

IMAGE RETRIEVAL BASED ON ENERGY HISTOGRAMS OF THE LOW FREQUENCY DCT COEFFICIENTS

Jose A. Lay and Ling Guan

School of Electrical and Information Engineering

University of Sydney, Australia

e-mail: jlay@ee.usyd.edu.au and ling@ee.usyd.edu.au

Abstract

With the increasing popularity of the use of compressed images, an intuitive approach for lowering computational complexity towards a practically efficient image retrieval system is to propose a scheme that is able to perform retrieval computation directly in the compressed domain. In this paper, we investigate the use of energy histograms of the low frequency DCT coefficients as features for the retrieval of DCT compressed images. We propose a feature set that is able to identify similarities on changes of image-representation due to several lossless DCT transformations. We then use the features to construct an image retrieval system based on the real-time image retrieval model. We observe that the proposed features are sufficient for performing high level retrieval on medium size image databases. And by introducing transpositional symmetry, the features can be brought to accommodate several lossless DCT transformations such as horizontal and vertical mirroring, rotating, transposing, and transversing.

Keywords: Image Retrieval, Energy histograms of low frequency DCT coefficients, Feature Extraction, Compressed domain based Image Retrieval, Real-time Image Retrieval.

1. Introduction

As more and more digital images are acquired into the internetworked multimedia-computing environment, searching and retrieving of image data based-on their information content are essential prior to the development and utilization of an effective image database. This necessity has then inspired many studies in the so-called Content-based Image Retrieval (CBIR) research area since the early 90's, and as a result many tools and techniques have been developed. A much comprehensive survey of the technical achievements is given in [1].

Although much of the images being used are of the compressed format, most of the works so far have been concentrated only on processing of the uncompressed image data. This discrepancy has then contributed to the high computational characteristic, as additional transform steps are required to bring the compressed image data into the uncompressed retrieval domain. The computational complexity issues are often the hindrance in the building of a practical CBIR application, especially in the real-time platform. To overcome this problem, a real-time capable compressed domain based image retrieval approach is being proposed in this work. Processing the image data directly in the compressed domain is favorable not only due to the abolition of the time consuming transform pre-processing but also in consideration of the possibility to work on a more compact data

set. However, as conventional features may not be directly accessible in the compressed domain, exploration of new compressed domain based features may become mandatory. Several recent works involving the use of DCT coefficients are reported in [2][3][4].

Meanwhile, an image database frequently contains many images that differ only in their visual representation. These images are often of transformed variations such as mirroring, transposition, rotation, and transversing. Identification of similarities among these images is significant for a high-level retrieval system.

In this paper we examine the use of energy histograms of the low frequency DCT coefficients for the high level retrieval of DCT compressed images. We design a feature set that is able to accommodate several transformed variations of DCT compressed images known to image databases. We then implement a retrieval system based on the real-time image retrieval approach. We confide that image retrieval systems outfitted with real-time processing capabilities are of greater suitability for the fast growing internetworked computing environment. Finally we test our system on a JPEG photograph database consisting of nearly 4.700 real-life images.

2. Image Retrieval Model

In order to avoid the high computational cost at run time, most of the current image retrieval systems are built based on a model that maintains separate feature and image databases as shown in Figure 1(a) [5][6]. In such a model, a feature database is built in addition to the image database itself during the database creation. The retrieval mechanism then performs the similarity measure by contrasting the features extracted from a query with the feature data stored in the feature database. Consequently, the model requires a feature extraction pre-processing step that in turn inhibits it from being used as a real-time application. Conversely, no feature pre-processing step is required during a database construction based on the real-time image retrieval model as shown in Figure 1(b). Though simple and fast feature extraction scheme will be obligatory as features are extracted on the fly within a retrieval cycle.

As Internet continues its tremendous growth, a large number of small to medium size independently maintained image and or multimedia databases will be remotely accessible but since many diversified indexing schemes are likely to continue their cohabitation, a real-time model that is able to handle the feature extraction locally will be of much preference in most cases. Correspondingly, a compact but reliable real-time capable image retrieval tool will also be much appreciated since it may ensure easier integration into a multimedia based retrieval application.

