[Issue,2(1):10oct,2013] ISSN: 2321-8134

IJFEAT

INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

Discrete Wavelet Transform Based Image Mining

Mr. Ashish Bandre[#], Prof. Ramanand Singh[#], Prof. Mohd. Ahmed[#]

[#]*Department of Electronics & Communication Engineering, OIST,*

Bhopal (M.P).

[1]ashish.bandre1@gmail.com

[2]ramanandsingh@oriental.ac.in

[3]mohdahmed@oriental.ac.in

Abstract

Typical content-based image retrieval (CBIR) system would need to handle the vagueness in the user queries as well as the inherent uncertainty in image representation, similarity measure, and relevance feedback. We discuss how Histogram set theory can be effectively used for this purpose and describe an image retrieval system called HIRST (Histogram image retrieval system) which incorporates many of these ideas. HIRST can handle exemplar-based, graphical-sketch-based, as well as linguistic queries involving region labels, attributes, and spatial relations. HIRST uses Histogram attributed relational graphs (HARGs) to represent images, where each node in the graph represents an image region and each edge represents a relation between two regions. The given query is converted to a FARG, and a low-complexity Histogram graph matching algorithm is used to compare the query graph with the FARGs in the database. The use of an indexing scheme based on a leader clustering algorithm avoids an exhaustive search of the FARG database. We quantify the retrieval performance of the system in terms of several standard measures.

Keywords- CBIR, Mining, Image Processing, HSV.

1.INTRODUCTION

Retrieval of required-query-similar images from abundantly available / accessible digital images is a challenging need of today. The image retrieval techniques based on visual image content has been in-focus for more than a decade. Many web-search-engines retrieve similar images by searching and matching textual metadata associated with digital images. For better precision of the retrieved resultant images, this type of search requires associating meaningful image-descriptive-text-labels as metadata with all images of the database. Manual image labeling, known as manual image annotation, is practically difficult for exponentially increasing image database. The image search results, appearing on the first page for fired text query rose black, are shown in Figure 1 for leading web search engines Google, Yahoo and AltaVista

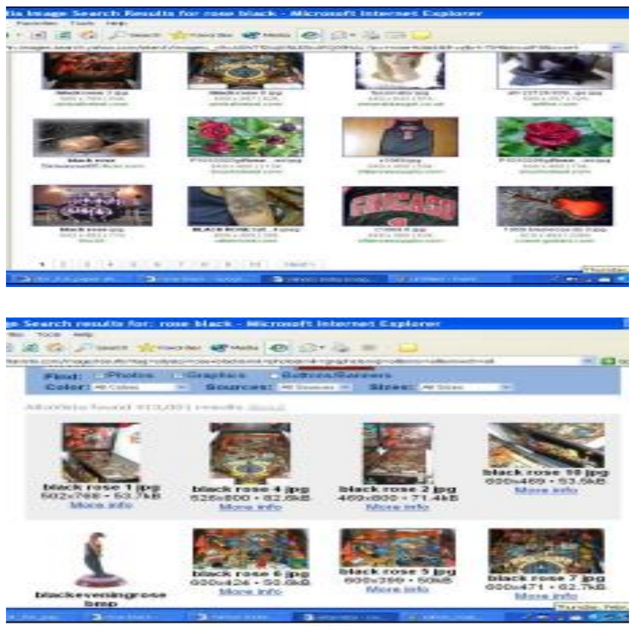

Fig. 1 Web Search Engines

Many resultant images of Figure 1 lack semantic matching with the query, showing vast scope of research leading to improvements in the state-of-art-techniques. The need evolved two solutions – automatic image annotation and content based image retrieval. The content based image retrieval techniques aim to respond to a query image (or sketch) with query-similar resultant images obtained from the image database. The database images are preprocessed for extracting and then storing –indexing corresponding image features. The query image also gets processed for extracting features which are compared with features of database images by applying appropriate similarity measures for retrieving query similar-images.

