A new approach for content-based image retrieval for medical applications using low-level image descriptors

Mohammed Q. Shatnawi[1], Mohammad Alrousan[2], Suzan Amareen[3]
[1,2,3]Faculty of Computer and Information Technology, Jordan University of Science and Technology, Jordan
[2]Department of Computer Information Systems, Jordan University of Science and Technology (JUST), Jordan

Article Info

ABSTRACT

Content based image retrieval (CBIR) has become an important factor in medical imaging research and is obtaining a great success. More applications still need to be developed to get more powerful systems for better image similarity matching, and as a result getting better image retrieval systems. This research focuses on implementing low-level descriptors to maximize the quality of the retrieval of medical images. Such a research is supposed to set a better result in terms of image similarity matching. In this research a system that uses low-level descriptors is introduced. Three descriptors have been developed and applied in an attempt to increase the accuracy of image matching. The final results showed a qualified system in medical images retrieval specially that the low-level image descriptors have not been used yet in the image similarity matching in the medical field.

Corresponding Author:

Mohammed Q.Shatnawi,
Faculty of Computer and Information Technology,
Jordan University of Science and Technology,
Ar Ramtha 3030، Ar-Ramtha, Jordan.
Email: mshatnawi@just.edu.jo

1. INTRODUCTION

The traditional way to find images is by assigning keywords to each image, and then using some textual query the needed image(s) will be retrieved. This approach sometimes can be time consuming when large number of images is searched, or when the assigned keywords for the images are not relevant to the image content. Content based image retrieval (CBIR) means that images can be searched by their visual content, such as graph, text, color [1], and local and global features [2]. CBIR has many methods for analyzing images; each method represents different aspects of the visual information of the image. Image searching and image archival can greatly reduce the time consumed by using automatic image analysis tools [1].

CBIR consists of different phases. In the first phase, the images are analyzed and inserted to the image database. In the second phase, the images can be queried based on color, shape, and texture. In the third phase, the search queries are issued by giving an initial image, or by starting with random images from current images in a specific database. In the last phase the query continues so that images can be marked as positive or negative samples to refine the search and to get better results [1]. Another classification for CBIR phases as mentioned in [3] the phases are feature extraction, feature storage in which the calculated features are stored somewhere, feature comparison to define the similar images, and the query interface which is used to initiate the search process.

CBIR has achieved a great success in medical images applications, since it has the advantage of having easy and efficient comparison ways between two images. These comparisons provide the specialists with the needed information in the fastest way it could, and then it helps to improve their decisions based on more information resources. Therefore, to perform these comparisons a huge amount of images already stored in specific database is required. There are still many limitations in storing, searching, and retrieving

images. Also there are limitations in processing large databases. Thus more researches on these limitations are still needed [4, 5].

Based on the above there is a need to do more researches on the field of image similarity matching. Medical sector is one of the major sectors that can be benefited from the progressive development in the field of image processing and analysis. The reason behind this is trying to get to a level that reduces the human intervention of analyzing the medical images. Human errors during the patient's diagnostic procedures are not accepted at all. An accurate and efficient automated system can provide the physicians with at least a preliminary diagnostic-based decision. The provided decision can be investigated further by either the physician themselves or other well-implemented automatic system. This research addresses the image retrieval process based on the images' content and provides a thoroughly discussion of this approach

2. RESEARCH METHOD

Explaining research chronological, including research design, research procedure (in the form of algorithms, Pseudocode or other), how to test and data acquisition [1, 3]. The description of the course of research should be supported references, so the explanation can be accepted scientifically [2, 4]. Although many approaches in finding images using CBIR have been developed, there is still a need to do more researches in CBIR for medical applications. The developed proposed approach works by extracting content descriptions for each image (feature extraction), after that the semantic metadata will be built by developing similarity matching algorithms. These algorithms are used to compute the matching ratio between images. By having all the needed input and output data, the retrieval accuracy rate for the system can be calculated. This method is much better than using text to retrieve any patient's information. Although textual methods are widely used due to its simplicity, it is not very efficient; as it requires a domain expert. Also, the textual interpretations of image content may vary from person to person, which leads to store a huge data about patients. In addition, the textual data do not have a standard base [3]. That's why finding the similar cases in medical applications based on image content gives better information and knowledge for specialists, and has proven great benefit in teaching [3] by helping instructors and students to browse and access educational medical datasets and view the retrieved results [2].

