A Compact and Efficient Color Descriptor for Image Retrieval

Ka-Man Wong, Lai-Man Po, and Kwok-Wai Cheung

kmwong@ee.cityu.edu.hk, eelmpo@cityu.edu.hk, kwcheung@ee.cityu.edu.hk
Department of Electronic Engineering, City University of Hong Kong,
83 Tat Chee Avenue, Kowloon, Hong Kong SAR of China.

Abstract — An important problem in color based image retrieval is the lack of efficient way to represent both the color and the spatial structure information with single descriptor. To solve this problem, a new Dominant Color Structure Descriptor (DCSD), is proposed. The descriptor combines the compactness of Dominant Color Descriptor (DCD) and the accuracy of Color Structure Descriptor (CSD) to enhance the retrieval performance in a highly efficient manner. The feature extraction and similarity measure of the descriptor are designed to address the problems of the existing descriptors such as color inaccuracy of DCD and redundancy of CSD. Experimental results show that DCSD has a significant improvement in retrieval performance and descriptor size over DCD. An eight-color DCSD (DCSD 8) gives an Averaged Normalized Modified Retrieval Rate (ANMRR) of 0.0993 using MPEG-7 common color dataset, outperforming compact configurations of Scalable Color Descriptor and Color Structure Descriptor with smaller descriptor size.

Keywords — Content Based Image Retrieval, Dominant Color Structure Descriptor, Dominant Color Descriptor, Color Structure Descriptor.

I. Introduction

Recently, the rapid development of digital image and video devices catalyzed the growth of digital multimedia contents. In response to the need of efficient management of multimedia database systems, MPEG-7 [1] has been developed as a standard for multimedia content description to facilitate interoperability between different searching engines and different databases for multimedia data searching and content exchange. For describing multimedia contents, low level features are usually used since they can be generated efficiently. Among the low level image features, color is the most expressive and effective in visual content description. Several color descriptors [2] have been defined in MPEG-7 for describing color features of multimedia contents. These color descriptors are applicable to a broad range of applications according to their complexity (in terms of extraction and similarity measure) and efficiency (in terms of descriptor size and accuracy).

Among the color descriptors, Dominant Color Descriptor (DCD) [2], is a compact color descriptor, it stores the dominant colors, their percentages, and several optional parameters. Since DCD stores only the dominant colors (max. 8 colors) instead of a color histogram, the storage requirement is very effective with relatively small redundancy. Unfortunately, DCD performance is the worst among various color descriptors in its existing usage as reported in [2-3]. Its performance is even worse than a more compact descriptor, Color Layout Descriptor. The performance of DCD is quite limited due to the lack of spatial color distribution information and the color accuracy problem. Among various MPEG-7 color visual descriptors, Color Structure Descriptor (CSD) always gives the best retrieval rate. The main reason is the capture of spatial color distribution information in CSD. However, CSD has redundancy as it uses a fixed color space for the histogram representation. The most compact CSD configuration takes about 32 bytes for each image. The size is approximately 4 times that of CLD and about double of DCD. This might be negligible but, for mobile or other bandwidth concerned applications, every bit should be saved. To address this problem we can combine the dominant color features of DCD and the spatial color distribution structure of CSD to design a new color descriptor, called Dominant Color Structure Descriptor (DCSD), which maintains the compactness of dominant colors while significantly improving the retrieval accuracy by using structured dominant color histograms for introducing spatial distribution information of the image.

The organization of the remaining sections of this paper is as follows. Section II gives an overview of DCD and CSD which forms the basis for the subsequent discussion. Section III describes the core of the proposed descriptor - dominant color structure histogram. Section IV presents the similarity measure developed for DCSD. Finally, experimental results and a conclusion are presented in Section V and Section VI, respectively.

