13.6 COLOR STRUCTURE DESCRIPTOR

The color structure descriptor (CSD) represents an image by both the color distribution of the image (similar to a color histogram) and the local spatial structure of the color. The additional information about color structure makes the descriptor sensitive to certain image features to which the color histogram is blind. Figure 13.7 illustrates this with a pair of images, each of which consists of two *iso-color planes*[3], one grey and one black. The grey iso-color plane on the left is highly structured, whereas the one on the right is less so. The *structure* of an iso-color plane is the degree to which its pixels are clumped together relative to the scale of an associated structuring element.

Each image contains exactly 50 pixels in its grey plane and 250 pixels in its black plane. Hence, they are indistinguishable, based solely on the information in their two-bin color histograms. But their two-bin CS Descriptors are very different and thus the images can be easily distinguished in an indexing or retrieval application based on the CSD.

The CSD is identical in form to a color histogram but is semantically different. Specifically, the CSD is a 1D array of eight bit−quantized values,

$$\text{CSD} = \overline{h}_s(m), \quad m \in \{1, \ldots, M\}$$

where M is chosen from the set {256, 128, 64, 32} and where s is the scale of the associated square structuring element. In the example of Figure 13.7, $s = 3^2$. The M bins (array elements) of $\overline{h}_s$ are associated in a bijective[4] manner to the M cells of the nonuniformly quantized HMMD color space (see Section 13.2).

13.6.1 CSD Interoperability

Descriptor interoperability was discussed generally in Section 13.1. There is, however, an aspect of interoperability that is peculiar to the CSD. In retrieval applications it may be

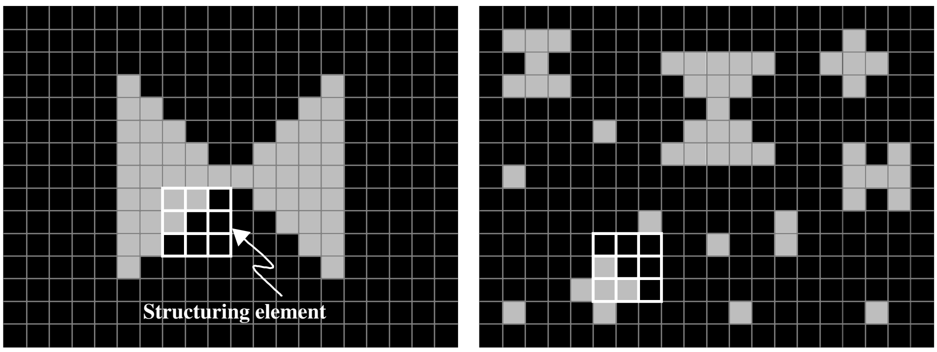


Figure 13.7 Two iso-color planes with differing amounts of structure

[3] An image quantized to N colors is composed of N iso-color planes. The n-th plane is the set of all pixels having the n-th quantized color, $n \in \{1, \ldots, N\}$.

[4] A map, $f: A \to B$, is said to be *bijective* if f maps set A onto set B in a one-to-one manner.

the case that a query descriptor presented (e.g. via the Web) to a remote search engine has a length that differs from the descriptors in the database. In order to compute the similarity between query and database descriptors, the lengths must be equalized.

Now in the case of color histograms, a length N histogram can be obtained either

- By extracting it directly from an image quantized to N colors or
- By extracting from the same image, quantized to $M > N$ colors, a length M histogram and then unifying (summing) appropriate subsets of its bin values to form the N-bin histogram.

Either method results in the *same* histogram as long as a *scalability condition* is met by the quantized color space as discussed in [13].

The CSD does not enjoy this property because the color quantization of an image affects its color structure. The reader is directed to [13] for a discussion of the somewhat subtle reason for this. The salient point is that a CS Descriptor obtained from image I by one scheme will, in general, lead to different retrieval results than a CSD from I by the other scheme. That is, the two extraction or resizing methods are not interoperable.

Consequently, and in contrast to most other MPEG-7 visual descriptors, extraction and resizing of the CSD is a *normative* process within the standard, by which we mean that the major steps are specified by the standard. Deviation from these steps risks breaking the interoperability of the descriptor.