In order to define the similarity between two images the access to image information must be performed at the perceptual level. Or in another words the low-level features such as color, texture, shape, etc. These features are also defined as low-level descriptors. Two or more features can be integrated together to define new descriptor [6]. The main advantage of using low-level descriptors is the computational efficiency which its attributes provide. This makes it widely preferred and used in the CBIR areas. Many advanced algorithms have been developed but sometimes it does not model the image semantic properly, that's why there is a need to use as many features as possible, or good enough features to define the best similarity between images [7].

2.1. Lucene image retrieval

One CBIR approach as described in [8] is Lucene Image Retrieval (LIRe). This approach is an extensible CBIR open source Java library. It extracts image features that are stored in specified repository and stores them in Lucene index for later retrieval. There are many image features included in this system such as color histograms in RGB, Tamura texture features coarseness, contrast and directionality, color and edge directivity descriptor (CEDD), fuzzy color and texture histogram (FCTH), and many other descriptors. The system is highly dynamic as new features added to the literature can be implemented and then integrated to the whole system framework.

The main low-level features are color, shape, and texture. There are many low-level global features, such as color, edge, texture, and average features. From them we can extract many descriptors such as:
- Color layout descriptor (CLD) which is a low level feature that represents the spatial layout of the images in a compact form. It's generated by applying discrete cosine transformation (DCT) on the image to form a 16-dimensional feature vector [6]. Using this feature, image features can localize the image in separate 4x4 sub-images [8].
- Edge histogram feature (EHD) is a low level feature which utilizes the spatial distribution of image edges; it's represented by dividing the image into 4 x 4 sub-images and then generates histogram from the edges of each sub image [6]. After generating the histograms, they are categorized into one of five types. These types are vertical, horizontal, diagonal with 45 and 135 degrees, and non-directional edges. Sometimes the sub-image can be categorized as non-edge block [9]. There are future hopes and attempts to develop a joint representation of EHD and CLD [10].
- Color of edge directivity descriptor (CEDD) which is a low level feature that can be extracted from the image. It incorporates color and texture information in one histogram by segmenting the image first into

24 color regions, and 6 textures regions, then creating the final histogram that includes 6 x 42=144 regions. As a low level feature, this feature is suitable to be used in large image databases since it costs low computational power. However, the size in this feature is limited to 54 bytes per image only; this makes the comparison between plenty of images an easier task [11]

- Fuzzy color and texture histogram (FCTH) is also a low level feature that can be extracted from the image. It incorporates both color and texture information in one histogram. This feature can be obtained by combining 3 fuzzy systems together. It is also suitable to be used for large image databases; it costs low computational power as well. The size of this feature is limited to 72 bytes per image which makes comparisons between thousand images an easy task. As in [12], this feature has proved an accurate retrieval feature even in hard cases when the images contain some noise or smoothing.
- Color structure descriptor (CSD) is based on color histogram [13], it's considered as an accurate descriptor in providing localized color distribution of each color cm. This descriptor is specified by:
- h(m), m=1, …, M where m ∈ {32, 64, 128, 264}.
- h(m) represents the final value of the number of positions where the structured elements contain color cm [14].

There are many other low-level descriptors such as grey level co-occurrence matrix (GLCM), and color coherence vector (CCV) that can be developed and used [11]. Although all of these descriptors have been developed in LIRe, only one descriptor can be used in one retrieval phase. The end-user must select the descriptor he/she wants to choose, and then the retrieval of the images will be based on this descriptor. An integration of all of these features has not been used yet. We took this advantage and designed a system that uses more than one descriptor in the retrieval phase. Also LIRe have not been used in the medical retrieval systems, so we decided to choose this system for medical retrieval objective, and compare the performance results between LIRe and our developed system.

2.2. ImageCLEFmed, medical image retrieval task test collection

In [15, 16] ImageCLEFmed a medical image set free of charges and copyrights of more than 300,000 images along with image annotations has been created. Searchers and learners may use it in many medical retrieval fields to accomplish many tasks such as evaluating the efficiency of existing systems, or by testing and evaluating new developed tools. Having such a test collection made it possible to evaluate and access the performance of medical IR systems with a real collection test, so the comparisons and assessments for these systems would be realistic. In addition, these test collections as mentioned in [17] are used in image retrieval outside medical fields.

As mentioned in [15, 16] creating such a test collection will help researchers to test how well new systems will operate in retrieving relevant documents. Common measures are used to evaluate the systems such as recall and precision. Most searchers focus on the precision values by calculating the final average of all calculated precision values for each test task for the system. This measure is called mean average precision (MAP) which is also known as the most frequent measure used in the TREC collection test mentioned in [18].

