

Color Content Matching of MPEG-4 Video Objects

Berna Erol and Faouzi Kossentini

Department of Electrical and Computer Engineering, University of British Columbia
2356 Main Mall, Vancouver, BC, V6T 1Z4, Canada
e-mail: {bernae, faouzi}@ece.ubc.ca

Abstract. Color histogram is one of the most widely used visual feature representations in content-based retrieval. When processing the image/video data in the JPEG/MPEG compressed domains, the DC coefficients are used commonly to form the color histograms without fully decompressing the image or the video bit stream. In this paper, we address the issues arising from the adaptation of DC based color histograms to the arbitrarily shaped MPEG-4 video objects. More specifically, we discuss the color space selection, quantization, and histogram computation with the consideration of the specific characteristics of the MPEG-4 video objects. We also propose a method for reducing the chroma keying artifacts that may occur at the boundaries of the objects. The experimental results show that great retrieval performance improvements are achieved by employing the proposed method in the presence of such artifacts.

1. Introduction

Color histograms are commonly used for image and video retrieval, as they are relatively easy to extract, not much sensitive to noise, and invariant to image scaling, translation, and rotation. Because digital image and video is available mostly in the compressed formats, several researchers suggested methods for obtaining the color histograms from the coded bit stream without requiring full decompression and reconstruction of the visual data [1]. In most of the current image and video coding standards, such as JPEG, MPEG-1/2/4, and H.263, each frame is divided into 8x8 blocks, followed by DCT, quantization, zigzag scan, and run length coding. The quantized DC coefficient of each 8x8 block could be easily extracted from the bit stream by only performing parsing of the headers and run length decoding. In the intra coded frames, with a simple scaling, the DC coefficient is equal to the mean value of the corresponding block. Therefore, the DC coefficients of the Y , C_b , and C_r components can be simply employed to extract color features, including color histograms, as presented in the literature [1]-[3].

The most recent MPEG video coding standard, MPEG-4, supports the representation of arbitrarily shaped video by allowing the coding of the shape information of the video objects along with their texture. In this paper, we look into selection of color space, the number of quantization bins, and the histogram computation for the arbitrarily shaped MPEG-4 video objects, which generally have lower resolutions than that of frame based video and have consistent color throughout their lifespan.

Chroma keying is one of the most popular methods to obtain arbitrarily shaped video objects. If the video object shape is not accurately extracted prior the MPEG-4 encoding and/or the

MPEG-4 encoder does not employ the LPE padding technique described in the MPEG-4 verification model [4], the chroma key value of the background could contribute to some color artifacts that would eventually affect the color histogram of the video object. In this paper, we also propose a method to detect and compensate for such artifacts in order to obtain a more accurate color histogram representation.

The remainder of the paper is as follows. In the next section, we discuss the extraction of the DC coefficients from the MPEG-4 bit stream. In Section 3, we address the issues rising from the use of DC based color histograms to represent the MPEG-4 video objects. In Section 4, our proposed method for reducing the chroma keying artifacts in histogram computation is presented. Experimental results and conclusion are given in Section 5 and Section 6, respectively.

2. DC Coefficient Extraction in the MPEG-4 Bit Stream

Intra (I) frames are commonly used to obtain color histograms from the video sequences, as they are not predicted from any other frames. In MPEG-1/2 and H.263 I-frames, the DC coefficient can be obtained simply by parsing of the headers and run length decoding. On the other hand, in the MPEG-4 intra coded Video Object Planes (IVOPs), the reconstruction of the DC coefficient is required as the DCT coefficients of macroblocks can be predictively coded (either from the left or above block). After the DC coefficients are extracted from the MPEG-4 bit stream, reconstructed, and dequantized, the mean Y , C_b , and C_r values of the corresponding blocks are obtained by

$$M_Y = \frac{DC_Y}{8}, \quad M_{Cb} = \frac{DC_{Cb}}{8} - 128, \quad M_{Cr} = \frac{DC_{Cr}}{8} - 128,$$

where DC_Y , DC_{Cb} , and DC_{Cr} , are the DC coefficients of the luminance, chrominance b, and chrominance r blocks, respectively.

