CHAPTER 3

COLOR FEATURE EXTRACTION

3.1 INTRODUCTION

Primitive features denote some general visual characteristics including color, shape, texture, and spatial relationships among objects, and these features can be used in most CBIR applications. The color feature, which is widely used in CBIR systems, will be discussed in detail in this chapter. Characters which discriminate the images from each other and used to find the similarity between them, are called Image Features. The most common Image Features are Color, Texture, Shape etc.,

Image features can be classified as

- General features
- Domain specific features

The former include color, texture, and shape features, while the latter are application-dependent and may include, features of human faces or finger prints.

The human eye is sensitive to colors, and color features are one of the most important elements enabling humans to recognize images. Color features are, therefore, fundamental characteristics of the content of images. Color features can sometimes provide powerful information for categorizing images, and they are very useful for image retrieval. Therefore, color-based image retrieval is widely used in CBIR systems.

Color feature is relatively robust to background complication and independent of image size and orientation. In image retrieval, the color histogram is the most commonly used color feature representation. Each image added to the collection is analyzed to compute a color histogram which shows the proportion of the pixels of each color within the image. The color histogram for each image is then stored in the database. During searching time, the user can either specify the desired proportion of each color (75% olive green and 25% red, for example), or submit an example image from which a color histogram is calculated. Either way, the matching process retrieves those images, whose color histograms match those of the query most closely. Considering that most color histograms are very sparse and thus sensitive to noise, Stricker and Orengo proposed using the cumulated color histogram. To overcome the quantization effects, as in the color histogram, Stricker and Orengo (1995) proposed using the color moments-based approach. The mathematical foundation of this approach is that any color distribution can be characterized by its moments.

3.2 COLOR SPACE

Colors are commonly defined in three-dimensional color spaces. Color space models can be differentiated as hardware-oriented and user-oriented. The hardware-oriented color spaces, including RGB, CMY, and YIQ, are based on the three-color stimuli theory.

Color spaces provide the method to manipulate colors. A color space is defined as a model for representing color in terms of intensity values. The following four models are used in a color image retrieval system.

- The RGB color model
- The HSV Color Model
- The HMMD color space
- The CIE lab color space

The HSV color model has been described in section 2.3 of this thesis. Transformations are available for conversion from one representation of color to other models. A complete set of the color models and the transformations between color models are presented in Appendix 1.

3.3 COLOR QUANTIZATION

In order to produce color histograms, color quantization has to be applied. Color quantization is the process of reducing the number of colors used to represent an image. A quantization scheme is determined by the color space and the segmentation (i.e., split up) of the color space used.

In applying a standard quantization scheme on a color space, each axis is divided into a number of parts. When the axes are divided into *k*, *l*, and *m* parts, the number of colors (n) used to represent an image will be $n = k. l.m$. A quantization of color space into n colors is often referred to as a n-bins quantization scheme. Figure 3.1 illustrates the effect of quantizing color images. The segmentation of each axis depends on the color space used. In the next section, different color spaces and their quantization methods will be described.