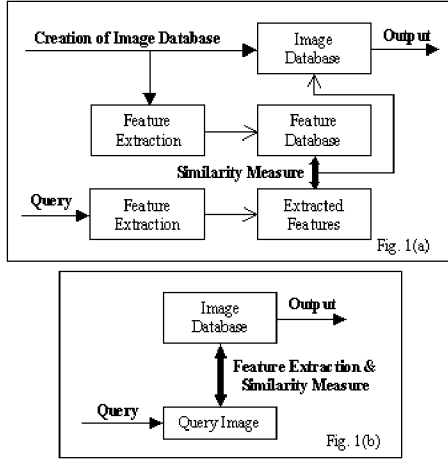


Figure 1: (a) A separate feature and image databases model
(b) A real-time model.

3. The DCT Coefficient Domain

The DCT transform is a unitary integer transform widely adapted in many image and video compression standards. The forward and inverse 2-D DCT transform for an 8x8 block sample used in JPEG and MPEG are given by:

Forward :

$$F(u, v) = \frac{CuCv}{4} \sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}$$

Inverse :

$$f(i, j) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 CuCv F(u, v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16}$$

where $C\tau = 1/\sqrt{2}$ for $\tau=0$ and 1 otherwise. $F(u, v)$ are the DCT coefficients and $f(i, j)$ are the samples of the input pixels. The DCT coefficients retrieved back from a JPEG image are lossy in nature unless a lossless JPEG codec is employed. This is due to the quantization and dequantization steps internal to the JPEG codec structure (Figure 2).

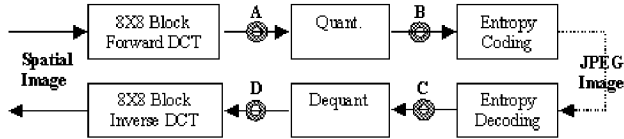


Figure 2: A JPEG codec structure.

Although the JPEG standard does not specify any particular color space for standard usage, the current implementation reinforces the use of YCbCr color space in order to maximize the compression ratio. On most of the implementation, image with different color-space is first converted into the YCbCr domain before it is forwarded to the usual compression steps. A YCbCr-based image can be sub-sampled by 2-1-1 ratio without generating any significant visible disparities. The 2-1-1 ratio is a rather counter-intuitive notation. It stands for 2Y1Cb1Cr in a single coding unit. But in terms of sampling frequency, it shall stand for 1 sample per pixel for the luminance component (Y) and 1 sample for every two pixels for each of the chrominance components (Cb and Cr).

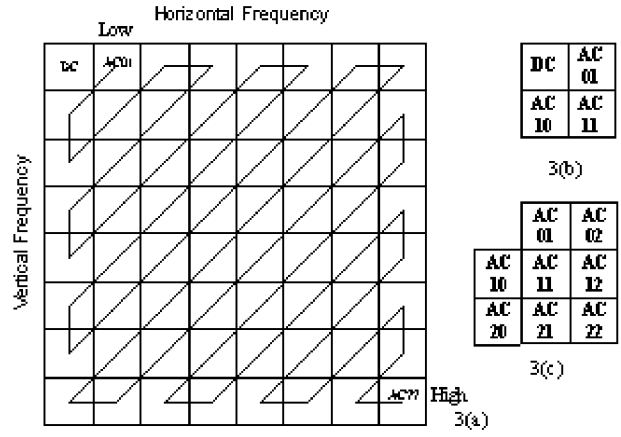


Figure 3: (a) The zigzag pattern of frequency increment in a DCT block,
(b) The $F2B$ feature block, (c) The $F3A$ feature block.

Low frequency DCT coefficients generally bear most of the energy in a typical DCT coefficient block. Even though the total energy of the samples remains unchanged subsequent to a DCT transform, the distribution of the energy will be immensely altered. A typical 8x8 block transform will have most of the energy relocated to its upper-left corner with the so-called DC coefficient (F_{00}) represents the proportional average of the 64 sample values of the block and other AC coefficients denote the intensity changes among the samples. A zigzag pattern depicting the frequency increment of an 8x8 DCT block used in JPEG is shown in Figure 3(a). An empirical study of the energy histograms for several DCT coefficient groups is given in [2].

