

Retrieval of Images From Artistic Repositories Using a Decision Fusion Framework

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

Abstract—The large volumes of artistic visual data available to museums, art galleries, and online collections motivate the need for effective means to retrieve relevant information from such repositories. This paper proposes a decision making framework for content-based retrieval of art images based on a combination of low-level features. Traditionally, the similarity among two images has been calculated as a weighted distance between two feature vectors. This approach, however, may not be mathematically and computationally appropriate and does not provide enough flexibility in modeling user queries. This paper proposes a framework that generalizes a wide set of previous approaches to similarity calculation including the weighted distance approach. In this framework, image similarities are obtained through a decision making process based on low-level feature distances using fuzzy theory. The analysis and results of this paper indicate that the aggregation technique presented here provides an effective, general, and flexible tool for similarity calculation based on the combination of individual descriptors and features.

Index Terms—Content-based image retrieval, feature combination, fuzzy aggregation operators, MPEG-7 visual descriptors, similarity calculations.

I. INTRODUCTION

THE ubiquity of digital imaging devices as well as the maturity of the worldwide network infrastructure has led to a dramatic change in the way which people share their visual media. Individuals are now opting to share and make their media available electronically. Professional photographers are choosing to make their portfolios available to the world by showcasing their work on web-sites as well as distributing their photographs on cost-effective reproducible optical media. Furthermore, museums, art galleries, and online collections are starting to embrace the digital revolution for various reasons. Digitization of art and historical treasures permits such institutions to keep visually informative catalogues of their holdings, perform virtual restorations and experiments, enhance their on-site experience through the use of digital kiosks, make available the viewing of items in storage due to space constraints, and enable the sharing of their cultural assets to a global audience through the use of the Internet. Large institutions such as these typically have a great number

of cultural assets on hand. This translates to a large quantity of digital media.

Without a doubt, huge digital collections such as these definitely require fast, effective, and flexible tools that are geared toward the search and retrieval of relevant items from what can now be referred to as a cultural digital library. This implies a need for computational techniques for the automatic (and even manual) extraction and description of content, as well as algorithms for effective search creation and user-interaction. In traditional search schemes, content media was manually annotated using text-based descriptions of the contents of each item (e.g. an image). Retrievals were then performed by using input keywords which were matched to the indexed descriptions using standard text-matching algorithms [1]. A text-based approach, however, is both expensive and ambiguous due to the fact that such descriptions of image content are highly subjective [2], extremely time-consuming and subject to inconsistencies during translation. The subjectivity associated with such an approach is also easily realized in the context of a digital art collection where specific works can inspire different feelings and emotions, or particular colors and color palettes are indexed using un-intuitive terms (e.g. ‘cadmium yellow’).

The limitations of text-based search schemes have motivated the use of content-based media retrieval as a tool for locating relevant items within large databases. In this approach, visual media is represented using descriptions of low-level visual features such as color, texture, shape, motion, etc. It must be noted that a visual feature can be described in various manners. For example, color can be represented as a histogram or a set of dominant color vectors. Each of these manifestations or representations of a feature is referred to as a *feature descriptor*. These low-level feature descriptions are easily understandable by both humans and machines and may be automatically extracted using computational techniques. The media is then compared based on one or on a combination of these low-level features. Due to the semantic gap between machine representations and human understanding, feature selection and extraction becomes a daunting process. As a result, much of the work in the area of content-based media retrieval has been focused on devising effective solutions for the generation of semantically meaningful representations of the media at hand (e.g. effective description of a particular painting’s color content). To unify these efforts, the International Standards Organization (ISO) has finalized a standard set of tools for providing effective representations of general multimedia data in terms of low-level features [3]. This standard which is referred to as MPEG-7, is not limited to visual data and also takes into account auditory data to

Manuscript received February 14, 2003; revised October 7, 2003. This work was supported by the National Science and Engineering Council (NSERC). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ioannis Pitas.

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

Digital Object Identifier 10.1109/TIP.2003.821350

ensure that it applicable to the entire range of all audio-visual media. It is important to note here that the scope of MPEG-7 is limited to the *description* of constituent features (e.g. the description of color, shape, motion) and does not dictate standard ways whereby these descriptions are to be compared (feature similarity), combined (query formulation) or further manipulated and processed.

In contrast to feature extraction and description, very little work has been done to address the problem of similarity measurement [4]. Although outside its scope, the MPEG-7 standard does provide guidelines for similarity measurements. These guidelines, however, are not specifically geared toward the applications or needs of the artistic community, and are used more out of convenience and simplicity. In order to allow users to formulate broad and intuitive queries which consist of inclusions and exclusions of desirable and undesirable features, the use of multiple features in some combination is necessary. Since the use of multiple-feature based queries results in higher retrieval accuracy, it is very desirable for a retrieval system to permit hybrid query generation. This is even more evident in cases where constituent features are of a complementary nature and thus more detailed descriptions of digital library elements are achievable. Although existing systems, such as those reviewed in Section II, do often provide means for retrievals based on multiple features, they lack a systematic approach for responding to queries based variegated combinations of these features. The need for the generation of intuitive hybrid queries is very overt in the context of visual digital libraries where the array of the content may be extremely wide. The use of hybrid queries permits the immediate exclusion of large portions of a cultural database through the exclusion of features common to unwanted database elements (e.g. intelligently selecting admissible textures could suppress the retrieval of sculptures/carvings while promoting the retrieval of paintings and drawings).

In light of the aforementioned shortcomings, this work proposes a Unified Framework for Similarity Calculation (UFSC) for the combination of low-level feature similarities encompassing a wide set of previous techniques. The problem of similarity calculation is reformulated as a problem of decision making where a decision must be made upon the similarity between two elements according to evidence provided by each feature-level similarity. Unlike approaches which directly combine feature similarities to arrive at a final, overall similarity value, *decisions* made according to individual similarities are *aggregated* to achieve the same effect but with superior flexibility and generality. It is important to mention here that similarity calculations are *not* performed in a high dimensional space, but rather, within a *decision space*. Lastly, because of the fact that the proposed system utilizes an entire family of aggregators for feature combination, it exhibits a high level of flexibility through the use of a wide range of tools for modeling user intentions. These tools are very general in nature, but for the purposes of this paper, they are applied in the context of a cultural database consisting of artistic images (e.g. paintings, engravings, etc.).

The rest of this paper is organized as follows: Section II provides an overview of the existing methods, discusses some

of their limitations and the motivation for using the proposed framework. Section III provides an overview of the proposed approach while image representation and the feature set used for indexing and retrieval are presented in Section IV. The details of the framework are presented in Sections V and the experiments as well as results are given in Section VI and discussed in Section VII. Finally, Section VIII concludes the paper and provides some directions for future work.

II. PRIOR WORK

Within the fine art world, it comes as no surprise that a field which relies so much on visual cues would be richly enhanced and empowered through digital image processing. Traditionally, the realm of image processing has provided effective image enhancement techniques for edge detection, and image sharpening, as well as image correction techniques such as restoration, and filtering. More recently, however, efforts by researchers from different disciplines have been focused on the combination of a variety of schema in order to perform specific tasks more complementary and in-tune with the needs of the fine art world. Automatic reconstruction of fresco fragments, and computer-guided cleaning [5] and crack-removal of paintings, are extremely useful and highly applicable to the reconstruction of artwork. Despite the fact that these methods are virtual in nature, their value lies in the fact that they furnish workers with an excellent means to obtain a priori estimates or targets for works being restored. Furthermore, such techniques are extremely useful in providing visually improved versions of damaged artworks for exhibition through the World Wide Web [6]. To further enhance the experiences for both web-based and on-site museum patrons, content-based indexing of cultural assets such as paintings and sculptures provides for richer, and more flexible visits for audiences. Through the use of interactive kiosks delivering informative presentations as well as instantaneous access to vast digital versions of cultural assets, visitors are able to immediately search for similar works using criteria such as actual color and texture contents which extend beyond simple facts and indexed keywords. It is important to note, however, that the content-based indexing of artwork also benefits the art community from both academic and administrative standpoints where the writing of commentaries and criticisms can be expedited through vast, shared, and easily searched digital art databases.

In Content-Based Image Retrieval (CBIR), certain visual features such as shape, color, and texture are extracted to characterize images in a machine-understandable format. Each of the visual features is represented using one or more *feature descriptors*. During the retrieval, users specify queries by providing the system with an example image, feature primitives, or even a sketch containing the concepts or features of interest. The visual features and descriptors of the query are compared to those of the images in the available database [4], [7] in efforts to rank each indexed image according to its similarity/distance to the query. Finally, a ranked set of the best matches to a user's query is returned.

Although retrievals can be performed based on single features [8], [9], more perceptually meaningful results are obtained if a combination of features is used [10]. This is due to the fact that complementary features convey information about various properties of the indexed images, thereby helping to bridge the semantic gap between the human interpretation of images versus their machine-based representation. The use of multiple features or descriptors gives rise to several possibilities for similarity calculations. In one approach, the dissimilarity between images is calculated directly from the combination of features. In this case, descriptors of the same feature are combined to form *feature vectors* and the feature-level dissimilarities calculated as the distance between these vectors are combined to obtain the overall dissimilarity. *PicHunter* [11] which employs the Manhattan distance in determining similarity between feature vectors, obtains an *overall* distance between two images by combining feature distances using a simple weighted average scheme.