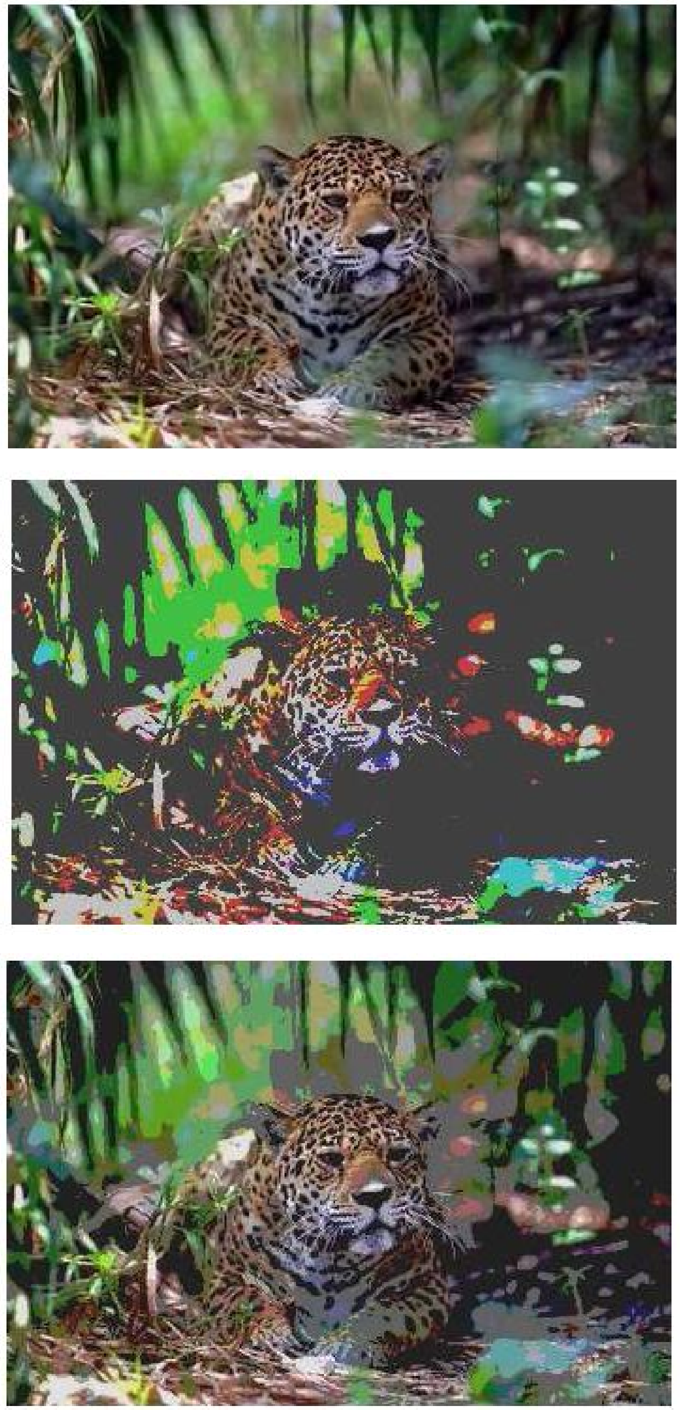


Figure 3.1 From top to bottom: The original image using 256^3 colors, quantized in 8 bins, and quantized in 64 bins, using RGB color space

3.4 COLOR HISTOGRAMS

Color histograms are defined as a set of bins where each bin denotes the probability of pixels in the image being of a particular color. A color histogram for a given image is defined as a vector:

$$H = \{H[0], H[1], H[2], \ldots, H[i], \ldots, H[N]\} \tag{3.1}$$

where i represents a color in the color histogram and corresponds to a sub-cube in the RGB color space, $H[i]$ is the number of pixels in color i in that image, and N is the number of bins in the color histogram, i.e., the number of colors in the adopted color model.

Typically, each pixel in an image will be assigned to a bin of a color histogram of that image, so for the color histogram of an image, the value of each bin is the number of pixels that has the same corresponding color. In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H' is defined as:

$$H' = \{H'[0], H'[1], H'[2], \ldots, H'[i], \ldots, H'[N]\} \tag{3.2}$$

where $H'[i] = \frac{H[i]}{P}$, P is the total number of pixels in an image (the remaining variables are defined as before).

An ideal color space quantization presumes that distinct colors should not be located in the same sub-cube and similar colors should be assigned to the same sub-cube. Using fewer colors will decrease the possibility of similar colors being assigned to different bins, but increase the possibility of distinct colors being assigned to the same bins, and the information content of the images will decrease by a greater degree as well. On the other hand, color histograms with a large number of bins will contain more information about the content of images, thus decreasing the possibility of distinct colors being assigned to the same bins. However, they increase the possibility of similar colors being assigned to different bins, increase the storage space of metadata, and the time for calculating the distance between color histograms. Therefore, there is a trade-off in determining how many bins should be used in color histograms. A typical figure found in the related literature is 64.