4. Implementing The System

Histogram is first introduced into the image retrieval field through the use of color histograms [7]. A color histogram is created by counting the number of times a color occurs in an image data set. Similarly an energy histogram of the DCT coefficients is obtained by counting the number of times an energy level appears in a DCT blocks set of a DCT compressed image. An energy histogram (h_c) of an 8x8 DCT block for a particular color component can be written as:

$$h_c[m] = \sum_{u=0}^7 \sum_{v=0}^7 \begin{cases} 1 & \text{if } Q(F[u, v]) = m \\ 0 & \text{otherwise} \end{cases}$$

with $Q(F[u, v])$ denotes the dequantized coefficient's energy level at the u, v location.

A histogram based retrieval scheme is generally tolerant to image rotation and modest object translation. Its versatility may also be extended to include scale invariant through the means of normalization [5]. The use of histogram as retrieval features may also simplify the similarity evaluation processing. A simple distance based similarity measure may be employed. However, a histograms based high-level image retrieval system is relatively computational inefficient as every data point is counted in the building of full image size histogram features. As one of the purposes of this work had been the building of a real-time capable retrieval system, any computational inefficiency should be avoided. Therefore instead of using the full block coefficient

data, we proposed to use only several low frequency DCT coefficients for the construction of our energy histogram features. The reduction is judicious with respect to the quantization properties of the DCT compression technique. However, partial employment of DCT coefficients may threaten the above-mentioned favorable characteristics of the histogram method. As it also had been our aim to have the proposed retrieval system capable of identifying similarities in changes of image-representation due to common image transformations, the invariant property for the features had to be acquired independently. To achieve this aim, we utilized the lossless transformation properties found in the DCT domain.

In the DCT coefficient domain, several transformations such as mirroring/flipping, rotating, transposing, and transverting of a compressed image can be attained without ever fully decoding the image. The transformations are realized by rearranging the DCT coefficients at node C on Figure 2. Hence a DCT domain based transformation is lossless in nature while it significantly reduces the transformation time needed to reshape an image into the intended format. For instances, horizontal and vertical mirroring of a JPEG image can be obtained by swapping the mirroring pairs of the DCT coefficient blocks and accordingly changing the sign of the odd-number columns or rows within each of the DCT block. Likewise, transposition of an image can be accomplished by transposing the DCT blocks following by numerous internal coefficient transpositions. Furthermore transverse and various rotations of a DCT compressed image can be achieved through the combination of appropriate mirroring and transpositional operations. For example, a 90-degree rotation of an image can be performed by transposing and horizontally-mirroring the image. A utility for performing several lossless DCT image transformations is provided in [8].

Bringing together all previous discussions, 6 square-like feature blocks were selected for our experiment:

$$F1 = [DC],$$

$$F2A = [AC_{01}, AC_{10}, AC_{11}],$$

$$F2B = [DC, AC_{01}, AC_{10}, AC_{11}],$$

$$F3A = [AC_{01}, AC_{02}, AC_{10}, AC_{11}, AC_{12}, AC_{20}, AC_{21}, AC_{22}],$$

$$F3B = [DC, AC_{01}, AC_{02}, AC_{10}, AC_{11}, AC_{12}, AC_{20}, AC_{21}, AC_{22}],$$

$$F4 = [DC, AC_{01}, AC_{02}, AC_{03}, AC_{10}, AC_{11}, AC_{12}, AC_{13}, AC_{20}, AC_{21}, AC_{22}, AC_{23}, AC_{30}, AC_{31}, AC_{32}, AC_{33}]$$

The square-like features were deliberately chosen as they are symmetrical to the transpositional operation, which is essential to the lossless DCT transformations. Low frequency coefficients were intended as they convey higher energy level in a typical JPEG image's DCT block. *F1* contains a bare DC component while *F2B*, *F3B*, and *F4* resemble the 2x2, 3x3, and 4x4 upper-left corner of a DCT coefficient block. *F2A* and *F3A* are obtained by removing the DC coefficient from the *F2B* and *F3B* blocks. *F2B* and *F3A* are consecutively illustrated in Figure 3(b) and 3(c). Note that counting the *F1* energy histograms alone is a direct resemblance of the color histograms technique [7] in the DCT coefficient domain. The introduction of *F2A* and *F3A* were meant to explore the contribution made by numerous low frequency AC components. And the use of *F2B*, *F3B*, and *F4*

were intended for evaluation on the impact of block size of the combined DC and AC coefficients.

