Volume 2, No. 4, July-August 2011
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Image Mining Issues and Methods Related to Image Retrieval System

D.N.D.Harini* and Dr.D.Lalitha Bhaskari
Research Scholar*, Associate Professor,
Dept. of CS&SE,
AUCE (A), Andhra University,
Visakhapatnam, AP, India
*hariniauphd@gmail.com
lalithabhaskari@yahoo.co.in

***Abstract*:** In this digital era with sufficiently high bandwidth and storage technologies, a huge volume of multimedia database is available over the web. With significantly large and increasing multimedia database often the users have to mine the available data to retrieve the relevant information. ImageRetrieval, which is an important phase in image mining, is one technique which helps the users in retrieving the data from the available database. This paper provides an overview of Image mining, its current research, and provides a brief study of the different phases in image mining and image retrieval. The approach is broad and focuses mainly on providing the different methods used in different phases of mining.

Keywords: Image Retrieval, Image mining, Feature extraction, Indexing, Object Recognition, Classification, Clustering.

I. INTRODUCTION

In this digital era, multimedia data plays a very vital role in every field such as e-commerce, entertainment, education, medicine, aerospace and so on. With the increasing use of internet, there is an enormous volume of multimedia data available to the users. Even though this huge availability of multimedia data like images, audio, video is useful and appreciable, sometimes it proves to be a bane as there are certain difficulties to gather the required useful data in an appropriate way. A huge volume of digital images are generated every day and if analysed properly there is a lot of useful information available to the users. With the drastic increase in the multimedia databases, the usefulness of such information is dependent on how well it can be accessed; searched and how well knowledge can be extracted from it. Due to lack of proper extraction methodologies, the users are unable to extract the relevant information from the available databases.

There has been a lot of research in text based retrievals, but image retrieval is also gaining its momentum with regard to both still and moving images. Efficient image database retrieval can be done only if we have a system that is able to automatically extract relevant features directly from the images stored in the database. So image mining proves to be efficient as it deals with complex operations like image retrieval, indexing and storing.There is a common misconception that image mining is an extension to data mining which is not so. Image Mining is an upcoming research field [4, 5] and is still is in its infancyas extracting relevant knowledge from image data still remains a difficult task. The next few sections in this paper deal with a brief introduction to image mining and a broad explanation of the various phases involved in an image retrieval system.

II. IMAGE MINING

Image Mining is focused on extracting patterns, implicit knowledge, image data relationship or patterns which are not explicitly found in the images from databases or collections of images [1]. Some of the methods used to gather knowledge are: image retrieval, data mining, image processing and artificial intelligence. These methods allow image mining to have two different approaches. First, is to extract only from databases or collections of images, and second, dig or mine a combination of associated alphanumeric data and collections of images [3].

Due to the richness of information present in an image, Image mining [7] comprises of various techniques in extracting the vital information from an image. The important activities in image mining aresearching and retrieval of images based onthe features and similarity of a given input query image from the image database.

There are many different approaches related to image mining based upon the applications [11].The various image mining tools available and detailed experimental survey is conducted about the various tools like iARM, CAViz, Web image-gathering task and the SVM classifier, B2S, DisIClass, MetaSEEk, PLSA, fully automated age estimation engine[11],QBIC, Photobook, SWIM, Virage, Visualseek, Netra, MARSand so on [53]. Image mining is proving to be an important research topic as it is having its applications in areas of Medical Imaging, Weather Forecasting, Management of Earth's Resources, Forest Fires and Criminal Investigation [54].

III. IMAGE RETRIEVAL SYSTEM

With the availability of voluminous multimedia data (images, audio and video) over the web, it is becoming a challenging task to retrieve the required information. Image

retrieving is the process of getting the required content based on the user query from the available multimedia database.An Image Retrieval System(IRS) can be categorised based on the type of searches as(i)Description of an image and (ii) Its visual content.

In the present scenario, the conventional user defined textsearches are basedon keyword, size, type, date and time of capture, identity of the owner etc. This search based techniques are successful but do not meet the user's final requirement in all cases. So, many researchers are concentrating on the search based on visual content i.e., finding the images similar to an input query image.

