# Content-Based Image Search Engine

**Mohammed Abdulshakoor Ameen**
mameen@kaau.edu.sa
*Department of Computer Science*
*King Abdul Aziz University,*
*80203 Jeddah 21589, KSA*

**ABSTRACT**. Image and video indexing have become important with recent increase in digital image collections. A majority of proposed indexing techniques in the literature are based on features extracted from the entire image. In this paper, images are segmented by the K-mean clustering algorithm in order to allow searching, and retrieving at region level. Further, the regions are classified into object and non-object classes as we may retrieve regions based on features tailored for the corresponding type. The object class contains clumped regions while the non-object class contains regions that scattered in the entire image scenes as constrained by $\chi^2$ statistic. The performance of the retrieval of four region indexing techniques, Histogram and three Wavelet-based techniques are evaluated. The four indexing techniques are suitable for general domain image collections. The evaluation result has shown that, although, the three wavelet-based region indexing techniques provide a good comparable performances for non-object region queries, the histogram-based region indexing technique outperforms the three wavelet-based region indexing techniques for object region queries. The histogram technique is more suitable for indexing object regions. On the other hand the wavelet-based techniques are more suitable for non-object region indexing.

**Keywords**: Internet Search Engine, Region identification, Wavelet indexing, Histogram indexing, Content-based image retrieval (CBIR).

## 1. Introduction

Since the World Wide Web revolution in the early of 1990s, the growth rate of digital image collections has increased exponentially. Every year, gigabytes of images are generated and archived. However, in order to browse and query images, appropriate indexing and organizing must be constructed so as to allow efficient retrieval. Storage and retrieval of these images may be accomplished by keyword-based indexing, content-based indexing, or multi-indexing that integrates both techniques. The keyword-based retrieval, also known as text-based or annotation approach is to store the digital object with some descriptive text assigned by human operators. There are several major problems with text-based indexing. The most essential one is the vast amount of annotation work required that is keyed in manually by human. To overcome this problem, content-based image retrieval (CBIR) system was proposed in which images are indexed by their own visual features that are extracted automatically. CBIR have become important with recent advances in image and video compression standards such as JPEG-2000 [1] and MPEG-4 [2]. They have wide applications in several areas such as multimedia information systems and digital libraries for education, medical imaging, remote sensing, movie industry, and video on demand and military. A major obstacle to achieve acceptable performance in multimedia retrieval system is that most content-based retrieval system relay on low-level descriptors that are difficult to link to higher concepts. The only visible solution to this linking problem to date is the manual annotation by human operator. Most Content-Based Image Retrieval (CBIR) systems existing today, both commercial and under research, index their general-domain image collection by local or global visual features. The visual features including color, texture, and shape; have been extracted automatically from the images at scene level (whole image). Some CBIR also retrieve the relevant image by similarity sequential search. When a user poses an image as a query by example, a CBIR system extracts the features from the entire query image and searches for all similar images by distance measure such as Euclidian distance.

If the features extracted from images at region or object levels then more semantically meaningful indexing of the images can be achieved. Figure (1) shows an outline of such system. However, segmenting general-domain images into accurate regions and error-free condition is the most challenging image processing task, especially when no priori information is available to define the regions. The following section presents the region identification process. In section 3, four region-indexing techniques are explained. The evaluations of the retrieval performance of the indexing techniques are given in section 4.

