Open Access

Medical Image Retrieval Algorithm Based on Content

Liying Du[1,*] and Nabila H. Saleh[2]

[1]Department of Computer Science and Engineering, Jilin Jianzhu University, Changchun 130118, China

[2]Pathology Department, National Research Center, Cairo, Egypt

Abstract: With the rapid development of science and technology and the improving medical service, medical image's role in clinical diagnosis and treatment becomes increasingly prominent. It has become a high-profile task to help the doctors pick out desired targets from massive medical images. Currently, techniques of text-based medical image retrieval have failed to meet the need of massive medical image retrieval. On the other hand, techniques of content-based medical image retrieval are already established and hold vast research potential. Starting with the introduction of matured techniques of medical image retrieval, the paper expounds on the evaluation criteria of the effectiveness of such techniques, then on the modified text-and-content-based medical image retrieval algorithm. The last part is the verification of the research conclusion with contrast experiments illustrated by the sample figures.

Keywords: Medical images, retrieval algorithm, content-based, image retrieval, segmentation, image gray value.

1. INTRODUCTION

With medical technology's rapid progress, massive medical images are being produced every day in medical institutions. The effective management of these images becomes a hot spot of research.

Depending on information such as the patients' name, the number of images or the serial number of their medical records, traditional techniques of text-based medical image retrieval are featured with evident retrieval trait and easy operation; they demand, however, the manual labeling on the images. For massive medical images, it is impossible to manually label all the images. Besides, human description is bound to be inaccurate and subjective. All these make techniques of text-based medical image retrieval fail to meet the needs of today's massive image retrieval [1].

Designed to solve the defects of such text-based techniques, techniques of content-based medical image retrieval, by processing the gray values, the topological and geometrical shapes, can extract these characteristics and establish respective retrieval criteria with them. This paper will focus on the algorithm of content-based medical image retrieval.

2. SYSTEM STRUCTURE OF CONTENT-BASED MEDICAL IMAGE RETRIEVAL

System of content-based medical image retrieval (illustrated by Fig. (**1**)) can be regarded as an information service system linking the user with the medical database.

Basically it works like this: first, it analyzes all the images in the database and extract the unique features from these images; then establish the medical image database as well as respective feature library. When retrieving images, it adopts the retrieved image's features and matches them with all the features in the library. Then, according to the matching result, picks out the desired images from the database. The system of content-based medical image retrieval is formed by the image processing module and the retrieval module. The processing module's main functions are: image labeling, extraction of features, establishment of index and storage of the image data. It also contains the techniques of extraction and marking of region of interest, classification of images and segmentation of various sized images.

3. COMMON TECHNIQUES OF CONTENT-BASED MEDICAL IMAGE RETRIEVAL

The key techniques of the system of content-based medical image retrieval are contained mainly the medical image segmentation, extraction and marking of region of interest, extraction of features, similarity match and retrieval as well as relevant feedback techniques [2].

3.1. Medical Image Segmentation

In general, medical images are focused on specific human tissues or organs. Medical image segmentation means to segment such specific sections and sort out regions of interest, which provides local features for clinical diagnosis.

In this way, the data processing amount is reduced, the processing speed is raised and the accuracy of the disease analysis and diagnosis are also improved.

3.2. Features of Medical Image

3.2.1. Features of Image Gray Value

Image gray value is one of the basic properties of images, which possesses certain robustness and stability. Its histo-

