ABSTRACT—Query-by-example (QBE) is a well-known method for image retrieval. In reality, however, an example image to be used for the query is rarely available. Therefore, it is often necessary to find a good example image to be used for the query before applying the QBE method. Query-by-layout (QBL) is our proposal for that purpose. In particular, we make use of the visual descriptors such as the edge histogram descriptor (EHD) and the color layout descriptor (CLD) in MPEG-7. Since image features of the CLD and the EHD can be localized in terms of a $4 \times 4$ sub-image, we can specify image features such as color and edge distribution on each sub-image separately for image retrieval without a query image. Experimental results show that the proposed query method can be used to retrieve a good image as a starting point for further QBE-based image retrieval.

Keywords—Content-based image retrieval, MPEG-7, image query, edge histogram descriptor, color layout descriptor.

I. Introduction

In MPEG-7 a set of visual descriptors are defined for multimedia contents description interface [1]. The international standard includes MPEG-7 visual descriptors, which can provide a rich and powerful content description technology to retrieve similar images using the query-by-example (QBE) framework. In reality, however, finding a good example image for the QBE system can be a challenging job. Moreover, since MPEG-7 has no standard query input format to capture the user’s search criteria, the problem becomes even more difficult when users want to issue the same single search query to multiple search engines [2]. The problems of polysemy and the unwanted connotations of the test images also make the problem of finding proper example images so difficult. To overcome the above limitations, a new procedure called MP7QF (MPEG-7 query format) was launched at the 77th MPEG meeting [3].

In this letter, we introduce a new query-by-layout (QBL) framework which allows users to specify edge and color layout in terms of sub-images. This overcomes the drawbacks of the QBE framework. To this end, two MPEG-7 descriptors, namely, the edge histogram descriptor (EHD) and the color layout descriptor (CLD), are employed. To the best of our knowledge, this is the first QBL-based natural image retrieval system which is compliant with the MPEG-7 standard.

II. Sub-image Based Manipulation of EHD and CLD

1. Edge Histogram Descriptor

The MPEG-7 EHD visual descriptor [4] represents the distribution of 5 edge types, namely vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-edge. To generate the EHD, an image is divided into non-overlapping $4 \times 4$ sub-images. Then, each sub-image serves a basic region to generate an edge histogram, which consists of 5 bins with vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edge types. Since there are 16 ($4 \times 4$) sub-images, each image yields an edge histogram with a total of 80 ($16 \times 5$) bins. These 80 normalized and quantized bins constitute the standardized EHD of MPEG-7. Specifically, arranging edge histograms for the sub-images in raster scan order (that is, $(0, 0), \ldots, (0, 3), (1, 0), \ldots, (3, 3)$), 16 local histograms are concatenated to have an integrated histogram with $16 \times 5 = 80$ bins. We denote $E_i(k)$ as a normalized and
quantized bin value for a sub-image at \((i, j) \in \Omega\), where 
\[ \Omega = \{(i, j) | 0 \leq i \leq 3, 0 \leq j \leq 3\} \] 
is a set of indices for sub-images and \(k \in \{1, \cdots, 5\} \) indicates one of 5 edge types.

2. Color Layout Descriptor

The MPEG-7 CLD visual descriptor [5] is designed to represent the spatial distribution of the color features in an image. To satisfy this requirement, the CLD is obtained by applying the discrete cosine transform (DCT) in a 2-D image space. Specifically, the image space is first divided into \(8 \times 8\) non-overlapping blocks and a representative color for each block is determined. Adopting the luma component \(Y\) and the blue and red chroma components \(Cb\) and \(Cr\), three \(8 \times 8\) representative color components are applied to obtain three \(8 \times 8\) DCT components of \(Y-Cb-Cr\). Then, a set of low frequency DCT components for each \(Y-Cb-Cr\) plane are selected and quantized for the standardized CLD. Note that to make a spatial correspondence with the \(4 \times 4\) grid structure of the EHD, we further take the \(8 \times 8\) inverse DCT (IDCT) to the CLD to obtain the localized \(8 \times 8\) color values. Then, these \(8 \times 8\) IDCT values can be grouped together such that one sub-image of the EHD spatially overlays \(2 \times 2\) of the \(8 \times 8\) IDCT values. Then, the color information as well as the edge histogram can be treated independently for each sub-image. Now, we denote \(Y_{\theta}, Cb_{\theta},\) and \(Cr_{\theta}\) as the IDCT values for \(Y, Cb,\) and \(Cr\) components at the DCT block \((i, j)\), respectively, where \(0 \leq i \leq 7\) and \(0 \leq j \leq 7\).

