Combining Text and Content Based Image Retrieval on Medical Resource Database

Wei Li[1] Bo Li Peng Cao Dazhe Zhao and Jinzhu Yang

Abstract. An approach of medical image retrieval by combining diagnosis text and image content features is presented in this paper. In this approach, the system firstly extracts fusion features and semantic content from images and related text in medical resource database. In the second stage, a multi-modal query strategy is generated based on the semantic model and medical ontology, and then the mixed retrieval rendering is interactively shown in intelligent user interface. We evaluate our approach on real medical image data collected from the hospital, and this method achieves encouraging results which are basically satisfying the clinical needs.

Keywords: Text retrieval. Medical ontology. Content based image retrieval. Semantic space

1 Introduction

With the use of digital medical image equipment and electronic medical information system becomes more common in hospital, so does the need for effective retrieval and knowledge discovery in massive medical data. The patients' medical images are essential for doctors to make right and immediate diagnosis in clinic. It will be more helpful to support similar medical image cases as reference or learning source for the young radiologists or less experienced clinical doctors. Therefore, it is necessary to build a medical image database which can support the image data organization, storage, retrieval and analysis [17]. It can promote the development of clinical reasoning and evidence-based medicine so as to improve the quality of the individual diagnosis and treatment.

[1] Wei Li (✉)
Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang, China
e-mail: l-w@neusoft.com
Bo Li, Dazhe Zhao Peng Cao and Jinzhu Yang
Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang, China
e-mail:{l.b, zhaodz, cao.p, yangjinzhu}@neusoft.com

Medical image database, which is based on the database management, image processing, computer network and medical domain knowledge, is a database technology to support the medical image data organization, storage, retrieval and mining effectively, so that medical image resource can be fully utilized [18,19]. Its development will contribute to the process of clinical diagnosis, medical education and research. Many works has been done on medical image retrieval. Generally speaking, there are two main approaches to medical image retrieval from medical image database, that are text-based medical image retrieval and content-based medical image retrieval.

- Text-based medical image retrieval is implemented by annotating the images with people's subjective impression of the image [13, 11]. Experts make description text about the disease representation of the full images, image ROIs (Region Of Interesting), physical features and image attributes including modality, size, storage format and so on. The visual characteristic of the image is not considered in this way. Text-based image retrieval has a highly efficiency and fully expression of the semantic content in images. However, manual annotations of images are labour intensive, time consuming and unscalable.
- Content-based medical image retrieval is completely different from the text-based image retrieval, which indexes images by visual characteristics rather than text descriptions [4, 7, 12, 14, 15]. Visual feature extraction and selection are critical to the performance of content based image retrieval. Query by example is one common way in content based image retrieval system that user gives an image as a sample. Then the system finds matched images sorted by similarity scores and renders them for user. However, the semantic in an image is usually too complex to distinguish the user's intention, and this approach relies on the accuracy of feature similarity comparative. So the result usually contains the unexpected images for users.

Although each of the above techniques are used in many success applications, such as Google, Baidu etc, the query based fusion features in image retrieval is still a challenging research problem in the literature especially in the medical domain. In clinical practice, accuracy is the first thing and an import factor in all the hospital activities. This paper presents a multi-modal retrieval method that combines structure information retrieval, keywords based text retrieval and content-based image retrieval together. This work is all based on one medical resource database, which contains all the medical data we need. This approach greatly improves the retrieval accuracy and efficiency which takes advantage of text and content based image retrieval techniques.

The remainder of this article is organized as follows. In Section 2, we discuss some related work briefly. In Section 3, the outline of our approach is presented. We will describe the fusion semantic space in Section 4. Multi-modal retrieval process and performance study are introduced in Section 5 make conclusion in Section 6.

2 Related Works

There are many research works on image retrieval domain and most of the algorithms are applied on the general images. It is less on medical image relatively. We discuss some useful ones in clinical practice. The popular prototypes in previous work are listed as follows.

Medical Image Resource Center (MIRC) is launched by RSNA, which is based on text based image retrieval technology [11, 13]. It supports collaboration of multiple image databases which are managed by each provider. Various types of digital data can be retried including teaching files, clinical and technical documentation, electronic display and image data set.