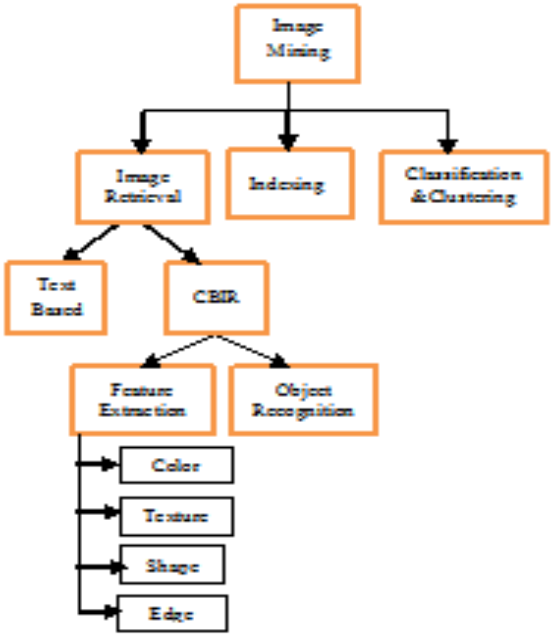

Figure 1: Phases involved in Image Mining System.

Similarity based Image Retrieval process is called Content Based Image retrieval (CBIR) wherein the images are searched and retrieved based on the visual content of the images.

The different steps in involved in Image retrieval system (IRS) are shown in figure 1.

IV. FEATURE EXTRACTION

There are no theoretical guidelines that suggest the appropriate features to use in extraction of knowledge data.Features are the measurements or properties used to classify the objects in an image. Feature extraction is a method of determining various attributes as well as properties associated with a region or objects for indexing and retrievalmainly on abstracted image information obtained through segmentation [1]. Usually the features considered with respect to images are color, shape, texture, edge.

A. General Criteria for Choosing the Appropriate Features[1]:

a. The features should be independent and should posses' maximum information about the image.
b. The extracted features should be easily computableeven for a large image collection and information retrieval should be very fast.
c. They should relate well with the human perceptual characteristics since users will finallydetermine the suitability of the retrieved images.

Because of perception subjectivity, there does not exist a single best representation for a feature.

B. Feature Extraction Techniques

The basic feature extraction technique deals with extraction of features by segmenting the foreground from the background. This is done by dividing the bigger image into smaller windows so that the features are easily extracted. There are several algorithms and many researches have followed their own methods. Some of the algorithms are present in[2, 37] and template matching is also one method used in digital image processing for finding small parts of an image which match a given template image. This technique is often used when two images of characters need to be compared[2].The coming sections deals with a brief explanation of the various features used in feature extraction.

a. Color as Feature:

Colorisone of the most widely used features in image retrieval.Retrieving images based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values.Current research in this area attempts to segment color proportion by region and by spatial relationship among several color regions as in [48].

Color Histogramis the most used in color feature representation and several approaches are used to calculate the color percentages. Some of the few techniques which are used under histogram based color feature extraction are HistogramDistances Measures[37], Color coherence vector [38], Color moment [38], Stastical Measures[41], Co-occurrence Histogram Computation[40], Histogram intersection [44],Average Color [14], Dominant color method[17],The Conventional Color Histogram Models[18], The Fuzzy Color Histogram[39], The Color Correlegram[18], Color Similarity Measure [39], Color Averaging Techniques [42], Color Discretization [43], Color Indices [43], Color Spatial Indices [43], Histogram Intersection [46] and Histogram Backprojection [46].

b. Texture as a Feature:

Textureis a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modelling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness,directionality, smoothness, granularity, repetitiveness,roughness, randomness and so on may be estimated.

Some of the important texture features that have been developed by the researchers are Markov Random Fields parameters, Gabor multi-channel features, fractal-based features and co-occurrence features, wavelet transform representations, global CW & CG approach [49, 50, 41], Multi Texton Histogram (MTH) and The Texton Co-occurrence Matrix (TCM) [51], Gabor wavelet features fortexture analysis is provided along with a comprehensive experimental evaluation in [47], Quasi-Gabor filter [45], the statistical scheme [26], Texture feature extraction algorithm for arbitrary-shaped regions [25], multiresolution multigrid

framework [40], Motif co-occurrence matrix (MCM), difference between pixels of scan pattern (DBPSP) [24],texture features in compressed domain [20] etc.

