Content Based Image Retrieval Utilizing HEVC Intra Coding Features

Kai-Wen Liang, Shin-Ru Hsu, and Pao-Chi Chang
Department of Communication, National Central University, Taoyuan City 32001, Taiwan
{ kwliang, srhsu, and pcchang}@vaplab.ce.ncu.edu.tw

Abstract— *In this paper, a content-based image retrieval technique applied to compression domain for the newest video coding standard, high efficiency video coding (HEVC), is developed. We propose to extract the features from intra frames, which are only partially decoded. The directions of intra mode are taken as a part of the feature when the corresponding sizes of prediction unit are divided into two groups. The second part of the feature is the histogram of the signal power of residuals. The experimental results show that the proposed method can reduce the resource consumption and achieve a good retrieval performance in Oxford 5K dataset.*

Index Terms—Compression domain, Content-based image retrieval, Intra prediction, HEVC

I. INTRODUCTION

Content based image retrieval (CBIR) technology is developed to search similar images in a database by the content of the query image. A typical CBIR system is divided into two parts: feature extraction and feature matching. The features such as color, shape, texture, and so on, are extracted from image contents in the spatial domain and stored in a feature database. Feature matching is a process when a user submits a query image. The CBIR system calculates the features for the query image and computes the similarity between the query images and database images. According to the similarity scores, the CBIR system ranks the retrieval results for a query image.

Traditional CBIR technologies can be categorized into two kinds of approaches: pixel domain approach and compression domain approach. CBIR in pixel domain means extracts the image features directly from the image data such as color, texture, shape, and so on. Since most of the media data is stored in compressed format, images need to be fully decoded to produce the feature before retrieving. It wastes computational and storage capacity in the system. On the other aspect, the CBIR methods in compression domain extracting features directly from the compressed image with partial decoding are more beneficial to process the huge amount of multimedia data.

II. PREVIOUS WORKS AND THE PROPOSED METHOD

A. Previous works

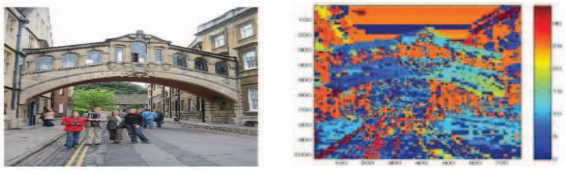
Most CBIR researches used characteristics such as color [1], texture [2] in pixel domain. On the other hand, some CBIR techniques in compression domain used information from compression domain such as statistics values of DC and AC coefficients [3]. Furthermore, [4] and [5] introduced the compressed domain image retrieval methods for I-frame coded images of H.264 standard.

B. Proposed Method

In this work, we propose a content based image retrieval technique in the HEVC compression domain. We adopt two kinds of information in I-frame as features: directions of intra mode and signal power of residual coefficients. Since the coding modes can represent the directions of edges in the images, the histogram of intra modes is built as a feature. On the other hand, the histogram of residual power is also considered as a feature since it provides rich information about the textural characteristics of the contents. The two features are combined to score the similarity in the retrieval process.

C. Direction mode;

As shown in Figure. 1, the map of the intra mode index can efficiently describe the texture characteristics.



(a) (b)

Figure 1. (a) Original image (b) Intra mode map

We divide each image into 16 patches with size

$$\textbf{patch size} = \frac{H}{4} \times \frac{W}{4} \quad (1)$$

where the height and width of image is $\boldsymbol{H}$ and $\mathbf{W}$. An image is divided into 16 patches, and the local mode histogram is calculated for each patch. In HEVC intra prediction, there are 5 PU sizes and 35 prediction modes. Each prediction mode with PU 4x4, 8x8 or 16x16 is regarded as one mode while others with PU 32x32 and 64x64 is considered as another in this work, which results in 0 to 34 bins corresponding to mode 0 to 34 with PU 4x4, 8x8 or 16x16 and 35 to 64 bins for mode 0 to 34 with PU 32x32 and 64x64. To form the histogram, we first map the mode index into prediction angle

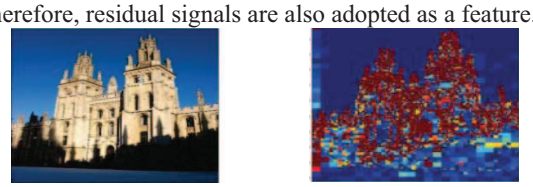
978-1-4799-8751-1/15/$31.00 ©2015 IEEE

nonlinearly and use a moving averaging method, which weighting neighbor modes in Gaussian function (2), to increase the correlation of neighboring mode.

$$GF(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

D. Residual map

The residual signal can efficiently indicate the texture characteristic of original frame as shown in Figure. 2. Therefore, residual signals are also adopted as a feature.



