Chapter II

The Impact of Low-Level Features in Semantic-Based Image Retrieval

Konstantinos Konstantinidis, Democritus University of Thrace, Greece

Antonios Gasteratos, Democritus University of Thrace, Greece

Ioannis Andreadis, Democritus Unviversity of Thrace, Greece

Abstract

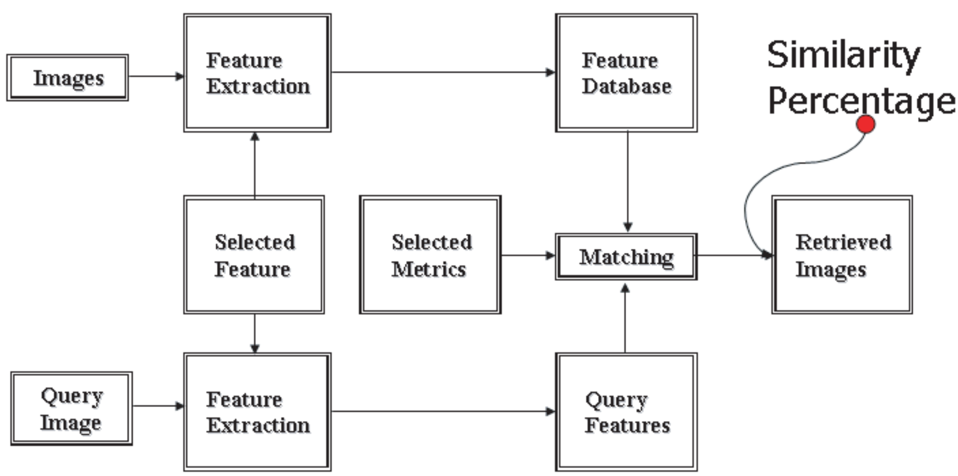
Image retrieval (IR) generally is known as a collection of techniques for retrieving images on the basis of features, either low-level (content-based IR) or high-level (semantic-based IR). Since semantic-based features rely on low-level ones, in this chapter the reader initially is familiarized with the most widely used low-level features. An efficient way to present these features is by means of a statistical tool that is capable of bearing concrete information, such as the histogram. For use in IR, the histograms extracted from the previously mentioned features need to be compared by means of a metric. The most popular methods and distances are, thus, apposed. Finally, a number of IR systems using histograms are presented in a thorough manner, and their experimental results are discussed. The steps in order to develop a custom IR system along with modern techniques in image feature extraction also are presented.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Introduction

Research in color imaging recently has emerged in a number of different applications, including military, industrial, and civilian, that generates gigabytes of color images per day. Moreover, recent improvements in information and communication technology have led to higher data transmission rates and, consequently, to a boom in networking. Therefore, more and more people have access to an increasing number of images. It is obvious that this will lead to a chaotic predicament, unless the enormous amount of available visual information is organized (Gagliardi & Schettini, 1997). Organization here means that appropriate indexing is available in order to allow efficient browsing, searching, and retrieving as in keyword searches of text databases. Associating a text to each image is one of the most popular and straightforward ways to index (Rowe, 2005). However, this means that prior to submitting each image into a database, a human agent must accompany it with a caption, thus leading to a lack of system automization. In many applications, such as in digital photography, area surveillance, remote sensing, and so forth, the images are labeled with automatically produced computerized names that are totally irrelevant to their semantic content. The best solution to such cases is the extraction and storage of meaningful features from each image for indexing purposes. In order to retrieve these images, a procedure known as query by example is performed; that is, the user has to present an image to the system, and the latter retrieves others alike by extracting features from the query image and comparing them to the ones stored in the database. The extraction of meaningful features, both content (Del Bimbo, 1999) and semantic (Zhang & Chen, 2003), is critical in IR and, therefore, an active field of research (Eakins, 2002; Smeulders, Worring, Santini, Gupta, & Jain, 2000). Nevertheless, while considering a semantic query (e.g., A Red Round Dirty Car), the descriptive components are based on low-level features; red on color, round on shape, and dirty on texture. Hence, in order for a semantic-based IR system to perform effectively, its lower features

Figure 1. Block diagram of the basic structure of a generic IR system



Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

must be extracted and indexed accordingly. The low-level features most commonly used by researchers are color (Castelli & Bergman, 2002, Del Bimbo, 1999, Gagliardi & Schettini, 1997; Konstantinidis & Andreadis, 2005; Konstantinidis, Gasteratos, & Andreadis, 2005; Liang, Zhai, & Chavel, 2002; Pass, Zabih, & Miller, 1996; Swain & Ballard, 1991), texture (Castelli & Bergman, 2002; Del Bimbo, 1999; Gasteratos, Zafeiridis, & Andreadis, 2004; Howarth & Rüger, 2004; Wang & Wiederhold, 2001), and shape (Brandt, Laaksonen, & Oja, 2002; El Badawy & Kamel, 2002; Jain & Vailaya, 1996). Color possesses a dominant function in vision, thus allowing the performance of complex tasks such as the discrimination between objects with similar shape characteristics but different color features, the tracking of moving objects as well as scene property analysis. Therefore, the use of color as a meaningful feature in IR is straightforward. Also, due to variations in viewing angles, texture patterns may vary in scale and orientation from image to image or even in different parts of the same image. Therefore, texture provides a useful cue for IR. Finally the role of shape is essential in IR, since there are examples in which two images with similar color and texture characteristics may exhibit highly dissimilar semantic content.

A useful tool in color image analysis is the histogram (Gonzalez & Woods, 1992)—a global statistical low-level descriptor that represents the color distribution for a given image. In color image processing methods, histograms usually enclose information of three components that correspond to the three components of the color space used. A histogram-based retrieval system (Figure 1) requires the following components: a suitable color space such as HSV, CIEBLAB or CIELUV; a feature representation such as classic, joint, or fuzzy histograms; and a similarity metric such as the Euclidean Distance, the Matusita distance, or the Histogram Intersection method. Similar to color, an image also can be indexed using a textural energy histogram. To this end, the image is processed by a method that may include convolution with Gabor filters (Zhang & Lu, 2000), wavelets (Bartolini, Ciaccia, & Patella, 1999), and Laws' masks (Laws, 1980).

IR using color histograms has both advantages and limitations (Konstantinidis & Andreadis, 2005):

- It is robust, since color histograms are rotation- and scale-invariant.
- Histograms are straightforward to implement.
- It is fast. The histogram computation has O(M2) complexity for an MxM image, while a histogram comparison has O(n), where n is the number of histogram bins, or quantization levels, of the colors used.
- It has low storage requirements. The color histogram size is much smaller than the size of the image itself.
- However, no spatial information of the color distribution is included, so the problem of two completely different images having similar histograms may arise.
- It is not immune to lighting variations.

