Content Based Image Retrieval and Information Theory: A General Approach

John Zachary and S. S. Iyengar
Louisiana State University, Department of Computer Science, Baton Rouge, LA

Jacob Barhen
Oak Ridge National Laboratory, Center for Engineering Science Advanced Research, Oak Ridge, TN

A fundamental aspect of content-based image retrieval (CBIR) is the extraction and the representation of a visual feature that is an effective discriminant between pairs of images. Among the many visual features that have been studied, the distribution of color pixels in an image is the most common visual feature studied. The standard representation of color for content-based indexing in image databases is the color histogram. Vector-based distance functions are used to compute the similarity between two images as the distance between points in the color histogram space. This paper proposes an alternative real valued representation of color based on the information theoretic concept of entropy. A theoretical presentation of image entropy is accompanied by a practical description of the merits and limitations of image entropy compared to color histograms. Specifically, the L_1 norm for color histograms is shown to provide an upper bound on the difference between image entropy values. Our initial results suggest that image entropy is a promising approach to image description and representation.

1. Introduction

Digital images are an increasingly important class of data, especially as computers become more usable with greater memory and communication capacities. As the demand for digital images increases, the need to store and retrieve images in an intuitive and efficient manner arises. Hence, the field of *content-based image retrieval* (CBIR) focuses on intuitive and efficient methods for retrieving images from databases based solely on the content contained in the images.

The corpus of CBIR research has focused on the definition of new visual feature representations for images that provide a meaningful discriminant for conducting similarity queries (Carson et al., 1997; Flickner et al., 1995; Gray, 1995; Jacobs et al., 1995; Pass et al., 1996; Pentland et al., 1996; Smith, 1997; Stricker, 1994; Zachary & Iyengar, 1999). Most current representations of visual features are based on vector forms. This paper expands existing visual feature representations based on vector spaces to improve retrieval performance and efficiency.

The representation of visual features can generally classified into several levels. As depicted in Figure 1, a visual object has several different levels of representation based on the complexity of the representation and the level of information aggregation. Our example assumes three levels of representations: the image level, which is the most general and complex; the vector level, which aggregates information into a vector representation; and the number level, which represents the highest level of aggregation and the least amount of complexity. As one aggregates information from the image level to the number level, the information contained in a given representation levels is not guaranteed to be unique to the representation at higher levels. Thus, it is with care that one must depend on the information content of a given level. Optimally, an automated method will incorporate multiple representation levels into computing a value or decision. This paper will present such a method.

Digital images are used throughout science, engineering, business, and personal computing. There are several reasons for the proliferation of images throughout general computer usage. The demilitarization of imaging and satellite technology has made it possible to capture data in high-resolution formats and from almost any region of the world. The emergence and explosive use of the World Wide Web (WWW) as a global network allows people to gather and share images en masse. Indeed, some estimates have conservatively put the number of images available on the WWW at between 10 and 30 million (Sclaroff et al., 1997). The impetus to merge television, entertainment, and computing technology into a cohesive platform is a forcing function for the miniaturization, low-cost fabrication, and increased capacity of memory and secondary storage devices.

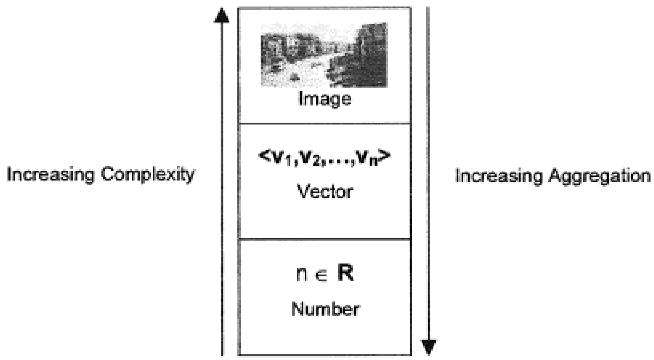

FIG. 1. Representation levels of visual objects.

Information theory has been an important application in image compression and coding. The initial work of (Shannon, 1948) formulated the foundation of information theory in terms of the information capacity of communication channels. The fundamental notion of *information entropy* describes a theoretical lower bound on the number of bits necessary to encode information. This concept has been useful in the development of image compression algorithms (Gonzalez & Woods, 1992; Pianykh, 1998). The use of information theory concepts to developed methods in image interpretation has received little attention. In Jagersland (1995), the entropy of an image was used to derive a description of scale in an image. The effort focused on the fact that in an image, the information content of a scene is typically confined to a small range of scales. In this paper, we describe an approach to the application of the information entropy of an image in computing a similarity value between pairs of images. We present experimental results of our approach given general unconstrained digital imagery.

2. Background

As the popularity of digital images grows, so does the need to organize, store, and retrieve images from collections or databases. The professional usage of digital image collections spans several fields.

- The health care field is increasingly adopting the digital storage of imaging technology (e.g., CT scan and MRI) over hard-copy film. As such, the need to retrieve information based on content plays a critical role in automated diagnostic and on-line educational systems.
- Law enforcement use digital imagery to store facial and fingerprint information on victims and suspects as well as historical records of crime scenes. The ability to retrieve "mug shots" from image databases based on the similarity analysis of content enables law enforcement to capture criminals in a more timely manner.
- Remote sensing technology from orbiting satellites and airplanes is used to monitor environmental conditions such as the erosion of coastal land. In addition, multi-spectral and hyper-spectral imaging modalities are used to monitor many environmental phenomena. Since many environmental processes are irreversible, CBIR may be employed to provide faster monitoring of the environment.

A fundamental distinction between textual and visual information is the nature of the retrieval process. The retrieval of textual information is based on discovering *semantic* and/or *syntactic similarity* between textual entities. Visual information retrieval, on the other hand, is concerned with discovering *perceptual similarity*. The concept of perceptual similarity is made clear by examining the kinds of queries that users are likely to expect to use when retrieving images from an image database. Although there is a marked lack of research on understanding the needs of users of CBIR systems, an analysis of the image features and attributes that can be used to construct effective CBIR queries might include the following items:

- the presence or absence of a particular color, texture, shape, or combination of these features (e.g., 30% Aubergine and Sunburst pixels)
- the presence, absence, or arrangement of specific types of objects (e.g., Royal Bengal Tiger)
- the depiction of a particular type of event (e.g., Football game)
- the presence of named persons, places, or events (e.g., Nick Saban at a press conference)
- the description of a subjective emotion or a personally significant characterisitic (e.g., LSU fans at a conference bowl game)

This list of image features and attributes is presented in increasing levels of subjectiveness and abstraction. A classification of query types based on a similar analysis of image features and attributes was developed by (Eakins & Graham, 1999). They aggregated queries founded on image content into three levels of increasing complexity:

Level 1 Queries. Queries of type level 1 are comprised of *primitive* features, such as color, shape, and texture. This type of query is objective and composed of features directly derived from images using image-processing algorithms. There is no need to consult external data sources for classification guidance. Examples of this type of query include "retrieve all images with red blobs in the middle of the image", "retrieve images that contain blue squares, rectangles, and diamonds", and "retrieve images that look similar to this image". This latter type of query is called *query by example* and is a major focus of CBIR research, including this paper. Level 1 queries correspond to the first item in the list above. Examples of the types of Level 1 queries are given in Figure 2.