3.4.1 Color histogram discrimination

The color distance formulas arrive at a measure of similarity between images based on the perception of color content. Three distance formulas that have been used for image retrieval include histogram Euclidean distance, histogram intersection and histogram quadratic (cross) distance.

Histogram Euclidean distance

Let ***h*** and ***g*** represent two color histograms. The Euclidean distance between the color histograms ***h*** and ***g*** can be computed as:

$$d^2(h,g) = \sum_A \sum_B \sum_C (h(a,b,c) - g(a,b,c))^2 \qquad (3.3)$$

Histogram intersection distance

The color histogram intersection was proposed for color image retrieval. The intersection of histograms ***h*** and ***g*** is given by:

$$d(h,g) = \frac{\sum_A \sum_B \sum_C \min(h(a,b,c), g(a,b,c))}{\min(|h|,|g|)} \qquad (3.4)$$

where |***h***| and |***g***| gives the magnitude of each histogram, which is equal to the number of samples. Colors not present in the user's query image do not contribute to the intersection distance. This reduces the contribution of background colors. The sum is normalized by the histogram with fewest samples.

Histogram quadratic (cross) distance

The color histogram quadratic distance was used by the QBIC system. The cross distance formula is given by:

$$d(h,g) = (h-g)^t A(h-g) \tag{3.5}$$

The cross distance formula considers the cross-correlation between histogram bins based on the perceptual similarity of the colors represented by the bins. And the set of all cross-correlation values are represented by a matrix ***A*,** which is called a similarity matrix. And a (*i*,*j*)th element in the similarity matrix ***A*** is given for RGB space,

$$a_{ij} = 1 - d_{ij} / \max(d_{ij}) \tag{3.6}$$

where d_{ij} is the distance between the color *i* and *j* in the RGB space.

Problems with histogram based retrieval

The first problem is the high dimensionality of the color histograms. This high dimensionality ensures that methods of feature reduction, pre-filtering and hierarchical indexing must be implemented. The large dimensionality also increases the complexity and computation of the distance function. It particularly complicates cross distance functions that include the perceptual distance between histogram bins.

3.5 COLOR MOMENTS

To overcome the quantization effects as in color histogram, Stricker and Orengo (1995) proposed to use the color moments approach. The color moments of an image are a simple yet effective feature for color-based image retrieval (Stricker and Orengo 1995, Ghosal and Mehrota 1997, Mandal et al 1998, Vailaya et al 1999). From the probability theory, it is observed that a probability distribution is uniquely characterized by its moments. Thus, if we interpret the color distribution of an image as a probability distribution, then the color distribution can be characterized by its moments as well. Furthermore, as most of the color distribution information can be captured by the low-order moments, using only the first three moments: mean, variance and skewness, it is found that these moments give a good approximation and have been proven to be efficient and effective in representing the color distribution of images (Stricker and Orengo 1995). These first three moments are defined as:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \tag{3.7}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(p_{ij}-\mu_i)^2} \tag{3.8}$$

$$S_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij}-\mu_i)^3\right)^{1/3} \tag{3.9}$$

where p_{ij} is the value of the i^{th} color channel of the j^{th} image pixel. Only 3 x 3 (three moments for each color component) matrices to represent the color content of each image are needed which is a compact representation compared to other color features. To overcome the shortcomings of having lower discrimination power (if used alone), color moments can be used as the first pass to narrow down the search space before other more elaborate measures are used for retrieval.

3.6 COLOR COHERENCE VECTOR

A color coherence vector (CCV) is a color histogram refinement scheme that divides each bin into coherent and non-coherent pixels (Pass et al 1996). A pixel in a bin is said to be coherent if it is part of a large similarity-colored region. An 8-neighbor connected component analysis is used to extract connected regions of the same color. Pixels in regions whose size exceed a threshold (1% of image size) are counted as coherent pixels, and those from smaller regions are counted as non-coherent regions.

