International Journal of Innovative Computing, Information and Control
Volume **2**, Number **4**, August **2006**
 ISSN 1349-4198
pp. 831–839

A CONTENT-BASED IMAGE RETRIEVAL SCHEME IN JPEG COMPRESSED DOMAIN

Zhe-Ming Lu[1,2], Su-Zhi Li[2] and Hans Burkhardt[1]

[1]Institute for Computer Science
University of Freiburg
Georges-Koehler-Allee 052, 79110 Freiburg, Germany
{ zheminglu; Hans.Burkhardt }@informatik.uni-freiburg.de

[2]Visual Information Analysis and Processing Research Center
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen 518055, P. R. China
zhemingl@yahoo.com; suzhili@dsp.hit.edu.cn

Received May 2005; revised October 2005

Abstract. *Nowadays, a large number of images are compressed in JPEG (Joint Photographic Experts Group) format. Therefore, content-based image retrieval (CBIR) for the JPEG images has attracted many people's attention and a series of algorithms directly based on the discrete cosine transform (DCT) domain have been proposed. However, the existing methods are far from the practical application. Thus, in this paper, a new image retrieval scheme for JPEG formatted images is presented. The color, spatial and frequency (texture) features based on the DCT domain are extracted for the later image retrieval. It doesn't require decompressing the images but directly retrieving in the DCT domain. Thus, compared with the spatial domain based retrieval methods for JPEG images, the computation complexity can be greatly reduced. In addition, this retrieval system is suitable for all color images with different sizes. Experimental results demonstrate the advantages of the proposed retrieval scheme.*
Keywords: Content-based image retrieval, JPEG, Discrete cosine transform

1. **Introduction.** Nowadays, CBIR is a hotspot of digital image processing techniques. CBIR research started in the early 1990's and is likely to continue during the first two decades of the 21st century. Many research groups in leading universities and companies are actively working in this area and a fairly large number of prototypes and commercial products are already available. However, the current solutions are still far from reaching the ultimate goal. In this paper, an image retrieval system based on the DCT domain for JPEG images is proposed. Traditionally, image retrieval for JPEG images needs firstly decompressing the images and then performing in the spatial domain. Wherefore, the computation complexity and the processing time are very high since the decompression course is time-consuming, especially for large image databases. With the development of the compression standard, JPEG formatted images alone account for more than 95% of the images over the Internet [1]. Thus, the CBIR performed directly in the compressed domain has become a hotspot. Retrieval in the compressed domain tries to extract the feature vector directly from the compressed data or partial decoded data. It can greatly improve the processing efficiency while reducing the requirements of the computer resources.

Until now, many retrieval methods have been presented in the DCT domain for JPEG formatted images. Most methods make use of the characteristics of DCT coefficients. Literature [2] presented a method based on the histogram of the low frequency coefficients and indicated that the feature vector only based on the DC energy histogram is only suitable for the images highly similar in color. Generally, only partial DCT coefficients are selected to extract the feature vectors. For example, 9 feature vectors can be selected from the 8×8 sub-block in the DCT domain [3]. In [4], the texture feature is extracted through firstly throwing away the high frequency DCT coefficients and then computing the logarithm values for the low and middle frequency DCT coefficients. Some other people use the discrete wavelet transform (DWT) structure to rearrange the DCT coefficients because of the advantage of the DWT [5].

In this paper, we take full advantage of the DCT coefficients and consider the color and texture information for the retrieval of JPEG formatted images. The feature vectors are computed from several DCT coefficients. And this operation is performed in the partial decoded domain. It can greatly decrease the retrieval complexity. What is more, the feature vectors proposed here comprehensively describe the image content. We design a retrieval system based on the presented features. It is a query-example-based image retrieval system. The simulation results illustrated by P-R curves demonstrate the promising advantages of the proposed retrieval scheme.

2. JPEG Compression Standard and CBIR.

In this section, we will give a brief introduction to the JPEG standard and Content-based image retrieval.