The ImageCLEFmed test has been created after three years of gathering, organizing, and storing since 2005 till 2007. The main ImageCLEFmed dataset includes many sub-collections that include set of images with its annotations, these collections came from four resources in the first two years, and in 2007 two additional resources were added. The structure of the ImageCLEFmed is shown in Figure 2 [15]. As shown in this figure ImageCLEFmed is presented as the main library which consists of multiple collections. These collections were gathered into one main repository, each collection consists of multiple images and annotations. Each annotation is an Extensible Markup Language (XML) file. The annotation has description of specific medical case, and may include one or more image, these images are related together, and with the description these images represent a specific medical case (e.g. Renal cell carcinoma). ImageCLEFmed has 6 main collections, as specified in Table 1 [15].

Along with creating the dataset, a number of 85 topics have been developed. These topics were generated from variant real-world medical search engine logs on the internet. Topics can refer to one or more of the following: imaging modality, anatomical location, view for some medical case, and disease finding. An image would be relevant to a specific topic if it meets all explicit mentioned terms. In other words, it is an (AND) relationship not an (OR). The synonyms when searching for some topics are considered. Synonyms in this system are not language-based that depends on only stemmers but rather medical-based synonyms.

Recall and precision measurements are used in evaluating the system with an aggregated measure which is mean average precision (MAP). MAP is computed by taking the average precision for all topics [19]. The similarity matching in this case can be stated as fusion-based similarity matching [20]. Using the ImageCLEFmed collection as a dataset containing the images and textual annotations; the result of retrieval of the top 10-30 images for each retrieving task reached 50% of precision in some cases. Retrieving

10-30 images aims to retrieve a good enough set of images, although the relevant images in the dataset might be much more. The main objective for this research is not only enhancing the performance results, but also to having more researchers use this collection to evaluate the performance of new developed systems and develop more approaches in the future [21]. We have used the ImageCLEFmed as a test collection for our system, we contacted the authors of ImageCLEFmed dataset, and they gave us an access to it [6].

Table 1. ImageCLEFmed collection description

Collection Name	Image type(s)	Cases	Images	Annotations	Languages
Casimage	Radiology and pathology	2,076	8,725	2,076	French, English
MIR	Nuclear medicine	407	1,177	407	English
PEIR	Pathology and radiology	32,319	32,319	32,319	English
PathoPIC	Pathology	7,805	7,805	15,610	German, English
myPACS	Radiology	3,577	15,140	3,577	English
CORI	Endoscopy	1,496	1,496	1,496	English

2.3. The proposed system

The main objective of this research is to design and implement a CBIR system that is capable to search for medical images using medical images as an input. Thus, the system framework will be similar to any other CBIR system consisting of the same phases. Although, many approaches for CBIR systems have been proposed in different ways, they are eventually the same. Every developed system goes through the same phases. However, the way in describing them only differs from a system framework view-point to another. The system in this research uses Low-level Features and uses CBIR to retrieve medical images. We choose LFCBIRM to define the system.

In the proposed system several algorithms were developed under the MATLAB platform. The proposed system consists of three main components. The first component is the one that is responsible for extracting the features. The features can be extracted from both the input image or the input query. The second component is the database that has the collection of images' features in general. The third component is the one that compare between the extracted features and the feature dataset (i.e. stored in the database). Based on the comparison result, the input will be diagnosed within the correct category.

2.4. Components of LFCBIRM system

The system, as shown in Figure 1, works as follows. The user who is assumed to be a medical specialist or a student (e.g. medical student) will enter an image query to the main system interface. In this system the query is a medical image (e.g. MRI image for head). Let us assume an X-ray image for the human chest. This image will be processed by calculating its low-level descriptors. After that, the results will be compared with previously calculated descriptors that are stored in the index database.

The indexed database (features data-file) was built by calculating all low-level descriptors of each image from the medical database; the ImageCLEFmed test collection created in [22] is used as a test dataset in this research. As described earlier, this dataset contains more than 66,000 images. A set of 5000 image have been used as a collection test dataset for the developed system. This dataset was selected from different collections in the ImageCLEFmed dataset. It includes diverse images for all the human body parts which are categorized into four main medical image types, which includes Pathology, Nuclear, Radiology, and Endoscopy. Examples of images are X-Ray image which is an example of Radiological images, MRI images, Ultrasound images, and Mammogram images.

