COMPRESSED DOMAIN IMAGE INDEXING AND RETRIEVAL BASED ON THE MINIMAL SPANNING TREE

Ch. Theoharatos, V.K. Pothos, G. Economou and S. Fotopoulos

Electronics Laboratory, Dept. of Physics, University of Patras, Rio 26500, Greece

ABSTRACT

In this paper, a method for content-based retrieval of JPEG images is presented, utilizing features directly from the discrete cosine transform (DCT) domain. Image indexing is achieved by extracting color and texture feature vectors, using an efficient technique applied on the DCT coefficients. Similarity between the query- and database- images is provided based on a statistical graph matching approach. The proposed measure makes use of the Wald-Wolfowitz test, a nonparametric test that assesses the commonality between two different sets of multivariate observations. Experimental results demonstrate the enhanced performance of our approach, compared to previously reported methods.

1. INTRODUCTION

Nowadays, due to space limitations and reduction in download speed, the large amount of digital images on the World Wide Web is stored in JPEG format [1]. While image compression successfully solves the problem of data size, the content access of those compressed files becomes a major issue for further research. To bridge the gap between the compressed and the pixel domain, where the majority of image processing and analysis algorithms are developed, research is directed to content indexing techniques in the compressed domain [2-5]. Here, we investigate the use of DCT coefficients, which are the representative components of JPEG standards, in content-based image retrieval (CBIR) [3-5].

In CBIR, the goal is to choose appropriate features from the database images, construct the representative distributions and estimate the similarity between them using a properly defined measure. In relation to JPEG images, color and texture features are the only attributes that can be directly extracted from the DCT coefficients [3], without having first to decompress them. Regarding distributions, the most suitable representation for color information is color histogram [6], which statistically denotes the joint probability of intensities of the different color channels. Texture, which is characterized by responses to spatial and orientation filters in a neighborhood of a pixel position, can also be represented using histogram methodologies [7]. In general, histogram-techniques provide useful clues for the subsequent expression of similarity between images, due to their robustness to background complications and object distortion. However, the necessary trade-off during the binning procedure has been characterized as the most significant drawback of these methods. Finally, concerning similarity, a profound number of measures have been proposed for estimating the distance between histograms [7].

In this work, a dual algorithmic procedure for image retrieval of compressed JPEG images is introduced, using multivariate graph matching. First, color and texture features are directly extracted from the DCT domain. Indexing is obtained by selecting the DC component from each *NxN* macroblock as the color attribute. In addition, a number of *k*-vectors are extracted from the diagonal zig-zag lines. The k-magnitudes of these vectors are afterwards estimated and used as a means of mining texture or color-texture attributes. Next, the comparison between database images is accomplished via a nonparametric test dealing with the "*Multivariate Two-Sample Problem*" [8]. The specific test is a multivariate extension of the classical *Wald-Wolfowitz test* (WW-test) and compares two different samples of vectorial observations (i.e. two sets of points in $\mathbf{R}^{\mathbf{P}}$) by checking whether they form different branches in the overall minimal spanning tree (MST) [9]. The output of this test can be expressed as the probability that the two point-samples are coming from the same distribution. Its immense advantage is that no a-priori knowledge of the distribution of points is a prerequisite [10].

2. FEATURE EXTRACTION OF JPEG IMAGES

In line with JPEG images, we limit our discussion and feature extraction scheme to DCT compression scheme, extracting appropriate features directly from the block DCT coefficients of the .jpg images, without having first to decompress them.

2.1. DCT compression scheme

The JPEG compression standard is based on the discrete cosine transform (DCT). First, an image is divided into fixed size *NxN* blocks (macroblocks) and the DCT is applied afterwards to each block to separate the high- and low-frequency information (*N=8* for JPEG images). Given a macroblock $f_{i,j}$ of *NxN* pixels, the 2-D DCT transform applied over the *i,j* pixels is given by:

$$F_{u,v} = \frac{1}{N} C(u) C(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_{i,j} \cos\frac{\pi u(2i+1)}{2N} \cos\frac{\pi v(2j+1)}{2N}$$

and $C(u), C(v) = 1/\sqrt{2}$ if $u, v = 0$, otherwise $C(u), C(v) = 1$. $F_{u,v}$ is the 2-D DCT coefficient computed from the $f_{i,j}$ image spatial value. As *u* and *v* varies along the horizontal and vertical direction inside [0, N-1], the DCT coefficients are obtained, covering different spectral bands. $F_{0,0}$ is the DC component which essentially represents the average brightness over all the *8x8* pixels (in the case of JPEG) of each corresponding