(a) (b)

Figure. 2 (a) original image (b) residual signal map

Mean of residual powers of each PU is defined as follow:

$$\sigma_R^2 = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K} R(i,j)^2 \quad (3)$$

$$P = 10log(\sigma_R^2 + 1)\ (dB) \quad (4)$$

where $R(i,j)$ is the residual coefficients, P is the sample of each PU in the histogram of the residual power.

E. Feature matching

In this work, we intersect the query image feature histogram (H_q) and database image feature histogram (H_d) to score their similarity:

$$S_i = \max_j \{H_q(i) \cap H_d(j)\} \quad (5)$$

Two proposed features, intra mode histogram and residual histogram are separately matched:

$$S_{mode} = H_{mode}(q) \cap H_{mode}(\mathrm{d}) \quad (7)$$

$$S_{resi} = H_{resi}(q) \cap H_{resi}(\mathrm{d}) \quad (8)$$

Then the two similarity are combined to get the final results.

$$\mathrm{S} = \lambda \times \mathrm{S}_{mode} + (1-\lambda) \times S_{resi} \quad (9)$$

$\lambda = 0.85$ in experiments.

Fig. 3 The proposed image retrieval system

III. Experiment Results

Oxford 5k database [6] is used in this experiment. The software is HM12.0 with RDO and QP = 30. As shown in Table 1, the size combination improves the MAP value obviously. That convinced the PU size grouping procedure in the section of direction mode.

As to weighting the neighbor modes, we test different parameters for Gaussian filter, as shown in Table 2.

Finally, we integrate the scores of mode histogram similarity and residual mode for ranking the result. In our work, the proposed system can achieve the performance MAP=0.247 and ANMRR=0.693. Different retrieval methods are compared in Table 3. Our pmethod achieves the best retrieval performance among these cases.

Table 1. Experiment results for different size combination

	H_{all}	H_{mode1} (4↑8), (16↑32↑64)	H_{mode2} (4↑8↑16), (32↑64)	H_{mode3} (4↑8↑ 16↑32↑64)	H_{mode4} 4, (8↑16↑32↑64)
MAP	0.181	0.206	0.199	0.192	0.174
ANMRR	0.768	0.740	0.746	0.748	0.774

Table 2. Experiment results of using different filter orders

	H_{mode1}	H_{mode1}*GF3	H_{mode1}*GF5	H_{mode1}*GF7
MAP	0.206	0.224	0.222	0.213
ANMRR	0.740	0.722	0.725	0.734

Table 3. Comparison of different methods in HEVC

	H_{color} [1]	DCT [3]	[4]HEVC	[5]HEVC	$H_{mode\ 1} * \mathrm{GF3} + H_{resi}$
MAP	0.139	0.076	0.139	0.156	0.247
ANMRR	0.809	0.898	0.809	0.785	0.693

IV. Conclusion

In this work, we analyze the performance of image retrieval by utilizing HEVC coding information in the compression domain. Experimental results show that the propose method that uses the intra mode combination and residual coefficients distribution in HEVC improves the performance of image retrieval. Experimental results show that the MAP can be increased up to 53.4% in average and the full decoding is avoided.

Reference

[1] M. J. Swain, and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, pp. 11-32, 1991.

[2] N. P. Doshi, and G. Schaefer, "A Comprehensive Benchmark of Local Binary Pattern Algorithms for Texture Retrieval," *2012 21st International Conference on Pattern Recognition (ICPR)*, pp.2760-2763, 11-15 Nov. 2012.

[3] (F. E. Malik?) and B. Baharudim, "Effective Content-Based Image Retrieval: Combination of Quantized Histogram Texture Features in the DCT Domain," *2012 International Conference on Computer & Information Science (ICCIS)*, pp.425,430, 12-14 June 2012.

[4] F. Zargari, M. Mehrabi, and M. Ghanbari, "Compressed Domain Texture Based Visual Information Retrieval Method for I-frame Coded Pictures," *IEEE Transactions on Consumer Electronics*, vol.56, no.2, pp.728-736, May 2010.

[5] R. J. Wang, Y. T. Yang, and P. C. Chang, "Content-Based Image Retrieval Using H.264 Intra Coding Features," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 963-969, July 2014.

[6] Oxford 5k ,dataset http://www.robots.ox.ac.uk/~vgg/data/oxbuild