



A comprehensive survey on content based image retrieval system and its application in medical domain

K. Srinivasa Reddy^{1*}, R. Anandan², K. Kalaivani³, P. Swaminathan⁴

¹Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.

²Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.

³Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.

⁴Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*Corresponding author E-mail: srmlgnlgap@yahoo.co.in

Abstract

Content Based Image Retrieval (CBIR) is an important and widely used technique for retrieval of different kinds of images from large database. Collection of information in database are available in different formats such as text, image, graph, chart etc. Here, our focus is on information which is available in the form of images. Searching and retrieval of the image from a large amount of database is difficult problem because it uses the image visual information such as shape, text and color for indexing and representation of an image. For efficient CBIR system, there is a need to develop different kinds of retrieval methods using feature extraction, similarity matching etc. Text Based Image Retrieval systems are used in many hospitals, but for large databases these are inefficient. To solve this problem, CBIR systems are proposed to retrieve matching images from database using automated feature extraction method. At present, medical imaging field finds extensive growth in the generation and evaluation of various types of medical images which are high inconsistency, usually fused and the combination of various minor composition structures. For easy retrieval, need to be development of feature extraction and image classification methods. Different methods are used for different kinds of medical images. The Radiology department and Cardiology department are the largest producers of medical images and the patient abnormal images can be stored with the normal images. CBIR uses query image as input and it retrieves the images, which are similar to the query more efficiently and effectively. This paper provides a comprehensive Survey about CBIR system and its one of the major application in medical domain.

Keywords: CBIR, image retrieval, feature extraction, medical images.

1. Introduction

Text based and visual based are the different types of image features. Keywords, tags, annotations are the textual features and shape, color, text etc. are visual image features. Image database consists of large number of images which are collected and stored. In this, data consists of raw images and can be extracted by computer-aided image analysis. Image content consists visual content as well as semantic content. An important research area in image processing is to develop image retrieval systems.

Text based Image Retrieval is a old method in which images are indexed using keywords which are used as retrieval keys and different users use different keywords for annotation, this makes as non-standardized. In this, there exist many problems of the query by text such as unexpressed feelings or emotions, synonyms and homonyms, misspellings etc. So, CBIR technique was developed to solve the problems of old method. CBIR is an important technique used for retrieving the images using feature extraction method both accurate and time efficiently. The term "content" refers to extract the features such as textures, colors, shapes or any additional information obtained from the image itself. The word "Content-based" means, the search searches the original contents of image rather than metadata such as tags,

keywords and descriptions related with the image. In CBIR, the features of the query image with each image from the database are compared. CBIR uses the feature components such as shape, texture, color or combination of these to achieve higher retrieval efficiency.

Earliest CBIR in 1980's is proposed by IBM is Query By Image Content (QBIC) which uses low level features only such as shape, text and color and does not include any semantic level. In QBIC, the way of image representation, query representation and comparison of the query are needed. CBIR also known as Content based Visual Information Retrieval (CBVIR) and uses the application of computer vision techniques to overcome the problems found in image retrieval. The basic components of CBIR are feature extractor which is used to generate metadata, a query engine used for similarity measurement and a user interface. The different kinds of CBIR are namely general and application specific. A general CBIR is used for matching a query from a large number of images. The CBIR which is used for matching a query from a large number of images of a particular type such as X-ray images of particular organs, finger prints etc. is known as application specific.

CBIR has a large number of applications in medical diagnosis and medical field. Radiology department and Cardiology department are the largest producers of medical images. In CBIR a sample database WANG is used for testing purpose which consists of 10

categories, if each category consists of N images, then WANG database consists of 10XN images. The performance of CBIR system depends on the parameters namely Precision, Recall, and Retrieval Time.

