International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 7, No. 6 (2014), pp. 41-50
http://dx.doi.org/10.14257/ijsip.2014.7.6.04

Survey on Content-based Image Retrieval and Texture Analysis with Applications

Ashwani Kr. Yadav, R. Roy, Vaishali and Archek Praveen Kumar

Asst. Prof. (ASET), Amity University Rajasthan, Jaipur
ashwaniy2@gmail.com, rroy@jpr.amity.edu, vaishaliyadav26@gmail.com, archekpraveen@gmail.com

Abstract

Content-based image retrieval is a very important area of research nowadays. Content Based mage Retrieval (CBIR) is a technique which uses visual features of image such as color, shape, texture, etc. CBIR technologies provide a method to find images in large databases by using unique descriptors from a trained image. A lots of research works had been completed in the past decade to design efficient image retrieval techniques from the image or multimedia databases. Large number of retrieval techniques has been introduced, but there is no universally accepted feature extraction and retrieval technique available. In this paper, we present a study of various content-based image retrieval systems and their behaviour, texture analysis and various feature extraction with representation.

Keywords: *CBIR, Texture, segmentation, classification, shape*

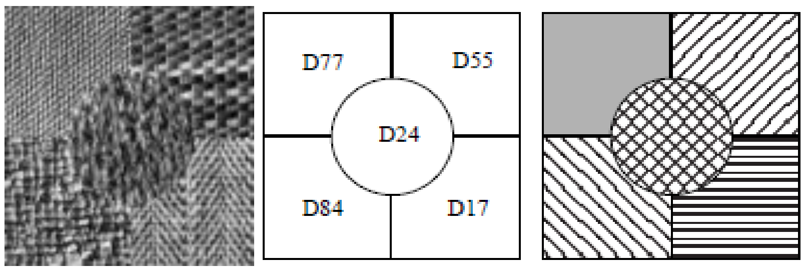
1. Introduction

The aim of CBIR is to neglect the use of textual descriptions. So in CBIR, retrieving of image based on similarities in their contents like textures, colors, shapes etc. are lower level features of image. CBIR is the application of computer vision techniques to the image retrieval problem, that is, the problem of finding of images from large databases. In Content based image retrieval the search will identifies the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The interesting images to the user are only a very small portion of the large image database, in which most images remain unlabelled. Much work regards the problem as a strict two class classification problem, with equal treatments on both positive and negative examples. It is reasonable to assume positive examples to cluster in certain way, but negative examples usually do not cluster since they can belong to any class [1].

In CBIR, retrieval of image is based on similarities in their contents, *i.e.*, textures, colors, shapes, *etc.*, which are considered the lower level features of an image. Milestone of CBIR system is low level feature extraction. Feature extraction may be done from region or an entire image [2]. These conventional approaches for image retrieval are based on the computation of the similarity between the users query and images. In CBIR each image stored in the large database and its features are extracted, compared to the features of the query image. Thus, broadly, it involves two processes, viz, feature extraction and feature matching [3]. Although, images of real things normally do not contains regions of uniform intensities. Let us take an example, image of a cloth, that is not uniform but having different intensities which form certain repeated patterns called visual texture. The patterns can be the result of cloth properties such as roughness or oriented strands, or they could be the result of

ISSN: 2005-4254 IJSIP

reflectance differences such as the color of cloth. We recognize texture when we see it but it is very difficult to define. This difficulty is demonstrated by the number of different texture definitions attempted by vision researchers. Image texture, defined as a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image Texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image. An application of image texture is the recognition of image regions using texture properties. According to Figure 1(a), we can identify the five different textures and their identities as cotton canvas, raffia, straw matting, pressed calf leather and herringbone weave. To identify homogeneous regions texture is very important. This is called texture classification.



(a) (b) (c)

Figure 1. (a) An image consisting of five different textured regions. [8] (b) The texture classification with textured region and category label. (c) The texture segmentation is to separate the regions in the image which have different textures and identify the boundaries between them.

