Efficient Use of MPEG-7 Edge Histogram Descriptor

Chee Sun Won, Dong Kwon Park, and Soo-Jun Park

MPEG-7 Visual Standard specifies a set of descriptors that can be used to measure similarity in images or video. Among them, the Edge Histogram Descriptor describes edge distribution with a histogram based on local edge distribution in an image. Since the Edge Histogram Descriptor recommended for the MPEG-7 standard represents only local edge distribution in the image, the matching performance for image retrieval may not be satisfactory. This paper proposes the use of global and semi-local edge histograms generated directly from the local histogram bins to increase the matching performance. Then, the global, semi-global, and local histograms of images are combined to measure the image similarity and are compared with the MPEG-7 descriptor of the local-only histogram. Since we exploit the absolute location of the edge in the image as well as its global composition, the proposed matching method can retrieve semantically similar images. Experiments on MPEG-7 test images show that the proposed method yields better retrieval performance by an amount of 0.04 in ANMRR, which shows a significant difference in visual inspection.

Manuscript received Apr. 11, 2001; revised Oct. 24, 2001.

Chee Sun Won (phone: +82 2 2260 3337, e-mail: cswon@dongguk.edu) and Dong Kwon Park (e-mail: dkpark@dongguk.edu) are with the Electronics Engineering Department, Dongguk University, Seoul, Korea.

Soo-Jun Park (e-mail: psj@etri.re.kr) is with Knowledge Retrieval Technology Research Team, ETRI, Daejeon, Korea.

1. INTRODUCTION

The histogram is the most commonly used structure to represent any global feature composition of an image. It is invariant to image translation and rotation, and normalizing the histogram leads to scale invariance. Exploiting the above properties, the histogram is very useful for indexing and retrieving images [1], [2].

Edges in images constitute an important feature to represent their content. Also, human eyes are sensitive to edge features for image perception. One way of representing such an important edge feature is to use a histogram. An edge histogram in the image space represents the frequency and the directionality of the brightness changes in the image. It is a unique feature for images, which cannot be duplicated by a color histogram or the homogeneous texture features. To represent this unique feature, in MPEG-7, there is a descriptor for edge distribution in the image. This Edge Histogram Descriptor (EHD) proposed for MPEG-7 expresses only the local edge distribution in the image. That is, since it is important to keep the size of the descriptor as compact as possible for efficient storage of the metadata, the normative MPEG-7 edge histogram is designed to contain only 80 bins describing the local edge distribution. These 80 histogram bins are the only standardized semantics for the MPEG-7 EHD. However, using the local histogram bins only may not be sufficient to represent global features of the edge distribution. Thus, to improve the retrieval performance, we need global edge distribution as well. This paper describes how to generate the semi-global and global edge histograms from the local histogram bins. Then, the global, semi-global, and local histogram bins are used to evaluate the similarity between images.

We begin in Section II with the definition of the standardized semantics of the EHD. The algorithms for EHD extraction and matching, which are non-normative parts of the standard, are then discussed in Sections III and IV, respectively. Experimental results with 11639 natural images are shown in Section V. Finally, we conclude the paper in Section VI.

II. DEFINITION AND SEMANTICS OF THE EHD

The EHD basically represents the distribution of 5 types of edges in each local area called a sub-image. As shown in Fig. 1, the sub-image is defined by dividing the image space into 4×4 nonoverlapping blocks. Thus, the image partition always yields 16 equal-sized sub-images regardless of the size of the original image. To characterize the sub-image, we then generate a histogram of edge distribution for each sub-image. Edges in the sub-images are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges (Fig. 2). Thus, the histogram for each sub-image represents the relative frequency of occurrence of the 5 types of edges in the corresponding sub-image. As a result, as shown in Fig. 3, each local histogram contains 5 bins. Each bin corresponds to one of 5 edge types. Since there are 16 sub-images in the image, a total of 5×16=80 histogram bins is required, (Fig. 4). Note that each of the 80-histogram bins has its own semantics in terms of location and edge type. For example, the bin for the horizontal type edge in the sub-image located at (0,0) in Fig. 1 carries the information of the relative population of the horizontal edges in the top-left local region of the image.