CBIR methodology consists in six steps namely Image acquisition, pre-processing, feature extraction, feature selection, classification, relevance feedback. The first step used to acquire an image from N number of images found in the database. To remove noise from an image, pre-processing step is used, which consists image segmentation stage by using noise removal filter, so thinning and cleaning noise are removed. Preprocessing is also used for object identification which effectively used for image mining application. In Feature Extraction step, features and characteristics of the objects are selected and used for the representation of maximum relevant information for a complete characterization of the image. To improve the prediction accuracy by reducing the feature space, feature selection step is used. And Classification process involves in two phases that are training phase and testing phase. Relevance feedback gives the feedback to the previous result thereby increases the efficiency of Image retrieval systems. In CBIR, the different types of challenges occur due to computational complexity occurs in retrieving process and to index, storage problems and usability.

One of the most important applications of CBIR is in the medical domain even though the medical diagnostics is the hardest. The CBIR system mainly contains two phases named as feature extraction phase and classification phase. In Medical Imaging, one of the important applications is MR brain imaging. In case-based or evidence based reasoning, when several images are available for diagnosis, CBIR can be guideline that might give a correct diagnosis. The earliest Computer aided diagnosis (CAD) CBIR has limited applications such as photography of the skin, microscopy, radiographs of spine, magnetic resonance Imaging of head. Image Retrieval Techniques are text and content-based. Text based algorithm comprises the Keywords and annotations which are given to all images in data base. Annotation requires more time in its operation, tedious, it requires more description. In some annotations the complete image features are not mentioned. In CBIR system, images are indexed by their content automatically. Both techniques together using different features can also be used to achieve higher performance compared to use each technique separately.

The organization of the survey is presented in different parts. Part 1 presents the detailed information about CBIR System. The major application of CBIR in Medical domain for retrieving similar medical images is presented in part 3. The discussion about the Literature survey is presented in part 4. The last part gives the conclusions about CBIR Systems and its application in medical image retrieval systems and also presented the suggestions regarding the scope of possible future work.

2. An overview of CBIR system

Feature extraction method

It is the basis of CBIR System. One of the visual features consists of primitive features such as Shape, Color, Texture.

Color: It is the fastest feature and requires low memory space in CBIR and can be used to provide much information about the visual content of an image. It is the important and most widely used feature in CBIR because of its easy and fast computation. The models of color are hardware based (RGB) and user based (HSV). Many images consists Red, Green, Blue (RGB), but RGB cannot separate luminance from chrominance component which is not suitable for perceiving human colors. So, Hue, Saturation, Value (HSV) color space can be used commonly for image retrieval and each component in this model contributes directly to visual perception. Block-based, Color moments, Color Correlogram, histograms are the common examples for the color feature. For computing distance based on the similarity of color

for entire image, a color histogram method is used. The key components of the color feature extraction are color quantification, color space and similarity measurement and this feature do not depend up on image size. Color indexing can be done by using color moments and is good under lighting conditions. The memory storage requirement for color moments is very less.

Shape: Shape contains the visual information for human perception in most attractive form in which data can derived from images directly. In CBIR, numerous systems uses shape feature after image have been segmented into objects or regions. So, Shape features can be separated into two different classes such as region based and boundary based which are generally used in similarity retrieval. Vertices to edges, contours, joints, polygon areas removed from an image and perceptually grouped geometric cubes are example representations of Shape feature. Consecutive Boundary sets is the example of local features, the examples of global features are moment invariants, circularity, aspect ratio are usually used in Shape feature. It is invariant to image scaling, rotation, translation. Fourier Descriptors, moment Invariants are used for representation of shape feature.

Texture: Texture is used to provide information about structural surface arrangement. A simple and effectively used method to represent as well as for analysis of texture is a gray level co-occurrence matrix. Study of texture and analysis can be done at three different levels namely statistical level, structural level and spectral. Statistics extracted set from image is represented as texture in statistical where as in structural level, image primitives and placement rules said to be texture. In spectral, texture can be described as coefficient set in transform technique. There are six properties for Texture namely Coarseness, contrast, dimensionality, line lankness, regularity and robustness.

