

1-D DCT Shape Feature for Image Retrieval

Seyed Alireza Seyedin^{#1}, Mohammad Faizal Ahmad Fauzi^{#2}, Fatahiyah Mohd Anuar^{#3}

^{#1,2,3} Faculty of Engineering, Multimedia University,
Cyberjaya, 63100, Malaysia

^{#1} Electrical Engineering Dept., Ferdowsi University of Mashhad,
Mashhad, 91779 48974, Iran

¹ seyedin@um.ac.ir

² faizal1@mmu.edu.my

³ fatahiyah@mmu.edu.my

Abstract: Content Based Image Retrieval (CBIR) is a difficult task. It needs different approaches and tools to get a set of images similar to a complex query image. Finding appropriate features is one of the important issues in this process. There are several tools to extract the appropriate features from the images. In this paper we explore the usage of 1-D Discrete Cosine Transform (DCT) as a tool for finding features of the boundary of objects in the query and database images. It is shown that the sensitivity of the DCT to rotation of objects can be reduced and the 1-D DCT of extended boundary signals can even give better results than the DFT in shape retrieval process.

I. INTRODUCTION

Efficient automatic image retrieval is a difficult task. Shapes of objects in the images are important elements for finding the best matches or retrieving the query image from database. Boundary of objects has the most information of the shapes. Features of the boundaries are usually used for decreasing the redundancies of the data for efficient matching and retrieval.

One of the important processes in trademark image retrieval is feature extraction. Good features are those that are robust under small affine transformation, scale, translation and rotation. Some features that has been used by researchers are Fourier descriptor [4], Hue's invariant moments [5], Eigen values [6], Zernike moments [7] and combination of some of them [8]. Amongst these features Fourier descriptor has the capability to be applied to the boundary of an object in the image assuming that coordinates of any pixel on the boundary is a complex number and the set of these numbers makes a 1-D array. Fourier descriptor is the Discrete Fourier Transform (the DFT) of this one dimensional complex array.

Except Fourier descriptors, other types of the mentioned features are usually applied to the 2-D regions and give the features of the contents of the image in that region. One of the problems with these features is that the descriptors extracted from similar shapes may be very different and vice versa. This means the value of a similarity measure for two similar shapes may be very different. So, the retrieved images may not satisfy human perception. These features are called non-geometric features [8].

One of the descriptors which can carry information of a boundary is the DFT. The DFT of the boundary of a region is one of powerful features that are usually used for object

recognition [9]. One of the special cases of the DFT (or the discrete time Fourier series) is the Discrete Cosine Transform (DCT). In this paper we intend to examine the effectiveness of 1-D DCT as the descriptor of a boundary in trademark images in compare with the DFT. In the Section II we explain more about the DCT and its relationship with the DFT. In Section III the signal modifications for getting the best result from the DCT coefficients are discussed. Section IV shows the results and at the end we bring the conclusion.

II. THE DCT

The DCT [10] is one of the famous transformations that has benefited communications and information technology community especially in the area of speech and image storage and transmission [11-13]. The wide applications of the DCT are because of energy compactness of a signal in its DCT coefficients and availability of algorithms and hardware for its fast computation [12, 13]. Reeves and Kubik [14] proposed techniques based on properties of the DCT with improvements in computation time and matching quality.

In color and gray level image retrieval, the DCT is usually used as an effective tool for indexing [15]. In these applications the 2-D DCT is used.

In CBIR, shapes in the image are important. The region or boundary of objects is usually extracted after a pre-processing step. Sometimes the boundary is converted to a complex 1-D array and its features are used for retrieval process. The DFT, as one of the effective transformations is utilized for feature extraction. The coefficients of the 1-D DFT are usually combined with other features like invariant moments to get a better result in retrieval process [16].

In the DFT implementation, multiplications and additions are done on complex numbers. But in the DCT all calculations are real numbers. In this paper we try to examine the 1-D DCT instead of the 1-D DFT for increasing the efficiency in trademark image retrieval process.