Level 2 Queries. Queries of this type are comprised of *logical* features that require some level of inference about the identity of things in the image. An outside knowledge base is required for this type of query. The field of computer vision, particularly the subfield concerned with model-based vision operations, falls into this category. Level 2 queries can be further classified as queries of objects of a

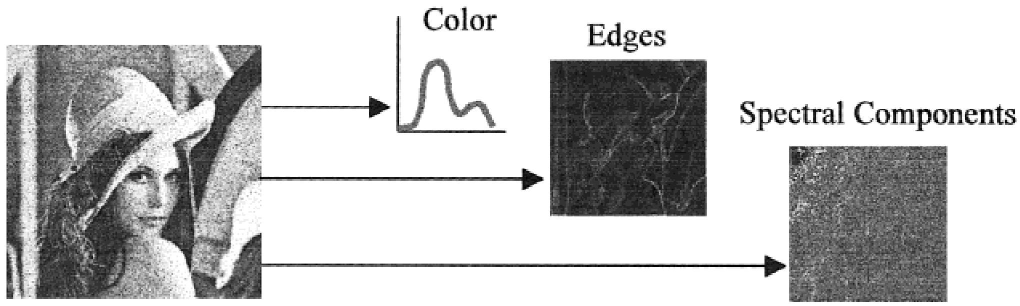

FIG. 2. Examples of Level 1 Queries for Lena Image.

given type (item two in the list above) and queries of individual objects or persons, as is depicted in Figure 3.

Level 3 Queries. Queries of this type of composed of *abstract* notions and attributes and require a significant amount of higher level reasoning about meaning and purpose. Typically, it is very difficult to automate this type of reasoning. Because the link between image content and abstract concepts requires complex reasoning and subjective judgment, systems based on this type of visual processing will incorporate a “human in the loop” to guide the computer to a correct solution. A query of this type is given in Figure 4.

Level 1 queries are generally considered to be the focus of CBIR research and systems development. Levels 2 and 3 are considerably harder to implement, as exemplified by several decades of computer vision research. They can be considered *semantic image retrieval*, a subcategory of CBIR. The distinction between Level 1 and Level 2 is not artificial; there exists a significant gap between them in terms of what computer science and cognitive modeling can currently deliver. This knowledge representation and modeling gap (or chasm, depending on who you ask) is commonly referred to as the *semantic gap.*

2.1. General System Structure

The basic problem addressed in this paper is the specification of unconstrained query images by a user to a CBIR system to search and retrieve a set of result images that are similar to the images initially specified. The search and retrieval process is based on the visual features contained in the images that comprise both the query set and the image database. The general computational framework of a CBIR system is depicted in Figure 5. The entire process starts with the construction of an image database. The images to be added to the database are processed by a feature extraction algorithm. The output of this algorithm is a feature representation, which is the data structure actually stored in the database and used to compute similarity.

The same feature extraction algorithm is used to process the query image and the images contained in the database. Hence, the same feature representation is computed for the query image as was for each image in the database. The similarity measure then compares the query feature representation with each of the feature representations in the database. Those feature representations deemed “similar” are returned to the user as a result set. It is not strictly necessary that an image be specified as part of the query. Queries can be specified by sketches or by graphical user interface tools (Flickner et al., 1995; Gray, 1995). However, the ultimate result of the query specification must be the same feature representation that is used by the database to store and index images. The specification of the query can be with an example image, a user drawn sketch, or explicit information from the user about the primitive features of interest.

2.2 Visual Features

The extraction of visual features from an image is one of the fundamental operations of CBIR. Visual features are properties of an image that are extracted using image processing, pattern recognition, and computer vision methods (Duda & Hart, 1973; Gonzalez & Woods, 1992). Most methods of feature extraction focus on color, texture, shape, and spectral properties of images and, thus, are considered required elements at the primitive level.

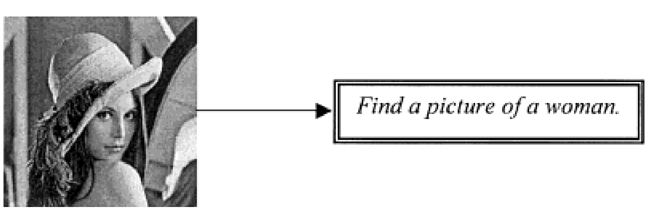

FIG. 3. Example of a Level 2 Query for the Lena Image.

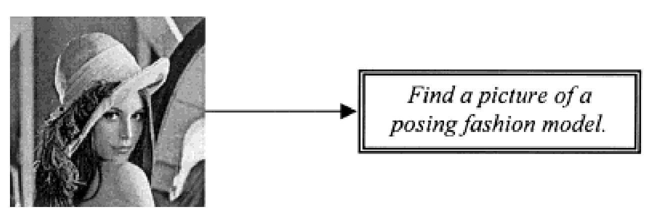

FIG. 4. Example of a Level 3 Query for the Lena Image.

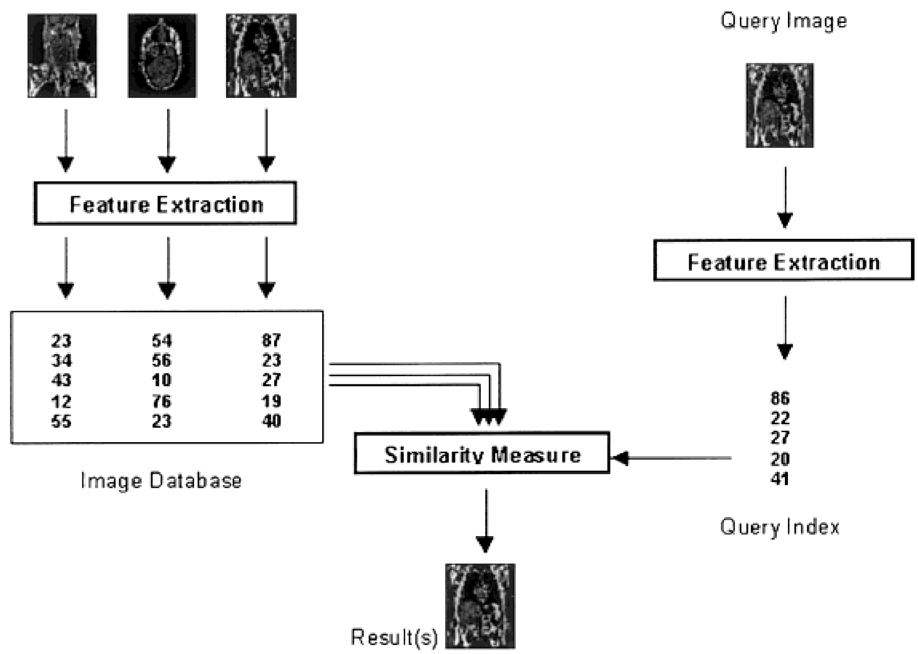

FIG. 5. Computational Framework of CBIR Systems.

Color is by far the most common visual feature used in CBIR, primarily because of the simplicity of extracting color information from images (Flickner et al., 1995; Gray, 1995; J. Huang, 1998; Pass & Zabih, 1996; Smith, 1997; Stricker, 1994; Swain & Ballard, 1991). (Stricker & Swain, 1994) present a thorough analysis of effectiveness of color histograms intersection for CBIR. Color histograms describe the distribution of pixels of each color in the color space of the image. The algorithms developed in (Gray, 1995; J. Huang, 1998; Pass & Zabih, 1996; Smith, 1997) augment color histograms with other derivative visual features, such as spatial coherence or edge information. (Carson et al., 1997) develop a region based color query method. These methods show impressive results for particular classes of image.