The semantics of the 1-D histogram bins form the normative part of the MPEG-7 standard descriptor. Specifically, starting from the sub-image at (0,0) and ending at (3,3), 16 sub-images are visited in the raster scan order and corresponding local histogram bins are arranged accordingly. Within each sub-image, the edge types are arranged in the following order: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional. Table 1 summarizes the complete semantics for the EHD with 80 histogram bins. Of course, each histogram bin value should be normalized and quantized. For normalization, the number of edge occurrences for each bin is divided by the total number of image-blocks in the sub-image.

The image-block is a basic unit for extracting the edge information. That is, for each image-block, we determine whether there is at least an edge and which edge is predominant. When an edge exists, the predominant edge type among the 5 edge categories is also determined. Then, the histogram value of the corresponding edge bin increases by one. Otherwise, for the monotone region in the image, the image-block contains no

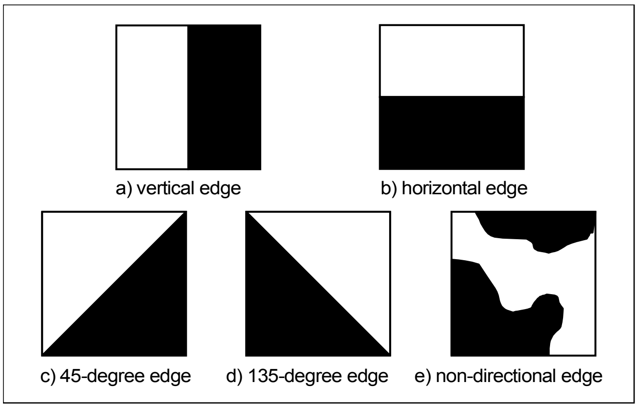

Fig. 2. Five types of edges.

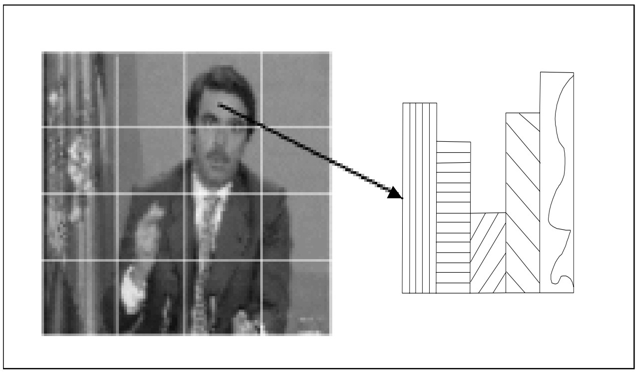

Fig. 3. Five types of edge bins for each sub-image.

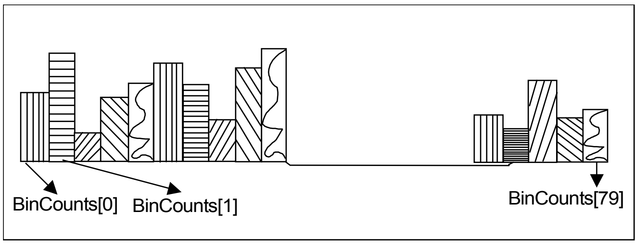

Fig. 4. 1-D array of 80 bins of EHD.

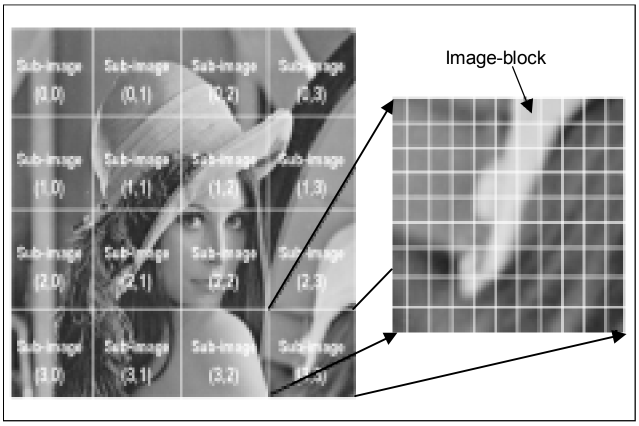

Fig. 1. Definition of sub-image and image-block.

Table 1. Semantics of local edge bins.