In other approaches, the similarity measurement is performed at the descriptor-level. *PicToSeek* [12] combines color and shape descriptors into a histogram and expresses the overall distance measure as a normalized histogram intersection. Another way of performing descriptor-level aggregation is to stack the descriptor values as a single, large vector and calculate the distance between these ‘*super-vectors*’ in a high dimensional feature space. To this end, the Euclidean distance is commonly used in a large number of CBIR systems [13]–[19] for finding the distance between two feature vectors. In the case where the feature vector contains only a single descriptor, this distance is used to calculate dissimilarity between the two images directly. On the other hand, if more than one descriptor is used, the Euclidean distance is employed not only to calculate individual descriptor distances but also to combine the descriptor-level distances and provide an overall *aggregated* dissimilarity between two images. The use of the Euclidean distance for distance combination, however, does not always provide perceptually meaningful results as the assumption of a Euclidean feature space is generally incorrect [7], [20].

In light of this important point, it is valuable to note that when distance measures are devised, the nature of the feature space needs to be taken into consideration since unilaterally employed distance measures may not necessarily be meaningful for all feature descriptors. With this in mind, it is encouraging to know that several systems use distinct distance measures at the descriptor-level before combining resulting distances via some kind of *aggregation operator* to form the overall distance [11], [21]–[27].

In performing this combination/aggregation of feature distances, the use of the weighted average imposes of a certain structure on the conceptual user queries that can be modeled by a CBIR system. Due to the diverse nature of user requirements in retrieval from artistic repositories, however, a wide range of queries may need be formulated by the user. For example, if the goal is to find paintings a certain time period, it may be required that the results match the query in color composition or texture as these features are indicative of the period of the artistic work. In order to further narrow down the search, the user may wish to specifically exclude certain texture patterns or color palettes.

Considering user requirements such as these, conceptual queries are classified into the following four categories:

- 1) *All* features of a given image should match those of the query.
- 2) *At least one* feature of the given image matches the query.
- 3) *Some* of the features match the query.
- 4) Certain features must *not* match the query.

It will be shown in Section VII that the disadvantage of employing the weighted average for obtaining overall image similarity from feature distances is that its logical function cannot be easily classified into any of the above mentioned categories. In contrast we have chosen a *fuzzy logic framework* as an alternative since it offers great flexibility in modeling the inherent vagueness in human perception and various logical concepts as understood by humans. In this framework, the four categories of conceptual queries can be readily modeled using the logical AND for *all*, OR for *at least one*, a mean operator for *some*, and the logical NOT operator for query exclusions. Systems such as *MARS* [28], *BlobWord* [29], and *Garlic* [30] attempt to address these issues by extending the concept of Boolean expressions used in text retrieval to image retrieval. *MARS* [28] represents the user query as a fuzzy Boolean expression over image features and similarity between images becomes the evaluation of this expression using the calculated similarities. *Garlic* [30] uses the min and max as the fuzzy logical AND and OR, respectively, to address conjunctive and disjunctive queries. Although Boolean expression type approaches can model a wide range of queries, they only provide the logical functionality of the *AND*, *OR*, and *NOT* connectives, and exhibit no behavior allowing them to model compromises. Yet, decision making which is based on the aggregation of criteria is generally not conjunctive or disjunctive [31], indicating a need for the availability of an operator which has the potential to make such compromises.

III. OVERVIEW OF THE PROPOSED APPROACH

Due to the fact that the distance measures employed at the descriptor-level are functions of the descriptors used, distances are calculated individually for each descriptor. Instead of performing a direct aggregation of these distances, *decisions* based on the descriptor-level distances are made in order to determine image similarity. Performing aggregation at the decision-level instead of a descriptor-level provides the added benefit of utilizing fuzzy logic principles that can be used to model conceptual queries. Conceptual queries are of great importance especially when the user has a very good idea of a specific image or of a particular type of image that needs to be retrieved. Searches which are based on the inclusion and/or exclusion of particular palette colors [33], queries using intelligently oriented shape primitives, and combinations of these and additional conceptual queries are all indispensable in the world of cultural databases. Since fuzzy conceptual operators stray away from traditional binary concepts, they are able to model some elements of human perception where things are not always black and white. Similarity decisions based on descriptor-level distances need to be aggregated in some way to provide an overall decision about the similarity between images. It will be shown in Section VI that

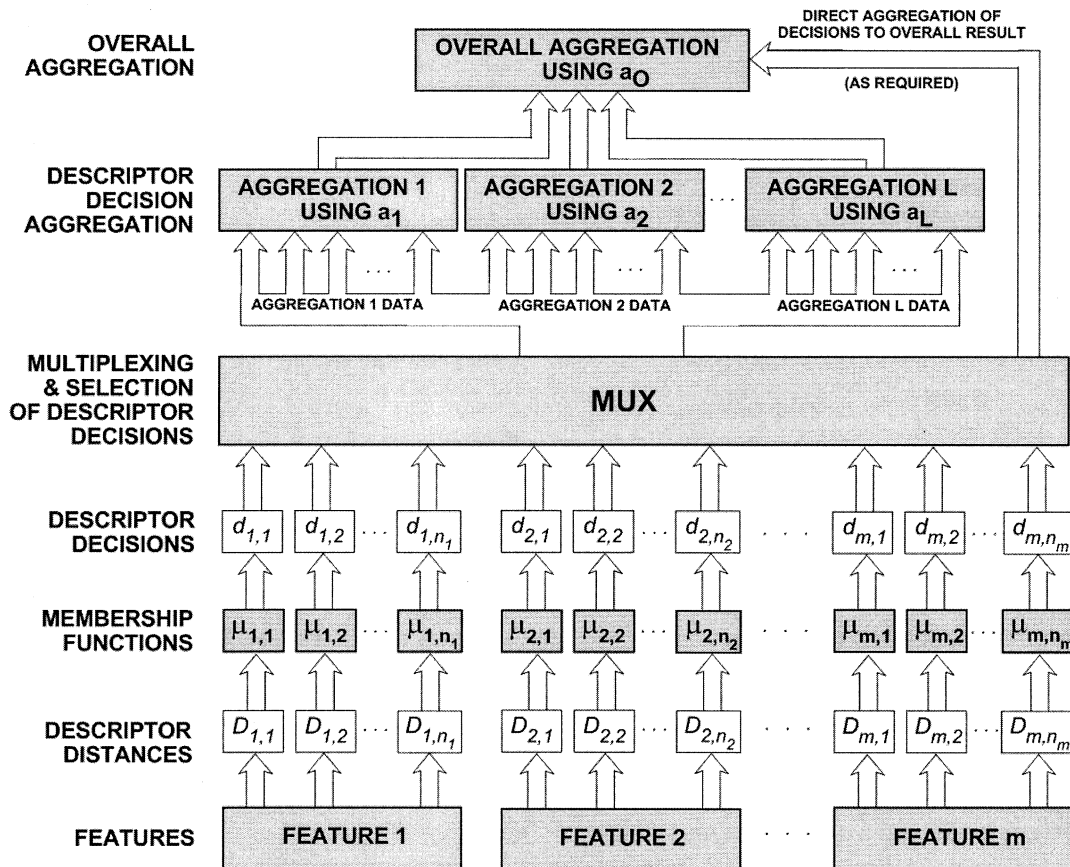


Fig. 1. UFSC overall structure. The multiplexing element (MUX) performs the actual selection of descriptor decisions $d_{i,j}$ and passes them to the appropriate aggregation mechanism a_k before the overall aggregation is performed (if necessary). Descriptor distances $D_{i,j}$ are passed through membership functions $\mu_{i,j}$ in order for descriptor decisions to be obtained.

the method employed to perform this combination has a great effect on the overall retrieval accuracy. Furthermore, it should be mentioned that the aggregation operator achieving the highest retrieval accuracy is not identical across different queries, but rather is dependent on both the features and distances used as well as the user. In light of this conclusion, the UFSC moves away from the use of a single aggregator for decision combination and employs a *set* of aggregation operators for addressing various logical conceptual queries, where the specific operator used is chosen according to the nature of the query. This is illustrated in Fig. 1 which depicts the overall organization of the framework discussed herein. It will be shown that the weighted average and the Boolean expression approaches can both be viewed as specific instances of the proposed framework.

IV. IMAGE REPRESENTATION

The low-level features selected for image representation are color and texture. Color is an important feature as it provides visual cues for the delineation and recognition of objects [34], and permits the creation of queries according to color distribution, or according to color palettes. Texture is an appropriate feature for use with artistic databases as it can be used to distinguish works such as sculptures and paintings through the textures which characterize particular artistic techniques (e.g. pointillism) or the medium used.

Although for the purposes of this paper we conform to the MPEG-7 standard in terms of the feature descriptors (all of which are outlined in [35]), it is important to state here that the proposed framework is applicable to *any* set of features and descriptors as chosen by the designer. This concept is illustrated in the bottom half of Fig. 2 which shows that MPEG-7 provides only *one* possible set of features and descriptors which can be used by the framework (features and descriptors spanning different feature sets may also be mixed and matched as desired). The upper half of this figure depicts the connection of the UFSC to a retrieval front-end which consists of the user, the retrieval interface, and a relevance feedback engine which would serve the purpose of updating UFSC parameters according to user input.

Three of the MPEG-7 descriptors are used for color; the Color Structure Descriptor (CSD), the Dominant Color Descriptor (DCD), and the Color Layout Descriptor (CLD). The MPEG-7 texture descriptors used are the Edge Histogram Descriptor (EHD) and the Homogeneous Texture Descriptor (HTD). Details regarding both the color and texture descriptors employed are discussed below.