Similarity measurement

Similarity function is also known as a matching function. It is used to compute the content difference among images (search parameter and the other images stored in the database) based on their features. Similarity measurement is a measurement of how close a vector to another vector and the similarity can be represented by computing the content difference among the features of the query image and all images in data base. The Image Retrieval systems return the image whose distance from the query is minimum at first and returns the images whose distance from the query is maximum at last. Similarity measures can also used to measure Color, shape, texture, object, relationship similarity.

Methods of CBIR

Shape Based Method: In this Method, feature extraction consists of N dimensional feature vector considered as a point in N dimensional space. Using the extracted feature vectors, images are indexed in database after calculation of distance among the features of query image with the images in the database. Image Retrieval systems are used for retrieval of images after finding the similarity among query and images in database. Different matching functions can be used.

Texture Based Method: Wavelets and Gabor Filters are used to capture the texture of images. The better performance of Gabor filters are used for capturing the image characteristics or its parts according to changes in scales and directions. It also used for images, regions using homogenous texture.

Continuous Feature Selection Method: It is used to solve Semantic gap problem. To overcome the Semantic gap, it uses a statistical association rule, to appropriate specialist's high knowledge with low features. According to their relevance to weight the features, these rules are used. By discarding irrelevant features, dimensionality reduction is performed. To improve query precision, experiments are conducted and proved that the complexity of the query decreases with the decrease in dimensionality and concludes that this is efficient method.

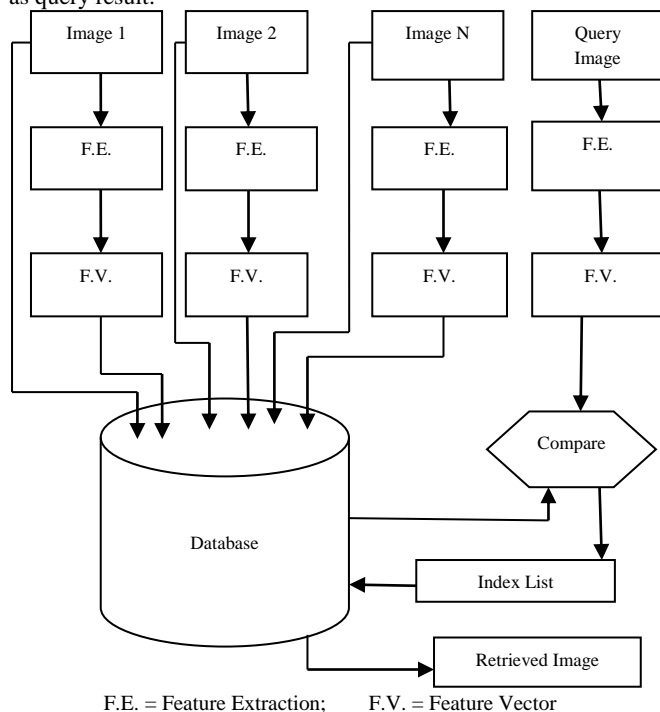
Storage and Access Methods: CBIR used as large system, which is used for storing large number of images which requires large database known as common storages also used as relational databases, inverted files, self made structures, indexing in memory. Space or dimension reduction, pruning techniques often need for quick accessing to the data efficiently. Principal of Component Analysis and KD trees are some indexing techniques used for feature space reduction or dimension reduction. Independent Component Analysis is also another space reduction technique. There is no common database with known relevance's for evaluation in CBIR, two databases are used. A CBIR task with arbitrary photographs is represented by using the WANG database. IRMA database is used to represent detailed content of objects involving on the images.

Image retrieval techniques

For searching and browsing digital images from database a technique named as Image Retrieval is widely used. An indexing scheme is used for effective searching and effective image retrieval. Among the important, a prominent technique is a relevance feedback which is used whenever sometimes unwanted images found in the result. Artificial intelligence, Gaussian mixture models are also used as Image Retrieval Techniques.

Architecture of CBIR system

The general architecture of CBIR System is as shown in figure 1. Input is given as query, system extracts its feature values and can be represented as feature vectors by the feature extraction stage and compares with the features of Image 1 to image N stored in the data base by using comparison stage. It gives the rank for the database images based on content difference among the query image, which can be used for retrieving similar images and given as query result.