III. Query Layout by Combining EHD and CLD

Since the EHD and the IDCT coefficients of the CLD are based on \(4 \times 4\) and \(8 \times 8\) grids, respectively, there exists a spatial one-to-one correspondence between the sub-image of the EHD and \(2 \times 2\) IDCT values of the CLD. That is, as shown in Fig. 1, the adjacent \(2 \times 2\) IDCT values correspond to one sub-image of the EHD. So, the visual features of each sub-image can be characterized by the \(2 \times 2\) representative colors of the IDCT of the CLD as well as the edge histogram of the EHD at that sub-image. Now, let us denote the density of the edges in the sub-image \((i, j) \in \Omega\) of the \(l\)-th test image as \(P_{\theta}^{EHD}(l)\):

\[ P_{\theta}^{EHD}(l) = \sum_{k=1}^{5} E_{\theta}^{(k)}(k). \]  
(1)

Since \(E_{\theta}^{(k)}(k)\) is normalized, \(0 \leq P_{\theta}^{EHD}(l) \leq 1\). Also, as \(P_{\theta}^{EHD}(l)\) approaches 0, we can expect less edges (more homogeneous/monotone regions) in the sub-image \((i, j) \in \Omega\). Conversely, as \(P_{\theta}^{EHD}(l)\) approaches 1, we can expect more edges (less homogeneous/monotone regions) in the sub-image \((i, j) \in \Omega\).

Note that it is possible to set separate parameters for different edge orientations instead of averaging the EHD bins as in (1). However, it is our intention to minimize the number of parameters (sliding bars) to control by using the edge density feature of (1). Similarly, for each sub-image we can define normalized IDCT values of \(Y, Cb,\) and \(Cr\) of the CLD as \(P_{\theta}^{Y}(l), P_{\theta}^{Cb}(l),\) and \(P_{\theta}^{Cr}(l)\), respectively, by grouping and normalizing \(2 \times 2\) IDCT values as follows:

\[ P_{\theta}^{Y}(l) = \frac{Y_{2i,2j} + Y_{2i+1,2j} + Y_{2i,2j+1} + Y_{2i+1,2j+1}}{4 \times 256}, \]  
(2)

\[ P_{\theta}^{Cb}(l) = \frac{Cb_{2i,2j} + Cb_{2i,2j+1} + Cb_{2i+1,2j} + Cb_{2i+1,2j+1}}{4 \times 256}, \]  
(3)

\[ P_{\theta}^{Cr}(l) = \frac{Cr_{2i,2j} + Cr_{2i,2j+1} + Cr_{2i+1,2j} + Cr_{2i+1,2j+1}}{4 \times 256}. \]  
(4)