13.6.2 Extraction

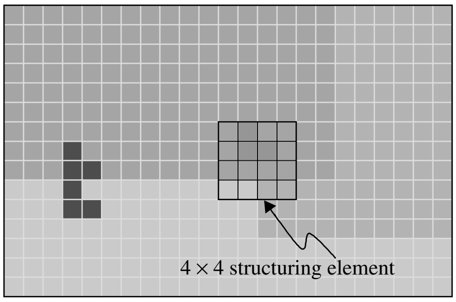
The CSD is best understood in terms of the *Color Structure Histogram*, h_s, upon which $\overline{h}_s$ is based. Extraction of a CSD is a three-step process:

1. A 256-bin CS Histogram is extracted (i.e. accumulated) from an image represented in the 256 cell–quantized HMMD color space. If the image is in another color space, then it must be converted to HMMD and requantized prior to extraction.
2. If $N < 256$ is desired, then bins are unified to obtain a N-bin CS Histogram.
3. The values (amplitudes) of each of the N bins are nonlinearly quantized in accordance with the statistics of color occurrence in typical consumer imagery.

We now discuss these steps in more detail, a full description of which are given in [5] and [14].

Accumulation of CS Histogram

In the context of the CSD, the length and color space of the CS Histogram are fixed. Outside this context, however, the CS Histogram can, in general, be of any length and can be accumulated from an image represented in any quantized color space. The procedure is depicted in Figure 13.8 where a simple five-color 'image' is shown together with a 4×4 structuring element. Also shown in tabular form on the right is an eight-bin CS Histogram,



Color	Bin value
c_1	h (1) + 1
c_2	h (2)
c_3	h (3)
c_4	h (4) + 1
c_5	h (5)
c_6	h (6) + 1
c_7	h (7) + 1
c_8	h (8)

Figure 13.8 (see also Plate 3) Accumulation of CS Histogram

$h_s(m)$, whose bins are associated with eight quantized colors, c_m, $m \in \{1, \ldots, 8\}$, in which the image is represented.

In nominal operation, the structuring element scans the image such that

- The element visits every position in the pixel grid and
- The element always lies entirely within the image.

At each position, the CS Histogram is updated on the basis of the colors present within the element. The operation is illustrated in Figure 13.8 where, in its current position, four colors are present within the structuring element. Therefore, each of the four corresponding bins of the CS Histogram is incremented by one. Observe that in any given position, the increase in $h(m)$ is determined by whether color c_m is present or absent within the element rather than by how much of c_m is enclosed. Hence, the final value of $h(m)$ is determined (up to normalization) by the number of positions at which the structuring element contains c_m.

It is interesting to note that the CS Histogram may be viewed as a *generalized* color histogram since it reduces precisely to an ordinary color histogram when a 1×1 structuring element is used.

Although a 4×4 element is shown in Figure 13.8, the MPEG-7 Standard defines the scale to be 8×8. This was determined by experiment to be the optimal scale. In conjunction with this, the Standard calls for images that deviate from a nominal size to be uniformly subsampled, both horizontally and vertically, in order to reduce the computational load. The subsampling factor is given by $K = 2^p$ where

$$p = \max\{0, \lfloor \log_2 \sqrt{W \cdot H} - 7.5 \rfloor\}$$

where W and H are the picture width and height, respectively, and where $\lfloor \cdot \rfloor$ is the *floor* operator. The reader is directed to References 5 and 14 for an equivalent formulation in which the accumulation requires no explicit subsampling of the image.

The CS Histogram (and hence the CSD) can be extracted from arbitrarily shaped, possibly disconnected, regions of an image. This is done in practice by means of a binary mask that defines the regions. Movement of the structuring element is as above (i.e. over the entire extent of the image), but the histogram is accumulated only with pixels that lie in the transparent portions of the mask.

Bin unification

When a CSD of length $N \in \{128, 64, 32\}$ is required, the 256-bin CS Histogram is reduced in length by bin unification. This process adds the values in each of N disjoint subsets of bins from the full-length histogram to form the N bins of the shorter histogram.

