University of Alberta

Cell Histograms Versus Color Histograms for Image Representation and Retrieval[1]

by

Renato O. Stehling
Mario A. Nascimento
Alexandre X. Falcao

Technical Report TR 01-01
March 2001

DEPARTMENT OF COMPUTING SCIENCE
University of Alberta
Edmonton, Alberta, Canada

[1]An extended abstract version of this paper has been published in the Proc. of the 2nd International Workshop on Multimedia Information Retrieval [28] (held jointly with ACM 2001 Multimedia), p. 171–174, November, 2001 (also available in electronic form at `http://www.acm.org/sigs/sigmm/MM2000/ep/stehling/index.html`).

Cell Histograms Versus Color Histograms for Image Representation and Retrieval

Renato O. Stehling[1]
renato.stehling@dcc.unicamp.br

Mario A. Nascimento[2†]
mn@cs.ualberta.ca

Alexandre X. Falcão[1]
afalcao@dcc.unicamp.br

[1]Institute of Computing, Univ. of Campinas, Brazil
[2]Dept. of Computing Science, Univ. of Alberta, Canada

Abstract

Color is a commonly used feature for realizing content-based image retrieval (CBIR). In this context, this paper presents a new approach for CBIR which is based on well known and widely used color histograms. Previous approaches have used a single global color histogram (GCH) for the whole image, or local color histograms (LCHs) for cells within a grid of fixed size. Our proposed approach is also based on a grid of cells, but unlike the latter it uses a cell histogram for each of the colors actually present in the images, representing how that color is distributed among the image cells – thus the name *Cell/Color Histograms*. Our experiments have shown that the actual number of colors present in images is often low. Thus we are able to achieve performanance comparable to using LCHs within a grid, but with a much smaller space overhead. Furthermore, the proposed approach is very flexible in the sense that the user may easily calibrate the trade-off between space overhead and retrieval effectiveness. In fact, we have been able to outperform GCHs (typically a compact representation) requiring less storage space.

1 Introduction

Image databases are becoming more and more common in several domains. The evolution of techniques for acquisition, transmission and storage of images have allowed the construction of very large databases, containing not only so-called traditional data, but also multimedia data, e.g., images. This has spurred great interest in the area of content based image retrieval (CBIR). This interest is partially motivated by the fact that representing images as textual data (e.g., keywords) is not adequate for image retrieval in large and heterogeneous image databases [4]. This inadequacy stems from three main problems: it is difficult for text to capture the perceptual salience of visual features, text is not well suited for modeling perceptual similarity, and text descriptors reflect the point of view of the annotator, which might be quite different from that of a user actually querying the system. We consider heterogeneous a large set of images that are not strongly related; they belong to potentially different domains, were acquired with different devices, have different size and resolution and are stored in different graphic formats.

A distinguishing aspect of CBIR is that it is based on the availability of a representation (metadata) of the visual content of the images. The complex nature of images, and the complexity of the analysis algorithms, do not allow a direct and objective comparison between two images. As a result, representative and summarized

†Corresponding author. Phone: (780) 492-5678. Fax: (780) 492-1071.

metadata, which may be obtained through automatic or semi-automatic processing, is needed. The type of metadata used to access images has direct impact on the internal organization of the retrieval system, on the way in which the retrieval is performed and on its effectiveness. CBIR is complex since different types of metadata may be associated with images. Such different types of metadata include [13]:

- Content-independent metadata - data which is not directly concerned with image content, but in some way related to it. Examples of such data are: the format, the authors name, date, location and ownership.
- Content-dependent metadata - data which refers to perceptual low/intermediate-level features, like color, texture, shape, spatial relationship and combinations thereof.
- Content-descriptive data - data which refers to content semantics. It is concerned with relationships of image entities with real-world entities or emotions, impressions and meaning associated with visual signs.

In general, the goal of CBIR is to retrieve images similar to an image/sketch provided by the user. This contrasts with traditional databases where the goal is an exact-match of a set of attributes. Matching is a binary partition, being intrinsically committed to deciding whether or not the object observed corresponds to a model. Similarity-based retrieval is, instead, the task of reordering database images according to their measured similarity to a query example. It is therefore concerned with ranking rather than classification [4].

There are two chief methods to query images given an image as an example [26]: finding the most similar images, or finding all images within a certain degree of similarity relative to the reference image. In order to realize image-based queries, we believe that one needs: (1) a domain-independent characterization of the visual content of the images, (2) image processing techniques to automatically extract such visual characteristics, (3) a compact yet representative abstraction for these characteristics, (4) a similarity metric to effectively compare images and (5) indexing techniques to efficiently access relevant images in the database. The first two requirements are related to image metadata extraction and representation. The last three are related to CBIR efficiency and effectiveness. A great challenge in this area is to find the best compromise between these conflicting requirements. In large and heterogeneous image databases, the above issues are even more critical. The universe of potential metadata is restricted by the constraint of efficient metadata extraction, as well as analysis and indexing. In other words, the metadata should be representative, simple and compact. The similarity metrics should also be simple from the computational complexity viewpoint. Finally, any semi-automatic image processing should be avoided as much as possible and not relied upon as they are not scalable. Thus, in order to tackle the problem of efficient retrieval in large heterogeneous image databases, one must focus on techniques which are automatic, efficient and scalable.

This paper deals with the first four requirements above, and content-dependent metadata, colors in particular. Our goal is to propose and evaluate a new technique for image abstraction and retrieval based on color distribution and location. Our basic motivation is the fact that global color histograms are not able to adequately capture image content, in particular the location (spatial component) of the colors. Alternative approaches, based on a set of local color histograms pose a large overhead in terms of metadata representation/storage. Our main contribution is also a histogram-based representation, though potentially much more compact, which is able to capture the spatial distribution of an image's colors. This paper improves on [28] in the following ways. A more thorough investigation of retrieval effectiveness, including a more robust effectivenes measure as well as traditional precision vs recall charts. A number of additional experiments were performed in order to further assess the robustness of the proposed contribution, which is also compared to another well known CBIR approach (namely, CCVs [18]) in addition to classical approaches.