Applications of CBIR

There are numerous applications of CBIR. Some of them are as follows:

- **Medical Domain:**
Medical diagnosis: Used to diagnosis by identifying similar past cases, tumor detection, to improve MRI and CT scan comprehension ability.

Medical applications of CBIR: PACS/Health database management, computer aided diagnosis, medical research, education and training, Blood cell detection.

- Digital forensics: Finger print Matching for crime detection.
- Investigation: face recognition systems on the internet used by police force in crime prevention.
- Intellectual Property: Trademark registration, for identification of property ownership.
- Shape Identification: Identification of fault and defect in Industrial automation.
- Remote sensing: Weather forecast, Satellite Imaging, Space programs, Geographical Information and Various other information systems
- Radar Engineering: helps in identification and detection of targets.
- Architectural and Engineering Design :Cartography Synthesis of Weather maps, mapping from photographs
- Publishing: Journalism, advertising media, fashion and graphic design.
- Miscellaneous Applications: Home entertainment, Trademark database, Art galleries, Archaeology, Image search, Social networking, Digital Libraries, Agriculture, Military, Copy right on the internet, Bio-diversity information systems, Tourist Information etc.,

Performance measurement

Euclidean Distance: It is defined as the distance computed from the difference between unknown and known feature vectors of the images & given as

$$D = \sum (A_i - B_i)^2.$$

Weighted Euclidean Distance: It is calculated by multiplying squared differences by corresponding weights and can be represented by using the formula as:

$$dx, y = \sqrt{\sum_{j=1}^i \frac{1}{s_j^2} (X_j - y_j)^2}$$

Performance Parameters: The performance parameters namely Precision and Recall are the most commonly used for measuring performance of retrieval systems in order to solve the evaluation problems in retrieval process found in CBIR.

The ratio of various retrieved similar images to the total number of retrieved images in response to the query is called as the Precision.

The ratio of various retrieved similar images in response to the query to the total number of similar images in the database is known as recall rate.

3. Application of CBIR in medical domain

Medical CBIR

Medical Images require a special design. Some medical image acquisition devices such as computed tomography scanners (CT), magnetic resonance imagers (MRI), ultrasound probes (US) and nuclear imagers generate images with varying different properties such as, resolution, contrast, signal to noise ratio on human anatomy and physiology. In medical filed flipping the image during the search retrieval process has been ignored by previous researchers.

Types of medical imaging

Medical Imaging has undergone major advancements in modern medicine and useful in clinical applications. Conventional X-ray, computed tomography (CT) and mammography are the X-ray based methods of Medical Imaging. Contrast agents are used to enhance the X-ray image, for example angiography examination. A variety of methods used to visualize biological process in

Molecular Imaging which used in Nuclear Medicine. Magnetic resonance Imaging (MRI) and ultrasound imaging are other type in Medical Imaging. MRI and Ultrasound operates without ionizing radiation and uses strong magnetic fields.

Need for CBIR in medical

In order to improve the quality and efficiency of care process, the goal of medical information systems need to deliver the needed information to right person, the right place at the right time. This goal need query by Patient name, Series, Study ID for the images and can be beneficial for the clinical decision making process. CBIRs have been implemented in several departments. Besides diagnostics, used in teaching and research are improved through the use of visual access methods. Mammographics are one of the most frequent application areas with respect to content based search and classification in the radiology department. Another area of classification is a high resolution computed tomography scans MRI of brain are used to demonstrate the image search algorithm. After segmentation, shape properties are used to search for medical tumors by CT brain Scanner. For Medical image retrieval spine x-ray and functional photon emission tomography (PET) are also used. When several images are available for diagnosis, there is a need for finding similar medical cases as references and the Image retrieval system can be used as a guideline that might gives a correct diagnosis. CBIR proposed for wide use in medical and medical management, but only a few CBIR systems have been developed specially for medical images. One of the important among them is IRMA (Image Retrieval in medical images). The seven consecutive steps such as categorization, registration, feature extraction, feature selection, indexing, identification and retrieval that are combined to form the image retrieval process in IRMA system.

