Medical Image Retrieval and Quantitative estimations via Ray-based Methods

Nasibeh Esmaeilishahmirzadi
Institute for Information Technologies, Mathematics and Mechanics
Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia

Abstract

The main goal of this paper is to review some existing text-based and content-based image retrieval systems and methods. The images should be indexed properly to be saved and retrieved. It is suggested although images have some features different from texts but some techniques in text retrieval such as artificial intelligence and relevance feedback can be used to improve the image retrieval systems. In addition, due to the importance of on-time diagnosis of lung cancer, an algorithm based on CT images processing is introduced. This proposed algorithm is described in order to determine area of tumor. The algorithm has three stages and result is shown in histogram of images.

Keywords: *Content-Based Image Retrieval, relevance feedback, text-based retrieval, Medical image retrieval, Review.*

1. INTRODUCTION

The most basic connection between human beings by the images. Kherfi and his colleagues have investigated the application of image in their paper.In their point of view text can't describe and convey information in the description of an event, medical images, describing a place, etc as well as image.Images are widely used in the medical field as diagnosis, education and research . Kherfi and his colleagues (2004) noted the increasing use of images in articles, newspapers, advertising, architecture, medicine, etc the needs of organization of this information is necessary[1]. Enser (1995) [2] describes the different methods of indexing and search archives of images, mostly text-based system.

Tang and his colleagues (1999) had researched on review of intelligent content-based indexing. They had investigated on trace the development of visual information systems for healthcare and medicine from Picture Archiving and Communications Systems (PACS) to retrieve images based on color, texture, and shape [3,4]. In this paper with regard to the importance of organization and image retrieval techniques, system features and applications based on the content and assessment methods are discussed. Examples of the image retrieval system are introduced, then notes the use of image retrieval systems in medical images database.

2. Image retrieval

Image retrieval subject has been discussed since the early 70s Y.Liu has introduced four major methods for searching and retrieving images: [5]

I. A search for images based on attributes such as file name, creation date and other categories such as subject and file creator of the file. The main problem is that information is automatically generated images, they are not enough to describe the content and add data manually takes time.

II. The second method of detecting objects automatically, which is often used in certain domains such as medical images. This method is also time-consuming and very hard to do in general.

III. A way of taking advantage of the text to interpret the images and text, followed by the use of data recovery techniques. This method can be done in two ways: (1) manual interpretation (2) The automatic interpretation. First, professionals must interpret each image. The problem with this method does not work when the volume of images is high, in addition interpretation is subjective and incomplete.

IV. The fourth one is content based. Due to the increasing volume of video databases, biased and subjective and time-consuming and costly process of assigning tags to images, text indexing has become a relatively inefficient way. Trying to solve this problem led to the creation of a content-based image retrieval systems [6]. A low-level features have characteristics such as color, shape and texture and a high level of features have characteristics image semantic features. In general the image retrieval systems act in 2 stages: firstly the base visual features of images automatically are extracted, and then with the help of them images are indexed and secondly, after receiving the user's query image, low-level features or visual features extracted and visual characteristics database try to find the nearest images to the image query [7].

3. Visual features used in a retrieval system based on content

In addition, content-based image retrieval techniques most used features (colors, layout, texture and shape) are briefly discussed.

I. Retrieval based on color feature
Color histogram is the most important color display used in image retrieval. The most important techniques are: Histogram Intersection, compared histogram, and the Cumulated Color histogram (to reduce the effects of noise) [8]. The other additional features can be mentioned is Color Moments and Color Sets [8,9].

II. Retrieval based on color spatial features
Use only the color characteristics of the databases with high amount of images leads to increase the rate of error in responses and become ineffective.so the color spatial features is presented. The general idea of the color spatial feature is to extract certain areas of the image [10].

III. Retrieval based on texture features
Retrieval based on texture do not use alone But the similarities between areas of tissue that have the same color (such as sky, sea or leaves and grass) is very useful. Using the matrix of occurrence [11] was one of the first efforts in this area has a history that dates back to long before the topic of image retrieval. One of the most important method in this area is Tamura[12].

IV. Retrieval based on shape features

Retrieval based on shape is the most powerful visualization technique in retrieval systems. Protecting its identity against rotation, scaling and transmission is the most important feature, which make it suitable for image processing applications [13].

