2014 IEEE Students' Conference on Electrical, Electronics and Computer Science

Fusion of Local and Global Features using Stationary Wavelet Transform for Efficient Content Based Image Retrieval

Manoj D. Chaudhary
P.G Student, Dept. of Electronics & Communication
L.D. College of Engineering
Ahmedabad, India
chaudharymanojd@gmail.com

Abhay B. Upadhyay
Asst. Professor, Dept. of Electronics & Communication
L.D. College of Engineering
Ahmedabad, India
abupadhyay1979@gmail.com

***Abstract*—In this paper we propose a hybrid approach for Content Based Image Retrieval that takes into account both global as well as local features of an image. Towards this, first Stationary Wavelet Transform is applied on query image to extract horizontal, vertical and diagonal detail matrices. Stationary Wavelet Transform is used because of its translational invariant property. After this global textural features are extracted using Gray level Co-occurrence Matrix for each of these sub-matrices. To aid the retrieval process, a local descriptor is also computed by splitting the image into sub-regions. Finally Euclidean distance is used to retrieve the relevant results. Experimental results show that the proposed approach provides significant improvement over existing methods.**

Keywords—Content Based Image Retrieval, Euclidean distance, Gray Level Co-occurrence Matrix, Stationary Wavelet Transform.

I. Introduction

Due to the rapid development of computing hardware, digital acquisition of information has become one of popular method in recent years. Every day, Giga-bytes of images are generated from various sources. Military, Medical, Journalism are few to name among them. Use of digital content necessitates the development of effective ways for management and retrieval of visual information. Basically images can be retrieved in two different ways, firstly, by text and secondly, by content. Text based approach is very well known and widely used. In this approach a user can search a desired image by entering keywords related to the image. Though the method is simple, it suffers from the problem of human perception based image annotation. This means that the keywords used to search the image are language dependent. An alternative to text based approach is Content Based Image Retrieval (CBIR) [1]. CBIR is a technique which uses visual features of an image such as color, shape, texture, etc…to search images from a large database similar to user's query [2]. The CBIR system have been used in variety applications such as Crime prevention, Architectural and Engineering design, Fashion and Interior design, Journalism and Advertising, Medical diagnosis, Home Entertainment and Web searching etc.

II. Literature Survey

Literature survey is important for understanding and gaining more knowledge about a specific area of the subject. Towards this a brief overview of the different existing techniques for CBIR is presented. The authors in [3] have used Ripplet transform for feature extraction followed by multilayered perceptron as a classifier for retrieving relevant images. Reference [4] and [5] describe methods based on Curvelet Transform for image retrieval. Work proposed in [6] presents a methodology based on Scale Invariant Feature Transform to aid the retrieval process. The authors in [7] have utilized a combination of dual-tree complex wavelet transform and generalized Gaussian density model for feature extraction. The method has been tested to give promising results on Messidor Retinal image database. Reference [8] evaluates the performance of adapted non-separable wavelet filter bank as well as separable wavelet filter bank on Diabetic Retinopathy Database consisting of 1045 photographs. The work proposed in [9] compares the effectiveness of several techniques on the basis of feature extraction in retrieving CT brain images.

The rest of the paper is organized as follows. Section III gives detailed description of the proposed methodology. Section IV provides experimental results showing effectiveness of the method. Section V concludes the paper by presenting the scope for future work.

III. Methodology

Feature extraction is an important step in any CBIR system. The retrieval accuracy highly depends upon the extent the given feature vector represents the image under test. In our approach we have considered both local as well as global features.

A. Computation of Global Features

1) Stationary Wavelet Transformation:

Discrete Wavelet transform has been widely used for feature extraction in CBIR systems [10]-[12]. However DWT suffers from time variant property. This means that DWT of a

978-1-4799-2526-1/14/$31.00 ©2014 IEEE

translated version of a signal X is not equal to the translated version of DWT of X [13]. Time invariance is desired in many applications such as Change detection, De-noising, Pattern Recognition, etc. To overcome the limitations of the traditional wavelet transform, we have used a multi-layer Stationary Wavelet Transform (SWT) [14] in our approach.