A. The DCT and the DFT formulas and their relationship

There are four different types of DCT [17]. In most cases, the second type is used as the DCT. The formulas of the type 2 - one dimensional DCT and IDCT are as follow:

$$X^C(k) = w(k) \sum_{n=1}^N x(n) \cos\left(\frac{\pi(2n-1)k}{2N}\right) \quad k = 1, 2, \dots, N \quad (1)$$

$$x(n) = \sum_{k=0}^{N-1} w(k) \cdot X^C(k) \cos\left(\frac{\pi(2n-1)k}{2N}\right) \quad n = 1, 2, \dots, N$$

where

$$w(k) = \begin{cases} \frac{1}{N} & k = 1 \\ \frac{2}{\sqrt{N}} & 2 \leq k \leq N \end{cases}$$

$X^C(k)$ is the unitary discrete cosine transform of $x(n)$, N is the length of x , and x and $X^C(k)$ are of the same size.

As a comparison, the DFT and IDFT formulas are as follow:

$$\begin{aligned} \hat{X}(k) &= \sum_{n=1}^N x(n) W^{jk} & k = 1, 2, \dots, N \\ x(n) &= \frac{1}{N} \sum_{k=1}^N \hat{X}(k) W^{-jk} & n = 1, 2, \dots, N \end{aligned} \quad (2)$$

where $w_N = e^{(-2\pi i)/N}$ and \hat{X} is the DFT of x

The DCT coefficients are the coefficients of a cosine series expansion of a band-limited, symmetrically extended discrete signal and the DFT coefficients are the coefficients of an exponential series expansion of the signal without making any odd or even symmetry in the expansion of the original signal. The DCT of a signal has a simple relationship with the real part of the DFT of the modified signal [17].

There are three major advantages on the DCT in compare with the DFT of a signal.

1. Computational complexity: Because of real multiplication and additions in the DCT and the IDCT formulas its calculation is done faster.
2. Energy compactness: Symmetrically extended original signal produces a smoother periodic signal. Because of this the coefficients belong to lower frequencies can represent the signal better than the DFT coefficients. A comparison has been done in [18] for a small size 2-D signal (image). This means that the energy of the signal is compacted in the low frequencies more than the DFT.
3. Gibbs phenomena: In lossy compression as well as image retrieval we use a limited number of coefficients as the descriptors of the signal. Reconstruction of a signal using a limited number of coefficients by the DCT produces some ringing caused by Gibbs phenomenon. The error produced by this effect in the DCT is less than the DFT.

However, there are two major disadvantages of the DCT in compare with the DFT.

- The DCT is not rotational invariant,
- The DCT is not shift invariant.

In the DFT, the absolute value of the coefficients are invariant under rotation and the normalized AC coefficients are invariant under shifting the signal (normalization is done by dividing the AC coefficients with the DC coefficient) [9].

Despite of the advantages of the DCT, these two disadvantages caused some limitations in the usage of the DCT as a descriptor in image retrieval.

In the next Section we examine the usage of the DCT in trademark image retrieval.

III. THE DCT IN IMAGE RETRIEVAL SYSTEM

Finding the boundary of objects and shape analysis play an important role in retrieval process. For increasing the accuracy in shape retrieval from digital image database, the following steps are taken:

- 1- Image is Smoothed for Noise removal
- 2- Smoothed image is converted to binary image
- 3- Binary image is segmented
- 4- Objects are detected and cropped
- 5- Size and rotation of each object is adjusted
- 6- Shape features of the objects (such as moments) and/or their Boundaries (such as DFT) are extracted
- 7- These features are matched with the corresponding features of the objects in the database
- 8- The images of the database are ranked based on matching of the features and the similarity measures

The first five steps are pre-processing steps. Some of researchers may not use all of these steps because of the nature of their images.

In our work, we used MPEG7_CE-Shape-1_Part_B database which consists of about 1400 binary images. For this reason we started from step 3. Each of these images has only one object and some of them have similar objects from different views or with different translations. In this work, the 1-D DFT or the 1-D DCT has been used for feature extraction from object boundaries. Because of sensitivity of the DCT to the shift and rotation, we used step 5 for getting a better result. For matching of two objects, we calculate the distance between their feature vectors. The norm of this vector is used as similarity measure for ranking the similarity of query image and the images in the database.

