Information Theoretic Similarity Measures for Content Based Image Retrieval

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

Content-based image retrieval is based on the idea of extracting visual features from image and using them to index images in a database. The comparisons that determine similarity between images depend on the representations of the features and the definition of appropriate distance function. Most of the research literature uses vectors as the predominate representation given the rich theory of vector spaces. While vectors are an extremely useful representation, their use in large databases may be prohibitive given their usually large dimensions and similarity functions. In this paper, we propose similarity measures and an indexing algorithm based on information theory that permits an image to be represented as a single number. When use in conjunction with vectors, our method displays improved efficiency when querying large databases.

1. Introduction

As digital images become an increasingly important class of data, the need to store and retrieve images in an intuitive and efficient manner must be addressed. 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.

In another paper in this issue, we expand existing visual feature representations based on vector spaces to improve retrieval performance and efficiency. Our approach is based on the information theoretic concept of image entropy. In this paper, we present two similarity measures that use information theory to compare images.

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 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

2.1. 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.

2.2. Color as a Visual Feature

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

TABLE 1. Common distance functions defined on an M-dimensional vector space.

Metric Name	Formula
Minkowski Metric (L_r norm)	$D_r(\mathbf{p}, \mathbf{q}) = \sqrt[r]{\sum_{i=1}^{n} \lvert p_i - q_i \rvert^r}$
Histogram Cosine (non-metric)	$D_\theta(\mathbf{p}, \mathbf{q}) = \grave{e} = \cos^{-1} \frac{\mathbf{p}^T\mathbf{q}}{\lvert\mathbf{p}\rvert\lvert\mathbf{q}\rvert}$
Quadratic Form	$D_Q(\mathbf{p}, \mathbf{q}) = (\mathbf{p} - \mathbf{q})^T\mathbf{A}(\mathbf{p} - \mathbf{q}), \quad a_{ij} = 1 - \frac{d_2(\mathbf{c}_i, \mathbf{c}_j)}{d_{\max}}$

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.

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 (Gupta, 1996), 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.

2.3. Color Representations and Similarity Measures

An appropriate representation that captures the distribution of the colors in an image is necessary in the computation of similarity between images. The most common representation is the *color histogram*. The color histogram captures the distribution of colors in an image or region of an image as a point in an M-dimensional vector space. 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. The definition of a normalized color histogram space we assume is

$$H = \{(\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_M) \mid h_i \geq 0, \sum_{i=1}^{M} h_i = 1 \quad h_i = \frac{N(c_i)}{N_1 \cdot N_2}$$

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. This suggests that the distances between color histograms of images with similar colors or images containing all the colors of the color space are small.

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**. Similarity between images can be defined in terms of D(**p**,**q**) as:

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**).*

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*. Recall that a function defined on two points in a metric space is called a metric function if it is non-negative and satisfies the identity, symmetry, and triangle inequality axioms. A list of common metrics is given in Table 1.

3. Information Theory, Image Entropy, and Similarity

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 chapter 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 $k > 1$.

This chapter develops 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.

3.1. Color Histograms as Probability Density Function

This section expands color histograms by describing a them 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_1xN_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

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

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.

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*.

Given a vector $\mathbf{v}$ 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 $\mathbf{v}$ is given by the formula

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

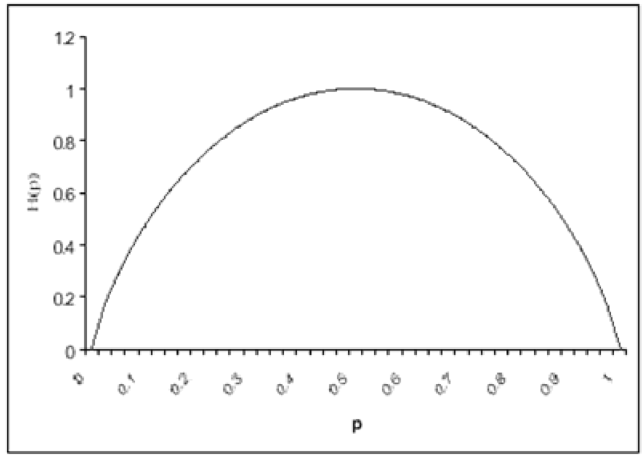

FIG. 1. Entropy function in two dimensions.

The mathematics describing $H(\mathbf{v})$ in the context of communications theory was developed in (Shannon, 1948) and is the most common definition of entropy in the literature. It should be clear that $H(\mathbf{v})$ 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 1, $H(\mathbf{v})$ is a continuous, positive, and concave function of $[0,1]^n \in \Re^n$ that maps to $[0,1] \in \Re$. The function $H(\mathbf{v}) = 0$ when $v_i = 1$ and $v_j = 0$ for all $j \neq i$.

The sensitivity of the entropy function to small perturbations in the probability distribution function is explained in our other paper in this issue.

3.2. 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 $\mathbf{v}$ 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.

Figure 2 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 $H(\mathbf{v})$, including

1. The average uncertainty of $\mathbf{v}$.
2. The theoretically least number of bits necessary to encode $\mathbf{v}$.
3. A measure of the randomness of the color distribution in $\mathbf{v}$.

