State of Libya

Libyan Academy – Misurata Branch

Department of Information Technology

Development of Medical Image Retrieval Model based on Minimum Cross Entropy Thresholding

Thesis Submitted in Partial Fulfilment of the requirement for the Degree of Master in Information Technology

Prepared by

Hanaa Mansour Al-bory

Supervised by

Dr. Mohammed Elsheh

Academic year 2018

Declaration for thesis

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the Academy library open access institutional repository or allow the library to do so on my behalf, subject to Libyan Academy copyright legislation.

Hanaa Mansour Al-bory

Acknowledgments

First of all, Praise be to Allah and all thanking to Allah for his help in this work.

I would like to thank my family, especially Mom and my Uncle, for the continuous support they have given me throughout my time in graduate school; I could not have done it without them.

I would like to express my sincere regard and thanks to my supervisor **Dr.Mohammed Elsheh**, for his continuous support constant guidance and valuable comments. Without which, this thesis would not have been successfully completed.

My gratitude also goes out to all my teachers for their support and assistance،

and for all utilizations provided by the Libyan Academy – Misurata branch.

My thanks also go to my friends for all the comfort, and guidance they have given me during my stay here at the Academy.

Hanaa Mansour Al-bory

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
CBIR	Content Based Image Retrieval
TBIR	Text Based Image Retrieval
SBIR	Semantic Based Image Retrieval
ECM	Eigen Co-occurrence Matrix
GLCM	Gray Level Co-occurrence Matrix
GUIDE	Graphical User Interface Design Environment

الملخـــــص

في هـذه الاطروحـة تم استخـدام احدى طرق كشف الحـواف الحديثـة فـي استرجـاع الصـور الطبيـة. كشف الحـــواف يعطـي القيمـــة الأساسيـــة فـي الصــــور ويهمـل باقـــي القيــم. هـذه الطريقـة تعتمد علـى خوارزميـة minimum cross entropy thresholding فـي كشف حافـة الصورة الطبية ، وذلك لما لها من قدرة عالية في كشف الحواف مقارنة بالطرق التقليديـة، حتـى فـي حالة الصور المشوشة كانت نتائجها دقيقة وسريعة.

تم استخدام البيانات التي تتمثل في مجموعة صور اشعة صينية تم جمعها من قسم الاشعة في مستشفى مصراته المركزي، هذه الصور تحتوي على مجموعة من التشويش كما أنها ذات دقة منخفضة لكي يتم اختبار النظام عن طريقها بصورة دقيقة. تم تصنيف هذه الصور إلى خمسة مجموعات (القفص الصدري، الركبة، الحوض، اليد، القدم) وكل مجموعة لا تقل عن ستة صور بحيث تكون لدينا قاعدة بيانات تحتوي على سبعين صورة طبية.

أيضا تم استخدام مجموعة من مقاييس التشابه لعملية الاسترجاع، كما تم قياس مدى كفاءة ودقة الخوارزمية في عملية استرجاع الصور الطبية باستخدام معدلات الدقة precision والاستدعاء recall، بالإضافة إلى دالة قياس زمن التنفيذ. وكانت النتائج المتحصل عليها جيدة ومرضية مقارنة بإحدى طرق كشف الحواف التقليدية (canny) في عملية الاسترجاع.

Abstract

This thesis is based on one of the modern methods of edges detection for medical images retrieval. The edge detection gives the useful image information and neglects the other data of the image. This method is based on Minimum Cross Entropy Thresholding in the medical image edge detection, because it has a higher edge detection capability than other conventional methods, it also in the noise images the results were accurate and fast.

X-ray images were used as the main input for the proposed study. These images were collected from the Radiology Department at Misurata Central Hospital. Since sample images came with some noise and low resolution, the accuracy of the system outputs will be challenge. These images were divided into five groups (chest, knee, pelvis, hand, foot), and each group of not less than six images. As a total, the used database accommodated seventy medical images.

Some similarity measurements were used for the retrieval process. The efficiency and accuracy of the algorithm was measured in the process of retrieval medical images using precision and recall rates, as well as the time measurement function. The results obtained from the development system shown good results interms of accuracy and retrieving speed compared to canny edge detection method in the retrieval process.

Chapter 1: Introduction

1.1 Introduction

The volume of information and medical images for individuals are growing very rapidally in recent years. These images are subject to heavy analysis in order to extract available information from them. As a result, the need to creative methods to archive and access these images motivates many researches to tackle this issue.

As that search in conventional databases only allow for textual search on metadata . Often, the database only includes references to the image data, which are stored on the file system as individual files[1]. Since the images are stored according to patient names, it makes searching on a special organ or ailing is very difficult task. To overcome this issue, mining and retrieving the information based on the images content has become an urgent necessity.

However, interpreting of a medical image goes through three steps: the first step is identifying contents of the image. The second step is the interpretation of these contents and in the third step adducing recommendation to medical administration is provided [2]. Moreover, for doctors to do these steps, they need to look in images similar to the case of their patient's injury from the database of medical images for comparison and conclusion. In case the database was very huge, as the case today, the need for an approach to help in retrieving images from the database similar to a specific picture (the query image) is raised. This process should be done timely and systematically to help in images interpretation. To accomplish this task, a massive database management for medical images processing is needed [2].

However, image mining is defined as searching and finding valuable information and knowledge in huge files of data [3]. In addition, the main goals of data mining are fast data/information retrieving and discovering knowledge from the databases [4]. Clearly, the image mining requires that images are retrieved based on several requirement specifications such as visual image features[5]. The study in [5] focuses on query image contents which considered a very important factor in retrieving the desired images.

The CBIR (Content Based Image Retrieval) system is based on the quality of the extracted features. Images edges are regions with strong intensity than other regions. Edges are a jump in intensity from one pixel to the next. In addition, using the edge detecting of an image can reduce the amount of data and filters out unuseful information. As a result, it preserves the important structural properties in an image [6]. However, entropy is used to distinguish the texture of the input image, its value and does not affected by rotation nor scaling [7].

1.2 Problem Definition:

Image retrieval approaches suffer from several issues that need more studies and investigation to enhance the way of image is retrieved.

TBIR (Text Based Image Retrieval) does not compare the visual content of images. This limits the access of the valuable information of images maintained in the database [8].

On other hand, CBIR uses the SBIR (Semantic Based Image Retrieval) to improve the image retrieval systems. However, its description of lesion does not correlate in the images with collection of elements[9].

Moreover, canny edge detection method for medical image retrieval removes the noise before it applied. Since images in the binary form might have several noise, this causes some problems in the edge detection of image [6]. Consequently, applying this algorithm may cause loss of information and increase the consumed time.

1.3 Thesis Statement

Evaluation of using minimum cross entropy thresholding in development of medical images retrieval model.

1.4 Aims and Objectives

This study aims to enhance the way in which medical images are retrieved based on edge detection using minimum cross entropy thresholding. This aim will be achieved by considering the following objectives to:

1. explore the area of image retrieval approaches based on edge detection and entropy.
2. develop an approach for medical image retrieval based on edge detection.
3. evaluate the proposed approach by comparing the obtained results with traditional methods such as (canny edge detection), in terms of accuracy, and the response time.

1.5 Outlines

The reminder of this thesis is constructed as:

Chapter 2: examines the current and past literature related to the image retrieval and edge detection approaches.

Chapter 3: introduces the edge detection algorithm based on minimum cross entropy thresholding for medical images edge detection.

Chapter 4: explains the development of the used model in details to clarify how the edge detection based on minimum cross entropy thresholding is applied for the medical images retrieval. Also, it presents findings and description of results.

Chapter 5: presents the conclusion, and future work.

Chapter 2: Literature Survey

2.1 Introduction

Image retrieval is one of the most exciting and fastest growing research areas in the field of computer science. Moreover, image retrieval by CBIR has become one of the most active research areas since the 1990s[10]. During the past decade, remarkable progress has been made in both theoretical research and system development regarding to image retrieval by entropy and edge detection. However, there are many challenging research problems still continue to attract researchers from multiple disciplines, for instance the medical image retrieval systems that researchers seek to integrate it with computer diagnostics systems. Many methods based on different features were proposed for this purpose.

2.2 Detection and Retrieval Methods

Some of the most remarkable methods for image detecting and retrieving are discussed as follows:

2.2.1 Bit Plane Slicing Method

Bit plane slicing is a method used to cleavage images into binary planes. It represents the intensity of each image pixel by each bit in a binary plane. The digital image is splited into eight bit planes for analyzing each bit plane in an image as shown in Figure (2-1). The bits which are present in the bit plane zero is the least significant bit, whereas the bits which are present in the bit plane seven are the most significant bits [11].

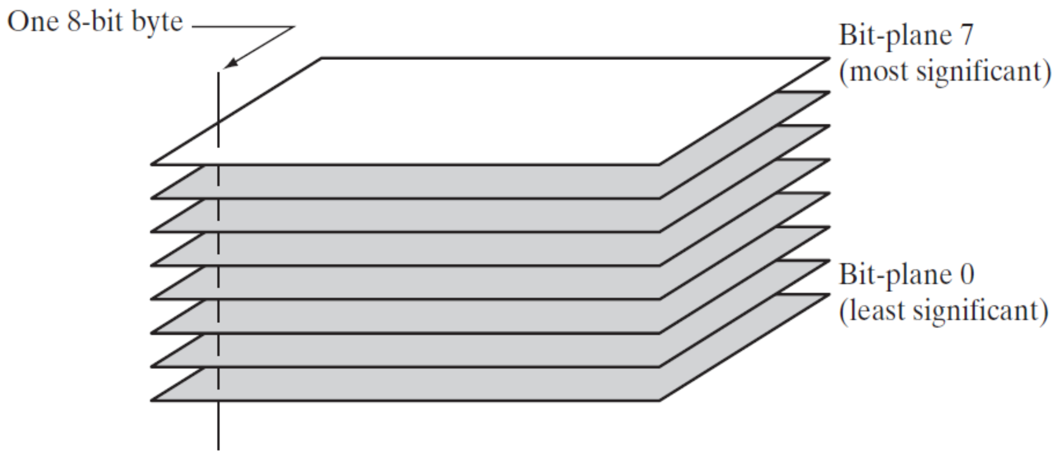


Figure (2-1) Bit-plane representation of an 8-bit digital image [12]

2.2.2 ECM (Eigen Co-occurrence Matrix) method

The ECM is a method for extraction a novel feature. Additionally, it represents features from within an event sequence in terms of a causal relationship between events. It suggests that the causal relationship becomes stronger when the distance between two events is closer. Similarly, causal relationship becomes stronger when an event-pair appears more frequently[13].