As stated before, the efficiency of the semantic representation depends heavily on how compact and robust the indexing of the low-level features is. Therefore, in this chapter, we approach the role of color and texture in IR by constructing compact, easy-to-handle, meaningful histograms. We also present examples in which color, texture, and spatial information

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

are combined in generalized histograms for indexing. Methods (e.g., joint, fuzzy, etc.) and tradeoffs in histogram making also are discussed as well as the metrics most frequently encountered in a histogram-based IR system. Recent advances are presented through exemplar IR systems. These are based on the aforementioned methods and are comparatively apposed. At the end of the chapter, the user will have a sufficient background with which to compose a custom IR system.

Background and Related Work

Color Histograms

In order to define a color histogram, let I be an n x n image. Each pixel, p=(x, y), of the image may have one of m colors of the set {c1,c2,…,cm}; that is, $I(p) \in \{c_1, c_2, \ldots, c_m\}$. Let $I^c \overset{\Delta}{=} \{p \in nxn | I(p) = c\}$ be the set of pixels of image I, which are of color c.
Using this notation a histogram, HI(.), for an image I is given by:

$$H_I(i) \overset{\Delta}{=} n^2 \Pr[p \in I^{c_i}] \qquad (1)$$

In most applications of the color histogram, the term $\Pr[p \in I^{c_i}]$ is estimated as the fractional number of pixels with color ci.

The meaning of this is that, given a discrete color space, a color histogram is a statistic that counts the frequency with which each color appears in the image.

Color Histogram Extraction Techniques

In the past, various scientists have attempted to tackle the problem of real-time IR. In view of the fact that content-based IR has been an active area of research since the 1990s, quite a few IR systems, both commercial and research, have been developed. The most popular systems are query by image content (QBIC) and MIT's Photobook and its new version of FourEyes. Other popular systems are VisualSEEK, MetaSEEK and WebSEEK, Netra, Multimedia Analysis and Retrieval Systems (MARS), and Chabot.

In IR systems, image content is stored in visual features that can be divided into three classes according to the properties they describe: color, texture, and shape. Color and texture contain vital information, but nevertheless, shape-describing features are also essential in an efficient content-based IR system.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

In addition to the complex real-time systems mentioned previously, several researchers also tried to present simpler and, therefore, computationally wise and much lighter solutions to the problem.

For example, using the approach in Swain and Ballard (1991), a histogram is created from the opponent color space by subdividing the three derived components: rg, by, and wb into eight, eight, and four sections, respectively, resulting in a 256-bin histogram.

Furthermore, Liang et al. (2002) presented a system that uses crisp values produced by a Gaussian membership function in order to characterize a similarity fuzzy set centered at a given color vector in the RGB color space. Simple histograms are created using the previously mentioned Gaussian membership function.

Lu and Phillips (1998) proposed the use of perceptually weighted histograms (PWH). Instead of dividing each color channel by a constant (quantization step) when obtaining a histogram, they find representative colors in the CIELUV color space. The number of the representative colors equals the required number of bins.

Pass et al. (1996) created a joint histogram by selecting a set of local pixel features, subsequently constructing a multidimensional histogram. Each entry in this histogram contains the number of pixels that are described by a particular combination of feature values. Just as a color histogram approximates the density of pixel color, a joint histogram approximates the joint density of several pixel features. The features used were selected empirically and can be implemented time efficiently: color, edge density, textureness, gradient magnitude, and rank.

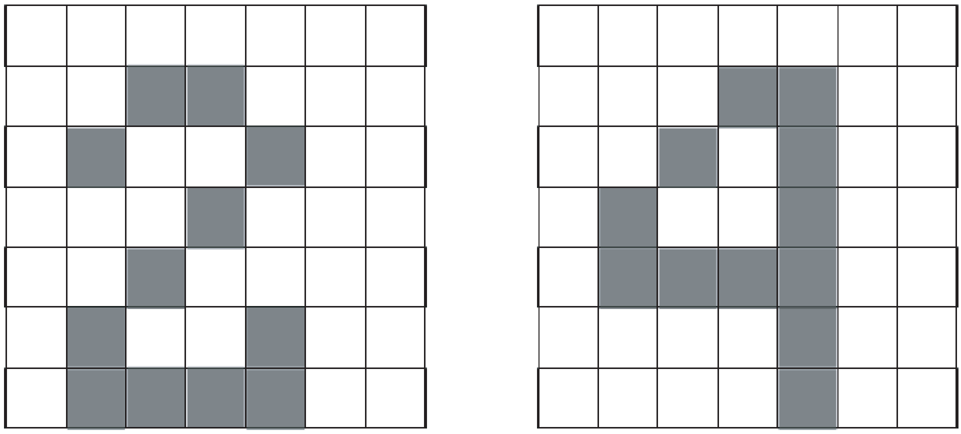
The idea of normalizing color images separately in each band as a reasonable approach to color constancy preprocessing in the context of indexing into an image database was adopted by Drew, Wei, and Li (1998). The information extracted from the images is transformed into a 2D representation by using histograms of chromaticity. Then, regarding the 2D feature space histograms as images, they apply a wavelet-based image reduction transformation for low-pass filtering, a square root operation, and discrete cosine transform (DCT) and truncation. The resulting indexing scheme uses only 36 or 72 values to index into the image database.

The weak spot of the indexing methods described earlier is the lack of spatial information in the histograms. For example, the two pictures shown in Figure 2 have identical histograms but different spatial distributions. Their semantic content is noticeably different, so evidently, one cannot assume that color distribution alone is always sufficient to represent the pictorial content of an image, since it comprises an abstract representation of it.

So, although the histograms become computationally heavier, in order to avoid false negatives, several methods have been introduced that enrich the color feature with geometric or spatial information. Stricker and Swain (1994) used boundary histograms to encode the lengths of the boundaries between different discrete colors in order to take into account geometric information in color image indexing. But this method may produce a large feature space (for a discrete color space of 256 elements, a boundary histogram of 32,768 bins) and is not robust enough to deal with textured color images. Gagliardi and Schettini (1997) investigated the use and integration of various color information descriptions and similarity measurements in order to enhance the system's effectiveness. In their method, both query and database images are described in the CIELAB color space with two limited palettes of perceptual importance of 256 and 13 colors, respectively. A histogram of the

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 2. Two images with different semantic content, but with identical histograms



finer color quantization and another of the boundary lengths between two discrete colors of the coarser quantization are used as indices of the image. While the former contains no spatial information (it describes only the color content in the image), the latter provides a brief description of the color spatial arrangement.

Pass and Zabih (1999) proposed another color histogram enhancement that also augments color histograms with spatial information. Color coherence vectors (CCVs) classify each pixel as either coherent or incoherent; based on whether it is part of a large color homogeneous region in the image or not. After the classification, the histogram is constructed (as in the original color histogram formulation), and the value of each bin is the number of coherent pixels. One shortcoming of this method is that it neglects the relationship of a connected component to its background. It also fails to capture the shape of the component. Thus, Zachary and Iyengar (1999) developed the so-called threshold edge vector (TEV) (an extension to the CCV method), which addresses these two issues, by storing an additional vector containing edge information.