0-7803-9332-5/05/$20.00 ©2005 IEEE

macroblock in the spatial domain, as depicted in Fig. 1. The remaining *63* coefficients, referred to as the AC coefficients, capture the frequency and directionality properties within the pixel-block. For most images, much of the signal energy lies at low-frequency components, which appear in the upper left corner of the DCT. This means that the coefficients of the high-frequency components are close to zero, having little visible impact and can therefore be neglected in most cases. In this way, compression is achieved by quantizing the AC coefficients within each macroblock to remove high-frequency components, using the zig-zag scanning (Fig. 1) to obtain an appropriate ordering from the lowest to highest frequency. The vectors resulting from the zig-zag ordering Z_k, $k=1:2N-2$ contain all the AC coefficients starting from the upper left location (0,1) to the bottom right (N-1,N-1). Specifically, in the case of JPEG images, there exist $k=14$ groups of zig-zag vectors inside the *8x8* macroblocks, as presented in Fig. 1.

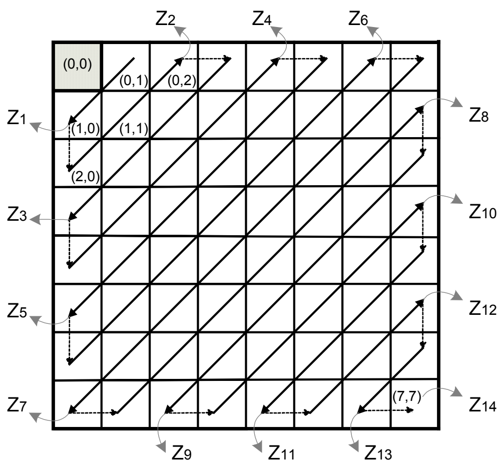

Fig. 1. Zig-zag ordering of the AC coefficients. Each straight diagonal line Z_k is considered as a vector.

2.2. Feature extraction

In JPEG standard, the YCrCb color space is used to encode color images. This model results from the RGB color space by applying a linear transformation [1], where the Y-channel represents the luminance information and the Cr- and Cb-frames represent the two chrominance differences. Working separately in each panel, color and texture features can be extracted directly in the compressed domain from the DCT coefficients.

In order to represent color information from each *NxN* macroblock of a given image, all DC components are separately extracted and used as input vectors in the WW-engine to form a 3-D vector space. Texture features, on the other hand, can be defined as the spectrum energies in different localizations of a local macroblock [3]. Since the DC coefficient $F_{0,0}$ represents the average grayscale value of each *N*x*N* macroblock, it is not considered to carry any texture information. The remaining AC coefficients can be considered to characterize image texture and be used as texture features. In our approach, a novel indexing method is used. It is based first on the extraction of *k*-vectors, obtained from the diagonal zig-zag coefficients of each macroblock, where a vector is defined by the AC components contained inside each diagonal line of the zig-zag scheme (see also Fig. 1). The *k*-magnitudes V_k, $k=1:2N-2$ of the corresponding zig-zag vectors are afterwards computed, from Z_1 to Z_k (in the case of *8x8*, $k=14$ as shown in Fig. 1), using the following formula:

$$\{V_k\} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{pmatrix} = \begin{pmatrix} \sqrt{F_{0,1}{}^2 + F_{1,0}{}^2} \\ \sqrt{F_{0,2}{}^2 + F_{1,1}{}^2 + F_{2,0}{}^2} \\ \vdots \\ \sqrt{F_{7,7}{}^2} \text{ or } |F_{7,7}| \end{pmatrix}, \text{ for } k = 1:2N-2 \quad (1)$$

The introduced feature extraction methodology was proven to be robust when similarity is examined in terms of image rotation. That is, by applying the DCT transform to an image macroblock and its rotated version, the set of the absolute values of the DCT coefficients is identical, whereas their positions in the zig-zag ordering scheme are different. Even though the DC component is used for color feature characterization and the remaining AC components for texture features, it is a common belief that color and texture attributes are mixed together in the (*N*x*N*)-1 coefficients contained inside a pixel-block [3]. In this way, we expect color to be present at several AC coefficients, packing most of its spectral energy in the fewest number of low-frequency coefficients at the upper left corner of the macroblock.

