

# A Review of Content Based Image Mining System

**Manar Abdulkareem Al-Abaji**

*Computer Sciences Dept. , College of Education for Pure Science, University of Mosul, Iraq*

[Manar2000@yahoo.com](mailto:Manar2000@yahoo.com)

---

**ARTICLE INFO**

**Submission date:** 5/2/2019

**Acceptance date:** 14/4/2019

**Publication date:** 10/7/2019

---

**Keywords:** *image retrieval, TBIR, CBIR, color, shape, texture.*

---

## Abstract

In recent years, with the spread of the internet, there is a large amount of data available at it. Therefore, it becomes necessary to find fast search engines to retrieve images and documents. Image retrieval is a very significant area in digital image processing. To understand and learn more about "image retrieval system", the current study presents a review to describe the types of image retrieval techniques, explain the advantages and disadvantages of them. Moreover, this paper reviews different research studies and methodologies that applied to the area of CBIR.

## 1- Introduction

the noticeable increase in online image collections, such as satellite, medical imagery and art works, has attracted users in several professional fields like medicine, advertising, medicine, design, fashion, geography and publishing [1]. "Retrieving a set of images that is similar to the query image needed by the user from a database images is called image retrieval process". since 1970s, image retrieval has become a very interesting research area, especially in two major research areas: computer vision & database management [2].

Examples of image retrieval applications: "Computer Assisted Tile Recovery System" Which retrieves tiles from a catalog of digital tiles based on the pattern and color of the query tile, "Medical diagnosis" as a retrieval of medical brain images from a database similar to the image of the query in order to provide diagnostic support by presenting relevant past cases, "Crime prevention" having a database containing fingerprints, facial images, hair style, and shoe printing is useful in crime prevention. When committing a crime, the evidence collected from the crime scene is compared with the records in their database [3].

The purpose of this report is to review the techniques used to retrieve images, describe the advantages and disadvantages of each, as well as the techniques used to extract features. This report also includes some previous studies on various techniques in image recovery.

## 2- Image retrieval techniques

Image retrieval techniques can be classified into three categories [1]:

### A. Text-based image retrieval (TBIR)

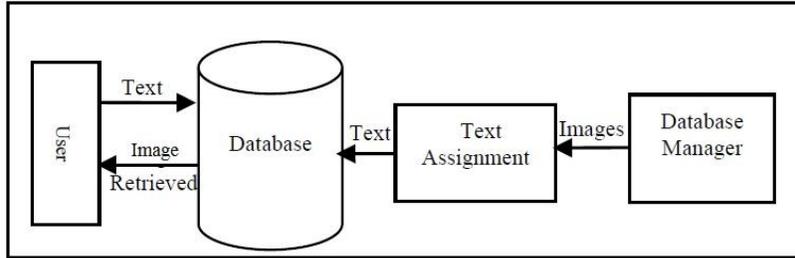
TBIR has been used since the late 1970s, Where the images are explained in a descriptive text firstly, and then they used the systems that manage the text-based database for the retrieval of images. TBIR uses the method of adding Metadata to images, such as captions, keywords or image descriptions. The retrieval that based on illustration words is complex because it requires addition of a huge annotation on the images manually and it also takes a long time, the typical TBIR system is illustrated in Figure (1) . However, there are many advantages and disadvantages of TBIR are shown below [4]:

Advantages of TBIR

1. The implementation of TBIR is easy
2. It is fast in retrieval

**Disadvantages of TBIR**

1. The difficulty of adding manual annotation for a large database
2. Many of the images are very rich in content and they need more explanation
3. Polysemy problem
4. The inaccuracy of writing the annotation



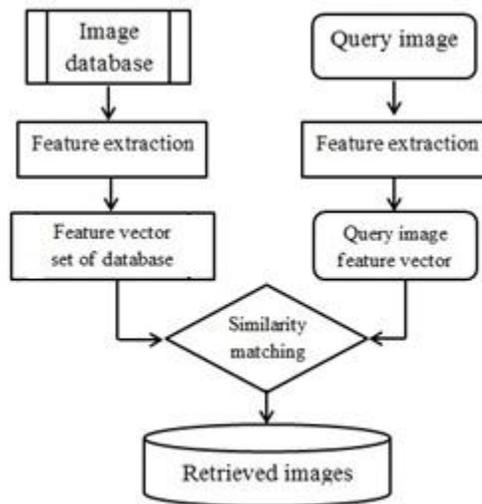
**Fig.1 Typical TBIR System**

**B. Content -based image retrieval (CBIR)**

In 1990, the method of retrieving image based on its content was introduced by T. Kato when he adopted the color and shape features to retrieve images from the database. After that, attention span to the use of the CBIR system was increased especially after the development of services provided by the internet, including the huge collections of images in industry, education, medicine, sciences and other fields available to all users. This has increased the difficulties faced by the TBIR method. All these needs are a driving force for the emergence and evolution of CBIR techniques [5]. the typical TBIR system is illustrated in Figure (2) .