Parsing of the MPEG-4 bit stream in order to extract the DC information is also more complex than parsing of the MPEG-1/2 and H.263 bit streams. In MPEG-4, the shape information is placed before the texture information in the bit stream [4]. Therefore, arithmetic decoding of the shape is required before obtaining the DC coefficient. Nevertheless, reconstruction of the shape is not necessary.

3. Video Object Retrieval Using Color Histograms

Video objects, different than frame based video sequences, generally have low resolution, less variation in color, and their color content usually remains consistent unless there is occlusion by a large object or the video object is entering to or exiting from the scene. Therefore, a color histogram representation that is optimal for the frame based video is not necessarily optimal for the object based video.

Here, we address the problems of color space selection, quantization of colors, and histogram computation for arbitrary shaped video objects. In order to justify our particular color space and quantization parameter selections, we evaluate the retrieval results based on a measure used during the MPEG-7 standardization activities: Normalized Modified Retrieval Rank (NMRR) and Average NMRR (ANMRR). NMRR and ANMRR values are in the range of [0, 1] and the lower values represent a better retrieval rate. The specific formulas of these measures are given in Section 5.1.

3.1 Color Space and Histogram Size Selection

Video in MPEG-4 domain is represented in $YCbCr$ color space, as in MPEG-1/2. While $YCbCr$ representation is good for efficient compression, it is not a desirable representation in visual retrieval as it is not a perceptually uniform color space. HSV (Hue, Saturation, Value) color space, which is adopted for the MPEG-7 color histogram descriptor, more closely resembles to human perception, but it is also a non uniform space [5]. MTM (Mathematical Transformation to Munsell) is a perceptually uniform color space that very closely represents the human way of perceiving colors [5]. In the MTM space, the colors are represented by Hue (H), Value (V) and Chroma (C) components [6].

Table 1 presents the video object retrieval performance comparison employing three different color spaces. The results are obtained by querying MPEG-4 video object planes in a database of more than a thousand VOPs. Uniform quantization is employed to reduce the number of histogram bins. Employing $YCbCr$ color representation does not require conversion to another color space, however it gives the lowest retrieval performance. Using the MTM representation clearly offers a superior retrieval performance. Table 2 shows the retrieval results when different number of quantization bins used to represent the color components of the MTM space. As can be seen from the table, employing a 128-bin histogram offers the best tradeoff between the retrieval performance and the memory requirements.

query video object plane	HSV H:8 S:4 V:4	MTM H:8 V:4 C:4	$YCbCr$ Y:5 C_b :5 C_r :5
bream	0.0007	0.0004	0.0097
fish	0.0876	0.0249	0.2400
stefan	0.0116	0.0208	0.1303
singing girl	0.2686	0.2006	0.2715
Average NMRR	0.0912	0.0617	0.1629

Table 1. NMRR values obtained by querying the first video object planes of the various video objects, employing color histograms computed in three different color spaces.

query video object plane	256 bins: H:16 V:4 C:4	128 bins: H:8 V:4 C:4	64 bins: H:4 V:4 C:4	32 bins: H:4 V:2 C:2
bream	0.0003	0.0004	0.0728	0.0202
fish	0.0411	0.0249	0.0425	0.0194
stefan	0.0302	0.0208	0.1841	0.1900
singing girl	0.0666	0.2006	0.3349	0.3812
Average NMRR	0.0346	0.0617	0.1586	0.1527

Table 2. The retrieval performance results (in NMRR) for using different number of quantization bins for the H, V, and C components of the MTM color histograms.

3.2 Histogram Computation for Video Objects

We obtain the color histograms of individual VOPs by using only the color components that correspond to the blocks that are either completely inside (i.e., opaque) or on the boundary (i.e., intra) of the video object planes. This information is directly available in the MPEG-4 bit stream. In average, only half of the pixels in a boundary block lie in a video object. Therefore,

when computing the color histograms for individual VOPs, we count the color components of the boundary blocks as half of the color components of the opaque blocks.

After constructing the histograms for the individual VOPs, the video object histogram can be formed by using one of the following techniques used for frame based video [7].

- Average histogram: It is obtained by accumulating the histogram values over a range of frames and normalizing that by the number of frames.
- Median histogram: The bin values of this histogram is computed by taking the median values of each corresponding histogram bin of the individual frames.
- Intersection histogram: This histogram contains only the colors that are common to all the frames.