Histogram bins	Semantics
BinCounts[0]	Vertical edge of sub-image at (0,0)
BinCounts[1]	Horizontal edge of sub-image at (0,0)
BinCounts[2]	45-degree edge of sub-image at (0,0)
BinCounts[3]	135-degree edge of sub-image at (0,0)
BinCounts[4]	Non-directional edge of sub-image at (0,0)
BinCounts[5]	Vertical edge of sub-image at (0,1)
:	:
BinCounts[74]	Non-directional edge of sub-image at (3,2)
BinCounts[75]	Vertical edge of sub-image at (3,3)
BinCounts[76]	Horizontal edge of sub-image at (3,3)
BinCounts[77]	45-degree edge of sub-image at (3,3)
BinCounts[78]	135-degree edge of sub-image at (3,3)
BinCounts[79]	Non-directional edge of sub-image at (3,3)

edge. In this case, that particular image-block does not contribute to any of the 5 edge bins. Consequently, each image-block is classified into one of the 5 types of edge blocks or a nonedge block. Although the nonedge blocks do not contribute to any histogram bins, each histogram bin value is normalized by the total number of image-blocks including the nonedge blocks. This implies that the summation of all histogram bin values for each sub-image is less than or equal to 1. This, in turn, implies that the information regarding non-edge distribution in the sub-image (smoothness) is also indirectly considered in the EHD.

Now, the normalized bin values are quantized for binary representation. Since most of the values are concentrated within a small range (say, from 0 to 0.3), they are nonlinearly quantized to minimize the overall number of bits. Table 2 shows the representative values for coded bits for each edge type. The normalized 80 bin values are nonlinearly quantized and fixed length coded with 3bits/bin as defined in Table 2. *Bin Counts*[*0*], …, and *BinCounts*[*79*] (Table 1) represent the final coded bits for the EHD.

III. EHD EXTRACTION

Since the EHD describes the distribution of non-directional edges and nonedge cases as well as four directional edges, the edge extraction scheme should be based on the image-block as a basic unit for edge extraction rather than on the pixel. That is, to extract directional edge features, we need to define small square image-blocks in each sub-image as shown in Fig. 1. Specifically, we divide the image space into nonoverlapping square image-blocks and then extract the edge information from them. Note that, regardless of the image size, we divide the image space into a fixed number of image-blocks. The purpose of fixing the number of image-blocks is to cope with the different sizes (resolutions) of the images. That is, by fixing the number-of-image blocks, the size of the image-block becomes variable and is proportional to the size of the whole image. The size of the image-block is assumed to be a multiple of 2. Thus, it is sometimes necessary to ignore the outmost pixels in the image to satisfy that condition.

A simple method to extract an edge feature in the image-block is to apply digital filters in the spatial domain. To this end, we first divide the image-block into four sub-blocks as (Fig. 5). Then, by assigning labels for four sub-blocks from 0 to 3, we can represent the average gray levels for four sub-blocks at (i,j)th image-block as $a_0(i,j)$, $a_1(i,j)$, $a_2(i,j)$, and $a_3(i,j)$, respectively. Also, we can represent the filter coefficients for vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges as $f_v(k)$, $f_h(k)$, $f_{d-45}(k)$, $f_{d-135}(k)$, and $f_{nd}(k)$, respectively, where $k=0,\ldots,3$ represents the location of the sub-blocks. Now, the respective edge magnitudes $m_v(i,j)$, $m_h(i,j)$, $m_{d-45}(i,j)$, $m_{d-135}(i,j)$, and $m_{nd}(i,j)$ for the (i,j)th image-block can be obtained as follows:

Table 2. Quantization table for 5 types of edges.