Texture is a pervasive yet ill-defined property of images; it can be difficult to define, but we know it when we see it. The analysis of texture in digital images has received much attention in the areas of machine vision, pattern recognition, and image processing (Gonzalez & Woods, 1993; Haralick et al., 1973; Picard & Minka, 1995). In (Picard, 1996), texture is described as lacking a specific complexity, containing high frequency information, and having a finite range of scalability. A statistical approach developed in (Haralick et al., 1973) is the gray level co-occurrence matrix. This method characterizes texture by generating statistics of the distribution of intensity values as well as position and orientation of similar valued pixels. Recent approaches compute texture present in images by employing spectral methods, namely Fourier and wavelet transforms (Chang & Kuo, 1993; Prasad & Iyengar, 1997).

The recognition of shapes is a fundamental perceptual activity, and it is natural that shape-based queries are a component of the primitive level. A number of features of an object's shape in an image are computed for each object in an image and stored. Like color histogram intersection, a query image is analyzed in terms of the same object characteristics and the computer features of the query image are compared to the features of the stored images. Those stored features that best match the query image features are used as the result set. Shape features take on many geometric and non-geometric forms, such as aspect ratio, circularity, moment invariants (Flickner et al., 1995). An example image (QBE) or user sketch is commonly used to construct the shape-based query.

Spectral methods, such as Fourier and wavelet transforms (Prasas & Iyengar, 1997), are used independent of texture analysis to extract features from images. In (Jacobs et al., 1995), wavelet coefficients of images are used to search an image database from a low-resolution example image or user-drawn sketch. Their approach created image signatures of the query and stored images from the Haar wavelet decomposition method. Each signature is a truncated and quantized version of the coefficients computed from the images. A similarity comparison is made by determining the number of significant coefficients in common between the example signature and the stored signatures in the database. The Fourier-Mellin transform is compared to the Haar wavelet decomposition method in (Cherbuliez, 1997).

3. Color as a Visual Feature

Virtually all CBIR systems allow searching capability based on color, an approach pioneered in (Chang & Fu, 1981). Most research and commercial CBIR systems that have been developed, such as QBIC (Flickner et al., 1996), Virage, Excalibur, and Photobook (Pentland et al., 1996) employ color together with other visual features as a search and retrieval mechanism. The results presented in (Stricker & Swain, 1994) placed color histograms on a firm theoretical foundation. In his doctoral research, (Smith, 1997) developed binary representations of color histograms. However, most previous work in color feature extraction and, to a large degree, feature extraction in general, focuses on an approach restricted to a single vector-based representation of features. In particular, representation of color in image has not been investigated much beyond color histograms.

The extraction of color features from digital images depends on an understanding of the theory of color and the representation of color in digital images. Color spaces are an important component of relating color to its representation in digital form. The transformations between different color spaces and the quantization of color information are primary determinants of a given feature extraction method.

The human eye, through the receptors present in the retina called *rods* and *cones*, perceives color as linear combinations of three *primary colors*. These primary colors, red (R), green (G), and blue (B), have specific wavelength values of 700nm, 546.1nm, and 435.8nm, respectively.

Chromatic light is colored light.[1] The basic terms used to describe chromatic light are *hue*, *saturation*, and *brightness*. Hue is used to describe the dominant wavelength or perceived color of an object.[2] It is the "redness" of an apple or the "yellowness" of a banana. Saturation refers to purity of a hue or the distance a color is from a gray of equal intensity. Red is highly saturated while pink is not. Brightness is the chromatic analogue of intensity for achromatic light. Hue and saturation are sometimes combined and referred to as *chromaticity*.

Given the response functions $f_r(\lambda)$, $f_g(\lambda)$, and $f_b(\lambda)$ for each of the primary colors, the following equation of the electromagnetic response for a wavelength λ is defined as

$$F(\lambda) = Xf_r(\lambda) + Yf_g(\lambda) + Zf_b(\lambda)$$

The values (X,Y,Z) are called the *tristimulus values* for color $F(\lambda)$ and denote the respective amounts or red, green, and blue necessary to form a color. Commonly, the tristimulus values are used to specify a color in terms of its trichromatic coefficients

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z}$$

Tristimulus values are, in general, normalized. Thus, the trichromatic coefficients are likewise normalized. While X, Y, and Z may all be equal to 1, the trichromatic coefficients are subject to the relation $x+y+z=1$. The primary colors and tristimulus color theory is the mechanism that allows televisions to display the colors we see. Cathode ray tubes (CRTs) have three channels of red, green, and blue. By varying the voltage of each channel and combing their outputs, each pixel on a television screen can output a large array of colors. Secondary colors are specified in terms of the primary colors. Magenta is formed from equal amounts of red and blue light. Yellow is formed from equal amounts of red and green light. Cyan is formed from equal amounts of green and blue light. For pigments used for print, the primary and secondary color designations are reversed (Gonzalez & Woods, 1992).

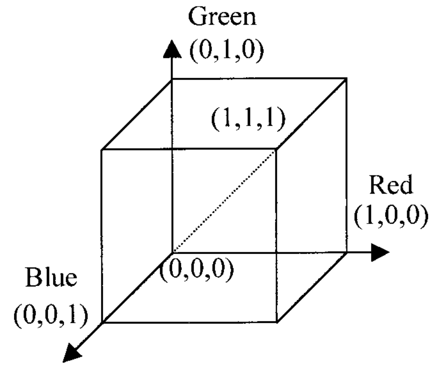

FIG. 6. RGB color cube. The red, green, and blue ordinates are labeled as unit vectors. The line defined by (0,0,0) and (1,1,1) represents the set of grayscale levels.

3.1. Color Spaces

A color space (or color model) is used to specify a three-dimensional color coordinate system and a subspace of the system in which colors are represented as points. The most common color space for digital images and computer graphics is the RGB color space (Figure 6) in which colors are represented as linear combinations of red, green, and blue color channels. The primary reason for the ubiquity of the RGB color space is due to the use of CRT monitors and color raster graphics devices. Additionally, most digital image formats store pixel values from the RGB color space. Thus, it would seem reasonable to based color feature extraction methods on the RGB color space. However, there are sufficient drawbacks to the RGB color space to warrant the use of transformations and quantizations (Smith, 1997) to other color spaces.

The RGB color space is not perceptually uniform. The distance between two points in the color space does not suggest that the two colors are similar or dissimilar. Additionally, the three color channels of the RGB color space do not vary consistently with one another with respect to brightness. Therefore, the pixels of the images in the image database and query examples must be transformed into an alternative color space that satisfies the properties of uni-

[1] Achromatic light is void of any color. It is characterized by the perception of intensities, such as the gray levels one may see on a black and white television.

[2] Some colors do not have a dominant wavelength. Purple is an example.

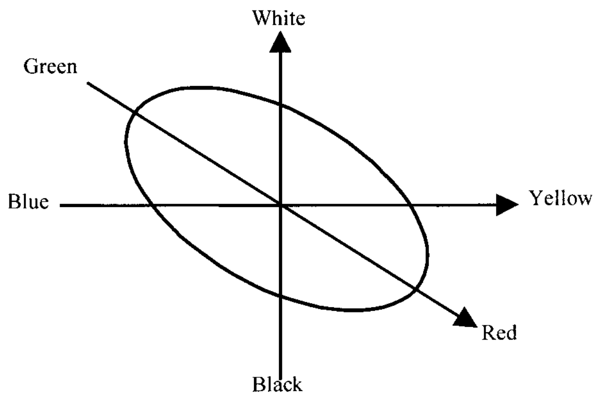

FIG. 7. CIELAB Color Space.

formity, completeness, and uniqueness. One common alternative color space that satisfies these conditions is the CIE family of color spaces.