To approach the real-time platform, we designed the retrieval system based upon the real-time model shown in Figure 1(b). No pre-processing step was required during the database creation in this experiment. Feature extraction and similarity judgment were all performed on the fly when a retrieval cycle was initiated.

For similarity measure, we embraced the normalized Minkowski-form (L_I) distance tailored for easy implementation:

$$d_c(q-m) = \frac{1}{n} \sum_{i=1}^n \frac{\|h_i^q - h_i^m\|^c}{M}$$

with $d_c(q-m)$ represents the normalized L_I distance for a particular component or color channel of the q and m images, h_i represents a particular histogram bin, n denotes the number of histogram bins used for the component, and M stands for the largest possible magnitude of the shifted coefficients. The coefficients were shifted for practical reason to have their possible minimum energy level set to 0. Note that the distance of an image from itself is 0 and the number of histogram bins along with its weight factor may vary across components if preferred.

5. Experimental Results

In the experiment we used a single uncategorized database of nearly 4,700 uniformly sized JPEG photographs consisting of a broad range of real-life images. The collection are produced and maintained by [9]. The DCT coefficients were extracted at node D in Figure 2 utilizing the library provided in [8]. A dequantization step was performed as the quantization table used in the JPEG compression could vary among images.

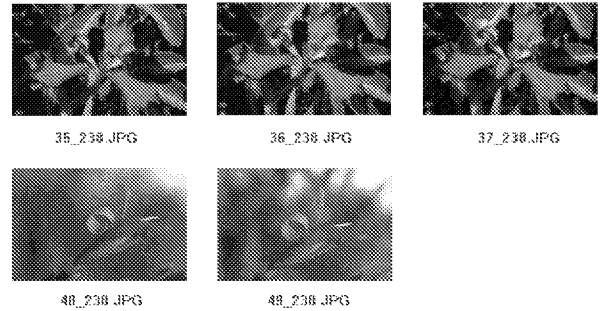


Figure 4: Query and similar images.

Some forty query images were pre-selected by hand to ensure similar images by human vision were properly identified in the database. Thresholding was not considered as the purpose had been to reveal the pre-recognized images' position on the retrieval-hit list. A sample of two query images and their similar associates are shown in Figure 4. Two query images, 36_238.JPG and 49_238.JPG are used here to illustrate the features' contribution on the retrieval performance. The first image group consist of 3 very similar images taken with delicate lens movement while two similar images with slight different

background compose the second image group. The retrieval-hit for the experiment are tabulated in Table 1.

Q: 36_238	F1	F2A	F2B	F3A	F3B	F4
Rank 1 (best)	37_238	37_238	37_238	37_238	37_238	37_238
Rank 2	35_238	X	35_238	35_238	35_238	35_238
Rank 3	X	X	X	X	X	X
Rank 4	X	35_238	X	X	X	X

Table 1(a): Retrieval-hit for Query 35_238.JPG

Q: 49_238	F1	F2A	F2B	F3A	F3B	F4
Rank 1 (best)	X	X	48_238	X	48_238	48_238
Rank 2	X	X	X	48_238	X	X
Rank 3	X	48_238	X	X	X	X
Rank 4	X	X	X	X	X	X
Rank 5	48_238	X	X	X	X	X

Table 1(b): Retrieval-hit for Query 49_238.JPG

We observed that energy histograms based only on the DC coefficient (*F1*) might only perform well on the retrieval of images with very high similarity of colors. Meanwhile the results of *F2A* and *F3A* suggested that histograms of low frequency AC coefficients, which carry the texture and edge information, are contributory to the similarity measure. Thus the combination of DC and numerous low-level AC coefficients (*F2B*, *F3B*, *F4*) yielded better result on both of the lists. On comparison of block-size effect, further examination using other queries in our experiment showed that in general *F2B* and *F3B* are much preferable to *F4*. This is due to the fact that as the feature block grows larger, heavily quantized coefficients are also taken into consideration, hence erroneous result may be generated.