*Address correspondence to this author at Department of Computer Science and Engineering, Jilin Jianzhu University, Changchun 130118, China; Tel: +8613944104396; E-mail: duliying@126.com
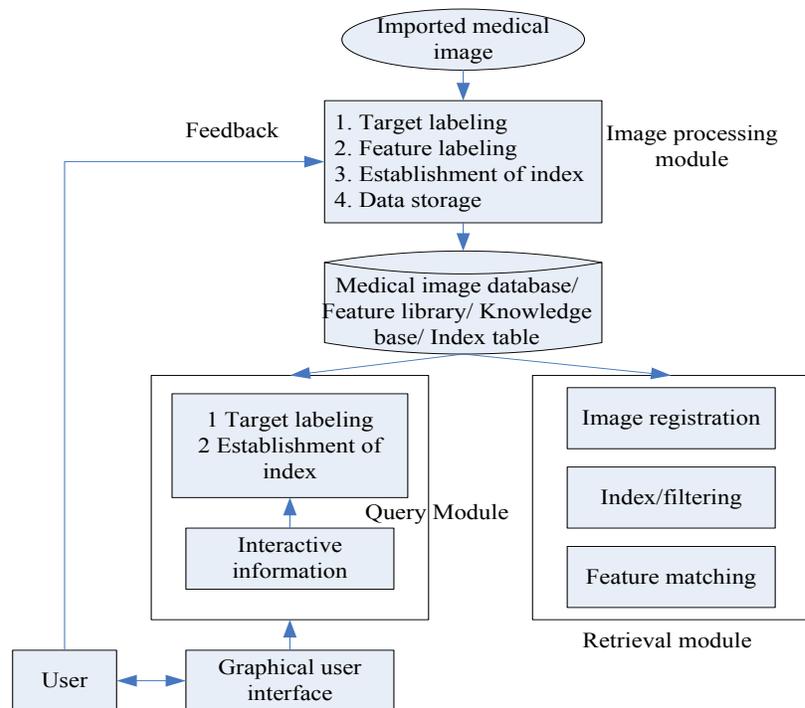
Fig. (1). Illustration of CBMIR system.

gram reflects directly features of the general distribution; however, it shows no information about gray value's spatial location. The mean gray value histogram acts as its image gray value function and the gray value distribution function of all normal image data within the medical image database. The histogram can also be used in the comparison of normal images and abnormal images [3].

3.2.2. Texture Features

As a feature related to the geometric distribution of medical gray value, texture feature is more friendly to the doctor's intuitive feelings. In terms of the relationship with medical image pixels, texture feature analysis could be categorized as the statistical method, the spectral method, the model method and the structural method. The statistical method, being the frequently-used one, depicts texture features such as the uniformity, the directionality and the thickness. The operation theory is to extract different statistics from the image's texture feature.

3.2.3. Image Shape Feature

Being an important feature of the medical images, the shape feature holds a significant position in the medical image retrieval system. Common methods to depict image shape features include Freeman chain code, quadratic curves, B line, super-quadratic-curves and shape approach method based on wavelet transform.

3.2.4. Similarity Match and Retrieval

Techniques of high dimensional medical image retrieval, employing multidimensional hash table and network spatial technology in its early development time, received only modest results. Nowadays there are mature techniques of

high dimensional medical image retrieval, such as Bucketing grouping algorithm, cluster, neural network, and quad-tree and k-d tree. Take self-organizing neural network tree index method (SOM) for example, it can memorize in the absence of human supervision, has dynamic clustering function and support arbitrary similarity measurement. All these make it an index technique with wide application potential.

3.2.5. Interactive Feedback Technique

System of content-based medical image retrieval's accuracy may be impaired by the inconsistency between the image's feature and the user's understanding toward the data. There with interactive feedback technique, the user can, without establishing weights for feature information, pick out images similar or dissimilar to the desired image. Then the system will automatically adjust the retrieval result according to the former feedback. Thus a better retrieval result is achieved and the inconsistency reduced.