The main objective of texture classification is to generate a classification map of the input/base image in which each uniform region has been identified with its texture class as shown in Figure 1(b). The use of image texture can be used as a description for regions into segments. Two main types of segmentation based on image texture are available, region based and boundary based. Though image texture is not a perfect measure for segmentation it is used with other measures, such as color, this will solve the segmenting problem, the main aim of texture segmentation is attain by boundary map shown in Figure 1(c).

2. Content-based Image Retrieval

Content-based image retrieval (CBIR) is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. The architecture of a CBIR system can be understood as a basic set of modules that interact within each other to retrieve the database images according to a given query. In typical content-based image retrieval system (Figure 2), the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. With the development of the Internet, and the availability of image capturing devices such as digital cameras, image scanners, the size of digital image collection is increasing rapidly. Efficient image searching, browsing and retrieval tools are required by users from various domains, including remote sensing, fashion, crime prevention, publishing, medicine, architecture, *etc.* For this purpose, many general purpose image retrieval systems have been developed. The table1 below gives some of the CBIR systems, with the features they extract and learning algorithms used to extract the features and similarity matching [4].

These models have brought this area of CBIR from its early stage to a matured stage. They study the various features of the images, make statistical analysis of color distribution as well as shape and texture and retrieve images from the contents of the image. Below we present the research works done so far starting with image retrieval then relevance feedback, existing gap between low-level and high-level features followed by relevant pattern analysis and data manipulation issues.

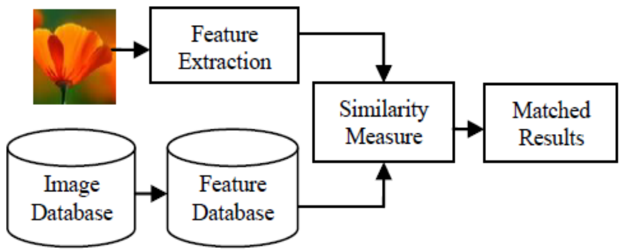


Figure 2. Content-based Image Retrieval System

3. Texture Analysis Problems

The various methods for modelling textures and extracting texture features can be applied in four broad categories of problems: texture segmentation, texture classification, texture synthesis, and shape from texture. We now review these four areas.

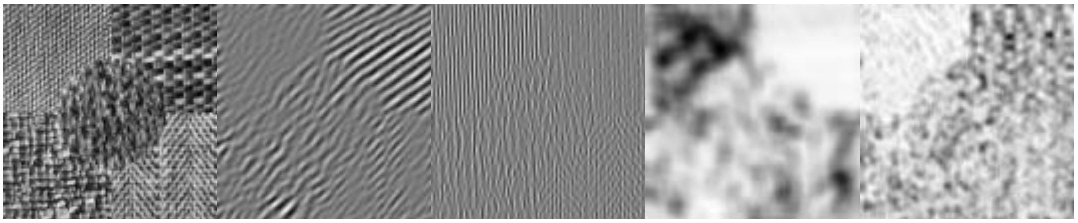
Table 1. CBIR System with Low Level Features and Learning Algorithm

CBIR System	Low level features	Learning algorithm	Similarity matching
RETIN	1. Color 2. Texture	Color histogram Output of gabor transform	Weighted minkowski distance
KIWI	1. Color 2. Shape	Color histogram Gabor filters	Euclidean space
i PURE	1. Color 2. Texture 3. Shape 4. Spatial	average color in CIE's LUV space Wold decomposition Fourier descriptor location- centroid	Euclidean space

3.1. Texture Segmentation

Texture segmentation is a difficult problem because one usually does not know a priority what types of textures exist in an image, how many different textures there are, and what regions in the image have which textures. In fact, one does not need to know which specific textures exist in the image in order to do texture segmentation. All that is needed is a way to tell that two textures (usually in adjacent regions of the images) are different.