We now describe the procedure for a general size reduction from M to $N < M$ bins. For the case at hand, one merely lets $M = 256$ and $N \in \{128, 64, 32\}$. Let $P = \{p_1, \ldots, p_M\}$ and $Q = \{q1, \ldots, qN\}$ be two scalable quantizations of a color space, S, where the p_m and q_n are the individual cells of the two quantizations and where $M > N$. Quantization *scalability* is equivalent to the conditions:

$$\bigcup_{m=1}^{M} p_m = S = \bigcup_{n=1}^{N} q_m$$

for each n, $$q_n = \bigcup_{m \in J_n} p_m, \quad \text{where } J_n = n_1, \ldots, n_{k_n}$$

$$J_i \cap J_j = \emptyset \quad \text{for } i \neq j$$

The first condition insures that both P and$Q$ cover the space, S. The second condition implicitly defines the index subsets, J_n. The third condition is a consequence of the fact that quantization cells are, themselves, disjoint. Hence, it is redundant, following from the second condition. We include it for clarity.

In light of the bijection between bins and color space cells, the bin unification is defined by

$$h_s^N(n) = \sum_{m \in J_n} h_s^M(m), \quad n \in \{1, \ldots, N\} \tag{1}$$

where the superscripts denote the respective histogram lengths. The index subsets, J_n, for reducing a 256 CS Histogram length to a shorter length can be derived from the four scalable quantizations of the HMMD color space defined in Section 13.2.2.

Bin value quantization

The final step in extracting an N-bin CSD is to normalize to the range [0, 1) the bin values (amplitudes) of the N-bin CS Histogram from the preceding step, and then to nonlinearly quantize the normalized values to eight bits according to the quantization table in [5]. The nonlinear quantization was derived using several heuristics and experiments conducted with the CCD and it dramatically increases the retrieval accuracy of the CSD. The chief effect of the nonlinearity is to give the small values greater weight in the similarity measure than they would otherwise have.

13.6.3 CSD Resizing

The extraction procedure of Section 13.6.2 insures that lengths, say N and $M > N$, of two different length CSDs can always be equalized. The resizing procedure adjusts the longer descriptor to match that of the shorter. First, the bin values must be de-quantized so that *linear* values participate in the bin unification. Next, the M bins are unified just as discussed for the general case in Section 13.6.2. Finally, bin values of the new N-bin histogram are nonlinearly requantized to obtain the desired N-bin CSD. It can be shown that this resizing process gives the same result as having extracted an N-bin CSD in the first place.

13.6.4 Retrieval Results

As with other histogram descriptors, the CSD uses the l_1-norm for matching in its Similarity Measure. The CCD was modified by the addition of a few more query images to further differentiate the retrieval performance between the CSD and SCDs. Table 13.6(a) shows CSD retrieval accuracy for the four lengths defined by the standard. The longest descriptors yield the best results.

To motivate the choice of the nonuniformly quantized HMMD color space, Table 13.6(b) lists the retrieval results in the case in which the CS histograms were extracted in the HSV color space followed by nonlinear bin value quantization. The (uniform) color space quantization of the HSV space for each descriptor length is shown in the second column of Table 13.6(b). A comparison of the results in the two tables clearly shows the performance gained by using the nonuniformly quantized HMMD color space.

13.7 COLOR LAYOUT DESCRIPTOR

The color layout descriptor (CLD) is a very compact and resolution-invariant representation of color for high-speed image retrieval. It is designed to efficiently represent spatial distribution of colors. This feature can be used for a wide variety of similarity-based retrieval, content filtering and visualization. It is especially useful for spatial structure–based retrieval applications, for example, sketch-based retrieval and video segment identification. The sketch-based retrieval is considered to be a very important functionality since it can offer very user-friendly interfaces, especially when the search is fast enough.

Table 13.6 CSD retrieval results using a) HMMD color space and b) HSV color space

a)

Descriptor size	ANMRR
256 bins	0.06799
128 bins	0.07613
64 bins	0.09374
32 bins	0.14438

b)

Descriptor size	H × S × V quant.	ANMRR
256 bins	16 × 4 × 4	0.08707
128 bins	8 × 4 × 4	0.09204
64 bins	8 × 2 × 4	0.10700
32 bins	8 × 2 × 2	0.14832