To summarize texture features can be measured based on

– Statistical measures:
- Energy
- Entropy
- Homogeneity
- Contrast
- Inertia
- Correlation
- DifferenceMoment

– Wavelets
– Fractals

c. Shape as a Feature:

Queries for shapes [9] are generally achieved by selecting an example image provided by the system or by having the user sketch a shape. The primary mechanisms used for shape retrieval include identification of features such as lines, boundaries, aspect ratio, circularityand by identifying areas of change or stability via region growing and edge detection methods. While dealing shape as a feature one of the problem is dealing with images having overlapping areas or touching shapes. Some of the approaches used are:

i. Global Features: (Moment Invariant, Aspect Ratio &Circularity)
ii. Local Features: Boundary segments

Some of the attributes considered for thecharacterizationof shape are Perimeter, Centroid, Norm Features, Hole-based and so on [52].

d. Edge as a Feature:

Edges correspond to discontinuities in a homogeneous region and they provide important visual clues that help during theobject recognition process. Edge detection is one of the most widely and a prominently used feature in the process of extraction. The output of the edge detection is usually an edge image or edge map, in which the value of each pixel reflects whether the pixel is an edge pixel or not. Edges play an effective role in image segmentation as they can be calculated using gradients. Most edge detection methods work on the assumption that an edge occurs where there is a discontinuity in the intensity (or depth) function or a very steep intensity (or depth) gradient in the image.

Some of the edge detection methods can be categorised as shown in figure 2.

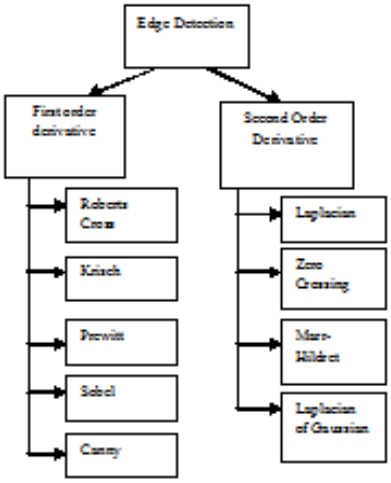

Figure2: Edge Detection Methods

Although shape, color, edge and texture are undoubtedly important visual features for image representation, there is still a little understanding of how best to implement these attributes for image retrieval. An understanding of what constitutes similarity for image retrieval purposes is also needed. The technology for content-based image retrieval is still an on-going research area which is attractingmanyresearchers. Object Recognition is discussed in the next section.

V. OBJECT RECOGNITION

Apart from feature extraction, another important phase during image retrieval is Object recognition which is basically an attempt to mimic the human capability to distinguish different objects in an image[10].In general, an object recognition model consists of four components, such as 1) model database, 2) feature detector, 3) hypothesizer, and 4) hypothesis verifier. The model database contains all the models known to the system that became trained. The models also contain several features or labels that describe the objects in the image being tested. The detected image primitive features in the gray scale (pixel) level help the hypothesizer to assign likelihood to the objects in the image. Finally, the verifier uses the models to verify the hypothesis, refine the object likelihood, and label the features in the tested image [3, 5].

Object recognition[15] in images is an important problem in computer vision with applications in image annotation, surveillance and image retrieval. Supervised or unsupervised object recognition algorithms have been developed recently which can be used for semantic-based image retrieval. Some of the few approaches given by researchers are Unsupervised scale-invariant learning method[15], semi-supervised version of the EM algorithm [21], a two-phrase generative/discriminative learning method [22],an incremental Bayesian algorithm[23] etc.

Applications of object recognition in different fields are medicine, forensics, remote sensing, and industrial applications.

VI. INDEXING

As the size of image database is increasing rapidly, retrieval speed will be an important factor to be concerned. Image indexing provides an efficient support for the

retrieval of information.Indexing can be done based on traditional and multi-dimensional approaches. A lot of research has been done in indexing and a few are mentioned here.

As the dimensionality of image features are usually high (up to tens or hundreds), traditional indexing algorithms such as k-d-b tree [28], quad-tree [29], and R-tree [27] are not suitable for image feature space indexing due to the well-known 'curse of dimensionality' problem [30]. The performance of these indexing algorithms degrades as the dimensionality of feature space increases.It is reported that when the dimensionality is above 10, the performance is no better than a simple sequential scan [31,32]. To overcome the curse of dimensionality, high-dimensional indexing algorithms such as X-tree [33], VA-file [32], and i-Distance [34] have been introduced. However, such algorithms focus only on how to index but not what to index. That is, they are designed without considering the specific properties of image features.

