International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Survey on Implementation of Radiological Image Retrieval in CBIR System

Nabin Newar[1], Sandeep Kumar.S[2]

[1]Department of Computer Science & Engineering, Mangalore Institute of Technology and Engineering Mangalore, Karnataka, India

[2]Assistant Professor, Dept. of Computer Science & Engineering, Mangalore Institute of Technology and Engineering Mangalore, Karnataka, India

Abstract: ***This article gives an overview of the retrieval of Medical images from the Health Care Database. This working model aimed on the concept of content-based image retrieval (CBIR) system. The main goal of this model is to retrieve relevant radiological medical images from the Health Care as given by query image. The model focuses on the two main methodologies – feature extraction and similarity measurement. Feature extraction and similarity measurement will help to identify the relevant images based on query image. Two Similarity measurement metrics is being used and compared to measure the vectors in order to retrieve the relevant images from the database.***

Keywords: CBIR; Euclidean distance metric; Feature Extraction; Manhattan distance metrics, Medical Images; Similarity Measurement

1. Introduction

Due In the 80's the first implementation of CBIR system was developed by IBM and was called as QBIC (Query Based Image Content), later CBIR system turn out to be one of the most fascinating topics in the world of computer prophecy [1]. The history of the textual information was followed in the medical field that was stored in the large database and that was not enough for medical staffs to achieve the task.

Later with the improvement in the medical science technology image information had become the most tremendous advancement in the field. In the today's world, the technology progresses in medical research and the rigorous use of digital multimedia information have produce data that has become the main wealth for medical diagnosis used by the experts. Hence, researcher aimed on the advance of tools to support medical diagnosis by taking the advantage of the multimedia information.

The scenario is based on the content-based image retrieval (CBIR) system. The CBIR system will retrieve the medical image and visualize the image similar to the give in the query image from the medical database and provides the precaution related information to that query image by using the feature extraction techniques like Color-Moments, the Gabor wavelet transform, Color-Autocorrelation, Haralick and hsv-Histogram. After extraction technique the original image remains in the database and feature vector is generated and is being store within an index. In order to store the vectors within an index a data structure is used. As a result, the new output images of CBIR system is being retrieved from the medical database and displayed to the users that gives the query images to search the relevant images.

Section 2 elaborate about the literature survey, Section 3 describes the methodology, Section 4 describes about feature extraction techniques, Section 5 discusses similarity measure and Section 6 shows the testing and evaluation, conclusions and is discussed in Section 7 and section 8 gives future work.

2. Literature Survey

Literature survey is the most imperative step in software development process.

[1] Herbert Chuctaya, Jvan Gutierrez represents that for "Medical Content based Image Retrieval using metric data-structure" for medical images, by using four feature extraction i.e. BIC, Gabor Transform, Haralick and Gray level Histogram. Via slim tree data structure for indexing purpose and for similarity measurement of the vectors Euclidean distance algorithm is being used.

[2] GauravJaswal, discussed about "Content based Image retrieval", here the task is performed in order to retrieve the images from the large collection of database on the basis of their own visual content. It reviews the main modules of a CBIR system, including feature representation, indexing and system design, while importance the past and current technical achievements.

[3] Elena Gonzolez, Francesco Bianconi, and Antonio Fernandez proposed of "A Comparative Review of Color Features for Content Based Image Retrieval", here color feature is being aimed on feature extraction. These models only takes into account the color of an image.

[4] "Content-Based Image Retrieval: Theory and Applications", Ricardo da Silva Torres proposed advancement for data storage and image acquisition technologies have enabled the creation of set of images for this purpose of system, it requires a system that can manage these collections. The most common method is CBIR systems. Based on content based properties i.e. Color, structure, texture, their aim is to retrieve image by getting feature vectors after feature extraction. The CBIR technology has been used in several applications such as

Licensed Under Creative Commons Attribution CC BY

finger print detection, crime avoidance, medicine historical search.