2.2.3 Statistical Moment Invariants

Region moment is interprets a function normalized gray level image as a probability intensity of a 2D random variable. This random variable properties can be described using statistical characteristics moments, and the moment is based on translation, scaling, and even on gray level transformations[7].

2.2.4 Edge Detection Method

The edge detection method is used to take edges from the images. In addition, it relies on an algorithm of sub divided images into the components and the objects of the image [14].

The important steps for the edge detection method are:

- **Detection:**

It concerns with determination of the pixels which should be discarded as a noise, and detection of the edge pixels which should be retained.

- **Localization:**

It deals with the determination of the exact location of an edge that is in the closer edge pixels.

- **Number of results:**

It applies a filter to enhance the quality of real edge and it should not result in more than one detected edge [14].

2.2.5 Canny Edge Detection Method

The canny edge detection is an edge detection method. Its optimal determination of edges is based on some norms. Finding edges by minimizing errors made detected edges close to the real edges. This means maximize location and determine edges only once when a single edge exists for minimal response [6]. As Figure (2-2) illustrates, this method runs in five sequence steps.

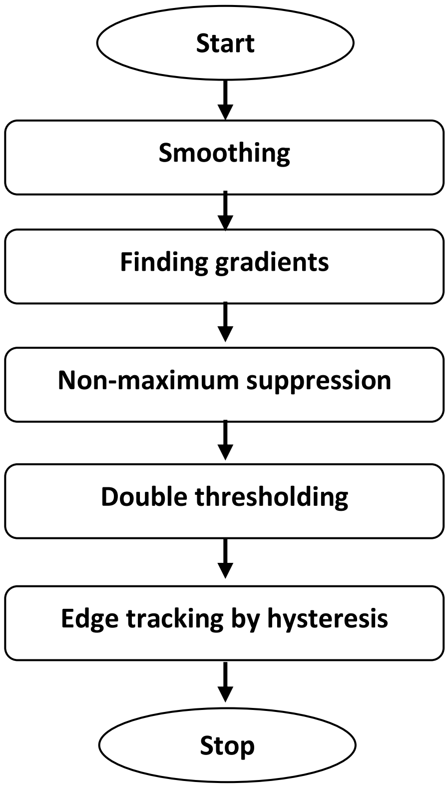


Figure (2-2) canny edge detection method diagram

The mentioned steps are described as follow:

- Smoothing: this step is blurring the image for noise removing.
- Finding gradients: this phase is finding gradients that is marked gradients when the images have large magnitudes.
- Non-maximum suppression: in this stage the edge maxima only should be marked.
- Double thresholding: this is convenes with of determination the potential edges by thresholding.
- Edge tracking by hysteresis: this step is suppressing all edges to determine final edges that are not connected to a very certain edge [14].

2.3 Related Works

Many studies have been done during last few years regarding to image retrieval. However the most relevant studies for our approach are discussed in the reminder of this chapter.

2.3.1 Image Retrieval using Entropy

The authors in [15] presented an effective approach for content based image retrieval (CBIR). It presented four methods for calculating the feature vector of image. Retrieving the images from the database using the entropy of sub blocks of one, three, four and twelve bit plane as a feature vector. Using threshold bit plane one for interband average image and three for color image on other side using bit plane slicing upper four bit plane for gray scale image and twelve bit plane (upper four for each color plane) for color image entropy feature vector is calculated and compared with each other. To improve the precision and recall performance of each method entropy sub block of each type of bit plane calculate which performance is better than full bit plane. Using a similarity measurement for computing the distance between each database image and query image on feature vector to find set of images falling in the class of query image. In this approach performance of twelve bit plane entropy feature vector of color image shown good results.

2.3.2 Content Based Medical Image Retrieval with Texture Content Using GLCM (Gray Level Co-occurrence Matrix) and K-Means Clustering Algorithms

In [16], the researchers proposed a model for the Content Based Medical Image Retrieval System by using texture feature in calculating the GLCM. It used a various statistical measures in sake of increasing similarities between query image and database images for improving the retrieval performance. This model was developed successfully in an effective manner by achieving the targeted output. This image retrieval model is capable of retrieving images based on the texture feature of the image. It takes into account the preprocessing as a traditional segmentation process, feature extraction, classification and retrieval steps in order to construct an efficient retrieval tool. The model was designed with a flexible and consistent flow for easy understanding. By achieved the targeted output, the results shown that the system has been developed successfully and in effective manner.

2.3.3 Using Statistical Moment Invariants and Entropy in Image Retrieval

The researchers in [7] suggested a method for automatic image retrieval using moment invariants and image entropy. This method consists of two parts, which are region selection and shape matching. In the first part, the image is partitioned into a set of sub images by using a local diffusive segmentation method. In the second part, the image entropy is computed and used to narrow-down the search space. The later, the moment invariants of the image are matched, which are independent to translation, scale, rotation and contrast, to every sub image and to the given template. To this effect, this method uses image moments and entropy experimental results demonstrate that the purposed technique is efficient.

2.3.4 Image Based Retrieval Using Edge Detection Algorithm

In [14] the Content based image retrieval was done by using the edge detection algorithm called canny edge detection algorithm. It used the color and texture for retrieving the image. Because color is the important cue in extracting the information from images, color histograms are commonly used in content based retrieval systems and they are very useful. Also texture plays an important role, because all the region is spatially distributed on the image. That regions are taken into consideration for the gradients calculation in the canny edge detection algorithm. That regions were smoothed and further steps done for selected regions. This make the work easy, thus the texture played an important role in the image retrieval. So that the image retrieval was accurate and the correct image was retrieved, all the steps are played a major role in the implementation to give good result.

2.3.5 Canny Edge Detection Method for Medical Image Retrieval

The study in [6] introduced a shape representation and retrieval method using the canny edge detection operator to shape features of medical images. At first, the color image is read from a database that contains a collection of medical images. These images are converted to grayscale images. Then global thresholding value was applied in order to obtain binary images. After that, a median filter was used to remove noise. Finally, canny edge detection algorithm was applied to produce the final images. This algorithm is considered to be a very good candidate for image retrieval. This is due to high generalization performance of canny edge detection. The experimental results revealed that the method gained good results.

Chapter 3: Minimum Cross Entropy Thresholding & Edge Detection Algorithm

3.1 Introduction

The image edges provides important information about the image contents since they constitute boundaries between objects in the image. However, the images are prone to artifacts and noise . To obtain valuable information from edge detection technique an accurate method for edge detection is needed [17]. The reminder of this chapter introduces entropy and minimum cross entropy thresholding.

3.2 Entropy

Entropy is the information content gauge in probability distribution [18]. It separates information into two different regions. One is above a specific intensity threshold and the other is below it [19].

Another way of using the entropy is a randomness statistical measure which can be used to characterize the input image texture [20].

3.2.1 Entropy measurement

For a high dimensional, separate random variable (image) is formulated as $X= (x_1,\ldots,x_d) \in R^d$ that has a probability mass function of $p(x_1,\ldots , x_d)$[21], The entropy H of an image is defined as shown in equation (1):

$$H(X) = \sum_{x1}^{xd} p(x_1 \ldots\ldots\ldots\ldots x_d) log \frac{1}{p(x_1 \ldots\ldots x_d})\quad \rightarrow \quad (1)$$

In addition, The estimation of the underlying probability of pixel intensities is based on the normalized histogram[21], the probability of pixel intensities is formulated as equation (2).

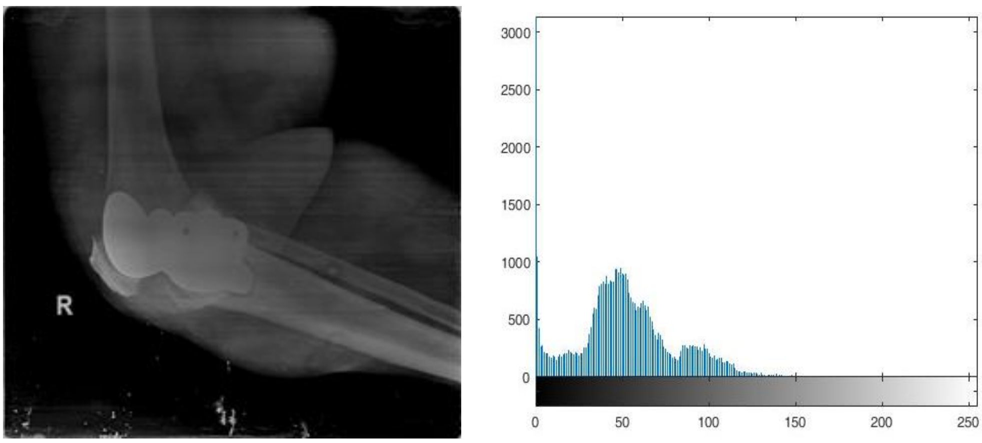
$$p(x_i) = h_I(x_i)/N \rightarrow (2)$$

where $h_I(x_i)$ denotes the histogram entry of intensity value x_i in image I and N is the total number of image pixels[21]. consequently, the entropy of the image can be computed as shown in equation (3):

$$H(I) = \sum_{x_i} h_I(x_i) log N / h_I(x_i) \quad \rightarrow (3)$$

3.2.2 Histogram

Histogram is the estimation of the probability distribution of a certain type of data [22], whereas an image histogram is a graphical representation of the number of image pixels as a function of their intensity [22]. Since image histograms are significant means for inspecting images, they allow to spot back ground and grey value range at a glance. Also, clipping and quantization noise in values of an image can be spotted immediately[22]. Figure (3-1'b') explains the histogram of an image displayed in Figure (3-1'a').