II. Limitations of Dominant Color Descriptor and Color Structure Descriptor

A. Dominant Color Descriptor

With target application of similarity retrieval in large image database, dominant color descriptor extracts the features from an image by clustering the colors in an image into a small number of colors and is defined as

$$F = \{(c_i, p_i, v_i), s\},\ (i = 1,..., N)\,. \tag{1}$$

The descriptor consists of the representative colors c_i , their percentages p_i, the optional color variances for each dominant color v_i, and the optional spatial coherency s of the dominant colors. The quadratic histogram distance measure (QHDM) [2] is used for similarity measure for DCD and it is defined as

$$D^2(F_1,F_2) = \sum_{i=1}^{N_1} p_{1i}^2 + \sum_{j=1}^{N_2} p_{2j}^2 - \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} 2a_{i,j} p_{1i} p_{2j}$$
$$a_{i,j} = \begin{cases} 1-\left\|c_{1i}-c_{2j}\right\|/\alpha T_d, & \left\|c_{1i}-c_{2j}\right\| \le T_d \\ 0, & \left\|c_{1i}-c_{2j}\right\| > T_d \end{cases} \tag{2}$$

where $F_1 = \{\{c_{1i}, p_{1i}\},\ i=1,...,N_1\}$ and $F_2 = \{\{c_{2i}, p_{2i}\},\ i=1,...,N_2\}$ are two DCD descriptors with N_1 and N_2 dominant colors, respectively. The dominant color and its percentage value are denoted by c_i and p_i. The sum of percentages is normalized to 1. The similarity coefficient, $a_{i,j}$, is used to take into account the closeness between the two dominant colors c_{1i} and c_{2j}. T_d is a threshold for determining the similarity between two colors and α is used for adjusting the importance of color distance, in the reference software XM T_d=15 and α=1.2 are used. With this simple and compact representation, DCD allows efficient indexing for similarity retrieval while sacrificing retrieval accuracy due to lack of spatial information of the description.

In DCD feature extraction, generalized Lloyd algorithm (GLA) [4] color quantization is applied on the image. Since GLA is a splitting process based on minimizing distortion of each cluster, the colors obtained by GLA can be very close. However, QHDM distance measure requires that the colors in a descriptor should have values with at least T_d away to maintain the metric of the similarity measure [2]. Thus, after the GLA quantization, agglomerative clustering is applied to the quantized colors by searching the closest pair of colors, c_1 and c_2 in CIE*Luv* color space. If the distance of the color pair is less than or equal to the threshold T_d , the two colors will be merged as

$$c_{new} = \frac{p_1 c_1 + p_2 c_2}{p_1 + p_2} \quad ; \quad p_{new} = p_1 + p_2 \qquad (3)$$

to give the combined color and its percentage, c_{new} and p_{new}. This merging process iterates until the minimum internal distance is larger than the threshold T_d so as to comply with the requirement of QHDM. The distortion due to the GLA process will not be reduced as this clustering process is not a reverse operation of GLA. Also, the additional distortion introduced is not minimized since the merging is not based on distortion minimization. Thus, the clustering makes the mixed colors become inaccurate and affects the retrieval performance.

B. Color Structure Descriptor

Color structure descriptor [2] is based on color histogram, but aims at providing a more accurate description by identifying localized color distributions of each color. CSD is characterized by a color structure histogram for M quantized color, c_m, and is expressed as

$$h(m),\ m = 1,\ldots,M \qquad (4)$$

where $M \in \{256, 128, 64, 32\}$ and the bin value $h(m)$ is the number of structuring elements containing one or more pixels with color c_m . Unlike the conventional histogram, the color structure histogram is extracted from an image by accumulation using an 8×8-structuring window. The structuring element scans the image and counts the number of times a particular color is contained with the structuring element. Let I denote be the set of quantized color indexes of an image and $S \in I$ be the set of quantized color indexes existing inside the subimage region covered by the structuring element. With the structuring element scanning the image, the color histogram bins are accumulated according to

$$h(m) \leftarrow h(m) + 1, \qquad m \in S \qquad (5)$$

Thus, the final value of $h(m)$ is determined by the number of positions at which the structuring element contains c_m .