BinCounts (3bits/bin)	Representative values for vertical edges	Representative values for horizontal edges	Representative values for 45-degree diagonal edges	Representative values for 135-degree diagonal edges	Representative values for non-directional edges
000	0.010867	0.012266	0.004193	0.004174	0.006778
001	0.057915	0.069934	0.025852	0.025924	0.051667
010	0.099526	0.125879	0.046860	0.046232	0.108650
011	0.144849	0.182307	0.068519	0.067163	0.166257
100	0.195573	0.243396	0.093286	0.089655	0.224226
101	0.260504	0.314563	0.123490	0.115391	0.285691
110	0.358031	0.411728	0.161505	0.151904	0.356375
111	0.530128	0.564319	0.228960	0.217745	0.450972

$$m_v(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_v(k) \right| \tag{1}$$

$$m_h(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_h(k) \right| \tag{2}$$

$$m_{\mathrm{d}-45}(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_{d-45}(k) \right| \tag{3}$$

$$m_{\mathrm{d}-135}(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_{d-135}(k) \right| \tag{4}$$

$$m_{nd}(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_{nd}(k) \right| \tag{5}$$

If the maximum value among 5 edge strengths obtained from (1) to (5) is greater than a threshold (T_{edge}) as in (6), then the image-block is considered to have the corresponding edge in it. Otherwise, the image-block contains no edge.

$$\max\{m_v(i,j), m_h(i,j), m_{d-45}(i,j), m_{d-135}(i,j), m_{nd}(i,j)\} > T_{edge} \tag{6}$$

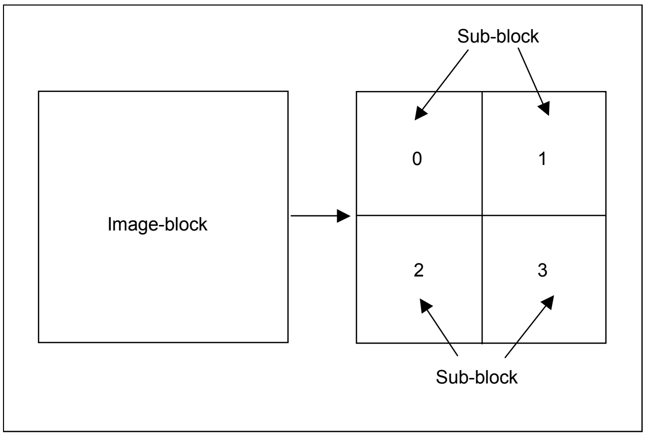

Fig. 5. Sub-blocks and their labeling.

In MPEG-7 XM Document [1], a set of filter coefficients, depicted in Fig. 6, is recommended. Note that the filter coefficients in Fig. 6, especially the non-directional edge filter, appear somewhat heuristic. In fact, the non-directional edges by definition do not have any specific directionality. So, it is hard to find filter coefficients that are applicable for all types of non-directional edges. To avoid this problem, we can first check whether the image-block can be classified into one of a monotone block and four directional edge blocks. If the image-block does not belong to any of the monotone or four directional edge blocks, then we classify it as a non-directional block. The flow chart of this method is shown in Fig. 7. Another edge ex-

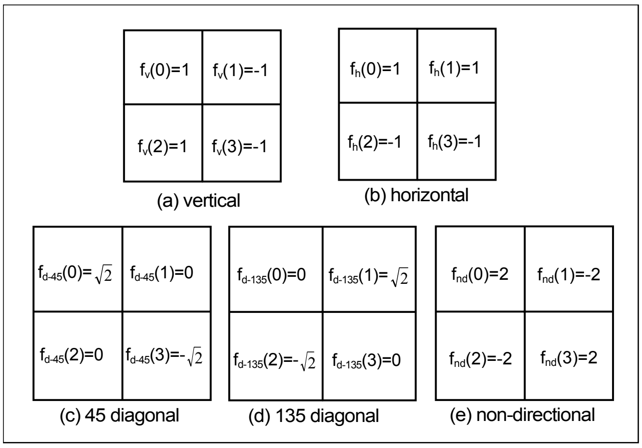

Fig. 6. Filter coefficients for edge detection.