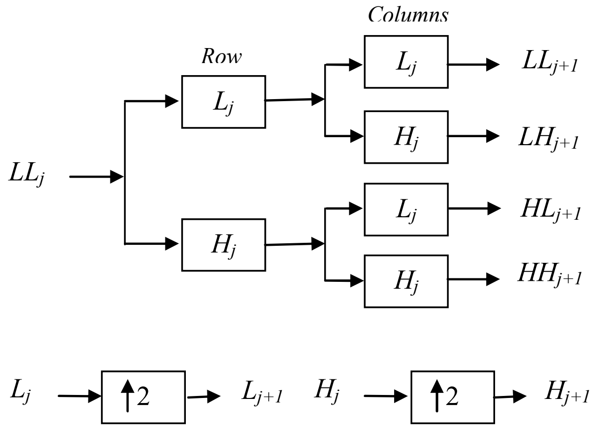

Fig. 1. Stationary Wavelet Decomposition of a two-dimensional image

In Fig. 1 H_j and L_j represent high-pass and low-pass filters at scale j, resulting from interleaved zero padding of filters H_{j-1} and L_{j-1} $(j > 1)$. LL_o is the original image and the output of the scale j, LL_j would be the input of scale $j+1$. LL_{j+1} denotes the low frequency (LF) estimation after the stationary wavelet decomposition, while LH_{j+1}, HL_{j+1} and HH_{j+1} denote the high frequency (HF) detailed information along the horizontal, vertical and diagonal directions, respectively. These sub-band images would have the same size as that of original image because no down-sampling is performed during the wavelet transformation [14]. In our approach we have used "Haar" wavelet to perform multi-layer stationary wavelet decomposition on the input 2D image. Mathematically, the wavelet decomposition can be described by (1).

$$
\begin{aligned}
LL_{j+1}(x,y) &= \sum_{m,n} L[n]L[m]LL_j(2^{j+1}m-x, 2^{j+1}n-y) \\
LH_{j+1}(x,y) &= \sum_{m,n} L[n]H[m]LL_j(2^{j+1}m-x, 2^{j+1}n-y) \\
HL_{j+1}(x,y) &= \sum_{m,n} H[n]L[m]LL_j(2^{j+1}m-x, 2^{j+1}n-y) \\
HH_{j+1}(x,y) &= \sum_{m,n} H[n]H[m]LL_j(2^{j+1}m-x, 2^{j+1}n-y)
\end{aligned} \tag{1}
$$

Where $L[.]$ and $H[.]$ represent the low-pass and high-pass filters respectively, and $LL_0(x, y) = f(x, y)$. Compared with the traditional wavelet transform, the SWT offers several advantages. First, each sub-band has the same size, so it is easier to get the relationship among the sub-bands. Second, the resolution can be retained since the original data is not decimated. Also at the same time the wavelet coefficients contain many redundant information which helps to distinguish the noise from feature. Using SWT we have extracted horizontal, vertical and diagonal detail sub-matrices.

II) Gray Level Co-occurrence Matrix:

A common technique in texture analysis involves the computation of GLCM (Gray Level Co-occurrence Matrix) as a second order texture measure [15]. GLCM describes the frequency of one gray level appearing in a specified spatial relationship with another gray level, within the area under investigation. Having computed GLCM several statistical parameters can be extracted from it. In [15], Haralick et al. described fourteen textural features that can be computed from GLCM. Reference [16] pointed out that out of fourteen only six of the textural features are considered to be the most relevant. These are *Energy, Entropy, Contrast, Variance, Correlation and Inverse Difference Moment.* In our algorithm we have employed four of them namely *Energy*, *Contrast*, *Correlation*, and *Inverse Difference Moment*. Energy is a measure of textural uniformity. When the gray level distribution over a window has either a constant or periodic form, the value of energy is maximum. Contrast is a measure of intensity contrast between a pixel and its neighbour over the whole image. Correlation is a measure of gray tone linear-dependencies in an image. High correlation values (close to 1) imply a linear relationship between the gray levels of pixel pairs. Inverse Difference Moment, also termed as Homogeneity achieves its largest value when most of the occurrences in GLCM are concentrated near the main diagonal. It is inversely proportional to contrast. The equations governing these measures are given by (2)-(5) respectively [16]. The Gray Level Co-occurrence Matrix is computed for each of horizontal, vertical and diagonal sub-matrices obtained after applying SWT on main image.