The two general approaches to performing texture segmentation are analogous to methods for image segmentation: region-based approaches or boundary-based approaches. In a region-based approach, one tries to identify regions of the image which have a uniform texture. Pixels or small local regions are merged based on the similarity of some texture property. The regions having different textures are then considered to be segmented regions. This method has the advantage that the boundaries of regions are always closed and therefore, the regions with different textures are always well separated. It has the disadvantage, however, that in many region-based segmentation methods, one has to specify the number of distinct textures present in the image in advance. In addition, thresholds on similarity values are needed.



(a) (b) (c) (d) (e)

Figure 3. Filtering results on an example texture image. (a) Input image of five textures. (b), (c) A subset of the filtered images each filtered with a Gabor filter. (d), (e) The feature images obtained corresponding to the filtered images in (b) and (c)

The boundary-based approaches are based upon the detection of differences in texture in adjacent regions. Thus boundaries are detected where there are differences in texture. In this method, one does not need to know the number of textured regions in the image in advance. However, the boundaries may have gaps and two regions with different textures are not identified as separate closed regions. Strictly speaking, the boundary based methods result in segmentation only if all the boundaries detected form closed curves. In Figure 3 there are five different images are shown which explains the Filtering results on an example texture image. (a) Input image of five textures. (b), (c) A subset of the filtered images each filtered with a Gabor filter with the following parameters. The filter in (b) has a central frequency at 16 cycles/image-width and 135° orientation. The filter in (c) has a central frequency of 32 cycles/image-width and 0° orientation. (d), (e) The feature images obtained corresponding to the filtered images in (b) and (c). The filtered image in (b) shows a lot of activity in the textured region of the top right quadrant and the image in (c) shows activity in the textured region of the top left quadrant. These are reflected in the feature images in (d) and (e).

Boundary based segmentation of textured images have been used by Tuceryan and Jain [5], Voorhees and Poggio [6], and Eom and Kashyap [7]. In all cases, the edges (or texture boundaries) are detected by taking two adjacent windows and deciding whether the textures in the two windows belong to the same texture or to different textures. If it is decided that the two textures are different, the point is marked as a boundary pixel. Du Buf and Kardan [8] studied and compared the performance of various texture segmentation techniques and their ability to localize the boundaries. Tuceryan and Jain [5] use the texture features computed from the Voronoi polygons in order to compare the textures in the two windows. The comparison is done using a Kolmogorov- Smirnoff test. A probabilistic relaxation labeling, which enforces border smoothness, is used to remove isolated edge pixels and fill boundary gaps. Voorhees and Poggio [6] extract blobs and elongated structures from images (they suggest that these correspond to Julesz’s textons). The texture properties are based on blob characteristics such as their sizes, orientations, etc. They then decide whether the two sides of a pixel have the same texture using a statistical test called maximum frequency difference (MFD). The pixels where this statistic is sufficiently large are considered to be boundaries between different textures. Jain and Farrokhnia [9] give an example of integrating a region-based and a boundary based method to obtain a cleaner and more robust texture segmentation method. They use the texture features computed from the bank of Gabor filters to perform a region-based segmentation. This is accomplished by the following steps: (a) Gabor features are calculated from the input image, yielding several feature images. (b) A cluster analysis is performed in the Gabor feature space on a subset of randomly selected pixels in the input image (this is done in order to increase computational efficiency. About 6% of the total number of pixels in the image is selected). The number k of clusters is specified for doing the cluster analysis. This is set to a value larger than the true number of clusters and thus the image is over segmented. (c) Step (b) assigns a cluster label to the pixels (pattern) involved in cluster analysis. These labelled patterns are used as the training set and all the pixels in the image are classified into one of the k clusters. A minimum distance classifier is used. This results in a complete segmentation of the image into uniform textured regions. (d) A connected component analysis is performed to identify each segmented region.