An *L1* distance measure is used to compute the dissimilarity between CSDs. CSD provides more accurate similarity retrieval because of the inclusion of spatial color information. This representation is more closely related to the human perception and, thus, is more useful for indexing and retrieval. Although the color structure histogram contributes to the high retrieval accuracy of CSD, the fixed color space requirement of the histogram results in redundancy in the representation. For example, a DCD with eight colors need 21.75 bytes in binary representation, and a DCD with 4 colors only need 11.75 bytes. In contrast, the most compact CSD 32 uses 32 bytes per descriptor, which is about 1.5-3 times of DCD. In the following section, a new dominant color structure descriptor is proposed to achieve both compact descriptor size of DCD and relatively high retrieval accuracy of CSD.

III. Dominant Color Structure Histogram

The proposed new descriptor uses a color structure histogram based on representative colors extracted from image instead of using a fixed color space. Although the histogram is called dominant color structure histogram, the representative or dominant colors used are not the same as that defined in DCD. This is because we want to minimize the color distortion introduced while building color structure histogram using DCD's dominant colors. This is mainly due to the GLA color quantization and agglomerative clustering process in the DCD's color extraction described in previous section. In order to tackle this problem, the two processes for generating of the proposed dominant color structure histogram will be described below.

A. Color Quantization

The number of quantizing colors and the descriptor size are limited by the maximum number of clusters allowed in GLA splitting. To maintain the minimum distortion achieved by GLA, no agglomerative clustering is applied to the resulted clusters as in DCD. As a result, there is no restricted minimum distance between quantizing colors.

Given a set of quantizing colors $C_q = \{c_1, c_2, \cdots, c_n\}$ obtained from GLA, a closest quantizing color,

$$c_{\min} = c_i \big| \min(d(c_{xy}, c_i)); c_i \in C_q \qquad (6)$$

is found for an image pixel with color value c_{xy} at location *(x,y)*. The pixel is quantized to

$$c_{xy} = \begin{cases} c_{\min} & \text{if } d(c_{xy}, c_{\min}) < T_d \\ \text{null} & \text{otherwise} \end{cases} \qquad (7)$$

where $d(c_{xy}, c_{min})$ is the distance between the original color and closest dominant color. The threshold T_d can be in the range of 15~20 without significant difference and $T_d = 20$ is selected in our experiments. It should be noted that this threshold is used only for quantizing image. The dominant color set is not changed at all. The color of each pixel c_{xy} is quantized to the closest color $c_{\min}$ in the dominant colors set if the color distance $d(c_{ij}, c_{min})$ between the original pixel and the closest dominant color is smaller than a threshold T_d in the CIE*Luv* color space. Otherwise, the pixel will be ignored. Some pixels are ignored because GLA only minimizes the overall distortion but not the distance between dominant colors. They might not be similar to any dominant colors. Thus, they are not considered in the feature extraction as they are irrelevant with the dominant colors.

B. Structure Block Scanning

In the proposed method, color structure bins are formed by accumulation using an 8×8 structuring element. In structure block scanning, the structuring element scans through every pixel position of the image while keeping the entire structuring element within the image. A set of color bins is maintained for the scanning. At each pixel position, the colors existing in the region enclosed by the structuring element are recorded by incrementing the corresponding color bin count by one. A histogram will be formed with the bin values counting the times that the structuring element enclosed the colors in the scanning. The histogram will then be normalized by the number of pixel positions that the structuring element scanned. It should be noted that each bin is being normalized. It is different from conventional histogram that the histogram sum is being normalized. This leads to a significant difference in the interpretation of the bins. Each color structure histogram bin represents the area covered by the color in the image while, in the conventional histogram, each bin represents the number of occurrence of the color in the image. Figure 1 illustrates the DCSD extraction procedures.

From the extraction process described above, DCSD has significant difference from CSD and DCD. As CSD uses a fixed color space for counting the histogram, the color set of the histogram will be the same such that a simple *L1* distance can be used for similarity measure. In DCD, dominant colors are computed in extraction process. Each image may have a unique dominant color set. DCD uses a quadratic-like similarity measure, which is based on an assumption that the sum of the percentages is normalized. As a different normalization method is used in the structure histogram, it may not be effective to directly apply the original similarity measure to DCSD. Therefore, a new similarity measure for the structure histogram is proposed in the next section.