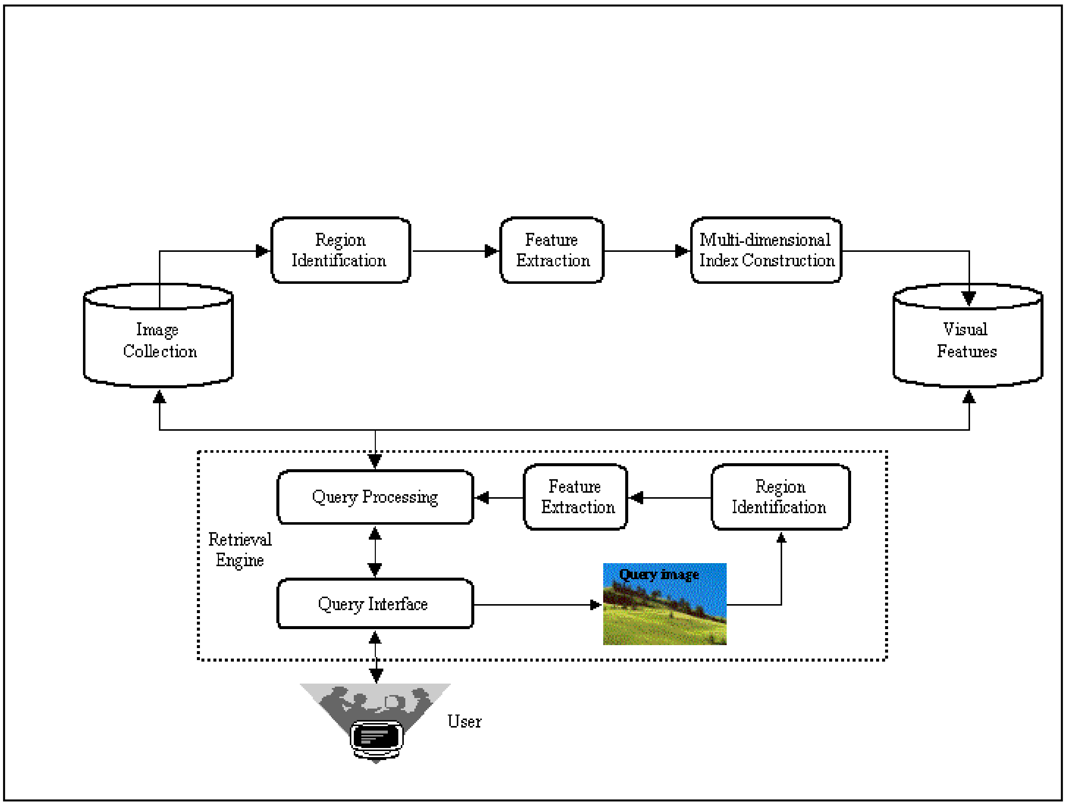



Figure (1) : Architecture of the proposed region based image retrieval system.

## 2. Region Identification

Region identification process involves image segmentation followed by region classification. Region identification process segments the image into regions such that:

- an image $I = \{R_1, R_2, \ldots, R_n)$ where $R_k$ is the k region of image I.
- $I = \cup R_i$ where $i = 1, 2, \ldots, n$.
- $R_i \cap R_j = \phi$ for all $i \neq j$.

In this research, the well known statistical clustering algorithm, K-mean,[3,4,5] is used to segment images into disjoint regions. Then regions that represent objects or meaningful parts are identified by the statistical goodness of fit test. The objective of the region identification process is to classify regions into object and non-object regions. Identifying image regions is important for indexing and searching of large-scale image databases.

In a general real-world image collection as shown in Figure 2, a scene may contain main object and background, many objects, or non-object. Texture images as shown in Figure (3) consist of two or more non-object regions.

Figure (2) : Samples of different scenes that contain more than one object.

Figure (3) : Samples of texture scenes with non-object regions

The classification of regions into object and non-object classes by human is subjective and incomplete. Meaning of the content of an image depends on the point-of-view and experience of the viewers. In this research the meaning of object is constrained by the scattering concept. Region identification automates an objective rather than subjective region classification as explained in the following subsections.

**2.1 *K-mean Image Segmentation***

This section presents image segmentation by the K-mean method a well-known statistical clustering algorithm. Many research are have used K-mean for image segmentation, Tou and Gonzalez [3] and Bezdek [4] classify the pixels into clusters based only on their intensity value. In this research, the K-mean algorithm has been applied to image segmentation using wavelet and color features. After the pixels are classified into regions, each region then will be labeled as object region or non-object region. Wang, et. al. [5] applied similar segmentation algorithm to classify the whole image as texture or non-texture. Another difference between Wang, et. al. [5] and the work in this research is that, in Wang, et. al. [5], features are extracted from 4x4 non-overlapped blocks. In this research, features are extracted from each region to reduce the distortion that results from other objects in the scene. In this way, the retrieval system will be scalable to adding new features or modifying for specific-domain.

The K-mean clustering algorithm is as follows:

Given n objects with f measures on each object, $X_{ij}$ for I=1, 2, …, n; j=1, 2, …, f.

The K-mean algorithm arranges the objects into k clusters such that within-cluster average distance:

$$\sum_{k=1}^{k} \sum_{i \in s_k} (\vec{x}_i - \vec{x}_k)^2 \qquad (1)$$

is minimized, where $s_k$ is the set of objects in cluster k and $x_k$ is the centeroid of cluster k. The algorithm starts with initial number of cluster and iteratively increased until one of the following criteria is satisfied.