A. Color

The color descriptors used in this work were selected so that they provide complementary functions to each other addressing

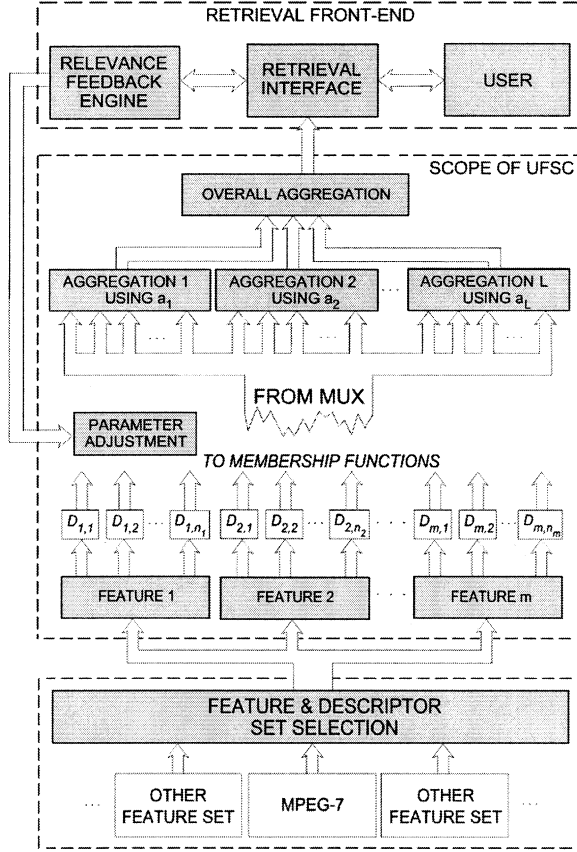


Fig. 2. Interfacing with the user and the selection of features and descriptors is outside the scope of the UFSC. The set of features and descriptors outlined by the ISO in the MPEG-7 standard are only *one* of many possible choices, and their selection and combination can be made at the discretion of the designer. The retrieval front-end module in the upper half, depicts a user-interface which could conceivably enable retrieval supporting user-provided feedback to iteratively adjust UFSC parameters.

a variety of needs in a CBIR system geared toward an artistic application. The DCD is used for a compact representation of the color composition of the images. Furthermore, this descriptor is of interest when retrieving from artistic repositories as it allows for the formulation of queries based on single or multiple colors. The CSD is extracted in the Hue-Max-Min-Difference (HMMD) space and thus closely matches the human perception of color. This can be contrasted to the use of the $YCbCr$ space in the extraction of the CLD descriptor that is resolution-invariant and very compact. Since these three color descriptors employ different color spaces, the information that cannot be captured in one space may be captured by another. Furthermore, one particular descriptor may perform best for a specific application or query. Yet, because of the fact that a CBIR system must address various applications and query types, it becomes necessary to have a *collection* of descriptors so that the most appropriate descriptors can be dynamically selected by the UFSC to ensure optimal retrieval performance. Thus, for each query, irrelevant descriptors can simply be dropped and not considered during similarity calculations.

The MPEG-7 Color Structure Descriptor (CSD) provides information regarding color distribution as well as localized spatial color structure in the image. The image is represented by a

modified color histogram which is different from a traditional histogram since it incorporates the spatial distribution of each color. This is achieved through the compilation of histogram data in the quantized HMMD space by sliding a structuring element across the image [36]. The distance between two CSD histograms for images Q and I is calculated using the generalized form of the Minkowski distance D_M shown in (1) with the parameter L set to unity, and where x and y are two vectors both of dimensionality N_{gen} ,

$$D_M(x, y) = \left[\sum_{i=0}^{N_{gen}} |x_i - y_i|^L \right]^{\frac{1}{L}}. \quad (1)$$

This L_1 -norm formulation of the distance D_{CSD} between CSD histograms is shown in (2) where $H_{Q,i}$ represents the i^{th} bin of the color structure histogram for image Q and a 256-bin histogram has been used for each image [37].

$$D_{CSD}(Q, I) = \sum_{i=0}^{255} |(H_{Q,i} - H_{I,i})|. \quad (2)$$

The Dominant Color Descriptor (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 and complements the local structure information provided by the CSD. The recommended distance measure [37] to be used with the DCD is:

$$D_{DCD}(Q, I) = \left(\sum_{i=1}^{N_Q} p_{Ii}^2 + \sum_{j=1}^{N_I} p_{Qj}^2 - \sum_{i=1}^{N_Q} \sum_{j=1}^{N_I} 2a_{Ii, Qj} p_{Ii} p_{Qj} \right)^{\frac{1}{2}}, \quad (3)$$

where the similarity coefficient $a_{k,l}$ between two RGB color vectors, c_k and c_l , is calculated as in (4).

$$a_{k,l} = \begin{cases} 1 - \frac{d_{k,l}}{d_{max}} & d_{k,l} \leq T_d \\ 0 & d_{k,l} > T_d \end{cases}. \quad (4)$$

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 Color Layout descriptor complements the CSD and the DCD by providing information about the the spatial color distribution within images. In order to extract this descriptor, the image is divided in 64 blocks. One representative color from each of the blocks is used to generate an 8×8 icon image. The Discrete Cosine Transform is then applied to the icon image in the $YCbCr$ space to generate a representation of the image. In this work, six coefficients corresponding to the Y component and three for each of the chromatic components are used. The matching of images based on the CLD is performed through the calculation of D_{CLD} as shown below [37]:

$$D_{CLD}(Q, I) = \sqrt{\sum_i w_{yi} (Y_{Qi} - Y_{Ii})^2} + \sqrt{\sum_i w_{cbi} (Cb_{Qi} - Cb_{Ii})^2} + \sqrt{\sum_i w_{cri} (Cr_{Qi} - Cr_{Ii})^2}. \quad (5)$$

In (5), w_i represents the weight associated with coefficient i . The low-frequency components receive the largest weights in order to reflect human perceptual characteristics, paralleling the approach of JPEG quantization [32].

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, nonhomogeneous areas (such as complicated natural scenery) while the HTD operates on homogeneous texture regions (e.g., portraits).

The Edge Histogram Descriptor captures the edge distribution within an image. It extracts edge information from 16 sub-images and categorizes it into five classes for each sub-image: horizontal, vertical, 45° , 135° and nondirectional [36]. The similarity between two histograms, D_{EHD} , is determined by calculating the L_1 -norm of the 80-dimensional feature vectors H_Q and H_I [37].

The Homogeneous Texture Descriptor (HTD) represents the mean and energy deviation of 30 frequency channels modeled using Gabor functions [37]. The distance D_{HTD} between two vectors T_Q and T_I is calculated as shown in (6) where $\alpha(k) = 1$ was determined experimentally.

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

C. Representative Sets

After calculating descriptor-level distances, a similarity decision is made based on each descriptor. Similarity decisions represent the grade of membership to the set of images similar to a given query and may be obtained from the distances using a membership function. These membership functions are illustrated in Fig. 1 as $\mu_{i,j}$. The form of the membership is strongly dependent on the features and distances used. A detailed discussion on the choice of the membership functions can be found in [38].

In [31], an exponential membership is used in conjunction with Euclidean-types of distances for color similarity calculation. The experiments of Section VI indicate that this membership function provides acceptable results for the descriptors used here. As a result, the exponential membership function will be employed for the generation of decisions:

$$\mu_I = d(Q, I) = \exp\left(\frac{-D(Q, I)}{\alpha}\right), \quad (7)$$

where $D(Q, I)$ is the distance between images Q and I with respect to a given descriptor, and α is a normalization factor. The form of the membership function and its parameters α and $D(Q, I)$, affect the retrieval accuracy. The distance $D(Q, I)$, is obtained from the MPEG-7 distance measures for each descriptor presented in the previous section. In Section VI we examine other possible forms of the membership function and report on the effects on the overall retrieval accuracy.

The normalization factor α is used to adjust the spread of the values in the range $[0, 1]$. In Section VI, different methods for calculating this parameter are considered. Based on these results, α is calculated as in (8) where $S = (s_1, s_2, \dots, s_n)$ and s_i represents the distance of the i th image to the query in the given descriptor.

$$\alpha = \frac{\text{median}(S)}{-\log(0.5)}. \quad (8)$$

This choice of α causes the image similarity distribution to be centered around 0.5 with an equal number of images on either side of the center, thus ensuring that all descriptors contribute equally to the final decision. If the exclusion of a particular feature is required, the similarity decision of (7) is further modified as shown in (9) to reflect the logical NOT operator:

$$\mu_N = 1 - \mu_P. \quad (9)$$

The similarity decisions obtained through (7) must be combined in some way to provide an overall decision. In other words, a decision regarding the similarity between images must be made based on the evidence provided by each descriptor. There are various possibilities for performing the aggregation depending on the nature of the features and descriptors used. For example, all descriptor-decisions can be aggregated directly, or certain representative descriptors can be chosen for each feature and combined to obtain the final decision. Thus, as illustrated in Fig. 1, a multiplexing (MUX) element is employed in our framework which empowers the designer with the flexibility and freedom to choose *which*, and *how* descriptor decisions are to be aggregated. This means that descriptor decisions can be combined in any fashion preferred by the designer or according to specific requirements of the application. In addition to generating various combinations of descriptors, the MUX may simply re-route each input to an output using a one-to-one mapping. This implies that individual descriptor decisions can be directly aggregated at the top-level either for partial contribution to the overall result, or for the creation of an overall result through a direct combination of appropriate decisions in a single step. Since descriptors of a given feature report information on various aspects of the same visual characteristic of an image (e.g. color as opposed to texture), we have chosen to perform the aggregation in a hierarchical manner. In this scheme, the descriptor decisions for each feature are first aggregated. These feature decisions are then combined to arrive at the final decision. It must be emphasized that the choice of the aggregation scheme is dependent on the nature of the features and requirements of the system. This hierarchical approach is shown in Fig. 3 which shows how the aggregation of color descriptor decisions is performed independently of textural descriptor decisions before a final result is obtained.