I. Literature Survey

The biggest issue for CBIR system is to incorporate versatile techniques so as to process images of diversified characteristics and categories. Many techniques for processing of low level cues are distinguished by the characteristics of domain-images. The performance of these techniques is challenged by various factors like image resolution, intra-image illumination variations, non-homogeneity of intra-region and inter-region textures, multiple and occluded objects etc. The other major difficulty, described as semantic-gap in the literature, is a gap between inferred understanding / semantics by pixel domain processing using low level cues and human perceptions of visual cues of given image. In other words, there exists a gap between mapping of extracted features and human perceived semantics.

The dimensionality of the difficulty becomes adverse because of subjectivity in the visually perceived semantics, making image content description a subjective phenomenon of human perception, characterized by human psychology, emotions, and imaginations. The image retrieval system comprises of multiple inter-dependent tasks performed by various phases. Inter-tuning of all these phases of the retrieval system is inevitable for over all good results. The diversity in the images and semantic-gap generally enforce parameter tuning & threshold-value specification suiting to the requirements. For development of a real time CBIR system, feature processing time and query response time should be optimized. A better performance can be achieved if feature-dimensionality and space complexity of the algorithms are optimized. Specific issues, pertaining to application domains are to be addressed for meeting application-specific requirements. Choice of techniques, parameters and threshold-values are many a times application domain specific e.g. a set of techniques and parameters producing good results on an image database of natural images may not produce equally good results for medical or microbiological images.

II. Proposed Algorithm

a. Select an Input Image
b. Form An Histogram Using DWT Transform.
c. Compare RGB Color Histogram & Calculate Their deviation Value(Ecludian Distance).
d. Arrange results in Ascending order
e. Display Result

A. *Image DWT Preprocessing algorithm*

Pre-processing Algorithm consists of two steps:

1) *Feature Extraction Method*

- Color Feature

The color feature is one of the most widely used visual features in image retrieval. Images characterized by color features have many advantages such as Robustness, Effectiveness, Implementation simplicity, Computational simplicity,Low storage requirements [6]. The color feature has widely been used in CBIR systems, because of its easy and fast computation. Typically, the color of an image is represented through some color model. There exist various color model to describe color information.A color model is specified in terms of 3-D coordinate system and a subspace within that system where each color is represented by a single point. The more commonly used color models are RGB (red, green, blue), HSV (hue, saturation,value) and Y,Cb,Cr luminance and chrominance). Thus the color content is characterized by 3-channels from some color model. One representation of color content of the image is by using color histogram. Statistically, it denotes the joint probability of the intensities of the three color channels.Color is perceived by humans as a combination of three color stimuli: Red, Green, and Blue which forms a color space.

This model has both a physiological foundation and a hardware related one. RGB colors are called primary colors and are additive. By varying their combinations, other colors can be obtained. The representation of the HSV space is derived from the RGB space cube, with the main diagonal of the RGB model, as the vertical axis in HSV. As saturation varies form 0.0 to 1.0, the colors vary from unsaturated (gray) to saturated (no white component). Hue ranges from 0 to 360 degrees,with variation beginning with red, going through yellow,green, cyan, blue and magenta and back to red. These color spaces are intuitively corresponding to the RGB model from which they can be derived through linear or non-linear transformations [6].Color descriptors of images can be global or local and consist of a number of histogram descriptors and color descriptors represented by color moments, color coherence vectors or color correlograms [6].Color is also an intuitive feature and plays an important role in image matching. The extraction of color features from digital images depends on an understanding of the theory of color and the representation of color in digital images. The color histogram is one of the most commonly used color feature representation in image retrieval [3].

- *Color Space Selection and Color Quantization*

The color of an image is represented, through any of the popular color spaces like RGB, XYZ,YIQ, L*a*b*, U*V*W*, YUV and HSV. It has been reported that the HSV color space gives the best color histogram feature, among the different color spaces. In HSV color space the color is presented in terms of three components: Hue (H), Saturation (S) and Value (V) and the HSV color space is based on cylinder coordinates. Color quantization is a process that optimizes the use of distinct colors in an image without affecting the visual properties of an image. For a true color image, the distinct number of colors is up to 2^24 = 16777216 and the direct extraction of color feature from the true color will lead to a large computation. In order to reduce the computation, the color quantization can be used to represent the image, without a significant reduction in image quality, thereby reducing the storage space and enhancing the process speed. The effect of color quantization on the performance of image retrieval has been reported by many authors [3].