Medical image retrieval techniques

These are text and content-based. Most CBIR systems use the query by examples needs an appropriate starting image for query purpose. Sometimes missing the starting page known as page zero problems, so image consists of text and this is used as a starting point. A generic Image Retrieval uses sketches and used in medical field also. This method is only being applicable for a very small subset of queries which consists the difficulty in exact drawing and also it needs the artistic skill and time. So, the combination of both visual features and content and context of the images with textual does have the potential to achieve good results. Medical patient records also contains details such as age, gender, also profession of the patient, these details can also used to put the images into content in information retrieval besides the free text.

Architecture of M-CBIR system

The architecture of CBIR to retrieve medical images (M-CBIR) given as shown in figure 2.

The different kinds of medical images can be stored as input1, input2,....., input N into the medical image database. For medical images, Image enhancement and also Image segmentation can be done before extraction of the features. After extracting the feature vectors of all kinds of medical images they are stored in the feature database. The same process can be used for a given query image, system extracts the feature vectors by using feature extraction stage. Similarity comparison stage compares features of query image with features of all stored images in feature database and returns similar images as result of the Retrieval process in response to the query image.

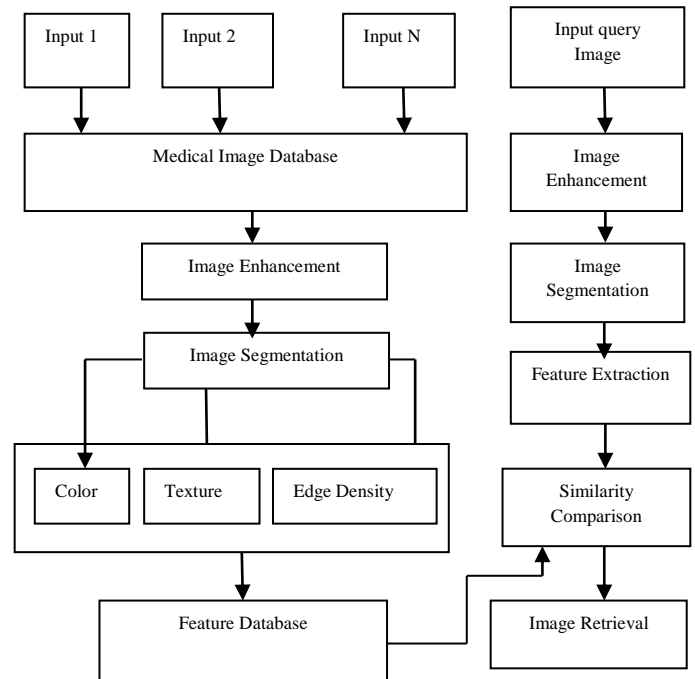


Fig. 2: A schematic diagram for a typical medical CBIR (M-CBIR) system

Retrieval process used for medical images

Following process is used for the retrieval of similar medical images from the medical database.

- Input the medical image to the M-CBIR system and extract the features by using Haar Wavelet (HW) or Fourier Descriptor (FD).
- From the Medical database, Collect and format the medical images and read all the medical images consecutively to extract the features using HW or FD technique.
- Using Euclidean or Canberra distance (CD) techniques compare the query image and database image features.
- Results can be stored in the data base.
- Results can be sorted and displayed according to the similarity.

4. Literature review

One of the earliest papers in image retrieval is reported in [1]. In their work, the authors proposed a method to test Color images available in COREL data base, and noted from the results that recall rate is increased by 5% , fall out rate decreased by 0.06% over the Legendre chromaticity distribution moments based method. Weidong Cai et al., [2] proposed the method for efficient 3-D CB neurological image retrieval based on CMRGlc texture patterns.