Following the hierarchical aggregation scheme, the image I is characterized by a group of *sets of descriptors decisions* known as the *representative set*, R_I . Each descriptor decision set in R is used to represent the similarity decisions made based on a particular feature. The concept of a representative set is demonstrated in (10) where $d_{i,j}$ represents the decision based on the i th descriptor of the j th feature and n_j and m represent the number

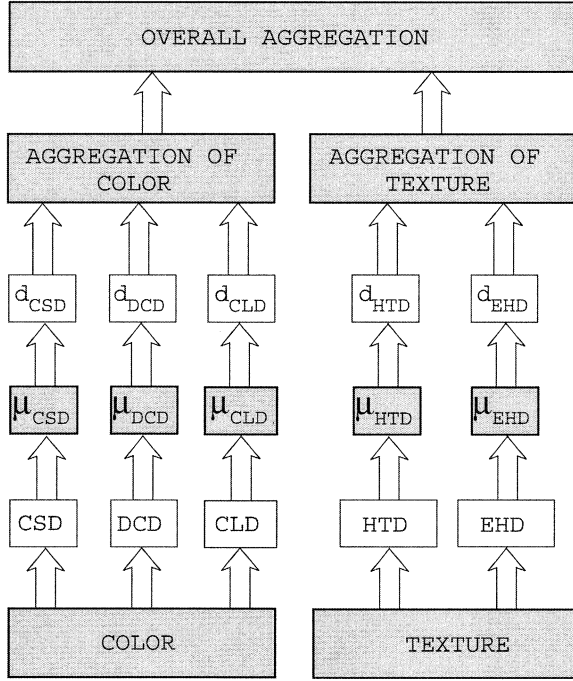


Fig. 3. Hierarchical formulation of the UFSC as discussed in Section IV-C. Descriptor decisions d_{CSD} , d_{HTD} , d_{DCD} , etc. are aggregated according to their features before color and texture aggregations are combined for a final, overall result.

of decision elements for feature i and the total number of features, respectively,

$$R_I = \left\{ \underbrace{\{d_{1,1}, d_{2,1}, \dots, d_{n_1,1}\}}_{\text{feature 1 decisions}}, \dots, \underbrace{\{d_{1,m}, d_{2,m}, \dots, d_{n_m,m}\}}_{\text{feature m decisions}} \right\}. \quad (10)$$

The form of (10) reiterates the fact that aggregation is to be performed in a hierarchical fashion on particular aggregated feature subsets before the generation of an overall feature decision. This concept is made more concrete by Fig. 3 which demonstrates the two-stage aggregation scheme using the descriptors employed herein. The representative set based on the features and descriptors used here is shown in (11).

$$\mathcal{R}_I = \left\{ \underbrace{\{CSD, DCD, CLD\}}_{\text{color decisions}}, \underbrace{\{EHD, HTD\}}_{\text{texture decisions}} \right\}. \quad (11)$$

V. AGGREGATION OPERATOR

The representative set formulation of (11) provides an indication of the degree of similarity between two images with respect to each descriptor. Yet, R_I , still remains as a set of decisions rather than a single value related to similarity that can be used to rank the images. In order to arrive at such a value, the decision elements must somehow be combined or *aggregated* using an operator. This aggregation operator should be symmetric (commutative) and monotonically nondecreasing as detailed in both [39] and [40]. Assuming that these criteria are satisfied, the choice of the aggregation operator used is ultimately left at the

discretion of the designer. In attempt to meet the needs of users, however, this choice should permit the creation of queries combining search criteria using basic and intuitive logical operators such as AND and OR.

Using a purely Boolean approach leads to two extreme aggregation cases, namely, the conjunctive and disjunctive. The conjunctive (AND) operator strictly requires the satisfaction of all criteria. In other words, the images are required to be similar with respect to all features. For example, in searching for all white marble statues, it would be required that the texture matches that of marble and the color is nearly white. The disjunctive (OR) requires that at least one of the criteria is satisfied or that the images are similar with respect to at least one descriptor. These two cases, respectively, represent severe (pessimistic) and indulgent (optimistic) behavior. In multi-criteria decision making, however, the aggregation function is generally neither purely conjunctive nor disjunctive [31] and lies somewhere between the two extremes. Depending on whether a more conjunctive or disjunctive behavior is needed, the operator should exhibit behavior closer to the logical AND or logical OR. In ensuring that all of the various types of concepts presented in Section II can be tackled, the aggregation operator should thus be chosen from an *array* of operators that stretch the full range between AND and OR. To this end, the set of *compensatory operators* [31] that allow the direct adjustment of the degree of closeness to the AND or OR are considered. Furthermore, for the purposes of this work, the family of quasilinear T-S compensatory operators introduced in [40] are used as the aggregation operators. This is a family of operators ranging between a t-norm and t-conorm, and the specific member of the family can be selected using a compensation parameter. The general form of this class of operators is:

$$QL = f^{-1}((1 - \gamma)(T(x_1, \dots, x_n)) + \gamma f(S(x_1, \dots, x_n))) \quad (12)$$

where x_1, \dots, x_n are the elements being aggregated, $\gamma \in [0, 1]$, $f : [0, 1] \rightarrow [-\infty, \infty]$ is a continuous and strictly monotonic function such that $\{f(0), f(1)\} \neq \{-\infty, \infty\}$ and T and S represent a t-norm and a t-conorm, respectively.

The t-norm operators are used to represent the intersection of fuzzy sets (AND operator) and must satisfy the properties shown in Table I [41]. In Fuzzy Theory, the t-conorm operators model the OR connective, and are characterized by the properties shown in Table I [41]. Therefore, the operator of (12) moves from the logical AND ($\gamma = 0$) to a logical OR ($\gamma = 1$). It is shown in [42] that the behavior of the aggregator varies depending on how close it is to a disjunction (OR) or to a conjunction (AND). The OR operator requires only the satisfaction of one criterion for concluding overall similarity. Thus, the decisiveness of the operator decreases when moving toward OR. For the logical AND operator, the aggregated decision is dominated by the weakest element and therefore, the discriminating power of the operator decreases as it moves toward the logical AND.

In order to gain insight into the logical behavior of the operators, the closeness to the logical AND and OR operator is considered. We, therefore, take an approach similar to [39] and define the following concepts.

TABLE I
PROPERTIES OF T-NORMS (*) & T-CONORMS (\perp)

Operator	Consistency	Commutativity	Associativity	Identity
t-norms	$0 * 0 = 0$	$x * y = y * x$	$(x * y) * z = x * (y * z)$	$x * 1 = x$
t-conorms	$1 \perp 1 = 1$	$x \perp y = y \perp x$	$(x \perp y) \perp z = x \perp (y \perp z)$	$x \perp 0 = x$

Definition 1: Let \mathcal{A} be an aggregation operator. Then, the *Andness* degree of \mathcal{A} is defined as:

$$Andness(\mathcal{A}) = \frac{\mathcal{A}(x_1, \dots, x_n) - \min(x_1, \dots, x_n)}{\max(x_1, \dots, x_n) - \min(x_1, \dots, x_n)}. \quad (13)$$

Definition 2: Let \mathcal{A} be an aggregation operator. Then, the *Orness* degree of \mathcal{A} is defined as:

$$Orness(\mathcal{A}) = 1 - Andness(\mathcal{A}). \quad (14)$$

The *Andness* measure is essentially an indicator of how close the operator is to the logical AND. Similarly, *Orness* measures the closeness of the operator to a logical OR.

The T and S functions in (12) can be chosen to be any t-norm or t-conorm. We have selected $T(x_1, \dots, x_n) = \min(x_1, \dots, x_n)$ and $S(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$ to reduce the complexity of the system. The function $f(x)$ is chosen to provide a compromise between T and S . In selecting this function, the following well known mean operators were considered: arithmetic mean $((x + y)/2)$, geometric mean (\sqrt{xy}) , harmonic mean $((2xy)/(x + y))$, and quadratic mean $(\sqrt{((x^2 + y^2)/2)})$. The *Andness* measure for each of these operators for the case of $n = 2$ (two arguments) and $x < y$ was calculated and results are summarized in Table II. It is seen that geometric, harmonic, and quadratic mean operators are biased in the sense that their *Andness* is a function of their arguments. The arithmetic mean, however, provides a constant *Andness* at 0.5 for the case of $n = 2$ and does not depend on its arguments. As result, we select $f(x) = x$ or the arithmetic mean. Using $f(x) = x$, $T(x_1, \dots, x_n) = \min(x_1, \dots, x_n)$ and $S(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$ in (12) the aggregation operator, \odot , becomes:

$$x_1 \odot \dots \odot x_n = \gamma \max(x_1, \dots, x_n) + (1 - \gamma) \min(x_1, \dots, x_n). \quad (15)$$

Theorem 1: Let \odot be the aggregation operator defined in (15). Then,

$$\begin{aligned} Andness(\odot) &= \gamma, \\ Orness(\odot) &= 1 - \gamma. \end{aligned}$$

Equation (15) represents a family of operators generated by a weighted mean of the logical AND and OR providing various degrees of compromise depending on the choice of the compensation parameter, γ . The choice of this family of operators ensures that the logical behavior of the operator is only affected by the grade of compensation, γ . As a result the *Andness* of this operator or the logical structure of the conceptual query can be directly adjusted by the user allowing for more control over query formulation. It must be emphasized that the function $f(x)$ and the T-S norms should be chosen based on the requirements

TABLE II
ANDESS OF WELL KNOWN MEAN OPERATORS

Operator	Andness
Arithmetic mean	0.5
Geometric mean	$\frac{\sqrt{xy} - x}{y - x}$
Harmonic mean	$\frac{xy - x^2}{y^2 - x^2}$
Quadratic mean	$\frac{\sqrt{x^2 + y^2} - x}{y - x}$

of the system and are not restricted to what is presented in this paper. Other manifestations of this operator can be found in [43].

We now investigate some desirable properties of the aggregator in (15).