It is presented in the literature that employing average histogram yields the best results for frame based video retrieval [7]. Video object color generally remains consistent during its life-span, therefore, in most cases, an average histogram represents the video object color content accurately. Median histogram is most useful if some frames in a video sequence differ in color significantly than the others, which is usually not the case for video objects. Also, there is an increased computational cost associated with the median operation because of the sorting performed for each bin. Intersection histogram is also not very suitable for video object color representation: When the objects are entering to/exiting from a scene or when they are occluded, only a small part of their color range is visible, which would be the only colors represented by an intersection histogram. In conclusion, considering the characteristics of the arbitrarily shaped video objects, the average histogram is clearly the most appropriate choice to represent the color histogram of video objects.

Average histogram can be computed using the individual histograms of all the IVOPs in an MPEG-4 video object. A better alternative that reduces the computational requirements would be computing the histogram on a temporally sampled subset of IVOPs or on key VOPs that represent the salient content of the video object [8].

4. Compensation for the Chroma Keying Artifacts

Chroma keying remains one of the most popular methods to obtain semantic video objects. In chroma keying, the foreground object is separated from the background by placing the object in front of a color screen that has a unique chroma key value (typically blue or green) and defining the pixels that belong to the screen as outside the video object. Ideally, the coded video object should not contain any pixels from the background. However, if an MPEG-4 encoder does not approximate the video object shape very accurately and/or it does not perform low pass extrapolation (LPE) padding technique prior to DCT, which is defined in the MPEG-4 verification model [4] but not part of the MPEG-4 standard, the boundary blocks of the video object could contain some severe chroma keying artifacts. These artifacts could result in the chroma DC values (DC_{Cb} and DC_{Cr}) of the boundary blocks include the chroma key color along with the actual video object color, resulting in an inaccurate computation of the color histogram. In order to overcome this problem, we propose to first detect the existence of such artifacts and then compensate for them accordingly.

Our experiments show that, if a video object plane has any chroma artifacts, it is likely to affect all the blocks on the video object boundary. Therefore, if such effects are detected in one or several boundary blocks, it is reasonable to assume that the most of the boundary blocks of the video object have such artifacts. We propose to detect the chroma artifacts at the decoder, assuming no apriori information about the encoder, by decoding the texture and the shape of the first boundary block of the video object plane and then computing the mean chroma values (C_b and C_r) for the pixels that are inside and outside the video object area using the shape mask for

that particular block. If the difference between the chroma values corresponding to the inside and outside of the video object is very small, than it could be concluded that the segmentation was done properly and the LPE technique was employed prior to encoding. Therefore, the DC values of the boundary blocks correctly reflect the real video object color and no further processing is required. However, if the inside and outside chroma mean values differ significantly, then we define the chroma key values (K_{Cb} , K_{Cr}) as equal to the mean chroma values of the outside pixels.

After chroma keying artifacts are detected and the chroma key values are determined, then the scaled DC coefficients (M_{Cb} and M_{Cr}) of the boundary blocks are adjusted to reduce the chroma artifacts. Considering that, in average, half of the pixels in a boundary belongs to the inside the object and the other half belongs to the outside the object, the following approximations can be made to find actual mean value of the pixels inside the video object.

$$M_{Cb} \approx \frac{V_{Cb} + K_{Cb}}{2}, M_{Cr} \approx \frac{V_{Cr} + K_{Cr}}{2} \Rightarrow V_{Cb} \approx 2M_{Cb} - K_{Cb}, V_{Cr} \approx 2M_{Cr} - K_{Cr},$$

where the M_{Cb} and M_{Cr} are the scaled chrominance DC coefficients extracted from the bit stream, K_{Cb} and K_{Cr} are the approximated chroma key values, and V_{Cb} and V_{Cr} are the mean chrominance values of the pixels that belong to the video object in a boundary block. Video object color histogram is computed by using the approximated V_{Cb} and V_{Cr} values of the chrominance components, along with the unmodified luminance component, M_Y .

5. Experimental Results

Here, we demonstrate the performance of our proposed technique in the presence of chroma keying artifacts. We present retrieval results for some individual VOPs as well as some video objects. Our database consists of over 20 arbitrarily shaped video objects, coded in 2 to 3 different spatial resolutions each, resulting in an MPEG-4 database of over 50 bit streams and over 1500 intra coded VOPs. We utilize the MTM color space and 128-bin uniform quantization for the color histograms of the VOPs. Video object histograms are formed by histogram averaging on their key VOPs. The key VOPs are found by the algorithm described in [8]. The color histogram distances between two VOs or two VOPs are computed using the L1 norm, which was demonstrated to have a superior performance for measuring the histogram distances [7][9].