Heidemann (2004) presented an approach to represent spatial color distributions using local principal component analysis (PCA). The representation is based on image windows, which are selected by two complementary data-driven attentive mechanisms: a symmetry-based saliency map and an edge/corner detector. The eigenvectors obtained from the local PCA of the selected windows form color patterns that capture both low and high spatial frequencies, so they are well-suited for shape as well as texture representation.

Overall, color is the most widely used feature in IR; sometimes enriched with spatial information or else with multiple feature vectors. In a number of applications however, even these enrichments are insufficient to provide the exact semantic content (e.g., the Red Round Dirty Car). Thus, a synergy of features becomes a necessity.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Texture Histogram Extraction Techniques

Liapis and Tziritas (2004) approached the IR problem based on a combination of texture and color features. Texture features are extracted using the discrete wavelet frame analysis, whereas histograms of the CIELAB chromaticity coordinates are used as color features.

A method for color texture classification using self-relative histogram ratio features was presented by Paschos and Petrou (2003). The method utilizes the 3-D xyY color histogram of a given image (xyY is derived from the CIEXYZ color space, where xy is chrominance and Y is luminance). The chrominance component (xy) turns out to contain sufficient information for the proposed method to adequately classify the set of 164 VisTex color textures. When any of the previously described histogram extraction phases comes to an end and when an adequate image descriptor has been produced for every image in the database, a way to compare the latter to the one from the query image is needed.

Histogram Comparison Methods (Metrics)

The histogram comparison methods presented in this chapter are the ones most frequently met in the literature for the purpose of IR. In the mathematical expressions presented next, H_Q is the query histogram, H_C is the histogram to be compared, and *(i)* is the number of bins.

The Bhattacharyya distance (Fukunaga, 1990) measures the statistical separability of spectral classes, giving an estimate of the probability of correct classification. This distance overpasses zero histogram entries. For highly structured histograms (i.e., those that are not populated uniformly), this can lead to the selection of matches in which there is a strong similarity between the structure of query and database histogram.

$$B(H_Q, H_C) = -\ln \sum_i \sqrt{H_Q(i) \times H_C(i)} \tag{2}$$

The divergence factor measures the compactness of the color distribution in the query histogram with respect to the histograms of the images in the database.

$$D(H_Q, H_C) = \sum_i \left[\left(H_Q(i) - H_C(i) \right) \ln \frac{H_Q(i)}{H_C(i)} \right] \tag{3}$$

The Euclidean distance is one of the oldest distances used for IR and can be expressed through equation (1).

$$L_2(H_Q, H_C) = \sqrt{\sum_i \left(H_Q(i) - H_C(i) \right)^2} \tag{4}$$

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

The Matusita distance (Fukunaga, 1990) is a separability measure that provides a reliable criterion apparently because, as a function of class separability, it behaves much more like the probability of correct classification. It is expressed as:

$$M(H_Q, H_C) = \sqrt{\sum_i \left(\sqrt{H_Q(i)} - \sqrt{H_C(i)}\right)^2} \tag{5}$$

Swain and Ballard (1991) introduced the histogram intersection method, which is robust in respect to changes in image resolution, histogram size, occlusion, depth, and viewing point. The similarity ratio that belongs to the interval [0, 1] is compared to a given threshold. It can be described by:

$$H(H_Q, H_C) = \frac{\sum_i \min\left(H_Q(i), H_C(i)\right)}{\min\left(\sum_i H_Q(i), \sum_i H_C(i)\right)} \tag{6}$$

A New Perspective To Image Retrieval

In this section, a novel perspective to the problem of IR is presented, and a number of solutions are proposed. In order to tackle the problem of IR, four exemplar systems are presented that deal with all the aspects of the matter except for shape, which is used in much more application-dependent systems. The features addressed are those of color, texture, and spatial information. The first example shows that the color spaces used can be addressed as fuzzy sets to construct a more efficient and compact image descriptor. Due to its robustness, it easily could be included as a basis for a semantic feature. In the second system, it is explained that although color is sufficient enough to perform efficient retrieval, there are times when dissimilar images have very similar color distributions, thus producing false positives. These can be avoided by embedding spatial information into the feature at hand. Textureness is introduced in the third system, which is an example of an application-specific IR system for aerial images. Nonetheless, it is a fair example of the synergy between color and texture. In order to avoid the dependency, the features of color and color-texture are interlaced, as presented in the fourth system.