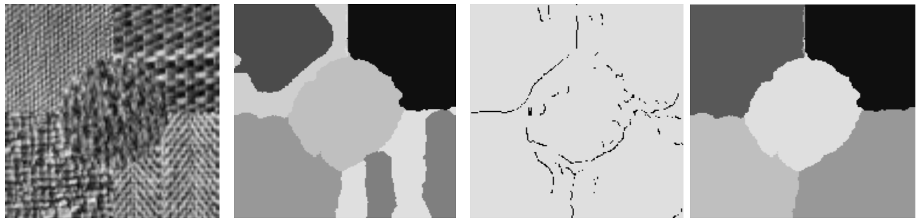
(e) A boundary-based segmentation is performed by applying the Canny edge detector on each feature image. The magnitude of the Canny edge detector for each feature image is summed up for each pixel to obtain a total edge response. The edges are then detected based on this total magnitude. (f) The edges so detected are then combined with the region-based segmentation results to obtain the final texture segmentation. The integration of the boundary-based and region-based segmentation results improve the resulting segmentation in most cases. For an example of this improvement see Fig. 4. This shows the results of integrating region-based and boundary-based processing using the multi-scale Gabor filtering method.

3.2. Texture Classification

Texture classification involves deciding what texture category an observed image belongs to. In order to accomplish this, one needs to have an a priori knowledge of the classes to be recognized. Once this knowledge is available and the texture features are extracted, classical pattern classification techniques in order to do the classification. Where texture classification has been applied as the appropriate texture processing method includes the classification of regions in images taken from satellites into categories of land use [10]. Texture classification was also used in automated paint inspection by Farrokhnia [11]. In the latter application, the categories were ratings of the quality of paints obtained from human experts. These quality rating categories were then used as the training samples for supervised classification of paint images using texture features obtained from multi-channel Gabor filters.

3.3. Texture Synthesis

Texture synthesis is a problem which is more popular in computer graphics. It is closely tied to some of the methods discussed above, so we give only a brief summary here. Many of the modelling methods are directly applicable to texture synthesis. Markov random field models discussed in Section 3.3.1 can be directly used to generate textures by specifying the parameter vector Ø and sampling from the probability distribution function [12]. The synthetic textures in Figure 2(b) are generated using a Gaussian Markov random field model and the algorithm in [13].



(a) (b) (c) (d)

Figure 4. (a) Original image consisting of five natural textures. (b) Seven category region based segmentation results. (c) Edge-based processing and texture edges detected. (d) New segmentation after combining region-based and edge-based results

4. Feature Extraction

Visual feature extraction is the basis of any content-based image retrieval technique. In a broad sense, features may include both text-based features (key words, annotations) and visual features (color, texture, shape, *etc.*). Within the visual feature scope, the features can be further classified as low-level features and high-level features. The selection of the features to represent an image is one of the keys of a CBIR system. Because of perception subjectivity and the complex composition of visual data, there does not exist a single best representation for any given visual feature. Multiple approaches have been introduced for each of these visual features and each of them characterizes the feature from a different perspective.

4.1 Color

Color is a perception that depends on the response of the human visual system to light and the interaction of light with objects. It is a product of the illuminant, surface spectral reflectance and sensor sensitivity (*i.e.*, of digital sensors or of cones in the human eye). Color is one of the most widely used visual features in content-based image retrieval. It is relatively robust to background complication and independent of image size and orientation. The key issues in color feature extraction include the color space, color quantization, and the choice of similarity function. Various studies of color perception and color spaces have been proposed [14-16]. Each pixel of the image can be represented as a point in a 3D color space. If we want to describe an image by its color features, we have to first determine the color space to use. There exist different space models such as RGB, HSV or opponent color. The best representation depends on the special needs of the application.

4.1.2. Color Space: There is a number of different color spaces currently used for the representation of images in the digital world. Choosing an appropriate color space for the implementation of a content based image retrieval system is not only important to the production of the accurate results, but to the accurate representation of color in the way that the human visual system perceives it. There are a number of color spaces in use of which some of the most commonly used are:

4.1.2.1 RGB: The most popular color space is RGB which stands for Red-Green-Blue. This space consists of the additive primary colors of light Red, Green and Blue. Varying levels of the three colors are added to produce more or less any color in the visible spectrum.