The CIELUV and CIELAB color spaces were created in 1976 as alternatives to color spaces that assumed luminance was constant for all colors. An equal amount of emphasis is placed on chromaticity and luminance. As a result, the three properties desirable of a color space in perceptually sensitive application, uniformity, completeness, and uniqueness, are satisfied. This paper proposes the use the CIELAB color space as a foundation for feature representation and similarity measurement. This choice is motivated primarily because of the almost perceptual uniformity of the space, a characteristic that is a departure from most other CBIR approaches. A detailed description of the transformation of a point in RGB color space to CIELAB color space is given in (Zachary, 2000).

The domain of values for L*, a*, and b* is from R and, hence, necessitates that a quantization be applied to partition the CIELAB color space into non-overlapping partitions which completely cover the original continuous space. Our quantization of the CIELAB color space strikes a balance between fidelity and the dimensionality of the resulting quantization (Figure 7). The axis defined by L* defines the *brightness* of the color, that is blackness to whiteness. The a* and b* axes are defined by an *opponent color theory* (Berger-Shunn, 1994) in which the a* ordinate describes the redness (+80) to greenness (-80) of a color, and the b* ordinate shows the yellowness (+80) to blueness (-80) of a color. Since a* and b* define the most significant characteristic of a color, namely its chroma, the quantization of these values will be higher than for L*. This is justified by another argument. The human visual system discerns changes in brightness by much larger gaps than changes in color. Especially as L* increase, the human eye cannot detect changes in brightness for small changes in the value of L*. Thus, the brightness values do not require as fine a resolution as the chroma colors. Additionally, L* can be quantized into equidistant bins as opposed to bins fit to logarithmic distances from zero because the logarithmic response of the human visual system has been accounted for in the transformation formulas from the RGB color space to the CIELAB color space.

The axis defined by L* defines the *brightness* of the color, that is blackness to whiteness. The L* axis is quantized to five equidistant bins corresponding to L* = {[0, 20), [20, 40), [40, 60), [60, 80), [80, 100]}. The a* and b* axes are defined by an *opponent color theory* (Berger-Shunn, 1994) in which the a* ordinate describes the redness (+80) to greenness (-80) of a color, and the b* ordinate shows the yellowness (+80) to blueness (-80) of a color. If both the a* and b* axes is partitioned into eight bins, i.e., {[-80, -60), [-60, -40), [-40, -20), [-20, 0), [0, 20), [20, 40), [40, 60), [60, 80]}, then the quantization of the CIELAB color space results in M = 5*8*8 = 320 distinct colors.

Many other color spaces exist, each with advantages and disadvantages. As mentioned, the RGB color space is an additive color space made popular by the ubiquity of CRTs to display digital images. While easy to implement, it is not linear with respect to human visual perception. Additionally, the RGB color space dependent on the device displaying the colors. The CMY (Cyan, Magenta, Yellow) color space is used mostly for printing output and is not perceptually uniform The HSL (hue, saturation, lightness) color space has several co-spaces it shares characteristics with HSV (hue, saturation, value) and HSI (hue, saturation, intensity). This family of color spaces, too, is not perceptually uniform. The main attraction of these color spaces is the separation of chromaticity (hue & saturation) from luminance (intensity, brightness, and value). The YIQ, YUV, and YCrCb color spaces are used for NTSC, PAL, and JPEG standards, respectively. They are highly device dependent and also perceptually non-uniform. A summary of the comparisons between the different color spaces is given in Table 1.

3.2 Color Representations

The transformation of points in the RGB color space to the quantized CIELAB color space requires an appropriate representation that captures the distribution of the colors in an image. The most common representation is the *color histogram*. The color histogram captures the distribution of colors in an image or region of an image, and its unnormalized definition is the following formula

$$\vec{h}_I = h_{m=1,\ldots,M}[m] = \sum_{x=1}^{X} \sum_{y=1}^{Y} \ddot{a}_{I_{i,j}m}$$

TABLE 1. Summary of color space comparisons.

	Uniform	Complete	Unique	Device independent
RGB	No	Yes	No	No
CMY	No	Yes	No	No
HSL, HSV, HSI	No	Yes	Yes	No
YIQ, YUV, YCrCb	No	Yes	No	No
CIELUV, CIELAB	Yes	Yes	Yes	Yes

where M is the number of quantized colors, X and Y are the width and height, respectively, of an image I, and δ is the Kronecker delta function. An analysis of the metrical properties of the color histogram space is given in (Stricker & Swain, 1992). Normalization of the color histograms is a necessary computation to ensure a unit variance between elements of a histogram, i.e., to eliminate the dependency on the number of pixels that comprise the histogram. Normalized histograms are computed by dividing each element of the histogram by the length of the histogram. The definition of a normalized color histogram space is

$$H = \{(\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_M) | h_i \geq 0, \quad \sum_{i=1}^{M} h_i = 1$$

The color histogram space H is a subset of an M-dimensional vector space and forms the face of an M-dimensional simplex (thus, it is an M-1-dimensional simplex). In order for two distinct histograms h_i and h_j to be distinguishable from one another, they must be separated by a non-zero distance t. This property is called *t-difference* and describes H as a Hausdorf space. The value of t depends on the composition of the image data set. However, Stricker & Swain discovered a bimodal shape to distance distributions for two large image data collection (one of which was randomly generated). The bimodal behavior of the distance distribution suggests that reasonable values of the variable t are found in the first interval of the distance distribution with a large slope. More importantly, this also suggests that the distances between color histograms of images with similar colors or images containing all the colors of the color space are small.

3.3 Similarity Measures

Once the feature representation space has been defined as an M-dimensional color histogram space, the problem of defining the similarity between two images is described as the distance between two points in the color histogram space, denoted as D(**p**,**q**) for points **p** and **q**. We now provide a formal definition of similarity.

Definition 2.2: *An image* ***v*** is more similar to ***u*** *than another image* ***w*** *is to* ***u*** *if D(**u**,**v**)<D(**u**,**w**).*

In (White & Jain, 1996b), similarity measurements are defined in terms of the following parameters:

- A query object **u** and similarity measurement function D(●,●) such that references are ordered in increasing value with respect to D(**u**,**v**), $\forall$ **v** $\in \Delta$.
- A parameter $k \in \mathbf{Z}^+$ is used as an upper bound for the cardinality of the set of nearest-neighbor references as computed by D.
- $T \in \mathbf{R}^+$ places an upper bound on the distance that database vectors can be from **u** in order to be included in the result set R, i.e., $R = \{\mathbf{v} \in \Delta \mid D(\mathbf{u}, \mathbf{v}) \leq T\}$.

The second item describes the *k*-nearest neighbor form of similarity while the third item describes the range-based form for similarity.

Once the abstract notion of similarity is defined in terms of distance, several mathematical formulas for the distance function can be defined. The terms *distance function* and *similarity function* are used interchangeably in this paper.