Since color histograms and color moments lack information about the spatial distribution of colors in an image, color coherence vectors have been proposed to incorporate spatial information into color histogram representations (Ma and Manjunath 1997, Vailaya et al 1999). Let α_i denote the number of coherent pixels in the i^{th} bin and β_i the number of incoherent pixels. The color coherence vector of the image is defined as $<(\alpha_1, \beta_1), (\alpha_2, \beta_2), ... (\alpha_N, \beta_N)>$. It is noted that $<\alpha_1+\beta_1, \alpha_2+\beta_2, ...\alpha_N+\beta_N>$ is the color histogram of the image. With additional distinguishing power, color coherence vectors have been shown to provide better retrieval performance than color histograms (Vailaya et al 1998).

3.7 COLOR CORRELOGRAMS

The color correlogram was proposed as an image feature vector for image indexing and retrieval by Huang et al (1997) and proved to be very effective and efficient for content-based image retrieval. Unlike purely local properties such as pixel position, gradient direction, or purely global properties such as color distribution, correlograms take into account the local color spatial correlation as well as the global distribution of this spatial distribution. Therefore, the color correlogram was shown to be a generic tool for spatial color indexing (Huang et al 1998) that could be applied to problems like image subregion querying, object searching, image localization, tracking, video cut detection, etc. The results in Huang et al (1997) indicate that in a large image database the overall performance of the correlogram was much better than that of the histogram, the color coherence vector, and the color coherence vector with successive refinement.

3.8 MPEG -7 COLOR DESCRIPTORS

Several approaches have been proposed, to describes the way colors in an image should be characterized, relating to their perception, coherency and spatial distribution. MPEG-7 has standardized a subset of these approaches in the form of color descriptors. MPEG-7 is a content representation standard for information search.

These descriptors cover different aspects of color and application areas. They are as follows:

- dominant color
- scalable color

- color layout
- color structure

In the following sections each of these descriptors are explained in detail.

3.8.1 Dominant color descriptor (DCD)

A set of dominant colors in a region of interest or in an image provide a compact description that is easy to index. The target application is similarity retrieval in large image databases using color. Colors in a given region are clustered into a small number of representative colors. The feature descriptor consists of the representative colors and their percentages in the region. A similarity measure, similar to the quadratic color histogram distance measure, is defined for this descriptor. The representative colors can be indexed in the 3-d color space thus avoiding the high-dimensional indexing problems associated with the traditional color histogram. For similarity retrieval, each representative color in the query image or region is used independently to find regions containing that color. The matches from all the query colors are then combined to obtain the final retrievals.

The difference between the dominant color descriptor and the color histogram descriptor is that the representative colors are computed from each image instead of being fixed in the color space, thus allowing the feature representation to be accurate as well as compact.

In order to compute this descriptor, the colors present in a given image or region are first clustered. This results in a small number of colors and the percentages of these colors are calculated. The percentages of the colors present in the region should add up to 1. The descriptor is thus defined by

$$F=\{C_i, P_i\}, \qquad (i =1,2,\ldots\ldots\ldots n) \tag{3.10}$$

where C_i i^{th} dominant color

P_i its percentage value

The number of dominant colors ‘n’ can vary from image to image and a maximum of eight dominant colors can be used to represent the region.

(i) Color Segmentation: Color segmentation is done by first segmenting the image according to the number of colors present in the image.

(ii) l*a*b* color space is used to segment the images.

(iii) K-means clustering that treats each object as having a location in space.

(iv) Label every pixel in the image using the results from k-means.

(v) Images are segmented based on color.

After labeling the pixels, it can separate the objects in the image by color, which will result in separate images according to the number of colors present.

Figure 3.2 gives of sample image segmented into four clusters. From top to bottom the first image is the input image, the rest of the images are segmented images for four clusters.