The remainder of this paper is organized as follows. Section 2 discusses some related works relative to color features and to their spatial distribution. Section 3 presents a new metadata representation for CBIR, named Cell/Color Histogram, which is more robust and flexible than those discussed in Section 2. Section 4 presents a similarity metric, based on L1 distance, and which is common to all approaches that will be compared in the experiments described in this paper. Section 5 evaluates retrieval effectiveness using Cell/Color Histograms, comparing it to three other techniques. We also show that the user may easily tune it in order to calibrate the trade-off between space overhead and retrieval effectiveness. Finally, Section 6 presents our conclusions and directions for future work.

2 Related Work

With respect to the use of colors as the visual feature of interest when performing CBIR, there are two main issues one should be concerned with. The first issue is the underlying model, which contains two other issues in itself: color space and a suitable representation. The second issue is related to the spatial layout of the color distribution. In this section we review previous work related to both of these issues.

2.1 Color features

Color is a visual feature which is immediately perceived when looking at an image. The most simple and well known approach to encode the low-level color information of an image is the global color histogram - GCH [10, 30]. It is obtained by discretizing an image's colors and counting how many pixels belong to each color. Histograms, by themselves, do not include spatial or shape information and as a result, images with very different layout can have similar representations. Hence, retrieval of images based on color histograms are prone to yield a large number of false hits, specially when used with a large database of heterogeneous images.

The QBIC system [3] uses a partition-based approach to extract and represent color features. For each cell relative to a rectangular grid superimposed on the image, the average Munsell color and the five most frequently occurring colors and their frequencies (i.e., a partial histogram) are computed. In Dimai's approach [8], the features that represent the color distribution in the L*a*b* color-space are the average color and the covariance matrix of the color channels. Androutsos et al [1] used color segmentation, in the HSV color-space, to extract regions of prominent color. Each region is described by its average color vector. Sciascio et al [21] also uses the average color to describe the cell content in a 4x4 grid superimposed on the image. In the Color-WISE approach [24], the dominant hue and saturation values in the HSI color space were used. These values are determined from different parts of an image through a process of block-based histogram building and peak detection.

Stricker and Orengo [29] propose two approaches to color representation. The first one, cumulative color histograms, is a more robust version of the commonly used color histogram. The cumulative histograms are always completely dense vectors, even if only a few colors of the discrete color space appear in each image. The robustness of this technique allows one to work with very coarsely quantized color spaces. The pertinence of cumulative histograms for CBIR lies on the property that the color similarity between two nearby histogram bins should be bigger than the color similarity between two further separated histogram bins. Zhang et al [31] proposed a variation of cumulative histograms. The modified procedure applies the cumulative histogram technique to each sub-range of the histogram. This technique was called Local Cumulative Histogram. The second approach presented by Stricker and Orengo [29] instead of representing the complete color distributions, represents only their dominant features, via the first three moments of each color channel in the HSV color space.

Appas et al [2] decompose an image into five regions and also represents the color feature of each region by the first three moments of each color channel in the HSV color space. Two differences distinguish this work from previous work: (1) the third moment is computed from the second moment (instead of the first one) and (2) the three moment values are combined into a single descriptor. Liang and Kuo [15] proposed an integrated wavelet coding system for CBIR. All the features (color, texture and shape) are based on wavelet coefficients and their energy distribution among sub-bands, across quantization layers and in the space. These features were extracted during the successive approximation quantization (SAQ) stage of the wavelet compression process. The obtained SAQ histograms are represented by their first three central moments (mean, variance and skewness), as proposed by Stricker and Orengo [29].

Pass et al [18] describe yet another technique based on color histograms called Color-Coherence Vectors (CCVs). The technique classifies each pixel in a color bucket as either coherent or incoherent, depending upon whether the pixel is part of a larger and connected similarly-colored region. The argument is that the comparison of coherent and incoherent feature vectors between two images allows for a finer distinction of similarity than when using color histograms The authors compare their experimental results with other techniques and show their technique to yield a significant improvement in retrieval performance.

2.2 Spatial Distribution of Colors

In order to take into account the spatial location of colors and their distributions, and consequently improve retrieval, several approaches have been proposed. In what follows we review the results of related research.

Spatial grids partition the images from space into equally sized blocks, where each block corresponds to a spatial portion of the image. The QBIC system [3] decompose an image using two approaches: partition-based and region-based. The partition-based approach is similar to the method described in [19]. The images are divided into a 6x8 or a 9x12 grid of cells. The region-based approach uses an approximate segmentation of each image into a hierarchical set of colored rectangles. Sciascio et al [21] also uses a 4x4 grid to partition the images. Androutsos et al [1] uses color segmentation, in the HSV color-space, to extract regions of prominent color. The approach presented by Appas et al [2] decomposes an image into five regions (the center and the four corners).

The approach proposed by Guibas et al [12], decomposes images using a fixed quadtree [25]. In Leung and Ng's approach [14], each image has a 4-level multi-resolution representation. At the first level, the image is represented by a single color histogram. In the second level, the image is divided into four non-overlapping blocks, each one represented by one color histogram. In the next levels, each block is successively divided into four new blocks. The idea is the same as the in quadtree-based approaches. The work of Sebe et al [23] decomposes images into three levels. The first level is the whole image itself. The second level is a 3x3 grid and the third level is a 5x5 grid. This decomposition results in 34 regions plus the image (level 1). The regions in this approach are of different sizes (according to their level) and overlap in different levels. The approach of Malki et al [17] is similar to the previous approach: they use a quadtree of three levels to decompose an image. The idea of hierarchical partitioning of images has also been proposed by Lin [16], where a three-dimensional hash table was used for efficient indexing of images. Color-WISE approach [24] uses a fixed image partitioning scheme which allows overlapping blocks.

Dimai [8] proposed an approach which combines a global vector describing the color features of the whole image with a set of inter-hierarchical distances, which are based on a fixed partition of the image (nine non-overlapping fixed-size regions). The inter-hierarchical distances are the difference measured between feature vectors of a region and its sub-regions. Thus, only the difference between the global vector and the vector of one sub-region is stored. Chen and Wong [5] proposed an augmented color histogram that captures the spatial distribution of pixels in addition to the color distribution. The spatial information is incorporated by computing features from the spatial distance between pixels belonging to the same intensity color. The mean, variance and entropy of the distances are computed to form an Augmented Image Histogram. The augmented histogram therefore captures the spatial distribution of pixels with the same color relative to each other, instead of containing information about absolute spatial location of the pixels. The SFGraph [27] is an approach that partitions the images symmetrically in space and frequency. The SFGraph view elements correspond to wavelet sub-bands of portions of the images. The SFGraph embeds many image representations, such as the multiresolution pyramid, wavelet, tiled-wavelet, wavelet packet, spatial quadtrees and Flashpix. Flashpix [7] is yet another approach that combines the wavelet and spatial grids by partitioning the images into equal-volume tiles, at different resolutions.