The shape features according to the extraction are divided into two categories: features based on edge and features based on areas. The most important characteristic features of objects that are used include aspect ratio circularity constant torque Fourier Descriptors (Use the edges of the Fourier transform) Consecutive Boundary Segments. Other methods have been proposed to adapt shapes are elastic deformation of templates, directional histogram for edge of image, skeletal representation of image objects which are compared with graph matching techniques, Finite Element Method, Turning Functions and wavelet descriptors[14].

4. MEDICAL IMAGE RETRIEVAL

One of the most important application of image is in medical. Muler reported that in of the University Hospital of Geneva alone produced more than 12,000 images a day [4]. Some content based system in medical and clinical field is ASSERT[8], IRMA and NHANES. Many medical image retrieval systems such as high-resolution images of CT-lungs, Mammography, CT images of the chest, X-ray images of the chest, X-ray images of the spinal cord and X-ray images of teeth are often used for some special part of body and specific imaging used And are not used for other applications. Conversely few of these systems (KMeD) and IRMA have been developed for general medical applications [15].

Most of the clinical retrieved image systems leads to content keywords like DICOM information header for retrieving. CBIR widely have been investigated in a variety of areas and provide a method to retrieve query based on image data with features such as color, shape and texture. With all these efforts available CBIR systems are not fully compatible in Health in wide database. At data entry time, numerical features are computed from each image stored within the database. Using the QBE approach, the same features are extracted from the query image and compared to the features stored within the database. The images that correspond to the most similar features are then retrieved from the database and presented to the user to answer his query [11].

4.1 IRMA (Image Retrieval in Medical Applications)

IRMA is a cooperative project of the department of Diagnostic Radiology, the Department of medical informatics, and division of medical image processing and chair of computer science at the Aachen University of Technology. Aim of the project is the development and implementation of high-level methods for content-based image retrieval with prototypical application to medico-diagnostic task on a radiologic image archive [5]. IRMA has 3 online demos that are IRMA Query demo which provides evaluating CBIR in several databases, IRMA Extended Query Refinement demo which is contain 10000 images of CBIR and Spine Pathology and Image Retrieval Systems (SPIRS) which is designed by NLM/NIH (USA) and have 17000 spine and chest X-ray image [16].

4.2 MedGIFT (GNU Image Finding Tool)

MedGIFT project began at the University of Geneva, Switzerland in 2002 and at the Institute for Business Information Systems at the University of auxiliary Switzerland (HES-SO). The MedGIFT project aims at providing an open source framework of reusable component for a variety of medical application. It's based strongly on the GNU image finding tool (GIFT) as its central piece. Main developments are on the integration of various new components to create domain specific search and navigation tool [17]. This project includes Talisman lung image retrieval system, case-based image retrieval system, Medical image retrieval system using grid computing, ImageCLEF for validation and evaluation and MedSearch for user interface.

5. Proposed Algorithm

5.1 Boundary points

In image of both normal and abnormal lungs, key points are determined and for this, points located in edges of the image are used so that initially, all boundary points are determined and then, points which are critical are extracted from them. Of course, this stage is completed manually by an expert or it can be automated using a method such as artificial neural networks and genetic algorithm.

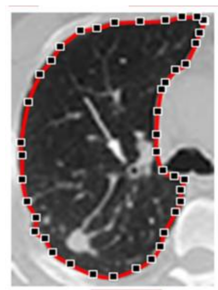

Fig 1: Boundary points

5.2 Lines between points

At this stage, points determined in previous step will be connected to each other. For this purpose, various algorithms based on MCP. Finally, points are considered as the vertices of the graph and the distance between two points is taken as the weight of the line between them. Final arrangement after applying algorithm is depicted in Fig. 2.

5.3 Determination of the area of tumor

In this step, graph obtained previously is processed. This step has two parts. First, boundaries of the tumor are highlighted by dashed line and second, area of the highlighted region is computed. These two stages are explained below.

This step is also accomplished in two parts. In first step, center of the image is determined and its distance from all vertices of the graph are computed and distances are arranged in array in a manner that highest point is taken as starting point and its distance from the center is written in the first cell and then, we move clockwise and find the values of other arrays. In Fig. 2, distance between points are determined.

At the end of this step, two arrays are created as *a* and *b* whose elements are the distances between the center of each image and graph vertices. In the next step, two arrays created in the first step are compared to each other and if the difference between two analogous elements exceeds a threshold value, it will be regarded as the sign of the presence of an abnormal tumor in lung. This threshold is obtained using a trial and error method.