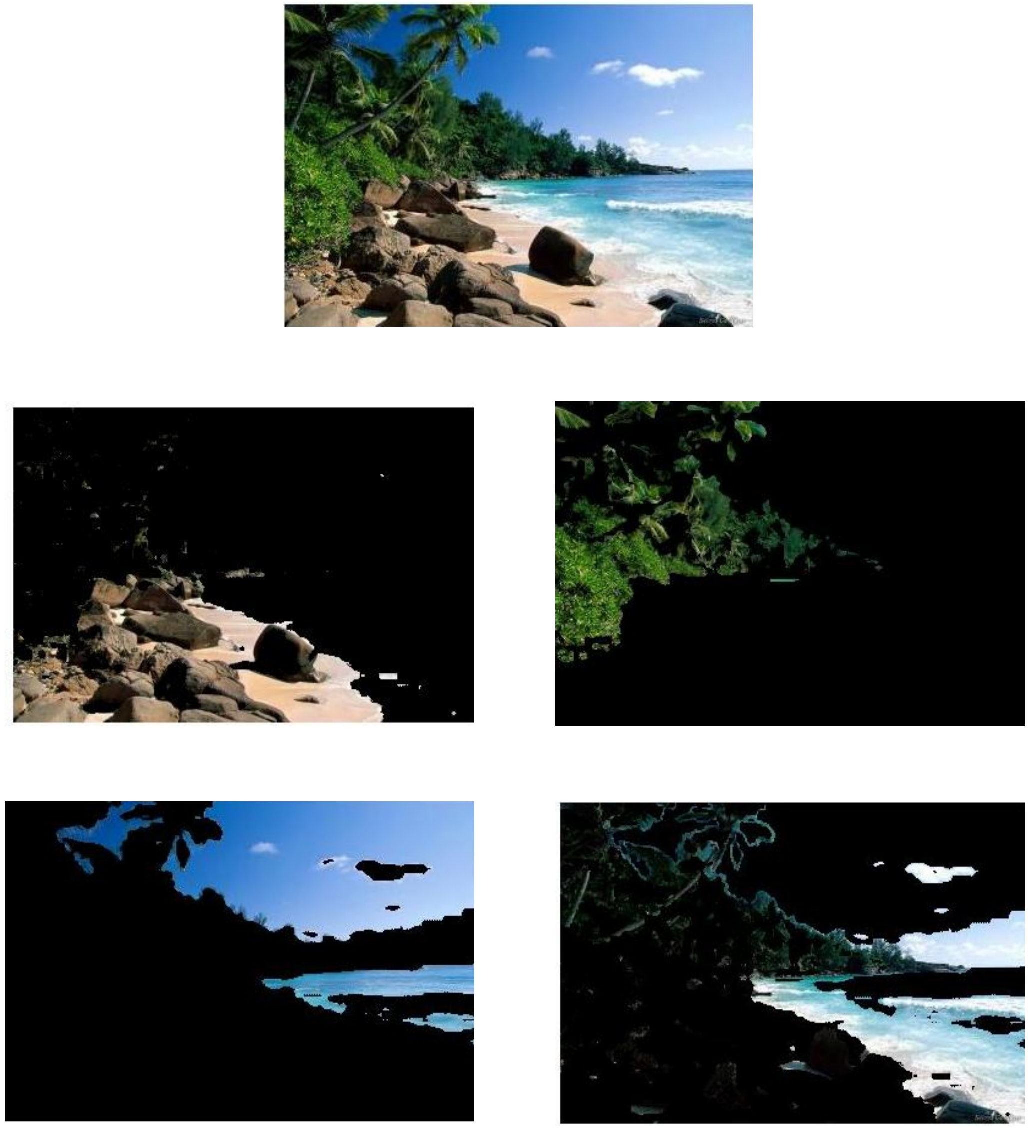


Figure 3.2 Dominant color separation with four clusters

3.8.2 Scalable color descriptor (SCD)

The SCD addresses the interoperability issue by fixing the color space to the HSV, with a uniform quantization of the HSV space into 256 bins. The bin values are non-uniformly quantized to a 11-bit value.

This method achieves full interoperability between different resolutions of the color representation, ranging from 16 bits/histogram at the low end to approximately 1000 bits/histogram at the high end. Of course, the accuracy of the feature description is highly dependant on the number of bits used.

The HSV space is uniformly quantized into a total of 256 bins. This includes 16 levels in h, four levels in s, and four levels in v. The histogram values are truncated into a 11-bit integer representation. To achieve a more efficient encoding, the 11-bit integer values are first mapped into a nonlinear 4-bit representation, giving higher significance to the small values with higher probability.