1. Optimal within-cluster average distance is reached by setting a threshold.
2. The derivative of the distortion with respect to k is minimum than a threshold.
3. The number of clusters exceeds an upper bound.

**2.2 *Feature Extraction for Image Segmentation***

In this subsection the feature extraction process for segmentation purpose is explained. Since the image retrieval system developed in this research is oriented to general-purpose images such as the images in World-Wide Web (WWW), six visual features are used. The first three features are the average color intensity in YCrCb color space. The International Telecommunications Union-Radio (ITU-R) as standard, ITU-R 601, has specified this color space. This standard is based on encoding the luminance signal (Y) and two color difference signals (Cr and Cb) that encode the chrominance color information. The YCrCb color space can be obtained from RGB color space by the following transform:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cr &= -0.1687R - 0.3313G + 128 \\ Cb &= 0.5R - 0.4187G - 0.0813B + 128 \end{aligned} \qquad (2)$$

This transform is used in many known color image compression such as JPEG algorithm because it has acceptable perception of color correlation properties more adequate for the purpose of computing visual features.

The three other features are the second moments of the high frequency bands of one-level Daubechies wavelet transform of the three-color components. The wavelet features capture the spatial color variations and texture properties of visual data effectively Unser [6]. Figure (4) and (5) show segmentation results for some images with non-object (texture) and object regions.

### 2.3 *Classification of Object and Non-Object Regions*

The classification of regions into object and non-object classes is based on region scattering concept. Region scattering can be mathematically evaluated by the statistical goodness of fit test. The goodness of match between the distribution of a region, $r_i$, and a uniform distribution is measured by the $\chi^2$ statistic. To compute the $\chi_i{}^2$ statistic for a region *i*, an image is partitioned into 16 zones, { $z_1, z_2, \dots z_{16}$}, each zone has a uniform distribution $p(z_j) = \frac{1}{16}$. The percentage of each region is the probability distribution $p(r_{ij})$ where

$$\sum_{j=1}^{16} p(r_{ij}) = 1. \tag{3}$$

The $\chi^2$ statistic for a region is *i* is given by:

$$\chi_i{}^2 = \sum_{j=1}^{16} \frac{((p(r_{ij}) - p(z_j))^2}{p(z_j)} \tag{4}$$

Region classified by thresholding the $\chi^2$ statistic into object or non-object class with each class corresponding to region. Figures (6 – 8) show an examples of object and non-object regions extracted from region identification process. Figure (9) shows two histograms of $\chi^2$ statistics for object and non-object regions over 759 regions extracted from 300 scenes. It is shown that the two histograms separate significantly around the decision threshold 0.36. The classification error is 0.07%.

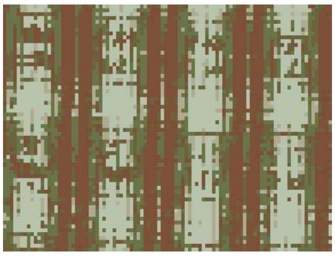
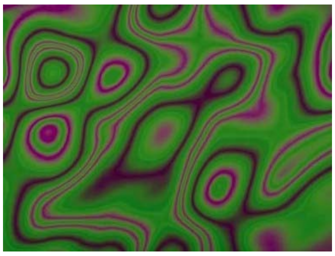
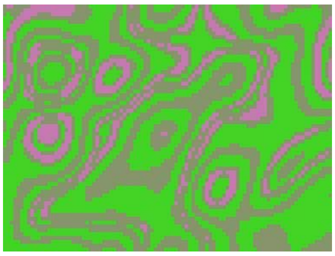
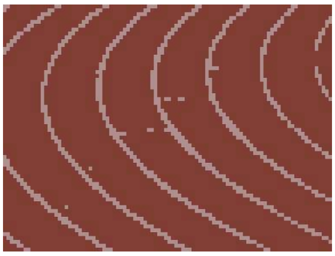
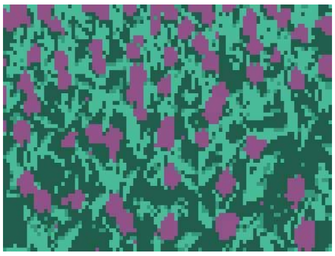
| Original Image | Segmented Image | Number of regions |
|---|---|---|
|  |  | **5** |
|  |  | **3** |
|  |  | **2** |
|  |  | **4** |
|  |  | **6** |

Figure (4) : Samples of segmentation results by the k-mean algorithm for images with no objects.