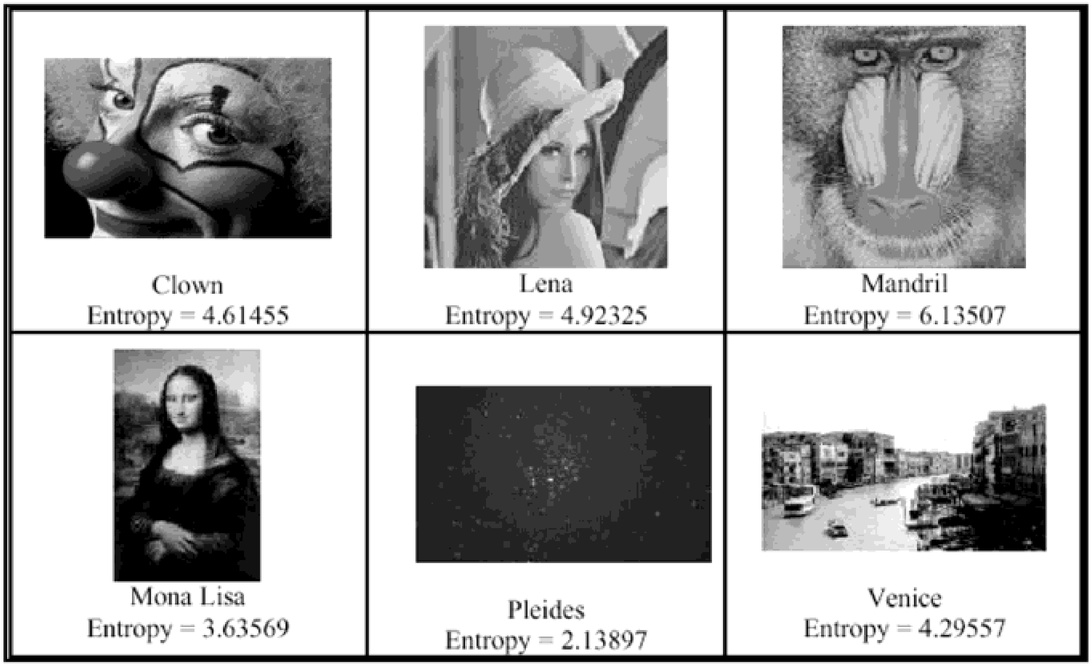

FIG. 2. Entropy values for some recognizable digital images.

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 2, 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 above, $H(\mathbf{v}) = 0$ 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.

3.3. Entropy Difference

The usual 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

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

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 3 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

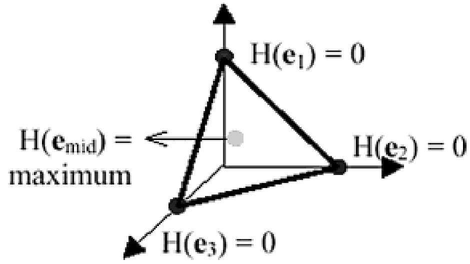

FIG. 3. A color histogram space of dimension 3 represents a 2-simplex, or a triangle.

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 $\mathbf{e}_3$.

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 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*

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

Then,

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

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 4, 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

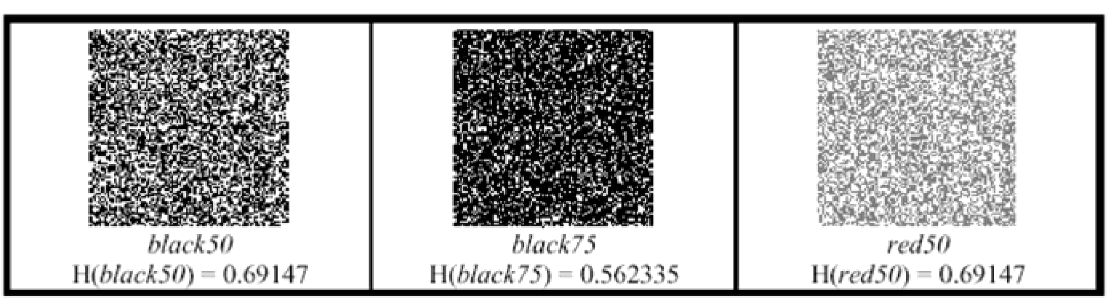

FIG. 4. Entropy values for three random images.

capable of providing a meaningful similarity measure based on entropy values alone. This should not be a 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 $|H(\mathbf{p}) - H(\mathbf{q})|$ as a similarity measure is that it suggests an extremely efficient method for retrieving images from a database. 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. The following section describes this algorithm in more detail.

3.4. Entropy Enhanced L1 Norm Algorithm

Color histogram indexing is based on the computation of distances between points in the color histogram space. Functions such as the L_1 norm take two vectors and compute the distance between them, effectively providing a mapping from $\Re^M$ to $\Re$. The maximum relative entropy function provides the same mapping. For very large databases that must be searched, sophisticated indexing methods are required to alleviate the computational effort in sequentially searching a list of images.

The *k-nearest neighbor rule* classifies a query histogram **v** based on the retrieval of the *k* nearest histograms in the image database. The alternative *range* query is a clustering method that labels as similar all database samples within a given distance *T* of the query histogram **v**. Because traditional Database Management Systems (DBMSs) do not handle multidimensional data very efficiently, methods such as point quad-trees, k-d trees, R-trees, R^*-trees, and R^+-trees have been proposed to index and retrieve data contained in multidimensional spaces (Duda & Hart, 1973; Sellis et al., 1987; Samet, 1990).