- Color Histogram

Color histograms are used to represent image color information in various CBIR systems. Color histogram describes the distribution of colors within a whole or within a interest region of image. The histogram is invariant to rotation,translation and scaling of an object but the histogram does not contain semantic information, and two images with similar color histograms can possess different contents [6]. A color histogram is a type of bar graph, in which each bar represents a particular color of the color space being used.The bars in a color histogram are referred to as bins and they represent the x-axis. The number of bins will be totally dependent on the number of colors in an image. Color histogram y-axis denotes the numbers of pixels of each bin. There are two types of color histograms, Global color histograms (GCHs) and Local color histograms (LCHs). A GCH represents one whole image with a single color histogram while the LCH divides an image into fixed blocks and takes the color histogram of each of those blocks [5]. The color histogram serves as an effective representation of the color content of an image if the color pattern is unique compared with the rest of the data set.The color histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image. In addition,it is robust to translation and rotation about the view axis and changes only slowly with the scale, occlusion and viewing angle. Since any pixel in the image can be described by three components in a certain color space (for instance,red, green, and blue components in RGB space, or hue, saturation,and value in HSV space), a histogram, i.e., the distribution of the number of pixels for each quantized bin, can be defined for each component. Clearly, the more bins a color histogram contains, the more discrimination power it has. However, a histogram with a large number of bins will not only increase the computational cost, but will also be inappropriatefor building efficient indexes for image databases[7].A color histogram represents the distribution of colorsin an image, through a set of bins, where each histogram bincorresponds to a color in the quantized color space. A colorhistogram for a given image is represented by a vector:

$$H = \{H[0],H[1],H[2],\ldots\ldots\ldots\ldots H[i],\ldots\ldots H[n]\}$$

Where i is the color bin in the color histogram and H[i]represents the number of pixels of color I in the image, and n is the total number of bins used in color histogram. Typically, each pixel in an image will be assigned to a bin of a color histogram.

Accordingly in the color histogram of an image, the value of each bin gives the number of pixels that has the same corresponding color. In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H' is given as:

$$H'=\{H'[0],H'[1],H'[2],\ldots\ldots\ldots\ldots H'[I],\ldots\ldots H'[n]\}$$

Where , p is the total number of pixels of an image[3].

2) Pre-processing Algorithm Steps

In Pre-processing algorithm, we use Color Histogram for color feature algorithm to obtain group of cluster of feature vectors. The steps to be followed are as follows:

Step1. Extract the Red, Green, and Blue Components from an image.

Step2. Decompose each Red, Green, Blue Component usingWavelet transformation at 1st level to get approximate coefficient and vertical, horizontal and diagonal detail coefficients.

Step3. Combine approximate coefficient of Red, Green, and Blue Component.

Step4. Similarly combine the horizontal and vertical coefficients of Red, Green, and Blue Component.

Step5. Assign the weights 0.003 to approximate coefficients,0.001 to horizontal and 0.001 to vertical coefficients (experimentally observed values).

Step6. Convert the approximate, horizontal and vertical coefficients into HSV plane.

Step7. Color quantization is carried out using color histogram by assigning 8 level each to hue, saturation and value to give a quantized HSV space with 8x8x8=512 histogram bins.

Step8. The normalized histogram is obtained by dividing with the total number of pixels.

Step9. Repeat step1 to step8 on an image in the database.

In order to compare histograms of two images, we first need to generate specific codes for all histogram bins. In this experiment, (r: 0-255, g: 0-255, b: 0-255) codes were generated for RGB histogram bins. When the images have been quantized into histograms, a method of comparing these is needed. Two Histogram are compared with an Equation

$$\text{Difference}=\sqrt{\sum_{i=1}^{N}(H-H1)^2}$$

Where ,

N=no of Pixels in Image

H=Histogram of Query Image

H1=Histogram of Directory Image.

The color histogram for an image is constructed by counting the number of pixels of each color. In project, algorithms follow

- Selection of a color space
- Quantization of the color space
- Computation of histograms.