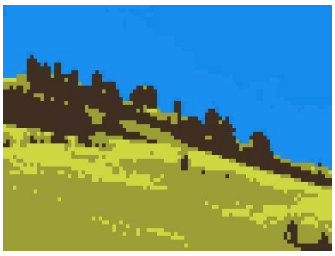
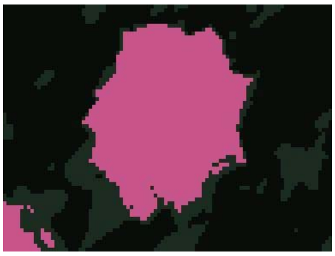
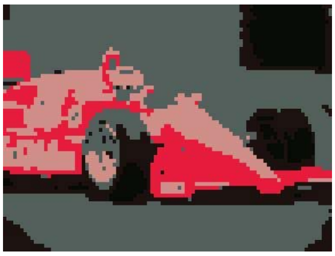
| Original Image | Segmented Image | Number of regions |
|---|---|---|
|  |  | **5** |
|  |  | **5** |
|  |  | **3** |
|  |  | **4** |
|  |  | **7** |

Figure (5) : Samples of segmentation results by the k-mean algorithm for images with objects.

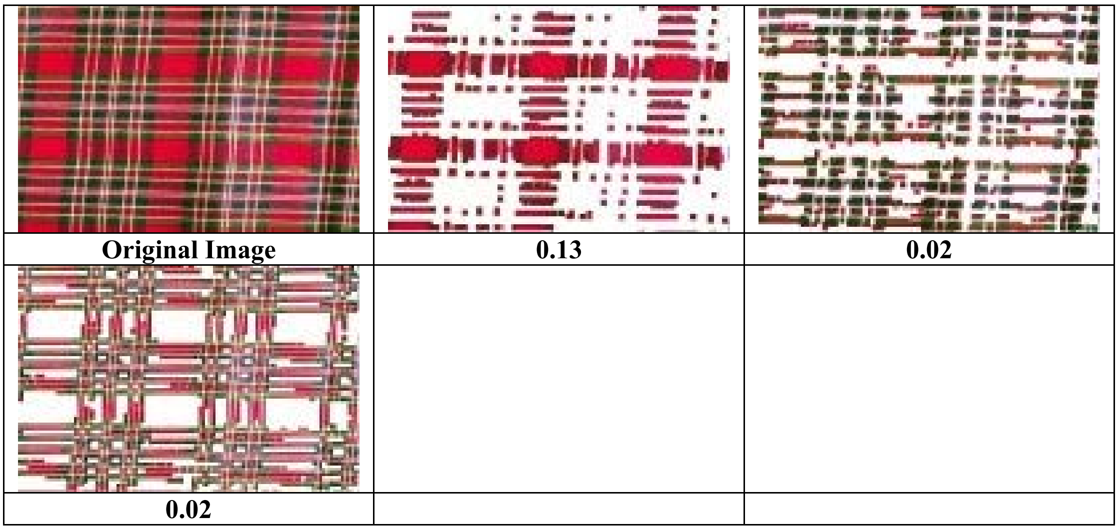



Figure (6) : Samples of non-object regions with $\chi^2$ values of each region

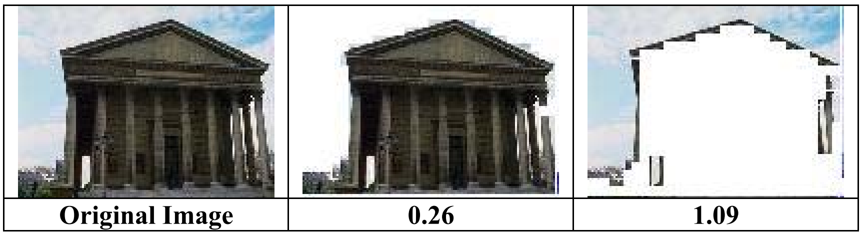



Figure (7) : Samples of object regions with $\chi^2$ values of each region

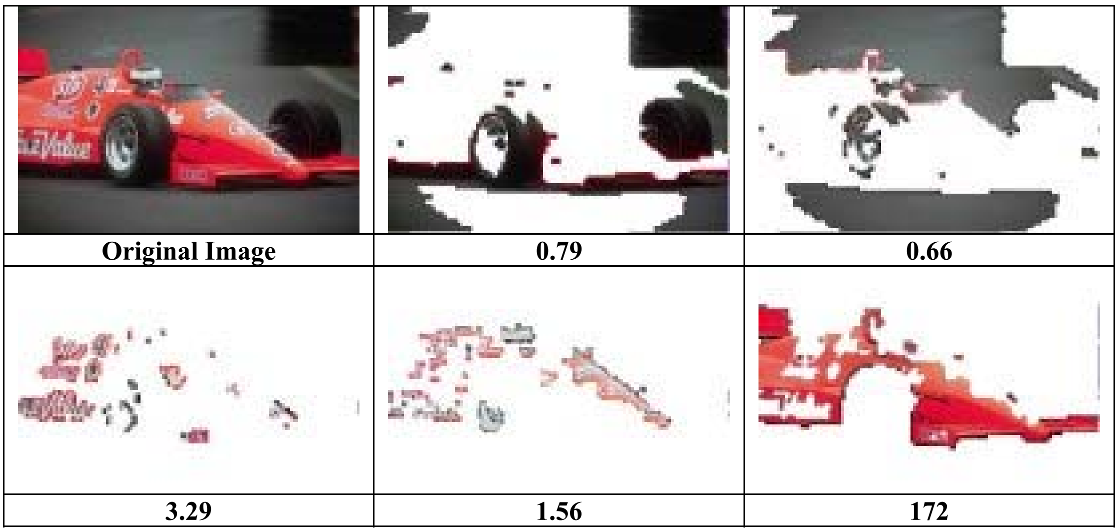



Figure (8) : Samples of object regions with $\chi^2$ values of each region

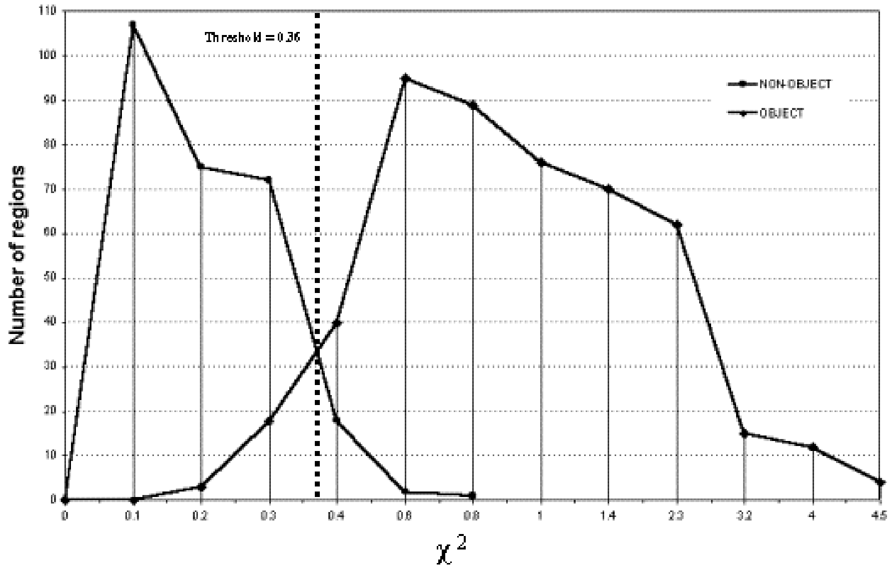



Figure (9) : the histograms of $\chi^2$ statistics over 759 regions extracted from 300 scenes.

### 3. Region Indexing and Feature Extraction

The regions that are result from the region identification process described in section 2 can be indexed by mathematical measures extracted from a window-constructed around each region by padding zeros [12]. The mathematical measures are also called features such as color, texture, and shape represent different properties. In this section, four general domain image-indexing techniques are presented. The histogram is presented in the next subsection. Subsection 3.2 presents the three other wavelet-based indexing techniques.

#### 3.1 *Histogram-Based Region Indexing*

An image can be modeled by a function of two variables f(x,y); where x and y are two co-ordinates in a plane. The values of the image function represent the brightness at image pixels. If the image is consisting of three component colors then the image function value is a three-dimensional function.