Advantages of CBIR

1. In CBIR, "features (color, texture and shape) are extracted automatically".
2. The retrieve of images is done depending on the distances between features.



**Fig.2 Typical CBIR System**

**C. Hybrid Approach.**

In this approach, the image retrieval process will depend on the text as well as the image content to improve the accuracy of the retrieval. Both of them have their advantages and the combination of them can overcome their disadvantage.

### 3- Feature Extraction Technique

"In image retrieval system, a feature of image is a certain visual property of it. Many features can be extracted from images such as color, texture and shape to create a vector of feature for each image" [5].

#### 3.1. Color Feature

The color feature is one of the most important features used to search an image in a database. This is because "color plays an important role in the human visual perception system". There are many ways to extract the color feature from the image. These ways are divided into two groups: the statistical method for color representation and the color histogram. Many color spaces like LUV, RGB, YCrCb, HSV, HIS and LAB were used to describe the image in image retrieval system [6]. Table(1) illustrates many techniques of extracting the color feature with the pros and cons of each one.

Color Feature	Pros	Cons
Conventional color histogram	Simple, fast computation, invariant to any transformation	No spatial information
Color moment	Low complexity	Precision low
Fuzzy color histogram	Fast computation. Robust to quantization noise.	More computation
Color Correlogram	Encode spatial info	Very slow computation
Color Autocorrelogram	Encode spatial info, low complexity	Precision low
Color Structure Descriptor	Both color distribution and local structure of the color, Hmmd color space	No spatial information
Scalable Color Descriptor	Simple, fast, Hsv color space	No spatial information
Color Co-occurrence Feature	Include spatial info, low complexity, high speed, reduce the feature dimensionality	Codebook needed
Color Coherence Vector	Include spatial information	Complexity high

Table(1) Many Techniques of Extracting the Color Feature [3]

#### 3.2. Texture Feature

the best characteristics that are used to describe the surface appearance of image is texture. it gives many features about the surfaces of the images and the structural arrangement of objects in images such as clouds , fabric, leaves and bricks, because it depends on the distribution of images intensity. There are several techniques used in image retrieval system to extracting the texture feature from database images and store them as feature vectors, in order to used them when the query image is entered [7]. Table (2) illustrates the most common texture feature extraction techniques.

Texture Feature	Pros	Cons
Tamura feature	Effective retrieval	Highly complex
Wavelet filter	Detect different frequency and orientation	Precision low
Gray level co-occurrence matrix	Including position of pixels having similar gray level values.	High dimensionality
Gabor filter	High retrieval	Computation expensive
Steerable pyramid	Support any no of orientation	Storage space is high
Edge histogram descriptor	Computation easy	Retrieval result poor
Gabor moment	Low dimensionality	Computation expensive
Bit Pattern feature	Low complexity, high precision, capture edge visual pattern, texture information	Codebook needed
LBP,LTP,LTRP	Best feature for texture retrieval	Only suitable for gray scale image

Table(2) Many Techniques of Extracting the Texture Feature [3]

### 3.3.Shape Feature

Shape is a very important feature in image retrieval system; it is image's local feature and it is closed to the human perception as humans can better recognize the shape of an object. There are two kinds of Shape extraction method: region based method and the contour based method. Region based method deals with the interior of the boundary while the contour based method deals with the contour or boundary of an object in an image. As a result of both methods, the shape is represented by a set of values called the shape feature vector [8].

### 4- Related Work

Manimala S., et. al in 2012 [9], present a novel approach to retrieve image based on its content. This approach combines two types of features; "color and texture", and it is called WBCHIR: "Wavelet Based Color Histogram Image Retrieval". The texture features are extracted using wavelet transform while the color features are extracted using color histogram. By using the Wavelet transformation, the computational steps are significantly reduced and this leads to the increase of the speed of retrieval. the present approach outperforms the other retrieval methods in terms of Average Precision.