The approach more frequently adopted for CBIR systems is based on the conventional color histogram (CCH), which contains occurrences of each color obtained counting all image pixels having that color. Each pixel is associated to a specific histogram only on the basis of its own color, and color similarity across different color dissimilarity in the same is not taken into account. Since any pixel in the image can be described by three components in a certain color space histogram, i.e., the distribution of the number of pixels for each quantized color, can be defined for each component. By default the maximum number of colors one can obtain using the histogram function is 256. The conventional color histogram (CCH) of an image indicates the frequency of occurrence of every color in an image. The appealing aspect of the CCH is its simplicity and ease of computation.

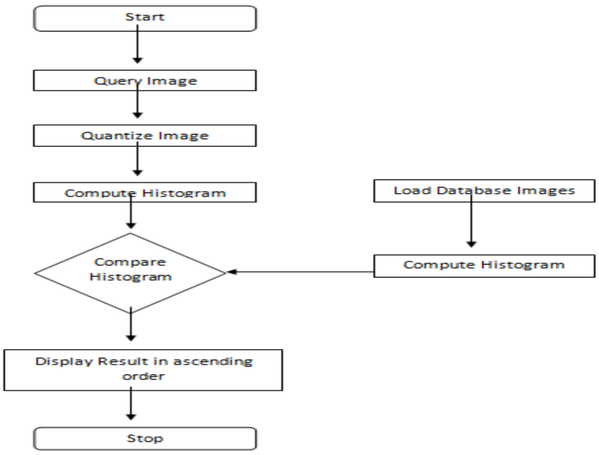

Fig. 2 DFD of Proposed Algorithm

The images data used in the experiment were taken from digital camera & few of the images were downloaded from a web site to create large database.