4. RETRIEVAL ALGORITHM OF CONTENT-BASED MEDICAL IMAGE RETRIEVAL

4.1. Introduction of Retrieval Algorithm of Medical Image Retrieval

With the textual information and the content of the medical image combined organically, the improved retrieval algorithm of content-based medical image retrieval proposed in this paper can extract automatically the textual information for the initial retrieving, and then conduct the content-based retrieval with comprehended feature information of the images. With key words for images not being necessary, the algorithm can extract the desired information from vast features with part of the images sifted by the initial text-based
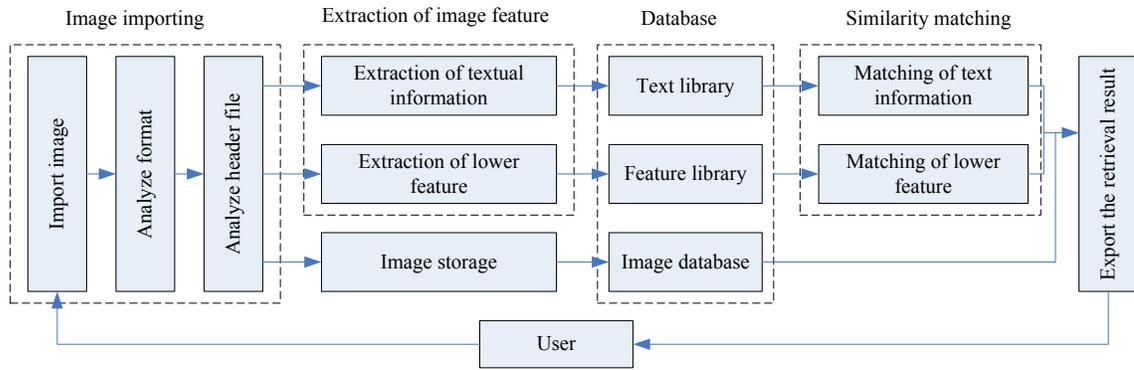
Fig. (2). System Frame of CBMIR.

Table 1.　Part of DICOM Text Information.

Label	Description	Example
(0008,0008)	Image Type	AXIAL
(0008,0060)	Modality	CT
(0010,0030)	Patient Date of Birth	19760101
(0010,0040)	Patient Sex	M
(0018,0015)	Body Part Examined	ABDOMEN
(0018,5100)	Patient Position	FFP
…	…	…

retrieval. The retrieving efficiency is increased from decreased amount of arithmetic. The paper will prove the algorithm's efficiency with several experiments based on CT images of human colons [4].

4.2. Construction of Content-Based Medical Image Retrieval

Fig. **(2)** shows the system frame of CBMIR, the CBMIR of this paper contains the following parts: 1. Database, of medical images, text information and image features: 2. Image import module, which filters images according to their formats and then separates text information from image data, filing them into respective database; 3. Image feature extraction module. This part extracts image's text information and file them in the library; 4 [5]. Similarity match module. After matching the eigenvector of the image with what is in the library, it retrieves the medical images in line with their similarity; 5. Dynamic retrieval. The user can filter out unrelated images by initial retrieval, and find qualified ones with secondary retrieval of lower features. The user can jump to the secondary retrieval, of course.

4.3. Implementation of Retrieval Algorithm of Content-Based Medical Image Retrieval

4.3.1. Extraction Technique of Image's Text Information

DICOM, being recognized all over the world, is the criteria of digital medical image's transmission, display and stor-

age. Its file consists of file header and image data, the file header containing relevant information of the identification data. Generally, CBMIR mainly extracts the patient's information from the header file, such as name, age, the hospital and the scanning time [6].

To make improvement on the retrieval, the algorithm proposed in this paper will move to location of the organ, image type, scanning posture and other more crucial information. To get better results, the user can first sift with text information, and then conduct a secondary retrieval with lower features of the medical images. Table **1** shows the text information of the CT image of human colon.

4.3.2. Extraction Technique of Medical Image's Lower Feature

The majority of medical images are gray level images with no apparent patterns and finer textures. Techniques of texture feature analysis, such as gray level co-occurrence matrix, can give clear image information in terms of direction and amplitude of variation, and depict medical image's direction of repetition and degree of roughness [7]. Gray level co-occurrence matrix assumes that in the θ direction, with the number of N pixels distance, the probability of a pair of pixels with their respective gray value as i and j is $P(i, j \mid n, \theta), (d = 1, 2, 3...; \theta = 0^0, 45^0, 90^0...)$. Adopting the probability in the matrix, we can get the following parameters:

1.　Ability

$$f_1 = \sqrt{\sum_i \sum_j (p(i,j)^2)} \; ; \tag{1}$$

1.   Contrast ratio

$$f_2 = \sum_i \sum_j (i-j)^2 \, p(i,j) \; ; \tag{2}$$

2.   Reversed distance

$$f_3 = \sum_i \sum_j \frac{p(i,j)}{1+(i-j)^2} \; ; \tag{3}$$

3.   Entropy

$$f_4 = \sum_i \sum_j (-\ln p(i,j)) p(i,j) \; ; \tag{4}$$

4.   Correlation coefficient

$$f_5 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{5}$$

4.3.3. Extraction of Image Shape Feature

Currently, there are two types of descriptors of medical image's shape: area-based shape descriptor and profile-based descriptor. Area-based descriptor covers all the shape area and has advantage in terms of the resistance to external disturbances such as noise. Profile-based descriptors are Hu moment invariant, Zernike moment and others. Taking Hu moment invariant for example, it has been proposed that the seven moments invariant have the invariance of measurement, rotation and translation. For these properties, it is widely applied [8].

4.3.4. Integration of Various Features of Medical Image

No retrieval based on one single image feature can take all factors into consideration. To compensate this defect, integration of various features with different weights in line with the user's demand is needed.

In retrieval with texture features or shape features, the user, after distributing different weights to respective features according to his need, and conducting linear weighted value summation, can get the similarity degree of the two images [9]. If the similarity of the text feature is S_1, and similarity of shape is S_2, and W_2 and W_2 are their respective weights, then the similarity summation is:

$$S = W_1 \cdot S_1 + W_2 \cdot S_2 \; .$$

4.4. Evaluation Criteria for Medical Image's Retrieval

4.4.1. Performance Evaluation for the Retrieval

There are several criteria for the evaluation of MIR, and the commonly used factors are the accuracy and the recall ratio. Because those criteria require preset threshold value of the image's similarity degree, and that the number of medical image is massive and is subjective to compare similar images, it is difficult to determine the threshold value. This research adopts the criteria recommended by the mature MPEG-7 to evaluate the experiments' results. The equation:

$$\varphi_N(q_i) = \sum_{j=1}^{N} \frac{\psi(p_j, R)}{N}, \; \psi(x,y) = \begin{cases} 1 & if \; x \in y \\ 0 & if \; x \notin y \end{cases} \tag{6}$$

φ_N means the correct rate of the former N image results; R is the image set with a certain feature; M is the preset number of sample images before the retrieval; q_i are the sample images.

4.4.2. Ordering Method for Medical Images

Another common evaluation criterion for medical image's retrieval is ordering method for medical images. Its working principle: the more similar the retrieval result with the sample image is, the higher it will be ranked. Sorting equation adopted in this paper is:

$$Rank = \frac{\sum_{i=1}^{R} P_i}{\sum_{j=1}^{T} P_j} \tag{7}$$

R is the set of relevant images, T is set of first N relevant images, and P is the serial number of the image. The closer to 1 its rank value is, the better is the ordering result [10].

5. EVALUATIONS OF THE EXPERIMENTS OF MEDICAL IMAGE'S RETRIEVAL

5.1. Preparation for the Experiments of Image Retrieval

The data is from the American database of human colonoscopy, each record contains the CT images of both side-lying posture and the supine posture. The resolution: 512×512, The number of scanning layer: from 316 to 497. 200 CT images of human colon are randomly chosen.

5.2. Procedures and Results of the Experiment

To test the retrieval algorithms' performance, the following experiments are conducted with three retrievals. 30 most relevant images are extracted to evaluate the result. Fig. (**3**) is an example of the processing system of medical images.

1) Single feature medical image's retrieval (experiment one): calculating the gray value co-occurrence matrix on this four directions $0^0, 45^0, 90^0, 135^0$, the texture feature values of the contrast ratio, reversed distance and entropy, then the retrieval was conducted by matching these texture feature values.

2) Integrated feature medical image's retrieval (experiment two): by distributing different weights to respective features according to his own need, the user retrieved the combined texture features and shape features. The weight ratio of the texture features and shape features could be 7:3.