Some efforts have been made in designing indexing algorithms specifically for image databases. To name, FIDS (Flexible Image Database System) proposes the bare-bones triangle inequality algorithm to index image data and to sharply reduce the number of images needed to be directly compared to a query image for a given distance measure [35].The other proposed indexing methods are RBIR system [36], ImageMap [19], Multidimensional Indexing [13], WBIIS [12]and graph mining technique [16].

Further work is still to be done in efficient high-dimensional image feature indexing for real-world image database retrieval**.**With the increasing multimedia database there is a lot to do with respect to multimedia indexing techniques. The next section deals with an overview of classification and clustering techniques.

VII. CLASSIFICATION & CLUSTERING

Another important phase in image mining is classification and clustering. This section deals with a brief explanation about the various approaches used in classification and clustering.

A. Image Classification

Image classification aims to find a description that can best describe the images in one class and to distinguish these images from all the other classes [6]. Image classification is the task of assigning objects to one of several predefined categories and is widely used in mining image information, especially spatial information from image databases. There are three processes involved in image classification [8] which are

a. Feature extraction - Here animage representation model is formed which extracts features from sample images that are already labelled and establish feature description for each image.
b. Training – In this, the samples of each class are trained and model description for each class is established.
c. Classification - Use the model to classify and index images that are not labelled.

The most commonly used classification techniques are Bayes, neural network, decision tree, support vector machine, K-nearest -neighbour-classifier, genetic algorithm, Artificial NeuralNetworks (ANN), Fuzzy measures, Fuzzy support Vector Machines (FSVM) and Genetic Algorithms with Neural Networks etc. Performance of a classifier is normally measured in terms of prediction accuracy, speed, robustness and extensibility.

B. Image Clustering

The process of grouping a set of images into classes of similar images without prior knowledge is called image clustering.Image clusteringis an unsupervised learning methodwhichgroups a given set ofunlabelledimages into meaningful clusters according to the image content. Images within a cluster have high similarity in comparison to one another but are very dissimilar to images in other clusters. The process normally comprises of 4 steps [8]:

a. Image pre-processing,feature extraction and selection.
b. Set up similarity metrics suitable for special application.
c. Image clustering
d. Labelling.

The available clustering algorithms are partitioning methods, hierarchical methods, grid-based methods, model-based methods etc.

VIII. CONCLUSION

One of the major tasks in Image mining is Image retrieval which comprises of different phases like image pre-processing, Feature extraction,Indexing, image classification/image clustering.During the past few decades, researchers are focusing towards image mining and its techniques. Image mining has its applicability in almost all the fields like medicine, remote sensing, entertainment, cyber forensics, DNA, Classification of objects etc.

This paper deals with a brief study of the various approaches dealt with different researchers in all the phases related to image retrieval and mining. It is up to the user to choose the appropriate method based on the application.

IX. REFERENCES

[1] Ji Zhang Wynne Hsu Mong Li Lee, "Image Mining: Trends and Developments", Journalof IntelligentInformation Systems, 19:1, 7-23, 2002, the Netherlands.

[2] Nor AmizamJusoh et.al, "Comparison between Techniques in Feature Extraction",

[3] B. J. Lei, Emile A. Hendriks, M.J.T. Reinders, "On Feature Extraction from Images" -Technical report on "inventory properties" for MCCWS- 1999.

[4] Simoff SJ, Djeraba C, Zaiane OR. MDM/KDD2002: "Multimedia Data Mining between Promises and problems". In: ACM SIGKKD Explorations. Vol.4 (2); 2002.

[5] Ji Zhang, Wynne Hsu and Mong Li Lee "Image Mining: Issues, Frameworks and Techniques"In: Second International Workshop on Multimedia Data Mining (MDM/KDD). San Francisco, USA; 2001.

[6] Han J, Kamber M., "Data Mining: Concepts and Techniques". Morgan Kaufmann; 2001.

[7] Aura Conc and Everest Mathias M. M. Castro,"Image mining by content",Expert Systems with Applications Volume 23, Issue 4, November 2002, Pages 377-383

[8] Hu Min and Yang Shuangyuan, "Overview of Image Mining Research", The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010.