III. EXPERIMENT RESULTS

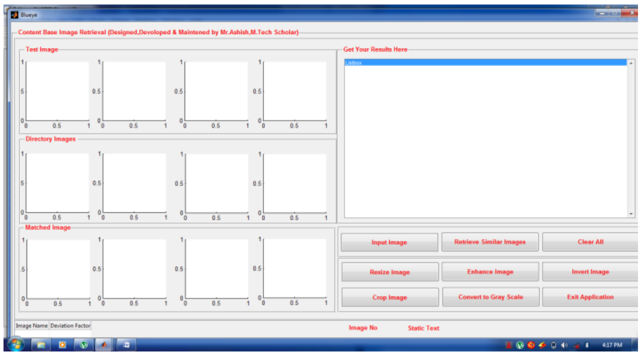

Fig. Main GUI

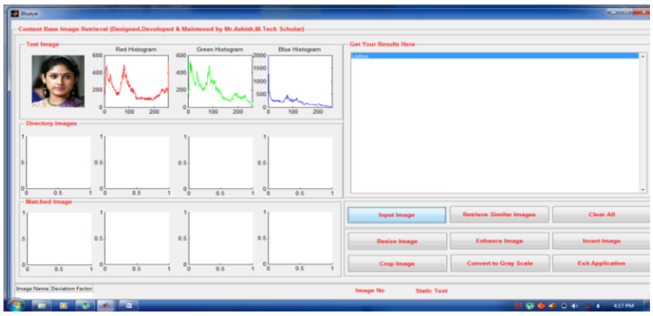

Fig. Load Query Image

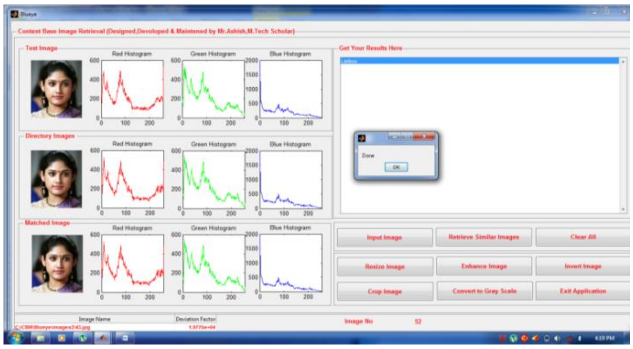

Fig. Query Image Retrieval Result

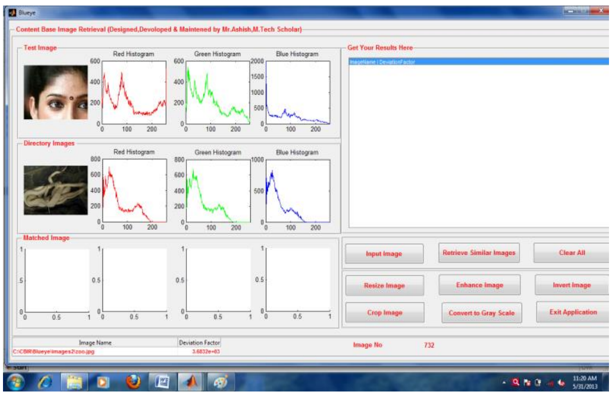

Fig. Query Image Retrieval Result after Crop

 ISSN: 2321-8134

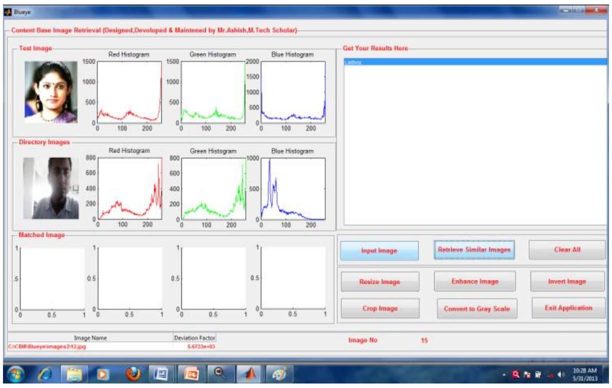

Fig. Query Image Retrieval Result after Enhancement

Table. Result Analysis with respect to parameters (After Resize the image)

Test	1	2	3	4
Image Type	jpg	jpg	jpg	Jpg
Size	471 X 600	150 X 100	75 X 50	15 X 10
Rotate	No	No	No	No
Image Format	24 Bits RGB	32 Bits RGB	32 Bits RGB	32 Bits RGB
Mining Turn Around time	00:00:14:65	00:00:04:01	00:00:03:02	00:00:03:07
Overall Image Retrieval Time	00:00:53:682	00:00:06:58	00:00:03:95	00:00:03:14
Deviation Factor	0	9.26	11.49	48.52
Overall Mining Precision	100%	100%	100%	100%

Table. Result Analysis with respect to parameters (After Crop the image)

Test	1	2
Image Type	Jpg	jpg
Crop	Yes	Yes
Size	167 X 77	142 x 84
Image Format	32 Bits RGB	32 Bits RGB
Mining Turn Around time	00:00:03:0484	00:00:03:054
Overall Image Retrieval Time	00:00:05:889	00:00:04:67
Deviation Factor	0	0
Overall Mining Precision	100%	100%

Table. Comparison of Our Method with Shrinivasa Kumar Result

Test	Total Number of Relevant Images	Number of Relevant Images Retrieved	Total Number of Retrieved Images	Recall	Precision	Result By Srinivasa Kumar [9]	
						Recall	Precision
1	46	42	45	**91.30**	**93.33**	63.04	64.44
2	47	42	46	**89.36**	**91.34**	78.94	68.00
3	45	38	42	**84.44**	**90.37**	66.66	80.00
4	42	35	41	**83.33**	**85.36**	83.33	69.44
5	40	25	35	**59.52**	**71.42**	71.80	65.70

IV. CONCLUSION

The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The images are preprocessed with various techniques and the texture calculation is highly focused. Here, images are clustered based on RGB Components .Histogram is used to compare the images with some threshold constraints. This application can be used in future to classify the medical images in order to diagnose the right disease verified earlier.

V. FUTURE SCOPE

This system is useful in future to detect the diseases related with human. More effort to be taken to reduce the Image retrieval time of an given input Query Image.

In future this system is also implemented in the field of computer Vision which is concerned with the automated processing of images from the real world to extract and interpret information on a real time basis.

In future these system is used in Astronomy to the study of celestial objects (such as stars, planets, comets, nebulae, star clusters and galaxies).