This 4-bit representation of the 256-bin HSV histogram would require 1024 bits/histogram, which is too large a number. To lower this number and make the application scalable, the histograms are encoded using a haar transform.

The basic unit of the haar transform consists of a sum operation and a difference operation, which correspond to primitive low- and high-pass filters. Summing pairs of adjacent histogram lines is equivalent to the calculation of a histogram with the half number of bins. If this process is performed iteratively, the use of subsets of coefficients in the haar representation is equivalent to histograms of 128, 64, 32…bins, which are all calculated from the source histogram.

It can be expected that the high-pass coefficients expressing differences between adjacent histogram bins usually have only small values. Exploiting this property, it is possible to truncate the high-pass coefficients to an integer representation with only a low number of bits.

(i) The only feature extracted is the upper co-efficients of the haar transform.

(ii) first perform the discrete wavelet transform on the image.

(iii) the discrete wavelet transform results in three detailed components.

(iv) form a matrix from the detailed components and find the mean of that matrix.

(v) this is the final mean value to be stored in the database.

The distance between two feature vectors $X^{(1)}$ and $X^{(2)}$ can be computed based on the L2-distance or the Euclidean Distance Measure.

$$D_1 = \sum_{I=1}^{N} (X_i^{(1)} - X_i^{(2)})^2 \tag{3.11}$$

3.8.3 Color layout descriptor(CLD)

The CLD is designed to capture the spatial distributions of color in an image. The spatial distribution of color constitutes an effective descriptor for sketch-based image retrieval, content filtering using image indexing and visualization.

The CLD is a compact descriptor that uses representative colors on an 8*8 grid followed by a Discrete Cosine Transform and encoding of the resulting coefficients. The feature extraction process consists of two parts; grid based representative color selection and Discrete Cosine Transform with quantization. An input image is divided into 64(8*8) blocks and their average colors are derived. It is implicitly recommended that the average color be used as the representative color for each block. This partitioning process is important to guarantee the resolution or scale invariance. The derived average colors are transformed into a series of coefficients by performing the 8*8 Discrete Cosine Transform; the color space adopted for the color layout descriptor is yc_rc_b.

The features extracted here are dy, dc_r, dc_b which are the Discrete Cosine Transform coefficients of the respective color component.

(i) the image is converted into yc_rc_b color space.

(ii) each of the color components is divided into 64 blocks and their average colors are derived.

(iii) the derived average colors are transformed into a series of coefficients by performing the 8*8 Discrete Cosine Transform.

(iv) the mean for each of the color components is calculated and stored in the database.

For matching two Color Layout Descriptors, $\{DY, dc_r, dc_b\}$ and $\{DY', dc_r', dc_b'\}$, the following distance measure is used:

$$D = \sqrt{\sum_i w_{yi}(DY_i - DY_i^{'})^2} + \sqrt{\sum_i w_{bi}(DCb_i - DCb_i^{'})^2} + \sqrt{\sum_i w_{ri}(DCr_i - DCr_i^{'})^2} \quad (3.12)$$

Here,(DY_i ,DC_{bi}, DC_{ri}) represent the i^{th} DCT coefficients of the respective color components. The distances are weighed appropriately, with larger weights given to the lower frequency components.

3.8.4 Color structure descriptor (CSD)

This descriptor expresses the local color structure in an image using an 8 x 8 -structuring element. It counts the number of times a particular color is contained within the structuring element as the structuring element scans the image. Suppose $c_0, c_1, c_2, \ldots, c_{M-1}$ denote the M quantized colors. A color structure histogram can then be denoted by $h(m)$, $m = 0, 1, \ldots, M-1$, where the value in each bin represents the number of structuring elements in the image containing one or more pixels with color c_m. The HMMD color space is used in this descriptor.

The CSD is defined using four color space quantization operating points: 184, 20, 64, and 32 bins. To construct a 184-level quantized color, the HMMD color space is quantized non-uniformly as follows. The whole HMMD color space is divided into five subspaces. This sub-space division is performed on the difference parameter. For the respective subspaces, uniform color quantization on the Hue and Sum values results in a 184-level color quantization.