The computation of the entropy of each image in the image database suggests an interesting possibility for improving the computational efficiency of search the database without using the sophisticated multidimensional indexing methods listed above.

The *Entropy Enhanced L_1 Norm* (EELN) algorithm is outlined as follows:

For each image I in the image database, computed H(I).
Sort the list of image entropies $\{H(I_j)\}$ in ascending order $\rightarrow$ L.
Compute $H(I_q)$ given a query image I_q.
Search L for the $H(I_j)$ such that $|H(I_j) - H(I_q)| < \epsilon$.
Insert $I_i \rightarrow R_{entropy}$.

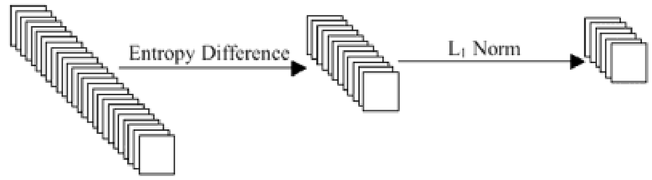

FIG. 5. EELN $\rightarrow L_1$ Norm Search Space Reduction Process.

Search $R_{entropy}$ for the k_2 closest color histograms using the histogram L_1-norm.

The fundamental idea of the EELN algorithm is graphically depicted in Figure 5.

The EELN algorithm is based on using the image entropy difference formula to decimate the number of items for an L_1 norm search in the image database. We have shown that using image entropy in the absence of other information does not discriminate among images satisfactorily. However, the entropy difference formula can be applied to the database to return an initial set of retrieved images. This initial set, which is smaller in size than the entire database, is then searched using the L_1 norm to retrieve a final set of images similar to the initial query image.

This algorithm will have maximum effectiveness if two conditions relating to performance are met. These conditions are as follows:

1. The set of images retrieved from the EELN algorithm must be a proper subset of the set of images retrieved from using the L_1 norm alone. This implies that the size of the results set from the EELN algorithm should be strictly less than the size of the result set from using the L_1 norm alone. Additionally, this implies that there should be no false positives contained in the result set of the EELN algorithm. Subjectively, we should expect that the quality of the result set from the EELN algorithm is greater than the quality of the result set from using the L_1 norm alone.
2. The EELN algorithm should execute faster than the L_1 norm, particularly for very large image databases.

It will be shown that both conditions are satisfied in Section 3.6.

3.5. Maximum Relative Entropy

The primary drawback of the entropy difference formula in the previous section is that it only measures similarity between two distinct probability density functions only after the entropies for the two distributions have been computed. Approaches such as the L_1-norm defined on the color histogram space perform the similarity measurement prior to aggregating a feature into a single number. We present an alternative approach to the L_1-norm that follows our theme of using concept from information theory to measure image similarity.

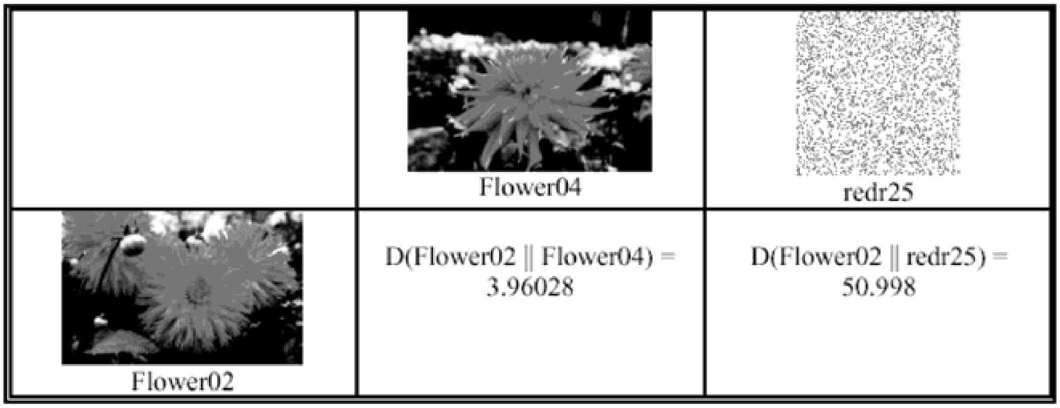

FIG. 6. Comparison of D(p || q) values.

The *relative entropy* or *Kullback-Leibler distance* measures the distance between two probability density functions. The relative entropy is given by

$$D(\mathbf{p}\|\mathbf{q}) = \sum_i p_i \log \frac{p_i}{q_i}$$

where, to ensure continuity, we assume that 0 log 0/q = 0 and p log p/0 = ∞.

The relative entropy D(**p** || **q**) between two probability density functions **p** and **q** captures an intuitive notion of contrast between two images. Recall that the entropy of an image captures the amount of information expressed by the colors present in an image. Images with more colors contain more information than images with fewer colors. The relative entropy captures the contrast in expressed information between two images. Two images with a similar representation of colors will have a lower relative entropy value than two images in which one has several more colors represented than the other image. Figure 6 shows the relative entropy values between two pairs of images. Flower02 is compared to Flower04 and redr25, an image with 25% red pixels uniformly distributed across the image. The relative entropy value D(Flower02 || Flower04) = 3.96028, which is less than the relative entropy value D(Flower02 || redr25) = 50.998. By observance, Flower02 and Flower04 contain a similar distribution of colors. However, there is an appreciable difference in the colors distribution between Flower02 and redr25, notably that redr25 does not contain any dark colors pixels.