2.1. JPEG standard and DCT.

JPEG [6,7] is a joint ISO (International Organization for Standardization) and CCITT (International Telephone and Telegraph Consultative Committee) standard for compressing images developed by the Joint Photographic Experts Group. We should first explain some techniques used in the JPEG scheme for the retrieval method in this paper. JPEG uses a combination of spatial-domain and frequency-domain coding. The image is divided into 8×8 blocks, each being transformed into the frequency domain using the DCT, which is an important step in the whole process. Therefore some redundant data due to the visual masking effect could be discarded to reach the compression aim. Then, through quantization and entropy coding of the DCT coefficients, the compressed image data can be obtained.

As the core transform for JPEG coding, DCT transform formulas are given by:
Forward:

$$F(j,k) = \alpha(j)\alpha(k)\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n)\cos[\frac{(2m+1)j\pi}{2N}]\cos[\frac{(2n+1)k\pi}{2N}] \tag{1}$$

Inverse:

$$f(m,n) = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} \alpha(j)\alpha(k)F(j,k)\cos[\frac{(2m+1)j\pi}{2N}]\cos[\frac{(2n+1)k\pi}{2N}] \tag{2}$$

where $\alpha(k) = \begin{cases} \sqrt{1/N} & \text{if } k = 0 \\ \sqrt{2/N} & \text{if } k = 1, 2, \cdots, N-1 \end{cases}$, $N = 8$. $F(j,k)(j = 1,2,...,8, k = 1,2,...,8)$ are DCT coefficients and $f(m,n)(m = 1,2,...,8, n = 1,2,...,8)$ are the samples of the input image. Some characters of the DCT coefficients can be concluded as follows:

1) Each block of the image is represented by 64 frequency components and the DC coefficient (the first coefficient) represents the average luminance of this block;

2) The low frequency coefficients concentrate on the upper left corner of the DCT block and the high frequency components are in the remaining part;

3) The AC coefficients denote the intensity changes among the samples.

Thus, various features can be extracted from the DCT coefficients.

2.2. **Content-based image retrieval (CBIR).** The general image retrieval system usually consists of three main modules, i.e., input module, query module, and retrieval module [8]. The retrieval process can be shown in Figure 1.

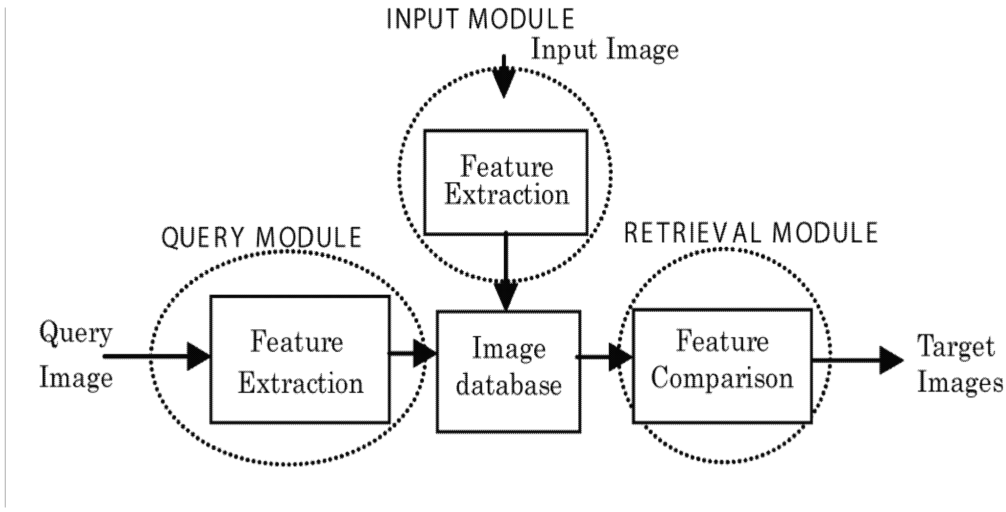

FIGURE 1. Block diagram of the content-based image retrieval system

In the query module, a query image is inputed, and its feature vector is extracted for the latter retrieval. While in the input module, the feature vector is extracted from the input image (the input image is stored in the image database) and the feature vector of each image in the database is also stored in the database. During the retrieval process, the feature vector of the query image is compared with each vector in the database, and the corresponding target (similar) images are outputed. Moreover, many CBIR systems now have the feedback parts and the retrieval performance can be improved accordingly.