Many well-known general content-based image retrieval systems have been developed, such as VisualSEEK developed by Columbia University [14, 15], QBIC developed by IBM [4], Photo-Book developed by MIT [10] and so on. In medical image domain, there are also some good prototypes. ASSERT is developed by Purdue university and Medical College of Wisconsin [12]. University Hospital of RWTH Aachen develops IRMA [7]. CasImage is developed by Geneva University [3] and DePaul University develops BRISC system [6].

These prototype are mostly in lab experiment other than in clinical process of the hospitals, and they do not make the combination of text and image content for searching. The approach in this paper as our knowledge is the emerging one in medical image for clinical practise, which integrates text and image into one semantic space to search in a process.

3 An Overview of Our Approach

Before we introduce the fusion semantic model and multi-modal query strategy, we first provide an overview of how the two steps are applied consequently to accomplish the overall task of combing text and content based image retrieval.

As shown in Figure 3.1, four parts are included in our approach. Semantic space is constructed from medical image database, and supports the semantic content queries. Search engine has the core role of multi-modal retrieval process, which is based on the knowledge base. User can interactively communicate with system in the upper level.

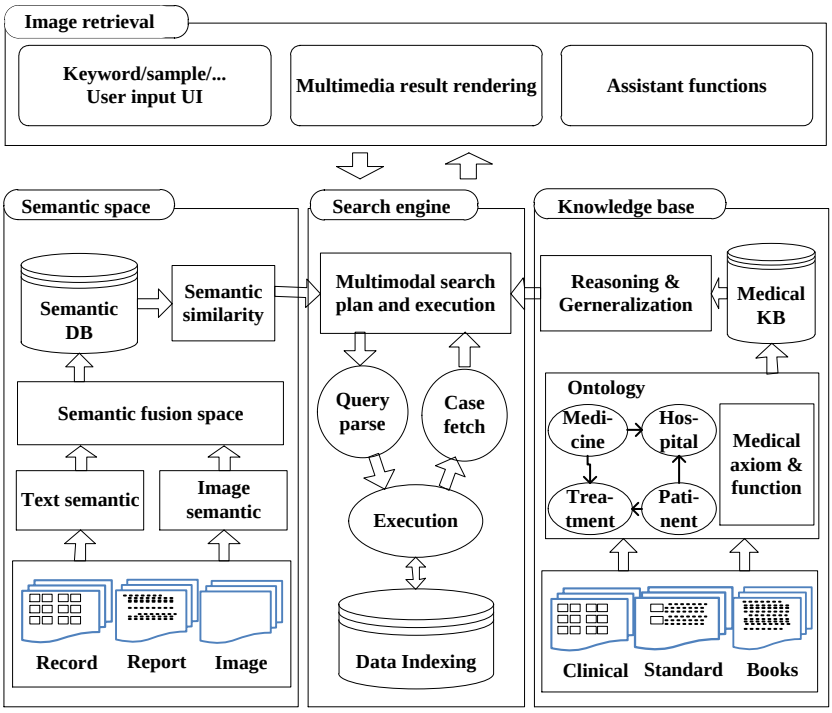


Fig. 3.1 Structure of our approach

The main two steps of our approach can be outlined as follows:

- **Fusion semantic space:** Fist of all, we extract the text semantic and image visual feature from the medical record, report and images. Then latent semantic analysis (LSA) is used to merge the text and image features as a semantic fusion space.
- **Multi-modal query strategy:** The search engine is the connector of user retrieval request, semantic space and knowledge base. User's search request is parsed with knowledge base which uses reasoning and generalizing on the request to generate a new search plan. The matched cases indexed in the semantic space are loaded to user.

This paper contributes two main innovations in medical image retrieval.

1. Our approach makes the search on fusion semantic space combing the text semantic and visual content of medical images.
2. We relax the user request by knowledge base which is based on medical ontology that we build in the paper.

According to the experimental results, the concurrency and response time of our approach basically satisfies the clinical applications.

4 Fusion Semantic Space

Fusion semantic space is based on the text semantic feature and image visual feature.

4.1 The Features

Text semantic feature extraction can be divided into two steps that are semantic term selection and feature vector generation. TF-IDF is a feature that commonly is used in information retrieval applications. TF-IDF feature considers the frequency of terms in the text. But it is not suitable in medical text because the terms occur in the medical report or record seldom, and the frequency of term may not be strong evidence of importance in documents. Usually, the term in the medical text plays determined factor of the main meaning. So the boolean model can be used in the text feature extraction.