Initially, the indexed database was created for all 5000 medical images by extracting the developed features for each image. When the user initializes the search the query image is inserted to the system and the index of the image is calculated (index=list of low-level descriptors values). After that, the index of the query image is compared to other images indices. The similarity list after that is calculated and is sorted in a descending order. A number of (30-40) top images are retrieved. These images are supposed to be the most similar images to the input query. We assumed earlier to retrieve a set of 40 images although the full set of relevant images might be much more in the dataset. Studies in [22] showed that the searcher mostly seeks for "few good cases" that satisfy his/her searching issue, and that the user by human nature usually will not look for more than 30-50 resulting image.

2.5. System phases

The LFCBIRM system has three main phases. As shown in Figure 2. The first phase is the feature extraction for the full dataset and storing it in a defined repository. In the second phase features are compared against each other (input image features vs. stored image features). In the final phase, images are retrieved through the user interface to the end user.
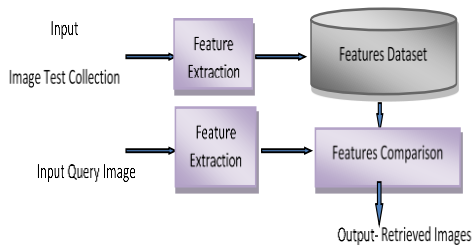
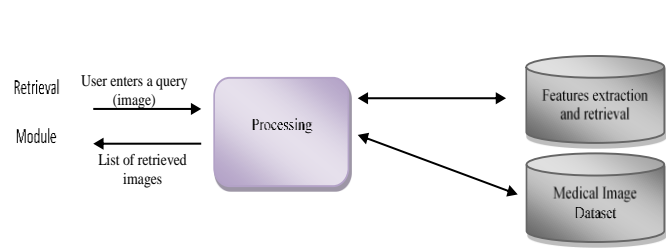Figure 1. The proposed system framewor   Figure 2. The proposed system phases

The system reads each image of the medical dataset, extracts the image features, and then stores them in a Matlab data-file. The data-file includes information about the image name, and the values of the remaining features. The extracted features are CSD, EHD, and CEDD descriptors. Table 2 shows an example on how the data is organized in the index (features) data-file, and the bit size of each descriptor.

Table 2. The features data-file

Image Name	Feature1: (CSD)	Feature 2: (EHD)	Feature 3: (CEDD)
(set of characters)	32 bits	150 histogram regions	166 histogram region
Image3445.jpg	double array	double array	double array

2.6. Low-level feature descriptors for the system

For CSD descriptor, M is assumed to be 32, this means that 32 colors are chosen to count the number of positions that contains the color cm. using only 32 colors is suitable for medical images, since most of them are either radiological, or homogeneous especially in the pathology images. 64 and 128 colors have been selected, and tests of using each one have been done, the retrieval results have changed significantly. Using 32 colors gives the best results among others. Three low-level descriptors, CEDD, CSD, and EHD, are selected as the feature extraction methods. No other modifications on these descriptors have been done. The system after that has been evaluated against LIRe using the same standard medical data set to measure and compare their performances.

2.7. System implementation

To build the system, a Matlab based application has been developed. Four low-level image descriptors have been implemented. Three of them were used and one was left because it's a time consuming descriptor compared to other descriptors. The achieved results using the three descriptors are good as the results prove. The implemented low-level descriptors are CLD, CEDD, DCD and EHD. The system was evaluated against LIRe using the same standard medical dataset to compare the accuracy rate between them.

2.8. Fusion-based similarity matching

The similarity between an inserted query image (Iq) and target image (Ij) is described as:

$$\text{Sim}(Iq, Ij) = \sum_F w^F \, \text{SimF}(Iq, Ij)$$

where F is one if the extracted features for the image, $F \in \{EHD; CLD; CSD\}$, and $\square F$ are the weights within the different image representations. These weights are assumed to be 1, but in this research EHD descriptor has been given twice importance than the other descriptors, so the factor 2 for EHD is used.

The similarity between the extracted features for the inserted image is compared against all the features of data-file values. Then the similarity list will contain all the differences between the query image and all other images. This list is sorted, so the top results are retrieved as the most matching images.