Color features can be extracted from an image or a region at two levels; global and local levels. A global feature is a feature extracted from the whole image, or a whole region including statistical moments and Histogram. Histogram denotes the probability of the intensities of image pixels. Let *I* be the gray levels in the region of interest. The histogram *h(I)* is defined as

$$h(I) = \frac{\textit{number of pixels with gray level } I}{\textit{total number of pixels in the region}} \tag{5}$$

Histogram indexing technique (HIST) is the most common visual feature that has been used by many image retrieval systems such as FINDIT system developed by M. J. Swain and his colleagues [7] and QUBIC system from IBM Corporation. The histogram computed from each region and retrieval is done by comparing the histogram of the query's regions with histograms of all regions in the image collection. A subset of images then are retrieved that have least differences of regions histograms. It has been shown that the histogram provides a good indexing and retrieval performance while being inexpensive in term of computation [8]. The histogram-based indexing is invariant to translation and rotation of regions. The histogram provides a good indexing and retrieval performance for regions of large objects but it fails to provide a good indexing for texture regions.

### 3.2 *Wavelet-Based Region Indexing*

Recently, wavelet-based indexing techniques are gaining popularity due to recent advances in image compression. Several wavelet-based indexing techniques have been proposed in the literature of CBIR. Histogram of wavelet coefficients of different bands has been proposed by Smith and Chang [9]. The discrete cosine transform of the wavelet lowest-frequency subband (WDCT) has been proposed by Abbadi [10]. The magnitude of the wavelet coefficients (WMGT) has been proposed by Wang et al. [11]. In the Wavelet Histogram Technique (WHST) proposed by Smith and Chang [9], the image is first decomposed M times to wavelet bands that form a pyramid of M levels. Each level consist of three high-pass wavelet bands. A three-level DWT decomposition is shown in Figure 10, where level-1 bands consists of {A7, A8, A9}, level-2 bands consists of {A4, A5, A6}, and level-3 bands consists of {A1, A2, A3}. The magnitudes of each band is up-sampled to the full-size image by inserting zeros and subsequently passed through appropriate filters to obtain a texture channel. The entire process for a three-level wavelet histogram generation is illustrated in Figure (11) where nine texture channels are generated from nine high-pass DWT bands. A texture point is then defined as a 9D vector by considering texture channel values from the same location of all nine bands. Thus for an MxN image, there will be MN 9D vectors. Each element of the 9D vector is thresholded to binary levels, 0 or 1. Due to the binary nature of the vector elements, 512 (=29) different vectors span the entire 9D feature space. The histogram (with 512 bins) of these vectors is employed as the index of the image.