[9] Abby A. Goodrum, "Image Information Retrieval: An Overview of Current Research", Special Issue on Information Science Research Volume 3 No 2, 2000.

[10] Ji Zhang, Wynne Hsu, Mong Li Lee, National University of Singapore, "An Information-driven Framework for Image Mining", In: Proceedings of 12th International Conference on Database and Expert Systems Applications (DEXA), Munich, Germany (2001).

[11] C. Lakshmi Devasena, T.Sumathi, Dr. M. Hemalatha, "An Experiential Survey on Image Mining Tools, Techniques and Applications" in International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397 Vol. 3 No. 3 Mar 2011.

[12] James Ze Wang, GioWiederhold, Oscar Firschein and ShaXin Wei, "Wavelet-Based Image Indexing Techniques with Partial Sketch Retrieval Capability" Proceedings of the Fourth Forum on Research and Technology Advances in Digital Libraries, 1997 1092-9959/97 Copyright c 1997 IEEE.

[13] Inés María, Gómez Muñoz and MihaiDatcu, "Image Information Mining Systems", German Aerospace Center (DLR), Germany.

[14] Patricia G. Foschi, Deepak Kolippakkam, Huan Liu and AmitMandvikar,"Feature Extraction for Image Mining".

[15] Ying Liua, DengshengZhanga, GuojunLua, and Wei-Ying Mab, "A survey of content-based image retrieval with high-level semantics", Science Direct, Pattern Recognition Volume 40, Issue 1, January 2007, Pages 262-282.

[16] GáborIváncsy, RenátaIváncsy and IstvánVajk, "Graph mining-based Image Indexing".

[17] M.Babu Rao, Dr.B.Prabhakar Rao, Dr. A. Govardhan, "Content Based Image Retrieval using Dominant Colorand Texture features", (IJCSIS) International Journal of Computer Science and Information Security,Vol. 9, No. 2, February 2011.

[18] Michele Saad, "Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval".

[19] Petrakis, E.G.M., Faloutsos, C., Lin, K.-I., "ImageMap: an image indexing method based on spatial similarity" in Knowledge and Data Engineering, IEEE Transactions on Sep/Oct 2002 Volume: 14 Issue:5 On page(s): 979 - 987 ISSN: 1041-4347.

[20] Pei-xuan Chen and Guo-can Feng; "Image Retrieval Based on Dominant Color and Texture Features in DCT Domain" in Pattern Recognition, 2009, CCPR 2009, Chinese Conference on Issue Date: 4-6 Nov. 2009, pages 1 - 5, Nanjing (s): ISBN: 978-1-4244-4199-0.

[21] Y. Li, J. Bilmes, L.G. Shapiro, "Object class recognition using images of abstract regions", International Conference on Pattern Recognition, August 2004.

[22] Y. Li, L.G. Shapiro, J. Bilmes, "A generative/discriminative learning algorithm for image classification", International Conference on Computer Vision, October 2005.

[23] L. Fei-Fei, R. Fergus, P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories", Computer Vision and Pattern Recognition, Workshop on Generative-Model Based Vision, 2004.

[24] M.Babu Rao, Dr.B.Prabhakara Rao, Dr.A.Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features" in International Journal of Computer Applications (0975 – 8887) Volume 18– No.6, March 2011.

[25] Y. Liu, X. Zhou, W.Y. Ma, "Extraction of texture features from arbitrary-shaped regions for image retrieval", International Conferenceon Multimedia and Expo (ICME04), Taipei, June 2004, pp.1891–1894.

[26] Minh N. Do, Member, IEEE, and Martin Vetterli, Fellow, IEEE, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback–Leibler Distance", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 11, NO. 2, FEBRUARY 2002.

[27] A. Guttman, "R-tree: a dynamic index structure for spatial searching", Proceedings of the ACM SIGMOD International Conference on Management of Data, Boston, MA, 1984, pp. 47–57.

[28] J. Robinson, "The k-d-b-tree: a search structure for large multidimensional dynamic indexes", Proceedings of the ACM SIGMOD International Conference on Management of Data, 1981, pp. 10–18.