Ashish obero et al.,[3] designed CBIR system using different techniques and evaluated its performance on different kinds of images. Jagpal Singh et al., [5] discussed the efficiency of different techniques and the combination of them to improve performance. They also described the methods which exhibits good precision and performance. Y. Fanid Fathabad et al., [6] proposed the retrieval of brain images from the data base using CBIR systems.

R. Senthil Kumar et al., [7] proposed a improved method for content based system used for medical field. Ms. Anchal et al., [8] proposed various methods of CBIR. A.R. Mahajan et al., [10] proposed the numerous kinds of images with advancements in medicine and image technology. Neha [11] provided a Comprehensive Review on CBIR and Image Retrieval work giving a detailed review of the traditional and modern CBIR.

Ashish Mohan Yadav et al., [12] proposed fundamental techniques for CBIR along with features and matching function

measurement. Amit Singh et al., [13] described the different methods used for content retrieval. Jahnvi Shukla et al., [14] described all feature extraction methods i.e., color, shape, texture and the researchers can combine any of these methods to get highest Precision and Recall by testing various combinations.

Marjan Mansourvar et al., [15] described a method to overcome a problem from retrieving images automatically from a large data base based on different feature combinations and provides an over view of CBIR, its application in medicine, also proposed retrieval techniques used for critical development in medical systems. Mohammed Hamzah Abed et al., [16] proposed an experiment which can be done with the help of Matlab to test different methods on color histogram.

L. Haldurai et al., [17] proposed a technique which explores various methodologies in extracting implicit knowledge, patterns and relationships focused in the images from the collection of images.

Devbrat Deya et al., [18] focus the research from algorithms of low to high-level visual feature extraction mechanism. S. Banuchitra et al., [19] presented the goal of the survey to provide an overview of the functionality of CBIR. Various features and their method of representations are described. Qusai Q. Abuein et al., [20] focused their research and concluded about the performance of the current and old systems used for retrieving the flipped images.

Ricardoda Silva Torres et al., [21] proposed content based information retrieval approach and proved the existence of automatic retrieval process. Henning Millier et al., [22] proposed a system that seem to be used in clinical practice with visual search tools. Sameer K. Antani et al., [23] describes the combining methods used in different medical applications for retrieving the images.

Epa Phrodite Uwimana et al., [24] proposed the automatic classification methods for medical images, the results with an error-rate of 1% on image modality. Joao Augustoda et al., [25] developed an application of the technique over an image data base using the Euclidean distance metric. Zobiari Raisi et al., [26] applied CBIR using CR algorithms for providing services to tourism using network and internet.

5. Conclusion and future scope

Here, we described the significance of CBIR systems and retrieval of images using feature extraction methods and also discussed about the overview of CBIR systems. This paper also suggested that CBIR is used for variety of applications and concluded that it has many applications in Medical field. This paper is also given a direction for future researchers there is a need to develop new methods and tools which are required to manage the increasing amount of medical images. Our future work will be directed to retrieve the medical images using space reduction techniques.