Hrudya P, et. al in 2012 [10], propose an architecture to retrieve the images using Semantic Web. their suggested text discovery method is dependent on the presence of temporary colors between the inserted text and its adjacent background. Firstly, they generated "The transition map based on logarithmical change of intensity and modified saturation", and then they generate linked maps to connect the components of each candidate area, after that they reshape every connected components to have smooth boundaries. The density of transition pixels were computed to distinguish between text and background.

S. Niranjanan, et. al in 2013 [11], introduced a novel method for CBIR, This method consists of many steps: resize image size to appropriate size [256,256], [128,128], [64,64], [32,32], [16,16] and [8,8], RGB Image convert to HSV, construct the histogram for Hue, Saturation and Value, generate n- signatures to them and store them. The same steps applied to query image to generate it's signatures. Finally vector cosine angle distance used to determine the distance between signature of query image and signatures of image database.

Md. Iqbal H. S., et. al in 2013 [12], proposed a CBIR method by using "texture and color features". Texture features are extracted by using a Haar wavelet transform and then they used F-norm theory to reduce the dimensions of the feature vectors. Color moments are used to extract the color feature. This method depends on Canberra distance to calculate the distance between database images and query image. the experiment results illustrate the higher accuracy of the proposed method than the methods that use a single feature.

MS. R. Janani, et. al in 2014 [13], presented a method that was using "color, texture feature, which are based on relevance feedback technique". In this method, The RGB image separated into its components, features are extracted using HOG descriptor, and they used DWT to extract the sub bands of images and get image's information at different

scales. The purpose of this paper is to study on papers complete by researchers in the area of the retrieval images based on the content based on color and texture.

Tie Hua Zhou, et. al in 2015 [14], presented “the Query-Frequency Pair (QF-P) algorithm” which is an integration of a probabilistic model to Combines image annotations to explore the relevance of one word to another to improve image annotations. QF-P algorithm improves image search results by mining keywords for representative images. They also discussed an engineering center model for querying several key words to study how to assign relevant comments, which can significantly improve the accuracy of retrieval. The integration between the two methods above is effective in the process of the query based on keywords or tags for various levels of classification.

D. Yuvaraj , et. al in 2016 [15], proposed a modified approach which was presented based on unsupervised learning, in this approach, image segmentation was computed and then features for color and texture and edge were calculated. After that histogram of gray scale image were constructed. "The experimental results significantly showed that region based histogram and colour histogram were effective as far as performance is concerned".

P. Kushwaha, et. al in 2016 [16], presented CBIR system that used multi kind techniques for feature extraction like edge histogram descriptor, GLCM and color moment to create a feature set. They used genetic algorithm to reduce the dimensionality and find the optimal features, and also they used K-Mean algorithm as a clustering technique. This research gave fast and better results.

M. G. Saeed, et. al in 2017 [17], proposed CBIR with a new approach of building feature vector. This vector consists of (140) values collected from many feature types like color moment, color histogram, GLCM, Gabor filters, moment invariants, wavelet transformation and tamura feature. To give for each image features their centroid , K-Mean was exploited here. Euclidian distance was used as a similarity measurement to retrieve the images that is similar to the query image.

M. K. Alsmadi, in 2018 [18], presented a study to extract important and robust features like color features using color histogram, shape features using canny edge detection method and color texture features using GLCM. At the same time, the genetic algorithm(meta-heuristic algorithm) has relied upon a similarity measure to retrieve the images from database that is similar to the query image. And gave good results by calculating the values of precision and recall from the results obtained in this study.

## 5- Conclusion

In this research, many techniques were introduced for image retrieval and a series of previous works were presented in the field of image retrieval in detail. It concluded that much work is needed to achieve high accuracy and shortest retrieval time possible.

The study also concluded that current techniques give good results when applied to a specific set (small set) of images, which means that they cannot be used on a wide range of images (in real applications) because the results will be low. Therefore, in the future, the image retrieval system needs to be optimized for use in real applications as well as to reduce the time it takes in retrieve process.