[29] R. Finkel and J. Bentley, "Quad-tree: a data structure for retrieval on composite keys", Acta Inf. 4 (1974) (1), pp. 1–9. Cited By in Scopus (233)

[30] H. Yamamoto, H. Iwasa, N. Yokaya, H. Takemura, "Content-based similarity retrieval of images based on spatial color distribution", 10th International Conference on Image Analysis and Processing, Venice, Italy, September 1999, pp. 951–956.

[31] R.Weber, H.-J.Schek, S. Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces", Proceedings of the 24th VLDB Conference, New York, USA, 1998, pp. 194–205.

[32] S. Berchtold, D.A. Keim, H.-P.Kriegel,"The X-tree: an index structure for high-dimensional data", Proceedings of the 22nd VLDB Conference Mumbai (Bombay), India, 1996, pp. 28–39.

[33] C.Yu, B.C.Ooi, K.-L.Tan, H.V.Jagadish, "Indexing the distance: an efficient method to KNN processing", Proceedings of the 27th VLDB Conference, Roma, Italy, 2001, pp. 421–430.

[34] A.P. Berman, L.G. Shapiro, "Triangle-inequality-based pruning algorithms with triangle tries". Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases, January 1999.

[35] I. Bartolini, P. Ciaccia, M. Patella, "A sound algorithm for region-based image retrieval using an index", International Workshop on Database and Expert Systems Applications (DEXA), 2000, pp. 930–934.

[36] H. T. Shen, B. C. Ooi, K. L. Tan,"Giving meanings to www images", Proceedings of ACM Multimedia, 2000, pp.39–48.

[37] SangohJeong, Chee Sun Won, and Robert M. Graya, "Image retrieval using color histograms generated by Gauss mixture vector quantization", S. Jeong et al. / Computer Vision and Image Understanding 94 (2004) 44–66, 2003 Elsevier.

[38] SushmitaMitra, TinkuAcharya, "Data Mining – Multimedia, Soft Computing, Bioinformatics".

[39] VarleyA.Petrushin and Latifur Khan (Eds), "Multimedia Data Mining and Knowledge Discovery", Springer-Verlag London Limited 2007.

[40] P. S. Hiremath and JagadeeshPujari, "Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement." in International Journal of Computer Science and Security, Volume (1): Issue (4)

[41] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, T. S. Huang, "Content-Based Image Retrieval Using Wavelet-based Salient Points."

[42] Dr.H.B.Kekre, Sudeep D. Thepade, Varun K. Banura, "Augmentation of Colour Averaging Based Image Retrieval Techniques using Even part of Images and Amalgamation of feature vectors." In International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5238-5246.

[43] RaimondoSchttini, GianhuigiCiocca, Silvia Zuffi,"A Survey of Methods for Color Image Indexing And Retrieval In Image Databases".

[44] Anil K. Jain and AdityaVailaya, "Image Retrieval using Color and Shape." 1995.

[45] Mira P., Jesse J., Laurence W., "Fast Content - Based Image Retrieval Using Quasi-Gabor Filter and Reduction of Image Feature Dimension." SSIAI, 2002, 178-182.

[46] Michael J. Swain and Dana H. Ballard, "Color Indexing" in International Journal of Computer Vision, 7:1, 11-32 (1991)© 1991 Kluwer Academic Publishers, Manufactured in The Netherlands.

[47] B.S. Manjunathi and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, August 1996.

[48] M. Stricker, and M. Orengo, "Similarity of Color Images," in Proc. SPIE Storage and Retrieval for Image and Video Databases, pp. 381-392, Feb. 1995.

[49] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distribution", Pattern Recognition, 29:51-59, 1996.

[50] T. Reed and H. Wechsler, "A review of recent texture segmentation and feature extraction techniques", CVGIP, 57(3):359-373, 1993.

[51] Guang-HaiLiua, Lei Zhangb, Ying-Kun Houc, Zuo-yongLic and Jing-Yu Yangc, "Image Retrieval Based on Multi-Texton Histogram".

[52] Luciano da Fontoura Costa and Roberto MarcondesCeasar Jr., " Shape Analysis And Classification", Image Processing series, 2001 by CRC Press LLC.

[53] http://en.wikiversity.org/wiki/DigitalLibraries/ Image_Retrieval.

[54] http://www.wikid.eu/index.php/Image_Mining.