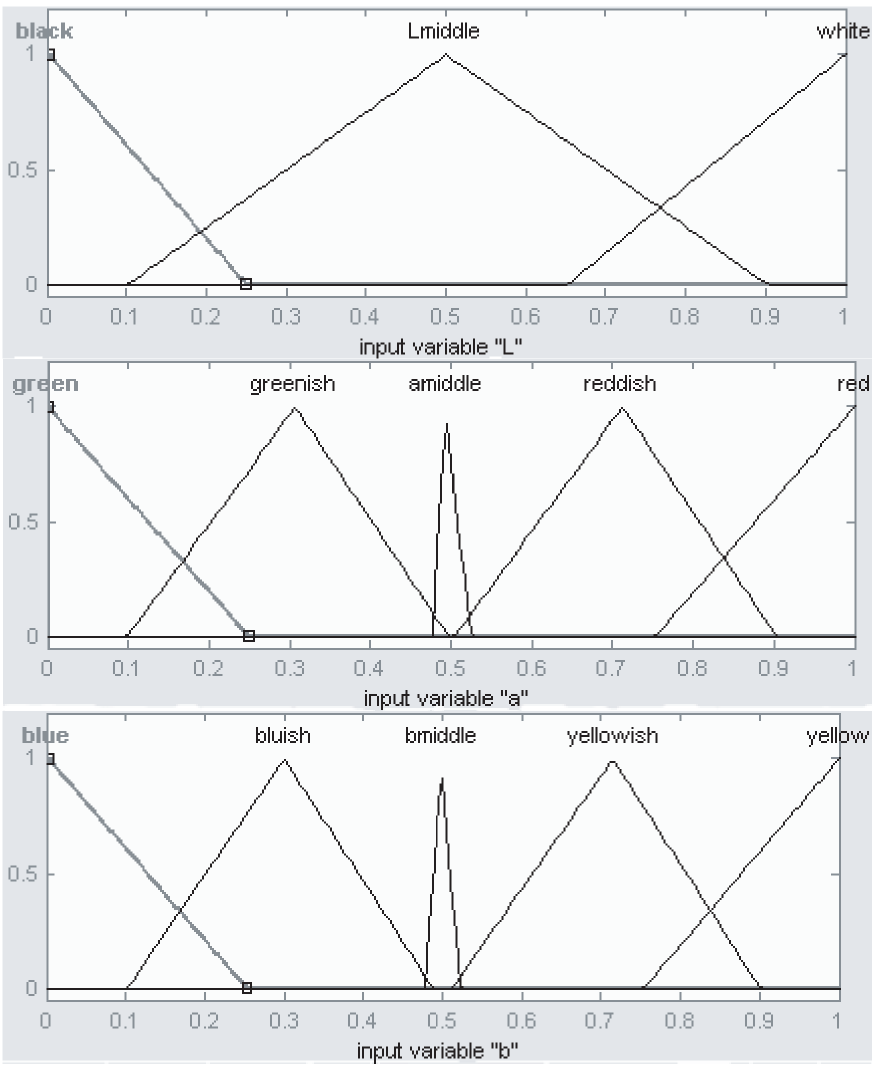
Fuzzy Histogram Image Retrieval

One might say that the largest image database in the world is that of the Worldwide Web as a whole, in which, since new sites are introduced daily, the images become truly countless. Thus, IR systems are needed that are fast and light but also robust and efficient. The most robust feature of all is that of color, and the easiest way to index it by also preserving memory is with histograms. However, a straightforward histogram IR system may be ac-

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

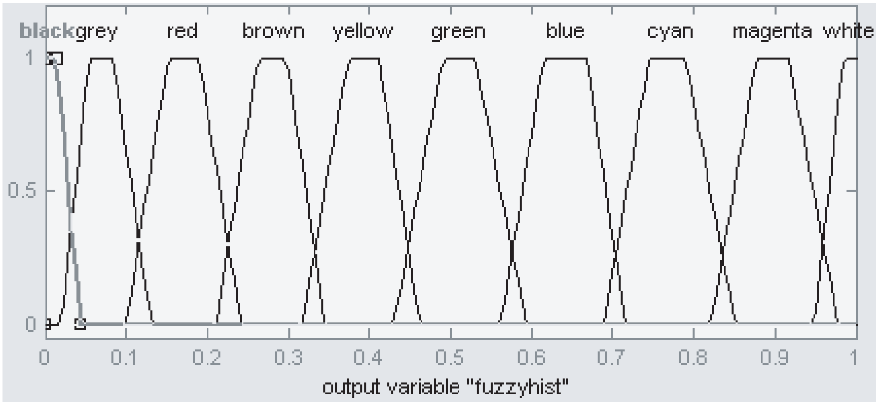
companied with a series of drawbacks; the number of bins required can be neither large, which presents the issue of perceptually similar colors belonging to adjacent bins and, thus, seemingly different, nor too small, since a few bins will result in a loss of resolution and, thus, of efficiency. In order to overcome this obstacle, a fuzzy histogram creation system is introduced that uses the CIELAB color space and produces a 10-bin histogram (Konstantinidis et al., 2005). The CIELAB color space is selected since it is a perceptually uniform color space, which approximates the way that humans perceive color. However, the main reason is that CIELAB was found to perform better than other color spaces in strenuous retrieval tests performed in the laboratory for this exact purpose. In CIELAB, the luminance

Figure 3. Membership functions of L, a*, and b**



Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 4. Membership functions of the output of the fuzzy system



is represented as L*, the relative greenness-redness as a*, and the relative blueness-yellowness as b*. The L* component does not contribute in providing any unique colors but for the shades of colors white, black, and grey. Thus, the L* component receives a lower weight with respect to the other two components of the triplet that represent the coloration. The tests performed on the regions of the CIELAB color space prove that in order for the IR system to work effectively the L*, a*, and b* components should be subdivided roughly into three, five, and five regions, respectively. The fuzzification of the input is accomplished by using triangular-shaped, built-in membership functions for the three input components, which represent the regions as shown in Figure 3. The reason for which the middle membership function even exists both in a* and b* is that in order to represent black, grey, and white, as seen in L*, then a* and b* must be very close to the middle of their regions.

The Mamdani type of fuzzy inference was used. The implication factor that determines the process of shaping the fuzzy set in the output membership functions based on the results of the input membership functions is set to min, and the OR and AND operators are set to max and min, respectively. The output of the system has 10 equally divided membership functions, as shown in Figure 4.

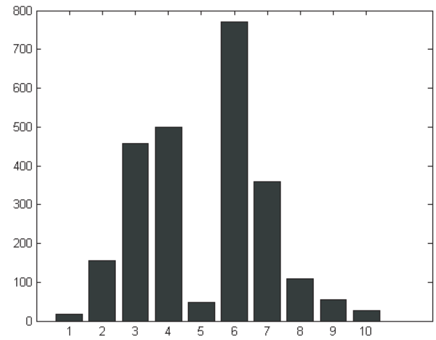
The defuzzification phase is best performed using the lom (largest of maximum) method along with 10 trapezoidal membership functions, thus producing 2,500 clustered bin values (all images were resized to a 50x50 pixel size to increase performance), which lead to the 10-bin final fuzzy histogram. The fuzzy linking of the three components is made according to 27 fuzzy rules, which leads to the output of the system. The rules were established through empirical conclusions that arose through thorough examination of the properties of a series of colors and images in the CIELAB color space.

In Figure 5, a query image (the well-known mandrill) and its respective fuzzy histogram are presented. The bins of the histogram shown in this Figure are in respect to (1) black, (2) dark grey, (3) red, (4) brown, (5) yellow, (6) green, (7) blue, (8) cyan, (9) magenta, and (10) white. The dominant colors in the image are easily noticed; bins 3, 4, 6, and 7 are activated mostly because of its fur and nose. The histogram in the proposed scheme,

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 5. (a) Query image 1 and (b) the respective 10-bin fuzzy linked histogram

(a)



(b)

though apparently rough, has proven to be an efficient tool for accurate IR, as presented in the experimental results section.

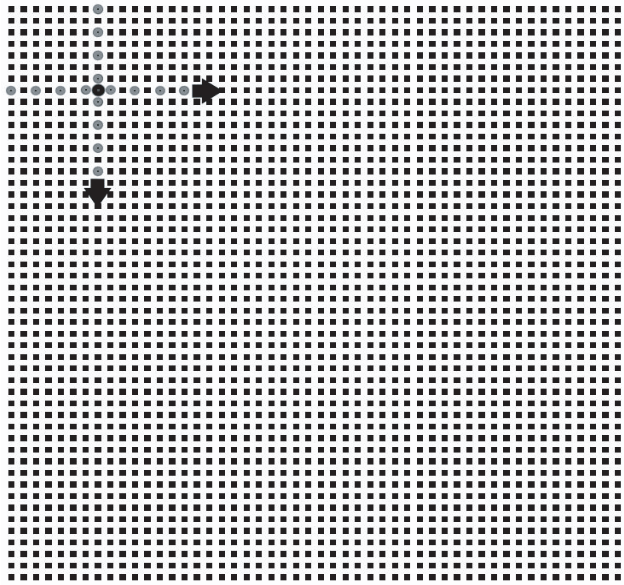
Once the histograms of the images in the database have been extracted, they are compared to the histogram derived from the query image by means of intersection (Swain & Ballard, 1991). This method was selected through a large number of tests performed in order to find the most suitable metric for this kind of histogram.