- 1) Let \odot be the operator of (15) and $\forall i x_i \in [0, 1]$, $x'_i \in [0, 1]$. Then \odot is continuous and monotonically nondecreasing in each argument, that is:

$$\forall i x_i < x'_i \Rightarrow x_1 \odot \dots \odot x_n \leq x'_1 \odot \dots \odot x'_n.$$

The continuity property is important to ensure that the aggregation function is well behaved and that it does not produce chaotic reaction to small changes in descriptor similarities [44]. The monotonicity property is essential to logical behavior of the operator as it means that if a stronger similarity is reported by a descriptor, the overall similarity between the images also increases.

- 2) The aggregation operator \odot is commutative. That is, if $\forall i x_i \in [0, 1]$ and σ is any permutation,

$$x_1 \odot \dots \odot x_n = x_{1(\sigma)} \odot \dots \odot x_{n(\sigma)}$$

The commutativity property states that the overall decision does not depend on the order that the descriptor values were aggregated. This is important since the similarity between images does not depend on the order that feature similarities are considered.

- 3) The aggregation operator \odot satisfies the strictness property. That is, $\forall i x_i \in [0, 1]$, $\gamma \in [0, 1]$,

$$x_1 \odot \dots \odot x_n = 1 \text{ iff } \forall i x_i = 1.$$

By satisfying the strictness property, the operator guarantees that the only time that a full similarity is achieved is when the two images are identical.

- 4) Let $X = (x_1, \dots, x_n)$ where the x_i 's are the elements being aggregated. Then, $\{x_i | x_i \neq \max(X), x_i \neq \min(X)\}$ are neutral elements of the \odot aggregator. The existence of a neutral element means that there exists an element that does not affect the result of aggregation. This property is especially useful with the chosen membership function. In this case, if the standard deviation of a particular descriptor is small, (7) forces the similarity values to be close to 0.5 or the neutral element causing this descriptor to have little or no effect during aggregation via (15). As a result, descriptors with low discrimination power for a particular query will automatically be ignored during aggregation.

Input:

Descriptors for query image Q , and candidate I .

Output:

Similarity between Q and I .

UFSC Algorithm:

Step 1. Calculate descriptor distances:

$$D_{CSD}, D_{CLD}, D_{DCD}, D_{EHD}, D_{HTD}.$$

(Section IV).

Step 2. Generate similarity decisions:

$$(d_{CSD}, d_{CLD}, d_{DCD}, d_{EHD}, d_{HTD}).$$

(Equation 7)

Step 3. Aggregate feature descriptor decisions:

$$S_{colour} = d_{CSD} \odot d_{CLD} \odot d_{DCD} \text{ (use } \gamma_{colour}\text{)}$$

$$S_{texture} = S_{EHD} \odot S_{HTD} \text{ (use } \gamma_{texture}\text{)}$$

Step 4. Aggregate feature decisions:

$$S(Q, I) = S_{colour} \odot S_{texture} \text{ (use } \gamma\text{)}$$

Fig. 4. Step-by-step outline of the UFSC operation.

A summary of the UFSC algorithm using the aggregation operator presented in this section is given in Fig. 4.

As shown in Section VI, the compensation parameter used with the aggregation operator has a great effect on the performance of the system. Thus, it is important that this parameter is chosen properly. This selection can be either performed manually or automatically. The system user selects the value of the compensation parameter considering the structure of their conceptual query (conjunctive, disjunctive, or compromise). Since the choice of this value is dependent on the database characteristics and requires expert knowledge of descriptors and the UFSC algorithm, the nonexpert user must be guided through the selection process if they are to choose the parameter properly. This can be done by choosing three extremes namely, 0, 0.5, 1 and presenting a subset of results of each case. The user can then further refine the value based on how well each result set matches his/her intentions.

The system may also possess the ability to automatically adjust the value of the compensation parameter. In the simplest case, if there is prior knowledge known regarding the database images or user queries to be expected in a system, the designer may wish to fix the compensation parameter to a predetermined value. The compensation parameter may also be adaptively set based on user needs during relevance feedback.

VI. EXPERIMENTS & RESULTS

To evaluate the effectiveness of the proposed framework and demonstrate its versatility for retrieval from art repositories, experiments were performed to compare the accuracy of our method to the weighted average based techniques. The experimental setup as well as results are given in Sections VI-A and VI-B, respectively. Furthermore, in Sections VI-C and VI-D, we report on the effects of system parameters such as the compensation parameter γ and the choice of membership functions on the retrieval results.

A. Experimental Setup

The effectiveness of the proposed scheme is evaluated with respect to accuracy and flexibility in modeling user queries using a database of 800 artistic images from the Corel collection. This collection contains paintings of various sizes belonging to several categories including art of antiquity, impressionist paintings, portraits, and decorative scenes.

During the experiments, the query is specified by providing an example image to the system containing the concepts or features of interest. The system is then expected to respond by retrieving and returning the set of images that are *similar* to this query. Since the relevance of each result is subjective in nature and highly dependent on the user, the objective evaluation of this result set is very difficult. In order to ensure a fair assessment of the retrieval accuracy, four categories of images were chosen from the database such that each contains a distinct semantic topic. The performance of the system is then evaluated with respect to a ground truth set determined by a human observer. To this end, a human subject was asked to browse through the entire database and select images similar to a given query. In this manner, a ground truth set was generated by the human for each query. Following this scheme, each retrieved result by the system is considered a positive match if it belongs to the ground truth set of the query. The selected categories are “Madonna”, “Female Portraits”, “Frescoes”, and “Impressionist Scenes” as listed in Table III together with the number of images contained in each ground truth class.

The criteria for performance evaluation is the *precision-recall* measure [45]. Precision is the ratio of relevant retrieved images to the total retrieved images and is an indication of the efficiency of the retrieval. Recall is the proportion of desired results retrieved within the first k results. Precision and recall are then defined as follows [45]:

$$R_k = \frac{RR_k}{N_i}$$

$$P_k = \frac{RR_k}{k} \quad (16)$$

where, R_k and P_k are recall and precision after k images are retrieved, respectively, RR_k is number of matches after k retrievals, and N_i is the total number of images in category i . The performance of the algorithm is evaluated by plotting precision versus recall as the number of retrieved results increases. The performance measure is then averaged over the four chosen queries (average precision-recall) to remove any bias that might have been introduced by choosing a single query.



Fig. 5. Retrieval results for “Madonna” using various features and descriptors. (a) Dominant color descriptor only, (b) all color descriptors, (c) color and texture (UFSC), (d) color and texture (weighted avg.).

TABLE III
THE QUERY CATEGORIES CHOSEN FOR PERFORMANCE TESTING

Category	Number of Images
“Madonna”	33
“Female portraits”	37
“Frescoes”	18
“Impressionist Scenes”	20

For brevity, the top 10 retrieval results are shown where the top left image is used as the query. In addition, we present the precision and recall after retrieving 10 and 20 image to provide a general idea with regard to the accuracy for each set of results.

Throughout the experiments of this section, the optimum values of the compensation parameters, γ_{Color} , $\gamma_{texture}$, γ_{All} , have been obtained through manual optimization using an exhaustive search. During this search, the result set for each value of γ was compared with human-selected ground truth set for the given. The compensation parameter that offered the best precision-recall was then chosen as the optimum value. The set of optimum values for the compensation parameter is shown in Table IV.

In Section VI-B the retrieval results using the weighted average and the UFSC are presented for each of these categories. In using the weighted average method, all features were assumed to be of equal importance as no weighting was used with the proposed method. It must be noted that optimization of the compensation parameter in the proposed scheme is not equivalent to optimization of feature weights in a weighted average-based approach. The compensation parameter only ad-

TABLE IV
OPTIMUM VALUES OF γ USED IN THE EXPERIMENTS

Category	Colour Features	Texture Features	Overall
Madonna	0.8	1	1
Female portraits	0.2	0	0
Frescoes	1	0.2	1
Impressionist Scenes	0.6	0.8	0.8

justs the similarity measurement operator and does not convey any information regarding the importance of different features to the user. It is possible to use a weighting scheme with the proposed method to indicate the importance of each feature descriptor in a manner similar to the weighted average. In this work, however, the use of weighting schemes were omitted from the experiments to demonstrate the sole effect of the aggregation operator on the results.

B. Retrieval Using UFSC

In this section, we report on the performance of the proposed scheme for retrieval of art images with respect to flexibility and accuracy. To demonstrate the flexibility of the proposed scheme, the retrievals are performed using one or more features and descriptors. To this end, the query category “Madonna” is randomly chosen for the following experiments. The retrieval is first performed based on a single descriptor of color, namely dominant color. This choice is made because a large majority of Madonna image within this collection are from the same artistic era and hence share a common color composition. The retrieval results for this query are shown in Fig. 5(a). In order to improve the retrieval accuracy, all descriptors of colors are used in the next query and results are shown in Fig. 5(b). Finally,

