Genetic Algorithm to find optimal GLCM features

Ruaa Mohammed Hamza   Dr. Tawfiq A. Al-Assadi
Department of Computer science   College of Information Technology

Abstract

This paper presents a novel method to Image Retrieval Based on optimal Texture Features extracted from GLCM using Genetic Algorithm . The basic approach used here is that the textures features values that extracted from gray level co-occurrence matrix (GLCM) gives the typical values for features analysis .The Genetic Algorithm finds optimal Texture Features extracted from GLCM based on the fitness function . The obtained results of different types of images like "texture" "non-texture" and as unknown images where characterized in a good range.

Keyword: optimal GLCM features, Gray-Level Co-occurrence Matrix (GLCM), Texture features.

1. <u>Introduction</u>

This paper introduces a new approach to find optimal GLCM features which are the most important in the field of image texture analysis using Genetic Algorithm (GA). The typical system performs three major tasks.

The first one is texture analysis for features extraction, There are several number of texture analysis techniques that have been used in image processing area. Generally, the texture study includes: structural, transform method, and statistical model . The most common second-order statistic is gray level co-occurrence matrix (GLCM) that used in this paper , the GLCM is essentially a two-dimensional histogram of the number of times that pairs of intensity values (or , more generally , arbitrary local features) occur in given spatial relationship . thus , it forms a summary of the sub patterns that could be formed by intensity pairs and the frequency with which they occur .

The second task is feature extraction (FE) from GLCM , For feature extraction in content based image retrieval there are mainly two approaches [5] feature extraction in spatial domain and feature extraction in transform domain [15]. The feature extraction in spatial domain includes the
CBIR techniques based on histograms [5], BTC [4, 8, 9], VQ [3,10,11]. The transform domain methods are widely used in image compression, as they give high energy compaction in transformed image. So it is obvious to use images in transformed domain for feature extraction in CBIR [11, 12]. Transform domain results in energy compaction in few elements, so large number of the coefficients of transformed image can be neglected to reduce the size of feature vector.

The third task is use Genetic Algorithm , where Genetic Algorithm (GA) is used for optimizing or selecting best features which gives a reduced feature set eventually results in high classification accuracy. Reducing the

dimensions of the feature space not only reduces the computational complexity, but also increases estimated performance of the classifiers. GA is biologically inspired and has many mechanisms resembling natural evolution [1] [6] [2].

There are other approaches that used GLCM with Genetic Algorithm in computer vision , Jestin V.K. and J.Anitha used GLCM with GA for Retinal Image Analysis [14] , Lijun Qian and Jianrong Xu used for Improvement of Feature Selection in multi-phase CT images of hepatic lesions [16]

When we propose Genetic Algorithm to find optimal GLCM features , it is necessary to allocate following points.

a. Divide image to number of blocks each with same size and give a label to each block .
b. Each chromosome is used to represent a sort of block Matrix .
c. Texture Features extraction using GLCM Matrix .
d. Calculate fitness function based on GLCM features .

2.Gray Level Co-occurrence Matrix (GLCM)

In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image. According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second-order and higher-order statistics.

The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features. GLCM introduced by Haralick . contains information about the positions of pixels having similar gray level values.

A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element $P(i, j \mid d, \theta)$ is the relative frequency with which two pixels, separated by distance d, and in direction specified by the particular angle (θ) , one with intensity i and the other with intensity j.
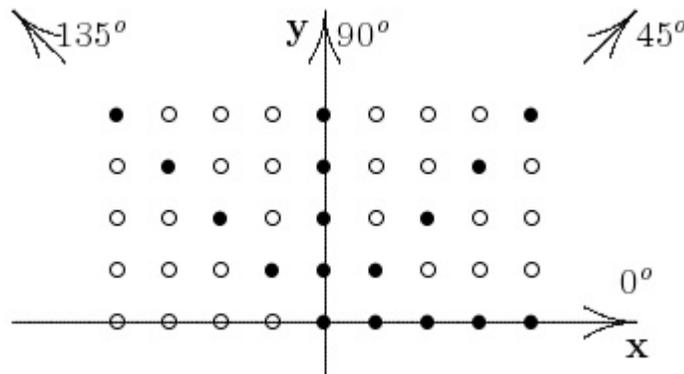
Figure 1: Geometry for measurement of gray level co-ocurrence matrix for 4 distances d and 4 angles   .