It can be shown that D(**p** || **q**) ≥ 0 with equality if and only if $p_i = q_i$ for all i. This is known as the *information inequality theorem*. However, the relative entropy is not a metric in the strict sense of the word since it does not satisfy the symmetry axiom or the triangle inequality. As discussed, there are similarity measures used in CBIR systems that are not true metrics. The triangle inequality is typically the condition that is relaxed in the definition of non-metric similarity measures. In practice, it is assumed that the condition D(**p**,**q**) = D(**q**,**p**) be valid for any similarity measure.

We define the *maximum relative entropy* function to be

$$D_{\mathrm{mre}}(\mathbf{p}, \mathbf{q}) = \max\{D(\mathbf{p}\|\mathbf{q}), D(\mathbf{q}\|\mathbf{p})\}$$

This definition has the following properties:

Proposition 3.4: D_{mre}(**p**,**q**) satisfies the identity and non-negativity axioms.
Proof. The *information inequality theorem* implies D_{mre}(**p**,**q**) ≥ 0 since D(**p** || **q**) ≥ 0 and D(**q** || **p**) ≥ 0. Additionally, if $p_i = q_i$ for all i, then D(**p** || **q**) = D(**q** || **p**) = 0. Thus, D_{mre}(**p**,**q**) = max{0,0} = 0.

Proposition 3.5: D_{mre}(**p**,**q**) satisfies the symmetry axiom.
Proof. D_{mre}(**p**,**q**) = max{D(**p** || **q**), D(**q** || **p**)} = max{D(**q** || **p**), D(**p** || **q**)} = D_{mre}(**q**,**p**).

It is not true that D_{mre}(**p**,**q**) satisfies the triangle inequality.

3.6 Discussion of Results

The experimental configuration for a process as subjective as computing similarity between images must be careful arranged to gauge results with other methods and remove any perceptual biases of the experimenter. The two experimental tools used in this work to minimize human subjectivity are random sampling and a large sample space in the form of a large image database. A large database size ensures that a particular class of images, such as medical images or images of people, does not affect the methods being tested. An additional purpose of a large database size is that the scalability of the methods under investigation can be tested.

We tested three similarity measures (L_1 norm for color histograms, the EELN algorithm, and the maximum relative entropy function) with respect to color as a valid visual feature to discriminate between images. It is important to note that we are comparing "apples to apples" in our experiment. Visual features such as shape, edges, and texture are not tested.

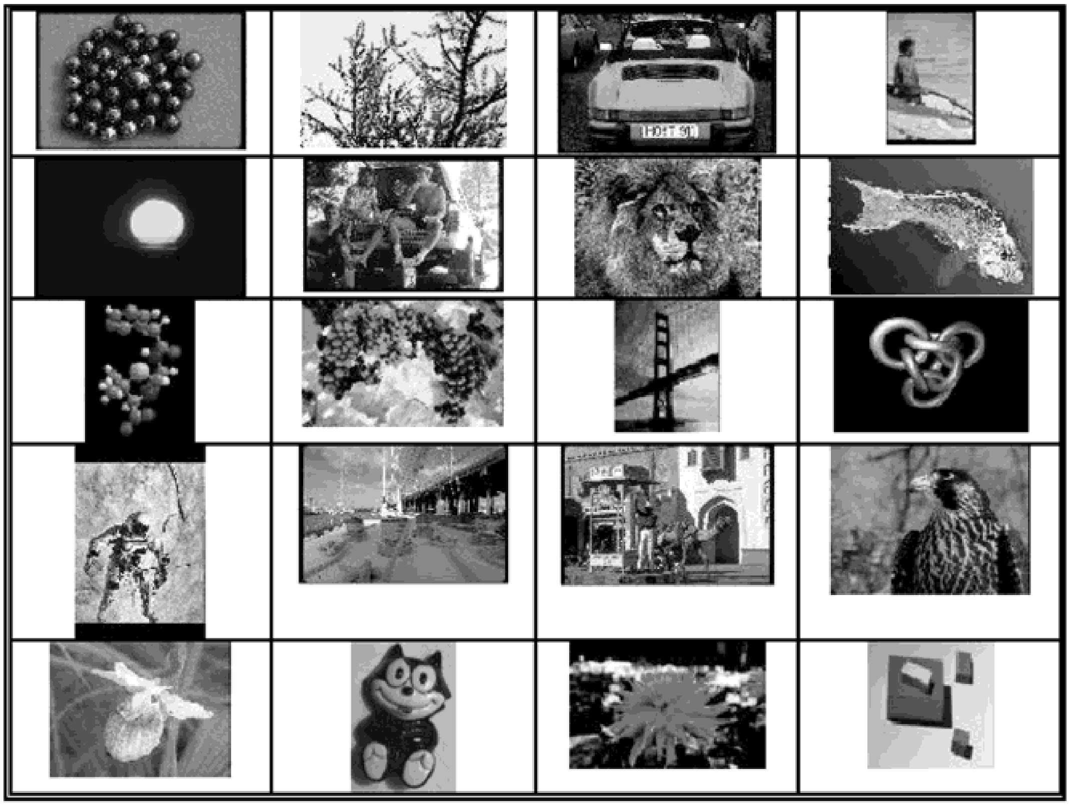

FIG. 7. Query images used to benchmark similarity measures.