[5] "Content Based Image Retrieval system in medical Applications", R. Senthil Kumar discussed about medical field, largely for digital images, created in ever increasing quantities and used for diagnosis and therapy. The Department of Radiology of Gevena University hospital produces almost 25000 images a day in 2012. Currently cardiology is the second to improve the overall CBIR system performance for medical applications. The components are extraction by creating the metadata and Query Engine calculation similarity, user interface- to view the images.

[6] "Content-Based Image Retrieval in Radiology: Current status and Future Dimensions" CoyhenBurakAkgul, Daniel L Rubin explains Diagnosis radiology requires accurate explanation of complex signals in medical images. This technique can be useful to radiologists in accessing medical images by identifying relevant images in large terms that can help in result building. In the feature extraction for images it uses hsv-histogram by distributing the color of the images and texture features, to get the feature vectors and for similarity of this vector.

[7] "A CBIR system for Human Brain Magnetic Resonance Image Indexing" Mina Raf. field with the advance of multimedia and technology of imaginary. Feature of images is being used by CBIR like color, texture and shape, to index images. Textual feature is the most powerful features. This research describes the retrieval of images using textual features, and to find out the images among the normal, abnormal conditions for the experts and helps the radiologists.

[8] "A Review of Content Based Image retrieval in Medical Diagnosis" Sh. Akbarpour focus on the medical domain mainly in cancer diagnosis. Here, two main components introduced – feature extraction, and classification Model. For extraction it include- shape Feature, color Moments, Color Histogram, and Textual Feature i.e. Gabor filter. Similarity is being measured of vectors is done by Euclidean distance.

[9] "Performance comparison of distance metrics in content-based Image retrieval applications", Euclidean distance metrics is being compared with the Manhattan distance for similarity measurement.

3. Methodology

1) The feature extraction technique must be used in the database in order to get the feature vectors. These feature vectors of database are being stored in the index by creating a dataset.
2) The features of query images are also extracted and the feature vectors that are being obtained use to compare the feature vectors of database from the index.
3) Similarity measurement between the query feature and the database feature is done using Manhattan metrics and Euclidean metrics.

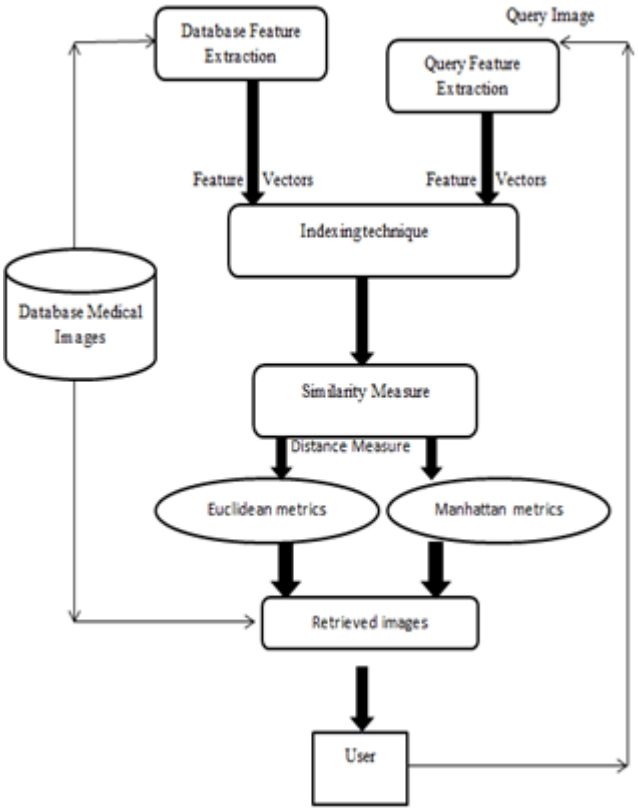

Figure1: Block diagram of the future system

System block diagram is divided into two phases – online phase and the offline phase. The operation extraction of features takes place for both the query image and the images that are present in the database so that based on their feature vectors both the images can be figure out which are the relevant images that are present in the database. Here, in the indexing technique similarity measurement is being process so that based on their vectors of database images and the query image less distance and more accuracy of relevant images will be obtained. Hence, required relevant images are obtained from the medical database. These images are being displayed on the screen to the users.