The basic GLCM algorithm is as follow:

1. Count all pairs of pixels in which the first pixel has a value i, and its matching pair displaced from the first pixel by d has a value of j.

2. This count is entered in the ith row and jth column of the matrix Pd[i,j]

3. Note that Pd[i,j] is not symmetric, since the number of pairs of pixels having gray levels[i,j]does not necessarily equal the number of pixel pairs having gray levels [j,i].

   4.The elements of Pd[i,j]can be normalized by dividing each entry by the total number of pixel pairs.

5. Normalized GLCM N[i,j], defined by:

$$\text{N}[i,j] = \frac{P[i,j]}{\sum_i \sum_j P[i,j]}$$

3. <u>Texture features</u>

In 1973, Haralick suggested a set of 14 textural features which can be extracted from the co-occurrence matrix, and which contain information about image textural characteristics such as Angular Second Moment , Contrast, Correlation, Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Mean of Correlation, and etc. in our investigation five Haralick feature functions were computed . They are Energy , Entropy, Contrast, Correlation , and Homogeneity. Energy Returns the sum of squared elements in the GLCM and the range will be in [0 1] , Entropy measures the randomness of intensity distribution , Contrast measure of the amount of local variation in the image , Correlation measure of image linearity , and Homogeneity Returns a value that measures the closeness of the distribution

of elements in the GLCM to the GLCM diagonal and range will be in [0 1].

The equations for these five textures features are shown in table (1) .

Energy	$\sum\limits_{m=0}^{G-1}\sum\limits_{n=0}^{G-1} P(m,n)^2$
Entropy	$\sum\limits_{m=0}^{G-1}\sum\limits_{n=0}^{G-1} p(m,n)\log p(m,n)$
Contrast	$\dfrac{1}{(G-1)^2}\sum\limits_{m=0}^{G-1}\sum\limits_{n=0}^{G-1}(m-n)^2\, p(m,n)$
Correlation	$\dfrac{\sum\limits_{m=0}^{G-1}\sum\limits_{n=0}^{G-1} mnp(m,n)-\mu_x\mu_y}{\sigma x\sigma y}$ *where* $\mu_x=\sum\limits_{m=0}^{G-1}m\sum\limits_{n=0}^{G-1}p(m,n)$ $\mu_y=\sum\limits_{n=0}^{G-1}n\sum\limits_{m=0}^{G-1}p(m,n)$ $\sigma_x=\sum\limits_{m=0}^{G-1}(m-\mu_x)^2\sum\limits_{n=0}^{G-1}p(m,n)$ $\sigma_y=\sum\limits_{n=0}^{G-1}(n-\mu_x)^2\sum\limits_{m=0}^{G-1}p(m,n)$
Homogeneity	$\sum\limits_{m=0}^{G-1}\sum\limits_{n=0}^{G-1}\dfrac{p(m,n)}{(1+|m-n|)}$

Table (1) : Equations of Texture Feature

4. <u>Genetic Algorithms</u>

The genetic algorithm (GA) is an optimization and search technique based on the principles of genetics and natural selection.GAs were first described by John Holland in the 1960s and further developed by Holland and his students and colleagues at the University of Michigan in the 1960s and 1970s.An algorithm is a series of steps for solving a problem . A genetic algorithm is a problem solving method that uses genetics as its model of problem solving . it's a search technique to find approximate solutions to optimization and search problems . Basically , an optimization problem looks really simple . GA handles a population of possible solutions. Each solution is represented through a chromosome , which is just an abstract representation . Coding all the possible solutions into a chromosome is the first part , but certainly not the most straightforward one of a Genetic Algorithm . A set of reproduction operators has to be determined , too . reproduction operators are applied directly on the chromosomes , and are used to perform mutations and recombination over solutions of the problem . Selection is supposed to be able to compare each individual in the population . Selection is done by using a fitness function . Each chromosome has an associated value corresponding to the fitness of the solution it represents . The fitness should correspond to an evaluation of how good the candidate solution is . The optimal solution is the one, which maximizes the fitness function . Once the reproduction and the fitness function have been properly defined , a Genetic Algorithm is evolved according to the same basic structure . It starts by generating an initial population of chromosomes . The first population must offer a wide diversity of genetic materials . The gene pool should be as large as possible so that any solution of the search space can be engendered . generally , the initial population is generated

randomly . Then , the genetic algorithm loops over an iteration process to make the population evolve .