References

- [1] Ananth Raj P & Venkataramana A, "Krawtchouk Chromaticity Distribution Moments for Content Based Image Retrieval", *NCC IIT Guwahati*, (2009).
- [2] Cai W, Liu S, Wen L, Eberl S, Fulham MJ & Feng D, "3D neurological image retrieval with localized pathology-centric CMRGlc patterns", *17th IEEE International Conference on Image Processing (ICIP)*, (2010), pp.3201-3204.
- [3] Oberoi A & Singh M, "Content Based Image Retrieval system for Medical Data bases (CBIR-MD)-Lucratively tested on Endoscopy, Dental and skull images", *International Journal of Computer Science Issues*, Vol.9, No.3, (2012).
- [4] Laban N, ElSaban M, Nasr A & Onsi H, "System refinement for content based satellite image retrieval", *The Egyptian Journal of Remote Sensing and Space Science*, Vol.15, No.1, (2012), pp.91-97.
- [5] Singh J, Kaleka JS & Sharma R, "Different approaches of CBIR Techniques", *International Journal of Computers & Distributed systems*, Vol.1, No.2, (2012).
- [6] Fanid Fathabadi Y & Balafar MA, "Application of content based image retrieval in diagnosis brain disease", *International Journal on Technical and physical problems of Engineering*, Vol.4, No.13, (2012), pp.133-138.
- [7] Senthil Kumar R & Senthilmurugan M, "Content based Image retrieval system in medical applications", *International Journal of Engineering Research & Technology*, Vol.2, No.3, (2013).
- [8] Solio AA & Ladhake SA, "A review of query image in content based image retrieval", *International Journal of Advanced Research in Computer Engineering & Technology*, Vol.2, No.4, (2013).
- [9] Bhende P & Cheran AN, "Content based image retrieval in Medical Imaging", *International journal of computational Engineering and Research*, Vol.3, No.8, (2013).
- [10] Mahajan AR, Zade SD & Raut P, "Content based image retrieval in medical images: current-status and future directions", *International Journal of Application or Innovation in Engineering & Management, Special issue for National Conference on Recent advances in technology and management for integrated growth*, (2013).
- [11] Neha, "Content-Based Image Retrieval: A Review", *International Journal of Science, Engineering and Technology Research*, Vol.3, No.5, (2014).
- [12] Yadav AM & Sengar BPS, "A survey on content based image retrieval systems", *International Journal of Emerging Technology and Advanced Engineering*, Vol.4, No.6, (2014).
- [13] Singh A, Shohani P & Kumar M, "A Review of Different Content Based Image Retrieval Techniques", *International Journal of Engineering Research and general science*, Vol.2, No.5, (2014).
- [14] Shukla J & Vania J, "A survey on CBIR Feature Extraction Techniques", *International Journal of Engineering and Computer Science*, Vol.3, No.12, (2014), pp.9555-9559.
- [15] Mansourvar M & Ismail MA, "Content based image retrieval", *International Journal of Information Technology*, Vol.20, No.2, (2014).
- [16] Abed MH & Al-Farhoosi Iraq DSJ, "Content based image retrieval based on Histogram", *International Journal of Computer applications*, Vol.110, No.3, (2015).
- [17] Haldurai L & Vinodhini V, "A study on content based image retrieval systems", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.3, No.3, (2015).
- [18] Arya D & Jha J, "A review on content based image retrieval using feature extraction", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol.6, No.3, (2016).
- [19] Banuchitra S & Kungumaraj K, "A Comprehensive Survey of Content Based Image Retrieval Techniques", *International Journal of Engineering and Computer Science*, Vol.5, No.8, (2016), pp.17577-17584.
- [20] Abuein QQ, Batiha R, Shatnawi MQ, Al-Aiad A & Amareen S, "Content Based Image Retrieval for Medical Applications with Flip-Invariant Consideration using Low-Level Descriptors", *International Journal of Advanced Computer Science and Applications*, Vol.7, No.4, (2016).
- [21] da Silva Torres R & Falcao AX, "Content-based image retrieval: theory and applications", *RITA*, Vol.13, No.2, (2006), pp.161-185.
- [22] Müller H, Michoux N, Bandon D & Geissbuhler A, "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions", *International journal of medical informatics*, Vol.73, No.1,(2004), pp.1-23.
- [23] Antani SK, Deserno TM, Long LR, Güld MO, Neve L & Thoma GR, "Interfacing global and local CBIR systems for medical image retrieval", *Bildverarbeitung für die Medizin*, (2007), pp.166-171.
- [24] Uwimana E & Ruiz ME, "Automatic classification of medical images for content based image retrieval system", *Proceedings of the AMIA Annual Symposium*, (2008), pp.1159-1165.
- [25] da Silva Júnior JA, Marçal RE & Batista MA, "Image Retrieval: Importance and Applications", *Workshop de Vis-ao Computacional-WVC*, (2014).
- [26] Raisi Z, Mohanna F & Rezaei M, "Applying Content-Based Image Retrieval Techniques to Provide New Services for Tourism Industry", *International Journal of Advanced Networking and Applications*, Vol.6, No.2, (2014), pp.2222-2232.