4. Feature Extraction

The User used to search for relevant images by giving a query image. Hence, in order to retrieve the images from the database all the images in the database will be feature extracted inside the feature extraction part same way the query image will also be feature extracted so that the vectors that has been obtained from the query image will be compared with the vectors of the database images. In the feature extraction part, there are some extraction techniques that will extract the features and feature vectors will be retrieved. Some of the features that are being used are –
a. Gray level Histogram,
b. Color moments,
c. Gabor transforms,
d. Haralick.

After feature extracting all this images, total feature vector of each image will be obtained.

a) Gray level Histogram Feature Extraction

Licensed Under Creative Commons Attribution CC BY

Histogram can be defined as the graphical representation of the distribution of data. It used to focus on two points over the image –

i) It used to bin the range of the colors of a particular image.
ii) It also used to see how much portion of each color in an image contains.

Gray level histograms are the most common technique to describe low-level properties of an image. The histogram of a digital image with intensity levels in then range [0;L□1], is a discrete function h(rk = Nk) , where rk is the kth value of intensity and nk is the number of pixels in image with intensityrk. It is common to normalize the histogram to divide each of its components by the total number of pixels in the image denoted by the product MxN, where it is usual that M and N are the dimensions of the image (row and column). The normalized histogram is given by P(rk) = Nk=MN to k = [0; 1; : : : ;L □ 1], which becomes a probability of occurrence of gray levels in rk, which implies that the sum of all must be 1.Hue – it is the combination of all the colors together.

b) Color-Moments Feature Extraction

Color-Moments feature extraction is being used for color distinguishing among two or more similar images. The input image has to be analysed and it extract 2 first moments from each R, G, B. The output image has 1x6 vectors containing the 2 first Color-Moments from each R, G, B. To extract the color channels from an image R, G,B matrix of an image is being used. Computing 2 first color moments are from each image, mean and standard deviation have to be obtain of R, G,B .Hence, to get output vector of Color Moments, matrix the values of mean and standard deviation of all R, G, B.

Mean -

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij}$$

Where, N is the number of pixels in the image andPijis the value of the j-th pixel of the image at the i-th color channel.

Standard Deviation –

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - E_i)^2\right)}$$

Where, Eiis the mean value, or first color moment, for the i-th color channel of the image.

c) Gabor Transforms

Gabor filter is a two-dimensional Gaussian function modulated with sinusoidal orientations at a particular frequency and direction. This technique extracts texture information from an image.. Expanding the mother wavelet Gabor forms a complete but non-orthogonal basis set. The non-orthogonality implies that there will be redundant information between different resolutions in the output data. This redundancy has been reduced by [3] with the following strategy:

Let's, Ut and Uh the high and low frequency of interest, S be the total numbers of scales, and K the total number of orientations (or translations) to be computed. Then the design strategy is to ensure that the half-peak magnitude support of the filter response in the spectrum of frequency of each contact, S = 4 and K = 6. The Gabor transform is then defined as:

$$W_{m,n}(x,y) = \int I(x_1, y_1) g_{mn} * (x - x_1, y - y_1) d_{x_1} d_{y_1}$$

Where, m; n are integers, m = [1, 2,..., S] and n = [1, 2, … ,K];

d) Haralick

Haralick presents a general statistical model to extract texture information from blocks belonging to an image. This design includes the construction of a spatial co-occurrences matrix $G_{N_g \times N_g}$, where Ng is the number of gray levels.

Each element G[i, j] is obtained verifying the amount of pixels with gray level i adjacent to pixels with gray level j. Thus, each entry G[i, j] can be considered as a probability. Spectrum frequency 2D with 4 states and 6 orientations that a pixel with gray level i is adjacent to a pixel with gray level j.