Our master database consists of 9,972 unconstrained images of various sizes collected from several sources. Image databases from Stanford, Caltech, INRIA, and IBM were combined with random images collected from the WWW into our master database. Our database contains realistic and synthetic images, such as images of animals, humans in various activities, landscapes, architecture, and space. Additionally, the database is not dominated by any class of images (e.g., medical images).

Our benchmarks are based on 20 query images given in Figure 7. Each query has a unique correct answer manually determined by inspection. The query images were randomly determined prior to the manual determination of the unique correct answer. As seen in Figure 7, these query images represent various situations.

The scalability of the methods presented in this paper was tested across several image databases sizes. The set of 20 queries were tested across 19 image database of sizes 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, and 9972. This resulted in 380 result sets for each similarity measure. Each database was randomly sampled from the same master database (except the database of size 9972, which represented a test on the master database). If the unique correct answer for each of the query images in Figure 7 was not included in a database, then an image was removed by random draw and replaced with the unique correct answer. The results of applying the 20 query images to a database were averaged to present a single measure of the effectiveness of a given method.

3.6.1 Retrieval Performance of Entropy Based Similarity Measures

Let D be an image database and Q be the query image. A query on D is expressed as function $R = f(D,Q)$ where R is (hopefully) a nonzero subset of D. Let $R_{L1} = f_{L1}(D,Q)$, $R_{EELN} = f_{EELN}(D,Q)$, and $R_{MRE} = f_{MRE}(D,Q)$ be result sets of image similar to Q by using the L_1 norm, EELN algorithm, and maximum relative entropy similarity measures, respectively.

As a global property of images, color histograms are susceptible to false positive matches. There are two basic questions to answer concerning the retrieval performance of the similarity measures presented in this paper compared to the L_1 norm retrieval method for color histograms:

1. Are false positives in R_{L1} removed in R_{EELN} and R_{MRE}?
2. Are new false positives introduced in R_{EELN} and R_{MRE} that aren't in R_{L1}?

The first question addresses whether the entropy-based methods increase the *accuracy* of the L_1 norm for color histograms. The determination of accuracy is a two-step process involving both quantitative measurement and subjective judgement. The first step is to ensure that $R_{EELN} \subset$

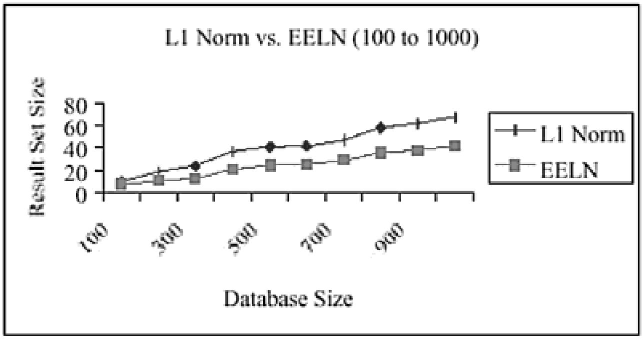

FIG. 8. Result set size comparison for L_1 norm and EELN for database sizes between 100 and 1000.

R_{L1} and $R_{MRE} \subset R_{L1}$. An inspection of the results to determine if the images in R_{L1} but not in either R_{EELN} or R_{MRE} or dissimilar to the query image is the second step. The *reliability* of the entropy-based methods compared to the L_1 norm is addressed by the second question and is checked by the same set relationships between R_{L1}, R_{EELN}, and R_{MRE}.

Our theory predicts that the entropy-based methods improve the accuracy of the L_1 norm and have the characteristic of being more reliable. The following subsections will show that experimentally, these predictions are confirmed.

3.6.2 Retrieval Performance of the EELN Algorithm

The EELN algorithm exhibits good accuracy and reliability. For each of the 20 query images in Figure 7, the intermediate search space produced by the entropy difference formula ranged was reduced to factors between 2.63:1 to 1.63:1.

The result sets for the EELN algorithm were commensurately smaller than the result sets for the L_1 norm, as shown in Figure 8 and Figure 9. The result sets for the L_1 norm and the EELN algorithm were analyzed to determine whether one or more images were contained in R_{EELN} but not in R_{L1}. It was determined that across all query images and database sizes that R_{EELN} was a true subset of R_{L1}. That is, $R_{EELN} \cap R_{L1} = R_{EELN}$.

A qualitative analysis of the set $R_{L1} - R_{EELN}$ was performed by visual inspection. The focus of the judgment made for this analysis was to determine how dissimilar the items in R_{L1} but not in R_{EELN} were to the query image. Each query image was inspected across all database sizes. It was discovered that, in general, the images not in R_{EELN} could be interpreted as dissimilar to the query image. This implies that the use of image entropy captures some characteristic of perceptual similarity.

Figure 10 is an example of the result sets for the L_1 norm and the EELN algorithm for a given query image across a database of size 100. The query image used in both retrieval methods is given in Figure 11. There are two dominating color characteristics present in this image. The first characteristic is the presence of a red/orange background. Dark pixels approximate to black dominate the foreground.

Thus, we should expect to retrieve images with these two themes, albeit with different spatial layouts.