In the next subsection we see how a 1-D DCT is used for feature extraction and shape analysis of a two dimensional object boundary.

A. Shape analysis using the 1-D DCT

Shape analysis using the 1-D DCT method is copped with 2-D boundary of objects. A shape in 2-D space can be described by a vector of N complex numbers (We represent vectors by bold letters):

$$\begin{aligned} \mathbf{z} &= [z_1, z_2, \dots, z_N]^T \\ z_i &= x_i + j \times y_i \quad i = 1, 2, \dots, N \end{aligned} \quad (3)$$

A prototype boundary and its n landmarks is shown in Fig. 1.

If we use the DFT as the descriptor of this boundary, vector \mathbf{z} can be used in (1) as a 1-D complex signal in time domain and the result will be its DFT signal in frequency domain.

But in the DCT, we can not use complex signals in (2). For this reason, we produce two real vectors from x and y values of signal \mathbf{z} as follows (we call them boundary signals):

$$\mathbf{x} = [x_1, x_2, \dots, x_N]^T \quad \mathbf{y} = [y_1, y_2, \dots, y_N]^T$$

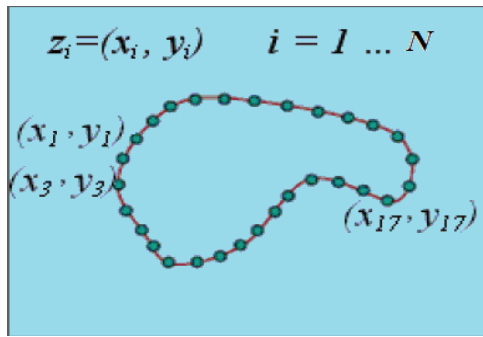


Fig. 1. A prototype boundary and its N landmarks.

In this case, we can calculate the DCT of each of these signals. We call $X^c(k)$ and $Y^c(k)$ the DCT of these two signals respectively. These signals are real and both of them or a combination of them can be used as a shape descriptor.

In the first approach we examined the following data as a feature vector in matching process.

$$\begin{bmatrix} X^c_1 & Y^c_1 & X^c_2 & Y^c_2 & \dots & X^c_N & Y^c_N \end{bmatrix}$$

The results of this approach were not good enough to bring in this paper.

In the second approach, a complex signal using $X^c(k)$ and $Y^c(k)$ has been made and its absolute value, after normalization, is utilized as feature vector for shape description.

Assume that the following vector is made from the two DCTs,

$$Z^c = [(X^c_1 + jY^c_1) \quad (X^c_2 + jY^c_2) \quad \dots \quad (X^c_N + jY^c_N)]$$

now a modified and normalized version of this signal means setting the DC element to zero and the X and Y in the AC elements is divided to the first corresponding non zero AC element of their vectors as shown in the following.

$$Z^c_m = [(0+j0) \quad (X^c_2 / X^c_2 + j Y^c_2 / Y^c_2) \quad \dots \quad (X^c_N / X^c_2 + j Y^c_N / Y^c_2)]$$

Then the absolute value of the new vector Z^c_m , is used as the one dimensional DCT shape descriptor.

This process is similar to the one done on the DFT descriptors for pattern recognition except that the AC elements of the vector z are simply divided by the first AC element which is a complex number.

In the next Section we modify the boundary signals and their DCT are used in the retrieval process.

IV. MODIFICATION IN THE SIGNAL AND SOME RESULTS

A. For getting a better result using the DCT, some modifications have been done on the boundary signals. For a better comparison, at the first attempt the original signals have been used in the DCT formula. At the second attempt, extended versions of the signals have been used in the retrieval process.

In our work, the general system shown in Fig. 2 has been used for trademark image retrieval. It means corresponding to any type of signal (extended and non extended) and the type of feature extraction system (the DCT or the DFT) we have to produce a corresponding image feature database.