IV. Color Matching Palette Similarity Measure

The proposed descriptor, DCSD, is defined as

$$F = \{(c_i, q_i), s\} \qquad \text{for } i = \{1, 2, \cdots, N\}. \tag{8}$$

It consists of the number of dominant colors (N), color values (c_i), structure values of dominant colors (q_i), and with the optional spatial coherency parameter (s). The major difference between DCSD and DCD is that, in DCSD, q_i represents color structure bins instead of color percentage. Each q_i is normalized instead of the sum of q_i is normalized. It is done by a new extraction method proposed. DCSD inherits the compactness of DCD. For example, a DCD with eight colors requires 21.75 bytes in binary representation, and a DCD with 4 colors requires only 11.75 bytes. The most compact CSD, however, needs 32 bytes.

Similarity measure methods [5,6] were proposed for DCD perform better than the original QHDM similarity measure. As all these methods involve splitting or merging the colors and their percentages, they may not be suitable for DCSD. In DCSD, the percentages represent the color structure properties and the meaning of merging or splitting two or more color structure bins are fundamentally different from conventional color histogram. Thus, a new similarity measure, Color Matching Palette Similarity Measure (CMPSM), which uses color matching based on the color distance between colors, is developed for DCSD.

For each color c_i in a DCSD descriptor, a matching color with the minimum color distance with within T_d is searched in the color set of the other DCSD descriptor. The color distances for these minimum distance pairs are denoted as cd_i and they are added to a color matching palette. The color distance is measured by the Euclidian distance between the matched DCSD colors in CIE*Luv* color space. Another part of the palette is the structure differences sd_i defined as the absolute difference between the structure bins of the corresponding minimum distance pairs. Accordingly, the Color Matching Palette is given as

$$P_m = \{sd_i, cd_i\} \quad \text{for } i = 1, \dots, N. \tag{9}$$

In the matching procedure, the maximum matching color distance cd_{max} in cd_i is found. As some colors may not be matched for descriptors with different number of dominant colors, the unmatched colors are appended to the matching palette by setting sd_i equals as its original structure bin q_i, and cd_i as the maximum matching distance cd_{max}. The distance between two DCSD descriptors is defined as

$$D(F_1, F_2) = 1 - \sum_i \left[(1 - sd_i)\left(1 - \frac{cd_i}{\beta T_d}\right) \right] \tag{10}$$

where β is a parameter used to adjust the importance of color distance. Setting $\beta=\infty$ means that the color distance is not considered. And setting $\beta=1$, which is the minimum value of β means that the importance of color distance and color structure are equal. By trying with different values of β, a range of $2<\beta<5$ is suitable for matching most natural images. The result of $D(F_1,F_2)$ is between 0 and 1, where smaller values represents more similar between images.

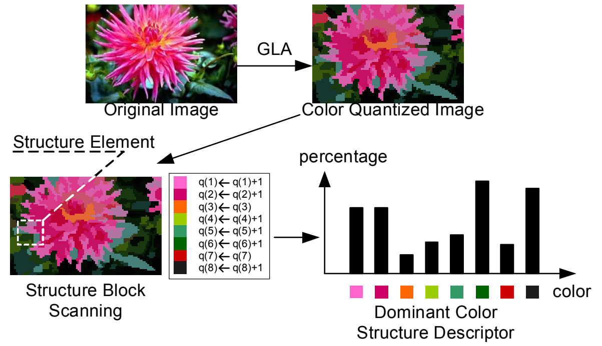

Fig. 1 Dominant Color Structure Descriptor extraction

V. Experimental results

The proposed method is implemented in a CBIR system, MIRROR [7]. The descriptor extraction and searching are based on MPEG-7 reference software XM with modifications. In our experiments, we use the Common Color Dataset (CCD) in MPEG-7 core experiments [8]. The dataset contains 5466 images and a common ground truth set with 50 sample queries defined.