3 Cell/Color Histograms

As described in the previous section, the most widely used technique to spatially locate visual features partitions images into fixed-size cells and then extracts visual features for each cell individually. Our main contribution in this paper is a simple and compact, yet effective, partition-based technique. To the best of our knowledge, it is original in the way that visual features are encoded. Our motivation was to reduce the space overhead of partition-based approaches taking advantage of the fact that only a relatively low number of distinct values of a particular visual feature is present in most images. In particular, we use color features to investigate the idea. However, it is possible to encode any other visual feature of an image with the same idea. We will refer to our proposed representation as *Cell/Color Histograms*[1] (CCH) approach.

Using 20,000 heterogeneous images in JPEG format, and the RGB color-space uniformly quantized in 64 colors, we obtained the data shown in Figure 1 – how much of an image is covered by a given number of colors. In the average, there were only 28,71 colors per image from a total of 64 colors. Moreover, 90% of the image content corresponds to only 9 colors. These values show that, at least 55% of a color histogram has bins with zero value and, using only 14% of such bins one is able to describe 90% of the color content of the image. In addition, consider the case where one superimposes a grid of cells over the image and extract (local) color histograms for each of such cells[2] in order to extract some notion of spatial location of the color distribution. Clearly, the amount of "useless" information (i.e., null bins) in this case grows proportionally to the number of cells in the grid, though better retrieval is likely to be achieved.

Figure 1: Color contribution for image content.

We aim at exploring the fact that colors not present in an image need not be represented. Note that Global Color Histograms (GCHs) do not allow this type of optimization directly. Thus, we propose the use of *Cell/Color Histograms* (CCHs). Using CCHs, the image abstraction is more compact than when using the Grid approach, yet just as representative. Consider an image partitioned into $I \times J$ non-overlapping cells and the RGB color-space uniformly quantized in C colors. A *cell histogram* for a given color c, $0 \leq c < C$, is a set of $I \times J$ bins (one for each the image's cells), where the bin values are described by the function $p(cell_k) = n_k/n$. In this function, $cell_k$ is the k^{th} cell of the image ($0 \leq k < I \times J$), n_k is the number of pixels in the $cell_k$ with color c and n is the total number of pixels in the image. An image composed by m colors is thus described by m cell histograms, each one describing the spatial distribution of one color. We call the set of all cell histograms of

[1]In [28] this representation was called *Color-Shape Histogram*. This name was changed to avoid the misconception that our technique would actually be able to perform retrieval based on color as well as shape, which is not the case.

[2]From here on, such an approach is referred to as the *Grid* approach.

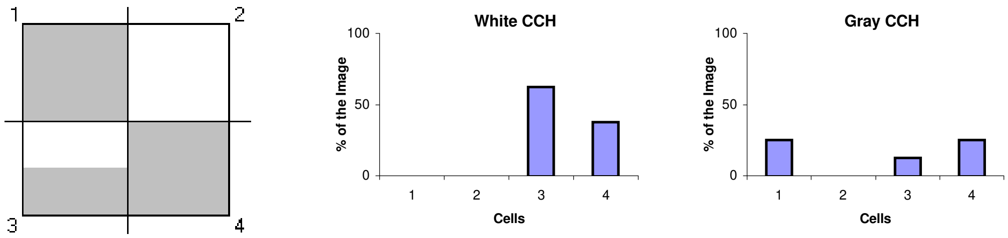

Figure 2: An image partitioned using a 2x2 grid and its CCH

an image, its *Cell/Color Histogram*. It is important to note that, in this type of decomposition, if a color is not present in an image, there is no cell histogram associated to it to be stored. Compare this to the case of using the Grid-approach, where one would need necessarily store $I \times J$ histograms, each with C bins, regardless of how many colors are actually present in the image. Both CCH and the Grid approach are able to store the same amount of information, but the amount of storage required by the former is likely much smaller than the amount required by the latter. Figure 2 illustrates this discussion, showing a two colors image with a 2x2 grid superimposed and its respective CCH. Note that the image's CCH would be the the same regardless of in many colors the original image space was quantized.

4 Similarity metric

In this section, we describe the similarity metric that will be used in all approaches compared in the experiments that follow. The chosen metric is based on the L_1 distance metric (Equation 1), where $h_q[i][j]$ and $h_d[i][j]$ represent the j^{th} bin of the i^{th} histogram used to represent the query image (h_q) and the database image (h_d), respectively. We assume that the histogram bins are normalized with respect to the image size, i.e., number of pixels.

$$D(h_q[i], h_d[i]) = \sum_{j=1}^{m} |h_q[i][j] - h_d[i][j]| \tag{1}$$

Recall that in the case of a GCH, there is only one histogram to be considered. In traditional partition-based approaches, there is a fixed number of local color histograms. In the CCH approach, there is a variable number of histograms per image.

To normalize the result obtained with the L_1 distance, we divide it by the sum of the areas (number of pixels) of the regions described by each histogram ($a_q[i] + a_d[i]$). These areas are also normalized with respect to the images' sizes. The normalized metric, D_n, is shown in Equation 2.

$$D_n(h_q[i], h_d[i]) = \frac{D(h_q[i], h_d[i])}{a_q[i] + a_d[i]} \tag{2}$$

Note that if we use GCHs, the denominator in Equation 2 is always 2. However, using CCHs, the areas being compared are, in general, smaller than the whole image, in fact, they are equal to the percentage of each image that the color being compared represents. This case requires one to use the sum of the areas explicitly to normalize the L_1 result (notice that the sum is not known *a priori*). So far, D_n measures the normalized distance between two histograms. The similarity between two histograms is the complement of the distance D_n. The similarity S between two images (Equation 3) is the weighted sum of the similarity between the histograms present in the images:

$$S(h_q, h_d) = \sum_{i=1}^{n} w[i] \times (1 - D_n(h_q[i], h_d[i]) \tag{3}$$

The goal of the weight values is to normalize the similarity between two images and describe the importance of each compared histogram. We chose $w[i] = min(a_q[i], a_d[i])$. Clearly, the sum of the $w[i]$s is at most 1.