$$ene = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} g^2(i,j) \tag{2}$$

$$con = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i-j)^2 g(i,j) \tag{3}$$

$$cor = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2} \tag{4}$$

$$idm = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} [\frac{1}{(1+(i-j)^2)}]g(i,j) \tag{5}$$

B. Computation of Local Features

A global descriptor uses the visual features of the whole image, while a local descriptor takes into account the regions or objects to describe the image. To compute local features we first divide the processed image into blocks and obtain a descriptor for each block. Fig. 2 shows the method of splitting the image into three different sub-regions regions.


(a) (b) (c)

Fig. 2. Different templates for splitting images into regions: (a) Vertical crop; (b) Horizontal crop; (c) Central crop

After splitting the image into sub-regions, two statistical measures are computed for each region. These measures are mean (μ) and standard deviation (σ) [17] and are given by (6) and (7) respectively.

$$\mu = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} f(i,j)}{M \times N} \tag{6}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} (|f(i,j)| - \mu)^2}{M \times N}} \tag{7}$$

Here *M, N* represent the dimensions of the image and $f(i,j)$ represents the intensity value at an index *(i, j)*. Mean describes the average intensity over the image, while standard deviation shows amount of variation or dispersion from the mean value. Once calculated these two parameters will serve as elements of feature vector for local descriptor. Thus in total local feature vector for each sub-matrix will comprise of 6 elements. Each of vertical, horizontal and central sub-images will provide 2 measurements. While forming the feature vector the central region is given more importance by assigning a suitable weight because, this is the region where the probability of the object being located is highest compared to boundaries.

C. Feature Vector Generation

Having computed the local as well as global features for the query image the next step is to combine all these features in a single feature vector which will be used for comparison during similarity matching. We obtain a total of 12 global features, 4 from each GLCM computed over horizontal, vertical and diagonal sub-matrices. Thus the dimension of global feature vector fG is 1×12 and is represented as in (8)-(11). In addition we consider 2 local measures for each region of image (Fig. 2) yielding in total a 6 element local feature vector for each of horizontal, vertical and diagonal sub-matrices. The dimension of feature vector for each sub-matrix is 1×6. Hence the final dimensions of local feature vector fL will be 1×18 and is represented as in (12)-(15). The next step is to concatenate both the feature vectors to generate a single feature vector f_{Querry} which will represent the query image. The dimensions of this main feature vector will be 1×30 and is represented as in (16). A similar feature vector $f_{DataBase}$ having identical dimensions to query image's feature vector is computed for every database image and is used during similarity matching process.

$$fG = (f_{GH}, f_{GV}, f_{GD}) \tag{8}$$

$$f_{GH} = (f_{GHene}, f_{GHcor}, f_{GHcon}, f_{GHidm}) \tag{9}$$

$$f_{GV} = (f_{GVene}, f_{GVcor}, f_{GVcon}, f_{GVidm}) \tag{10}$$

$$f_{GD} = (f_{GDene}, f_{GDcor}, f_{GDcon}, f_{GDidm}) \tag{11}$$

$$fL = (f_{LH}, f_{LV}, f_{LD}) \tag{12}$$

$$f_{LH} = (f_{LHh\_crop}, f_{LHv\_crop}, f_{LHc\_crop}) \tag{13}$$

$$f_{LV} = (f_{LVh\_crop}, f_{LVv\_crop}, f_{LVc\_crop}) \tag{14}$$

$$f_{LD} = (f_{LDh\_crop}, f_{LDv\_crop}, f_{LDc\_crop}) \tag{15}$$

$$f_{Querry} = (fG, fL) \tag{16}$$

In the above equations f_{GH} , f_{GV} and f_{GD} represent the vectors that contain all the four GLCM measures for horizontal, vertical and diagonal sub-matrices respectively. The subscripts *ene, cor, con,* and *idm* indicate *Energy*, *Correlation*, *Contrast* and *Inverse Difference Moment* as computed from Gray Level Co-occurrence Matrix of horizontal, vertical, and diagonal sub-matrices respectively. Similarly, the vectors f_{LH} , f_{LV} , and f_{LD} contain the local measures of *Energy* and *Standard deviation* for horizontal, vertical and diagonal sub-matrices. Subscripts *v_crop, h_crop c_crop,* denote the splitted sub-images as shown in Fig. 2. Fig. 3 shows a schematic for generation of main feature vector.