Averaged Normalized Modified Retrieval Rate (ANMRR) and Normalized Modified Retrieval Rate (NMRR) [9] are used to measure the performance of retrieval. To measure the performance of the proposed method, several configurations of DCSD with 4 to 8 colors with $\beta=5$, $T_d=20$ is used in the experiments. The proposed DCSD is compared with the color descriptors included in the current XM 6.1 including Dominant Color Descriptor with the original similarity measure (DCD-QHDM), Color Structure Descriptor (CSD), Scalable Color Descriptor (SCD) and Color Layout Descriptor (CLD). In addition we also compared with the DCD with Merged Palette Histogram Similarity Measure (DCD-MPHSM) [5]. It should be noted that all DCD and DCSD results are with the optional spatial coherency parameter enabled.

A. Ultra Compact Configurations

Using ultra compact configurations, DCSD outperforms DCD in both retrieval accuracy and descriptor size. TABLE I shows that DCSD 4 and DCSD 5 have less than 16 bytes per descriptor. DCSD can give an acceptable retrieval result. Although DCSD can use even fewer colors to make the descriptor more compact, it is not possible to satisfactorily describe an image with such small number of colors. Although DCSD 4 gives ANMRR of 0.2371,which is not the best among the ultra compact descriptors and does not perform better than a more compact CLD, it already achieves a retrieval result very close to the original DCD having 0.2328 ANMRR but with a descriptor size smaller by 27.74%. DCSD 5 gives an ANMRR of 0.1878, which is one of the best retrieval rates among the ultra compact descriptors. DCSD 5 is also very competitive with the best ultra compact SCD configurations such as SCD(16,2), SCD(64,4), SCD(16,0), SCD(32,3), where, in SCD(x,y), x is the number of colors and y is the number of bit-planes discarded for each histogram bin. Based on the above analysis, CLD, DCSD 5, and SCD(16,2) are good choices for the ultra compact color descriptors.

B. Size-accuracy Configurations

From TABLE I, DCSD 8 gives an ANMRR of 0.0993, which is the best among the descriptors in this range. It outperforms the most accurate SCD(64,2) by 18.54% with descriptor size smaller by 25%. DCSD 8 also slightly outperforms CSD 32 by 2.55% with descriptor size substantially smaller by 32%. Some visual results comparison between DCSD 5, DCSD 8, DCD QHDM and CSD 32 are also given in Figure 2 in which DCSD 5 and DCSD 8 retrieved more visually similar images than CSD and DCD QHDM. TABLE I also shows that DCSD outperforms every configuration of SCD with the same descriptor size. For examples, DCSD 6, DCSD 7 and DCSD 8 outperform SCD(128,8), SCD(32,2) and SCD(128,4) by 6.78%, 14.65% and 19.73%, respectively. Generally, DCSD is more accurate than SCD with same descriptor size.

TABLE I: Results of MPEG-7 color descriptors

	Descriptor	Avg. Descriptor Size (bytes)	ANMRR
Ultra compact configurations	CLD	8.00	0.2264
	SCD(64,8)	8.75	0.2511
	SCD(16,3)	9.50	0.2108
	SCD(16,2)	11.50	0.1928
	DCSD 4	**11.75**	**0.2371**
	DCSD 5	**14.25**	**0.1878**
	SCD(64,4)	15.00	0.1816
	SCD(16,0)	15.50	0.1813
	SCD(32,3)	16.00	0.1826
Size-accuracy configurations	DCD-QHDM	16.26	0.2328
	DCD-MPHSM	16.26	0.2297
	SCD(128,8)	16.75	0.1648
	DCSD 6	**16.75**	**0.1536**
	SCD(32,2)	18.50	0.1481
	DCSD 7	**19.25**	**0.1264**
	SCD(64,3)	21.00	0.1583
	DCSD 8	**21.75**	**0.0993**
	SCD(128,4)	23.00	0.1237
	SCD(32,0)	26.50	0.1367
	SCD(64,2)	29.00	0.1219
	SCD(128,3)	31.00	0.1278
	CSD 32	32.38	0.1019
	SCD(256,8)	32.75	0.1516
	SCD(256,6)	33.75	0.1527