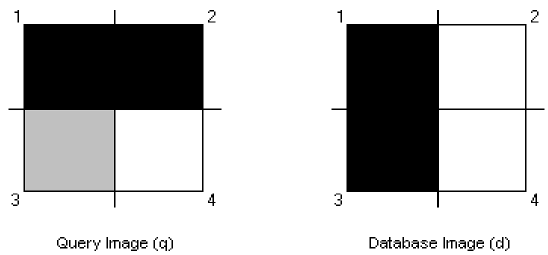

Figure 3: Sample images with a superimposed 2x2 grid

The two images in Figure 3 will be used to exemplify the application of the similarity metric in three histogram-based approaches for image representation: GCH, Grid, CCH. For simplicity, we divide the images into 4 cells (2x2 grid) in order to spatially locate colors. The cells are compared from top to bottom, left to right, and the color space has only three colors: black, gray and white, in this order. In Figure 3, q is the query image, and d is the database image to be compared against q. Figure 4 shows the metadata (histograms) obtained from the query image by each approach. The top row shows the single GCH. The second row depicts the set of four LCHs obtained in the Grid-approach, one for each cell. The bottom of the figure shows the three cell histogram, one for each of the present colors in q. Similarly, Figure 5 shows the same histograms for the database image (d).

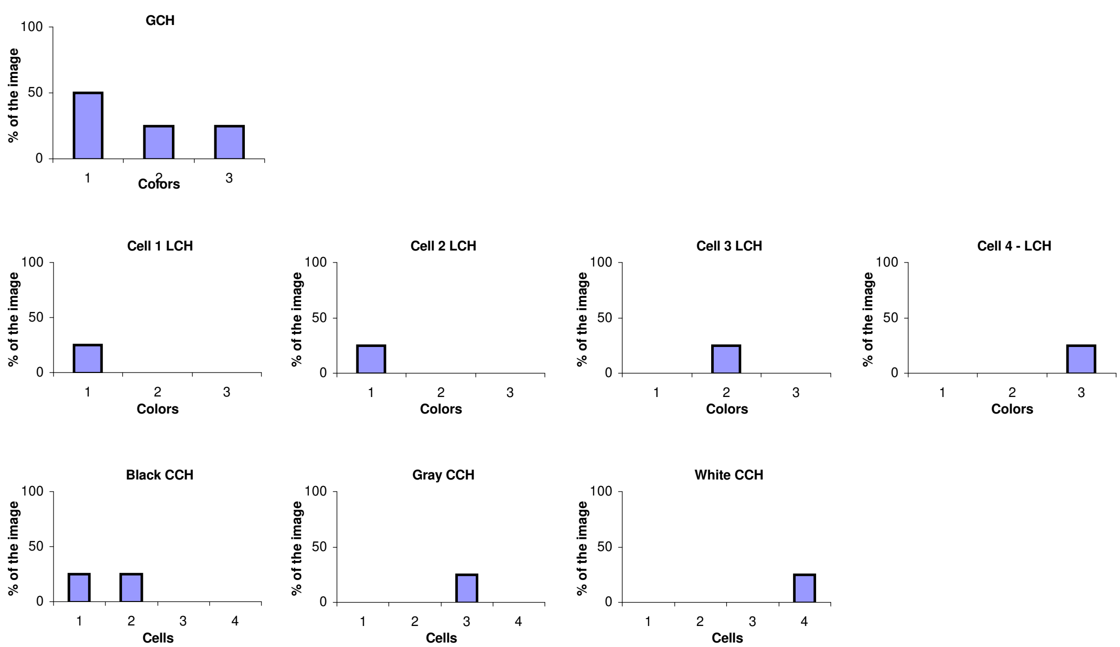

Figure 4: Histograms of the query image (q in Figure 3) in each approach

.

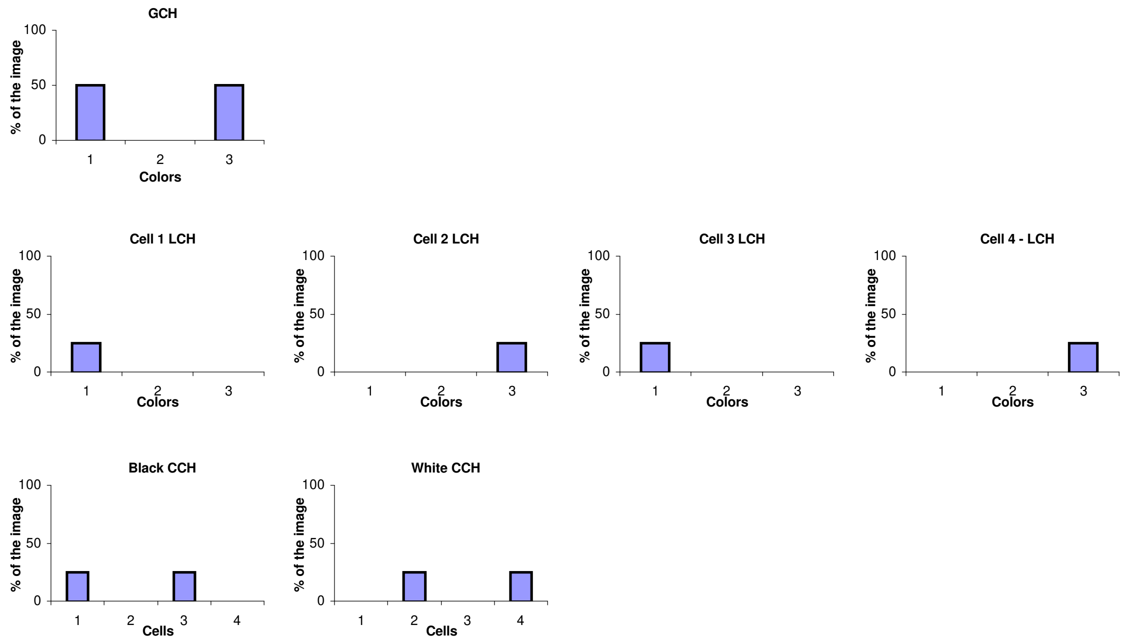

Figure 5: Histograms of the database image (d in Figure 3) in each approach.