Therefore, at the end of this step, difference between distances are computed and compared to the threshold value and if it is more than the threshold, analogue vertices are marked as shown in Fig. 3 and consequently, tumor region is determined.

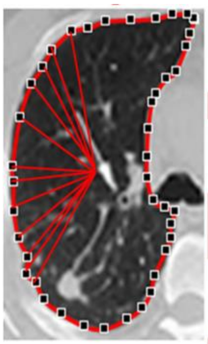

Fig 2: Representation of the distance between graph vertices and image center

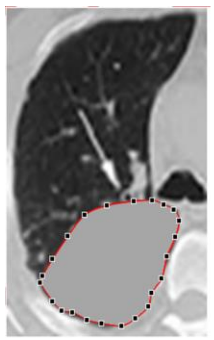

Fig 3: Marking the margins of the tumor and its color separation

5.4 Area of tumor

For this step, triangulation method is used. As can be seen in Fig. 3, each image is divided into many triangles whose area can be calculated and by summing them up, overall area can be found. For this end, overall area relation is used. Calculation is performed for both images and then, area of both image is obtained and their difference will give the area of the tumor.

Previous stages are done using graphing method and area of the tumor is obtained. To determine that how precise is the work, above stages are completed using a method common in image processing. At the end, results are compared and the precision is evaluated. The method used here includes two steps. First, boundary of tumors is determined so that the region of tumor can be clearly separated from the other parts of the image. Then, area of the extracted region is computed.

5.5 Separation of the tumor

For this purpose, separation by colors is used. In this way, difference in color of the normal and abnormal images determine the tumor. As can be observed in Fig.3, dashed lined region of the Fig.4 is separated from the other parts using the method of color separation. Since middle colors are moderated, values of the elements of the array are in white and black and middle and gray points are deleted. Consequently, by counting the number of pixels available in white region, area of the tumor will be obtained.

Fig 4: The mask for the separation of tumors for the area calculation

Lung image database consortium (LIDC) has a database, which includes CT images of lung and information about nodules, which are shown in these images including notes of 9 physicians about characteristics of nodules: calcification, internal structure, level of precision, lobules, margin, sphericity coefficient, malignity and speculation.

All of these characteristics are ranked in 1-5 range with integers except calcification, which is ranked in 1-6 range. Looking to the histograms of these characteristics (Fig. 5), we will find out that many of them including calcification, internal structure, precision and speculation are mainly in one or two main values. Therefore, when we try to evaluate the correlation between characteristics of image and ranking of physicians, these rankings don't help considerably. Data were extracted from XML and centroid calculation was used for determination of images with the same nodule. After that, nodule images were extracted from full-size CT scans of lung. In this way, files were extracted from nodules together with a set of XML files with all data corresponding to features, physicians notes and metadata for each of the nodule images. All nodule images which were smaller than 5×5 (about 3×3mm) were eliminated since images which are so small cannot provide meaningful data about texture (same minimum size is used by Kim et.al as well). After elimination of these images and those having numerous lines, the average size of images is 15×15 pixels and actual average size of the image is about 10×10mm. smallest nodules are about 3×3mm while largest size is more than 70×70mm. 88% of the images are below 20×20mm.

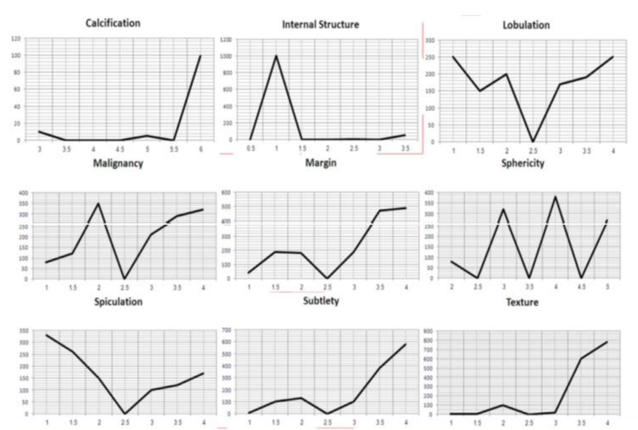

Fig. 5: Histogram of images

6. Conclusion

Images as a kind of information can transfer a large volume of information, especially in medicine, in education, research and diagnosis. The purpose of medical information retrieval system is providing relevant information at the right time to the right user. Images in research, indexing, storage and retrieval have been investigated. Content-based methods can be used on a large variety of images and in a wide area of applications. As mentioned in this paper text-based and content-based image retrieval systems, the application of the image systems especially in medicine. Some existing systems are described.