(a) **(b)**

Figure (3-1) An image and its histogram

3.2.2.1 Histogram Processing

Image histogram represents relative frequency of occurrence of several gray levels [23]. $P(x_i)$ gives an estimation of the probability of occurrence of gray level, it's given by the equation (2) [12].

3.2.3 Cross Entropy Thresholding

It measures distance between two distributions. This distance is a theoretic information. Cross entropy is used as a constraint for imposition the total intensity in the thresholded image. It is intended to be identical to the original image. This achieved by minimization of cross entropy between input image and output binary image. Since image thresholding is considered one of the methods used for separating objects from the background, the most widely used thresholding technique is depended on the grey level histogram[24]. Thresholding of an image accurately allows examining and analyzing the isolated regions in better details. Moreover, the thresholding is lesser rigorous computation than most of the current available methods, and it easier for implementation successfully. It has various applications including target detection in medical images and document analysis [19]. To clarify this aspect, let $F = \{ f_1 ,f_2, \dots , f_N \}$ and $G = \{ g_1 ,g_2 ,\dots, g_N \}$ are two probability distributions on the same set [17].

As reported in[17], the cross entropy between F and G is defined by equation (4):

$$D\ (F, G) = \sum_{i=1}^{N} f_i\ log\frac{f_i}{g_i} \quad \rightarrow (4)$$

3.2.3.1 Minimum Cross Entropy Thresholding

The minimum cross entropy thresholding algorithm selects several thresholds by minimizing the cross entropy between the original image and the resulting image [17]. To explain this aspect, let *I* be the original image and the corresponding histogram is *h(i), i=1,2,….,L* when L is being the number of gray levels. Then the resulting image, denoted by I_t using t as the threshold value is constructed by equation (5) as informed in [17]:

$$I_t\ (x, y) = \begin{cases} \mu\ (1,t) & I(x,y) < t \\ \mu\ (t,L+1) & I(x,y) \geq t \end{cases} \quad \rightarrow (5)$$

Where a mean of numbers from *a* to *b* is:

$$\mu\ (a, b) = \frac{\sum_{i=a}^{b-1} i\ h(i)}{\sum_{i=a}^{b-1} h(i)}$$

The cross entropy is then calculated by equation (6):

$$D\ (t) = \sum_{i=1}^{L} i\ h(i) \log(i) - \sum_{i=1}^{t-1} i\ h(i) loge\big(\mu(1,t)\big) - \sum_{i=t}^{L} i\ h(i) \log\big(\mu(t, L+1)\big) \quad \rightarrow (6)$$

As reported in [17] the Minimum Cross Entropy thresholding algorithm determines the optimal threshold by minimizing the cross entropy is formulated based on equation (7):

$$t^* = argmin\{D(t)\} \rightarrow (7)$$

3.3 Edge Detection using Minimum Cross Entropy Thresholding

It maybe defined a spatial filter mask as a matrix w of size $m \times n$. The spatial filtering process is consisting simply by moving a filter mask w of order $m \times n$ from point to point in an image. At each point (x,y), the response of the filter at that point is calculated a predefined relationship. Assume that $m=2a+1$ and $n=2b+1$, where a,b are nonnegative integers[17]. For this, smallest meaningful size of the mask is 3×3, as shown in Figure (3-2).

W(-1,-1)	*W(-1,0)*	*W(-1, 1)*
W(-0,-1)	*W(0,0)*	*W(0, 1)*
W(1,-1)	*W(1,0)*	*W(1, 1)*

Figure (3-2) Mask coefficients showing coordinate arrangement

f(x-1,y-1)	*f(x-1,y)*	*f(x-1, y+1)*
f(x,y-1)	*f(x,y)*	*f(x, y+1)*
f(x+1,y-1)	*f(x+1,y)*	*f(x+1, y+1)*

Figure (3-3) Image region under the mask

Image region under the above mask is shown in Figure (3-3). In order to detect an edge, mostly classified of all a pixels that satisfy the criterion of homogeneousness. After that, all the pixels on the borders between different homogeneous areas are detected. This scheme is preformed through creating a binary image by choosing a suitable threshold value using minimum cross entropy. Window is applied on the binary image, to find the edge pixels[17]. Set all window coefficients equal to 1 except center, center equal to x as shown in Figure (3-4).

1	1	1
1	X	1
1	1	1

Figure (3-4) Window coefficients

After that move the window on the whole binary image and find the probability of each central pixel of image under the window. Then, the entropy of each central pixel (CP) of image under the window is calculated by equation (8):

$$H\ (CP) = -p_c \ln(p_c) \quad \rightarrow (8)$$

Where, p_c is the probability of central pixel, and CP of binary image under the window. When the probability of central pixel $p_c = 1$ then the entropy of this pixel is zero. Thus, if the gray level of all pixels under the window homogeneous, then $p_c = 1$ and H = 0. In this case, the central pixel is not an edge pixel. Other possibilities of entropy of central pixel under window are shown in Table (3-1).

Table (3-1) p_C and H(CP) of central pixel under window

p_c	1/9	2/9	3/9	4/9	5/9	6/9	7/9	8/9
H	0.244	0.334	0.366	0.360	0.326	0.270	0.195	0.104

In cases p_c = 8/9, and p_c = 7/9, the diversity for gray level of pixels under the window is low. So, in these cases, central pixel is not an edge pixel. In remaining cases $p_c \leq 6/9$ the diversity for gray level of pixels under the window is high. So, for these cases, central pixel is an edge pixel. Thus, the central pixel with entropy greater than and equal to 0.244 is an edge pixel, otherwise is not. The Algorithm that shown in Figure (3-5) summarize this technique for calculating the optimal threshold values and the edge detector[17].

Algorithm: Edge Detection

1. **Input:** Image I of size $M \times N$ and t^{opt}, that has been calculated by Minimum Cross Entropy.
2. Create a binary image: For all x, y,
   if $I(x,y) \leq t^{opt}$ ***then*** $f(x,y) = 0$ ***else*** $f(x,y) = 1$.
3. Create a mask w of order $m \times n$, in our case ($m = 3, n = 3$)
4. Create an $M \times N$ output image g: For all x and y, Set $g(x,y) = f(x,y)$.
5. Checking for edge pixels:
   Calculate: $a = (m-1)/2$ and $b = (n-1)/2$.
   For all $y \in \{b+1, \dots, N-b\}$, and $x \in \{a+1, \dots, M-a\}$,
   $sum = 0$;
   For all $l \in \{-b, \dots, b\}$, and $j \in \{-a, \dots, a\}$,
   if $(f(x,y) = f(x+j, y+l))$ ***then*** $sum = sum + 1$.
   if $(sum > 6)$ ***then*** $g(x,y) = 0$ ***else*** $g(x,y) = 1$.
6. **Output**: The edge detection image g of I.

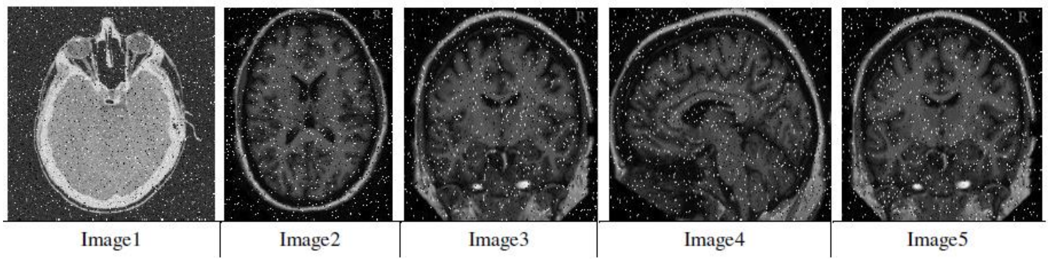
Figure (3-5) Edge Detection using Minimum Cross Entropy Thresholding

The edge detection using minimum cross entropy thresholding algorithm is applied as follows:

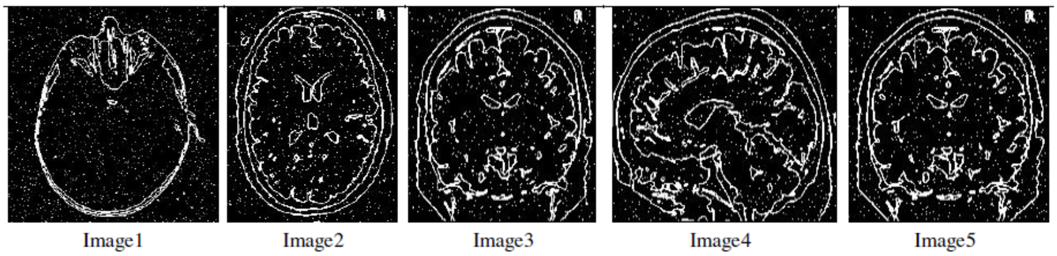
Stage 1: Find optimal threshold value (*t*) using Minimum Cross Entropy.