3. **Feature Vector Extraction.** The extraction process can be divided into two parts: (1) The color feature vector extraction; (2) The spatial and frequency (texture) feature vector extraction, which can be described as follows.

3.1. **Color feature extraction.** Color information is a kind of important visual information for images and it was pioneered using the content-based image indexing [9]. The color feature definition is so clear that it is widely recognized and studied in the CBIR. Traditionally, we should first decompress the compressed image and then extract the color information. But in this paper, we try to extract the color feature of the original image from few DCT coefficients (the coefficients are obtained through partial decoding of the compressed image). Here, the YCbCr color space is adopted for two reasons: (1) The YCbCr color space, where Y component represents the luminance information and CbCr represent the chrominance information, is used in the JPEG standard and so the complexity will be reduced; (2) Compared with the RGB color space, YCbCr is much more suitable for the human visual system and it is also used in the proposed color layout descriptor in the international standard MPEG-7 [10].

In the spatial domain, the color information of the image in a small block does not vary a lot. So for each 8×8 DCT block, we further divide it into four 4×4 sub-blocks. The average values of all 4×4 blocks in one 8×8 block are defined as M_{11}, M_{12}, M_{21} and M_{22}, as shown in Figure 2.

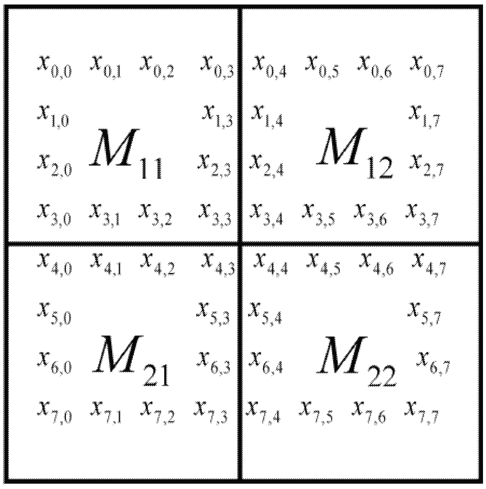

FIGURE 2. Four sub-blocks of the 8×8 DCT block

We assume that the color information is homogenous in each 4×4 sub-block since it is very small. Thus based on the literature [11], we know that the approximate value of M_{11}, M_{12}, M_{21} and M_{22} can be calculated through the four upper left coefficients in the 8×8 DCT block, and the corresponding equations can be expressed as follows:

$$\begin{aligned} M_{11} &= \frac{2C(0,0) + 2C(1,0) + 2C(0,1) + 2C(1,1)}{16} \\ M_{12} &= \frac{2C(0,0) + 2C(1,0) - 2C(0,1) - 2C(1,1)}{16} \\ M_{21} &= \frac{2C(0,0) - 2C(1,0) + 2C(0,1) - 2C(1,1)}{16} \\ M_{22} &= \frac{2C(0,0) - 2C(1,0) - 2C(0,1) + 2C(1,1)}{16} \end{aligned} \tag{3}$$

where $C(0,0)$, $C(1,0)$, $C(0,1)$ and $C(1,1)$ are the four upper left coefficients of the 8×8 DCT block. Because M_{11}, M_{12}, M_{21} and M_{22} are the mean values of the four sub-blocks, all of them can well represent the color information of the image. For each color component, we can calculate the four average values of sub-blocks for each 8×8 DCT block. We then respectively calculate the four normalized accumulative histograms H_{M11}, H_{M12}, H_{M21} and H_{M22} for each color component in the YCbCr color space. At last, for each color component, a color feature vector based on 4 histograms is constructed.