The similarity of the image in Figure 11 to the result sets in Figure 10 was judged to be acceptable compared to the entire database from which the results sets were obtained. The result set for the EELN algorithm a proper subset of the result set for the L_1 norm. The two images in R_{L1} and not in R_{EELN} have properties that suggest a reasonable explanation for exclusion from R_{EELN}. The first image marked “Not Included” has very little red and orange color. However, it does contain a large region of white pixels. The second image marked “Not Included” has reddish-orange pixels as well as dark pixels. However, it also contains a large region of blue pixels corresponding to the sky above the mountain peaks. An inspection of the entropy values and the color histograms shows that the presence of the bright region in the first image and the blue region in the second image causes the entropy difference to be a large enough filter so that these two images are not part of the intermediate result set upon which the L_1 norm is subsequent applied in the EELN algorithm.

This general theme was repeated over several query images and databases sizes. In general, it appears that the entropy difference formula acts as a filter for images with a well-represented count of pixels for some color not contained in the query image. This is the behavior predicted by the theory for image entropy in our companion paper in this issue. We thus conclude that the EELN algorithm offers an improvement in retrieval performance in terms of accuracy and reliability over the sole use of the L_1 norm.

3.6.3. Retrieval Performance of the Maximum Relative Entropy Measure

The retrieval performance of the maximum relative entropy measure was analyzed for the same 20 query images as the EELN algorithm. It was discovered that the maximum relative entropy measure provides some improvement over the L_1 norm. However, in some cases, the result set from using maximum relative entropy measure was not a proper subset of the result set from the L_1 norm. It was determined that R_{L1} and R_{MRE} shared between 85% and 93% of the

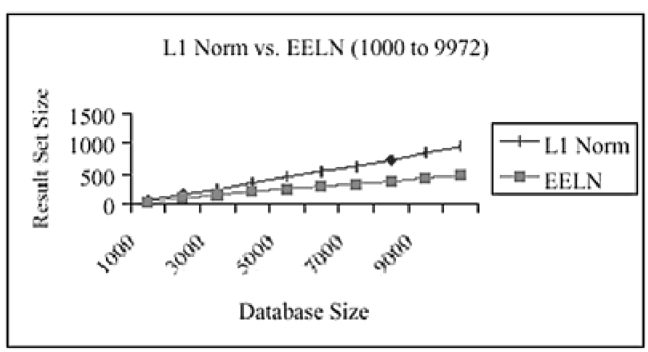

FIG. 9. Result set size comparison for L_1 norm and EELN algorithm for database sizes between 1000 and 9972.