We use binary vector to expression the feature of medical text. We define the function F: $F(i, j) = 1$ if the document d_j has the key term k_i. Otherwise the F equal zero. The boolean model has two shortages in application. One is that the similarity between two documents is not easy to be sorted. The other one is Boolean model doesn't support the partly comparative on two feature vectors. We make a new binary process on the TF-IDF method as equation 4.1.

$$\begin{aligned} w_{i,j} &= btf_{i,j} \cdot idf_j = btf_{i,j} \cdot \log_2\left(\frac{N}{df_j}\right) \\ that\ btf_{i,j} &= \begin{cases} 0 & if\ F(i,j) = 0 \\ 1 & if\ F(i,j) = 1 \end{cases} \end{aligned} \tag{4.1}$$

the df_j is the number of documents that contains the term j, N is the total documents size. The equation above decreases the affection of term frequency by binary the frequency of term in TF-IDF model. This model puts more attentions on the special term and fewer attentions on the common ones.

In order to fully describe medical images visual content, four different features are extracted from the medical images in this paper, which are gray level histogram, Haralick texture; shape context and FFT based frequency features [2].

- Gray-level histogram is the most common feature of image. Medical images are usually grey images with high level, for example, CT image has 4096 grey levels. We suppose that image with gray levels in the range [0, L-1], which are divided into L equal parts. Then we can get the feature vector of the ratio of pixels with the gray-level value in each L parts.
- Gray level co-occurrence matrix is the basis for the Haralick texture features [5]. This matrix is square with dimension N_g, that N_g is the number of gray le-

vels in the image. Element [i, j] in the matrix is generated by counting the number of times a pixel with value i is adjacent to a pixel with value j and then dividing the entire matrix by the total number of such comparisons made. Haralick described 14 statistics that can be calculated from the co-occurrence matrix with the intent of describing the texture of the image. In this paper, we use the normalized co-occurrence matrices that are calculated in four directions and five displacements (d = 1, 2, 3, 4, 5) to generate 20 matrices per medical image.

- The shape context is intended to be a way of describing shapes that allows for measuring shape similarity and the recovering of point correspondences [1, 9]. The Sobel operator is used for edge detection in our approach. Diagram of log-polar histogram bins are used in computing the shape contexts. We use 6 bins polar radius and 12 bins for circle divisions [8]. Therefore, we get a 60-dimensional histogram vector as the shape feature.
- FFT based frequency feature consists of coefficients obtained in the frequency domain after each image is transformed by the FFT. The coefficients in each quarter are then arranged, in which the coefficients are divided into four groups and each group is subdivided into three levels [16]. A feature vector with 10 dimensions is finally obtained. Since the FFT coefficient values contain both real and imaginary parts, its absolute value is computed into the feature vector.

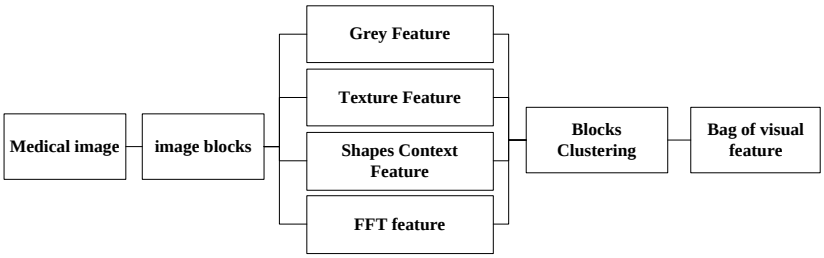


Fig. 4.1 Image feature extraction flow

We use the four visual features on every block of all images using visual topic word model, and get the cluster center of every image type [2]. So each medical image can be represented as a visual feature vector eventually. The process flow chart is shown as Figure 4.1. The clustering algorithm used in this paper is k-means.

4.2 Semantic Space

We study semantic space based on LSA to fuse the features of multi-modal data. The feature matrix is built by combining the medical text and image features, and then the medical semantic space is constructed with dimension reducing. There are

coupled relationships between the medical text and image information, so that the latent semantic layer is created to describe multi-modal semantic information.