In order to compute the CSD, an 8 x 8-structuring element is used. Even though the total number of samples is kept fixed at 64, the spatial extent of the structuring element scales with the image size. The following simple rule determines the spatial extent of the structuring element (equivalently, the sub sampling factor) given the image size:

$$\begin{aligned} &p = \max\{0, round(0.5\log_2 WH - 8)\} \\ &K = 2^p, \qquad E = 8K \end{aligned} \tag{3.13}$$

where

W, H	image width and height, respectively
E x E	spatial extent of the structuring element
K	sub-sampling factor

For images smaller than 256 x 256 pixels, an 8 x 8 element with no-sampling is used. As another example, if the image size is 640 x 480, then $p = 1$, $K = 2$, and $E = 16$. So every alternate sample along the rows and columns of a 16 x 16-structuring element is then used to compute the histogram.

The structuring element slides over the image. Sub-sampling of the image is carried out in both directions and subsequently applying the same 8 x 8 structuring element. Each bin of the CSD $\boldsymbol{h}(\boldsymbol{m})$ represents the number of locations of the structuring element at which a pixel with color $\boldsymbol{c_m}$ falls inside the element. The origin of the structure element is defined by its top-left sample. The locations of the structure element over which the descriptor is accumulated are defined by the grid of pixels of the possibly sub-sampled input image.

The bin values $h(m)$ of the CSD are normalized by the number of locations of the structuring element and lie in the range [0.0, 1.0]. The bin values are then nonlinearly quantized to 8 bits/bin.

CSDs containing 120, 64, or 32 bins are computed, based on approximations computed using the 184-bin descriptor. The mapping of the 184-bin descriptor to a descriptor with a lower number of bins is defined by re-quantizing the color represented by each bin of the 184-bin descriptor into the more coarsely quantized color space.

Similar to the other histogram descriptors, an L distance measure is used to compute the dissimilarity between two CSDs. The common color dataset was slightly modified by the addition of a few more query images so as to illustrate the qualitative difference in the retrieval performance between the color structure and scalable color histograms.

3.9 EXPERIMENTATION

The above mentioned algorithms are experimented with a number of images The images considered are collected from the Corel and Cires data sets. Their sizes are 128 x 128. The algorithms are implemented in the P-IV personal computer @ 2.44 GHz.

In this work, the color moments descriptors and MPEG-7 color descriptors are adopted for experimentation. The features evaluated for some sample images from the corel data set are provided in this section.

Color moment features

The image is clustered into 5 regions (r_1,r_2,r_3,r_4,r_5) and for each region the color moment values(μ,σ,S)as in equation numbers (3.7) – (3.9) are calculated. In Tables 3.1 and 3.2 each row contains the 9 moment values corresponding to a region in the two sample images.

Table 3.1 Color moment features for sample image 1

Sample Image 1	Moment feature vector values								
	R			G			B		
	μ	σ	S	μ	σ	S	μ	σ	S
	27.07	15.53	11.29	6.41	9.04	10.98	3.45	4.34	4.95
	16.73	16.80	14.65	6.93	9.85	12.05	3.93	4.96	5.67
	27.60	22.05	17.24	7.21	10.11	12.29	3.74	4.69	5.35
	38.33	32.37	22.09	7.03	9.94	12.16	3.67	4.62	5.29
	17.53	21.70	15.17	6.70	9.55	11.65	3.57	4.51	5.16

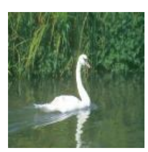
Table 3.2 Color moment features for sample image 2

Sample Image 2	Moment Feature vector values								
	R			G			B		
	μ	σ	S	μ	σ	S	μ	σ	S
	23.633	31.13	22.87	8.92	12.68	15.53	4.33	5.45	6.24
	17.97	18.93	18.40	4.98	7.04	8.63	2.92	3.67	4.21
	20.58	26.22	17.00	8.00	11.34	13.86	4.01	5.05	5.78
	5.44	7.28	3.92	4.01	5.68	6.95	2.53	3.18	3.64
	15.69	25.48	16.22	8.13	11.59	14.21	4.08	5.15	5.89

Dominant color features

In Dominant color feature evaluation, the image is segmented into 3 clusters. The average of the region colors and their percentage of the corresponding region colors are given in Table 3.3.