3. RESULTS AND ANALYSIS

Some sample runs have been performed in order to explain the details of the implementation and the description of this system. The input query is a frontal chest X-ray image as shown in Figure 3, and the resulting outputs using 40 retrieved images in Figure 4.
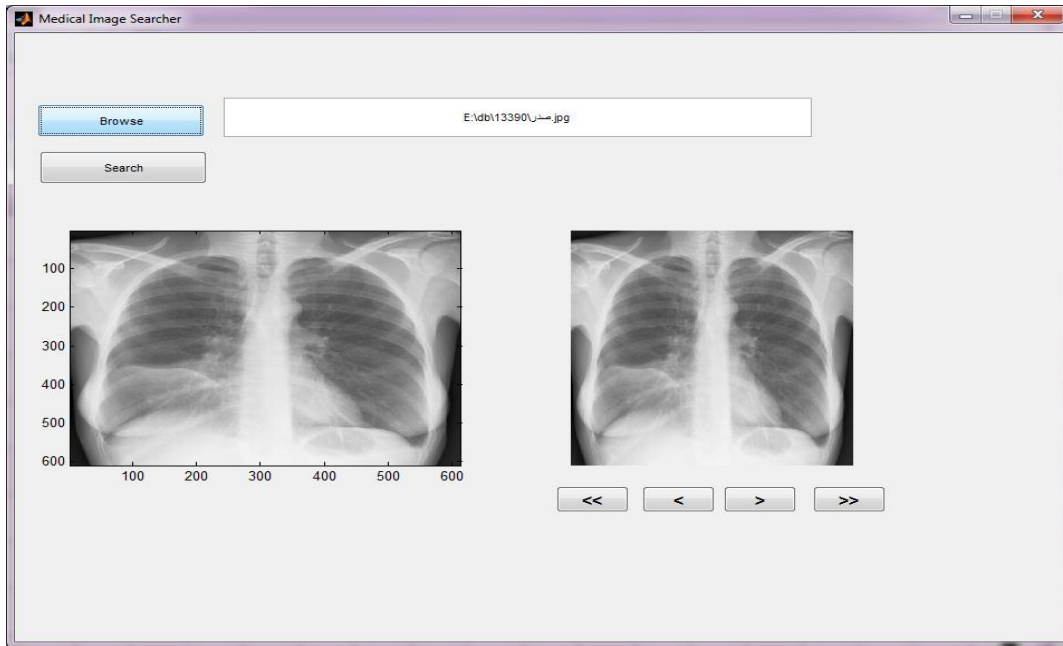
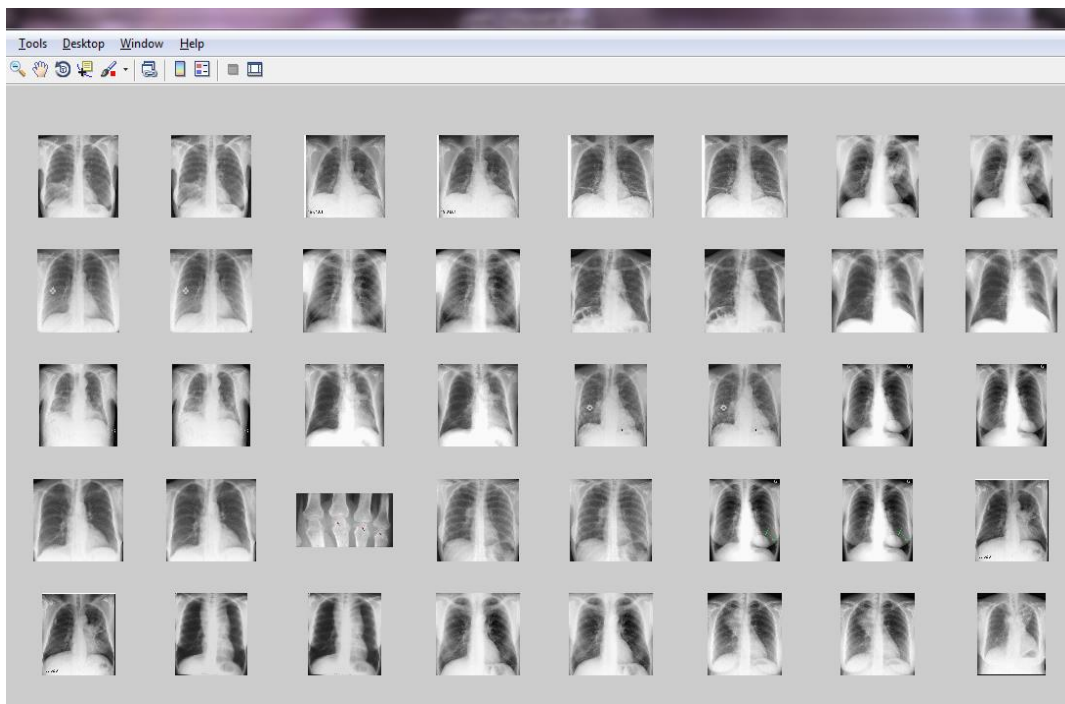Figure 3. Input query image is X-ray for chest