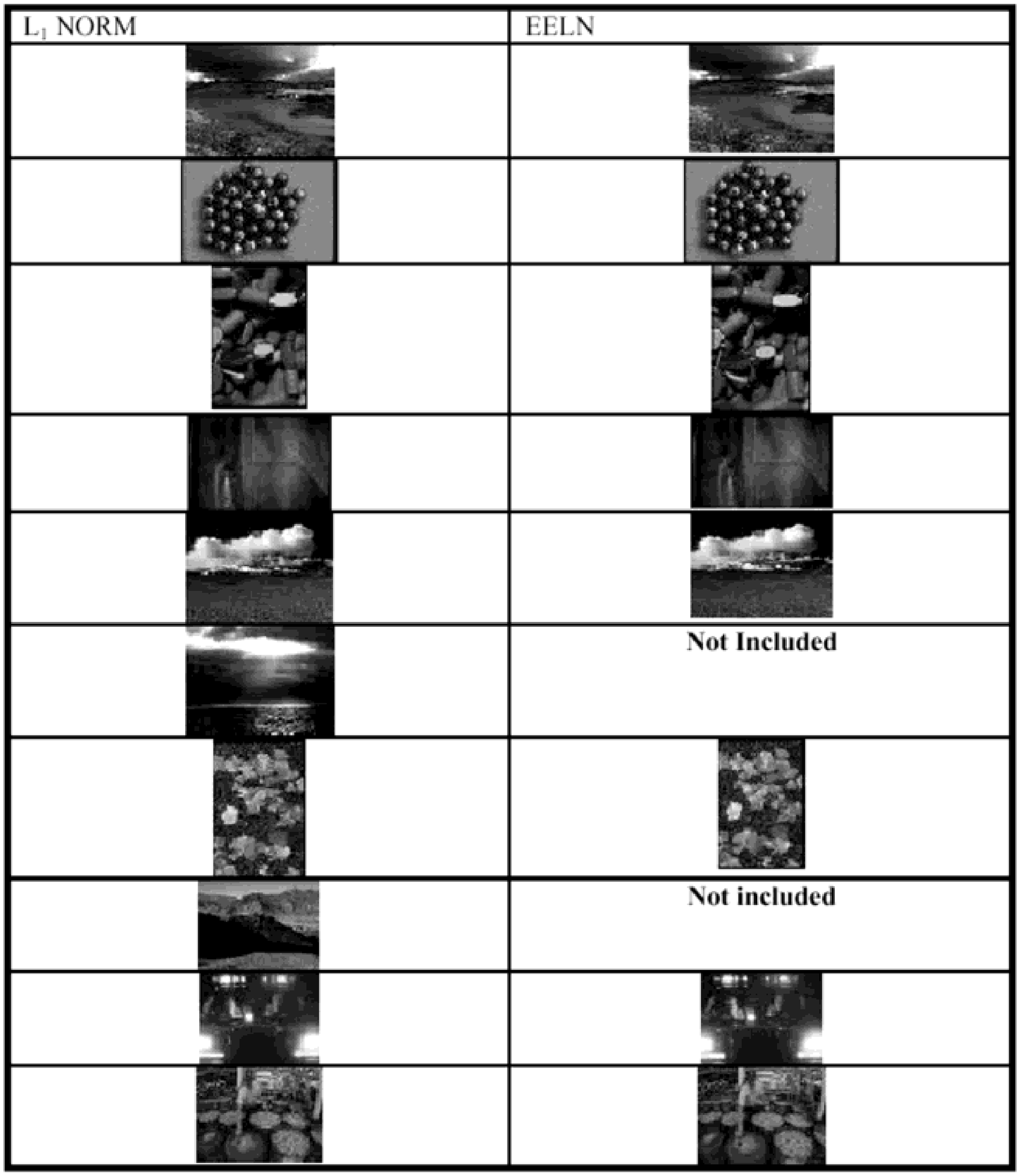

FIG. 10. Result sets from L_1 Norm and EELN to an image database of size 100.

same images. But in every case, the size of R_{MRE} was less than the size of R_{L1}. Figures 12 and 13 depict the behavior of the result set sizes for both methods.

The conclusion drawn from these experiments is that the EELN algorithm provides a viable alternative to the use of the L_1 norm as a similarity measure. The filtering behavior of the entropy difference not only reduces the search space for the L_1 norm part of the algorithm, but it also serves to remove images that have color characteristics that make them dissimilar to the given query image.

The maximum relative entropy function has some utility as a similarity measure for content-based image retrieval. Further studies should focus on the ability of the function to mimic human perceptual abilities. Additionally, studies are required to determine if images included in the result set for the maximum relative entropy measure but not included in the result set for the L_1 norm are similar than to the query image but not captured by the L_1 norm.

3.6.4. Runtime Performance of Entropy Based Similarity Measures

The runtime performance of the EELN algorithm and the maximum relative entropy measures compared to the L_1 norm across the 19 databases is depicted in Figure 14. The run time values were normalized to between 0.0 and 1.0 in order to remove the bias of the software and hardware.

The EELN algorithm executes significantly faster than either the L_1 norm or the maximum relative entropy function. This is the expected behavior since the initial step of the EELN algorithm performs a single subtraction operation for each element in the database. This is in contrast to

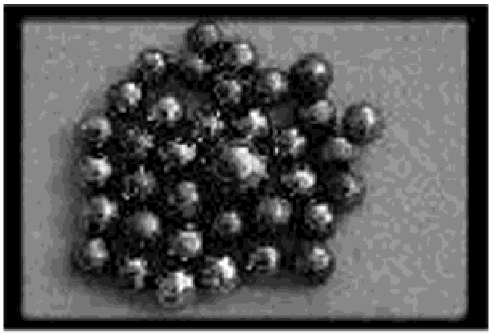

FIG. 11. Query image used for L_1 norm and EELN queries in Fig. 6.

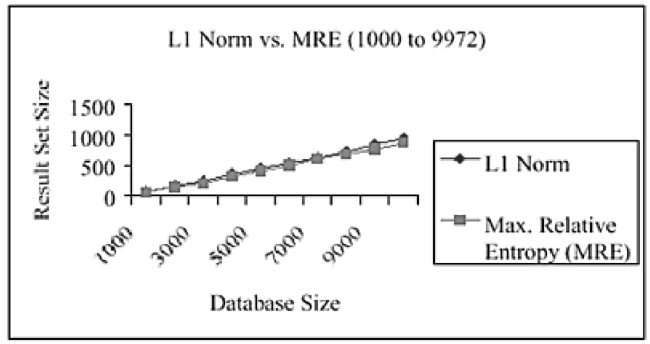

FIG. 13. Result set size comparison for the L_1 norm and MRE measures for image databases of size 1000 to 9972.

applying vector operations for each element in the image database. The next stage of the EELN algorithm then applies the L_1 norm to the intermediate result set of items deemed similar to the query image. Comparing the difference of the entropy values to a threshold makes the determination of similarity. Therefore, we conclude that a reduction in the search space for the L_1 norm similarity measure produces a result faster than searching the entire image database with the L_1 norm similarity measure.

The second behavior to notice is the slower performance of the maximum relative entropy measure compared to the L_1 norm. This is not unexpected since the calculation of the maximum relative entropy measure for each element of the probability density function required two divisions and the computation of the logarithm of the result. The L_1 norm, on the other hand, requires the absolute value of the difference between each element in the probability density function.

The performance of the three similarity measures as a function of the queries is depicted in Figure 15. This graph is interesting to study because it depicts the stability of the three similarity measures across the 20 queries given in Figure 7.

While the L_1 norm and the maximum relative entropy function remains essentially stable as the query image changes, the EELN algorithm exhibits a wider variation in run time performance. The range of normalized run time values for the EELN algorithm as a function of the query image was between 0.19934 and 0.484502. This behavior may be accounted for by software overhead based on the use of the C++ Standard Library container class as a part of the EELN implementation.

4. Summary

We have presented a new indexing algorithm called EELN that combines image entropy with the L_1 norm. The application of the entropy difference formula to the entire database results in a markedly smaller search space of images for the second phase of the algorithm employing the L_1 norm. This results in improved runtime performance as the size of the image database increases. However, its stability across query images may fall within a wider than expected range compared to the L_1 norm and maximum relative entropy functions.