The GCH for q could be represented as $h_q = [0.5, 0.25, 0.25]$ meaning that it has 50% of black, 25% of gray and 25% of white pixels, respectively. Similarly, $h_d = [0.5, 0.0, 0.5]$. Using Equation 3 we have:

$$S_{GCH}(q,d) = 1 \times (1 - \frac{|0.5 - 0.5| + |0.25 - 0.0| + |0.25 - 0.5|}{2}) = 0.75$$

As for the Grid-approach, the normalized distance for the first cell is $D_n^1 = 0$, because both cells have only black pixels. For the other three cells, the distances are $D_n^2 = 1, D_n^3 = 1$ and $D_n^4 = 0$, respectively. In addition, the weights $w[i]$ are all the same, because all cells have the same relative area. Thus, we have:

$$S_{Grid}(q,d) = 0.25 \times (4 - (D_n^1 + D_n^2 + D_n^3 + D_n^4)) = 0.5$$

Finally, using the proposed approach, we compare the three CCHs (for black, gray and white) using Equation 2. We have 25% of black pixels in cells 1 and 2 of q, and in cells 1 and 3 of d (recall that the quantity of pixels in a cell is normalized with respect to the image size), hence:

$$D_n^{black} = \frac{|0.25 - 0.25| + |0.25 - 0| + |0 - 0.25| + |0 - 0|}{0.5 + 0.5} = 0.5$$

Likewise, we obtain $D_n^{gray} = 1$, and $D_n^{white} = 0.33$.

Lastly, the normalized CCHs distances are complemented and weighted according to Equation 3 (notice that, unlike in the GCH and Grid approaches, the weights $w[i]$ are now variable, depending on the areas occupied by each color). Therefore, the similarity between the two images according to their CCHs is:

$$S_{CCH}(q,d) = 0.5 \times (1 - 0.5) + 0 \times (1 - 1) + 0.25 \times (1 - 0.33) = 0.42$$

It is interesting to note that $S_{GCH} > S_{Grid} \sim S_{CCH}$, as one could expect. There are several alternative distances that may be used to compare color histograms in an image-based query. We believe that the relative results should not depend on the particular metric used. That is to say that we expect any possible bias, if existent, to affect all approaches equally. We plan to investigate alternative similarity metrics in our future research.

5 Evaluation of Retrieval Effectiveness

The dataset used in our experiments was a set of 20,000 JPEG images from a stock CD by Corel Corp. The images' content are heterogeneous but, at the same time, there is a large number of small subsets where the images are similar, allowing one to distinguish them from the others without ambiguity. Some of these subsets were used to evaluate query results. Out of the 20,000 images in the data set, we chose 15 images to be used as query images. Those images not only had a similar color distribution but also had the same semantics. For instance, there is a set of several distinct Halloween pumpkins, a set of Bonzai trees, and so forth. The answer sets for these query images were also built *a priori* in order to facilitate our evaluation. We call these sets the "required result set" (RRset) for each query image[3]. The average number of images in such RRsets was approximately 11. The database creation and query times were not measured. At this point, our focus is only the quality evaluation of the image retrieval. The efficiency of the retrieval is an aspect which is tightly related to indexing structures/techniques, which is subject of further research.

In order to evaluate retrieval performance in a similarity-based scenario such as the one we are concerned with, we need a performance measure that embodies the position in which target items appear in the retrieval sequence (ordered by some similarity measure). We compared the retrieval effectiveness of each of the investigated techniques using two distinct measures: (1) Precision vs. Recall and (2) Normalized Average Rank ($NavgR'$) – a more robust derivative of the measure originally presented in [9]. The effectiveness of a retrieval system is a measure related to the users satisfaction with the system output. Since it is generally difficult for users to distinguish among a large number of degrees of relevance, we choose to work with binary values: either a image is relevant or it is not. Precision and Recall [20] were chosen because they are the best known and widely used measure in information retrieval systems. Although they are not the most adequate measure in the context of ranked output [9, 20, 22], they has been widely used also in comparison of CBIR systems.

The similarity-based retrieval is concerned with ranking rather than classification [4]. In this context, a effectiveness measure should consider the total ranking imposed on all the collection, independent of a retrieval threshold. The Normalized Average Rank (NavgR) [9] is a single-value measure that satisfies this restriction. It is a variation of the Normalized Recall measure [20] which was first used in the QBIC project [9] and then, in some other CBIR approaches [22, 28]. NavgR measures how close to the top of the ranked result, the set of relevant items appears. This value is normalized considering an ideal retrieval in which the relevant images appear ahead of any non-relevant one in the ranked result. The relative ordering of the RRset elements is irrelevant because we are using a binary judgment of relevance. The measure used in the QBIC project was defined as $NavgR = A/I$, where A and I are the actual average rank and the ideal average rank I of the RRset, respectively. In this paper, we propose a slightly different definition of the NavgR measure, inverting the original ratio (Equation 4):

$$NavgR' = \frac{I}{A} = \frac{\sum_{i=1}^{m}(i-1)}{\sum_{i=1}^{m} rank(i)} \tag{4}$$

In this equation, $m = |RRSet|$ and $rank(i)$ is a function that returns the rank of the i^{th} relevant image. The rank values varies in the range $[0, n-1]$, where $n = |DB|$. With this alternative definition, the $NavgR'$ value ranges from 0 (worst case) to 1 (best case). as occur in precision and recall measures. A problem with the $NavgR'$ measure, is that, the value obtained for a single query is still very sensitive to the rank of the last relevant document retrieved. Our future research will include an investigation of possible solutions to this problem. In this paper, we deal with this problem considering only a predefined portion of the RRSets for the purpose of

[3]The sets of image queries and respective RRsets can be seen at `http://www.cs.ualberta.ca/~mn/CBIRdataset`.

$NavgR'$ computation. We consider only the ranks of the first 80% portion of an RRSet. The last 20% of the RSSets (with the highest numerical ranks) are discarded, hence any possible mislead judgment of relevance (that could affect the $NavgR'$ value) will not affect the measurements. Furthermore, it is reasonable to assume that most users will inspect only the first few images returned by the CBIR system.

We compared the use of Cell/Color Histograms to three other approaches: (1) Global Color Histogram – GCH, (2) Color-Coherence Vector – CCV, and (3) a Grid-based approach (also referred to as Grid for short). A GCH [30] is described by a set of bins (one for each color in an image), each one with a height given by the function $p(c_k) = n_k/n$, where c_k is the k^{th} color (bin) in quantized color space, n_k is the number of pixels whose color is c_k and n is the total number of pixels in the image. Using CCVs [18], each color k in a quantized color-space, has two values associated to it: the number of coherent pixels vs. the number of incoherent pixels with this color. A pixel is coherent if it is part of a large similarly-colored connected region. Grid is a partition-based approach which decompose an image in a set of fixed size cells, accordingly with a grid superimposed on the image. For each cell a local color histogram (LCH) is extracted. A LCH is equivalent to a global color histogram; the only difference is that n_k is the number of pixels whose color is c_k into the cell being represented, not in the whole image. The normalization is also relative to the whole image (n).