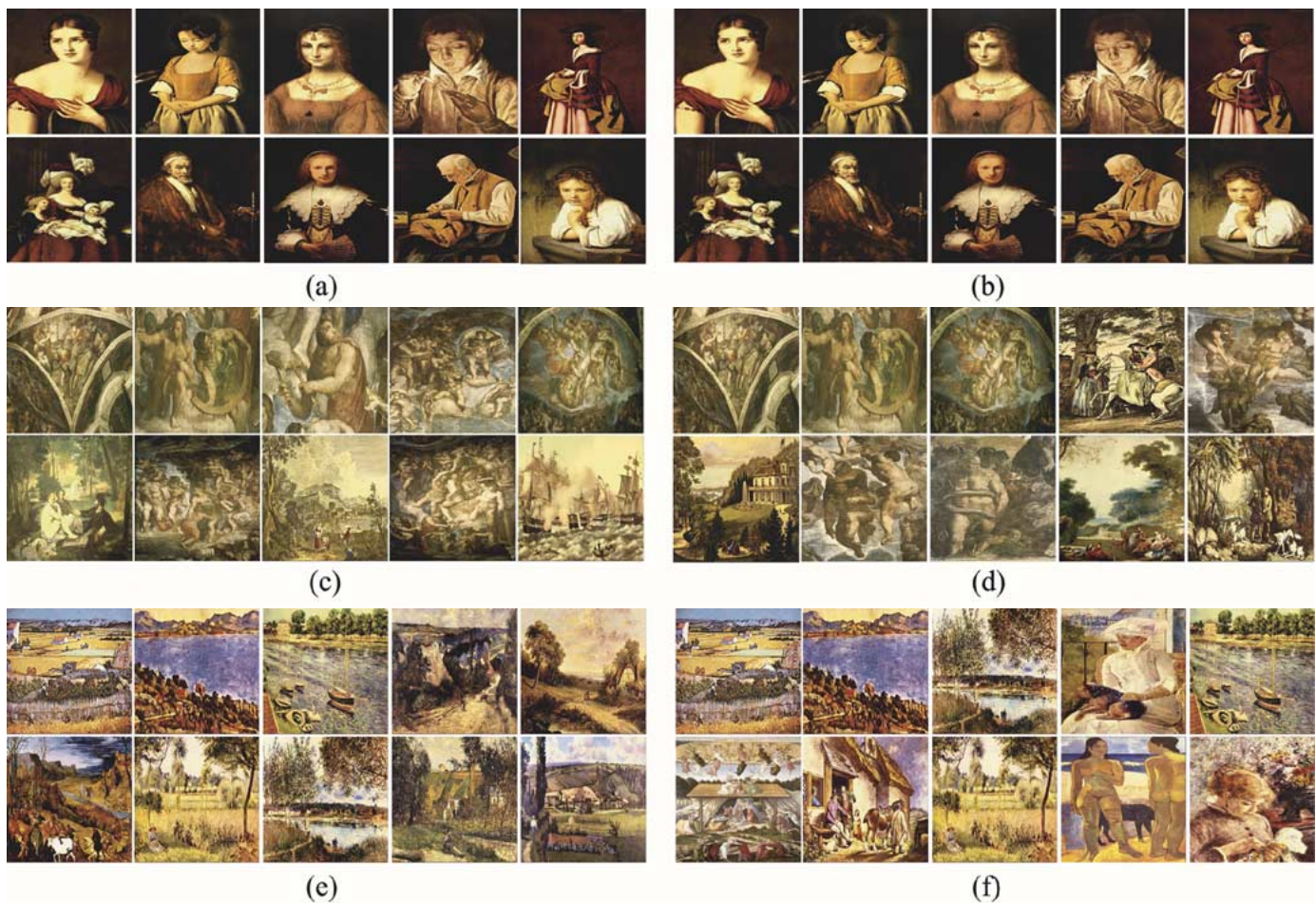


Fig. 6. Top 10 retrieval results for three chosen query categories using the UFSC and the Weighted Average (WA). (a) “Female Portraits” (UFSC), (b) “Female Portraits” (WA), (c) “Frescoes” (UFSC), (d) “Frescoes” (WA), (e) “Impres. Scenes (UFSC)”, (f) “Impres. Scenes (WA)”.

TABLE V
AVERAGE PRECISION AND AVERAGE RECALL AFTER 10 AND 20 RETRIEVALS FOR THE “MADONNA” QUERY

Query	$P(10)$	$P(20)$	$R(10)$	$R(20)$
dominant colour only	0.4	0.25	0.12	0.15
all colour descriptors	0.5	0.3	0.15	0.18
colour & texture (UFSC)	0.8	0.45	0.24	0.27
colour & texture (weighted avg)	0.4	0.35	0.12	0.21

the retrieval results using both color and texture for the proposed approach and the weighted average are shown in Fig. 5(c) and (d), while the precision and recall measures after retrieving 10 and 20 images are presented in Table V. In addition, the retrieval of results for each of the remaining three queries based on color and texture for both the weighted average and the proposed scheme are shown in Fig. 6 and precision and recall measures for each case is summarized in Table VI. From the results in Fig. 6(a) through (d), it can be said that the UFSC performs better than weighted average technique. This is due to the fact

TABLE VI
AVERAGE PRECISION AND AVERAGE RECALL AFTER 10 AND 20 RETRIEVALS FOR VARIOUS QUERIES

Query	P_{10}	P_{20}	R_{10}	R_{20}
“Madonna” (UFSC)	0.7	0.45	0.21	0.27
“Madonna” (weighted avg.)	0.4	0.35	0.12	0.21
“Frescoes” (UFSC)	0.7	0.6	0.39	0.66
“Frescoes” (weighted avg.)	0.6	0.4	0.33	0.44
“Female Portraits” (UFSC)	0.5	0.45	0.14	0.24
“Female Portraits” (weighted avg.)	0.4	0.25	0.11	0.14
“Impressionist Scenes” (UFSC)	0.8	0.5	0.4	0.5
“Impressionist Scenes” (weighted avg.)	0.5	0.35	0.25	0.35

that proposed framework offers direct control on the compensation parameter allowing the user to specify the logical structure of the query as desired. A more detailed description of the

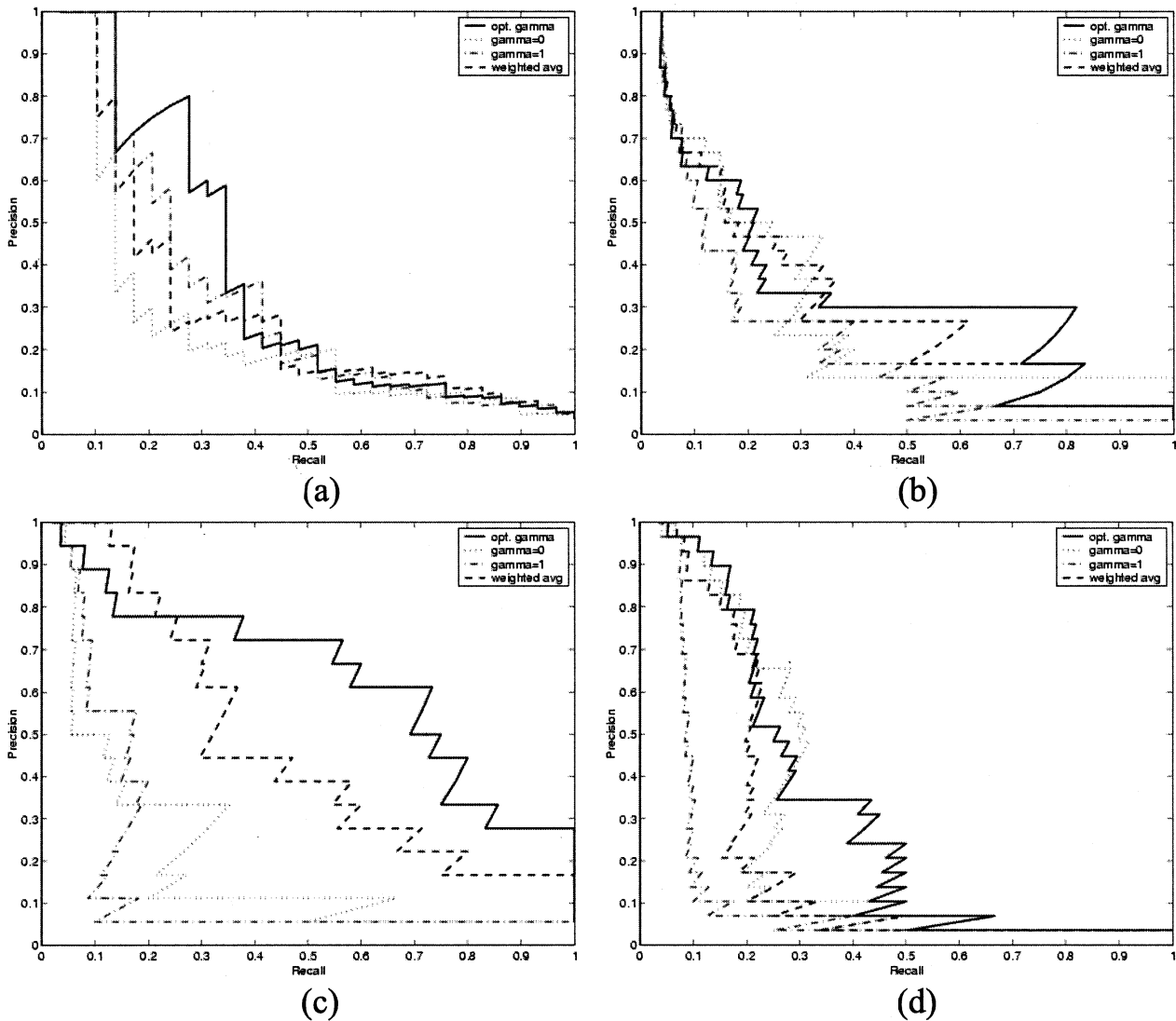


Fig. 7. Precision-recall graph for queries. (a) "Impres. Scenes," (b) "Madonna," (c) "Frescoes," and (d) "Female Portraits."

intended query by the user allows the CBIR system to achieve a higher retrieval accuracy.