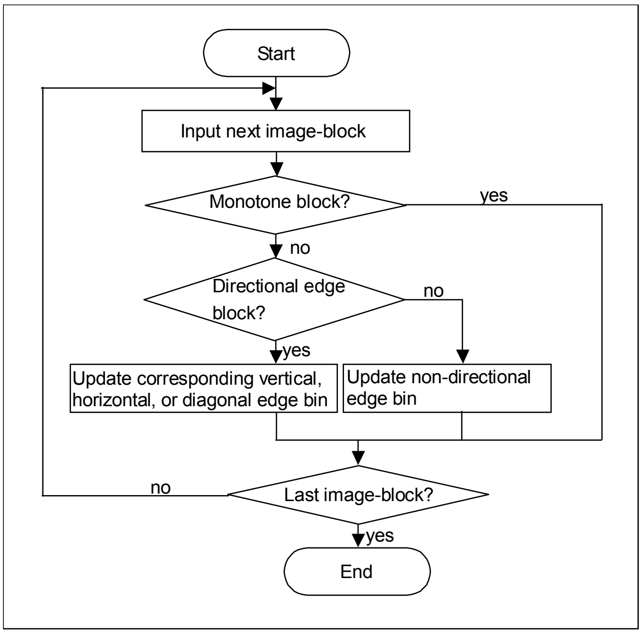

Fig. 7. Flowchart of edge classification without using the non-directional edge-filter.

traction method that follows this flow chart can be found in [4]. Here, the edge classification is based on the model-fitting criterion. For more detail, we refer the readers to the reference [4].

IV. SIMILARITY MATCHING

The 80 bins of the local edge histogram in Table 1 (i.e., *BinCounts*[*i*], *i=0,…79*) are the only normative semantics for the EHD. However, the local edge histograms alone may not be sufficient to yield efficient image matching. That is, we may need edge distribution information for the whole image space and some horizontal and vertical semi-global edge distributions as well as local ones. Fortunately, at the matching step, we can generate the global edge histogram and some semi-global edge histograms directly from *BinCounts*[*i*], *i=0,…79*. The global edge histogram represents the edge distribution for the whole image space. Since there are 5 edge types, the global edge histogram has 5 bins and each bin value is obtained by accumulating and normalizing the dequantized bin values of the corresponding edge type of *BinCounts*[]. Similarly, for the semi-global edge histograms, we can group some subsets of *BinCounts*[](Fig. 8). We define 13 different segments (i.e., 13 different subsets of the image-blocks) and for each segment we can generate edge distributions for five different edge types from the 80 local histogram bins. Consequently, we have a total of 150 bins (80 bins (local) + 5 bins (global) + 65 bins (13×5, semi-global)) for the similarity matching.

For the similarity matching, we calculate the distance *D(A,B)* of two image histograms *A* and *B* using the following measure:

$$D(A,B) = \sum_{i=0}^{79} \left| Local\_A[i] - Local\_B[i] \right| + 5 \times \sum_{i=0}^{4} \left| Global\_A[i] - Global\_B[i] \right| + \sum_{i=0}^{64} \left| Semi\_Global\_A[i] - Semi\_Global\_B[i] \right| \quad (7)$$

where *Local_A*[*i*] represents the reconstructed value of *BinCount*[*i*] of image A obtained by referring to Table 2. Similarly, *Local_B*[*i*] is the reconstructed value of *BinCount*[*i*] of image B. *Global_A*[] and *Global_B*[] represent the normalized bin values for the global edge histograms of image A and image B, respectively. Similarly, *Semi_Global_A*[] and *Semi_Global _B*[] represent the normalized histogram bin values for the semi-global edge histograms of image A and B, respectively. Since the number of bins of the global histogram is relatively smaller than that of local and semi-global histograms, a weighting factor of 5 is applied in (7).

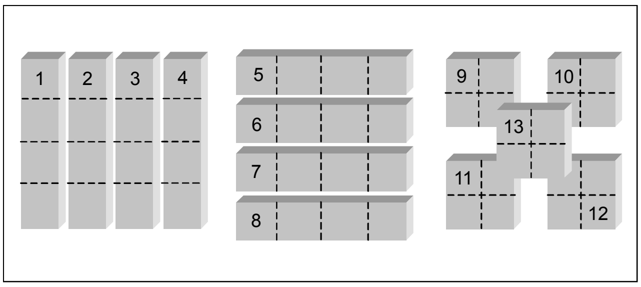

Fig. 8. Segments of sub-images for semi-global histograms.