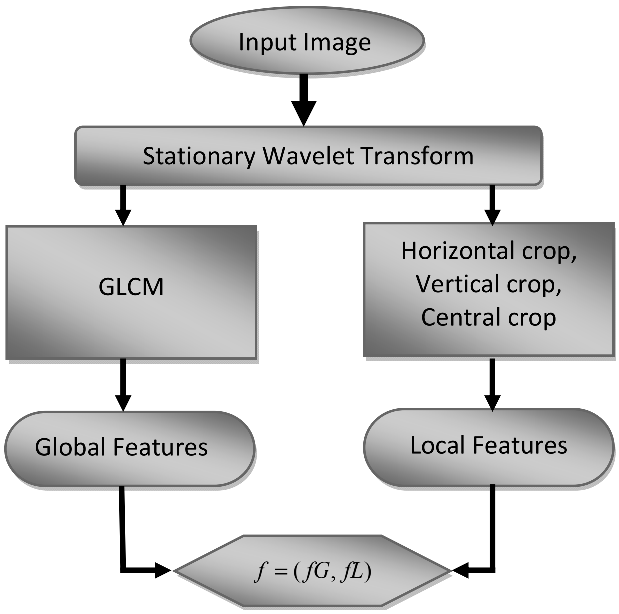

Fig. 3. Schematic for Feature Vector generation

D. Similarity Measurement

The characteristic features of the query image and all the images in database are stored in terms of feature vectors. Based on these vectors the similarity or dissimilarity between the images is computed. Methods used to classify images either measure the difference or similarity between two vectors. Two vectors with small difference will have large similarity. We have used Euclidean distance [17] which is the most common metric for measuring the distance between two vectors. Given two vectors Q and D, where

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ . \\ . \\ . \\ q_n \end{bmatrix} \text{ and } D = \begin{bmatrix} d_1 \\ d_2 \\ . \\ . \\ . \\ d_n \end{bmatrix}$$

Then the Euclidean distance between them is given by (17),

$$ED = \sqrt{\sum_{i=1}^{n} (Q_i - D_i)^2} \qquad (17)$$

Where, Q represents the feature vector of query image and D represents the feature vector of database image. Lower the distance value, larger the similarity between the two images. After computing the Euclidean distance the images are sorted in the increasing order of their distance from the query image. Finally the top 50 images are displayed as relevant to the query.

IV. EXPERIMENTAL RESULTS

The proposed method is implemented on MATLAB 7.8.0 (R2009a) on a PC with Intel Dual-Core 3rd Generation Processor having 4 GB of RAM capacity. The Database used for testing the performance of our algorithm is *Wang's Image Database* [18]. This database consists of 1000 images belonging to 10 different classes. There are 100 images from each category. The images are of size 384 × 256 or 256 × 384 pixels. Out of these 100 images from each class, 20 images are used as query, while rest 80 images serve as database from which the similar images are retrieved. Apart from above 20 images 5 additional images from sources other than Wang's Database are also used as query. Thus in all 80 images from each class are tested against 25 different queries and the performance of the algorithm is analyzed. Table I and Table II show values of global and local feature vectors for query image "602.jpg" from the Class of "Flowers".