According to the number of zig-zag vectors that need to be drawn from each macroblock so as to represent effectively and efficiently the color and texture attributes, much work has been done recently [3-5]. By testing with several JPEG images and using standard statistical methods [11] such as entropy estimation, the number of extracted zig-zag vectors from each chromatic channel was approximately found to be $k=3$, therefore using Z_1, Z_2 and Z_3 as provided by (1). In addition, $k=8$ zig-zag vectors need to be drawn from the luminance Y-frame in order to efficiently represent the color and texture attributes. It should be noticed here that the extraction of features (i.e. the DC component and the *k* zig-zag vectors $\{V_k\}$), increases the dimensionality of the derived feature space. However the computational complexity is not increased due to the fact that the WW-test, which will be presented next, is a function of the number of input vectors and not of their dimensions. On the other hand, the similarity measure is normally optimized by the higher number of extracted image features.

3. MULTIVARIATE GRAPH MATCHING

A nonparametric test dealing with the *"Multivariate Two-Sample Problem"* [8], has been adopted here for estimating image similarity. The specific measure is a multivariate extension of the Wald-Wolfowitz test and compares two different samples of vectorial observations (i.e. two point-sets in $\mathbf{R^d}$). The output of the test can be expressed as the probability that two point-samples are coming from the same distribution. Its great advantage is that it is model-free and this stems from the graph-theoretic origin of the test, which is actually based on the concept of minimal spanning tree (MST) graph [9].

WW-test can be used to test the hypothesis H_o, whether any two given multidimensional point samples $\{X_i\}_{i=1:m}$ and $\{Y_i\}_{i=1:n}$ are coming from the same multivariate distribution. At first, the sample identity of each point is not encountered and the MST of the overall sample is constructed. Then, based on the sample identities of the points, a test statistic *R* is computed.

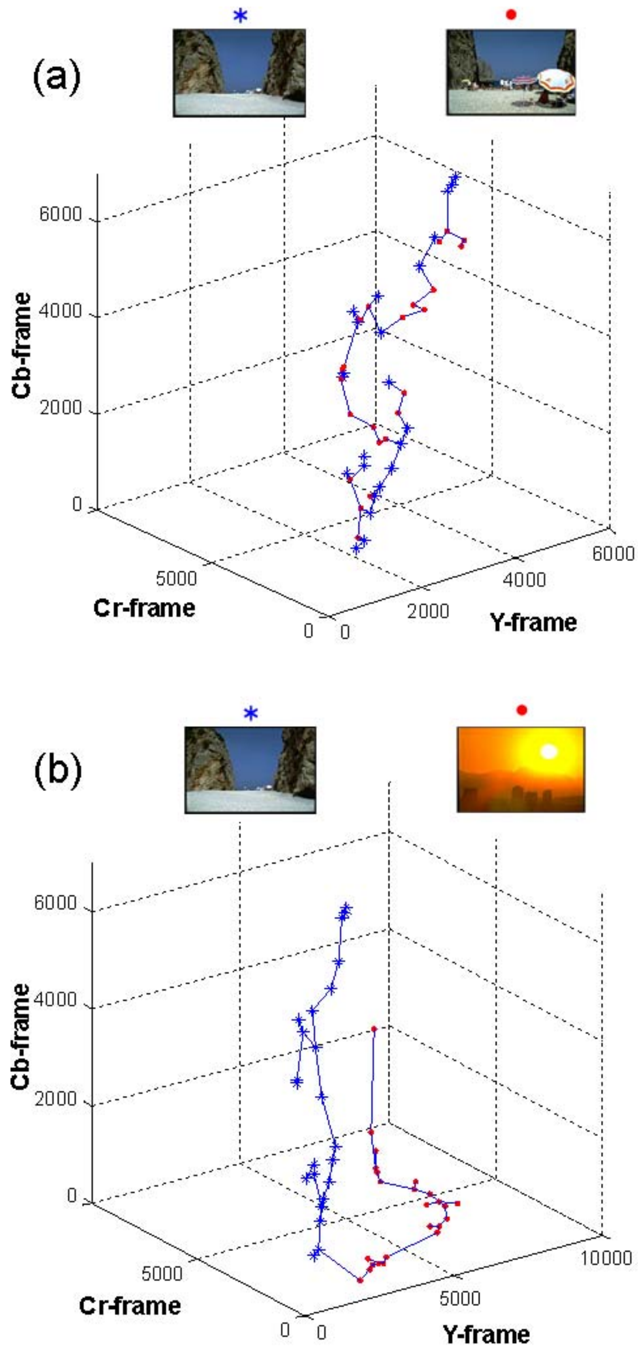

Fig. 2. WW-test for a pair of (a) similar and (b) dissimilar images, based on 24 extracted DC-coefficients from each image (labeled by '*' and 'o' for each different image).