Spatially Biased Histogram Image Retrieval

Most color histogram creation methods like the one previously described contain global attributes; that is, the histogram at hand describes the overall statistics of the color in the images. Although such systems have proved to be efficient in most cases and are insensitive to rotation and scaling of the images, they also can present deficiencies in cases in which the images have similar colors but are spatially distributed differently. This problem can be overcome with the use of local histograms (i.e., splitting each image into smaller regions (Konstantinidis & Andreadis, 2005). These, on the other hand, suffer from a severe lack in speed due to the rapid increase in computational burden, which results from the repetitiveness needed to produce them. This led to the need of adopting global histograms with embedded local characteristics, such as a spatially biased histogram. The main gist of this method is to create a histogram by taking into consideration the color distribution in the image along with the concentrations of the dominant colors. The suggested histogram creation method has a straightforward algorithm, and only the hue component is enriched with spatial information so as to maintain the original histogram speed. The main reason the HSV color space was selected is because it reflects human vision quite accurately and because it mainly uses only one of its components (hue) to describe color in an image. The other two components (i.e., saturation and value) are significant only when describing black, white, gray, and the various shades of the colors. Thereby, in this method, hue is used mostly, being divided into eight regions, whereas saturation and value are divided into four

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 6. Graphical presentation of the presented method: A 50x50 pixel image I with the cross scanning the neighboring pixels of I (8,8). (The gray dots are the pixels enlarged for presentation purposes, and the black dot is the pixel at hand)

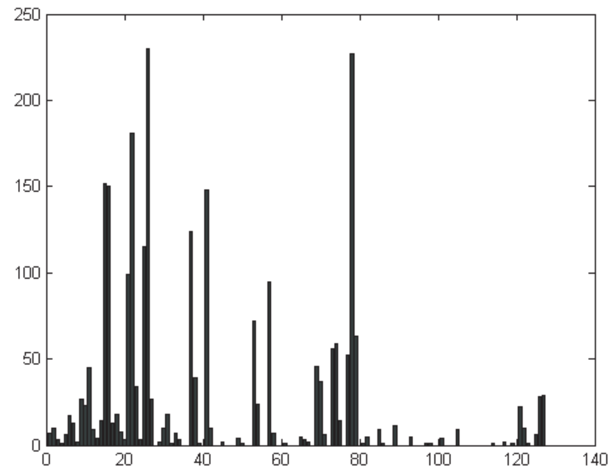


each. The three color components then are linked, thus creating a (8x4x4) histogram of 128 bins. During the reading of every pixel's value in the hue frame from left to right and top to bottom, the algorithm also searches in the manner of a cross having 15 pixels width and 15 pixels height around the current pixel for pixels with similar colors, as shown in Figure 6. If any pixel that is included in the vicinity of the full length of the cross possesses a color similar to the one of the central pixel, then instead of increasing the value of the specified bin of the particular color by one, it increases it by two, thus considerably enlarging the number of pixels in the bins that contain colors with significant concentrations in the image. The similarity of any pixel on the cross and the central pixel is measured by whether or not they belong to the same color region (bin).

Nonetheless, the pattern of the neighborhood for color vicinity may be replaced by any regular geometrical shape, such as a rectangle, a circle, or a rhombus, or even by an arbitrary geometrical shape. However, the cross was chosen due to its simplicity and also due to the gain in computational burden compared to using other shapes. The rest of the histogram, including saturation and value, is created in a straightforward manner so that even the slightest of information is included. The key idea of the method is that an 8x4x4 histogram is created, and then extra values are added to bins whose colors have significant concentrations anywhere in the image, thus solving the problem of noticeably different images having approximately the same histograms. The respective histogram of the mandrill for this method is depicted in Figure 7. One might notice the four distinct areas of peaks, which represent the four colors of the image that dominate both in volume and concentration. In this case, the peaks have slightly spread out due to the number of adjacent bins representing similar colors.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 7. The respective spatially biased histogram of the mandrill image



As a final instruction, when using this method, experimental results show that the size of the image need not be larger than 50x50 pixels in order to produce satisfying results, thus preserving disk space and speed, but increasing the size might produce a slight accuracy boost.

Having concluded the histogram extraction process and following numerous tests, the supposition was made that the similarity metric to be used in this method was that of the Bhattacharyya distance (equation (2)). As presented in the experimental results section in Table 2, one might notice the increase in accuracy compared to that of straightforward histogram creation methods.

Nevertheless, in some cases, as mentioned in the previous section, color alone is not sufficient to provide an efficient IR system, and it might seem more sensible to combine all the existing features into a single, possibly semantic, strong one. Yet this would lead to a nonrobust, computationally wise, heavy image descriptor, which most probably would be application-dependent. In order to overcome this predicament, the features need to be combined, but in a selective manner.

In the following sections, two systems are presented that are based on the synergy of color and texture.

Fuzzy Color-Texture Classification and Retrieval

The method presented is an example of semantic-based application dependency, as stated in the previous section. It produces a very efficient indexing and retrieval system, but it is confined in the range of classes with which it has been trained (i.e., deserts, clouds, sea, plantation). Overall, it is a fuzzy system in which the features of color and texture are combined via a

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

least mean square (LMS) technique. The texture features of the images are extracted using Laws' (1980) convolution method. However, instead of extracting a new image in which each of its pixels describes the local texture energy, a single descriptor is proposed for the whole image. Each class of scenes corresponds to a certain band in the descriptor space. The color similarity is examined by means of its characteristic colors (Scharcanski, Hovis, & Shen, 1994) in the RGB color space. The same feature set also can be used for image classification by its semantic content. The classification is performed by a fuzzy system. The membership functions of the proposed method are constructed by statistical analysis of the training features. As an example, a system that classifies aerial images is described, through which it can be noticed that the redundancy of texture information decreases the classification uncertainty of the system.

Specifically, the texture feature extraction of the presented system relies on Laws' (1980) texture measures in which where the notion of "local texture energy" is introduced. The idea is to convolve the image with 5x5 kernels and then to apply a nonlinear windowing operation to the convolved image. In this way, a new image results in which each pixel represents the local texture energy of the corresponding pixel of the original image. Laws (1980) proposed 25 individual zero-summing kernels, each describing a different aspect of the local texture energy. These kernels are generated by the one-dimensional kernels, as shown in Table 1. As an example of how the two-dimensional kernels are generated, L5S5 results by multiplying the one-dimensional kernel L5 with S5. Experiments with all the 25 kernels showed that, as far as our application is concerned, the most potent ones are R5R5, E5S5, L5S5, and E5L5. More specifically, by applying each of these four masks to images of a certain class (sea, forest, etc.), the global texture descriptors were more concentrated than was the rest of the masks. These kernels were used to extract the four texture descriptors of the proposed system.