Stage 2: Applying Edge Detection Procedure with threshold value [17].

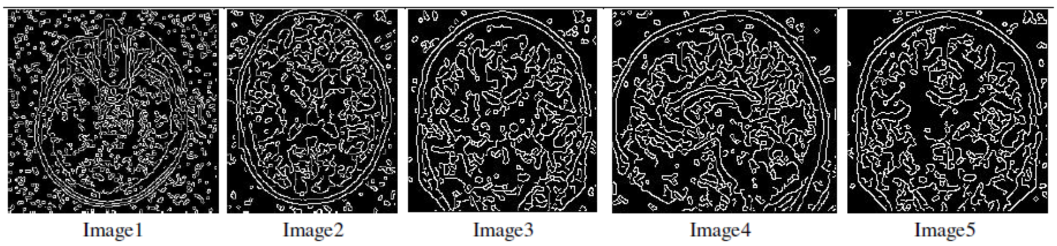
In [17] the authors proved that this algorithm is able to detect highest edge pixels in images, and decrease the computation time generating a high quality edge detection. Also it gives smooth and thin edges without distorting the shape of images as shown in Figure (3-6).



Original images with various noisy



minimum cross entropy thresholding method



Canny Method

Figure (3-6) Performance of edge detection using minimum cross entropy thresholding method

Chapter 4: System Implementation, Result and Discussion

4.1 Introduction

This thesis outlines a description of the development of medical image model retrieval based on minimum cross entropy thresholding. The minimum cross entropy thresholding algorithm used to edge detection of images for extracting the shape features of these images. Some similarity measures were used for the retrieval process.

The prototype system was built using a MATLAB R2013b software application on Intel® Pentium® CPU 2020M @ 2.40GHZ with 4 GB RAM. The structure of the final software application is illustrated to out the remainder of this chapter. Furthermore, the obtained result of its performance are illustrated by figures and charts.

The experiments were conducted on x-ray images which collected from department of radiology in Misurata central hospital.

4.2 Analysis and classification of data

A sample of seventy x-ray images were selected to test the proposed approach. Chosen images were divided into five categories (chest, foot, hand, knee, pelvic). Each group contains at least six images. Moreover, for serious testing the selected images came with some noise, and less resolution.

4.3 Detect edges of images

Once the images are collected and classified , edge detection algorithm was applied in order to extract images features. This process goes through several steps:

The first step, some pre-processes were applied such as convert the images to a fixed size (256x256), and same extension (bmp), which were in gray level type. In second step, the threshold (minimum cross entropy thresholding) was calculated for the x-ray images. In third step, the images were converted to binary by selected threshold. The image pixels less than threshold placed equal zero, whereas the image pixels larger than threshold equal one. In the final step, the images edges were detected as shown in Figure (4-1).

(a) original image (b) binary image (c) image edge

Figure (4-1) convert an image to binary and detect the edges

4.4 Design

After obtaining the features of the sample images by detecting their edges, and selecting the query image to get its features, the similarity was measured between features of database images and query image using several scales, the closest images to the tested image were retrieved.

Some pre-processes were applied on the selected images sample for testing the proposed method. This sample was saved in same path. The retrieval process calls and deals with the images easily via text database, which containing the images name and extension. Then, the images were called for detection their edges, also the query image was called from the same database for detecting its edges. In order to measure the similarity between the query image and the database images features, the retrieval process needs some similarity parameters. The proposed model of the work is illustrated in Figure (4-2).

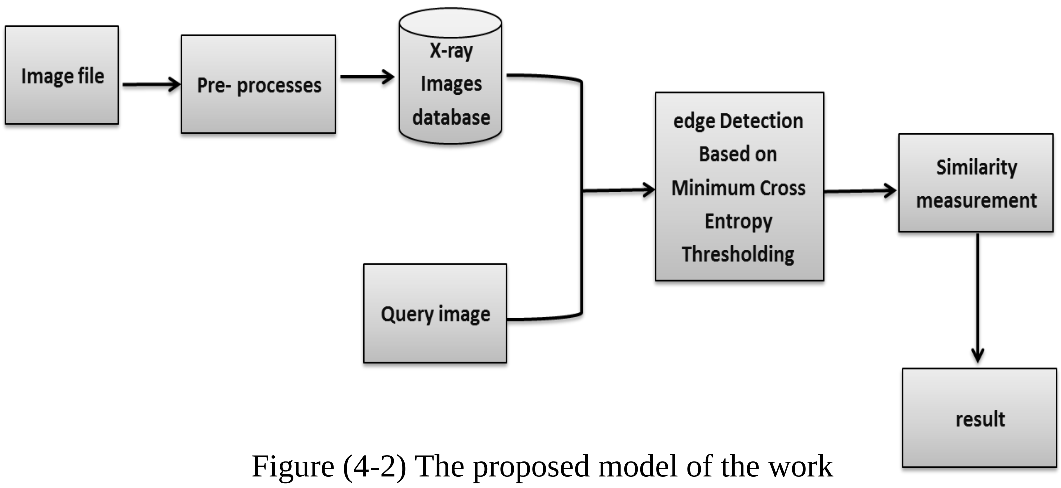


Figure (4-2) The proposed model of the work

4.5 Similarity Measurement

Estimation of image similarity is an important factor for retrieving images. Some similarity measures were used to retrieve the database images which are close from query image such as:

4.5.1 Correlation Coefficient (corr2)

The correlation coefficient between A and B, where A and B are matrices or vectors of the same size (m×n), where $\bar{A}$ is mean of A and $\bar{B}$ is mean of B, the correlation coefficient computed by equation (9):

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{[\sum_m \sum_n (A_{mn} - \bar{A})^2][\sum_m \sum_n (B_{mn} - \bar{B})^2]}} \quad \rightarrow (9)$$

In an equal case, it returned one, and the returned value is decreased as the difference increases[20].

4.5.2 Mean Squared Error (mse)

The mean-squared error (MSE) between the arrays X and Y is calculated by equation (10):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i)^2 \quad \rightarrow \quad (10)$$

Where X and Y can be any arrays of any dimension, but must be of the same size (n) . In an equal case, it returned zero, and the returned value is increased as the difference increases[20].

.4.5.3 Absolute Difference of two Images (imabsdiff)

It subtracts each element in array Y from the corresponding element in array X and returns the absolute difference in the corresponding element of the output array (Z) [20]. It is formulated by equation (11)

$$Z = |X - Y| \quad \rightarrow (11)$$

4.5.4 Image Entropy

It returns a scalar value representing the entropy of grayscale image. The returned value is formulated as equation (1). Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. This study make use of it to characterize the texture of the absolute difference array for two images. In an equal case, it returns zero, and the returned value is increased as the difference increases[20].

4.5.5 Structural Similarity Index (ssim)

It computes the Structural Similarity Index value for an image "x" using "y" as the reference image by equation (12):

$$Ssim(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \rightarrow (12)$$

Where μ_x average of *x*, μ_y average of *y*, σ_x^2 variance of *x*, σ_y^2 variance of *y*, σ_{xy} the covariance of *x* and *y*, $c_1=(0.01L)^2$, $c_2=(0.03L)^2$ are two variables to stabilize the division with weak denominator and *L* is dynamic range of the pixel-values. In an equal case, it returns one, and the returned value is decreased as the difference increases[20].

4.5.6 Norms (norm)

It returns the norm or maximum singular value of vector X with n elements by equation(13)

$$\|X\| = \sqrt{\sum_{k=1}^{n} |X_k|^2} \rightarrow (13)$$

This study make use of it on array of difference between two images. In an equal case, it returns zero, and the returned is value increased as the difference increases[20].

4.6 The Prototype System

To evaluate this model, a graphical user interface is needed. It was built using MATLAB GUIDE (Graphical User Interface Design Enviroment), and designed by using the layout tools provided by GUIDE as explained in Figure (4-3).

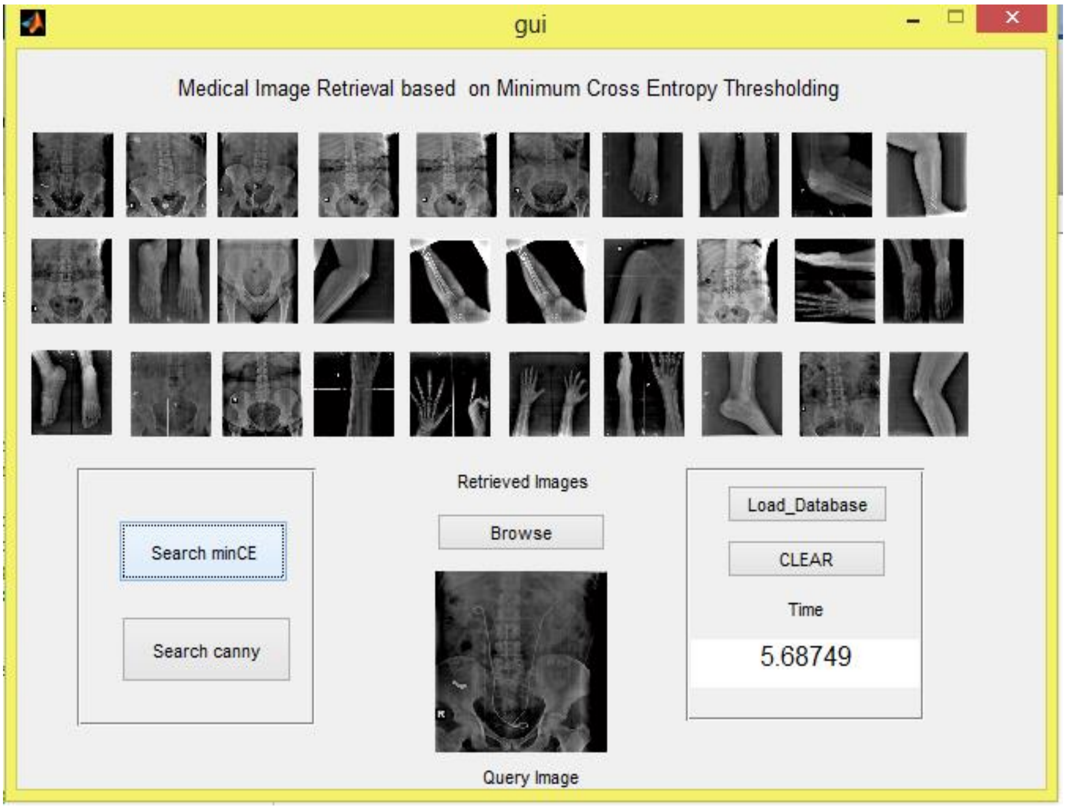


Figure (4-3) The user interface for medical image retrieval using minimum cross entropy thresholding