TABLE I. VALUES OF GLOBAL FEATURE VECTOR FOR FLOWER AS QUERY IMAGE

Vector Type	Energy	Correlation	Contrast	Inverse Difference Moment
f_{GH}	f_{GHene}	f_{GHcor}	f_{GHcon}	f_{GHidm}
	0.3236	0.6943	6.7730	0.8206
f_{GV}	f_{GVene}	f_{GVcor}	f_{GVcon}	f_{GVidm}
	0.2404	0.4524	12.3700	0.7116
f_{GD}	f_{GDene}	f_{GDcor}	f_{GDcon}	f_{GDidm}
	0.2030	0.1879	14.4076	0.6012

TABLE II. VALUES OF LOCAL FEATURE VECTOR FOR FLOWER AS QUERY IMAGE

Vector Type	Horizontal Crop		Vertical Crop		Central Crop	
	μ	σ	μ	σ	μ	σ
f_{LH}	0.013	8.628	0.075	8.286	0.052	9.183
f_{LV}	0.032	9.396	0.049	8.349	0.047	10.35
f_{LD}	0.015	4.429	0.092	4.211	0.016	5.456

Fig. 4, Fig. 5, and Fig. 6 show simulation results for three different classes. The images used as query are image "602.jpg" from the class of Flowers, "159.jpg" belonging to class of Sea and "520.jpg" from the class of Elephants.

Fig. 4. Query image from class of Flower (top) and top 12 retrieved images

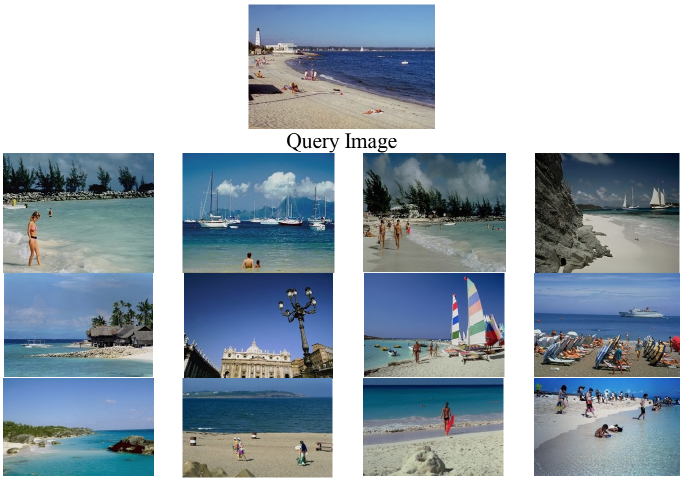

Fig. 5. Query image from class of Sea (top) and top 12 retrieved images

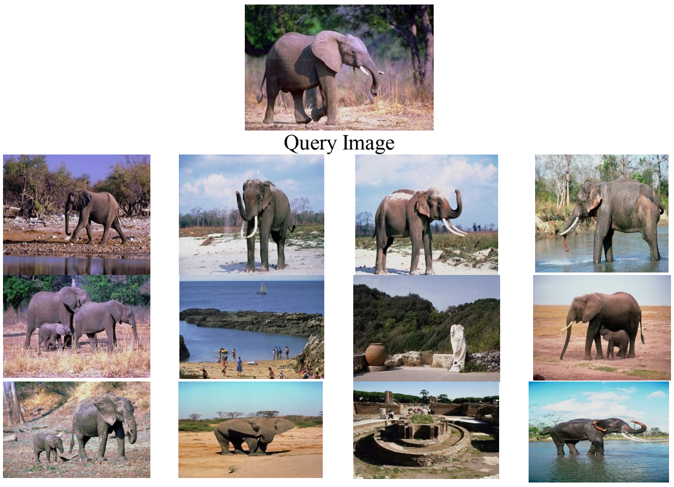

Fig. 6. Query image from class of Elephant (top) and top 12 retrieved images

A. Performance Evaluation

As a measure of performance we have used two widely used metrics of Precision and Recall [19]. Precision is a measure of ability of CBIR algorithm to retrieve only relevant images, while Recall decides the ability of CBIR algorithm to retrieve all relevant images as defined by (18) and (19) respectively.

$$P = \frac{No.\,of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \tag{18}$$

$$R = \frac{No.\,of\ relevant\ images\ retrieved}{No.\,of\ relevant\ images\ in\ the\ database} \tag{19}$$

Table III shows the values of average precision for different retrieval rates. Fig. 7 shows the Precision vs. Recall curves for four different categories. The Recall values are computed against total no. of images retrieved at a time. The values considered are 10, 20, 29, 36 and 44 respectively.