Each iteration consists of the following steps :

Selection : The first step consists in selecting individuals for reproduction . This is done randomly with a probability depending on the relative fitness of the individuals so that best ones are often chosen for reproduction than poor ones .

Reproduction : in the second step , offspring are bred by the selected individuals . for generating new chromosome , the algorithm can use both recombination and mutation .

Evaluation : Then the of the new chromosomes is evaluated .

Replacement : During the last step , individuals from the old population are killed and replaced by the new ones .

The algorithm is stopped when the population converges toward the optimal solution .

The basic genetic algorithm is as follow:

1. [**Start**] Generate random population of n chromosomes.

2. [**Fitness**] Evaluate the fitness f(x) of each chromosome x in the population.

3. [**New population**] Create a new population by repeating the following :

   a) [**Selection**] Select two parents from a population according to their fitness (the   better fitness , the bigger chance to get selected) .

   b) [**Crossover**] with a crossover probability ,cross over the parents to form a new offspring (children).

   c) [**Mutation**] with a mutation probability , Mutate new offspring at each locus   (position in the chromosome).

4. [**Replace**] Use new generated population for a further sum of the algorithm

5. [**Test**] If the end condition is satisfied(i.e. desired result is optimal), stop, and return the best solution in current population

   [**Loop**] Go to step2 for fitness evaluation .

5. <u>The Proposed System</u>

The proposed system consists of two stages as below:-
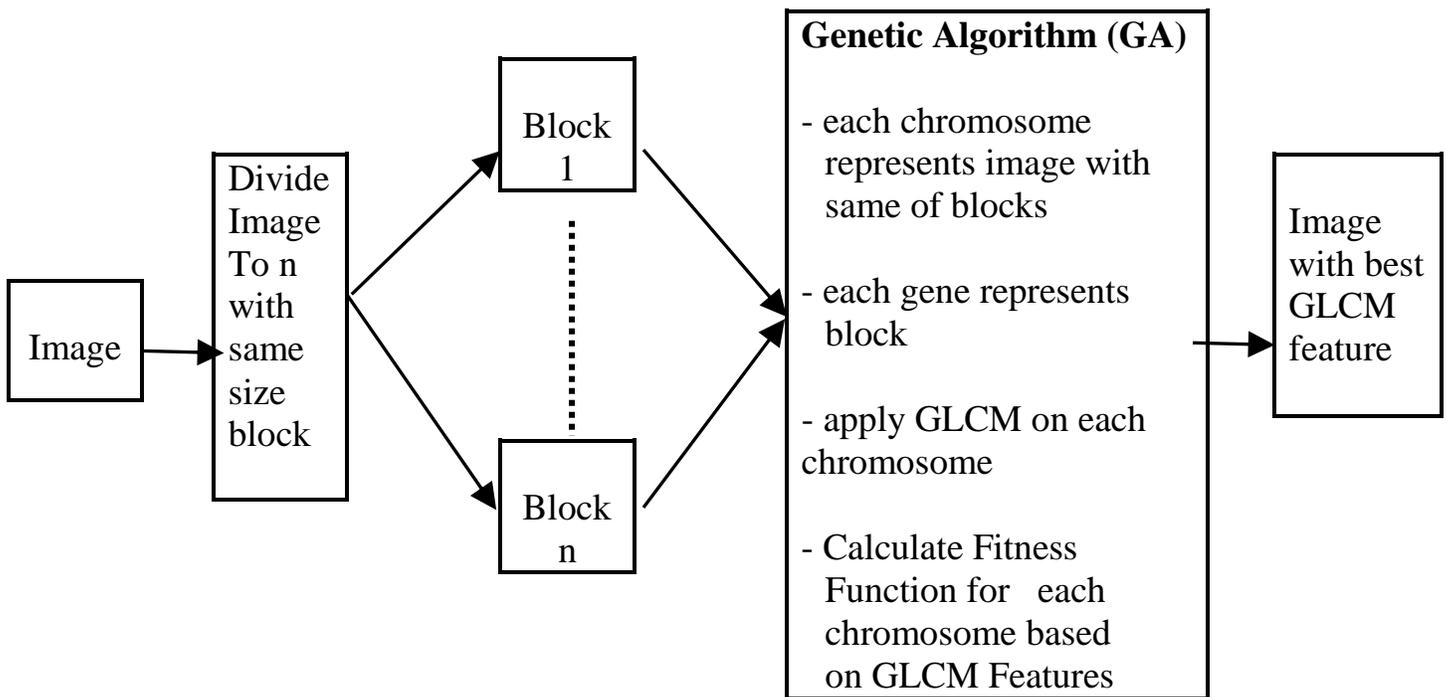
Figure 2: block diagram for the proposed system