3.2. **Texture feature vector extraction.** In this paper, a 12-dimensional feature vector is extracted from each image inspired by literature [3]. The DCT coefficient selection is illustrated in Figure 3. For the coefficient selection, we consider the following three facts: (1) the first coefficient (DC component) represents the energy information; (2) The AC coefficients of the DCT block reflect the frequency information; (3) The DCT coefficients in some regions can reflect the direction information of the image. In Figure 3, ① is the DC component, representing the energy information; Groups ② and ③ represent the frequency information; Groups ④,⑤ and ⑥ respectively represent the vertical, horizontal and diagonal direction information. For each 8×8 DCT block, we calculate the summation value of all coefficients in each group, totally 6 values for each block. We obtain all values from the same group in all blocks, and then calculate the mean and standard deviation over these values. Thus for each color component, we can obtain 12 feature values that can be denoted by the vector $F = [f_1, f_2, ..., f_{12}]^T$.

The block diagram of the feature vector extraction in this paper is shown in Figure 4.

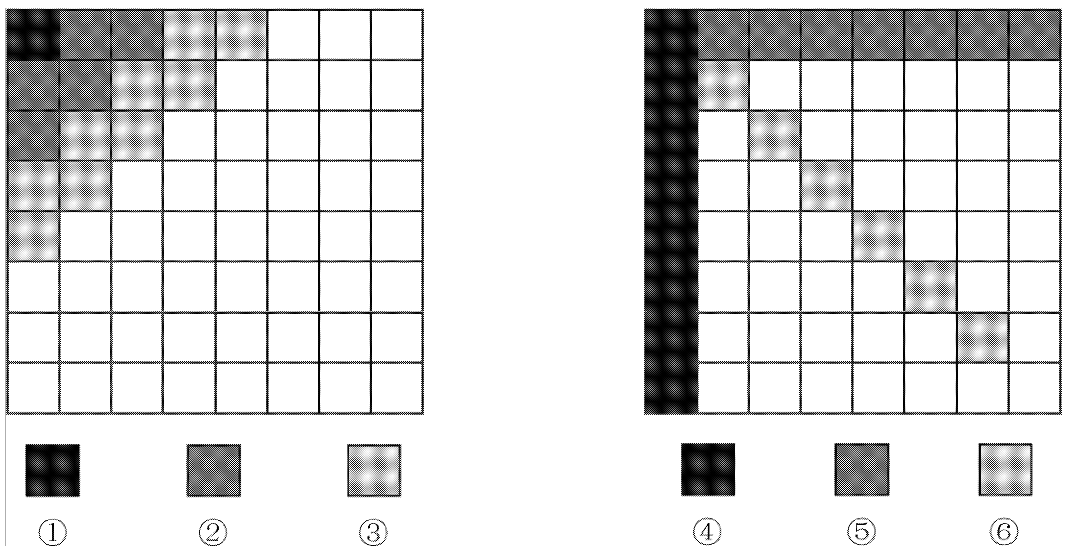

FIGURE 3. Illustration of DCT coefficient selection

3.3. **Similarity measure.** As described above, for each color component, we can get a color feature vector and a 12-dimensional spatial and frequency feature vector. Here we use the Euclidean distance to evaluate the distance between the query image and the images in the database.

Firstly, for all color histograms obtained from Section 3.1, the histogram matching method is adopted to compute the distance as follows:

$$M_E(Q,R) = \sqrt{\sum_{i=0}^{L-1} [Y_Q(i) - Y_R(i)]^2} + \sqrt{\sum_{i=0}^{L-1} [Cb_Q(i) - Cb_R(i)]^2} + \sqrt{\sum_{i=0}^{L-1} [Cr_Q(i) - Cr_R(i)]^2} \quad (4)$$

where Q, R respectively represent the query image and the image in the database, and $M_E(Q,R)$ is the distance function. $Y_Q(i)$, $Cb_Q(i)$ and $Cr_Q(i)$ are the normalized accumulated histograms of Y, Cb and Cr color components of the query image respectively, while $Y_R(i)$, $Cb_R(i)$ and $Cr_R(i)$ are for the image in the database. L is the number of bins for describing the histogram and here we set L=256.