To demonstrate the prototype model, we need to run the MATLAB GUI and go through several steps:

For loading the medical database images, the user is asked to click on load database button, following successful loading, the confirming message will be displayed as show in Figure (4-4).

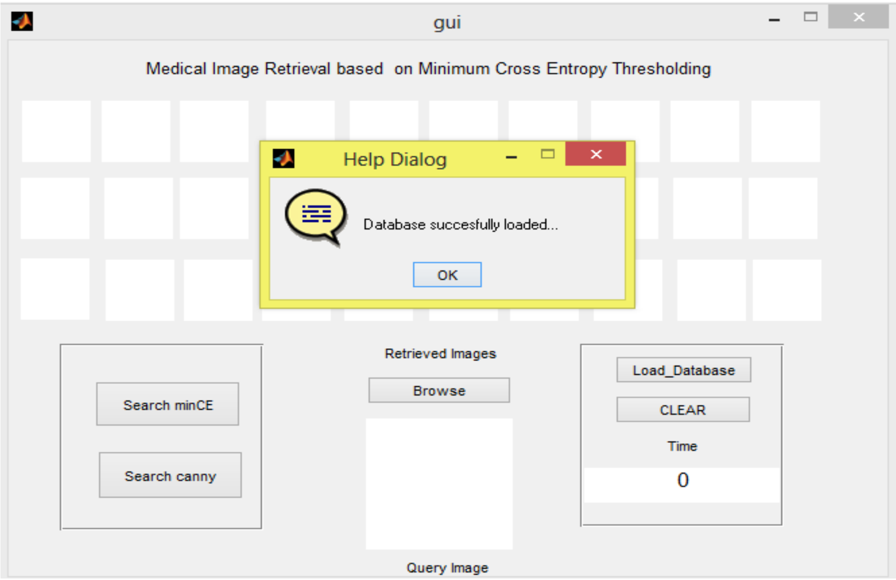


Figure (4-4) The load database confirming message

If the database was already loaded, a message will be displayed showing that the database is already loaded as show in Figure (4-5)

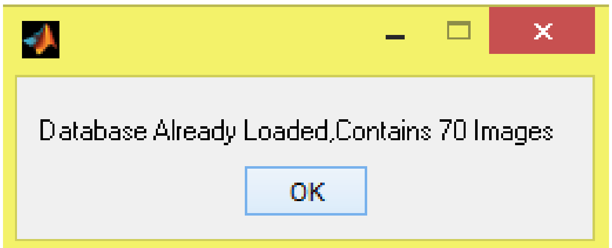


Figure (4-5) The database already loaded message

Also to delete the database, the user is asked to click on the clear button.

If the operation is don successfully, a message will be displayed as in Figure (4-6)

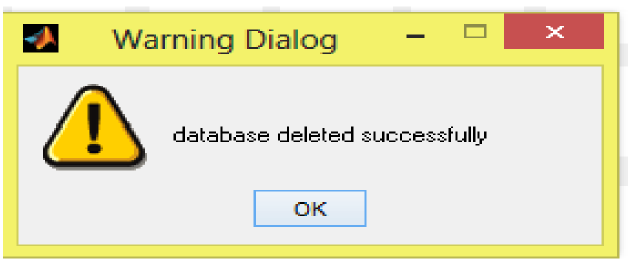


Figure (4-6) The delete database message

For retrieving process, the query image will be selected. If no image was selected, an error message will be displayed as show in Figure (4-7)

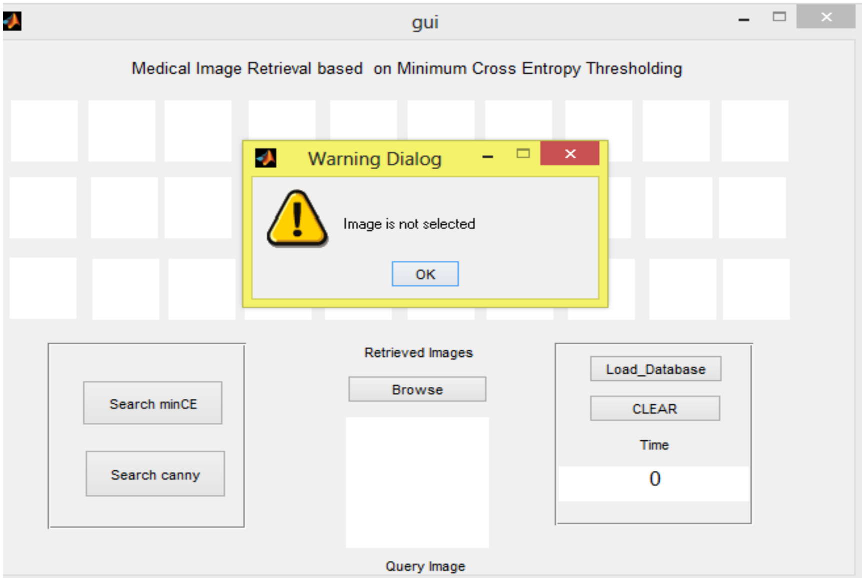


Figure (4-7) The error message of not selected query image

The prototype system offers two edge detection methods which are (Canny and minCE) for retrieving images purpose. Among seventy medical image that accommodated in the database, the prototype model retrieves thirty images. These images are viewed in GUI axeses. In addition the execution time of each retrieval method is displayed in GUI.

4.7 Results and Discussion

The proposed model was evaluated by comparing its results to one of the conventional edge detection model which is (Canny) in terms of the retrieval process. Some similarity measures were used to retrieving process.

The efficiency and execution time of the proposed algorithm was demonstrated by testing it over different noisy medical images and comparing it with traditional algorithm (Canny).

4.7.1 Retrieval Efficiency

For retrieval efficiency calculation, precision and recall values were calculated for randomly selected query images from seventy medical images database, these values were better when the number of relevant images retrieved were higher. The following benchmark equations (14)(15) have been used to calculate the precision values and recall respectively.

$$\text{Precision} = \frac{\text{No.of relevant images retrieved}}{\text{Total No.of images retrieved}} \rightarrow (14)$$

$$\text{Recall} = \frac{\text{No.of relevant images retrieved}}{\text{Total No.of relevant images in the Database}} \rightarrow (15)$$

4.7.1.1 Correlation Coefficient Similarity Measure (Corr2)

A comparison of the precision and recall values between canny and proposed methods using correlation coefficient similarity measure. It has been found that, the proposed method was better in (Knee, Hand, Foot) image categories, and it was similar in (chest) image category, however it was lower in (pelvic) image category comparing with Canny method. These values are shown in Tables (4-1) and (4-2).

Table (4-1) The precision by Corr2 similarity measure

Corr2					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.433333333	0.266666667	0.133333333	0.2	0.333333333
minCE	0.4	0.3	0.133333333	0.233333333	0.4

Table (4-2) The recall by Corr2 similarity measure

Corr2					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.764705882	0.666666667	0.666666667	0.545454545	0.769230769
MinCE	0.705882353	0.75	0.666666667	0.636363636	0.923076923

After, the precision and recall were calculated , the linear relationship between recall of Canny and proposed methods by Corr2 similarity measure for each images category was calculated as shown in Figure (4-8).

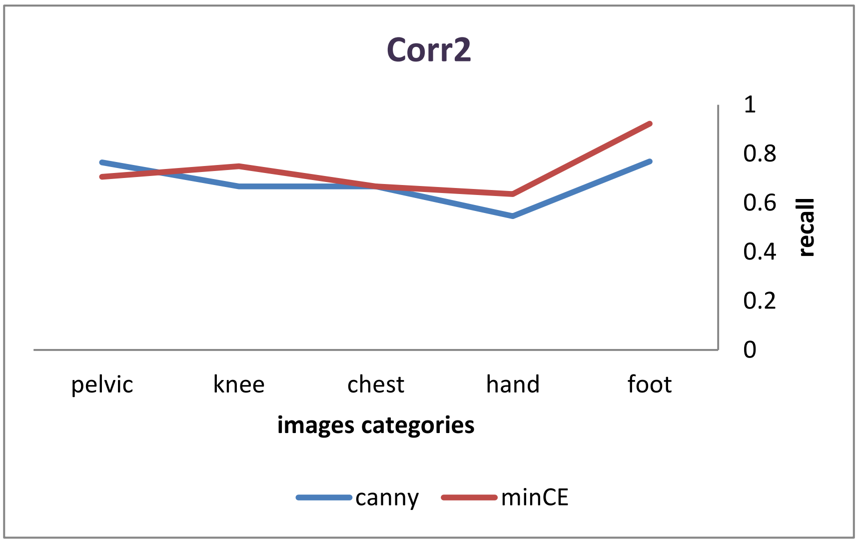


Figure (4-8) The linear relationship for recall by Corr2 similarity measure

4.7.1.2 Mean Squared Error Similarity Measure (Mse)

A comparison of the precision and recall values between canny and proposed methods by mean squared error similarity measure was conducted. It has been found that, the proposed method was better in (pelvic, Knee, chest) image categories, but it was lower in (hand, foot) image categories comparing with Canny method. These values are shown in Tables (4-3) and (4-4).