The first texture descriptor of the image is extracted by convolving it with the first kernel (R5R5). The descriptor is the absolute average of the convolved image pixel values. Thus, instead of measuring local texture descriptors by averaging over local windows (typically 15x15), as proposed in Laws' (1980) original work, we keep one global texture descriptor by averaging the whole image. This descriptor is normalized by the maximum average found among a database of 150 training images. If, for a query image, the absolute average of the convolved image is greater than the maximum value, then the descriptor is 1.

Table 2. 1-D kernels; the mnemonics stand for level, edge, spot, wave and ripple, respectively

L5 =	[	1	4	6	4	1	]
E5 =	[	-1	-2	0	2	1	]
S5 =	[	-1	0	2	0	-1	]
W5=	[	-1	2	0	-2	1	]
R5 =	[	1	-4	6	-4	1	]

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

$$d_1 = \begin{cases} \dfrac{\left(\dfrac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| I(i,j) * \text{R5R5} \right| \right)}{d_{1\max}} & \text{f} \quad \left(\dfrac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| I(i,j) * \text{R5R5} \right| \right) \le d_{1\max} \\ 1 \quad \text{otherwise} \end{cases} \tag{7}$$

The same procedure is followed in order to extract the other three texture descriptors, d_2, d_3, and d_4, by replacing in equation (7) kernel R5R5 with the kernels E5S5, L5S5, and E5L5, respectively.

According to Scharcanski et al. (1994), in order to extract the characteristic colors of an image, the frequency of appearance of each color is assigned. Next, the colors are sorted in descending order according to their frequency of appearance. Given a color and a certain radius, a spherical volume is constructed in the RGB color space. The first color in the descending order comprises the first characteristic color of the image. Starting with the second color, it is examined whether it lies within the volume of any color above it. If so, then the examined color is merged with the color in the volume in which it lies. Otherwise, it comprises a new characteristic color of the image.

Considering the set of the characteristic colors as a vector, the color similarity of two images is computed by means of the angle between these two vectors. More specifically, the ratio of the inner product to the product of the measures of the two vectors corresponds to the cosine of the angle of these two vectors. The greater the value of the cosine, the smaller the angle and the more similar the two images (in terms of their color prospect). Therefore, the cosine could be used as the color descriptor of similarity. However, because the angle is the absolute descriptor and the cosine is a nonlinear function, the descriptor used in the proposed system is:

$$d5 = \frac{2}{\pi} \arccos \frac{\overline{C}_1 \cdot \overline{C}_2}{\left\| \overline{C}_1 \right\| \cdot \left\| \overline{C}_2 \right\|} \tag{8}$$

where and   are the set of the characteristic colors of images 1 and 2, respectively.

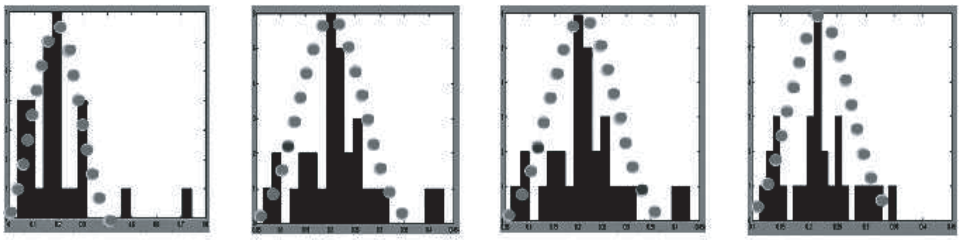
After extracting the descriptors both for the query and the database images, retrieval is performed by minimizing the following distance:

$$m = \frac{1}{\sum_{i=1}^{5} w_i} \sqrt{\sum_{i=1}^{4} w_i \left(din_i - ds_i \right)^2 + w_5 (d_5)^2} \tag{9}$$

where din_i (i=1,…4) are the four texture descriptors of the input image, resulting according to equation (7); ds_i is the corresponding texture descriptor of the sought image; d_5 is the color descriptor according to equation (8), and w_i is a weight tuning the retrieval process according to the importance of each descriptor. By comparing equations (7) and (8), it can be observed that although d_5 is a differential descriptor (i.e., it presents the difference of

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 8. The histogram of the global texture energy distribution for the training images belonging to the class of sea

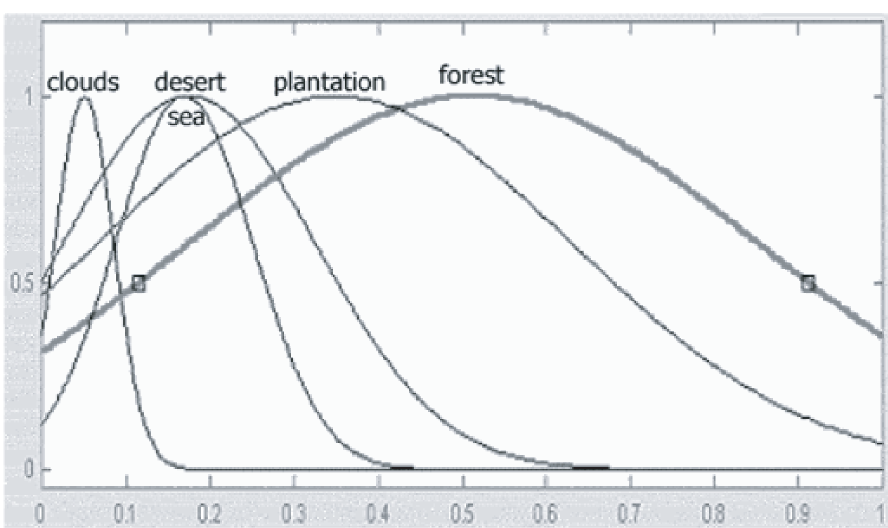


(a) (b) (c) (d)

two images by means of their color aspect) $d_1,\ldots d_4$ are absolute ones. This is the reason the difference of the latter appears in equation (9).

The same feature set described previously and used for IR may be used to classify images according to their texture and color properties by fusion of the descriptors previously presented. The system is tailored to meet the needs of the target application (i.e., the categorization of aerial images into five classes. However, with slight variations, it might be applied to other applications of image classification as well. The inputs of the fuzzy system are the five descriptors presented beforehand. In order to construct the membership functions for the inputs, a statistical analysis was carried out. Distinctively, five classes of photographs were used; namely, sea, clouds, desert, forests, and plantations. As training data, 100 images of each class were used, and for each image, the four texture descriptors were extracted. In Figure 8, the histograms of the distribution of the four descriptors for the class of the sea are presented.