The four compared approaches used the RGB color-space, uniformly quantized in 64 colors. For the Grid and CCH approaches, each image was decomposed into a fixed 8x8 grid, resulting in 64 cells per image. Although one may dispute the suitability of such parameters, we contend that they affect the three approaches in the same way, because all three approaches describe only color features, are histogram-based, and use the same similarity metric. In our judgment, this fact allows a fair comparative analysis.

At this point, an analysis of all approaches with respect to the required storage overhead is in order. Using GCHs, each image was represented by one color histogram with 64 bins (one for each of the quantized colors). In the CVV approach, each image was represented by one color histogram with 128 bins (two bins for each of the quantized colors). Using Grid, each image was described by 64 color histograms (each with 64 bins), one for each of the 8x8 grid cells. When using CCH, each image was described by a variable number of cell histograms, depending on the number of colors present in the images. Each such cell histogram had 64 bins, one for each of the 8x8 image cells. GCH is the most compact image abstraction. Using the proposed CCHs, the maximum number of cell histograms is 64. However, in the average, for the 20,000 database images less than 29 histograms were needed per image (see Figure 1). Therefore, CCHs required in average 55% less space than the LCHs in the Grid-based approach. Note that this means potential savings of over 50% in space when the image abstraction is stored for further processing, e.g., indexing, which is a desirable feature of the proposed approach.

5.1 Experiment I - Comparison with traditional approaches

As can be seen, from Table 1, CVVs yield a $NavgR'$ 37% higher than GCHs, whereas using Grid yields a $NavgR'$ 3.5 times higher (thus better) than using GCHs. This demonstrates clearly that the use of some notion of spatial location of colors improves significantly the retrieval effectiviness. Note the the $NavgR'$ obtained by using CCHs is slightly better (7%) than if we were using the Grid Approach. The CVV approach uses much more information than GCH to represent an image and the Grid approach uses even more information. Thus, the improvement in $NavgR'$ can be thought of as a result of the additional information, i.e., space overhead. Interestingly, CCHs uses less space overhead (i.e., number of histogram bins) than Grid. It combines the global information of GCH with the local information of Grid in an elegant and more compact way. These two types of information yield in a better $NavgR'$ for the Color-Shape approach. The Precision vs. recall curves of the first experiment can be viewed in Figure 6. As one can see these curves are consistent with the values obtained with the $NavgR'$ measure (Table 1).

Table 1: Average $NavgR'$ values for the first experiment.

Approach	**Avg.** $NavgR'$	**Advantage wrt GCH**
GCH	0.08	—
CCV	0.11	37,5%
Grid	0.28	250%
CCH	0.30	275%

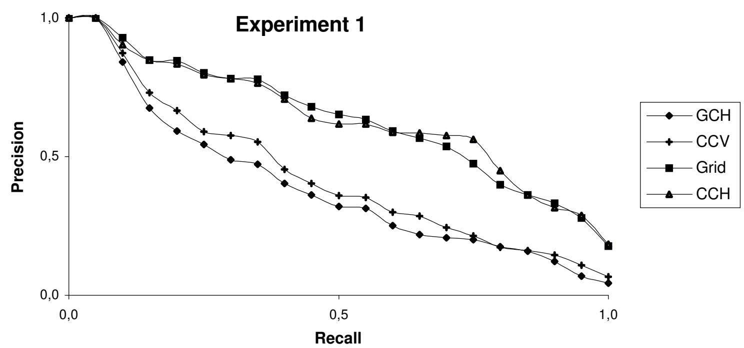

Figure 6: Precision vs. Recall curves for Experiment I

5.2 Experiment II - Reducing the number of cells

The next two experiments explore the flexibility of using CCHs. As argued earlier one of our main goals is to reduce the space overhead necessary to represent the spatial color distribution. We explore this issue here, reducing the number of cells in which we partition the images.

Table 2 shows the effect (in terms of CCH's $NavgR'$) of reducing the number of cells in which an image is partitioned. We compared the use of 8x8, 6x6, 4x4, 3x3 and 2x2 grids superimposed on the images. Obviously, the coarser the grid the smaller the space overhead, but the trade-off is evident in terms of retrieval effectiviness. The results show that, as the number of cells decreases, the $NavgR'$, thus the retrieval effectiveness, also decreases due to the loss of spatial information. There is less information to distinguish images, increasing the number of false-hits, and thus decreasing the retrieval effectiveness. The Precision vs. Recall curves showed in Figure 7 confirms these results, but also shows that for smaller values of recall the precision does not vary as much, which is an interesting fact, specially for applications where one is mostly interested in the first few images returned.

Table 2: Effects of the reduction of the grid size in CCH's $NavgR'$.

Grid Size	**Avg.** $NavgR'$	**Loss wrt 8x8 Grid**
8x8	0.30	—
6x6	0.28	7%
4x4	0.23	23%
3x3	0.20	33%
2x2	0.12	60%