Table (4-3) The precision by Mse similarity measure

Mse					
	Pelvic	Knee	Chest	Hand	foot
Canny	0.033333333	0.3	0.1	0.233333333	0.233333333
minCE	0.266666667	0.4	0.133333333	0.066666667	0.133333333

Table (4-4) The recall by Mse similarity measure

Mse					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.058823529	0.75	0.5	0.636363636	0.538461538
MinCE	0.470588235	1	0.666666667	0.181818182	0.307692308

The linear relationship between recall of Canny and proposed methods by Mse similarity measure was calculated as shown in Figure (4-9).

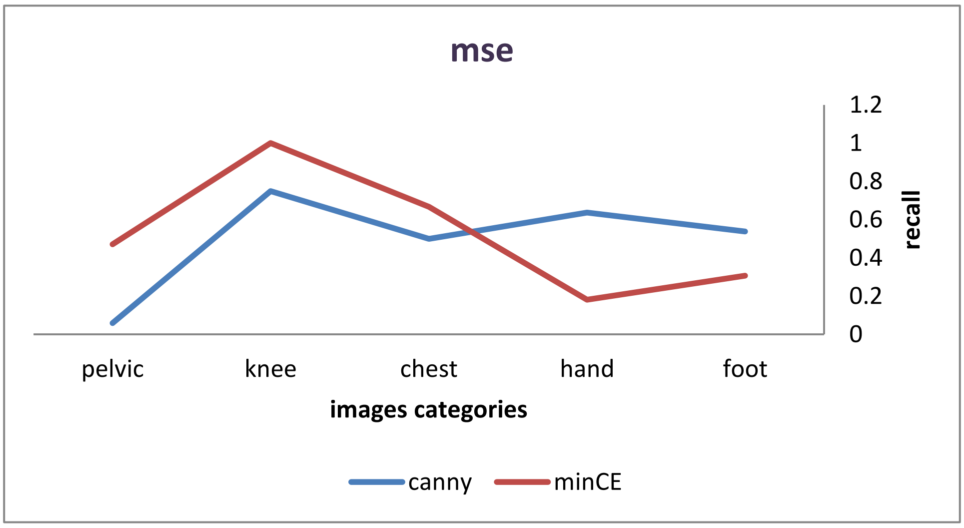


Figure (4-9) The linear relationship for recall by Mse similarity measure

4.7.1.3 Entropy Similarity Measure (Entropy)

A comparison of the precision and recall values between canny and proposed methods by Entropy similarity measure was carried out. It has been found that, the proposed method was better in (pelvic, Knee, chest) image categories. Nevertheless, it was lower in (hand, foot) image categories comparing with Canny method, the result is explained in Tables (4-5) and (4-6).

Table (4-5) The precision by Entropy similarity measure

Entropy					
	Pelvic	**Knee**	**Chest**	**Hand**	**foot**
Canny	0.033333333	0.333333333	0.1	0.233333333	0.233333333
minCE	0.266666667	0.4	0.133333333	0.066666667	0.133333333

Table (4-6) The recall by Entropy similarity measure

Entropy					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.058823529	0.833333333	0.5	0.636363636	0.538461538
MinCE	0.470588235	1	0.666666667	0.181818182	0.307692308

The linear relationship between recall of Canny and proposed methods by Entropy similarity measure was computed as shown in Figure (4-10).

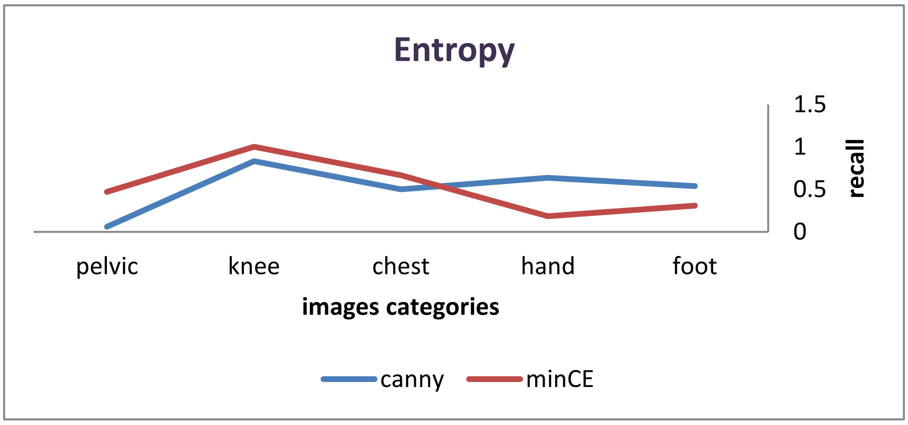


Figure (4-10) The linear relationship for recall by Entropy similarity measure

4.7.1.4 Structural Similarity Index Similarity Measure (Ssim)

A comparison of the precision and recall values between canny and proposed methods by structural similarity index similarity measure was done, it has been found that, the proposed method was better in (pelvic, Knee, chest) image categories. On other hand, it was lower in (hand, foot) image categories comparing with Canny method, the result is explained in Tables (4-7) and (4-8).

Table (4-7) The precision by Ssim similarity measure

Ssim					
	Pelvic	**Knee**	**Chest**	**Hand**	**foot**
Canny	0.066666667	0.333333333	0.066666667	0.2	0.233333333
MinCE	0.266666667	0.366666667	0.133333333	0.066666667	0.166666667

Table (4-8) The recall by Ssim similarity measure

Ssim					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.117647059	0.833333333	0.333333333	0.545454545	0.538461538
MinCE	0.470588235	0.916666667	0.666666667	0.181818182	0.384615385

The linear relationship between recall of Canny and proposed methods by Ssim similarity measure was determined as shown in Figure (4-11).

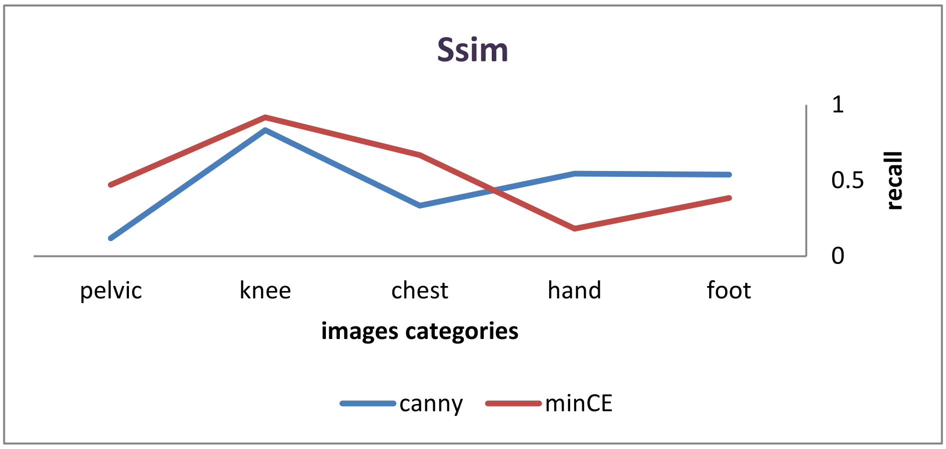


Figure (4-11) The linear relationship for recall by Ssim similarity measure

4.7.1.5 Norms Similarity Measure (Norm)

A comparison of the precision and recall values between canny and proposed methods by Norm similarity measure was performed. It has been found that, the proposed method was better in (Knee, Hand) image categories, it was similar in (Foot) image category, but it was lower in (pelvic, chest) image categories comparing with Canny method. These values are shown in Tables (4-9) and (4-10).

Table (4-9) The precision by Norm similarity measure

Norm					
	Pelvic	**Knee**	**Chest**	**Hand**	**foot**
Canny	0.533333333	0.266666667	0.2	0.066666667	0.333333333
MinCE	0.466666667	0.333333333	0.166666667	0.1	0.333333333

Table (4-10) The recall by Norm similarity measure

Norm					
	Pelvic	**Knee**	**Chest**	**Hand**	**Foot**
Canny	0.941176471	0.666666667	1	0.181818182	0.769230769
minCE	0.823529412	0.833333333	0.833333333	0.272727273	0.769230769

The linear relationship between recall of Canny and proposed methods by Norm similarity measure was calculated as shown in Figure (4-12).