The distance between two points can be classified as either *metric* or *non-metric.*

Definition 2.3: *A set* X *with elements called points is called a metric space if for any two points* **p** *and* **q** *in* X *there is a number* D(**p**,**q**)$\in\Re$ *called the* distance *from* **p** *to* **q** *such that*

$$\begin{array}{ll} D(p, q) > 0 \text{ if } p \neq q & \text{(non-negativity);} \\ D(p, q) = 0 & \text{(identity);} \\ D(p, q) = D(q, p) & \text{(symmetry);} \\ D(p, q) \leq D(p, r) + D(r, q) \ \forall \ r \in X & \text{(triangle inequality)} \end{array}$$

Any function **D(p**,**q)** *satisfying these three properties is called a* distance function *or* metric *(Rudin, 1976).*

A general class of distance metrics is the Minkowski metrics or L_r*-norms*:

$$D_r(p, q) = \sqrt[r]{\sum_{i=1}^{n} |p_i - q_i|^r}$$

The most commonly used Minkowski distances correspond to r-values of one, two, and infinity. The D_1(●,●) distance function is called the *Hamming* distance, and it corresponds to the L_1-norm.

$$D_1(p, q) = \sum_{i=1}^{n} |p_i - q_i|$$

The D_2(●,●) distance function is the well-known *Euclidean* distance, well known by school children in the earliest algebra courses. The corresponding norm is the L_2-norm.

$$D_2(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

The D_∞(●,●) distance reduces to

$$D_\infty(p, q) = \max_j |p_j - q_j|$$

which is the maximum coordinate or *Chebyshev* distance.

The histogram cosine distance function is closely related to the L_2-norm and is commonly used to compute similarity between text documents (Smith, 1997). The inner product of two vectors **p** and **q** is given by

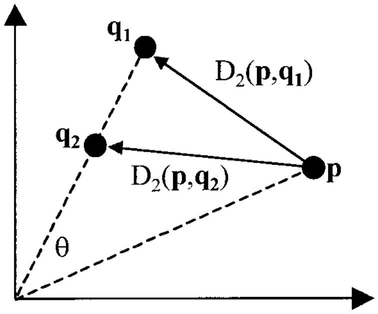

FIG. 8. Comparison of Histogram Cosine Distance to the L2-Norm.

$$p \cdot q = p^T q = |p|\ |q| \cos \grave{e}$$

Thus, solving for θ, we have

$$D_\theta(p, q) = \grave{e} = \cos^{-1} \frac{p^T q}{|p|\ |q|}$$

The histogram cosine distance function measures the difference in direction between two vectors irrespective of the magnitude. The relationship between the histogram cosine distance and the L_2-norm is depicted in Figure 8. We see that the histogram cosine distance between p and q1 is equal to the histogram cosine distance between p and q2. However, $D_2(\mathbf{p},\mathbf{q_1})$ and $D_2(\mathbf{p},\mathbf{q_2})$, are clearly not equal. Therefore, the histogram cosine distance is not a true distance function in the strictest sense since it fails to satisfy the triangle inequality. Researchers have discovered that satisfying the triangle inequality is not necessary in order to define a similarity measure that models human perception.

The general quadratic form for the distance between two vectors is

$$D_Q(p, q) = (p - q)^T A(p - q)$$

where **A** is typically a symmetric square matrix. The form for **A** varies for different distance calculations. In (Niblack et al., 1993), the matrix **A** is defined as a color similarity matrix with

$$a_{ij} = 1 - \frac{d_2(c_i, c_j)}{d_{\max}}$$

The vectors $\mathbf{c}_i$ and $\mathbf{c}_j$ are the i^{th} and j^{th} colors in the histogram space. The function $d_2(\mathbf{c}_i,\mathbf{c}_j)$ is the Euclidean distance between colors $\mathbf{c}_i$ and $\mathbf{c}_j$, and d_{max} is the maximum distance between any two colors in the color space. The effect of computing $D_Q(\bullet,\bullet)$ is the magnitude of the distance between **p** and **q** weighted by the distance between the colors in the color space. The difference between color amounts and similar colors is accounted for in this formula. Other quadratic forms include the Mahalanobis form (Duda & Hart, 1973).

4. Image Entropy and Color

Color histograms have been shown to be a promising method for indexing into image databases. However, for very large image databases and histogram spaces with large dimensions, the computational cost of performing distance calculations can be prohibitive. This section suggests an alternative viewpoint of color histograms based on information theory that offers the potential for a substantial increase in retrieval performance.

The motivation for this section is the desire to reduce the dimensionality of the color histogram space in order to provide a substantial improvement in retrieval performance. Several dimension reduction techniques have been developed, such as principle component analysis (Gerbrands, 1981; Gonzalez & Woods, 1992) and column-wise clustering (Duda & Hart, 1973). Generally, these techniques reduce the dimensionality of the histogram space from n to $k > 1$.

We develop the theory necessary to reduce the dimensionality of the color histogram space to one. The *entropy* of an image is a measure of the information content of the image. As will be seen, the Shannon entropy function maps an n-dimensional vector to the set of real numbers, and, hence, it can be regarded as a dimension reduction to the set of real valued numbers.

4.1. Color Histograms as Probability Density Function

This section expands the discussion of color histograms by describing a color histogram as an estimation of the first-order joint probability density function of an image. This description is important in allowing us to use methods from information theory to expand the characterization of images on the basis of their color contents.

A discrete image $I=F(N_1,N_2)$ of size $N_1 x N_2$ can be statistically characterized as the joint probability density function

$$p(I) \equiv p\{F(1, 1), F(1, 2), \ldots, F(N_1, N_2)\}$$

If each pixel value is statistically independent from all other pixels values, then the joint probability density function is factored into the following form

$$p(I) = p\{F(1, 1)\}p\{F(1, 2)\} \cdot \cdot\ p\{F(N_1, N_2)\}$$

which is the product of its first-order (one-dimensional) marginal densities. For a discrete set of values, the interpretation of p\{F(i,j)\} is developed on the basis of the finite range of possible values for F(i,j). For a digital image source, these values are the possible colors at each pixel, or *reconstruction levels*. It is generally assumed that the distribution of colors across an image follows a uniform distribution, i.e., each color has a 1/M probability to be assigned to a pixel.

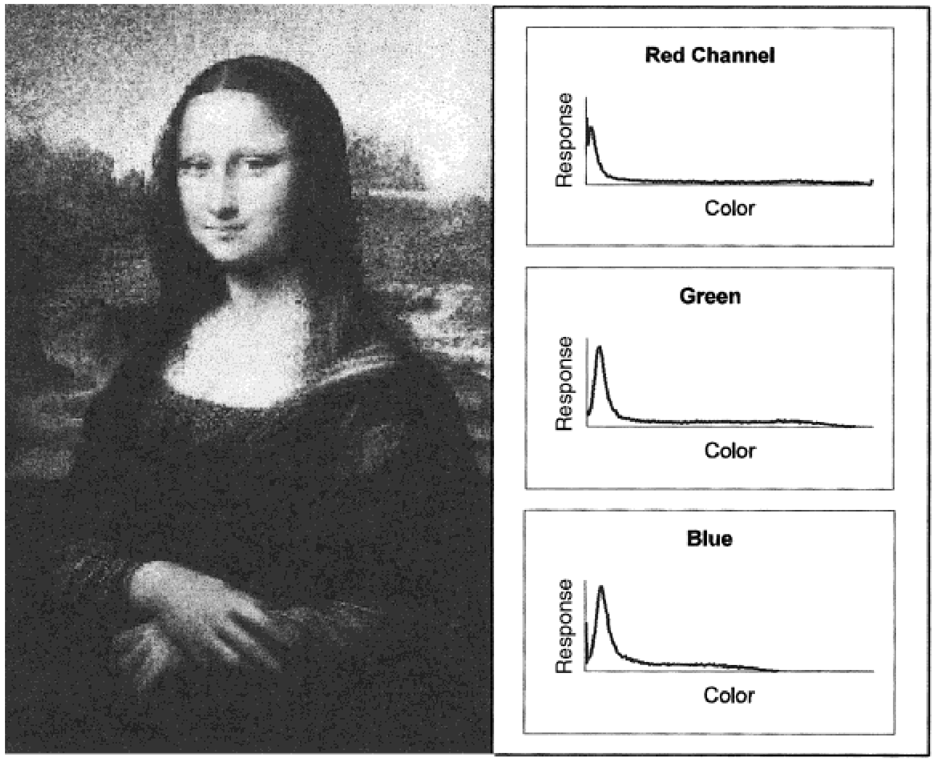

FIG. 9. Mona Lisa and the Histograms of the Tristimulus Channels.