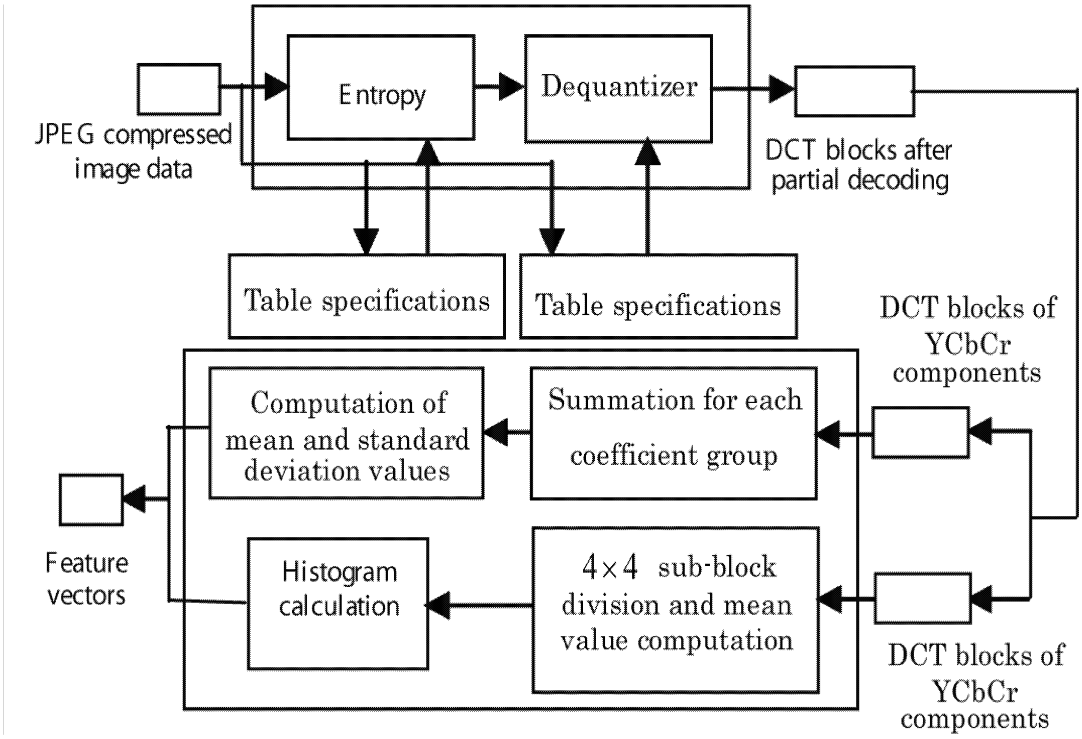

FIGURE 4. Feature vector extraction process

Secondly, for the 12-dimensional spatial and frequency vector, we calculate the Euclidean distance between the normalized feature vectors as follows:

$$D(Q,R) = \sqrt{\sum_{i=1}^{12} [\bar{f}_{Yi}^Q - \bar{f}_{Yi}^R]^2} + \sqrt{\sum_{i=1}^{12} [\bar{f}_{Cbi}^Q - \bar{f}_{Cbi}^R]^2} + \sqrt{\sum_{i=1}^{12} [\bar{f}_{Cri}^Q - \bar{f}_{Cri}^R]^2} \quad (5)$$

where Q, R respectively denote the query image and the image in the database, $D(Q,R)$ is the distance function, and i is the index of the vector. $\bar{f}_i = \frac{f_i - m_i}{\sigma_i}$, where m_i and σ_i are the mean and standard deviation values of f_i calculated over all the images in the database for the same feature.