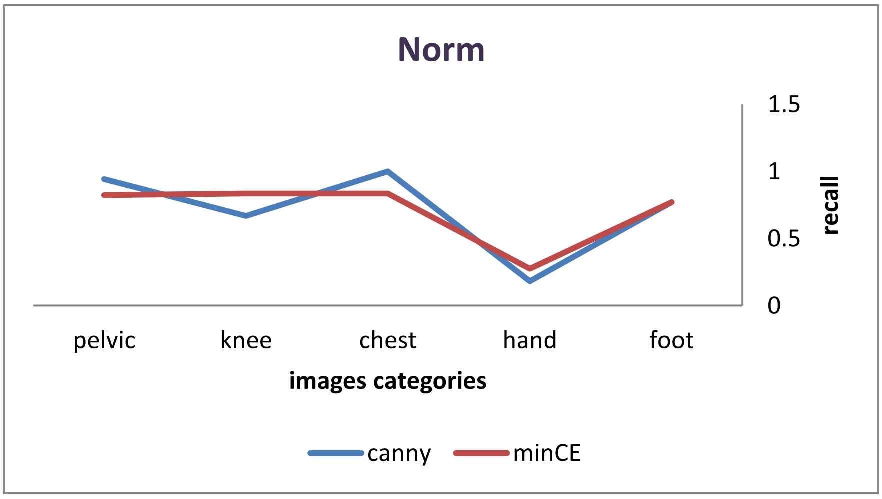


Figure (4-12) The linear relationship for recall by Norm similarity measure

4.7.1.6 Overall Average of Proposed Method

The overall average (precision / recall) of proposed method for image retrieval also computed in order to evaluate the proposed method performance for image retrieval. It has been found that, the proposed method show good results in terms of accuracy as Figure (4-13) demonstrates.

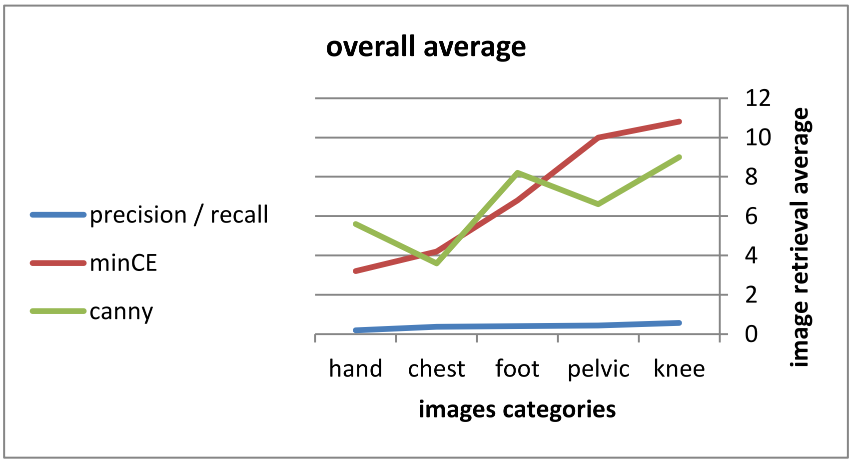


Figure (4-13) The linear relationship of overall average for image retrieval

4.7.2 Execution Time

There is another interesting finding, it has been observed that the proposed algorithm works effectively comparing to the run time of other method (Canny) with all similarity measures. Execution Time was measured by tic and toc command in matlab.

4.7.2.1 Correlation Coefficient

The prototype system execution time in second of canny and proposed methods using Corr2 similarity measure. The obtained result has been shown that, the proposed method was speeder in all image categories comparing with Canny method, the result is explained in Table (4-11).

Table (4-11)The execution time of the prototype system by Corr2 similarity measure.

Corr2-time (seconds)					
	Pelvic	**Knee**	**Chest**	**Hand**	**foot**
Canny	5.65455	5.65388	5.65409	5.65263	5.64723
MinCE	5.6469	5.6445	5.6439	5.6437	5.63904

Figure (4-14) show the linear relationship between proposed algorithm run time and the Canny algorithm run time using Corr2 similarity measure.

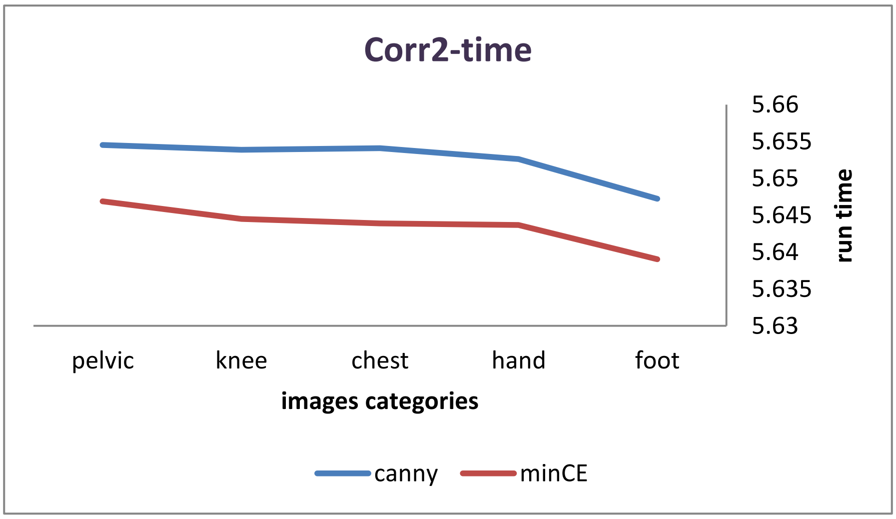


Figure (4-14) The linear relationship between proposed algorithm run time with the Canny algorithm run time using corr2 similarity measure

4.7.2.2 Mean Squared Error

The prototype system execution time in second of canny and proposed methods using Mse similarity measure. The obtained result has been shown that, the proposed method was speeder in (Knee, chest, Hand, Foot) image categories, but it was slower in (pelvic) image category comparing with Canny method, the result is described in Table (4-12).

Table (4-12)The execution time of the prototype system by Mse similarity measure.

Mse-time (seconds)					
	Pelvic	**Knee**	**Chest**	**Hand**	**foot**
Canny	5.4614	5.584	5.463	5.8234	5.943
MinCE	5.4947	5.583	5.457	5.68224	5.64554

Figure (4-15) show the linear relationship between proposed algorithm run time and the canny algorithm run time using Mse similarity measure.

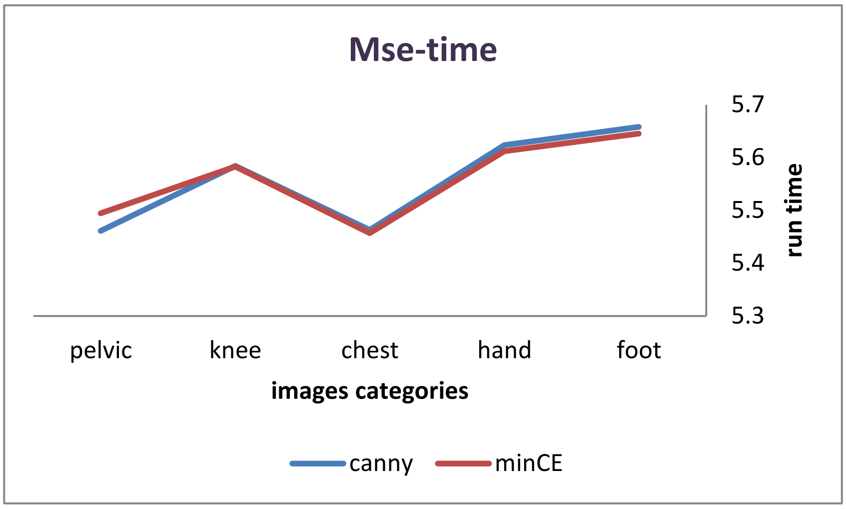


Figure (4-15) The linear relationship between proposed algorithm run time with the canny algorithm run time using Mse similarity measure

4.7.2.3 Structural Similarity Index

The prototype system execution time in second of canny and proposed methods using Ssim similarity measure. The obtained result has been shown that, the proposed method was speeder in (Knee, chest, Hand, Foot) image categories, but it was slower in (pelvic) image category comparing with Canny method, the result is explained in Table (4-13).

Table (4-13)The execution time of the prototype system by Ssim similarity measure.

Ssim-time (seconds)					
	Pelvic	**Knee**	**Chest**	**hand**	**foot**
Canny	6.184	6.314	6.087	6.211	6.024
MinCE	6.256	6.2127	6.0112	6.126	6.013

Figure (4-16) show the linear relationship between proposed algorithm run time and the canny algorithm run time using Ssim similarity measure.

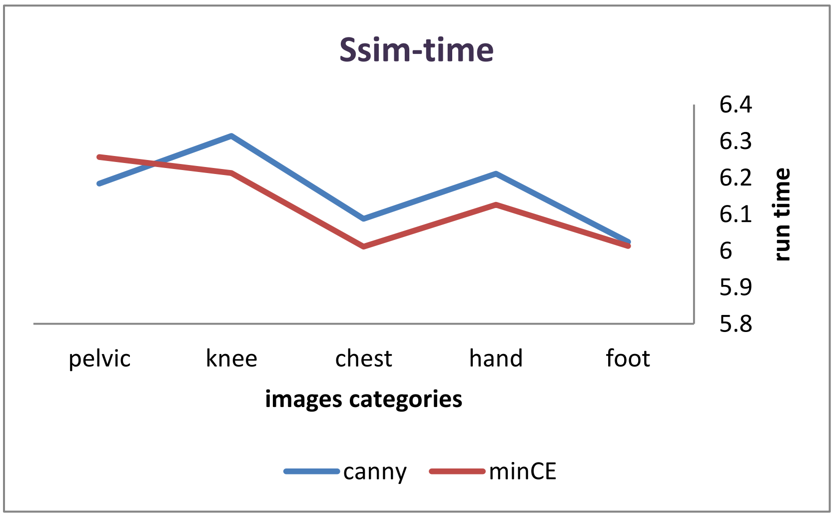


Figure (4-16) The linear relationship between proposed algorithm run time with the Canny algorithm run time using Ssim similarity measure

4.7.2.4 entropy

The prototype system execution time in second of Canny and proposed methods using entropy similarity measure. The obtained result has been shown that, the proposed method was speeder in (Pelvic, Knee, chest, Hand) image categories, but it was slower in (Foot) image category comparing with Canny method, the result is explained in Table (4-14).