As mentioned before, we used MPEG7_CE-Shape-1_Part_B database which consists of about 1400 binary images of 70 classes. For a better performance, size of all images has been converted to 100x100 pixels.

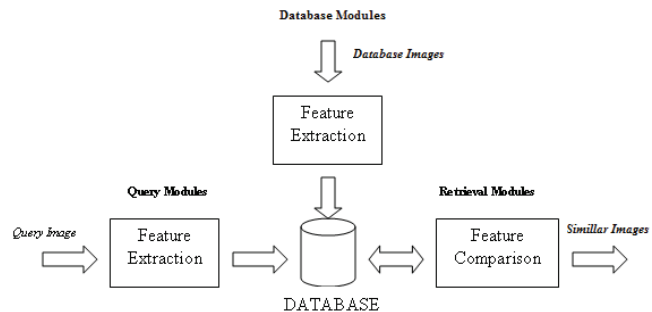


Fig. 2. Block diagram of image retrieval system

One of the images in this database is personal_car-12.gif. We use this image as query image for showing the procedure (Figs 3-4)

In Fig 3-c, The angle of rotation for the object in image is the angle (in degrees) between the x -axis and the major axis of the ellipse that has the same second-moments as the region. This calculation is supported by MATLAB.

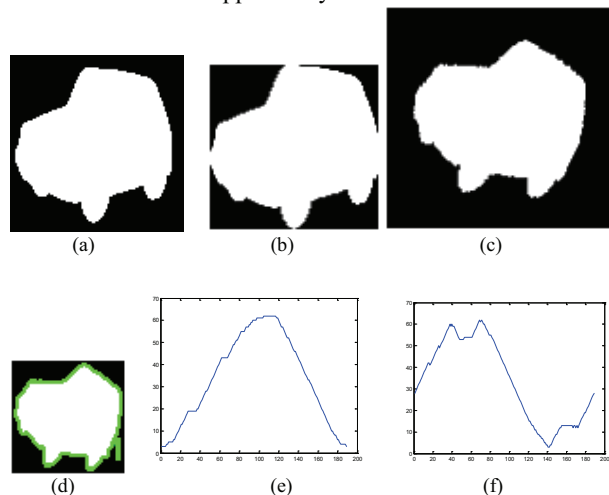


Fig.3. The procedure. a-original image personal_car-12.gif, b-The cropped car, c- Rotation, d- again car is cropped and its boundary is extracted.(Number of pixels of the boundary is 189 and its starting point is (3, 28)), e- plot of the vector X of the boundary, f- plot of the vector Y of the boundary

If we complete the retrieval procedure using 20 coefficients of the DCT, we get the result as shown in Fig. 5-a. This result is not comparable to the result obtained by the DFT shown in Fig. 5 -b.

Most of the problem in the result of the DCT is its sensitivity to the direction or rotation. To decrease this sensitivity we examined the extension of the x and y signals such that the rotated shape of object (car) also is considered in the procedure. For example, some of the possibilities for extension of the x and y signals are signal production with two, three, or four times of the original length (i.e. $2N$, $3N$, or $4N$) with the following relationships. In these cases the actual boundaries under process are as shown in Fig. 6.

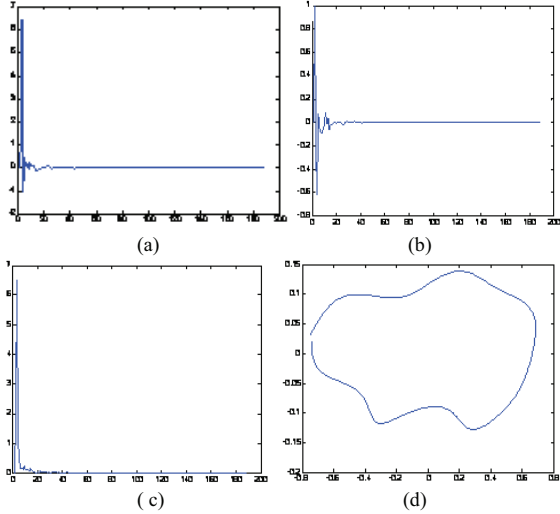


Fig. 4. The normalized DCT of the signals. a- the DCT of signal x shown in Fig. 3(e), b- the DCT of signal y shown in Fig. 3(f), c- the absolute value of signal $DCT(x)+j*DCT(y)$, d- reconstructed car boundary using the IDCT of the first 20 coefficients of the DCTs of x and y with length 189.