For digital images, the probability density function is a joint probability density function because the pixels, as discrete random variables, are not functions of one another. Additionally, pixels are assumed to be statistically independent because the *value* of a pixel is not a function of other pixel values. Furthermore, the digital image source is assumed to be *ergodic* in the sense that successive samplings of a certain pixel do not determine or affect the outcome of future values at that pixel. Another way to regard this property is that image sources are *memoryless*.

The Laplacian and Rayleigh joint probability density models are used as statistical descriptions in analog analysis of image systems (Pratt, 1978). However, for quantized discrete random variables, i.e., digital images, histograms of the color distribution in an image provide an adequate estimation of first- and second-order joint probability density functions for the image.

For an ergodic image source, the first-order joint probability density function is estimated by the first-order spatial histogram for an image

$$\vec{h}(i) = \frac{N(c_i)}{N_1 \cdot N_2}$$

where N(i) is the number of pixels in the image that are of color c_i. Despite a change in notation, this formula is identical to the formula of color histograms given previously.

Figure 9 displays the histograms for the red, green, and blue responses of the Mona Lisa. The shape of the histograms demonstrates a pattern in which the response for each of the tristimulus values at the darker end of the histogram is greater than the response at the brighter end. The Mona Lisa is a good example of the color distribution typical of most natural images.

The approximation of the second-order joint probability density function plays a significantly less role than the approximation of the first-order probability density function in image analysis. For completeness, the second-order joint probability density function of an image is estimated by the second-order spatial histogram of an image. The latter is a measurement of the occurrence of pairs of pixels at given color values separated by a specific distance. The interested reader is referred to (Gonzales & Woods, 1992; Pratt, 1978) for details of the second-order spatial histogram formula.

4.2. Information Theory and the Entropy Function

Given a vector **v** of numbers from a set $\{x_1, x_2, \ldots, x_n\}$ where the probability that $x_i \in \mathbf{v}$ is $\mathbf{p}_i = P(x_i)$, the entropy of **v** is given by the formula

$$H(v) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

The mathematics describing H(**v**) in the context of communications theory was developed in (Shannon, 1948) and is the most common definition of entropy in the literature. It

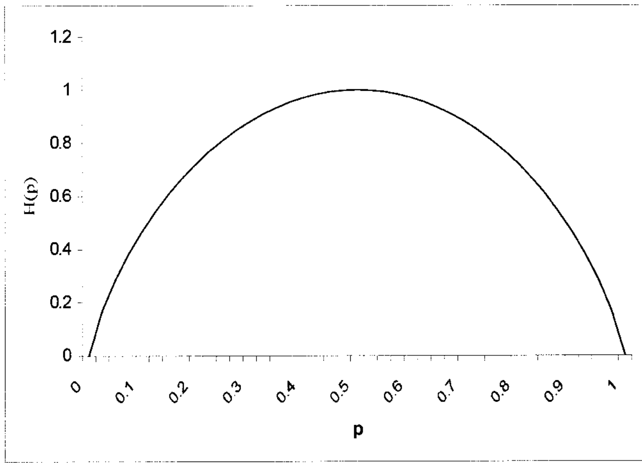

FIG. 10. Entropy function in two dimensions.

should be clear that H(**v**) is a function of the probability distribution of some random variable and not a function of the actual values the variable may assume. As seen in Figure 10, H(**v**) is a continuous, positive, and concave function of $[0,1]^n \in \Re^n$ that maps to $[0,1] \in \Re$. The function H(**v**) = 0 when $v_i = 1$ and $v_j = 0$ for all $j \neq i$.

Before pursuing a quantitative description of similarity between images represented as entropy values, we investigate the sensitivity of the entropy function to small perturbations in the probability distribution function. Given a uniform probability distribution **v** $=\{p_1=1/M,\ p_2=1/M, \ldots, p_M=1/M\}$ associated with the maximum entropy, assume that a new vector **u** assumes the form

$$u_i = v_i + \ddot{A}v_i, \quad \sum \ddot{A}v_i = 0$$

The Taylor polynomial expansion to the second derivative of H(**u**) is

$$H(u) = H(v) + \sum \frac{\partial H}{\partial v_i} \ddot{A}v_i + \frac{1}{2!} \sum \frac{\partial^2 H}{\partial v_i^2} (\Delta v_i)^2$$

The first partial derivative with respect to H(**u**) evaluates to

$$\frac{\partial H(u)}{\partial v_i} = -1 - \sum \ln\left(\frac{1}{M} + \ddot{A}v_i\right)$$

The second partial derivative with respect to H(**u**) evaluates to

$$\frac{\partial^2 H(u)}{\partial v_i^2} = -\frac{1}{\dfrac{1}{M} + \ddot{A}v_i}$$

If these partial derivatives are evaluated at $\Delta v_i=0$, then the Taylor polynomial for H(**u**) becomes

$$H(u) = H(v) - \frac{M}{2} \sum (\Delta v_i)^2$$

The term associated with the first derivative becomes zero based on the assumption that $\Sigma \Delta v_i = 0$. Therefore, we conclude from this sum that if **u** represents a small change in the probability distribution **v**, then the corresponding difference |H(**u**) – H(**v**)| is likewise small.

4.3. Image Entropy as a Visual Feature

The definition of color histograms as first-order joint probability density functions suggests that the entropy of an image can be calculated. In fact, this is exactly the case. The definition of **v** is derived from the interpretation of first-order spatial histograms as a joint probability density function. An element $\mathbf{v}_i$ is the percentage of pixels in the image that belong to the quantized color *I* and is also a close approximation to the value of the joint probability density function value p_i at *i*. The correlation of each histogram bin $\mathbf{v}_i$ to a probability function value p_I yields the function

$$H(v) = -\sum_{i=1}^{M} v_i \log(v_i)$$

Figure 11 gives the entropy values calculated by the formula for some recognizable digital images. Images such as Clown, Lena, and Mandril have complex color distributions and, hence, have higher entropy values. An image with a simple color distribution, such as Pleides, has a smaller entropy value.

For a digital image source, there are many interpretations of H(**v**), including

- The average uncertainty of **v**.
- The theoretically least number of bits necessary to encode **v**.
- A measure of the randomness of the color distribution in **v**.

An increase in image entropy corresponds to more uncertainty and more information contained in an image. Thus, the use of image entropy as a discriminant between two images is based on the idea that a meaningful difference between two image entropy values corresponds to a meaningful difference between the two source images. For example, in Figure 11, a meaningful difference between the entropy values for the Pleides and Venice images corresponds to a meaningful difference between the images themselves.