5.1 Performance Evaluation Criteria

We present our retrieval results by utilizing the Normalized Modified Retrieval Rank (NMRR) measure used in the MPEG-7 standardization activity. NMRR not only indicates how much of the correct items are retrieved, but also how highly they are ranked among the retrieved items. NMRR is given by

$$NMRR(n) = \frac{\left(\sum_{k=1}^{NG(n)} \frac{Rank(k)}{NG(n)} \right) - 0.5 - \frac{NG(n)}{2}}{K + 0.5 - 0.5 * NG(n)},$$

where NG is the number of ground truth items marked as similar to the query item, Rank(k) is the ranking of the ground truth items by the retrieval algorithm. K equals to $\min(4*NG(q), 2*GTM)$, where GTM is the maximum of NG(q) for the all queries. The NMRR is in the range of [0 1] and the smaller values represent a better retrieval performance. ANMRR is defined as the average NMRR over a range of queries.

5.2 VO Retrieval Results

Table 3 demonstrates the retrieval results for several video object (VO) queries. The first column shows the results when there are no chroma keying artifacts. The second column gives the retrieval performance when the query VO and several VOs in the database are coded by simulating chroma keying artifacts. Simulation of such artifacts are done by simply imposing a blue background to the objects and encoding the video objects with no LPE padding. As seen from the Table 3, chroma keying artifacts results in a poor retrieval performance. After applying the proposed technique to reduce these effects, the retrieval performance improves significantly.

query video object	Without artifacts	With artifacts	With reduced artifacts
children 1	0. 0000	0. 6396	0. 3333
stefan	0. 0741	0. 0410	0. 0370
hall monitor 2	0. 0250	0. 4250	0. 1500
Average NMRR	0. 0330	0. 3685	0. 1734

Table 3. Video object retrieval results (in NMRR) without any chroma artifacts, with chroma artifacts, and after compensation of the artifacts with the proposed method.

6. Conclusions

In this paper, we discussed the issues arising from employing the DC based color histogram technique in the MPEG-4 compressed domain and proposed a method to compute color histograms for the arbitrarily shaped MPEG-4 video objects. We also proposed a technique for reducing the chroma keying artifacts that may occur at the boundaries of these video objects. Our experimental results show that great retrieval performance improvements are obtained by employing our method in the presence of such artifacts.

References

- [1] R. Chang, W. Kuo, and H. Tsai, "Image Retrieval on Uncompressed and Compressed Domains", IEEE ICIP, 2000.
- [2] J. Lay and L. Guan, "Image Retrieval Based on Energy Histograms of Low Frequency DCT Coefficients", IEEE ICASSP, vol. 6, pp. 3009–3012, 1999.
- [3] M. Shneier and M. Abdel-Mottaleb, "Exploiting the JPEG Compression Scheme for Image Retrieval", IEEE Trans. on Pattern Anal. and Machine Intelligence, pp. 849-853, 1996.
- [4] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Video VM 12.2", doc no. M4576, March 1999.
- [5] Del Bimbo, A., "Visual Information Retrieval", Morgan Kaufmann Publishers, 1999.
- [6] M. Miyahara and Y. Yoshida, "Mathematical transform of (R,G,B) color data to Munsell (H,V,C) color data," in SPIE Visual Com. and Image Proc., vol. 1001, pp. 650-657, 1988.
- [7] M. Ferman, S. Krishnamachari, M. Tekalp, M. Abdel-Mottaleb, and R. Mehrotra, "Group of Frames Color Histogram Descriptors for Multimedia Applications", IEEE ICIP, 2000.
- [8] B. Erol and F. Kossentini, "Automatic Key Video Object Plane Selection Using the Shape Information in the MPEG-4 Compressed Domain", IEEE Trans. on Multimedia, June 2000.
- [9] R. Brunelli and O. Mich, "On the Use of Histograms for Image Retrieval", Proceedings of ICMCS, vol. 2, pp. 143-147, 1999.