This method consists of 13 features obtained from the co-occurrence matrix calculated:

1) Energy
2) Correlation
3) Inertia
4) Entropy
5) Inverse Difference Moment
6) Sum Average
7) Sum Variance
8) Sum Entropy
9) Difference Average
10) Difference Variance
11) Difference Entropy
12) Information measure of correlation
13) Information measure of correlation

5. Similarity Measure

In order to retrieve the radiological images the features of images must be extracted from both the database and the query images. Based on this extraction technique the images will be differentiated from each and every image. When feature is being extracted, feature vector is being obtained from each image. These vectors will help to retrieve the relevant images. Here, feature extraction techniques that has been used are – gray level histogram, haralick, color moments, Gabor transforms. Each of this extraction task differently. Indexing technique allows all the feature vectors to be arranged in an index. This indexing will help in searching of images by comparing the distance between database vectors and query vectors. Based on the minimum distance the more relevant images will be displayed. The indexing will make the searching more easily.

Similarity measures the distance vectors of the database images with query vectors. If the distance is less, more relevant images will be retrieved from the database by comparing the vectors. Hence, similarity is being measured by distance metrics – Euclidean metrics and the Manhattan metrics. Here, two metrics is being used based on their

Licensed Under Creative Commons Attribution CC BY

searching ability the images will be retrieved and depending on the best performance the images will be retrieved.

5.1. Manhattan Distance

The Manhattan distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance between two items is the sum of the differences of their corresponding components

Let us take, a=(U1,V1) and b=(U2,V2) are the two points, then the Manhattan Distance between a and b is given by-
MH(a,b) = |U1-U2| + |V1-V2|

Instead of 2-Didensions, if the points have n-dimensions, each as p = (U1,U2,U3,….,Un) and b = (V1,V2,V3,…,Vn) then by defining the Manhattan distance between p and q –
MH(p,q)=|U1-V1| + |U2-V2| + …. + |Un-Vn|
MH(p,q)= ∑ |U-V|.

5.2. Euclidean Distance

The Euclidean or Euclidean metric is the "ordinary" distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space.

The Euclidean distance between points **p** and **q** is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).In Cartesian coordinates, if **p** = (p_1, p_2,..., p_n) and **q** = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance (d) from **p** to **q**, or from **q** to **p** is given by the Pythagorean formula:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

6. Testing and Evaluation

Some features are dependent on the properties of the image, such as mean, median and standard deviation, where:

Mean: Average or treats the columns of the image as vectors, then returning a row vector of mean values.
Median: Treats the columns of the image as vectors, then, returning a row vector of median values.
Standard deviation ( ): Standard deviation of the pixels of each column of the image matrix then returning a row vector containing standard deviation. The figure below showing the results for the proposed system when applied on one of the test image as a model, also tables showing the results measured for the proposed system:

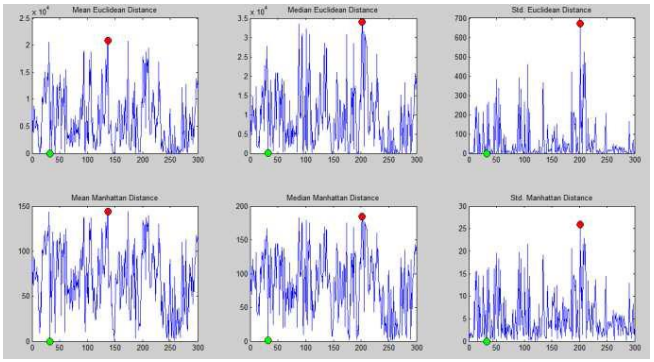

Figure 2: Mean, median, standard deviation of the Euclidian and Manhattan distance.

Table 1: Shows the similarity measure for both Euclidean & Manhattan distance based on Mean, Median and Standard Deviation

Features test image (1-30)	Mean		Median		Std.	
	Eclud.	Manh.	Eclud.	Manh.	Eclud.	Manh.
Rate of matching	100%	100%	13%	13%	100%	100%

Table1 shows the similarity measure for both Euclidean and Manhattan distance the rate of matching is 100% based on the mean and standard deviation have maximum average value so they are good for image retrieval and the images are retrieved for Euclidean, Manhattan distance.