Our interest will focus on the third interpretation of H(**v**) since it seems to hint that entropy captures a characteristic of an image meaningful in making a determination of whether images are similar. The Shannon definition of H(**v**) assigns information based on “sharpness” of the distribution that an event, or a group of pixels will have a given color value, will occur. Based on the mathematical properties

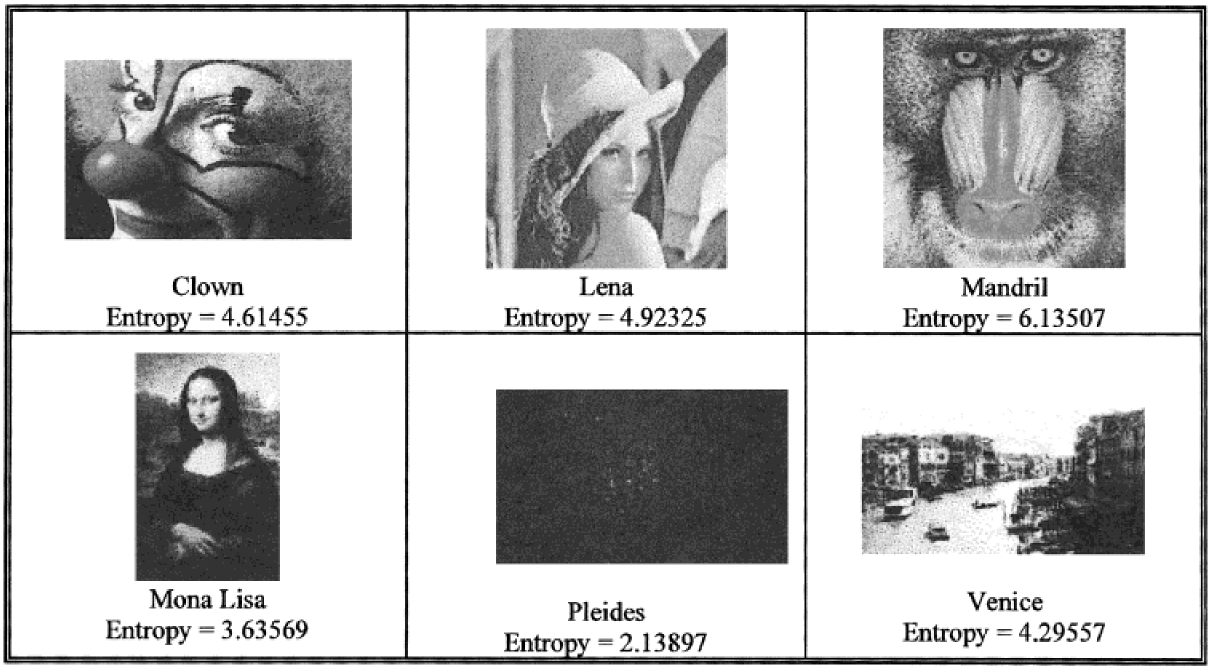

FIG. 11. Entropy values for some recognizable digital images.

above, H(**v**) = 0 implies a digital image has all pixel values set to the same value. Additionally, H(**v**) is maximized when all possible colors in the color space of the image are equally represented. Intuitively, this means we can express more information in an image that has more colors than in an image with fewer colors.

A fundamental element of comparing images that are in a certain representation is the definition of similarity. The definition of the similarity function depends on the metrical properties of the space in which the representations are defined. For color histogram spaces, the definition of similarity in terms of *norms* is natural given the theory of finite dimensional vector spaces. The definition of similarity between points in entropy space must be based on an understanding of the metrical properties of the space regardless of whether a metric or non-metric similarity function is defined.

4.4. Entropy Difference

The definition of similarity between color images is based on the L_1-norm between two points in the color histogram space. In the entropy space, this definition degenerates to the absolute value of the difference between two entropy values. The formula is given by

$$D_{L_1\_\mathrm{Entropy}}(p, q) = |H(p) - H(q)|$$

which is a straightforward application of the definition of a Minkowski metric given above. As such, the similarity metric $D_{L1\_Entropy}$ possesses the four properties of any distance function on a metric space, namely the non-negativity property, the identity axiom, the symmetry axiom, and the triangle inequality property.

This rather simple formulation has some interesting implications and properties. It is obvious that since this definition is simply subtraction over values in the interval [0,1], then the space is *t*-different for some value of *t* greater than zero.

The color histogram space *H* forms the faces of an M-dimensional simplex. Recall that a set of points $v_1, v_2, \ldots, v_M$ in $\Re^M$ spans a hyperplane defined by the linear combinations $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_M v_M$ such that $\lambda_1 + \lambda_2 + \ldots + \lambda_M = 1$. Figure 12 shows a 2-simplex defined by three unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. Any linear combination $\mathbf{v} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \lambda_3 \mathbf{e}_3$ where $\lambda_1 + \lambda_2 + \lambda_3 = 1$ translates to a point on the face of the triangle. If the entropies of the points on the face of the 2-simplex are plotted as a contour, then the distribution is such that the minima are found at the vertices of the 2-simplex. The maximum entropy corresponds to the point at the center of the 2-simplex corresponding to 1/3 $\mathbf{e}_1$ + 1/3 $\mathbf{e}_2$ + 1/3 $\mathbf{e}_3$.

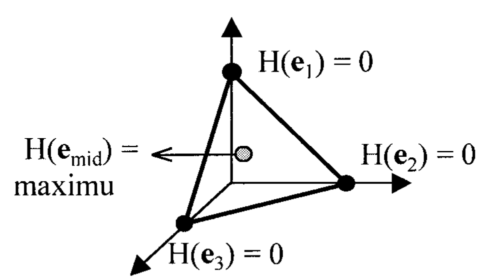

FIG. 12. A color histogram space of dimension 3 represents a 2-simplex, or a triangle.

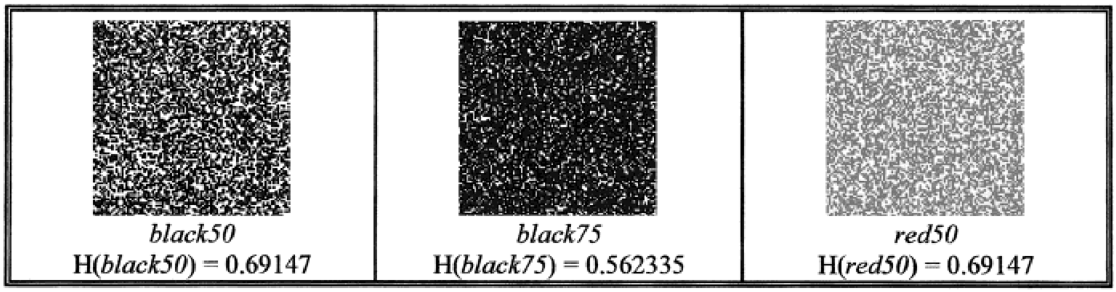

FIG. 13. Entropy values for three random images.

Geometrically, the entropy minima correspond to points in the color histogram space that are a maximal distance from one another. The interpretation in terms of the content of digital binary images is a completely white image and a completely black image which are more similar to one another (with entropies equal to zero) than to any other image. This includes a white image with a single back pixel. This will have a serious implication for using the use of entropy values in an indexing algorithm for color images.

An interesting relationship exists as a quantitative description of the bounds on the entropy function by the L_1-norm of two probability density functions **p** and **q**. This bounds is expressed in the following theorem from (Cover & Thomas, 1991).

Theorem 3.3 *(L_1 Bound on Entropy)*
Let **p** *and* **q** *be two probability density functions over a space* H *such that*

$$|H(p) - H(q)| \leq -\|p - q\|_{L_1} \log \frac{\|p - q\|}{|H|}.$$

Then,

$$\|p - q\|_{L_1} = \sum |p_i - q_i| \leq \frac{1}{2}.$$

This upper bound on |H(**p**) – H(**q**)| provides an important insight into the expected results of using entropy as an indexing key for image in an image database. We would expect that fewer results be retrieved for the entropic L_1-norm than for the color histogram L_1-norm.