REFERENCES

[1] Ji Zhang Wynne Hsu and Mong Li Lee, “Image Mining: Trends and Developments”.
[2] A.Kannan,Dr.V.Mohannd N.Anbazhagan,”Image clustering and Retrieval using Image Mining Techniques”, IEEE International Conference on Computational Intelligence and Computing Research, 2010.
[3] D.S. Zhang and G.Lu, “Content Based Image Retrieval Using Texture Features”, In Proc. Of First IEEE Pacific-Rim Conference on Multimedia PCM’2000, pp.392-395.
[4] Yixin Chen, James Z.Wang and Robert Krovetz “Cluster Based Retrieval of Images by Unsupervised Learning”. IEEE Transaction on Image Processing, August 2005, Vol 14, pp.1187-1199.
[5] Hewayda M.Lofty, Adel S.Elmaghraby, “CoIRS: Cluster Oriented Image Retrieval System”, IEEE conference on tools with Artificial Intelligence, 2004.
[6] M. Datcu and K. Seidel,” Image Information Mining: Exploration of Image Content in Large Archives”, IEEE Conference on Aerospace, 2000, Vol.3.
[7] Zhang Y.J,”A Survey on evaluation methods for image segmentation,” vol.29, no.8, 2000, pp.1335-1346.
[8] Ji Zhang Wynne Hsu Mong Li Lee,” Image Mining: Issues, framework and techniques”.
[9] Srinivasa Kumar Reddy “Content Based Image Retrieval System”, Georgian Electronic Scientific Journal No.5, 2009.

[10] A. Kitamoto "Data Mining for Typhoon Image Collection", In Proceedings of the Second International Workshop on Multimedia Data Mining, August, 2001, San Francisco, CA, USA.

[11] M. Sasikalal and N. Kumaravel "Comparison of Feature Selection Techniques for Detection of Malignant Tumor in Brain Images", IEEE Conference, 11 -13 Dec 2005. Chennai, India.

[12] C.Ordonez and E.Omiecinski,"Discovering Association Rules Based on Image Content" Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99).

[13] C.Ordonez and E.Omiecinski,"Discovering Association Rules Based on Image Content" Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99).

[14] U. M. Fayyad, S. G. Djorgovski, and N.Weir," Automating the Analysis and cataloging of Sky Surveys" Advances in Knowledge Discovery and Data Mining, 1996, 471-493.

[15] O. R. Zaiane and J. W. Han," Mining Multimedia Data", Meeting of Minds, November 1998, pp 83-96, Toronto, Canada

[16] Arnold W.M. Smeulders, Senior Member, IEEE, Marcel Worring, Simone Santini, Member, IEEE, Amarnath Gupta, Member, IEEE, and Ramesh Jain, Fellow, IEEE(2000), "Content-Based Image Retrieval at the End of the Early Years", IEEE Transactions on pattern analysisand machine intelligence, VOL. 22, NO. 12, December 2000

[17] Hiremath P. S, Jagadeesh Pujari, "Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image."

[18] Shriram K V and P.L.K.Priyadarsini (2012), "CBIR – An analysis and suggestions for improvement", International Journal of Computer Applications (0975 – 8887) Volume 42– No.14, March 2012

[19] B.V. Ramana Reddy, A. Suresh, M. Radhika Mani3, and V.Vijaya Kumar(2009), "Classification of Textures Based on Features Extracted from Preprocessing Images on Random Windows", International Journal of Advanced Science and Technology Volume 9, August, 2009

[20] Noah Keen ,"Color Moments" February 10, 2005

[21] Greg Pass Ramin Zabih_ Justin Miller, "Comparing Images Using Color Coherence Vectors" Cornell University Ithaca, NY 14853

[22] P. S. Hiremath and Jagadeesh Pujari , " Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement"

[23] Vaibhav Gupta ,(2012), "Evolution of CBIR approaches for differently sized images" International Journal on Computer Science and Engineering (IJCSE), Vol. 4 No. 01 January 2012

[24] Pradnya Rane ,Pallavi Kulkarni, Suchita Patil and B.B.Meshram, "Feature based image retrieval of images for CBIR", IJCEM International Journal of Computational Engineering & Management, Vol. 14, October 2011

[25] Swati V. Sakhare & Vrushali G. Nasre, "Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture", International Journal of Computer Science & Informatics, Volume-I, Issue-II, 2011

[26] A laa Al-Hamami and Hisham Al-Rashdan(2010), "Improving the Effectiveness of the color coherence vector",The International Arab Journal of Information Technology, Vol. 7, No. 3, July 2010