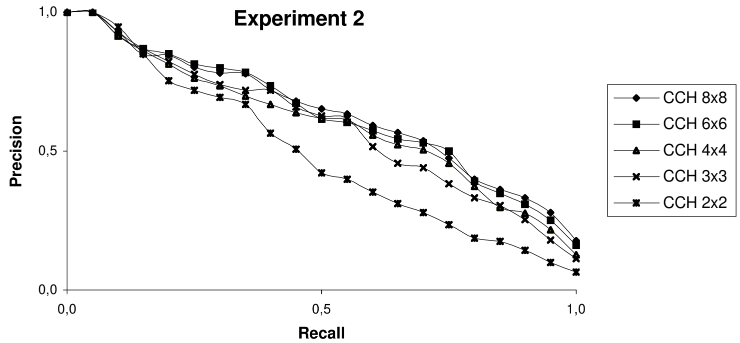

Figure 7: Precision vs. Recall curves for Experiment II

5.3 Experiment III - Partial representation of images

As can be seen in Figure 1, on average only about 29 colors (out of 64 possible) were present in our 20,000 heterogeneous database images. The same Figure also shows that only a small number of colors is responsible for the majority of the image content (this is in line with the reasoning used in QBIC). The experiments summarized in Table 3 explore this fact, when we aim at optimizing CCH's space overhead – The smaller the number of represented colors, the smaller the number of cell histograms needed to describe an image. Still according to Figure 1, on average 100% of the image content corresponds approximately to 29 colors, 95% corresponds to 12 colors, 90% corresponds to 9 colors, 80% corresponds to 6 colors and 70% corresponds to 4 colors. We use these values to compute the effect of comparing a partial content of the images in CCH's retrieval effectiveness ($NavgR'$). The results show that, representing only 80% of the images content results in an decrease of 33% in the CCH's $NavgR'$. The range between 90% and 100% results in smaller decreases of the $NavgR'$ in comparison with the space overhead reduction. For instance, for 90% of the image content we have on average 9 colors, resulting in 9 cell histograms. This number is 69% smaller than the number of cell histograms required to represent 100% of the image content (29). The $NavgR'$ decreases 13%, as expected, from 0.30 to 0.26, but the gain is storage space seems to offset well the loss in retrieval effectiveness. The corresponding Precision vs. Recall curves can be viewed in Figure 8.

Table 3: Use of partial content of images for CCHs. Images are partitioned by a 8x8 grid

% of the image	**Avg. Number of Colors**	**Avg. $NavgR'$**	**Loss wrt 100% coverage**
100%	29	0.30	—
95%	12	0.28	7%
90%	9	0.26	13%
80%	6	0.20	33%
70%	4	0.11	63%

5.4 Experiment IV - Fine Tuning CCHs

In the last batch of experiments, we compare the GCH, Grid and CCV approaches with some variations of the CCH approach, using different number of cells and representing different percentages of the images. Our goal is to demonstrate how one can fine tune the CCH approach, analysing the various opportunities for trade-off it offers. This also serves to summarize the previous experiments, comparing simultaneously the space overhead

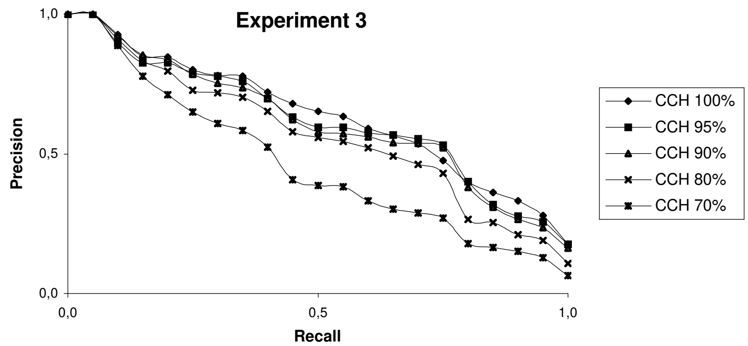

Figure 8: Precision vs. Recall curves for Experiment III

and the retrieval effectiveness of all investigated approaches. The traditional partition-based approaches, as well as the GCH and CCV, do not allow for space reduction by partially representing the content of an image. As the proposed approach will use different number of cell histograms according to the chosen parameters, we will measure the space overhead by the total number of bins required.

The results of the last experiment are shown in Table 4. CCHs representing 100% of the image colors and partitioning the images in 8x8 cells, offer the best retrieval effectiveness, with a reduction of 55% in space overhead when compared with Grid. Alternatively, CCHs representing 80% of the image colors and partitioning the images into 3x3 cells implies the smaller space overhead, 15% smaller than the single GCH, yet yielding a $NavgR'$ 75% better.

It is also possible to obtain intermediate results, for instance, using CCHs with 90% of the colors and a 4x4 grid results in a respectable 96.5% of reduction in space overhead when compared to the Grid approach. As well, using twice the number of bins that the GCH uses, it is possible to obtain a $NavgR'$ 150% better. Other intermediate results may be obtained by choosing an adequate compromise between the number of cells and the percentage of the images (number of colors) being represented. It is possible to emphasize retrieval effectiveness or space overhead reduction. These results confirm the potential and the flexibility of the using CCHs in comparison to the traditional global and partition-based approaches. The Precision vs. Recall curves of this experiment are also shown in Figure 9.

Table 4: Comparing the performance of different CCH configurations

	Parameters		**Results**		
Approach	**Grid Size**	**% of the image**	**# of bins**	$NavgR'$	**Gain wrt GCH**
GCH	1x1	100% (64 colors)	64	0.08	—
CCV	1x1	100% (64 colors)	128	0.11	37%
Grid	8x8	100% (64 colors)	4096	0.28	250%
CCH	8x8	100% (29* colors)	1856*	0.30	275%
CCH	6x6	95% (12* colors)	432*	0.26	225%
CCH	4x4	90% (9* colors)	144*	0.20	150%
CCH	3x3	80% (6* colors)	54*	0.14	75%

*Average values

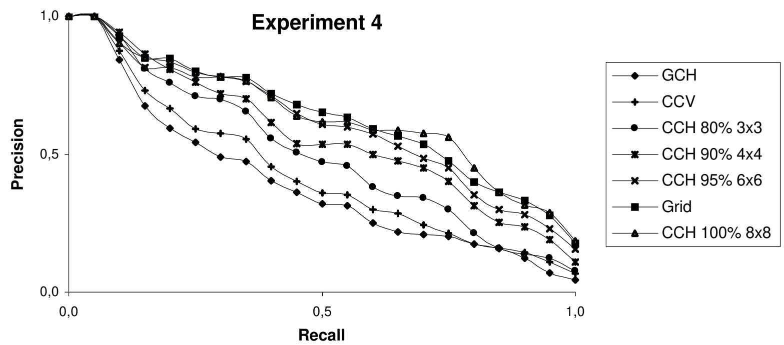

Figure 9: Precision vs. Recall curves for Experiment IV

6 Conclusions and Future Work

Our main contribution in this paper is a simple, yet very effective, variation of the partition-based techniques called Cell/Color Histograms. To the best of our knowledge, it is original in the way that visual features are encoded. Our motivation was to reduce the space overhead of partition-based approaches taking advantage of the fact that only a relatively low number of distinct values of a visual feature are present in most images. In particular, we used color features to verify the idea. We also proposed a similarity metric based on the L_1 distance metric.