It was shown that a meaningful difference between image entropy values for two images implies a meaningful difference between the images themselves. This is primarily a function of the entropy definition as a measure of the information for a given source. However, from a perceptual perspective, the converse is not necessarily true. That is, a gross perceptual difference in images does not imply a difference in entropy values. The value |H(**p**) – H(**q**)| can approach zero for two very dissimilar images and, yet, be greater than zero for two very similar images. For example, in Figure 13, three images are shown. Two of these images display randomly distributed black pixels on a white background in proportions of 50% and 75%. They are named *black50* and *black75* respectively. The other bicolor image, named *red50*, has a random distribution of red pixels over 50% of the image. The entropy differences are |H(*black50*) – H(*red50*)| = 0.0 and |H(*black75*) – H(*red50*)| = |H(*black75*) – H(*black50*)| = 0.130812. Even to the most casual of observers, *black50* and *black75* are much more perceptually similar than *black50* and *red50*.

From an information theoretic point of view, however, this is not true. The reason is that *black50* and *blackwhite* have identical distributions of black and white pixels, namely there is a 50% allocation to the black pixel bin, a 50% allocation to the white pixel bin, and a 0% allocation to all other colors. The image *black75*, on the other hand, has a 75% allocation to the black pixel bin, a 25% allocation to the white pixel bin, and a 0% allocation to all other color bins. Thus, from the information theoretic perspective, there is no difference in the information necessary to code *black50* and *red50*. However, there is a difference between the information necessary to code *black50* and *black75*. Hence, the entropy values are different for *black50* and *black75* but not for *black50* and *red50*.

From this discussion, we can conclude that the use of |H(**p**) – H(**q**)| as the sole measure of similarity may be inappropriate. Color histogram comparisons using the L_1 norm can distinguish the difference between *red50* and *black75*. Therefore, we do not assert that |H(**p**) – H(**q**)| is capable of providing a meaningful similarity measure based on entropy values alone. This should not be very surprising to the reader since such an assertion would suggest that a single real number contains more information than a vector for distinguishing between two images. The vector always contains more information than the single real number, particularly since the single number is an aggregation of the vector via the entropy function.

This paper asserts that the main benefit of using |H(**p**) – H(**q**)| as a similarity measure is that it suggests an extremely efficient method for retrieving images from a database. We highlight a basic entropy indexing method that is present in more detail in this issue (Zachary & Iyengar, 2000). The strategy is to use the entropy number as a filter to generate an interim result set of images. This interim image result set is then indexed based on the standard retrieval method using the L_1-norm between points in the color histogram space. It should be clear that for all but the most pathological of image databases, the interim result set will be much smaller in size than the entire image database.

5. Summary

The focus of this paper is on an information theoretic description of the color contained in a set of images. The goal was to derive a more compact yet expressive description of images that can be used as discriminant in CBIR systems. The interpretation of digital images as probability density functions enabled us to define the concept of image entropy. Image entropy was described in terms of the randomness in the distribution of colors in an image.

We believe that information theory and entropy have not received an adequate amount of research attention in the image interpretation and pattern recognition fields. Areas for future work include information theoretic descriptions of the other visual features of images, including spatial and geometric features. We are also interested in expanding the gamut of image representations beyond vectors. A companion paper in this issue explores different similarity measures based on the image entropy concept; we suspect that other similarity measures are possible and likely to be defined in the future.

References

Ash, R. (1990), Information Theory, Dover Publications.

Berger-Shunn, A. (1994). Practical Color Measurement: A Primer for the Beginner, A Reminder for the Expert. New York: John Wiley & Sons.

Carson, C., Belongie, S., Greenspan, H., & Malik, J. (1997). Region based image querying, CVPR'97 Workshop on Content-Based Access of Images and Video Libraries. Berkeley, CA: University of California.

Chang, N. & Fu, K. (1981). Picture query languages for pictorial database systems, IEEE Computer, 14(11).

Chang, T. & Kuo, J. (1993). Texture analysis and classification with tree structured wavelet transform, IEEE Transactions on Image Processing, 2(4).

Cherlbuliez, M. (1997). Content based image querying, University of Geneva dissertation.

Cover, T. & Thomas, J. (1991). Elements of Information Theory, Wiley Series in Telecommunications, New York: John Wiley & Sons.

Duda, R. & Hart (1973). Pattern Classification and Scene Analysis, New York: John Wiley & Sons.

Eakins, J & Graham, M. (1999). Content based image retrieval: A report to the JISC technology applications programme, Institute for Image Data Research, University of Northrumbia.

Flickner, M. et al. (1995). Query by image and video: The QBIC system, IEEE Computer, 28(9).

Gerbrands, J. (1981). On the relationship between SVD, KLT, and PCA, Pattern Recognition, 14.

Gonzalez, R. & Woods, R. (1992). Digital Image Processing, New York: Addison Wesley.

Gray, R. (1995). Content-based image retrieval: Color and edges, Dartmouth College Department of Computer Science Tech Report TR 92-252.

Gupta, A. (1996), The virage image search engine: An open framework for image management, Proceedings of the SPIE Storage and Retrieval for Image and Video Libraries, IV, 2670.

Huang, J. (1998). Color spatial image indexing and applications, Cornell University Ph.D. Dissertation.

Jacobs, C., Finkelstein, A., & Salesin, D. (1995). Fast multiresolution image querying, SIGGRAPH'95 Conference Proceedings.

Jagersand, M. (1995). Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach, Proceedings of the 5th International Conference on Computer Vision.

Niblack, W. et al (1993). The QBIC project: Querying images by content using color, texture, and shape, SPIE Vol. 1908.

Pass, G., Zabih, R., & Miller, J. (1996). Comparing images using color coherence vectors, ACM Conference on Mulitmedia.

Pentland, A., Picard, R., 7 Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases, Proceedings of the SPIE Storage and Retrieval for Image and Video Databases II, February.

Pianykh, O. (1998). Lossless set compression of correlated information, Louisiana State University Ph.D. Dissertation.

Poynton, C. (1999). Frequently Asked Questions about Color, [ONLINE] available at http://www.inforamp.net/~poyton.

Prasad, L., & Iyengar, S.S. (1997), Wavelet Analysis with Application to Image Processing, CRC Press.

Pratt, W. (1976). Digital Image Processing. New York: John Wiley & Sons.

Rudin, W. (1976). Principles of Mathematical Analysis, International Series in Pure and Applied Mathematics. McGraw-Hill.

Santini, S. & Ramesh Jain, R. (1999). Similarity measures, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(9).

Sclaroff, S. et al. (1997). ImageRover: A content based image browser for the World Wide Web, Proceedings of the IEEE Workshop on Content Based Access of Image and Video Libraries.

Shannon, C. (1948). A mathematical theory of communication, Bell Systems Technical Journal, 27.

Smith, J. (1997). Integrated spatial and feature image systems: Retrieval, analysis, and compression, Columbia University Ph.D. Dissertation.

Stricker, M. & Swain, M. (1994). The capacity and sensitivity of color histogram indexing, University of Chicago Department of Computer Science CS-94-05.

Swain, M. & Ballard, D. (1991). Color indexing, International Journal of Computer Vision, 7(1).

Zachary, J. (2000). An information theoretic approach to content based image retrieval, Louisiana State University Ph.D. Dissertation.

Zachary, J., & Iyengar, S. (1999). Content based image retrieval systems, IEEE ASSET, March 27–29, Dallas, Texas.

Zachary, J., & Iyengar, S. (2001), Information theoretic similarity measures for content based image retrieval. Special Topic Issue on Visual Based Retrieval Systems and Web Mining, Journal of the American Society for Information Science and Technology, 52, 857–868.