The main objective to use this feature are: (i) Filter out images with larger distance at first stage when multiple feature queries are involved.(ii) It uses a small number of data to represent the feature vector and (iii) It also uses less computation as compared to others.

Although, the accuracies of query result could be significantly impact if the feature is not combined with other features. This space is device dependant and perceptually non-uniform. This means that a color relative close together in the RGB space may not necessarily be perceived as being close by the human eye. RGB space is normally used in Cathode Ray Tube (CRT) monitors, television, scanners, and digital cameras. For a monitor the phosphor luminescence consists of additive primaries and we can simply parameterize all colors via the coefficients (α, β, γ), such that C = αR+βG+γB. The coefficients range from zero (no luminescence) to one (full phosphor output). In this parameterization the color coordinates fill a cubical volume with vertices black, the three primaries (red, green, blue), the three secondary mixes (cyan, magenta, yellow), and white as in Figure 5.

4.2. Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrices (GLCM) is a popular representation for the texture in images. They contain a count of the number of times a given feature (*e.g.*, a given gray level) occurs in a particular spatial relation to another given feature. GLCM, one of the most known texture analysis methods, estimate image properties related to second-order statistics.

We used GLCM techniques for texture description in experiments with 14 statistical features extracted from them. The process involved is follows:

1. Compute co-occurrence matrices for the images in the database and also the query image.

Four matrices will be generated for each image [17].

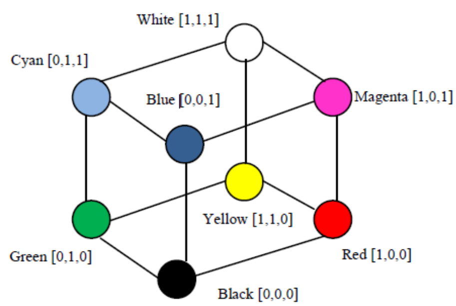


Figure 5. RGB Color Space

2. Build up a 4×4 features form the previous co-occurrence matrices as shown in Table 2

Table 2. Four Main Features used in Feature Extraction

Feature	Formulae
Energy	$\sum_i \sum_j P^2(i, j)$
Entropy	$\sum_i \sum_j P(i, j) Log\, p(i, j)$
Contrast	$\sum_i \sum_j (i - j)^2\, p(i, j)$
Homogeneity	$\sum_i \sum_j \frac{P(i, j)}{1 + (i + j)}$

4.3. Color Histogram

Color is the most widely used "feature" owing to its intuitiveness compared with other features and most importantly, it is easy to extract from the image. The color histogram depicts color distribution using a set of bins. However, a CBIR system based on color features shows distorted results sometime, because it uses global color feature, It cannot capture color distributions or textures within the image in some cases. To improve the method of the color extraction the color histogram feature can be divided into global and local color extraction.

Using Global Color Histogram (GCH), an image will be encoded with its color histogram, and the distance between two images will be determined by the distance between their color histograms. Local color histogram (LCH) can give some sort of spatial information, however the con associated with it is that it uses very large feature vectors.LCH includes information concerning the color distribution of regions. The first step is to segment the image into blocks and then to obtain a color histogram for each block. An image will then be represented by these histograms. When comparing two images, we calculate the distance, using their histograms, between a region in one image and a region in same location in the other image. The distance between the two images will be determined by the sum of all these distances.

However, it does not include information concerning the color distribution of the regions, so the distance between images sometimes cannot show the real difference between images. Moreover, in the case of a GCH, it is possible for two different images to have a very short distance between their color histograms. This is their main disadvantage.

4.4. Geometric Moments

An image moment is a certain particular weighted average (moment) of the image pixels' intensities, or a function of such moments, normally chosen to have some attractive properties. Image moments are useful to describe objects after segmentation.

Simple properties of the image which are found via image moments include area (or total intensity), its centroid, and information about its orientation. This feature use only one value for the feature vector, however, the performance of current implementation isn't well scaled, [18] which means that when the image size becomes relatively large, computation of the feature vector takes a large amount of time. The pros of using this feature combine with other features such co-occurrence, which can provide a better result to user.