<u>First Stage</u> : Image Dividing

take the image to be analyzed and Divide it to number of blocks each with same size and give a label to each block (we use the label number as value of gene in the correspond chromosome) and we save the pixels of each

block in temporarily storage in order to retrieve these pixels when we construct the original image matrix .

Second Stage : Genetic Algorithm

In the second stage the genetic algorithm will be used to find optimal GLCM features which are :

1. **Initialization** : generate the initial population randomly, the initial population consists of a number of chromosomes , each chromosome consists of a number of genes (number (label) of block in the blocks matrix that results from the first stage) , then represent each chromosome to correspond which a case of sort block of the matrix .

2. **Gray Level Co-occurrence Matrix (GLCM)**: apply GLCM to each matrix with the corresponding chromosome , in this work we use GLCM with distances d=1 and angle of $3\pi/4$.
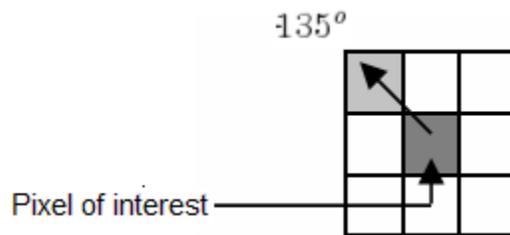
Figure 3 : Geometry for measurement of gray level co-ocurrence matrix for distances d=1 and angles $\theta= 3\pi/4$.

3 . **Fitness Evaluation** : for each GLCM matrix , we calculate the fitness function based on GLCM features using equations in table (1) .

4. **Selection (Rank Selection**) : select two individuals from population at random . the individual with the highest evaluation become the first parent . repeat to find a second parent.

10

5 . Crossover (Cycle Crossover) : apply crossover method (we use Cycle Crossover) to the two parents from previous step , the operator works as follows :

Each allele comes from one parent together with its position.

Informal procedure:

1. Make a cycle of genes from parent1 in the following way.

(a) Start with the first gene of parent1.

(b) Look at the gene at the *same position* in parent2.

(c) Go to the position with the *same* gene in parent1.

(d) Add this gene to the cycle.

(e) Repeat step b through d until you arrive at the first gene of parent1.