Table (4-14) The execution time of the prototype system by entropy similarity measure.

Entropy-time (seconds)					
	Pelvic	**Knee**	**Chest**	**hand**	**foot**
Canny	5.65455	5.65388	5.65409	5.75263	5.63723
MinCE	5.6469	5.6445	5.6439	5.7437	5.63904

Figure (4-17), shows The linear relationship between proposed algorithm run time and the Canny algorithm run time using Entropy similarity measure.

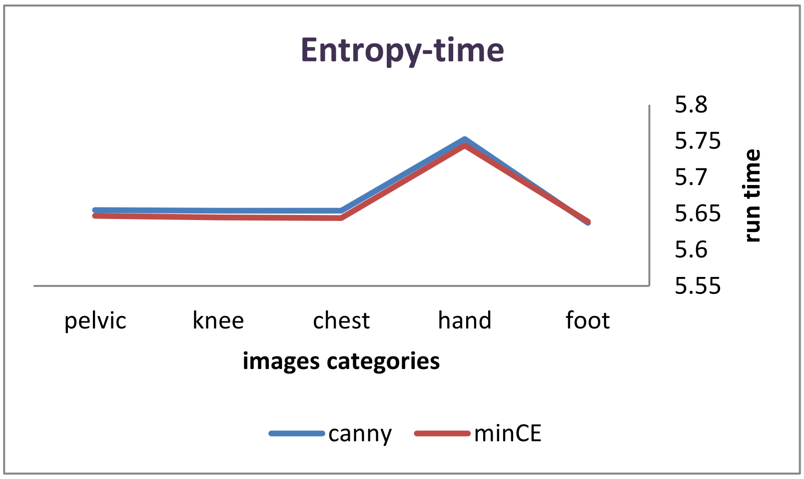


Figure (4-17) The linear relationship between proposed algorithm run time with the Canny algorithm run time using Entropy similarity measure

4.7.2.5 Norm

The prototype system execution time in second of Canny and proposed methods using Norm similarity measure. The acquired result has been explained that, the proposed method was speeder in (Pelvic, Knee, chest, Foot) image categories. However, it was slower in (Hand) image category comparing with Canny method, the result is explained in Table (4-15).

Table (4-15) The execution time of the prototype system by Norm similarity measure.

Norm-time (seconds)					
	Pelvic	**Knee**	**Chest**	**hand**	**Foot**
Canny	6.532	6.561	6.346	6.326	6.541
minCE	6.476	6.542	6.251	6.342	6.457

Figure (4-18), shows The linear relationship between proposed algorithm run time and the Canny algorithm run time using Norm similarity measure.

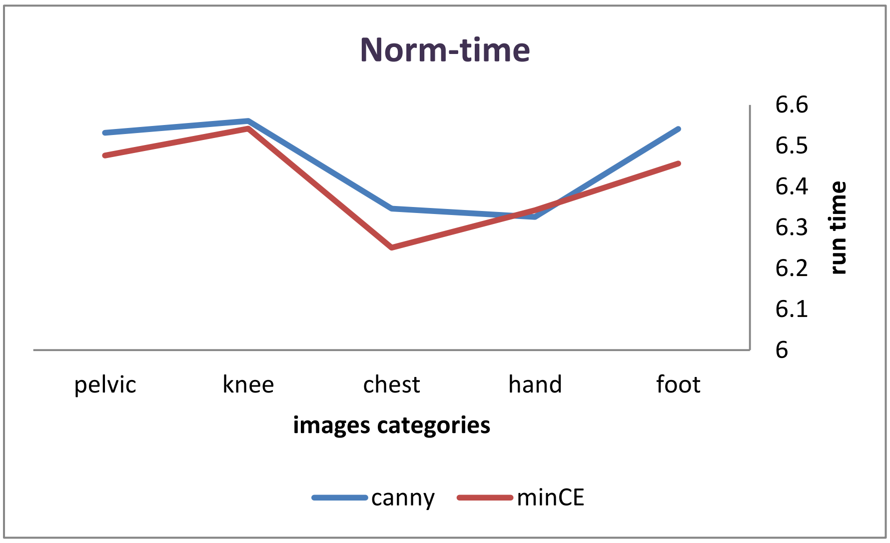


Figure (4-18) The linear relationship between proposed algorithm run time with the Canny algorithm run time using Norm similarity measure

Chapter 5: Conclusions and Future Work

5.1 Conclusions

The proposed method concerns with retrieving the medical images from database using the edge detection for extracting the shape features of these images. Edge detection was obtained by using the minimum cross entropy thresholding algorithm. Also some similarity measures were used for the retrieval process. In order to evaluate the proposed method performance, the precision and recall values were computed.

The linear relationship of recall between proposed method and the other method (Canny) shows good results for the proposed algorithm.

To find out minimum cross entropy thresholding algorithm performance for image retrieval, the overall average (precision / recall) value was computed. The overall average value demonstrates good results for the proposed algorithm performs.

It has been observed that the proposed algorithm works effectively comparing to the run time of the other method (Canny) with most similarity measures.

5.2 Future Work

Although the proposed model showed positive results in the medical image retrieval process where the basic features of the image were satisfied, more precise features such as fracture type or disease status were outside the focus of this study. It is recommended that the researchers improve the threshold selection in order to obtain more accurate features of the medical images.

References

[1] Lehmann, T. M., Güld, M. O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., & Wein, B. B. (2005). Automatic categorization of medical images for content-based retrieval and data mining. *Computerized Medical Imaging and Graphics*, *29*(2), 143-155.

[2] Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. *Journal of Digital Imaging*, *24*(2), 208-222.

[3] Nisia, T. G., & Rajesh, S. (2016). Survey on Selection of Features for Content Based Image Retrieval in Image Mining. *Journal of Data Mining and Management*, *1*(3).

[4] Chouhan, P., & Tiwari, M. (2016). Feature Extraction Techniques for Image Retrieval Using Data Mining and Image Processing Techniques. *IJARCCE*. Vol 5, Issue 5.

[5] Shukla, V. S., & Vala, J. A. (2016). Survey on Image Mining, its Techniques and Application. *International Journal of Computer Applications, Volume 133.*

[6] Jadwa. S. k. (2016). Canny Edge Detection Method for Medical Image Retrieval. *international Journal of Scientific Engineering and Applied Science Vol-2, Issue-8.*

[7] Amr, I. I., Amin, M., Kafrawy, P. E., & Sauber, A. M. (2010). Using Statistical Moment Invariants and Entropy in Image Retrieval. *International Journal of Computer Science and Information Security*, *Vol. 7*.

[8] Madugunki, M., Bormane, D. S., Bhadoria, S., & Dethe, C. G. (2011, April). Comparison of different CBIR techniques. *IEEE ,Vol 4.*

[9] Zhang, D., Islam, M. M., Lu, G., & Hou, J. (2009, December). Semantic image retrieval using region based inverted file. In *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.* (pp. 242-249). IEEE

[10] Joshi, F. (2013). Analysis of existing CBIR Systems: improvements and validation Using Color Features. *International journal of Emerging technology and advanced engineering*, *3*(5).

[11] Aarthi, R., & Kavitha, M. S. (2017). Image Encryption Using Binary Bit Plane And Rotation Method for an Image Security. *IJEDR.* Volume 5, Issue 2.

[12] Gonzalez, R. C, & Woods, R. E. (1992). Digital image processing, P(87,88).

[13] Oka, M., Oyama, Y., & Kato, K. (2004). Eigen co-occurrence matrix method for masquerade detection. *Publications of the Japan Society for Software Science and Technology*.

[14] Reddy, A., Kumar, K., & Ramyakrishna, I. (2015). Image Based Retrieval Using Edge Detection Algorithm, *International Journal & Magazine of Engineering, Technology, Management and Research, Volume 2, Issue 4* .

[15] Sai, N. S. T., & Patil, R. C. (2011). Image Retrieval using Entropy. *International Journal of Computer Applications*, *Volume 24*.

[16] Ramamurthy, B., & Chandran, K. R. (2012). Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms. *Journal of Computer Science*, *8*(7).

[17] Elaraby, A. E. A., El-Owny, H. B. M. A., Heshmat, M., & Rady, A. A. (2014). New Algorithm For Edge Detection in Medical Images Based on Minimum Cross Entropy Thresholding. *International Journal of Computer Science Issues (IJCSI)*, *11*(2).

[18] Al-Ajlan, A., & El-Zaart, A. (2009). Image segmentation using minimum cross-entropy thresholding. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on* (pp. 1776-1781).

[19] Alattas, R. (2014, October). Oil spill detection in SAR images using minimum cross-entropy thresholding. In *Image and Signal Processing (CISP), 2014 7th International Congress on* (pp. 709-713). IEEE.

[20] https://www.mathworks.com, Sep\ 2017.

[21] Sabuncu, M. R. (2004). Entropy-based image registration, *(Doctoral dissertation, Princeton University)*.

[22] Krutsch, R., & Tenorio, D. (2011). Histogram equalization. *Freescale Semiconductor, Document Number AN4318, Application Note*.

[23] Won, C. S., Park, D. K., & Park, S. J. (2002). Efficient use of MPEG-7 edge histogram descriptor. *ETRI journal*, *24*(1).

[24] AlSaeed, D. H., El-Zaart, A., & Bouridane, A. (2011, November). Minimum Cross Entropy Thresholding using Entropy-Li based on log-normal distribution for skin cancer images. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on* (pp. 426-430). IEEE.