For the combination of the above two kinds of feature vectors, we use the weighted summation of the two obtained distances, i.e., the final distance is calculated by:

$$S(Q,R) = \alpha M_E(Q,R) + \beta D(Q,R) \tag{6}$$

where (α, β) are the weighting factors and in this paper they are selected to be (0.99999995, 0.00000005) after plenty of experiments.

4. Experimental Results.

4. Experimental Results. The proposed method has been implemented using Visual C++ 6.0 platform. Most common evaluation measures used in image retrieval (IR) are precision and recall [12] (see Eq.7), usually presented as a precision *vs.* recall graph (PR graph). Thus, to give a just and impersonal evaluation of the retrieval efficiency of our method, the PR graph is used here. The method is tested on a certain database with 1000 images. It consists of ten categories including people, beach, building, bus, dinosaur, elephant, flower, horse, mountain and food, each containing 100 images. These images are JPEG formatted color images of different sizes. From each category, five images are randomly selected to be the query images, and for each query image we take down the precision and recall values when the user retrieves 16, 32, . . . , 160 images (in our retrieval system, one page displays 16 images). For each number of retrieved images, we compute the average precision and the average recall value over all test query images, thus we can get a PR graph with ten points.

In the simulation, we respectively test the system for three cases: (1) solely using the color feature, (2) solely using texture feature and (3) synthetically using the two kinds of features. The results are shown in Figure 5. To give a performance comparison with other DCT-domain retrieval algorithms, we also perform the experiment on the same database with the method in the literature [2], which uses 9 coefficients to calculate the histogram. The number of bins is also selected to be 256. The comparison results are shown in Figure 6.

$$precision = \frac{No.\ relevant\ documents\ retrieved}{Total\ No.\ documents\ retrieved},$$

$$recall = \frac{No.\ relevant\ documents\ retrieved}{Total\ No.\ relevant\ documents\ in\ the\ collection} \tag{7}$$

From Figure 5, we can see that the retrieval performances of the methods using two kinds of features together (labeled as the two features) and the texture features (labeled as texture feature) singly outperform the method using the color features (labeled as color feature) singly because only using color information can not adequately describe the image. In addition, the first 7 points of the method using the texture features are better than the method using the two kinds of features together. Meanwhile, these two better cases by and large perform equally. Compared with the other algorithm in the DCT domain (see Figure 6, labeled as other people's algorithm), we can see that the retrieval using the texture features and the two kinds of features in our paper both outperform the method proposed in [2].