2. Put the genes of the cycle in the first child on the positions they have in the first parent.

3. Take next cycle from second parent

for example :
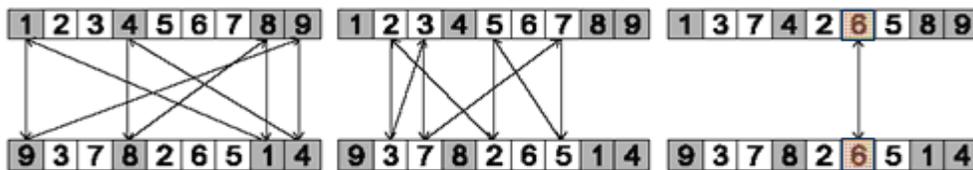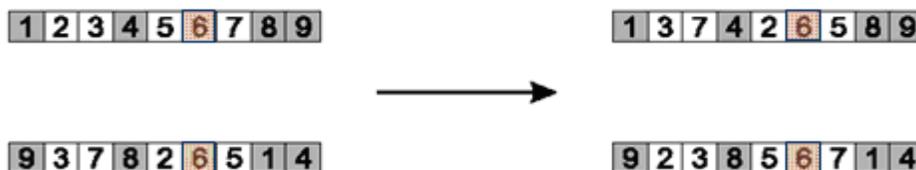
Figure 4 : Cycle Crossover

6 . **Mutation (Reciprocal Exchange)** : after crossover the genes are subjected to mutation (we use Reciprocal Exchange method) , in reciprocal exchange method each gene in both child subject to mutation with a probability rate (mut-rate =0.05), the operator works as follows :

   a. Start with the first gene of first child.

   b. generate a random float 'mut' .

   c. examine ,if mut is smaller than mut-rate , select a random position in child1, swap between a gene in this position and a gene from step (a) . if mut is greater than mut-rate , no mutation .

   d. repeat steps a through c until completing the first child's genes .

   e. repeat steps a through d to the second child .

after mutation apply GLCM to the offspring , calculate the fitness function to each GLCM matrix .

7 . **Replacement** : the last step is replacement , we use replacement to determine which of the current members of the population, if any, should be replaced by the new solutions ,there are two strategies are generational and steady state replacement , we use steady state replacement with weak method . the operator works as follows :

   a. start with the first child .

   b. select two chromosomes (chromosome1, chromosome2) at random from the previous population to the current population .

   c. compare , chromosome with highest fitness with the first child and select from them that has highest fitness to current population .

   d. repeat a through c to second child .

8. in this work we use steady state , new individuals are inserted in the population as soon as they are created , repeat 4 through 7 until the one population is completed .

9. continue in generate new populations by repeat steps 4 through 8 until each populations is completed ,in other word until reach to last population (the population contains the optimal GLCM feature).

5. <u>Experiments and Results</u>

To test the performance of our system , we take texture image and passing it to proposed algorithm in chapter three which divide image to number of blocks each with same size, generate initial population at random consists of a number of chromosomes , each chromosome consists of the same blocks but different from each other in order of blocks , each chromosome is passing to GLCM for extracting features , selection ,crossover ,replacement ,continued in generate number of generations until reached to the last generation , the result we obtained is one of the chromosomes in the last generation is optimal chromosome (optimal GLCM features) , after that we take image for other type , the result we obtained is optimal GLCM features

 (a) Original Image   (b) Optimal Features   (c) Image After Retrieve

Figure 5 : Genetic Algorithm find optimal GLCM features for these image
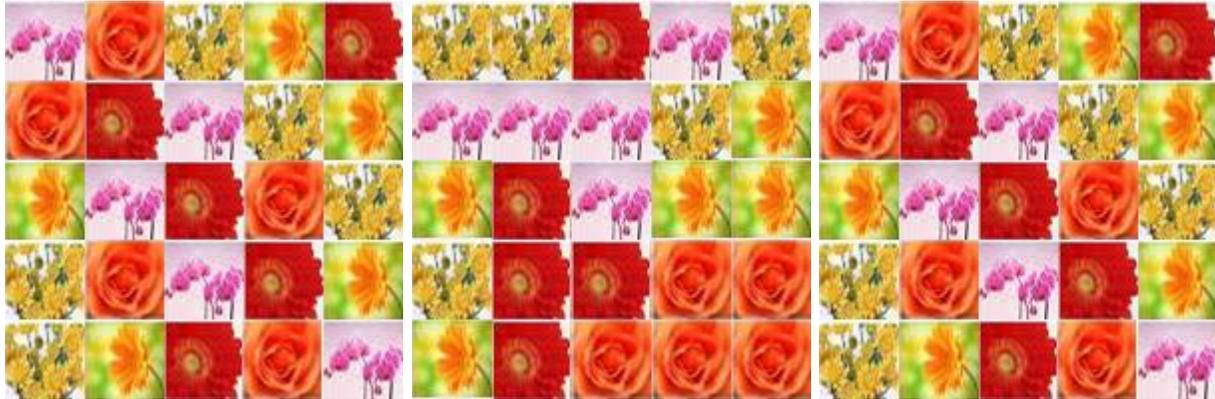
(d) Original Image   (e) Optimal Features   (f) Image After Retrieve