$$\begin{aligned}
 X_{e2} &= x(n) - x(-n + 2N) \\
 Y_{e2} &= y(n) - y(-n + 2N) \\
 X_{e3} &= x(n) + x(-n + 2N) - x(n - 3N) + x(-n + 4N) \\
 Y_{e3} &= y(n) - y(-n + 2N) - y(n - 3N) + y(-n + 4N) \\
 X_{e4} &= x(-n + N) + x(n - N) - x(-n + 3N) - x(n - 3N) \\
 Y_{e4} &= y(-n + N) + y(n - N) - y(-n + 3N) - y(n - 3N)
 \end{aligned}$$

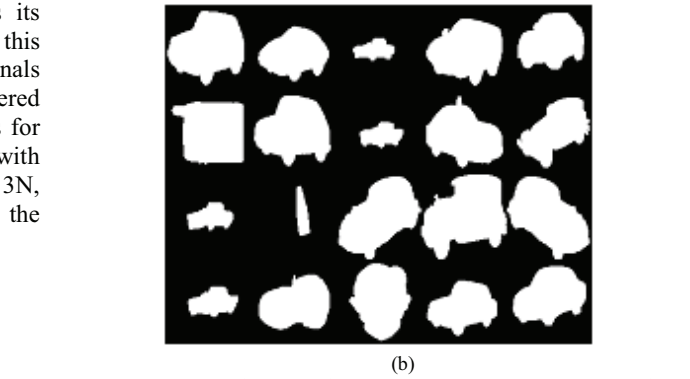
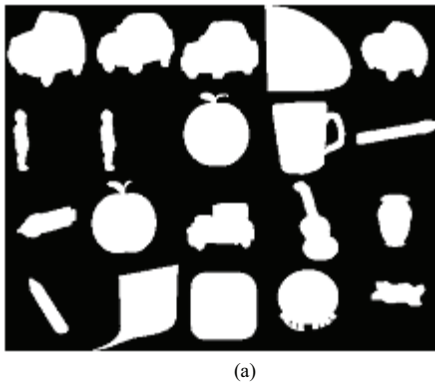


Fig. 5: The result of the retrieval. a-using 20 out of 189 coefficients of the 1-D DCT, b- using 20 out of 189 coefficients of the DFT.

We know that there are a lot of possibilities for signal extensions. One of the main reasons for choosing the type of extension of the boundary signals is the effect of Gibbs phenomena. We have to choose an extended sequence that has the least discontinuities. In this case, we may have a better approximation when we reconstruct the object by a limited number of coefficients. Other reasons for selections are the number of coefficients that is sufficient for incorporating in the matching process and reduction of sensitivity of the DCT to rotation of the object. Please note that in these types of extensions some of the DCT coefficients always are zero. For example, if we extend our signals by Equ. (5) odd coefficients will be zero. If we select 40 coefficients from X^c and Y^c , only 20 of them are incorporated in calculations. In this case memory needed for saving the descriptors of the database images will not be extended in compare with the usage of 20 Fourier descriptors.