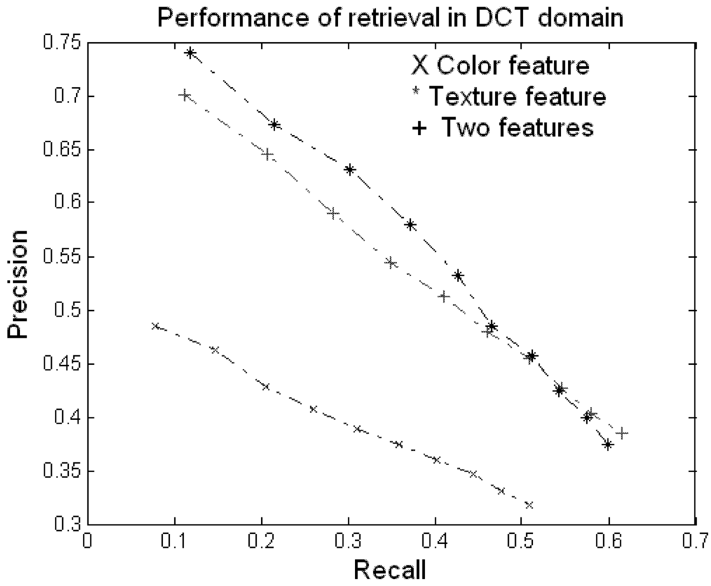

FIGURE 5. Retrieval performance comparisons for three cases

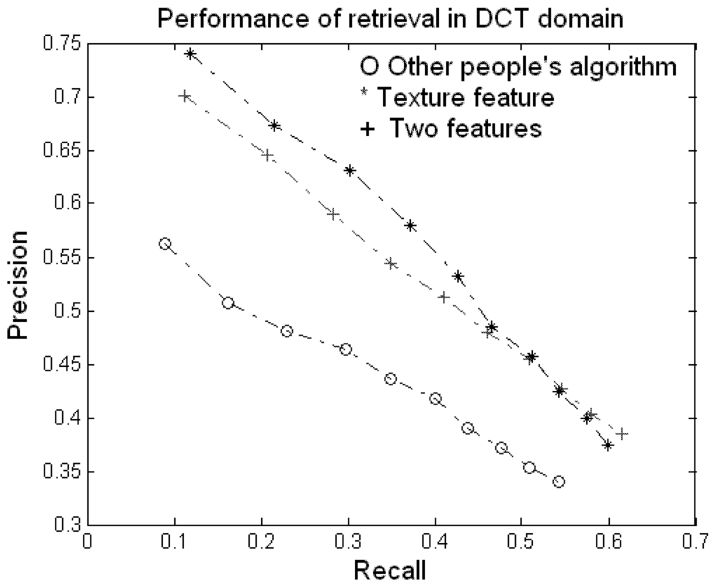

FIGURE 6. Comparisons of the retrieval performance

5. **Conclusions.** This paper proposed an image retrieval scheme in the DCT domain. It is suitable for retrieval of color JPEG images of different sizes. Two kinds of features, color and texture, based on DCT coefficients, are proposed. Experimental results indicate that the retrieval based on the texture features and the two kinds of features together perform well on a 1000-image database. Compared with the other algorithm, they both show much better performance. To improve the retrieval results, the subsequent work will focus on the feedback method for our retrieval system.

Acknowledgment. This work was supported by the Alexander von Humboldt Foundation Fellowship (Germany) grant, ID: CHN 1115969 STP.

REFERENCES

[1] Jiang, J., A. Armstrong and G. C. Feng, Direct content access and extraction from JPEG compressed images, *Patten Recognition*, pp.1-9, 2001.

[2] Lay, J. A. and L. Guang, Image retrieval based on energy histograms of the low frequency DCT coefficients, *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.6, 3009-3012, 1999.

[3] Bae, H. J. and S. H. Jung, Image retrieval using texture based on DCT, *Proc. of the International Conference on Information and Communications Security*, vol.2, pp.1065-1068, 1997.

[4] Sim, D. G., H. K. Kim and R. H. Park, Fast texture description and retrieval of DCT-based compressed images, *Electronic Letters*, vol.37, no.1, pp.18-19, 2001.

[5] Huang, Y. L. and R. F. Chang, Texture features for DCT-coded image retrieval and classification, *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.6, pp.3013-3016, 1999.

[6] Wallace, G. K., The JPEG still picture compression standard, *Comm. ACM*, vol.34, pp.30-44, 1991.

[7] Pennebaker, W. B. and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.

[8] Niblack, W., R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos and G. Taubin, The QBIC project: Querying images by content using color, texture, and shape, in *Storage and Retrieval for Image and Video Databases*, Proc. SPIE 1908, pp.173-187, 1993.

[9] Swain, M. J. and D. H. Ballard, Color indexing, *International Journal of Computer Vision*, vol.7, no.1, pp.11-32, 1991.

[10] Kasutani, E. and A. Amada, The MPEG-7 color layout descriptor: A compact image feature decryption for high-speed image/video segment retrieval, *Proc. of the IEEE International Conference on Image Processing*, vol.1, pp.674-677, 2001.

[11] Feng, G. C. and J. Jiang, Image extraction in DCT domain, *IEE Proceedings: Vision, Image and Signal Processing*, vol.150, no.1, pp.20-27, 2003.

[12] Mueller, H., W. Mueller, D. Squire, S. Marchand-Maillet and T. Pun, Performance evaluation in content-based image retrieval: overview and proposals, *Pattern Recognition Letters*, vol.22, no.5, pp.593-601, 2001.