Figure 9. The first input of the fuzzy system is the descriptor d1. The membership functions from left to right are: clouds, desert, sea, plantations, and forests



Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

As might be noticed, the distribution can be approximated by a trapezoidal or even a triangular membership function. However, a Gaussian function is also a good approximation, far better than the two latter ones. The reason is that its curve is not as steep as those of a triangular or a trapezoidal one, and therefore, it also includes the sided values. Experiments with several membership functions proved this intuition. For each descriptor and for each image class, the mean value and the standard deviation were calculated. The membership functions were computed as the normal distribution for the previous values (see Figure 8). The membership functions for the d_1 descriptor are depicted in Figure 9 as an example of the four first inputs of the fuzzy system.

For the color descriptor, five inputs were used. The characteristic colors of the 100 training images of each class were merged in the same way as already described for a single image. The result is a color codebook containing the characteristic colors of the whole image class. Equation (8) is used to compute the similarity between the characteristic colors of the input image and the codebook of each of the classes. The result of each of the color similarity values is used as an input to the fuzzy system (inputs from five to 10). Similarly, five sigmoid membership function outputs (one for each class) were used. In order to evaluate the performance of both the retrieval and the classification systems, several tests were carried out. The first experiments were run in order to acquire the weights of equation (9) that give optimum results. Each time, six more relevant images were asked to be retrieved. The precision (i.e., the ratio of the correctly retrieved images) over the total retrieved images was used to measure the efficiency of the retrieval system. It was observed that the best results occurred when $w_1=w_2=w_3=w_4=1$ and $w_5=11$. In particular, the retrieval precision was measured in the range of 35% to 100%, while no other combination of weights ever had resulted to 100% precision. This is to say that the color information plays a dominant role in the IR process, as the ratio of color to texture coefficients in the optimization of equation (9) is 11/4. Therefore, when the classification system is used in cascade to the retrieval system, it elevates the precision of the latter, as it rejects the majority of the scenes that do not belong to the same class as the retrieved one.

Color-Texture Histograms

The second color-texture-based system, unlike the previous one, is robust enough for usage in a nonclassified image database. It is a histogram creation method that uses both the CIELAB and HSV color spaces. The color-texture feature, which is extracted from the a* and b* components in synergy with the Hue component from the HSV color space, produce a single 256-bin histogram as the image descriptor.

In the first part of the method, all the images are transformed into the CIELAB color space, and only the a* and b* components are reserved. The same is performed for the HSV color space from which only the Hue component is stored. The color-texture feature is extracted from the image by means of convolution with Laws' (1980) energy masks similar to the previously presented system. The a* and b* components are convolved with the 5x5 masks, hence producing the respective ENa* and ENb* components, which represent the chromaticity texture energy of the image. The reason of having 25x2 color-texture components for

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

each image is that only the components with the greatest energy are reserved. The energy volume for each convolved image is computed by equation (10). Experimental results with all the 25 kernels show that the mask that most commonly results in the greatest energy is the L5L5.

$$E_{Vol} = \sum \left(\sum_{x=1,y=1}^{m,n} ENa^*(x,y),\ \sum_{x=1,y=1}^{m,n} ENb^*(x,y) \right) \tag{10}$$

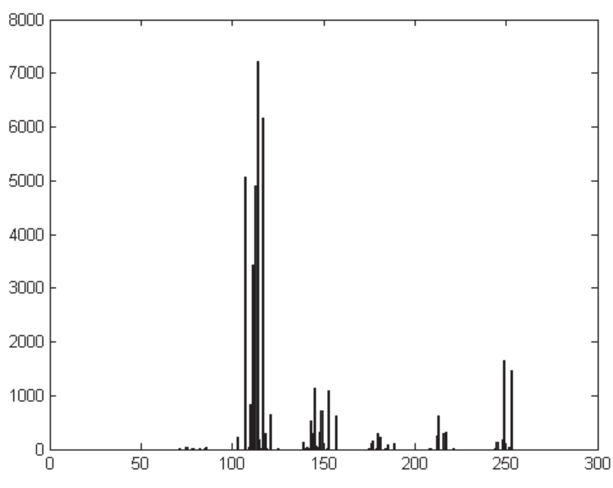
When the components with the greatest energy are found, the discretization process is activated during which all three components (ENa*, ENb*, and Hue) are divided into sections. The color-texture components are split up into eight parts each and the hue component into four. This way, they all are interlinked to each other, thus creating a histogram of (8*8*4) 256 bins. The respective histogram for the mandrill image is presented in Figure 10, where, in some cases, the dominant color peaks have been strengthened by the color-texture insertion, and in others, they have been weakened, thus producing a very strong peak and four other weak ones. The strong peak at around 110 represents the color textureness of the image, which is very strong for all colors.

Experiments with different metrics showed that the one performing best in conjunction with the histogram at hand was the Bhattacharyya distance. Comparative results are presented in the experimental results section.

Experimental Results

The systems introduced have been implemented and tested on a database of 1,100 images for comparison purposes. The images in the collection are representative of the general

Figure 10. The respective color-texture histogram for the mandrill image



Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Table 3. Experimental results

Image Set	Classic CIELAB Histogram	Swain & Ballard	Fuzzy Histogram	Spatially Biased Histogram	Color-Texture Histogram
1	65%	75%	90%	100%	100%
2	75%	70%	80%	80%	85%
3	90%	90%	95%	100%	100%
Overall Time(sec)	226.95	511.80	525.54	364.36	703.25

requirements of an IR system over the Internet. Therefore, the range of topics present in the image database is quite wide and varies from several different landscapes, flowers, sunsets, and everyday life images to sports, concerts, and computer graphics, which usually confuse IR systems. (The images are online, available at the following URL: http://utopia.duth.gr/~konkonst). The presented systems' effectiveness and efficiency, their precision, and time cost performances for three representative image sets are displayed next to Swain and Ballard's (1991) and the global histogram extraction method in the CIELAB color space in Table 2 with the intension of giving the reader an idea of the increase in accuracy together with the respective increase in time cost (i.e., computational burden).

Future Trends and Conclusion

Past and recent trends in content-based IR with the use of histograms were reported here as well as new heuristic methods to create histograms. One of the reasons just the low-level image descriptor of histograms was taken into consideration is that it constitutes the cornerstone of any semantic-based IR system. Moreover, the field of IR has become so broad that in order to cover all methods and techniques for all the features and their combinations using all kinds of descriptors would produce a very lengthy survey. However, the main reason for choosing color and texture histograms is that they are very easy to produce, compute, and store, thus making them an attractive feature descriptor to whomever is interested in constructing a custom IR system. To this end, it could be said that having read this chapter, the reader will have acquired sufficient knowledge to select an adequate color space, the

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

tools with which to create a robust low-level image descriptor, and the know-how of which comparison method to use in order to create a custom IR system.

In conclusion, the reader has been presented with four systems that cover all the aspects of a low-level feature-based IR system. At first, by presenting fuzzy histograms, the reader is walked through a color space discretization phase, which is one of the most important steps in IR. Moreover, it is shown that a histogram need not have a large number of bins in order to be effective. In addition, spatially biased histograms are presented, introducing the idea of embedding supplementary information into the straightforward histogram and, therefore, making the system even more efficient. Then, the importance of combining the features of color and texture is stressed out and discussed by means of an application-dependent system that makes use of color and texture as a whole, as well as a robust system that combines only selected attributes from the latter features.