4.5 Color Moments

To overcome the quantization effects of the color histogram, the color moments as feature vectors for image retrieval are used. So color distribution can be characterized by its moments and most information is concentrated on the low moments, only the first moment (mean), the second moment (variance) and the third moment (skewness) are taken as the feature vectors. With a very reasonable size of feature vector, the computation is not expensive [19].Color Moments are measures that can be differentiate images based on their feature of color, however, the basic concept behind color moments lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. The advantage is that, the degree of asymmetry in the distribution can be measure by its skew-ness.

5. Methods of Representation

Each feature may have several representations. For example, color histograms [14], color moments [20], color or color correlograms [21] coherence vectors [22], etc are representations of the image color feature. Moreover, numerous variations of the color histogram itself have been proposed, each of which differs in the selected color-quantization scheme. The color descriptors are related to mathematical operations of the pixel values represented in a certain color space. Some of the most popular descriptors are: Color histogram, Color Moments, Color Correlogram and Color coherence vector.

Texture representation methods can be classified into three categories:

- Statistical techniques characterize texture using the statistical properties of the gray levels of the pixels comprising an image. Normally, in images, there is periodic occurrence of certain gray levels. The spatial distribution of gray levels is calculated.
- Structural techniques characterize texture as being composed of texels (texture elements). These texels are arranged regularly on a surface according to some specific arrangement rules.
- Spectral techniques are based on properties of the Fourier spectrum and describe global periodicity of the grey levels of a surface by identifying high-energy peaks in the Fourier spectrum.

Statistical techniques are most important for texture classification because it is these techniques that result in computing texture properties. Some of the statistical representations of texture are tamura features, co-occurance matrices, and multi-resolution filtering techniques such as Gabor and wavelet transform.

6. Applications of CBIR

There are various possible applications for CBIR technology has been identified. Some of these are mentioned below:

- Investigations: face recognition systems, copyright on the Internet
- Shapes identification: identification of defect and fault in industrial automation.
- Medical diagnosis: Tumours detection, Improve MRI and CT scan Understand ability.
- Journalism, advertising Media, Fashion and graphic design.
- Remote sensing: Various information systems, weather forecast, satellite images.
- Trademark databases, Art galleries, museums and archaeology.
- Architectural and engineering designs.
- Cartography: map making from photographs, synthesis of weather maps.
- Digital Forensics: finger print matching for crime detection.
- Radar engineering: helps in detection and identification of targets.

7. Conclusion

CBIR is a fast developing technology with considerable potential. Research in CBIR in past has been focused on image processing, low level feature extraction etc. It has been believed that CBIR provides maximum support in bridging ‘semantic gap’ between low level feature and richness of human semantics. This paper provides comprehensive survey on feature extraction in various CBIR systems and texture analysis with various applications. Various features with their method of representation are discussed. The area of content-based image retrieval is a hybrid research area that requires knowledge of both computer vision and of database systems. The technology is exciting but immature, but few operational image archives shown interest in adoption. The field appears to be generating interesting and valid results, even though it has so far led to few commercial applications.

References

[1] T. Dharani and I. Laurence Aroquiaraj, "A Survey on Content Based Image Retrieval", Proceedings of the 2013 Int'l Conf. on Pattern Recognition, Informatics and Mobile Engineering, (**2013**) February, pp. 21-22.

[2] M. Rehman, M. Iqbal, M. Sharif and M. Raza, "Content Based Image Retrieval: Survey", DOI: 10.5829/idosi.wasj.2012.19.03.1506, World Applied Sciences, vol. 19, no. 3, (**2012**), pp. 404-412, ISSN 1818-4952; IDOSI Publications, 2012.

[3] D. A. Kumar and J. Esther, "Comparative Study on CBIR based by Color Histogram, Gabor and Wavelet Transform", Int'l Journal of Computer Applications (0975 – 8887), vol. 17, no. 3, (**2011**) March, pp. 37-44.