Owing to the importance of on-time diagnosis of lung cancer, an algorithm based on CT images processing is introduced. The algorithm has three stages. In first step, using the method of graphing, location and area of the tumor are determined. This stage is accomplished here by a manual way with using the an expert knowledge. It can be exchanged by the automated one using neural networks or genetic algorithm. On the next stage, using the image processing method, location of tumors are recognized, determined, and they are completely separated from other parts of the image. On the third stage the area of the separated part is calculated. Presented method is very effective in reduction of the human errors in diagnosis of tumors via images. In future works, we plan to reduce of the role of a human.

7. REFERENCES

[1] Kherfi ML, Ziou D, Bernardi A. *Image retrieval from the world wide web: Issues, techniques, and systems // ACM Computing Surveys (CSUR). 2004; 36(1):35-67*

[2] Enser, Peter. *Image databases for multimedia projects //Journal of the American Society for Information Science. January 1995. Vol. 46, Issue 1, pages 60–64*

[3] Tang, L. H. Y. et al. *A review of intelligent content-based indexing and browsing of medical images //Health Informatics Journal. March 1999. Vol.5, No.1, 40-49*

[4] Henning Müller*, Nicolas Michoux, David Bandon, Antoine Geissbuhler. *A review of content-based image retrieval systems in medical applications – clinical benefits and future directions //Service of Medical Informatics, University Hospital of Geneva, Rue Micheli-du-Crest 24,1211 Geneva 14, Switzerland*

[5] Lehmann, Thomas, et al. *"A Content-Based Approach to Image Retrieval in Medical Applications." Emerging Trends and Challenges in IT Management. Achen University of Technology (2006): 911-912*.

[6] Long, Fuhui, Hongjiang Zhang, and David Dagan Feng. "Fundamentals of content-based image retrieval." Multimedia Information Retrieval and Management. Springer Berlin Heidelberg, 2003. 3-4.

[7] Liu Y, Zhang D, Lu G, Ma WY. A survey of content-based image retrieval with high-level semantics. Pattern Recognition. 2007; 40(1):262-82.

[8] Stricker M, Orengo M, editors. *Similarity of color images1995: Citeseer.1-12*

[9] Smith, John R., and Shih-Fu Chang. *"Single color extraction and image query." Image processing, 1995. Proceedings. International conference on. Vol. 3. IEEE, 1995;1-4.*

[10] Faloutsos C, *Barber R, Flickner M, Hafner J, Niblack W, Petkovic D, et al. Efficient and effective querying by image content. Journal of intelligent information systems. 1994; 3(3):231-62.*

[11] Lehmann, Thomas M., et al. *"Automatic categorization of medical images for content-based retrieval and data mining." Computerized Medical Imaging and Graphics 29.2 (2005): 143-155.*

[12] Tamura H, Mori S, Yamawaki T. *Textural features corresponding to visual perception. Systems, Man and Cybernetics, IEEE Transactions on. 1978;8(6):460-473.*

[13] Rui Y, Huang TS, Chang SF. *Image retrieval: Current techniques, promising directions, and open issues. Journal of visual communication and image representation. 1999; 10(1):39-62.*

[14] Rui, Yong, Alfred C. She, and Thomas S. Huang. "*Modified Fourier descriptors for shape representation-a practical approach." Proc of First International Workshop on Image Databases and Multi Media Search. 1996;1-8*

[15] Antani S, Lee D, Long LR, Thoma GR. *Evaluation of shape similarity measurement methods for spine X-ray images. Journal of Visual Communication and Image Representation. 2004; 15(3): 285-302.*

[16] Analoui, Mostafa, Joseph D. Bronzino, and Donald R. Peterson, eds.*Medical imaging: principles and practices. CRC Press, 2012;6-14.*

[17] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: *a physician-in-the-loop content-based retrieval system for HRCT image databases. Computer Vision and Image Understanding. 1999; 75(1):111-132.*

About the author

Nasibeh Esmaeilishahmirzadi is a Ph.D. student at Lobachevsky State University of Nizhny Novgorod. Her contact email is nasibe.smaeili@gmail.com