R is the total number of *runs*, while a *run* is defined as a consecutive sequence of identical sample identities. Rejection of H_o is for small values of R. The null distribution of this statistic has been derived, based on combinatorial analysis [8]. It has been shown that the quantity:

$$W = \frac{R - E[R]}{\sqrt{Var[R \mid C]}} \qquad (2)$$

approaches (asymptotically) the standard normal distribution, while the mean E[R] and variance Var[R|C] of R are given in closed form [8]:

$$E[R] = \frac{2mn}{N} + 1\,, \quad Var[R \mid C] = \frac{2mn}{N(N-1)} \times$$
$$\times \left\{ \frac{2mn - N}{N} + \frac{C - N + 2}{(N-2)(N-3)} [N(N-1) - 4mn + 2] \right\}$$

where N=m+n and C is the number of edge pairs sharing a common node. The simple formula in (2) enables the computation of the *significance level* (and *p-value)* for the acceptance of the hypothesis H_o. The great advantage of WW-approach is that since it involves a "distributional distance" acting on samples of image constituents, the emerging similarity measure possesses desirable invariant characteristics, such as rotation and translation invariance.

In the present work, the above test is utilized as follows. With the feature extraction step, a representative point-sample is selected for the two compressed images to be compared. W is then computed and used as a similarity measure in a way that the more positive its value is, the more similar the two images are [10]. Fig. 2 presents the performance of WW-test for a pair of similar and a pair of dissimilar color images, when only color information is considered, by extracting the 3-D DC-component vectors for $N=32$ size macroblocks. By contrasting the overall MSTs, it becomes evident that in the case of similar images shown in Fig. 2(a) the selected vectors tend to mix together. In this case, there are 19 edges having differently labeled nodes as endpoints, resulting in R=20 (and W=-1.1531). On the contrary, when dissimilar images are considered as depicted in Fig. 2(b), the extracted vectors are clearly separated from each other, forming clusters at different locations in the corresponding feature space. Here, there is only 1 edge having different endnodes, resulting in R=2 (and W=-7.5235).

4. EXPERIMENTAL STUDY

Images from the Corel gallery [12] were used for demonstrating and validating our approach. The utilized dataset was formed by pre-assigning the images into 20 distinct classes of $S=50$ similar images, including in this way an overall of $D=1000$ JPEG images. A subset of $Q=50$ query images was randomly selected from this heterogeneous set. In order to justify the effectiveness of our approach, *precision* (*Pr*) and *recall* (*Re*) [1] were adopted for our experiments. In what follows, the above indices were used for comparing our method, which is based on the use of W-index as similarity measure, with the histogram intersection (HI) methodology [6] and the earth's mover distance (EMD) [7], both applied on an 64-bin histogram representation of the selected attributes. Experiments have been conducted using the same number of extracted color and texture features from each query and database image, authorizing in this way the direct comparison between the different approaches.

The Pr and Re indices were first evaluated for different sizes T of the selected list ($T=5:5:50$) and the computed values were used in the curves of Fig. 3. As we can perceive, the WW-test outperforms the other similarity methods, having in all cases of the selected list T of retrieved images significantly higher precision rate. In particular, considering the T=10 retrieved images (depicted by the second label-point in each curve), the proposed measure exhibits ~9% higher retrieval performance compared to the EMD and ~11% in respect to the HI-approach.

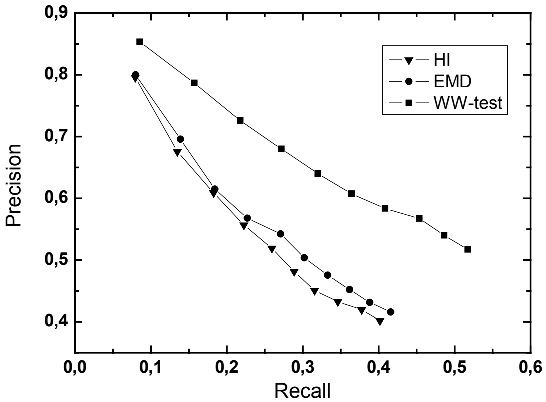

Fig. 3. Precision vs. Recall diagrams for the proposed similarity measure, along with the HI- and EMD- methodologies, for the same set of selected attribures.