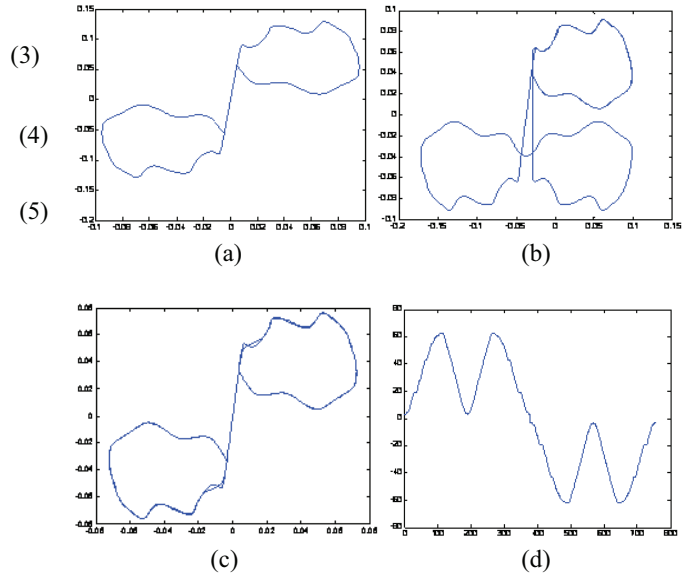
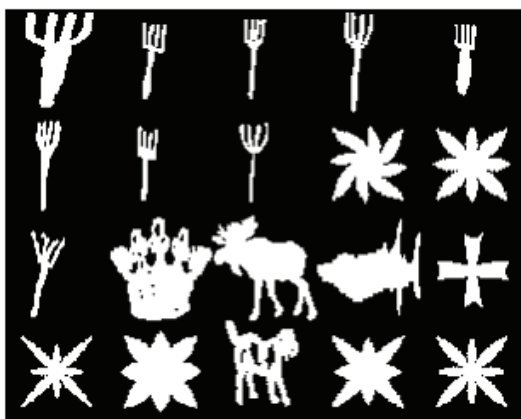


Fig. 6. Extension of signals and their corresponding boundaries made by the IDCT of a limited number of coefficients. a- based on Equ. (3), b- based on Equ. (4), c- based on Equ. (5), d- extension of the signal x in Fig 3(e) based on Equ. (3-X).

Fig. 7 shows the images retrieved by 40 coefficients (20 of them are zeros) of the DCT using (5). In the images retrieved here, there are two non car objects. In images retrieved by the DFT descriptors (Fig. 5) there are 4 non car objects. Despite reduction of sensitivity to shape changes including rotation in our work, the DCT is still variant under rotation. This might be useful for some cases. For example, when we used image fork-1.gif as the query image, the result using the proposed method is much better than the result using the DFT as descriptor (Fig. 8). In this result 20 coefficients from the Fourier descriptors (the DFT) and 20 nonzero coefficients with even indexes (40 coefficients including zero values) have been used. The result using the DFT shows that the fork has been considered as an object with a body and four legs. If the number of coefficients is increased its result will be better.



Fig. 7. Images retrieved by 40 coefficients (20 of them are zeros) of the DCT of the signals extended by (5).

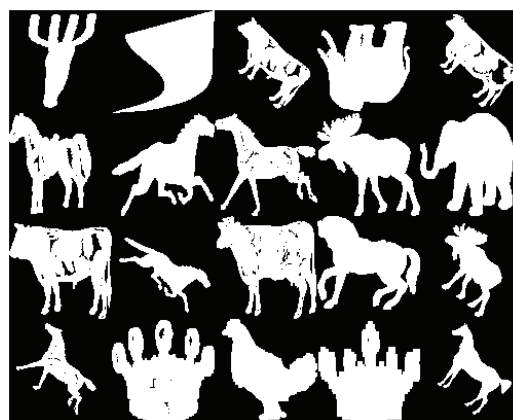


(a)

B. The execution time for the total 1400 objects included in the database is as follows (the pre-processing time is not included):

Using the DFT with 20 coefficients for each object, the total time is 7.3568 seconds. In this case total matching time is about 1.3 seconds.

Using the DCT with 20 coefficients (without signal extensions) for each object, the total time is 2.0995 seconds (It is three times faster than the DFT).



(b)

Fig. 8. The image fork-1.gif is the query image. The coefficients are the first 20 non zero values. a- The result using the proposed method (the DCT), b- the result using the DFT as descriptor.

Using the DCT with 40 coefficients (with signal extensions) for each object, the total time is 6.7525 seconds. In this case total matching time is about 1.11 seconds.