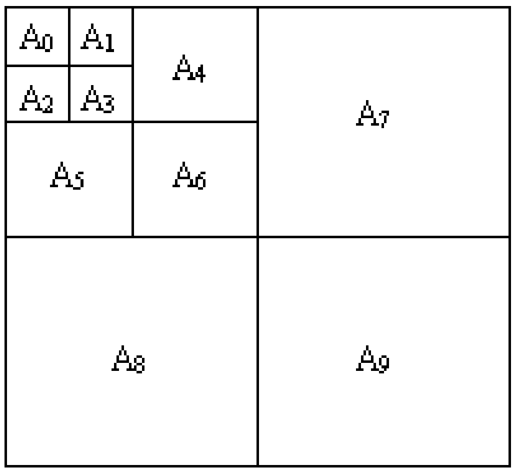



Figure (10) : Band structure of the three-level wavelet transform

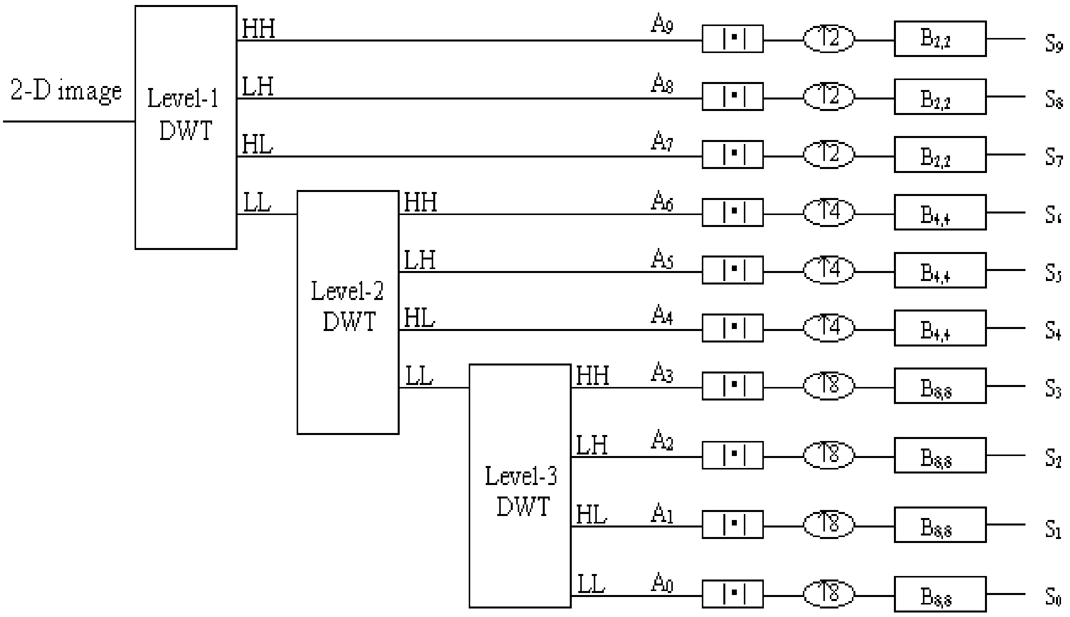



Figure (11) : Schematic of wavelet histogram generation by Smith and Chang [9]

## 4. Experimental Results

In this section, the evaluation of retrieval efficiencies of the three wavelet-based indexing techniques, WDCT, WMGT, and WHST, and the histogram indexing technique, HIST, is presented. To evaluate the retrieval performance a database of images has been created, IDB, which contains 25,000 images of various color images from data set such as Corel and from the web that have been used for testing other CBIR. IDB has a variety of images, including natural scenes, animals, humans, automobiles, flowers, texture, and buildings. Unfortunately there is no existence for test data set and queries as a benchmark for image retrieval performance evaluation, which makes it difficult to conduct standard performance evaluation. The evaluation is based on two query region sets, object and non-object regions. The standard retrieval performance measures, recall and precision, are used to evaluate the retrieval performances. Let A denote the number of retrieved images in a query, B is the number of relevant images retrieved, and C is the number of relevant images for a query then

$$\text{Pr}\,ecision = \frac{B}{A} \qquad (6)$$

$$\text{Re}\,call = \frac{B}{C} \qquad (7)$$

The query accuracy evaluated based on a subset of the IDB contains 60 images randomly selected. Five images have been driven from each image corresponding to normal, translated rightward, translated leftward, rotated clockwise, and rotated counterclockwise. As a result, a total of 300 images. The K-mean segmented the images into 455 object regions and 295 non-object regions. 50 regions have been randomly selected from each type and manually listed the 5 similar regions found in the collection. As a result, the number of relevant images, C, is fixed to be 5 for a given query. The number of retrieved images, A, depends on system design where set to 20, however, since each query region has only 5 relatives, among the 20 retrieved regions, only 5 regions can be retrieved for each query. For this reason A=C and Precision is equal to Recall. Daubechies (db2) wavelet has been used for the three wavelet-based indexing. The similarity of regions is based on $L^1$ metric. The performances of HIST, WHST, WDCT, and WMGT techniques are shown in Table 1.

Table (1) : Efficiency of four region-based indexing techniques

| Technique | Object region query | Non-object region query |
|---|---|---|
| HIST | 92 | 73 |
| WHST | 75 | 85 |
| WDCT | 80 | 77 |
| WMGT | 78 | 82 |

For object region indexing, HIST provide a good retrieval performance. It outperforms the three wavelet techniques. However, the three wavelet techniques outperform the histogram. This is due to the strong directional property of the texture regions that is captured by WHST and WMGT. The WDCT provides lower performance in non-object region indexing since the lowest-frequency subband are transformed by DCT so the high-frequency information is missed, however WDCT perform better than WHST and WMGT in object region indexing since the object regions tend to have less directional property and more low-frequency information.