The maximum relative entropy measure was compared to the L_1 norm and EELN similarity measures. Our experimental results across various subsets of a 9972 image database suggest that the maximum relative entropy similarity formula may be an effective measure of similarity between pairs of images, although the run time performance is not as good as the L_1 norm. A primary conclusion of our test show that the maximum relative entropy measure warrants further research attention, particularly to determine if

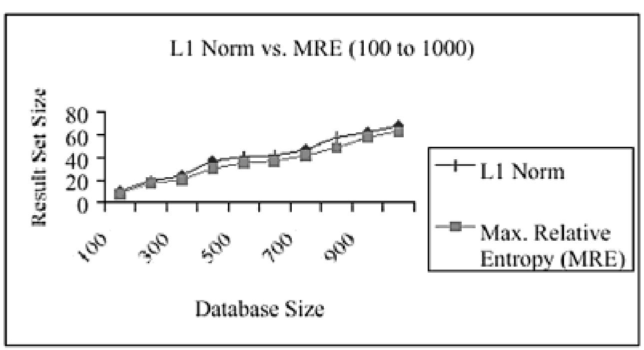

FIG. 12. Result set size comparison for the L_1 norm and MRE measures for image databases of size 100 to 1000.

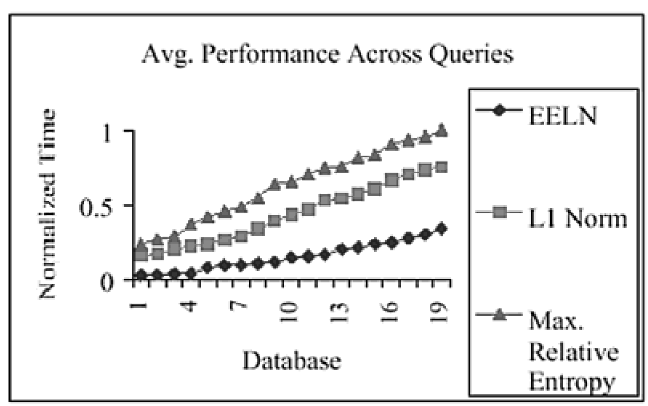

FIG. 14. Runtime performance of the three similarity measures across 19 databases.

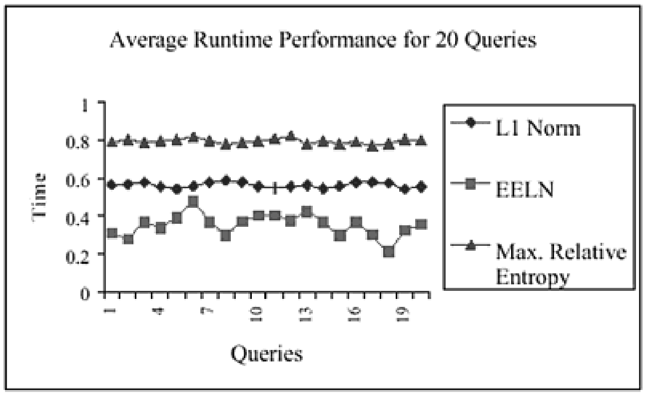

FIG. 15. Runtime performance of the L1 Norm, EELN, and MRE retrieval methods.

the maximum relative entropy function models human perceptual abilities better than the L_1 norm as our qualitative conclusions suggest.

Acknowledgment

This project in part was funded from ONR grant, DOE-ORNL—Lockhead Martin contract (1999–2000).

References

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

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).

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

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

Flickner, M. (1995). Query by image and video: The QBIC system, IEEE Computer, 28(9), 23–32.

Gerbrands, J. (1981). On the relationship between SVD, KLT, and PCA, Pattern Recognition, 14(1–6).

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 TRA 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 (pp. 76–87).

Huang, J. (1998). Color spatial image indexing and applications. Cornell University Ph.D. Dissertation, Dept. of CSC, Ithaca, New York.

Jacobs, C., Finkelstein, A., & Salesin, D. (1995). Fast multiresolution image querying, SIGGRAPH'95 Conference Proceedings (pp. 277–286).

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

Greg Pass, Ramin Zabih, and Justin Miller (1996), Comparing images using color coherence vectors, ACM Conference on Mulitmedia, Boston, MA, Nov. 5–9 (pp. 65–73).

Pentland, A., Picard, R., & 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. Vol. 18, No. 3 (pp. 223–254).

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

Samet, H. (1990). The Design and Analysis of Spatial Data Structures. New York: Addison-Wesley.

Sellis, T., Roussopoulos, N., & Faloutsos, C. (1987). The R^+ tree: A dynamic index for multi-dimensional objects, 13th International Conference on VLDB, Brighton, England (pp. 507–518).

Shannon, C. (1948). A mathematical theory of communication, Bell Systems Technical Journal, 27, 379–423; 623–656.

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

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, Baton Rouge, Louisiana.

Zachary, J. & Iyengar, S. (1999). Content Based Image Retrieval Systems, IEEE ASSET, Dallas, Texas.

Zachary, J., Iyengar, S.S., & Barhen, J. (2001). Content based image retrieval and information theory: A generalized approach. Image Retrieval: A General Approach. Special Topic Issue on Visual Based Retrieval Systems and Web Mining, Journal of the American Society for Information Science and Technology, 52, 841–853.