The experimental results we obtained confirm the potential and the flexibility of the CCH approach in comparison to traditional global and partition-based approaches. It is possible to emphasize retrieval effectiveness or space overhead reduction. For instance, if warranted, one could use the CCHs and obtain the smallest space overhead, 15% smaller than a GCH while still having a $NavgR'$ 75% better. The experiments discussed in this paper, show that under "usual" (e.g., using default parameters) circumstances CCH performs consistently well.

Future research will concentrate in alternative color spaces, for example, HSV and uniform spaces such as L*a*b* and L*u*v* [4]. Accordingly to the color-space, we will also investigate some alternative quantization schemes and alternative similarity metrics. Another topic of interest is alternative representations for the CCHs, for instance, cumulative histograms or initial moments of their spatial distribution. Experiments using object-based queries and the study of using automatic image segmentation techniques to decompose images in a more consistent way will also be pursued.

Finally, in order to optimize query processing, we will also investigate indexing structures, e.g., the M-tree [6]. The M-tree is an interesting approach in our context because, unlike traditional spatial access structures (e.g., R-trees [11]) it it requires only the notion of distances between objects. As such, (i) it does not suffer from the problems related to the higher dimensionality of the indexing space (usually histograms are mapped onto high-dimensional Euclidean spaces), and (ii) it would allow one to indeed save storage space by actually storing a variable number of histogram cells per image.

7 Acknowledgments

Renato O. Stehling was supported by a Graduate Scholarship from FAPESP, Brazil. Mario A. Nascimento was partially supported by a Research Grant from NSERC, Canada. Alexandre X. Falcão was partially supported by a Research Fellowship Grant from CNPq, Brazil.

References

[1] D. Androutsos, K.N Plataniotis, and A.N. Venetsanopoulos. Vector angular distance measure for indexing and retrieval of color. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases VII*, volume 3656, pages 604–613, 1999.

[2] A.R. Appas, A.M. Darwish, A.I. El-Desouki, and S.I. Shaheen. Image indexing using composite regional color channels features. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases VII*, volume 3656, pages 492–500, 1999.

[3] J. Ashley, R. Barber, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic. Automatic and semi-automatic methods for image annotation and retrieval in QBIC. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 24–35, 1995.

[4] A. Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann, 1999.

[5] Y. Chen and E.K. Wong. Augmented image histogram for image and video similarity search. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases VII*, volume 3656, pages 523–429, 1999.

[6] P. Ciaccia, M. Partella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proc. of VLDB'97 Intl. Conf.*, pages 426–435, 1997.

[7] Kodak Company. Flashpix format and architecture white paper. http://www.kodak.com/, July 1996.

[8] A. Dimai. Spatial encoding using differences of global features. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases IV*, volume 3022, pages 352–360, 1997.

[9] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, 1994.

[10] B. Funt and G. Finlayson. Color constant color indexing. *IEEE PAMI*, 17(5):522–529, 1995.

[11] V. Gaede and O. Guenther. Multidimensional access methods. *ACM Computing Surveys*, 30(2):123–169, 1998.

[12] L.J. Guibas, B. Rogoff, and C. Tomasi. Fixed-window image descriptors for image retrieval. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 352–362, 1995.

[13] V. Kashyap, K. Shah, and A. Sheth. *Metadata for building the multimedia patch quilt*. Springer-Verlag, 1995.

[14] K.-S. Leung and R. Ng. Multiresolution subimage similarity matching for large image databases. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases VI*, volume 3312, pages 259–270, 1998.

[15] K.-C. Liang and C.-C.J. Kuo. Implementation and performance evaluation of a progressive image retrieval system. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases IV*, volume 3312, pages 37–48, 1998.

[16] S. Lin. An extendible hashing structure for image similarity searches. Technical Report TR-00-06, Dept of Computing Science, University of Alberta, 2000. (MSc Thesis).

[17] J. Malki, N. Boujemaa, C. Nastar, and A. Winter. Region queries without segmentation for image retrieval by content. In *Proc. of VISUAL'99 Intl. Conf.*, pages 115–122, 1999.

[18] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *Proc. of ACM Multimedia Conference*, pages 65–73, 1996.

[19] H. Sakamoto, H. Suzuki, and A. Uemori. Flexible montage retrieval for image data. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases II*, volume 2185, pages 25–33, 1994.

[20] G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.

[21] E. Di Sciascio, G. Mingolla, and M. Mongiello. Content-based image retrieval over the web using query by sketch and relevance feedback. In *Proc. of VISUAL'99 Intl. Conf.*, pages 123–130, 1999.

[22] S. Sclaroff. Distance to deformable prototypes: Encoding shape categories for efficient search. In *Image Databases and Multi-Media Search*. World Scientific, 1997.

[23] N. Sebe, M.S. Lew, and D.P Huijsmans. Multi-scale sub-image search. In *Proc. of 1999 ACM Multimedia Intl. Conf.*, pages 79–82, 1999.

[24] I.K. Sethi, I. Coman, B. Day, F. Jiang, D. Li, J. Segovia-Juarez, G. Wei, and B. You. Color-wise: A system for image similarity retrieval using color. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases IV*, volume 3312, pages 140–149, 1998.

[25] E. Shusterman and M. Feder. Image compression via improved quadtree decomposition algorithms. *IEEE Trans. Image Processing*, 3(2):207–215, 1994.

[26] J.R. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis*. PhD thesis, Graduate School of Arts and Sciences, Columbia University, Feb 1997.

[27] J.R. Smith and S.-F. Chang. Joint adaptive space and frequency graph basis selection. In *Proc. of IEEE Int. Conf. Image Processing (ICIP)*, pages 702–705, 1997.

[28] R.O. Stehling, M.A. Nascimento, and A.X. Falcão. On 'shapes' of colors for content-based image retrieval. In *Proc. of the ACM Intl. Workshop on Multimedia Information Retrieval - MIR'00*, pages 171–174, 2000.

[29] M. Stricker and M. Orengo. Similarity of color images. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 381–392, 1995.

[30] M.J. Swain and D.H. Ballard. Color indexing. *Intl. J. Computer Vision*, 7(1):11–32, 1991.

[31] Y.J Zhang, Z.W. Liu, and Y. He. Comparison and improvement of color-based image retrieval tchniques. In *Proc. of SPIE -Storage and Retrieval for Image and Video Databases VI*, volume 3312, pages 371–382, 1998.