Figure 4. Retrieval of 40 images for the input query in Figure 3

3.1. Results and discussion

Measuring the performance of this system and any other developed system must be carried out to judge them. To define the performance measurement of a new developed system it must be compared to other similar existing systems within the field. The developed LFCBIRM system is compared to LIRe. An indexing file for the same test collection dataset was created for both systems. Then, a number of 50 test cases for the two systems were applied. The measurements, such as recall, precision, F-measure function, and fallout were calculated to define the performance enhancement. Measurements such as recall and precision are widely used in measuring Information Retrieval (IR) systems [23, 24].

Precision is used to measure the accuracy of the search process [25], it can be defined as the ratio between the number of relevant retrieved images to the total number of retrieved images. The equation of precision is showed below.

$$Precision = \frac{number\ of\ relevant\ items\ retrieved}{total\ number\ of\ items\ retrieved}$$

Precision directly evaluates the correlation of the query image to the test collection, and indirectly evaluates the completeness of the feature extraction algorithm [25]. The value of precision is between [0.1-1]. Precision value equals 1 (or 100%) when every image retrieved to the user is relevant [26].

Recall is used to measure the ability of the developed system to retrieve all the related items in the test collection [25]. Recall can be viewed as the probability that a retrieved image is relevant [25]. It can also be defined as the ratio between the number of relevant retrieved images to the total number of the relevant images in the collection dataset. The equation of precision is showed below.

$$Recall = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ relevant\ items\ in\ collectio}$$

Recall has also a value that is between, [0.1-1.0]. Recall value is 1.0 (or 100%) when every relevant image in the test collection is retrieved in the test case [26, 27]. F-measure is defined as a harmonic mean of recall (R) and precision (P) [28]. This measure has been introduced 20 years ago [16]. This measure was firstly introduced by C. J. van Rijsbergen in [29]. The F-measure combines recall and precision with an equal weight equation that is called F1 measure [29] F-measure equation is below:

$$F = \frac{2PR}{P + R}$$

Fallout measure is viewed as the inverse of recall [25]. It is defined as the ratio of the irrelevant retrieved images to the total irrelevant images in the test collection. This measure is defined in the following equation.

$$Fallout = \frac{number\ of\ irrelevant\ items\ retrieved}{number\ of\ irrelevant\ items\ in\ collection}$$

Experimental results

A sample of 50 images from different collections was selected. 50% of this collection was images not from the data set. Those images were downloaded from the internet, and the rest images were from the ImgCLEFmed collections. A comparison between the LFCBIRM system and LIRe was made. Using the same dataset which is a test collection of 5000 images that are selected from ImgCLEFmed, and retrieving 40 images every time for both systems.

The time to build the index with a laptop of processor i3 CPU and 3 GB RAM was 11 hours and 26 minutes, where it took about 4 hours to build the index with LIRe. However, the time to find the retrieved images was almost the same for both systems, which usually took about 3 seconds to 40 seconds in some cases. The time to get the results depends on the size of the images, the details in the images, and the colors of the images. Our observations showed that the colored images take longer time for the features to be extracted than the radiology (grey-level) images, this can be justified as the CSD descriptor for the colored images has more computations than the grey-level images.

The system has achieved more than 70% accuracy level using precision measure over all test cases. The next figures specify the results of the system performance. The measurement results of the developed LFCBIRM system are shown in Figure 5. The precision value is 71%, the recall value is 27%, and the F-function value is 39%. The reason why the recall value is small although the precision value is high will be specified in the next section.