3) Text- and content-based medical image retrieval in this paper (experiment 3): The sample images being the supine posture image, the key words were retrieved first, and then retrieval based on integrated image features was conducted, with the weight ratio of the texture features and shape features being 7:3.

Table 2.   **Experimental Results and Statistics.**

Method of Experiment	N	PN	Rank
Texture feature-based retrieval	19	64.48	1.397
Shape feature-based retrieval	23	77.8	1.247
Retrieval with integrated lower features	25	83.3	1.191
Retrieval with integrated lower features	26	84.4	1.190
Retrieval with integrated lower features	27	88.9	1.137
Retrieval with lower and text features	29	97.8	1.033

5.3. Evaluation of the Experiments of Medical Image's Retrieval

By ranking the above experiments' mean retrieval correct rate P_N and mean ordering value, we got the following conclusion:

1) Shape-based image feature retrieval is better than texture-based image feature retrieval.

For gray value co-occurrence matrix reflects the overall feature, and the images' texture varies much, shape-based image feature retrieval, possessing no great advantage, is still superior to other images. The comparative data is shown in Table **2**.

2) Retrieval algorithm of integrated texture and shape features is better than algorithms based on single image feature. As the texture-based algorithm and the shape-based algorithm are complementary to each other, the integrated use of two methods improved the correct ratio and the ordering.

3) Text- and lower-feature-based retrieval has the best performance. As the user can retrieve the key words in the text database to extract all the needed images first and then match with lower features, their efficiency improved as the retrieval range is diminished.

CONCLUSION

Based on the study of text-based and content-based retrieval algorithms, the paper proposed an improved medical image retrieval system. Combining the advantages of text-based and content-based retrievals together, the research can help the doctor analyze and compare medical images more comprehensively and accurately. Its retrieval result is also suitable for visual perception.

Emphasizing on retrievals with integrated features, the paper has only discussed two features (texture and shape features), thus more characterized lower features will be taken in the future retrieval. Also, the retrieval of doctor's region of interest will be discussed, so as to achieve even better retrieving results.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Y. F. Zhao, L. Z. Xia, and H. Pan., "Contourlet-based feature extraction for texture images", *Signal Processing,* vol. 14, no. 2, pp. 227-241, 2010.

[2] H. Wang, Y. Q. Zhu, and W. Wu, "Implementation of both information extraction and image display from Dicom Medical Image File", *Medical Equipments*, Information, vol. 22, no. 9, pp. 1-3, 2007.

[3] W. H. Song and J. Han, "Implementation of CBIRA with integrated lower and semantic features", *Journal of Huangshan University,* vol. 10, no. 5, pp. 66-69, 2008.

[4] J. Luo and H. Kuang, "Content-based image retrieval using combination features", *Computer Engineering and Applications*, vol. 45, no. 1, pp. 153-197, 2009.

[5] S. J. Zhang and Q. M. Wang, "Research on image retrieval based on multi-features, *Computer and Modernization,* no. 1, pp. 124-127, 2010.

[6] Y. Song and F. A. Liu, "An image retrieval combining color and shape feature, Microelectronics & Computer, vol. 24, no. 12, pp. 103-105, 2007.

[7] Y. X. Xie, L. D. Wu, and X. D. Luan, "Research of content-based image retrieval technology", *Computer Engineering and Applications,* vol. 38, no. 1, pp.35-38, 2002.

[8] X. B. Huang, "Content-based image retrieval technology", *Journal of Academic Libraries,* vol. 22, no. 4, pp.15-27, 1999.

[9] Z. X. Yin and J. G. Li, "Texture image retrieval based on planning strategy", *Mini-Micro-System*, vol. 28, no. 6, pp.25-28, 2001.

[10] S. E. Xia, W. R. Mo, and X. Y. Wang, "Retrieving technique of medical images based on feature fusion and relevance feedback", *Space Medicine & Medical Engineering,* vol. 17, no. 6, pp. 429-433, 2004.