C. Effect of γ on Retrieval

In order to examine the influence of the compensation parameter γ on the retrieval accuracy, the precision-recall graphs for various values of γ for the four queries are plotted and shown in Fig. 7. The first curve in each image demonstrates the retrieval results for the optimum value of γ which is determined experimentally. The other curves correspond to two chosen values of this parameter, namely 0 and 1. These values were chosen to demonstrate the sensitivity of UFSC to its compensation parameter if it was chosen with no prior knowledge of a user's intentions. For comparison purposes, the precision-recall graph for the weighted average is shown in each case as well. From these plots it can be said that the retrieval performance of UFSC is highly dependent on the compensation parameter γ as the logical structure of the user query is essentially determined by this parameter. This sensitivity demonstrates that the retrieval accu-

racy is greatly affected by the aggregation operator employed. We take advantage of this conclusion to improve the results by choosing the best operator for each given query. It must be emphasized here that even though an optimum value of γ exists, this value is query and user dependent and must be determined based on the user needs.

We claim that γ can be used to specify and model various conceptual queries. To that end, we present the results for the "Impressionist Scenes" query using conjunctive and disjunctive combination of features. The retrieval results for this case are shown in Fig. 8. This figure reiterates the fact that the aggregation operator used during similarity calculation has a great effect on the retrieval results as it determines the conceptual user query. Fig. 8(a) demonstrates the results for the query color OR texture. It is seen that the results correspond to images of farm-like landscape. This, however, is not the case for the query using the AND operator (Fig. 8(b)) as the results of this query correspond to natural scenery. Furthermore, the results using only two descriptor include mostly general impressionist paintings whereas the results of the query based on a compromise



Fig. 8. Retrieval results for the “Impres. Scenes” query using various aggregation operators. (a) Texture or color, (b) texture and color, (c) color and texture (compromise), (d) color and texture (dominant color and edge histogram).

between descriptors shown in Fig. 8(c) are impressionist natural scenery.

D. Effect of the Membership Function on Retrieval

The proposed aggregation scheme is tested with various membership functions in this section. To this end, the compensation parameter is fixed to a value of 0.5 for all operators and the following two membership functions are considered [31]:

$$\mu_1 = \exp \frac{-d(Q, I)}{\alpha} \quad (17)$$

$$\mu_2 = \frac{1}{1 + \frac{-d(Q, I)}{\beta}} \quad (18)$$

For the exponential function of (17), the parameter α was determined as in (7), as the *max*, *mean* and the *median*. The average precision-recall graph for each case is in Fig. 9. The graphs show that the method of (7) performs slightly better than the other methods. The insensitivity of the UFSC to the membership function may be due to the fact that in this work at most three descriptors are aggregated together since a hierarchical scheme is used. Therefore, the possibility that the different membership functions drastically change the min and max arguments in this case is low.

VII. DISCUSSION

The experiments of the previous section demonstrate that the aggregation framework presented in this paper is flexible in the sense that it offers the ability to model a wide variety of conceptual queries. This result is of significant importance for retrieval from artistic databases since even though describing the contents of a piece of art is inherently subjective, common visual characteristic of art work can be used as a measure of similarity among them. For example, each artist or artistic period is set apart from others by certain visual features such as color composition, or texture. By taking advantage of this fact and formu-

lating queries based on the desired combination of features, a higher retrieval accuracy may be achieved.

The proposed approach allows for modeling a wider range of conceptual queries than systems using Boolean expressions, Euclidean distance, or the weighted average. Boolean expression-based systems such as *MARS* can be thought of as specific realizations of the UFSC where the compensation parameter is restricted to only taking on binary values of 0 or 1. In this sense, the proposed framework provides a generalization of such systems.

Metrics such as the Euclidean distance are appropriate for use with low-dimensional vectors of fixed size. This imposes a severe limitation on the descriptors that can be used in a CBIR system. In addition, a Euclidean-distance based approach to similarity measurement dictates that the same distance is used with all descriptors. This distance, however, may not produce perceptually meaningful results for all descriptors. This leads to a degradation in the quality and accuracy of the retrievals. The UFSC framework, on the other hand, allows each descriptor to employ a distinct distance measure. In addition to the ability to incorporate various descriptor distance measures, the UFSC framework allows descriptors of different nature and size to be utilized in a system. The superiority of the proposed technique, thus, is seen in its flexibility to easily adapt to different design constraints.

In order to compare the weighted average method to the proposed aggregation scheme, we consider the *Andness* measure for each operator. Without loss of generality assume $\alpha_1 \leq \dots \leq \alpha_n$ are the descriptor distances. Then,

$$\begin{aligned} \text{andness}(WA) &= \frac{WA - \alpha_1}{\alpha_n - \alpha_1} \\ &= \frac{\frac{1}{n}(\alpha_1 + \dots + \alpha_n) - \alpha_1}{\alpha_n - \alpha_1} \end{aligned} \quad (19)$$

It can be seen from (19) that the *Andness* of the arithmetic average is a function of its arguments. In general, if the arguments are closer to the minimum term, the operator takes on

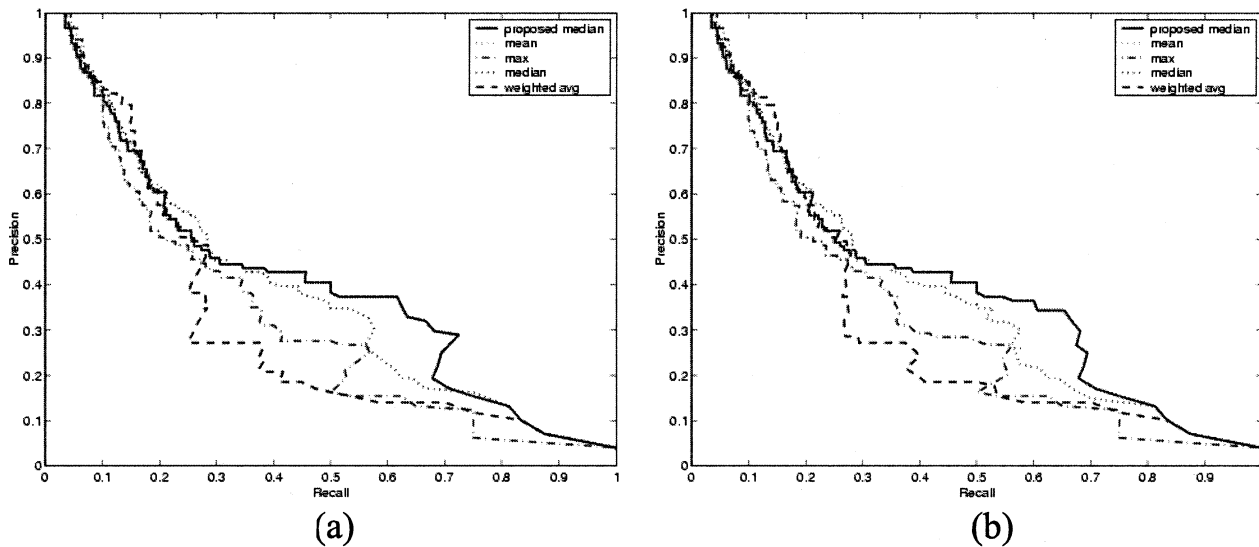


Fig. 9. Precision-recall graph using different normalization methods. (a) Using the exponential membership function, and (b) the reverse distance membership function.

a pessimistic view (conjunctive behavior). Similarly, if the average of the arguments is close to the maximum of the terms, the operator acts closer to a logical OR (disjunctive behavior). This automatic mood adaptivity is not always desirable since it results in a decreased ability to adjust the behavior of the system based on a user's needs. It is, therefore, strongly desirable to decouple the mood adjustment of the operator from its arguments. In contrast to the weighted average, the *Andness* of the compensative operator is simply γ , and does not depend on the arguments of the aggregator. This provides direct control over the conceptual operation of the operator.

Our system offers the functionality of a set of aggregation operator through the adjustment of a compensation parameter. This framework offers various options for the adjustment of this parameter. The user of the system can directly adjust the mode of the operator in a supervised manner. This is done based on the nature of the query and whether an optimistic, pessimistic, or compromise [42] behavior is required. Another option is for the system designer to set this parameter to a predetermined value if prior knowledge of system requirements, queries, and database employed is available. Finally, the operator can be selected adaptively based on user input using automatic techniques such as relevance feedback.

The fuzzy aggregation technique proposed here does not impose any requirements on the feature set. The features can be dynamically added or dropped as required. The computational complexity of it is linear at worst since it only requires that the min and max of the arguments to be calculated. Finally, we avoid the problem of a high-dimensional feature space and the cost and complexity associated with them since each feature distance is calculated separately and combined with other distances hierarchically.

VIII. CONCLUSIONS

The problem of similarity calculations between images is one of the key challenges in the area of content-based image retrieval as the common approaches to this problem suffer from several limitations. Techniques using the weighted average or the Eu-

clidean distance do not provide enough flexibility for modeling user queries since the logical functionality is dependent on their argument and cannot be directly controlled. In this paper, a similarity calculation framework is proposed based on decision combination. It was shown theoretically and experimentally that this approach provides an effective and flexible means for similarity calculation.

The proposed framework uses a *compensatory operator* for combining different feature similarities in order to provide the user with a set of tools for modeling various conceptual queries. Instead of using a fixed operator for feature combination, the operator is chosen from a family of aggregators by means of a compensation parameter and the functionality is directly controlled. In addition, the UFSC is well suited for use with a relevance feedback engine. As an example, the choice of the particular operator to be used as well as the importance of each feature can be decided according to feedback obtained from the user. Investigating iterative feedback-based adjustment of UFSC parameters for the purpose of improving retrieval results is one of the current research topics that we are pursuing.

By no means is the UFSC discussed in this paper is limited to what is presented here. The compensatory operator can also be used to combine local (region) similarities. Furthermore, the generality of this tool allows it to be applied to any feature or descriptor set.