V. EXPERIMENTS

For our experiments, we set the total number of image-blocks at 1100 and the threshold for edge detection (T_{edge}) at 11. For all our experiments, we used the image data set from the MPEG-7 Core Experiment (CE) [5], which has 11639 images in the database. Most of the images in the database are natural images from "Correl One Million Gallery" and others are from images provided by the proponents of other image descriptors such as color and homogeneous texture descriptors. From 11639 images, we used 51 images, which were selected by MPEG-7 CE participants [5], as query images. The ground truths that we used in our experiments were determined by three participants of the MPEG-7 CE. Each participant proposed from 3 to 33 ground truth images for each query image and they were approved by the other two CE participants [5]. As a measure of retrieval accuracy, we used the Average Normalized Modified Retrieval Rank (ANMRR) [6]. Precision and Recall are well-known measures for the retrieval performance. They are basically a "hit-and-miss" counter. That is, the performance is based on the number of retrieved images, which have similarity measures that are greater than a given threshold. For more specific comparisons, however, we also need rank information among the retrieved images. ANMRR is the measure that exploits the

Table 3. Retrieval performance (ANMRR).

	2bits/bin	3bits/bin	4bits/bin	5bits/bin
Matching with local-only histogram	0.396012	0.336060	0.317815	0.324698
Matching with local, semi-global and global histograms (proposed)	0.363593	0.296225	0.285961	0.284346

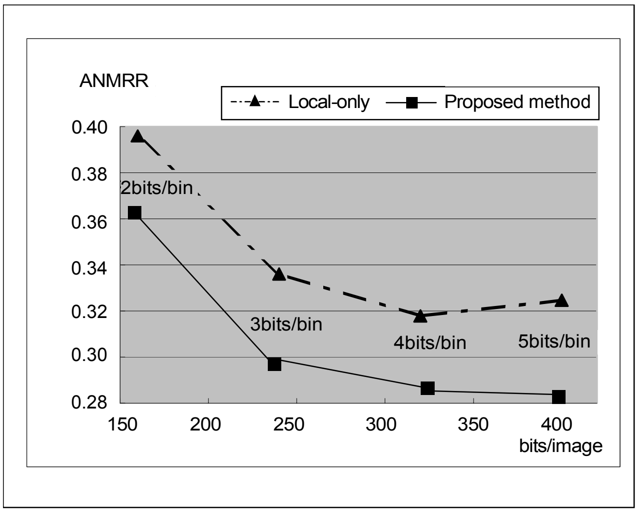

Fig. 9. Comparison between local histogram and proposed method.

Fig. 10. Retrieval results for query-by-“sphinx image.”