VI. Conclusions

An efficient color structure, dominant color structure histogram, for color image representation is described. A new color descriptor, Dominant Color Structure Descriptor (DCSD), based on the structure histogram is developed for color image retrieval. The descriptor inherits the compactness of Dominant Color Descriptor (DCD) and the retrieval accuracy of Color Structure Descriptor (CSD). The new similarity measure algorithm based on matching similar colors to generate a common palette, instead of using a fixed histogram space, is developed for DCSD. Experiment results show that DCSD gives better retrieval performance than the original DCD. Moreover, the effective configuration, DCSD 8, outperforms the most accurate SCD and CSD configurations with even smaller descriptor size. With its compactness and accuracy properties, DCSD is highly competitive as compared with the current color descriptors in a wide range of descriptor sizes.

Acknowledgment

The work described in this paper was substantially supported by a grant from City University of Hong Kong, Hong Kong SAR, China. [Project No.7001853].

References

[1] Special issue on MPEG-7, IEEE Trans. Circuits and Systems for Video Technology, vol.11, no. 6, June 2001.

[2] B.S. Manjunath, J. R. Ohm, V. V. Vasudevan, A. Yamada, "Color and texture descriptors," IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 6, 703 -715, June 2001.

[3] T. Ojala, M. Aittola, and E. Matinmikko, "Empirical evaluation of MPEG-7 XM color descriptors in content-based retrieval of semantic image categories," in Proc. of Int'l Conf. on Pattern Recognition, vol. 2, pp. 1021-1024, August 2002.

Fig. 2a Reference image and ground truth set of query #35

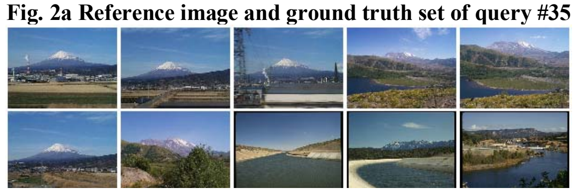

Fig. 2b Visual result of DCSD 5, NMRR = 0.0286, 4 of 4 ground truth images found in first 10 retrieval.

Fig. 2c Visual result of DCSD 8, NMRR = 0.0714, 4 of 4 ground truth images found in first 10 retrieval.

Fig. 2d Visual result of DCD QHDM, NMRR = 0.1857, 3 of 4 ground truth images found in first 10 retrieval.

Fig. 2e Visual result of CSD 32, NMRR = 0.2429, 3 of 4 ground truth images found in first 10 retrieval.

[4] A. Gersho and R.M. Gray, "Vector Quantization and Signal Compression," Kluwer Academic Publishers, 1992.

[5] L. M. Po and K. M. Wong, "A new palette histogram similarity measure for MPEG-7 Dominant Color Descriptor," in Proc. of IEEE Int'l Conf. on Image Processing 2004, vol. 3 pp. 1533-1536. 24-27 Oct. 2004, Singapore.

[6] J. Chen, T. N. Pappas, A. Mojsilović and B. E. Rogowitz, "Adaptive perceptual color-texture image segmentation," IEEE Trans. on Image Processing, vol. 14, no. 10, October 2005. pp. 1524-1536.

[7] K. M. Wong, K. W. Cheung and L. M. Po, "MIRROR: an interactive content based image retrieval system," in Proc. of IEEE Int'l Symposium on Circuit and Systems 2005, vol. 2, pp. 1541-1544. 23-26 May 2005, Kobe, Japan. (http://www.ee.cityu.edu.hk/~mirror)

[8] D. Zier and J. R. Ohm, "Common Data Sets and Queries in MPEG-7 Color Core Experiments," ISO/IEC JTC1/SC29/WG11 Doc. M5060, October 1999.

[9] P. Ndjuki-Nya et al., "Subjective Evaluation of the MPEG-7 Retrieval Accuracy Measure (ANMRR)," ISO/WG11, Doc. M6029, May 2000.