TABLE III. AVERAGE PRECISION AT DIFFERENT RETRIEVAL RATES

Average Precision				
Image Class	**No. of retrieved Images**			
	R_10	***R_20***	***R_30***	***R_40***
African	0.6	0.55	0.5	0.45
Sea	0.7	0.7	0.66	0.625
Building	0.7	0.6	0.56	0.55
Bus	1	1	0.93	0.925
Dinosaur	1	1	1	0.975
Elephant	0.6	0.55	0.4	0.375
Flower	1	1	0.967	0.925
Horse	0.8	0.75	0.73	0.65
Mountain	0.5	0.4	0.36	0.35
Food	0.5	0.5	0.467	0.425

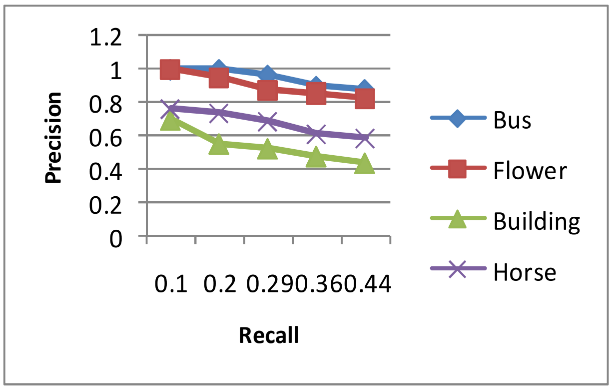

Fig. 7. Precision vs Recall plot

The accuracy of the proposed technique is compared with three different existing methods as proposed by S. Liapis et al. [20], J. Ahmed et al. [21], and Zhin-Chun et al. [22] respectively. The retrieval performance in terms of average precision for all the four methods including our method is shown in Table IV. The results clearly indicate that retrieval accuracy of our method is higher compared to above three methods. This is mainly because the proposed technique takes into account both local and global features.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a hybrid feature descriptor that utilizes both global and local properties of the image region for retrieving similar images. Both local and global features are made effective by utilizing the translational invariant property of Stationary Wavelet Transformation yielding better efficiency. Experimental results validate the strength and superiority of the proposed method over existing techniques utilizing various other textural measures. The results show that the method is superior to three earlier proposed techniques. However, still higher accuracy can be achieved by considering additional measures. As a part of future work, we look forward to incorporate color measure in a suitable form for further improving the retrieval accuracy.

TABLE IV. COMPARISON OF PROPOSED APPROACH WITH EXISTING TECHNIQUES

Average Precision in %				
Image Class	**Method Used**			
	S. Liapis et al's[20]	***J. Ahmed et al's[21]***	***Zhi-Chun Huang et al's[22]***	***Proposed Approach***
Africans	60	60	54	63
Sea	34	70	42	69
Building	32	80	16	62
Bus	92	70	67	94
Dinosaur	88	70	99	96
Elephant	28	50	40	51
Flower	86	70	97	93
Horse	80	50	96	67
Mountain	24	50	46	40
Food	33	40	79	50
Average	**55.7**	**61**	**63.6**	**68.5**

REFERENCES

[1] D. Feng, W. C Siu, and H.J. Zhang, "Fundamentals of Content Based Image Retrieval, in Multimedia Information Retrievaland Management-Technological Fundamentals and Applications", New York: Springer, 2003.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain,"Content-based image retrieval at the end of the early years," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 1349–1380, Dec-2000.

[3] Nivya Sasheedran, C. Bhuvaneswari, "An Effective Content Based Image Retrieval Approach Using Ripplet Transform", International conference on Circuits, Power and Computing Technologies, pp. 917-922, 2013.

[4] Ishrat Jahan Sumana, Md. Monirul Islam, Dengsheng Zhang, Guojun Lu, "Content Based Image Retrieval using Curvelet Transform", IEEE 10th Workshop on Multimedia Signal Processing, pp. 11-16, October 2008.