The next Figure 6 compares the precision, recall, and the f-function results between systems. The LFCBIRM is the developed system in this research. Two LIRe tests have been made. The first system is LIRe using the EHD descriptor as the main descriptor and refines the results based on CEDD descriptor, it is specified as (LIRe 1) in the next figures. The second system is LIRe using CLD as the main descriptor and refining the results based on FCTH descriptor, it is specified as (LIRe 2) in the next figures. LIRe actually lacks for a combination between descriptors, each descriptor is a standalone, where in LFCBIM a combination of the three developed descriptors is made.
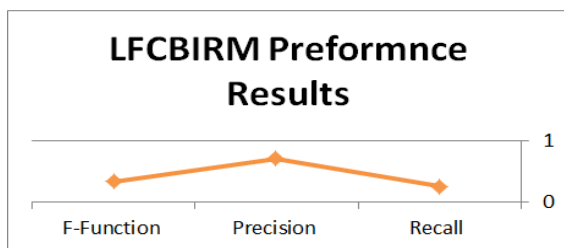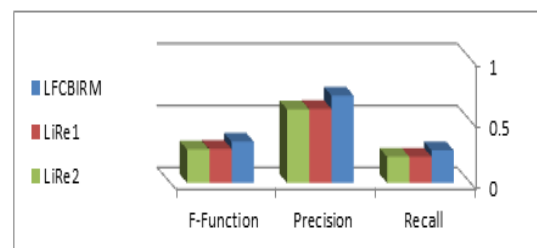
Figure 5. LFCBIRM performance results

Figure 6. Measurement comparisons
between systems

4. MAIN CONTRIBUTIONS

The proposed system provides the users with an accurate automated system to categories the input images. The resulted outputs are very similar based on the archived accuracy above. The proposed system is compared with the Lucene Image Retrieval system (i.e. LIRe). The comparison shows that the LFCBIRM system archives better result than the LIRe system. Moreover, the existing system has complicated APIs where the proposed one has a very simple and easy to use APIs. The future plan is to enhance the capabilities of the LFCBIRM system to a level that makes it a distinguished one in terms of the accuracy and its suitability for different kinds of users.

5. CONCLUSION AND FUTURE DIRECTION

In this content, building a medical CBIR system is introduced and implemented. This system can be built upon existing framework by making it specific to the medical applications by integrating and implementing four low-level descriptors to define similarity between images then to retrieve the most ranked images. The used descriptors provided good matching results between the images with more than 70% average precision. Thus, we can conclude that this system has gained good results in medical area as proved, and it promises for more future hopes for more systems and more ideas to be developed.

There are so many future directions which can be applied in the near future; especially that CBIR for medical research is still a hot new research area. The following is a brief list of directions in which researchers can peruse (pursue): Making clusters for the similar images. Simply a defined number of clusters can be defined using famous algorithms such as K-nearest neighbor algorithm (KNN).

Text-based with content-based images retrieval system can be developed. Each image is supported with an annotation file, using the description in these files to retrieve the related images within the same case. Synonyms must be considered as well. Building a system that considers the flipped or rotated images as relevant ones. Generating feedbacks from the users about the relevant images and the best image retrieving. The system can be expanded by developing more descriptors.

REFERENCES

[1]   Tom Weidong Cai, Jinman Kim, David Dagan Feng, "Content-Based Medical Image Retrieval," *Biomedical Information Technology*, 1st edition, 2008.
[2]   Yogita Mistry, D. T.Ingole, M. D. Ingole, "Content based image retrieval using hybrid features and various distance metric," *Journal of Electrical Systems and Information Technology*, pp. 1-4, Jan. 2017.
[3]   Chandan Singh, Kanwal Preet Kaur, "A fast and efficient image retrieval system based on color and texture features," *Journal of Visual Communication and Image Representation*, vol. 41, Nov. 2016.
[4]   K. Rajakumar, Dr.S. Muttan, "Medical image retrieval using modified DCT," in *Procedia Computer Science,* vol. 22, pp. 298-302, 2010.
[5]   Marcelo Costa Oliveiraa, Walfredo Cirne, Paulo M. de Azevedo Marques, "Towards Applying Content-Based Image Retrieval in the Clinical Routine," in *Future Generation Computer Systems*, vol. 23, no. 3, pp. 466–474, 2007.
[6]   Matthew Simpson, *et al.*, "Text- and Content-based Approaches to Image Retrieval for the ImageCLEFmed 2009 Medical Retrieval Track," *Aminer,* 2009.
[7]   A. Mojsilovic and B. Rogowitz, "Capturing image semantics with low-level descriptors," *Proceedings International Conference on Image Processing* (Cat. No.01CH37205), vol. 1, pp. 18-21, 2001
[8]   Mathias Lux, Savvas A. Chatzichristofis, "LIRe: Lucene Image Retrieval-An Extensible Java CBIRLibrary," in *Proceedings of the 16th International Conference on Multimedia,* 2008.
[9]   Julián Moreno-Schneider, Paloma Martínez, José L. Martínez-Fernández, "Combining heterogeneous sources in an interactive multimedia content retrieval model," *Expert Systems with Applications*, vol. 69, pp. 201-213, 2017.