[4] H. Tamura, S. Mori, and Y. Yamawaki, "Textural Features Corresponding to Visual Perception", IEEE Transactions on Systems, Man, and Cybernetics, SMC-8, (**1978**), pp. 460-473.

[5] M. Tuceryan and A. K. Jain, "Texture Segmentation Using Voronoi Polygons", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12, (**1990**), pp. 211-216.

[6] H. Voorhees and T. Poggio, "Detecting textons and texture boundaries in natural images", In Proceedings of the First International Conference on Computer Vision, London, (**1987**), pp. 250-258.

[7] K.-B. Eom and R. L. Kashyap, "Texture and Intensity Edge Detection with Random Field Models", In Proceedings of the Workshop on Computer Vision, Miami Beach, FL, (**1987**), pp. 29-34.

[8] D. Buf, J. M. H. M. Kardan and M. Spann, "Texture Feature Performance for Image Segmentation", Pattern Recognition, vol. 23, (**1990**), pp. 291-309.

[9] A. K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation Using Gabor Filters", Pattern Recognition, vol. 24, (**1991**), pp. 1167-1186.

[10] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, (**1973**), pp. 610-621.

[11] F. Farrokhnia, "Multi-channel filtering techniques for texture segmentation and surface quality inspection", Ph.D. thesis, Computer Science Department, Michigan State University, (**1990**).

[12] A. Khotanzad and R. Kashyap, "Feature Selection for Texture Recognition Based on Image Synthesis," IEEE Transactions on Systems, Man, and Cybernetics, vol. 17, (**1987**), pp. 1087-1095.

[13] R. Chellappa, S. Chatterjee and R. Bagdazian, "Texture Synthesis and Compression Using Gaussian- Markov Random Field Models," IEEE Trans. on Systems, Man, and Cybernetics, SMC-15, (**1985**), pp. 298-303.

[14] M. J. Swain and D. H. Ballard, "Color indexing", International Journal of Computer Vision, vol. 7, no. 1, (**1991**), pp. 11–32.

[15] Y. Rui, T. S. Huang and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues", Journal of Visual Communication and Image Representation, vol. 10, no. 1, (**1999**), pp. 39–62.

[16] F. Long, H. J. Zhang and D. D. Feng, "Fundamentals of Content-based Image Retrieval, Multimedia Information Retrieval and Management", D. Feng Eds, Springer, (**2003**).

[17] J. W. Bala, "Combining Structural and Statistical Features in a Machine Learning Technique for Texture Classification", IEA/AIE '90 Proceedings of the 3rd international conference on Industrial and engineering applications of artificial intelligence and expert systems, vol. 1, (**1990**), pp. 175-183.

[18] Y. Liu, D. Zhangt, G. Lu and W.Y. Ma, "Study on Texture Feature Extraction in Region-Based Image Retrieval System", Multi-Media Modelling Conference Proceedings, 12th International, (**2006**), pp. 8-15.

[19] S. Deb and Y. Zhang, "An Overview of Content-Based Image Retrieval Techniques", Proc. IEEE Int. Conf. on Advanced Information Networking and Application, vol. 1, (**2004**), pp. 59-64.

[20] M. Stricker and M. Orengo, "Similarity of color images. In Proceedings of Storage and Retrieval for Image and Video Databases", (SPIE), (**1995**), pp. 381-392.

[21] J. Huang, R. Kumar, M. Mitra, W. Zhu and R. Zabih, "Image Indexing Using Color Correlograms", In Proceedings of CVPR, (**1997**), pp. 762-768.

[22] G. Pass and R. Zabith, "Histogram refinement for content-based image retrieval", IEEE Workshop on Applications of Computer Vision, (**1996**), pp. 96-102.

[23] J. H. Su, W. J. Huang, P. S. Yu and V. S. Tseng, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns", IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 3, (**2011**) March, pp. 360-372.