7. Conclusion

As a conclusion, CBIR using the proposed technique shows that it uses four different feature extraction to retrieve medical images and for measuring the feature vectors, similarity measurements – Euclidean and Manhattan metrics are being used and gives almost the same high performance while comparing with the properties of the image. Hence, they are good for image retrieval in order to retrieve relevant images either of the distance metrics can be used depending on the users need.

8. Future Scope

The future work includes for similarity measurement by using another method of distance measurement like standardized Euclidean distance and Normalized Euclidean distance metrics.

References

[1] "M-CBIR: A medical content-based image retrieval system using metric data-structures", Herbert Chuctaya, Christian Portugal, C´esarBeltr´an Cathedra Concytec in TIC National University of San Augustin, UNSA Arequipa, Peru Fjuanherbert ,christ.pz.cs ,cesarbcg@gmail.com and Juan Guti´errez, CristianL´opez, Yv´anT´upac Cathedra Concytec in TIC National University of San Augustin, UNSA Arequipa, Peru fjcgutierrezc,criloal23,yvantvg@gmail.com, 2011

Licensed Under Creative Commons Attribution CC BY

30th International Conference of the Chilean Computer Science Society

[2] "Content Based Image Retrieval-A literature review" GauravJaswal, AmitKaul , NIT, Hamirpur-177001, Himachal Pradesh(CCC-09), National Conference on Computing, Communication and Control (CCC-09) , 2009.

[3] "A Comparative review of color features for Content-Based Image retrieval", Elena Gonzalez, FranescoBianconi and Antonio FernandezUniversidad de Vigo , Departamento de Diseño en la Ingeniería Vigo / España E-mail: {elena,antfdez}@uvigo.es, UniversitàdegliStudi di Perugia DipartimentoIngegneriaIndustriale Perugia / Italia E-mail: bianco@unipg.it.

[4] "Content Based Image Retrieval:Theory and Applications", Ricardo da Silva Torres, Alexander Xavier Falco, Institute of computing, State university of Campinas, SP, Brazil, rtorres, afalcao@ic.unicamp.br , 2006.

[5] "Content Based Image Retrieval System in Medical Applications, R.Senthilkumar, Asst. Proff. Dept od Computer Science Dharmapuramadhiram ,Mayiladuthurai(Tamil Nadu), IJERT, ISSN: 2278-0181, vol-2 Issue 3, March-2013

[6] "Content Based Image Retrieval(CBIR) in Radiological: Current status and Future Directions, Journal of Digital Imaging, Published online: 08 April 2010, email-acarbu@bounecu.tr

[7] "A Content Based Image Retrieval system for Human Brain Magnetic Resonance Image Indexing", Mina Rafi Nazari electrical engineering Department, Sharif University of Technology International Campus Tehran, Iran IJCA 0975-8887. Volume 7- No.14, October 2010.

[8] "A Review on Content Based Image Retrieval in Medical diagnosis", Sh.Akbarpour, Department of computer, Faculty of engineering, Shabestar Branch, Islamic Azard University, Shabestar, Iran , email: sh.akbarpour@gmail.com, June 2013.

[9] "Performance comparison of distance metrics in content-based Image retrieval applications", A VadivelDept. of Computer Science and Engineering, Indian Institute ofTechnology, Kharagpur, India vadi@cc.iitkgp.ernet.in , A K Majumdar Dept. of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, India akmj@cse.iitkgp.ernet.in, ShamikSural School of Information Technology, Indian Institute of Technology, Kharagpur, India shamik@cse.iitkgp.ernet.in

[10] "Content-based Image Retrieval (CBIR) using Hybrid Technique" Zainab Ibrahim Abood Electrical Engineering Department, University of Baghdad, Iraq , 2Israa JameelMuhsin Physics Department, College of Science, University of Baghdad, Iraq, NabeelJameelTawfiq Remote Sensing Unit, College of Science, University of Baghdad, Iraq, International Journal of Computer Applications (0975 – 8887)Volume 83 – No 12, December 2013.

Licensed Under Creative Commons Attribution CC BY