APPENDIX

PROOF OF THEOREM 1

Let $X = [x_1, x_2, \dots, x_n]$ and $\odot(X) = x_1 \odot \dots \odot x_n$. Then,

$$\begin{aligned}
 \text{Andness}(\odot) &= \frac{\odot(X) - \min(X)}{\max(X) - \min(X)} \\
 &= \frac{\gamma \max(X) + (1 - \gamma) \min(X) - \min(X)}{\max(X) - \min(X)} \\
 &= \frac{\gamma \max(X) - \gamma \min(X)}{\max(X) - \min(X)} \\
 &= \gamma
 \end{aligned}$$

■

REFERENCES

- [1] Z. Chen and B. Zhu, "Some formal analysis of Rocchio's similarity-based relevance feedback algorithm," *Inform. Retrieval*, vol. 5, pp. 61–86, 2002.
- [2] Y. Rui, T. Huang, and S. Mehrotra, "Relevance feedback techniques in interactive content-based image retrieval," *Proc. SPIE*, vol. 3312, 1997.
- [3] J. M. Martinez, R. Koenen, and F. Pereira, "MPEG-7: the generic multimedia content description standard, part 1," *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 78–87, Apr./June 2002.
- [4] 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, Dec. 2000.
- [5] A. De Rosa, A. M. Bonacehi, V. Cappellini, and M. Barni, "Image segmentation and region filling for virtual restoration of artworks," in *Proc. Int. Conf. Image Processing*, vol. 1, 2001, pp. 562–565.
- [6] M. Barni, F. Bartolini, and V. Cappellini, "Image processing for virtual restoration of artworks," *IEEE Trans. Multimedia*, pp. 34–37, 2000.
- [7] 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, pp. 945–965, 2002.
- [8] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content—the QBIC system," *Computer*, vol. 27, no. 9, pp. 23–32, 1995.
- [9] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Trans. Multimedia*, vol. 2, no. 4, 2000.
- [10] S. Jeong, K. Kim, B. Chun, J. Lee, and Y. J. Bae, "An effective method for combining multiple features of image retrieval," in *Proc. IEEE Region 10 Conf.*, vol. 2, 1999, pp. 982–985.
- [11] I. Cox, M. Miller, T. Minka, T. Pappas, and P. Yianilos, "The Bayesian image retrieval system, PicHunter: theory, implementation, and psychological experiments," *IEEE Trans. Image Processing*, vol. 9, pp. 20–37, Jan. 2000.
- [12] T. Gevers and A. Smeulders, "PicToSeek: combining color and shape invariant features for image retrieval," *IEEE Trans. Image Processing*, vol. 9, pp. 102–119, Jan. 2000.
- [13] 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. on Image Analysis and Processing*, 2001.
- [14] S. Aksoy, R. Haralick, F. Cheikh, and M. Gabbouj, "A weighted distance approach to relevance feedback," in *Proc. 15th Int. Conf. on Pattern Recognition*, vol. 4, 2000, pp. 870–876.
- [15] J. Shih and L. Chen, "A context-based approach for color image retrieval," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 2, pp. 239–255, 2002.
- [16] B. Mehtre, M. Kulkarni, and W. Lee, "Content-based image retrieval using a composite color-shape approach," *Inform. Process. Manage.*, vol. 34, no. 1, pp. 109–120, 1998.
- [17] D. Lee and H. Kim, "A fast content-based indexing and retrieval technique by shape information in large image database," *J. Syst. Softw.*, vol. 56, pp. 165–182, 2001.
- [18] G. Guo, A. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Networks*, vol. 13, no. 4, 2002.
- [19] J. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 947–963, Sept. 2001.
- [20] S. Santini and R. Jain, "Similarity measures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 9, pp. 871–883, 1999.
- [21] K. Idrissi, J. Ricard, A. Anwander, and A. Baskurt, "An image retrieval system based on the local and global color descriptors," in *Proc. Lecture Notes in Computer Science*, vol. 2195, 2001, pp. 55–62.
- [22] H. Yoo, D. Jang, S. Jung, J. Park, and K. Song, "Visual information retrieval via content-based approach," *Pattern Recognit.*, vol. 35, pp. 749–769, 2002.
- [23] I. Fudos and L. Palios, "An efficient shape-based approach to image retrieval," *Pattern Recognit. Lett.*, vol. 23, pp. 731–741, 2002.
- [24] G. Gaguadakis and L. Rosin, "Incorporating shape into histograms for CBIR," *Pattern Recognit.*, vol. 35, pp. 81–91, 2002.
- [25] A. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233–1244, 1996.
- [26] D. Squire, W. Muller, H. Muller, and T. Pun, "Content-based query of image databases: inspirations from text retrieval," *Pattern Recognition Lett.*, vol. 21, pp. 1193–1198, 2000.
- [27] S. Newsam, B. Sumengen, and B. Manjunath, "Category-based image retrieval," in *Proc. Int. Conf. on Image Processing*, vol. 3, 2001, pp. 596–599.
- [28] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, and T. Huang, "Supporting ranked Boolean similarity queries in MARS," *IEEE Trans. Knowledge Data Eng.*, vol. 10, no. 6, pp. 909–925, 1998.
- [29] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "BlobWorld: image segmentation using expectation-maximization and its applications to image querying," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 8, pp. 1026–1038, 2002.
- [30] R. Fagin, "Combining fuzzy information from multiple systems," *J. Comput. Syst. Sci.*, vol. 58, no. 1, pp. 83–99, 1999.
- [31] K. N. Plataniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*: Springer-Verlag, 2000.
- [32] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*. Boca Raton, FL: CRC, 2000.
- [33] P. Androutsos, A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Fuzzy aggregation of palette colors for hybrid querying of fine art image databases," in *Proc. Int. Conf. on Digital Signal Processing*, vol. 1, 2002, pp. 115–118.
- [34] H. Geerts, J.-M. Geusebroek, A. W. M. Smeulders, and R. van den Boomgaard, "Color invariance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1338–1350, Dec. 2001.
- [35] B. S. Manjunath, J.-R. Jens-Rainer Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 703–715, June 2001.
- [36] "Multimedia Content Description Interface—Part 3: Visual," ISO/IEC 15938-3:2001, ver. 1, 2001.
- [37] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. New York: Wiley, 2002.
- [38] K. N. Plataniotis, D. Androutsos, and A. N. Venetsanopoulos, "Adaptive fuzzy systems for multichannel signal processing," *Proc. IEEE*, vol. 87, no. 9, pp. 1601–1622, 1999.
- [39] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.
- [40] A. Pradera, E. Trillas, and T. Calvo, "A general class of triangular norm-based aggregation operators: quasilinear T-S operators," *Int. J. Approx. Reas.*, vol. 30, pp. 57–72, 2002.
- [41] D. Dubois and H. Prade, "A review of fuzzy set aggregation connectives," *Inform. Sci.*, vol. 36, no. 1–2, pp. 85–121, 1985.
- [42] I. Bloch, "Information combination operators for data fusion: A comparative review with classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 1, pp. 52–67, 1996.
- [43] H. J. Zimmermann, *Fuzzy Sets, Decision Making and Expert System*. Boston, MA: Kluwer, 1976.
- [44] J. Marichal and P. Mathonet, "On comparison meaningfulness of aggregation functions," *J. Math. Psychol.*, vol. 45, pp. 213–223, 2000.
- [45] J. Smith, "Quantitative assessment of image retrieval effectiveness," *J. Amer. Soc. Inform. Sci. Technol.*, vol. 52, no. 11, pp. 969–979, 2001.

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 pursuing the Ph.D. degree.

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

Panagiotis Androutsos (S'04) received the B.A.Sc. and M.A.Sc. degrees from the University of Toronto, Toronto, ON, Canada, in 1997 and 1999, respectively, where he is currently pursuing 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 (Florida Tech), Melbourne, in 1992 and 1994, respectively.

He was with the Digital Signal and Image Processing Laboratory, Department of Electrical and Computer Engineering University of Toronto, Toronto, ON, Canada, from 1995 to 1997. From August 1997 to June 1999, he was an Assistant Professor with the School of Computer Science at Ryerson University, where he served as a Lecturer in 12 courses to 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 and Computer Engineering where he researches and teaches adaptive systems and multimedia signal processing. He is the Bell Canada Junior Chairholder in Multimedia, a Nortel Institute for Telecommunications Associate, and Adjunct Professor with the School of Computer Science at Ryerson University. He co-authored (with A. N. Venetsanopoulos) the book *Color Image Processing & Applications* (New York: Springer-Verlag, 2000). He is a contributor to six books, and he 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 13–16, 2001. He is the Technical Program Co-Chair of the Canadian Conference on Electrical and Computer Engineering, (CCECE 2004), May 2–5 2004 and the Chair-elect for the IEEE Canada-Toronto Section.

Anastasios N. Venetsanopoulos (F'88) received the B.S. degree in electrical and mechanical engineering from the National Technical University of Athens (NTU), Athens, 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 he was promoted to Assistant Professor in 1970, Associate Professor in 1973, and Professor in 1981. He 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 of Graduate Studies in the Department of Electrical and Computer Engineering and was Acting Chair during the spring term of 1998–1999. In 1999, a Chair in Multimedia was established in the Electrical and Computer Engineering Department, made possible by a donation of \$1.25M from Bell Canada, matched by an equal amount of University funds. He became Inaugural Chairholder in July 1999 and two additional Assistant Professor positions became available in the same area. 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 National Technical University 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. 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 IEEE boards, councils, and technical conference committees, 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 TRANSACTIONS 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 of IEEE, 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) and Greece. He was elected 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” of 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 the IEEE. In April 2001, he became a Fellow of the Canadian Academy of Engineering. In 2003, he was the recipient of the highest award of the Canadian IEEE, the MacNaughton Award, and, in 2003, he served as the Chair of the Council of Deans of Engineering of the Province of Ontario (CODE).