## 5. Conclusions

Advanced content-based image retrieval has been proposed in this paper to allow querying images at region level. The K-mean clustering algorithm has identified regions. The Histogram and three other features that are based on wavelet domain have been used for region indexing. These techniques store and retrieve images and video, based on their contents. The performance of the retrievals of the four indexing techniques has been evaluated. The HIST provide a superior performance in the object region queries compared to the other wavelet techniques. The Histogram indexing provides a good indexing due to region identification process that isolates objects from other texture and objects in the scene that decrease the performance of the histogram as a global indexing. However, it fails to provide good indexing performance for non-object regions, since different texture can have very similar histograms. In the other hand for non-object region queries, the three wavelet techniques outperform the HIST technique. In both queries type, the WHST provide a performance comparable to WDCT and WMGT at higher complexity. The HIST is more suitable for object region indexing while wavelet techniques are more suitable for non-object region indexing.

### References


[1] **ISO/IEC JTC1/SC29/WG1**, *Document N390R, New Work Item: JPEG 2000 image coding system*, March 21, 1997.

[2] **IEEE Trans**. Circuits Systems Video Technol., *special issue on MPEG-4*, Feb. 1997.

[3] **Tou, J. T.** and **Gonzalez, R. C.**, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[4] **Bezdek, J. C.**, *Pattern Recognition with Fuzzy Objective Function Algo-rithms*. New York: Plenum, 1982.

[5] **Wang, J. Z., Li, J., Chan, D.,** and **Wiederhold, G.**, Semantic-sensitive Retrieval for Digital Picture, *D-Lib Magazine*, Nov, Vol. **5** Number **11** ISSN 1082-9873, 1999.

[6] **Unser, M.**, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Processing*, vol. **4**, no. **11**, pp. 1549-1560, Nov. 1995.

[7] **Swain, M.** and **Ballard, D**. Color indexing. *International Journal of Computer Vision*, **7(1)**:11-32, 1991.

[8] **B. M. Mehtre, M. S. Kankanhalli, A. D. Narasimhalu,** and **G. C. Man**, Color matching for image retrieval, *Pattern Recognit. Lett.* **16**, 1995, 325-331.

[9] **Smith, J. R.** and **Chang, S. F.**, Automated binary texture feature sets for image retrieval, in *Proc. ICASSP, Atlanta*, May 1996, Vol. **4**, pp. 2239-2242.

[10] **Abbadi, M. A.**. Content-based indexing and searching in image databases, *Thesis Diss. The George Washington Univ*. May 21, 2000.

[11] **Wang, J. Z.**, et al., Wavelet-based image indexing techniques with partial sketch retrieval capability, in *Proc. Forum on Res. And Tech. Adv. In Dig. Lib.*, May 1997.

[12] **Ameen, M.,** "Region clustering and identification for advanced Content-Based Image retrieval systems" *The George Washington University* 2002

# محرك بحث للصور مبني على محتوى الصورة

**محمد عبد الشكور أمين**

*جامعة الملك عبدا لعزيز، جدة، المملكة العربية السعودية*

**المستخلص** . لقد أصبحت فهرسة الصور والفيديو من الأمور المهمة مع الزيادة العددية في مجموعات الصور الرقمية مؤخرا. معظم طرق الفهرسة المقدمة في الأبحاث تعتمد على خواص تم حسابها من كامل الصورة. في ورقة البحث هذه، تم تجزئة الصورة بطريقة ك-مين العنقودية وذلك لتمكين البحث والاسترجاع على مستوى جزء من الصورة. بعد ذلك، تم تصنيف جزء الصورة إلى كائن أو غير كائن حتى نتمكن من استخدام خواص متناسبة مع النوع. إن صنف كائن يحتوي على أجزاء متكتلة بينما صنف غير الكائن منتشر في كامل مشهد الصورة ، كما تم تحديدها رياضيا بالكاي- سكوير الإحصائي. لقد تم تقييم أداء أربعة طرق فهرسة لأجزاء الصور الهستوقرام وثلاثة طرق مبنية على وافليت. الطرق الأربعة مناسبة لمجموعات الصور العامة. على الرغم من أن اداء الطرق المبنية على أساس وافليت متقارب لغير الكائن، إلا أن أداء طريقة الهستقرام كان أفضل من طرق الفهرسة الثلاث المبنية على أساس وافليت. الهستقرام أكثر ملائمة لفهرسة أجزاء تحتوي على الكائن، بينما طرق الفهرسة المبنية على وافليت أكثر ملائمة للأجزاء التي لا تحتوي على الكائن.