With reference to the methods apposed, it is fair to say that they perform satisfactorily in terms of speed and accuracy.

Finally, general instructions to produce an IR system are as follows:

- The histogram should consist of a number of bins as small as possible so that the computational burden will be as light as possible, while at the same time preserving all the significant information.
- The color-space selection depends very much on the application, but uniform and human color perception spaces result in enhanced retrievals.
- The metric selection also depends on the application and only should be selected after running exhaustive tests by use of the extracted feature.
- The textural information somewhat improves the performance of the IR system but also significantly increases its computational burden; therefore, the method applied should be made as simple as possible. This can be solved by a weighted combination of the features.

Future work should be based either on implementing some new way of representing the features (as histograms once were) or on producing an optimal combination of the low-level features at hand (e.g., color, texture) in order to keep the systems as robust as possible; it is noted that specific applications (e.g., face recognition) do not belong in this area.

Additionally, as computer speed and memory, including disk space, become greater and cheaper, evaluations and experiments on databases of a much larger scale become feasible.

Acknowledgments

Mr. Konstantinidis is funded partially by: (1) the Greek-Slovenian Bilateral Project "Development of New Techniques for Recognition and Categorization" and (2) the Archimedes-II

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Project of the EU and the Greek Ministry of Education EPEAEK (code 2.2.3.z, subprogram 10 in the ATEI of Thessaloniki).

References

Bartolini, I., Ciaccia, P., & Patella, M. (1999). WINDSURF: A region-based image retrieval system. Proceedings of the *10th International Workshop on Database & Expert Systems Applications,* 167–173.

Brandt, S., Laaksonen, J., & Oja, E. (2002). Statistical shape features for content-based image retrieval. *Journal of Mathematical Imaging and Vision, 17*(2), 187–198.

Castelli, V., & Bergman, L.D. (2002). *Image databases: Search and retrieval of digital imagery*. John Wiley & Sons, Inc.

Del Bimbo, A. (1999). Visual information retrieval. San Francisco, CA: Morgan Kaufman Publishing.

Drew, M.S., Wei, J., & Li, Z.N. (1998). Illumination-invariant color object recognition via compressed chromaticity histograms of color-channel-normalized images. Proceedings of the *IEEE International Conference on Computer Vision (ICCV'98)*, 533–540.

Eakins, J.P. (2002). Towards intelligent image retrieval. *Pattern Recognition, 35*, 3–14.

El Badawy, O., & Kamel, M. (2002). Shape-based image retrieval applied to trademark images. *International Journal of Image and Graphics, 2*(3), 375–393.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). Academic Press.

Gagliardi, I., & Schettini, R. (1997) A method for the automatic indexing of color images for effective image retrieval. *The New Review of Hypermedia and Multimedia, 3*, 201–224.

Gasteratos, A., Zafeiridis, P., & Andreadis, I. (2004). An intelligent system for aerial image retrieval and classification. In G.A. Vouros, & T. Panayiotopoulos (Eds.), *Lecture notes in artificial intelligence* (vol. 3025) (pp. 63–71). Berlin: Springer-Verlag.

Gonzalez, R.C., & Woods, R.E. (1992). *Digital image processing*. Reading, MA. Addison-Wesley.

Heidemann, G. (2004). Combining spatial and color information for content based image retrieval. *Computer Vision and Image Understanding, 94*, 234–270.

Howarth, P., & Rüger, S. (2004). Evaluation of texture features for content-based image retrieval. In P. Enser et al. (Eds), *Lecture notes in computer science* (pp. 326–334). Berlin: Springer-Verlag.

Jain, A.K., & Vailaya, A. (1996). Image retrieval using color and shape. *Pattern Recognition, 29*, 1233–1244.

Konstantinidis, K., & Andreadis, I. (2005). Performance and computational burden of histogram based color image retrieval techniques. *Journal of Computational Methods in Sciences and Engineering (JCMSE), 5*, 1–7.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Konstantinidis, K., Gasteratos, A., & Andreadis, I. (2005). Image retrieval based on fuzzy color histogram processing. *Optics Communications, 248*, 375–386.

Laws, K. (1980a). *Textured image segmentation* [doctoral dissertation]. Los Angeles, CA: University of Southern California.

Laws, K. (1980b). Rapid texture identification. *SPIE, Image Processing for Missile Guidance, 238*, 376–380.

Liang, Y., Zhai, H., & Chavel, P. (2002). Fuzzy color image retrieval. *Optics Communications, 212*, 247–250.

Liapis, S., & Tziritas, G. (2004). Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Transactions on Multimedia, 6*(5), 676–686.

Lu, G., & Phillips, J. (1998). Using perceptually weighted histograms for color-based image retrieval. Proceedings of the *4th International Conference on Signal Processing, China, 1150–1153*.

Muller, H., Muller, W., Squire, D. McG., & Marchand-Maillet, S. (2001). Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters, 22*(5), 593–601.

Paschos, G., & Petrou, M. (2003). Histogram ratio features for color texture classification. *Pattern Recognition Letters, 24*, 309–314.

Pass, G., & Zabih, R. (1999). Comparing images using joint histograms. *Multimedia Systems, 7*(3), 234–240.

Pass, G., Zabih, R., & Miller, J. (1996). Comparing images using color coherence vectors. Proceedings of the *4th ACM Multimedia Conference*, 65–73.

Rowe, N.C. (2005). Exploiting captions for Web data mining. In A. Scime (Ed.), *Web mining: applications and techniques*. Hershey, PA: Idea Group Publishing.

Scharcanski, J., Hovis, J.K., & Shen, H.C. (1994). Representing the color aspect of texture images. *Pattern Recognition Letters, 15*, 191–197.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval: The end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.

Stricker, M., & Swain, M. (1994). The capacity of color histogram indexing. Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, 704–708.

Swain, M.J., & Ballard, D.H. (1991). Color indexing. *International Journal of Computer Vision, 7*, 11–32.

Wang, J.Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on PAMI, 23*, 947–963.

Zachary Jr., J.M., & Iyengar, S.S. (1999). Content-based image retrieval systems. Proceedings of the *IEEE Symposium on Application-Specific Systems and Software Engineering and Technology*, 136–143.

Zhang, C., & Chen, T. (2003). From low level features to high level semantics. In O. Marques & B. Furht (Eds.), *Handbook of video database design and applications* (pp. 613–624). CRC Press.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Zhang, D.S., & Lu, G. (2000). Content-based image retrieval using gabor texture features. Proceedings of the *First IEEE Pacific-Rim Conference on Multimedia*, Sydney, Australia, 392–395.

Copyright © 2007, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.