5. CONCLUSIONS

In this study, a novel methodology for content-based indexing and retrieval of JPEG compressed images is proposed, using a dual segregation algorithmic step. First, color and texture attributes are directly considered in the DCT-domain, following an efficient indexing scheme for extracting representative feature vectors. Estimating the magnitudes of the diagonal zig-zag vectors from the reordered DCT coefficients, provides enhanced performance prior to image's geometric transformations. The similarity of images is assessed using the statistical multivariate WW-test. Part of the flexibility is due to the multivariate orientation of the core procedure, since different image characteristics can be combined in one type of query (i.e. color plus texture features). The main advantage comes from the statistical nature of WW-test, considering that different types of queries can evolve independently and their results can be compared across types, as in the case of an image retrieval system. The latter is a direct consequence of the fact that the computed W-index relates directly to significance level and therefore can be used as an absolute measure to rank among the results of different types of queries. The only seemingly weak point of the proposed scheme is that it relies on the formation of MST, which is known to be a computationally demanding procedure. To provide some insight about the computational complexity of the methodology, the MST- construction requires computational time $O(N^2)$ using Prim's algorithm [9], while the test statistic can be evaluated in time $O(N)$, where N is the number of data points in both cases.

A major issue that needs to be noticed here, is that the WW-test can be applied to operate on variable-length representations of the two distributions to be compared, avoiding in this way the well-known quantization problems related with the binning procedure of the histogram-matching techniques. However, this was not engaged in our work due to the nature of the compressed application domain, extracting an equal number of feature vectors simultaneously from the DCT-coefficients. By estimating the performance of the proposed similarity measure in an image retrieval task, we show that it outperforms histogram-based techniques, which are considered as classical approaches for image similarity.

Among our future objectives is to investigate whether features characterizing shape attributes can be directly extracted from the compressed domain, so as to be included in the WW-engine. Also, the extraction of more reliable texture features that capture structural fields beyond the DCT macroblock, need to be examined. Finally, we are inspecting to adapt our approach to JPEG2000 images as well, by utilizing wavelet coefficients in the selection procedure.

6. ACKNOWLEDGEMENTS

The work of the first author was supported by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), under the Program HERAKLEITOS.

7. REFERENCES

[1] V. Castelli and L. D. Bergman, *Image Databases: Search and Retrieval of Digital Imagery*, John Wiley & Sons, New York, 2002.

[2] M.K. Mandal, F. Idris and S. Panchanathan, "A critical evaluation of image and video indexing techniques in the compressed domain", *Image Vision Comput.*, vol. 17, no. 7, pp, 513-529 , 1999.

[3] Y. Zhong and A.K. Jain, "Object localization using color, texture and shape", *Pattern Recognition*, vol. 33, pp. 671-684, 2000.

[4] C.-W. Ngo, T.-C. Pong and R.T. Chin, "Exploiting image indexing techniques in DCT domain", *Pattern Recognition*, vol. 34, pp. 1841-1851, 2001.

[5] J. Jiang, A. Armstrong and G.C. Feng, "Web-based image indexing and retrieval in JPEG compressed domain", *Multimedia Systems*, vol. 9, pp, 424-432, 2004.

[6] M.J. Swain, and D.H. Ballard, "Color indexing", *Int. J. Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[7] Y. Rubner, J. Puzicha, C. Tomasi and J.M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture", *Comput. Vision Image Understand.*, vol. 84, pp. 25-43, 2001.

[8] J.H. Friedman and L.C. Rafsky, "Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests", *Annals of Statistics*, vol. 7, no. 4, pp. 697-717, 1979.

[9] C.T. Zahn, "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters", *IEEE Trans. on Computers*, vol. C-20, no. 1, 1971.

[10] Ch. Theoharatos, N. Laskaris, G. Economou and S. Fotopoulos, "A Similarity Measure for Color Image Retrieval and Indexing based on the Multivariate Two Sample Problem", *in Proc. of EUSIPCO 2004*, Vienna, Austria, Sept. 2004.

[11] R.O. Duda, P.E. Hart and D.G.Stork, *Pattern Classification*, Second Edition, John Wiley & Sons, New York, 2001.

[12] *Corel stock photo library, Corel Corp.*, Ontario, Canada.