[5] L. Ni and H. C. Leng, "Curvelet Transform and Its Application in Image Retrieval", 3rd IEEE International Symposium on Multispectral Image Processing and Pattern Recognition, Proceedings of SPIE, Vol. 5286, 2003.

[6] Zhuozheng Wang, Kebin Jia, Pengyu Liu, "An Effective Web Content Based Image Retrieval Algorithm by Using SIFT Feature", World Congress on Software Engineering, pp. 291-295, 2009.

[7] Christina George Baby, D. Abraham Chandy, "Content Based Retinal Image Retrieval using Dual-Tree Complex Wavelet Transform", IEEE International Conference on Signal Processing, Image Processing and Pattern Recognition, pp. 1-5, 2013.

[8] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, "Adaptive Non-Separable Wavelet Transform via Lifting and its Application to Content-Based Image Retrieval", IEEE Transactions on Image Processing, Vol. 19, No. l, pp.25-35, 2010.

[9] Wan Siti, H. Munirah, W. Ahmad, M. Faizal and A. Fauzi, "Comparison of Different Feature Extraction Techniques in Content-Based Image Retrieval for CT Brain Images," 10th IEEE workshop on Multimedia Signal Processing, pp. 503-508, 2008.

[10] Swati Agarwal, A. K.Verma, Preetvanti Singh, "Content Based Image Retrieval using Discrete Wavelet Transform and Edge Histogram Descriptor", International Conference on Information Systems and Computer Networks, pp. 19-23, 2013.

[11] Tian Yumin, Mei Lixia, "Image Retrieval Based on Multiple Features Using Wavelet," 5th IEEE International Conference on Computational Intelligence and Multimedia Applications, pp. 137-142, 2003.

[12] P. S. Hiremath, S. Shivashankar, J. Pujari, "Wavelet Based Features for Color Texture classification with Application to CBIR," International Journal of Computer Science and Network Security, Vol. 6, No.9A, September 2006.

[13] Yudong Zhang, Zhengchao Dong, Lenan Wu, Shuihua Wang, Zhenyu Zhou, "Feature extraction of brain MRI by stationary wavelet transform", International conference on Biomedical Engineering and Computer Science, pp. 1-4, April 2010.

[14] G. P. Nason and B. W. Silverman-The stationary wavelet transform and some statistical applications in wavelet and statistics. In: Antoniadis A ed. Lecture Notes in Statistics. Berlin: Spinger Verlag, 281-299, 1995

[15] R. M. Haralick, K. Shanmugam, I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man, and Cybernetics, pp.610-621, 1973.

[16] Andrea Baraldi, Flavio Parmiggiani, "An Investigation of the Textural Charactersitics Associated with Gray Level Cooccurence Matrix Statistical Parameters", IEEE Transactions on Geoscience and Remote Sensing, Vol. 33, No. 2, pp.293-304, March 1995.

[17] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, Third Edition, Copyright © 2008.

[18] J. Z. Wang, "Wang Database," [Online], Available at: http://wang.ist.psu.edu/. Last accessed February 15, 2014.

[19] Hennig Muller, Wolfgang Muller, David McG. Squire, Stephane Marchand-Maillet, Thierry Pun, "Performance Evaluation in Content Based Image Retrieval: Overview and Proposals", Image/Video Indexing and Retrieval (Elsevier), Vol. 22, No. 5, pp. 593-601, April 2001.

[20] S. Liapis and G. Tziritas, "Color and texture image retrieval using chromaticity histograms and wavelet frames," IEEE Transactions on Multimedia, Vol.6, pp. 676–686, October 2004.

[21] Ahmed J. Afifi, Wesam M. Ashour, "Content-Based Image Retrieval Using Invariant Color and Texture Features", International Conference on Digital Image Computing Techniques and Applications, pp.1-6, December 2012.

[22] Zhi-Chun Huang, Patrick P.K. Chan, Wing W. Y. NG, Daniel S. Yeung, "Content-Based Image Retrieval using ColorMoment and Gabor Texture Feature", Ninth International Conferene on Machine Learning and Cybernetics, pp. 719-724, July 2010.