The retrieval system is implemented in a platform of Pentium 4 with 1.7 GHz CPU and 1G memory. MATLAB Ver. 7.4.0(R2007a) has been used for implementing the programs.

V. CONCLUSION

In this paper, we examined the usage of the 1-D DCT in shape based image retrieval. We showed that an appropriate extension of the boundary signals reduces the sensitivity of the DCT to object rotation. Because of energy compactness property of the DCT, a limited number of the 1-D DCT coefficients of an object boundary signals can represent the shape of the object. Our experiments show that, with a limited number of coefficients, using the appropriate extension of the boundary signals, in trademark image retrieval, the DCT can act as good as the DFT and sometimes even better than the DFT. The DCT is faster and simpler for implementation in shape based image retrieval than the DFT. Finding the best extension of the boundary signals for reduction of the sensitivity of the 1-D DCT to rotation of object, in trademark image retrieval application, is still under investigation.

REFERENCES

- [1] Z. Hong, Q. Jiang, "Hybrid Content-based Trademark Retrieval using Region and Contour Features", 22nd IEEE International Conference on Advanced Information Networking, 2008.
- [2] C. Venters, M. Cooper, "A Review of Content-Based Image Retrieval Systems", Manchester Visualization Centre, Manchester Computing, University of Manchester, 2000.
- [3] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, 2000.

- [4] Sheng, “Orthogonal Fourier-Mellin moments for invariant Pattern Recognition”, J. Opt. Soc. Am, A/Vol. 11, No.6, pp. 1748-1757, 1994.
- [5] Hu, “Visual pattern recognition by moment invariants”, IRE Transaction on Information Theory, pp 179-187, 1962.
- [6] D. Mohamad, G. Sulong, “Trademark matching using Hotelling Transform”, First Int. Conf. of Engineering Computation and Computer Simulation, Changsa, China, 1995.
- [7] Z. Khotanzad, Y.H. Hong, “Invariant image recognition by Zernike moments”, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12, pp. 489–497, 1990.
- [8] H. Jianga, C. Ngoa, H. Tana, “Gestalt-based feature similarity measure in trademark database”, Pattern Recognition 39, pp. 988 – 1001, 2006.
- [9]J. T., Tou, R.C. Gonzalez, “*Pattern Recognition Principles*”, Addison-Wesly Pub. Co., 1974.
- [10] K. Rao, P. Yip, *Discrete Cosine Transform—Algorithms, Advantages, Applications*, Academic Publishers, San Diego, 1990.
- [11] G.K. Wallace, *The JPEG still-picture compression standard*, Comm. ACM 34, pp. 31–44, 1991.
- [12] K.-H. Tzou, “Video Coding Techniques: An Overview”, P. Pirsch (Ed.), *VLSI Implementations for Image Communications*, Elsevier Science Publishers, Amsterdam, pp. 1–47, 1993.
- [13] S. Lee, “Improved algorithm for efficient computation of the forward and backward MDCT in MPEG audio coder”, IEEE Trans. Circuits Syst. II—Analog Digital Signal Process. Vol. 48, pp 990–994, 2001.
- [14] R. Reeves, K. Kubik, “Shift, scaling and derivative properties for the discrete cosine transform”, Signal Processing, Vol. 86, 1597–1603, 2006.
- [15] Z.-M. Lu, H. Burkhardt, “Colour image retrieval based on DCT-domain vector quantisation index histograms”, *Electronic Letters*, Vol. 41 No. 17, 18th August 2005.
- [16] J. P. Eakins, K. J. Riley, J. D. Edwards, “Shape Feature Matching for Trademark Image Retrieval”, School of Informatics, University of Northumbria, CIVR 2003, LNCS 2728, pp. 28-38, 2003.
- [17] Oppenheim, A. V. and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1989.
- [18] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 2002.
- [19] H. Bae, S. Jung, “Image Retrieval Using Texture Based on DCT”, International Conference on Information, Communications and Signal Processing, ICICS '97 Singapore, 9-12 , September 1997.