer Engineering where he researches and teaches adaptive systems and multimedia signal processing. He is the Bell Canada Junior Chair in Multimedia, a Nortel Institute for Telecommunications Associate, and Adjunct Professor with the School of Computer Science at Ryerson University. He coauthored, with A.N. Venetsanopoulos, *Color Image Processing and Applications* (New York: Springer Verlag, May 2000). He is a contributor to six books, and has published more than 200 papers in refereed journals and conference proceedings in the areas of multimedia signal processing, image processing, biometrics, communications systems and stochastic estimation.

Dr. Plataniotis was a member of the IEEE Technical Committee on Neural Networks for Signal Processing. He was the Technical Co-Chair of the Canadian Conference on Electrical and Computer Engineering (CCECE 2001), May 2001. He is the Technical Program Co-Chair of the Canadian Conference on Electrical and Computer Engineering (CCECE 2004), May 2004, and the Chair-elect for the IEEE Canada—Toronto Section.

Anastasios N. Venetsanopoulos (S'66–M'69–SM'79–F'88) received the B.Eng. degree in electrical and mechanical engineering from the National Technical University of Athens (NTU), Greece, in 1965, and the M.S., M.Phil., and Ph.D. degrees in electrical engineering from Yale University, New Haven, CT, in 1966, 1968, and 1969, respectively. He joined the Department of Electrical and Computer Engineering of the University of Toronto, Toronto, ON, Canada, in September 1968 as a Lecturer and was promoted to Assistant Professor in 1970, Associate Professor in 1973, and Professor in 1981. He has served as Chair of the Communications Group and Associate Chair of the Department of Electrical Engineering. From July 1997 to June 2001, he was Associate Chair: Graduate Studies of the Department of Electrical and Computer Engineering (ECE) and was Acting Chair during the spring term of 1998–1999. In July 1999, he became the Inaugural Chairholder of the Chair in Multimedia, ECE Department. Since July 2001, he has been serving as the 12th Dean of the Faculty of Applied Science and Engineering of the University of Toronto. He was on research leave at the Imperial College of Science and Technology, the NTU of Athens, the Swiss Federal Institute of Technology, the University of Florence, and the Federal University of Rio de Janeiro, and has also served as Adjunct Professor at Concordia University, Montreal, QU, Canada. He has served as Lecturer in 138 short courses to industry and continuing education programs and as Consultant to numerous organizations. He is a contributor to 31 books, a co-author of *Nonlinear Filters in Image Processing: Principles Applications*, *Artificial Neural Networks: Learning Algorithms, Performance Evaluation and Applications*, *Fuzzy Reasoning in Information Decision and Control Systems* , and *Color Image Processing and Applications*, and has published over 740 papers in refereed journals and conference proceedings on digital signal and image processing and digital communications.

Dr. Venetsanopoulos has served as Chair on numerous boards, councils, and technical conference committees of the IEEE, including the Toronto Section (1977-1979) and the IEEE Central Canada Council (1980-1982). He was President of the Canadian Society for Electrical Engineering and Vice President of the Engineering Institute of Canada (EIC) (1983-1986). He was a Guest Editor or Associate Editor for several IEEE journals and the Editor of the Canadian Electrical Engineering Journal (1981-1983). He is a member of the IEEE Communications, Circuits and Systems, Computer, and Signal Processing Societies, as well as a member of Sigma Xi, the Technical Chamber of Greece, the European Association of Signal Processing, the Association of Professional Engineers of Ontario (APEO). He was elected as a Fellow of the IEEE "for contributions to digital signal and image processing." He is also a Fellow of the EIC, and was awarded an Honorary Doctorate from the National Technical University of Athens in October 1994. In October 1996, he was awarded the Excellence in Innovation Award from the Information Technology Research Centre of Ontario and Royal Bank of Canada, "for innovative work in color image processing and its industrial applications." In November 2000, he received the Millennium Medal of IEEE. In April 2001, he became a Fellow of the Canadian Academy of Engineering. In 2003, he received the MacNaughton Award, the highest award of the Canadian IEEE, and in 2003 he was the Chair of the Council of Deans of Engineering of the Province of Ontario (CODE).