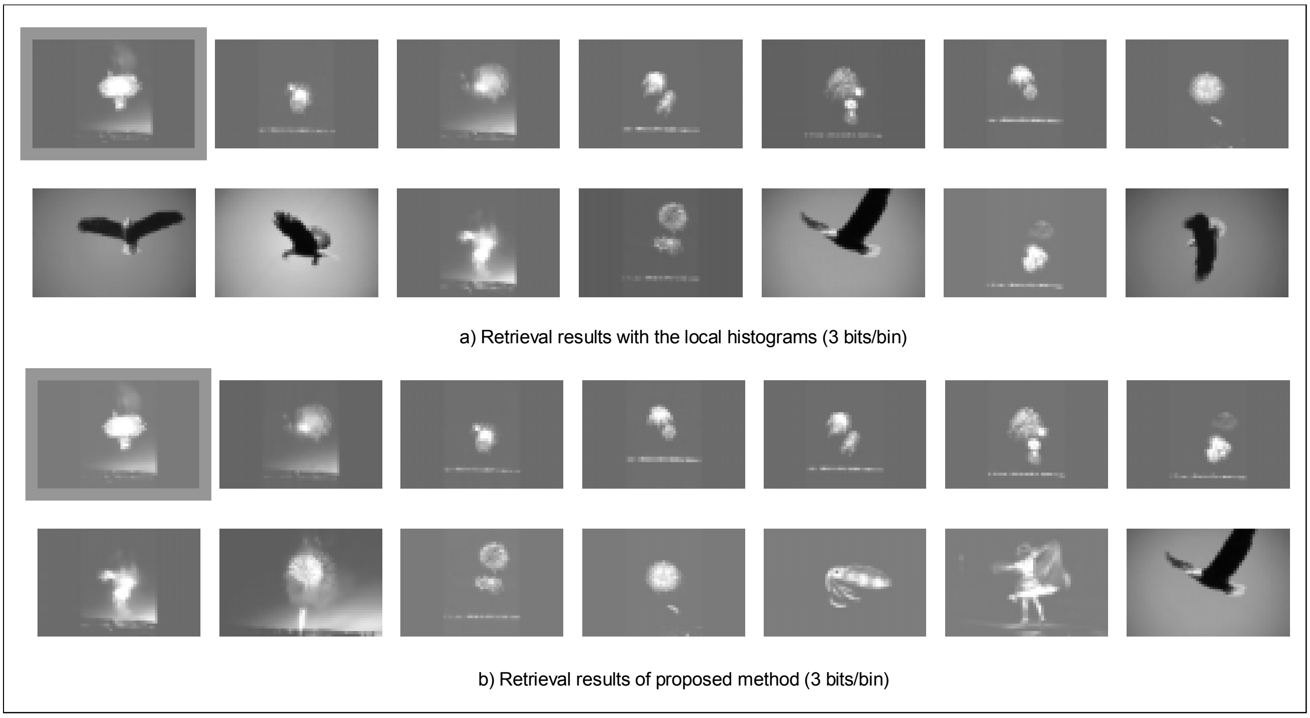

Fig. 11. Retrieval results for query-by-“firework image.”

rank of the retrieved images as well. It was developed during the MPEG-7 standardization activity and was used for the MPEG-7 Core Experiments (CE) [5]. Note that lower ANMRR values indicate more accurate retrieval performance.

Table 3 shows the results of the retrieval accuracy for different bits-per-bins numbers. Table 3 reveals that the proposed method with semi-global and global histograms yields significantly better retrieval performance. As the bits-per-bin increases the ANMRR decreases. However, Fig. 9 demonstrates that a further decrease in the ANMRR is not significant beyond