Table 3.3 Dominant color and color layout features

Sample 1	Dominant color features		Color layout features Avg (Cv, Ch, Cd)	Sample 2	Dominant color features		Color layout features Avg (Cv, Ch, Cd)
	Dominant color value	Dominant color %			Dominant color value	Dominant color %	
	23.8987	32.8573	10.4509		37.4270	42.5171	11.6049
	47.9088	31.1442	15.2803		22.6565	28.5950	14.3171
	19.5993	18.3940	16.4453		17.8810	18.6035	14.7759

Color layout features

The color layout features are the average values of the diagonal, vertical and horizontal components from the Discrete Cosine Transform applied. These values for two sample images are provided in Table 3.3.

Color Structure Features

Using Equation (3.13) discussed under section 3.8.4 the results of the experimentation were tabulated as in Table 3.4. There are 81 values resulting for the structuring element used in this feature evaluation.

Table 3.4 Color Structure Descriptor Features (80 bin values and DC component)

Image and Feature values	29.77	191.01	873.18	2721.23	191.01	162968
921575	98694.4	64273.79	873.18	744997	451174	293823
600166	2721.23	2321768	1406075	915692	1123.71	958761
1771.15	580630	378130	910.38	776744	470400	306343
1511160	828.52	706901	428103.	278798.	987.28	842355
	510134	332220	1071.62	914315	553714	360601
	1230.38	1049769	635746	414023	1408.98	1202155
728031	474123	1582.63	1350308	817754	532554	1681.85
1434967	869023	565943	1724.02	1724.02	1470947	890813
580133	1704.18	1454015	880559	573455	1783.56	1521742
915166	1676.89	595993	1761.23	1502694	910039	592654
1430734	866460	564273	1609.91	1373589	831853	541736

Scalable Color Features

The scalable color feature values for the images considered in the above cases are -0.3162, 0.0478 for the tiger and water bird images (sample 1 and sample 2 images) as per the explanation given in section 3.8.2.

3.10 CONCLUSION

The color moments descriptor has a compact representation. The moment descriptor includes the average, variance, and the third-order moment of the colors in the image. It is observed that the color moments descriptor performs slightly worse than a high-dimensional color histogram. One drawback of the moment descriptor is that the average of all the colors might be quite different from any of the original colors. Given a color moment feature description, it is difficult to recover the actual colors in the image.

The SCD and the CSD are two histogram-based descriptors, the dominant color descriptor, and the CLD. The histogram descriptors capture the global distribution of color whereas the dominant color descriptor represents the dominant colors present. The CLD captures the spatial distribution or layout of the colors in a compact representation. While MPEG-7 standards accommodate different color spaces, most of the color descriptors are constrained to one or a limited number of color spaces for ensuring inter-operability.

The dominant descriptor is also quite compact, and is based on the observation that a small number of colors is usually sufficient to characterize the color information in an image region. Since the descriptor captures the representative or dominant colors in a given region, it is referred to as the dominant color descriptor. The dominant color descriptor consists of the representative colors and their relative distribution in a given region. A similarity measure is defined for this color descriptor and is shown to be equivalent to the popular quadratic color histogram distance measure. However, the difference between the new descriptor and the color histogram descriptor is that the representative colors are computed from each image instead of being fixed in the color space, thus allowing the feature representation to be accurate as well as compact. Unlike the compact color moments descriptor, the dominant color representation allows simple visualization of the color distribution in the image.

The results demonstrate that the CLD is quite effective in image retrieval. The results also compare favorably with a grid based dominant color approach wherein the image is partitioned and dominant colors for these partitions are used to represent the layout. Feature extraction techniques of texture content in images are discussed in the next chapter.