We also noticed that the retrieval performance for feature *F2B* and *F3B* were relatively unaffected by translation of small objects in a globally uniform images. A retrieval sample is shown in Figure 5.

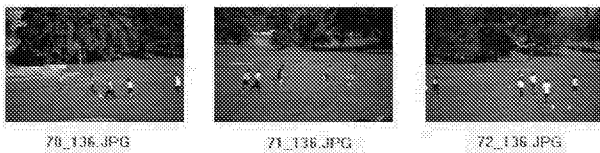


Figure 5: Retrieval of images with translated object contents.

As for retrieval of lossless DCT transformed images, a query was chosen and transformed using the jpegtran utility provided in [8]. The transformed images were then added into the image database prior to the retrieval test. We observed that all features were able to recognize the transformed images. An image group used in our experiment is shown in Figure 6.

Lastly, on the issue of real-time application, we noticed that computational complexity could be significantly reduced through attentive selection of features in the compressed domain. A 2x2- block feature used in the *F2B* may reduce the feature extraction complexity by a factor of $1/16$ with respect to the color histograms method [7]. A typical full cycle retrieval processing time of the *F2B* feature in our experiment was recorded at 250 seconds on a Sun Ultra-10 machine. Note that

there were 4,672 images in the database. This may have opened up an opportunity for a real-time DCT compressed video stream retrieval application.

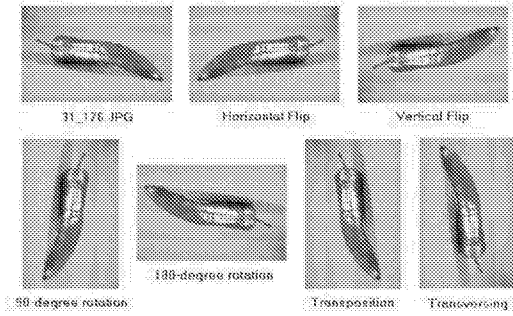


Figure 6: Retrieval of transformed images.

6. Conclusions

We have investigated the use of the energy histograms of the low frequency DCT coefficients as features for the retrieval of DCT compressed images. We conclude that the features are sufficient for performing high level real-time image retrieval on medium size databases. We have also shown that by introducing trans-positional symmetry, a vigorous retrieval system that is able to accommodate variations in lossless DCT transformed images can be constructed. On the real-time processing issue, we confide that an expansion of the system to deal with real-time MPEG video stream retrieval is a justifiable extension.

7. References

- [1] T.S. Huang and Y. Rui, *Image Retrieval: Past, Present, and Future*, Proc. of Int. Symposium on Multimedia Information Processing, Taiwan, Dec. 1997
- [2] A. Vellaikal and C.-C. J. Kuo, *Joint Spatial-Spectral Indexing for Image Retrieval*, Proc. of the 1996 IEEE Int. Conference on Image Processing, Lausanne, Switzerland, part 3, page 867-870.
- [3] H.J. Bae and S.H. Jung, *Image Retrieval Using Texture Based on DCT*, Proc. of Int. Conf. on Information, Comm., and Signal Processing, Singapore, 1997, page 1065-1068.
- [4] S.-S. Yu, J.-R. Liou, and W.-C. Chen, *Computational Similarity Based on Chromatic Barycenter Algorithm*, IEEE Trans. on Consumer Electronics, Vol. 42, No.2, 1995, page 216-220.
- [5] A.K. Jain and A. Vailaya, *Image Retrieval Using Color and Shape*, Pattern Recognition, Vol. 29, No.8, 1996, page 1233-1244.
- [6] V.N. Gudivada, and V.V. Raghavan, *Modeling and Retrieving Images by Content*, Information Processing & Management, Vol.33, No.4, 1997, page 427-452.
- [7] M.J. Swain and D.H. Ballard, *Indexing Via Color Histograms*, Proc. of 3rd Int. Conf. on Computer Vision, Osaka, Japan, 1990, page 390-393.
- [8] T. Lane, *Independent JPEG Group (IJG) JPEG Library*, <http://ftp.uu.net/graphics/jpeg/jpegsrc.v6b.tar.gz>.
- [9] Media Graphics International, *Photo Gallery 5,000 Vol. 1 CD-ROM*, <http://www.media-graphics.net>.