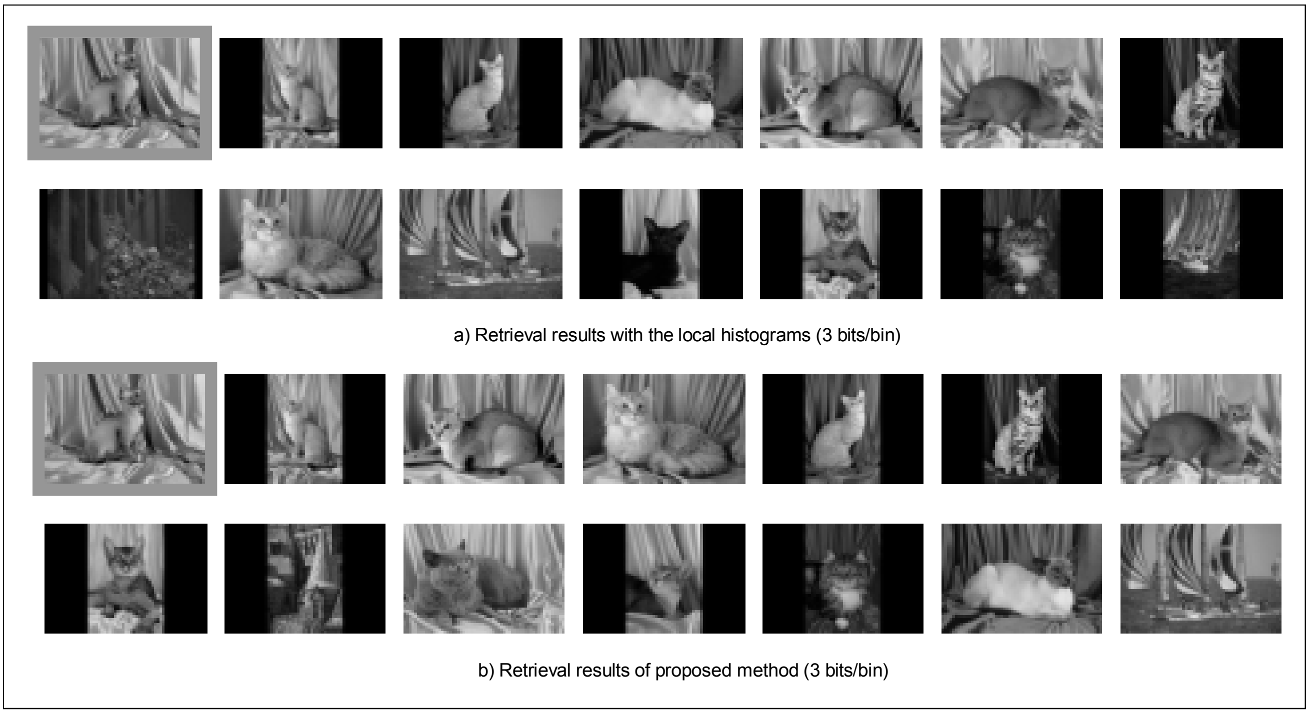

Fig. 12. Retrieval results for query-by-“cat image.”

3 bits-per-bin. This is why 3 bits-per-bin was chosen in MPEG-7.

Figures 10-12 demonstrate retrieval results for some query images. As you can see, the proposed method retrieves semantically more similar images. In Figs. 10-12, the left-upper solid-lined image is a query and its first ranked image. The other images are displayed in a raster scan order according to the retrieval ranks.

VI. CONCLUSIONS

The EHD is a compact descriptor with a fixed length. It always takes 240 bits per image. The EHD is also very flexible; it consists of local edge histograms only. However, using them, we can generate various patterns of semi-global edge histograms and a global edge histogram. In this paper, we show how to construct global and semi-global edge histogram bins from the local histogram bins. Among various possible clusters of sub-images, we used 13 patterns for the semi-global histograms. These 13 semi-global regions and the whole image space are adopted to define the semi-global and the global histograms, respectively. The extra histogram information can be obtained directly from the local histogram bins without an additional feature extraction process. Experimental results show that the semi-global and global histograms generated from the local histogram bins help to improve the EHD-based retrieval performances.

REFERENCES

[1] M.J. Swain and D.H. Ballard, “Color Indexing,” *Int. J. of Computer Vision*, vol.7-1, 1991, pp. 11-32.

[2] A.K. Jain and A. Vailaya, “Image Retrieval Using Color and Shape,” *Pattern Recognition*, vol. 29, no. 8, 1966, pp. 1233-1244.

[3] ISO/IEC/JTC1/SC29/WG11: “MPEG-7 XM Document: MPEG-7 Visual Part Experimentation Model Version 10.0,” *MPEG Document N4063*, Singapore, Mar. 2001.

[4] Chee Sun Won and Dong Kwon Park, “Image Block Classification and Variable Block Size Segmentation Using a Model-Fitting Criterion,” *Optical engineering*, Aug. 1997, pp. 2204-2209.

[5] ISO/IEC/JTC1/SC29/WG11: “Core Experiment Results for Spatial Intensity Descriptor (CT4),” *MPEG document M5374*, Maui, Dec. 1999.

[6] ISO/IEC/JTC1/SC29/WG11: “Description of Core Experiments for MPEG-7 Color/Texture Descriptors,” *MPEG document N2929*, Melbourne, Oct. 1999.

Chee Sun Won received his BS degree in electronics engineering from Korea University, Seoul, in 1982 and his MS and PhD degrees in electrical and computer engineering from the University of Massachusetts at Amherst in 1986 and 1990. From 1989 to 1992, he was a Senior Engineer with GoldStar Co., Ltd. (LG electronics) in Seoul. In 1992, he joined Dongguk University, where he is currently a Professor in the Electronics Engineering Department. He has been a Visiting Associate Professor at Stanford

University since July 2001. His research interests include MRF image modeling, image segmentation, content-based image retrieval, and image watermarking.

Dong Kwon Park was born in Republic of Korea on September 17, 1970. He received his BS and MS degrees in electronic engineering from Dongguk University in 1996 and 1998. He is currently a PhD student in the Department of Electronic Engineering at Dongguk University, Seoul, Korea. His areas of interest are image segmentation, retrieval, video browsing, and MPEG-7.

Soo-Jun Park was born in Korea on January 12, 1966. He received his BS degree in biochemistry from the University of Iowa, USA, in 1991. He received his MS degree in computer science from Lehigh University, USA, in 1994. He has been a Senior Researcher at Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, since 1994. His areas of interest are text and multimedia information retrieval and bioinformatics.