Figure 6 : Genetic Algorithm find optimal GLCM features for these image

6. Conclusion

The ultimate goal of this study is to evaluate genetic algorithms in object-based analysis to help find optimal GLCM features . This method is designed and tested for texture images then developed for any type of images . it is based on the concept of the texture feature for digital images Using GLCM . The data being analyzed on the images should be minimized to only the useful features value that is considered important in evaluate function. So the most important value was in homogeneity values return a value that where measures the closeness of the distribution of elements in the GLCM in different viewing direction such as 00, 450, 900 and 1800 is considered as a good feature extracted from the images to evaluate function in genetic algorithm .

Reference

[1] Michael L. Raymer, William F. Punch, Erik D. Goodman, Leslie A. Kuhn, And Anil K. Jain, July 2000, ―Dimensionality reduction using genetic algorithms‖, IEEE Transactions On Evolutionary Computation", Vol. 4, No. 2, pp.164-172.

[2] Il-Seok Oh, Member IEEE, Jin-Seon Lee, and Byung-Ro Moon, Member IEEE, Nov 2004, ―Hybrid genetic algorithms for feature selection‖, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 26, No.11, pp.1424-1437

[3] B.G.Prasad, K.K. Biswas, and S. K. Gupta, "Region –based image retrieval using integrated color, shape, and location index",International Journal on Computer Vision and Image Understanding Special Issue: Colour for Image Indexing and Retrieval, Volume 94, Issues 1-3, April-June 2004, pp.193-233.

[4] H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET Int. Journal of Electrical, Computer and System Engineering (IJECSE), Vol.2, No.3, pp. 172-180, Summer 2008. Available online at http://www.waset.org/ijecse/v2/v2-3-23.pdf

[5] H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), www.ceser.res.in/iji.html, Volume 1, No. A08, pp. 31-46, Autumn 2008.

[6] Enrique J. Carmona , Mariano Rincon , Julian Garcia- Feijoo Jose M. Martinez-De-La-Casa, 2008, ―Identification of the optic nerve head with genetic algorithms‖, Artificial Intelligence In Medicine,Vol.43, pp. 243-259.

[7] Fritz Albregtsen , "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices " , November 5, 2008 .

[8] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM Int. Conference on Advances in Computing, Comm. and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. CRCE, Mumbai. Is uploaded at ACM portal.

[9] H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Vol 2, No A09, Autumn 2009, pp. 2-14. Available online at www.ceser.res.in/iji.html.

[10] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Vol 2, No. A09,

Autumn 2009,pp. 55-65. Available online at www.ceser.res.in/iji.html.

[11] H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), Volume 9, Issue 5, pp.: 1-8, September 2009. Available online at http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html.

[12] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Non-Involutional Orthogonal Kekre's Transform", International Journal of Multidisciplinary Research and Advances in Engineering (IJMRAE), Ascent Publication House, 2009, Volume 1, No. I, 2009.

[13] H.B.Kekre, Sudeep D. Thepade, "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image ", International Journal of Information Retrieval, Serials Publications, Volume 2, Issue 1, 2009, pp. 72-79 (ISSN: 0974-6285).

[14] Lijun Qian ,Jianrong Xu , "Improvement of Feature Selection in multi-phase CT images of hepatic lesions" , 2009 .

[15] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo, "Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform", Int. Journal of Engineering Science and Technology (IJEST), Volume 2, Number 4, 2010, pp.362-371. (ISSN:0975-5462) Available online at http://www.ijest.info/docs/IJEST10-02-04-05.pdf.

[16] Jestin V.K. , J.Anitha , "Genetic Algorithm for Retinal Image Analysis" , 2011 .