## CONFLICT OF INTERESTS

**There are no conflicts of interest.**

## References

- [1] Alaa M. Raid, Hamdy K.Elminir, Sameh Abd-Elghany, “A Literature Review of Image Retrieval Based on Semantic Concept”, International Journal of Computer Applications, vol. 40, no. 11, pp: 0975-8887. 2012.
- [2] Vaishali D.Dhale, A.R.Mahajan, Uma Thakur,“ A Survey of Feature Extraction Methods for Image Retrieval”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, issue 10. 2012.
- [3] M.S. Meharban and Dr.S. Priya, " A Review on Image Retrieval Techniques", Bonfring International Journal of Advances in Image Processing, Vol. 6, No. 2, pp:7-10. 2016.
- [4] Gulfishan F. Ahmed and Raju Barskar, “A Study on Different Image Retrieval Techniques in Image Processing”, International Journal of Soft Computing and Engineering, vol. 1, issue 4. 2011.
- [5] H. Wang, D. Mohamad and N.A.Ismail, “Approaches, Challenges and Future Direction of Image Retrieval”, Journal of Computing, vol. 2, issue 6. 2010.
- [6] Fuhui Long, H. Zhang and D.D.Feng, “Fundamentals of Content-Based Image Retrieval”, Multimedia Information Retrieval and Management”, Springer. 2003.

- [7] Felci Rajam, S.Valli, "A Survey on Content Based Image Retrieval", Life Science Journal. 2013.
- [8] Neha Sharma, "Retrieval of Image by Combining the Histogram and HSV Features Along with Surf Algorithm", International Journal of Engineering Trends and Technology (IJETT), Vol. 4, Issue 7, pp: 3137-3140. 2013.
- [9] Manimala Singha and K. Hemachandran, "Content Based Image Retrieval using Color and Texture", Signal & Image Processing : An International Journal (SIPIJ) Vol.3, No.1. 2012.
- [10] Hrudya P and Gopika N.G, 2012," Embedded Text Based Image Retrieval System Using Semantic Web", Int.J.Computer Technology & Applications, Vol 3 (3), pp: 1183-1188
- [11] S. Niranjanan and S.P. Raja Gopalan, "Performance Efficiency of Quantization Using HSV Colour Space and Vector Cosine Angle Distance in CBIR with Different Image Sizes", International Journal of Computer Applications, Vol. 64, no. 18, 2013, pp: 39-47. 2013.
- [12] Md. Iqbal Hasan Sarker and Md. Shahed Iqbal, "Content-Based Image Retrieval Using Haar Wavelet Transform and Color Moment", Smart Computing Review, vol. 3, no. 3, pp: 155-165. 2013.
- [13] MS. R. Janani and Sebhakumar.P, "An Improved CBIR Method Using Color and Texture Properties with Relevance Feedback", International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014, pp:47-54. 2014.
- [14]Tie Hua Zhou, Ling Wang and Keun Ho Ryu, "Supporting Keyword Search for Image Retrieval with Integration of Probabilistic Annotation", Sustainability, 7, pp: 6303-6320. 2015.
- [15] Duraisamy Yuvaraj and Shanmugasundaram Hariharan, "Content-Based Image Retrieval Based on Integrating Region Segmentation and Colour Histogram", The International Arab Journal of Information Technology, Vol. 13, No. 1A, 2016 ,pp: 203- 207. 2016.
- [16] Preeti Kushwaha and R.R.Welekar, 2016, "Feature Selection for Image Retrieval based on Genetic Algorithm", International journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, no. 2, pp:16-21, 2016.
- [17] Mostafa G. Saeed, Fahad Layth Malallah and Zaid Ahmed Aljawaryy, "Content-Based Image Retrieval by Multi-Features Extraction and K-Means Clustering", International Journal of Electrical, Electronics and Computers(EEC Journal), Vol. 2, issue 3. 2017.
- [18] Mutasem k. Alsmadi, 2018, "Query-Sensitive Similarity Measure for Content-Based Image Retrieval Using Meta-Heuristic Algorithm", Journal of King Saud University – Computer and Information Sciences, 30, pp:373-381. 2018

## استعراض نظام التعدين على أساس المحتوى

منار عبد الكريم

جامعة الموصل، كلية التربية للعلوم الصرفة، قسم علوم حاسبات

Manar2000@yahoo.com

الخلاصة

في السنوات الأخيرة، مع انتشار الإنترنت، هناك كمية كبيرة من البيانات المتاحة فيه. لذلك، يصبح من الضروري العثور على محركات بحث سريعة لاسترداد الصور والمستندات. استرجاع الصور هو مجال مهم جداً في معالجة الصور الرقمية. لفهم ومعرفة المزيد حول "نظام استرداد الصور"، تقدم الدراسة الحالية مراجعة لوصف أنواع تقنيات استرجاع الصور، وشرح مزايا وعيوبها. علاوة على ذلك، تستعرض هذه الورقة الدراسات البحثية المختلفة والمنهجيات التي تنطبق على مجال CBIR.

الكلمات الدالة: استرجاع الصور ، TBIR ، CBIR ، اللون ، الشكل ، الملمس.