[10] Laila H. Shoukry, "Visualizing MPEG-7, Color Layout and Edge, Histogram Descriptors for CBIR Systems," Institute for Visualization and Interactive Systems, Intelligent Systems Group, Stuttgart University, 2007.

[11] Savvas A. Chatzichristofis and Yiannis S. Boutalis, "CEDD: Color and Edge Directivity Descriptor. A Compact Descriptor for Image Indexing and Retrieval," *International conference on computer vision systems*, pp. 312-322, 2008.

[12] Jesús Chamorro-Martínez, *et al.*, "A fuzzy approach for modelling visual texture properties," *Information Sciences*, vol. 313, pp. 1-21, 2015.

[13] Yesubai Rubavathi Charlesa, Ravi Ramraj, "A novel local mesh color texture pattern for image retrieval system," *AEU - International Journal of Electronics and Communications*, vol. 70, no. 3, pp. 225-233, 2016.

[14] Ka-Man Wong, Lai-Man Po, and Kwok-Wai Cheung, "Dominant Color Structure Descriptor for Image Retrieval," *International Conference Image Processing*, vol. 6, 2007.

[15] William Hersh, MD, Henning Müller and Jayashree Kalpathy-Cramer, "The ImageCLEFmed Medical Image Retrieval Task Test Collection," in *Journal of Digital Imaging*, vol. 22, no. 6, pp. 648-655, 2008.

[16] Yutaka Sasaki, "The Truth of the F-measure, School of Computer Science," University of Manchester, 2007.

[17] Jayashree Kalpathy-Cramer, *et al.*, "Evaluating performance of biomedical image retrieval systems-An overview of the medical image retrieval task at ImageCLEF 2004-2013," *Computerized Medical Imaging and Graphics,* vol. 39, pp. 55-61, 2015.

[18] Buckley C, Voorhees EM: Retrieval system evaluation, "TREC: Experiment and Evaluation in Information Retrieval. Cambridge," In: *Voorhees EM, Harman DK Eds, MA: MIT*, pp. 53-75, 2005.

[19] Stefan Büttcher, Charles L. A. Clarke, Gordon V. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines," *MIT Press,* pp. 1-624, 2010.

[20] Xiaoyan Luo, Jun Zhang, Qionghai Dai, "A regional image fusion based on similarity characteristics," *Signal Processing*, vol. 92, no. 5, pp. 1268-1280, 2012.

[21] Dimitrios Markonis, Markus Holzer, Frederic Baroz, Rafael Luis Ruiz De Castaneda, Célia Boyer, Georg Langs, Henning Müller, "User-oriented evaluation of a medical image retrieval system for radiologists," *International Journal of Medical Informatics*, vol. 84, no. 10, pp. 774-783, 2015.

[22] William Hersh, MD, Henning Müller and Jayashree Kalpathy-Cramer, "The ImageCLEFmed Medical Image Retrieval Task Test Collection," in *Journal of Digital Imaging*, vol. 22, no. 6, pp. 648-655, 2008.

[23] Paul H. Lewis, *et al.*, "An integrated Content and Metadata Based Retrieval system for art," *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 1057-7149, 2004.

[24] Mohammed Shatnawi, Qusai Abuein, Anas Shatnawi, "A modified framework for labeling images based on human computation and robust technique," *International Conference on Computer, Electrical, and Systems Science, and Engineering,* 2011.

[25] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, "Introduction to Information Retrieval," *Cambridge University Press*, pp. 1-381, 2008.

[26] Al-Tamimi M., "Enabling Search for Mathematical Expressions," Dissertation, Washington: The George Washington University; 2005.

[27] M. Shatnawi and A. Youssef, "Equivalence detection using parse-tree normalization for math search," *2nd International Conference on Digital Information Management, Lyon*, pp. 643-648, 2007.

[28] Nancy Chinchor, "MUC-4 Evaluation Metrics," in *Proceeding of the Fourth Message Understanding Conference*, pp. 22-29, 1992.

[29] C. J. van Rijsbergen, "Information Retrieval," *Butterworths*, London, 1979.