Note that \(P_{\theta}^{Y}(l), P_{\theta}^{Cb}(l),\) and \(P_{\theta}^{Cr}(l)\) take values between 0 and 1. As these values approach 0, the sub-image becomes darker. On the other hand, the values close to 1 indicate brightness of the corresponding sub-image. Knowing the physical meanings of (1) to (4), the user can set the values of \(P_{\theta}^{EHD}(Q), P_{\theta}^{Y}(Q), P_{\theta}^{Cb}(Q),\) and \(P_{\theta}^{Cr}(Q)\) manually to provide a query layout, where \(Q\) represents the query. Determining these values for all 16 sub-images, we can specify the edge and the color layout for the query. For the similarity measure, \(S(l)\), between the query layout and the \(l\)-th test image, we can use the city block distance or the Euclidean distance

\[ S(l) = \sum_{i,j \in \Omega} \sqrt{\sum_{k=1}^{5} w_{\theta} P_{\theta}^{EHD}(Q) - P_{\theta}^{EHD}(l)^2}, \]  
(5)
where \( w_{ij}^{\text{EHD}}, w_{ij}^{\text{Y}}, w_{ij}^{\text{Cb}}, \) and \( w_{ij}^{\text{Cr}} \) are the weighting factors, taking values from 0 to 1. For example, if we want to exclude the color components from the query layout, then we can set \( w_{ij}^{\text{Cb}} = w_{ij}^{\text{Cr}} = 0 \). Note that those parameters including \( P_{ij}^{\text{EHD}}(Q) \), \( P_{ij}^{\text{Y}}(Q) \), \( P_{ij}^{\text{Cb}}(Q) \), \( P_{ij}^{\text{Cr}}(Q) \), \( w_{ij}^{\text{EHD}}, w_{ij}^{\text{Y}}, w_{ij}^{\text{Cb}}, \) and \( w_{ij}^{\text{Cr}} \) are determined for each sub-image by the user to form the query image layout. Calculating \( S(l) \) for all test images \( l = 1, \ldots, L \), we can find the best match \( l^* \), such that

\[
l^* = \arg \min_{l \in \{1, \ldots, L\}} S(l).
\]

Then, \( K \) best ranked images with (6) are retrieved for the given query layout.

IV. Experiments

We used 4409 natural images \((L = 4409)\) for our experiments. For all our experiments, we set \( w_{ij}^{\text{EHD}} = w_{ij}^{\text{Y}} = 1 \), and \( w_{ij}^{\text{Cb}} = w_{ij}^{\text{Cr}} = 0 \) for all \((i, j) \in \Omega\). That is, only the edge layout of the EHD and the luminance layout of the CLD were used for our query. In our QBL graphic user interface (GUI) board, we have sliding bars for each sub-image. As shown in upper-left area of Fig. 2, there are four sliding bars for each sub-image. They represent the edge popularity of the EHD, and the luminance, \( \text{Cb} \), and \( \text{Cr} \) strength of the CLD, respectively. The left end of each bar indicates 0 and the right end is 1. The status of the sliding bars is visualized for all 16 sub-images below the sliding bars in Fig. 2. That is, the sliding bar of the edge population is shown numerically in the center of each sub-image. The decimal point is omitted between the first and the second digit (i.e., 027 means 0.27). Also, the status of the luminance sliding bar is displayed by the corresponding brightness for each sub-image. For example, in Fig. 2(a), the query layouts for the second and the third columns are bright and homogenous, while the sub-images in the first and the fourth columns are dark and edge regions. In Fig. 2(b), the query layouts of all 16 sub-images are bright (the lower 8 sub-images are brightest). Also, those sub-images in the first column, the fourth column, and the first row are chosen to be homogeneous (no edge). As shown in the figures, retrieved images (ranked from top to bottom and left to right) are sufficiently compliant with the chosen query layouts, demonstrating the power of the proposed QBL system.

V. Conclusion

The QBL paradigm has been proposed for natural image retrieval. Our QBL system is MPEG-7 compliant, utilizing the EHD and CLD MPEG-7 visual descriptors. The inverse DCT of \( Y, \text{Cb}, \) and \( \text{Cr} \) values of the CLD and \( 4 \times 4 \) edge histograms of the EHD are used to characterize each of the \( 4 \times 4 \) sub-images exclusively. The experiments show that the layouts of the retrieved images match the query layouts quite well, enabling more content-oriented retrieval even without example images for the query.

References