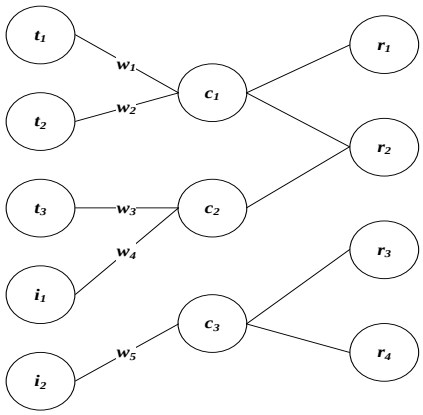


Fig. 4.2 Fusion semantic model

As shown in Figure 4.2, t_x is the text feature, i_x is the image feature, w_x is the weight, c_x is the element of latent semantic layer, r_x is medical image record (case), we can define the semantic space as following equation.

$$N = (W_t \cdot N_t, W_i \cdot N_i) = U\Sigma V^T \tag{4.2}$$

where W_t is the text weight vector, W_i is the image weight vector, N_t is the text feature matrix, N_i is the image feature matrix. In the semantic space, the similarity between two medical cases can be computed by the distance with the transformation of Singular value decomposition (SVD) which is shown as the equation 4.2.

5 Multi-modal Query Strategy

Multi-modal information retrieval in this work uses the medical ontology built from real clinical data that make higher accuracy on the search result in medical.

5.1 Ontology and Data Index

Ontology formally represents knowledge as a set of concepts and the relationships between those concepts within a domain. It can be used to reasoning about the entities and may be used to describe semantic relation in the domain.

According to the clinical needs, there are mainly three base classes defined in our medical ontology, that are person, institution and medicine. The roles of

people in the clinics, e.g. doctor, nurse, patient, are defined as sub-classes based on the person base class. We define hospital, department and other sub-organization sub-classes based on institution for effective management of the person information. Disease and clinical manifestation are defined based on medicine. A variety of object properties and data properties are defined to introduce the relations between these classes. The main classes and properties of our medical ontology are shown as Figure 5.1 (the secondary definitions are shown in the figure).

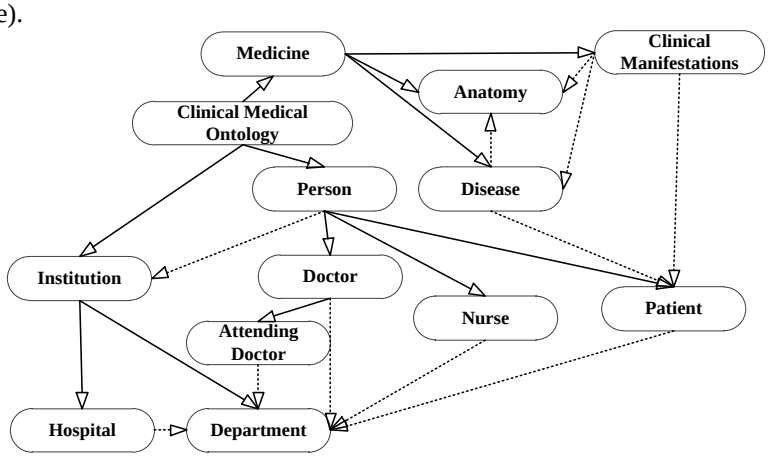


Fig. 5.1 The definition of medical ontology

Every data item of the medical reports is defined as a property of the patient, and more than 600 properties chosen from more than 20 information sections are defined for a patient in our ontology. We extract and store the image header information of each image as an ontology property in the image organization process simultaneously. The data property is defined, so that medical images are associated with their reports by the medical ontology.

With the help of domain experts, we extract the medical knowledge according medical literatures. Different properties are defined for disease including Chinese name, English name, ICD10 and so on. We also defined the object properties between disease and clinical manifestation, disease and department, disease and patient.

All ontology instance is indexed as three types of medical resource data including structured text such as demographic information, blood test results, free text data such as diagnostic reports, progress notes and medical image data such as CT, MR, PET and so on. Different types of query index structures are defined according to different data types that include B+ tree for structure data search, inverted files for full text search, and high dimensional tree for feature vector search. We can search clinical medical image or reports by data items of the reports, and the information about the disease.

5.2 Multi-modal Retrieval

Multi-modal retrieval in medical image database requires a combination of clinical image and clinical text retrieval process. The procedure of multi-modal search is shown as Figure 5.2. The retrieval process is divided into three phases: query condition phase, query plan generation and execution phase, query results visualization phase.

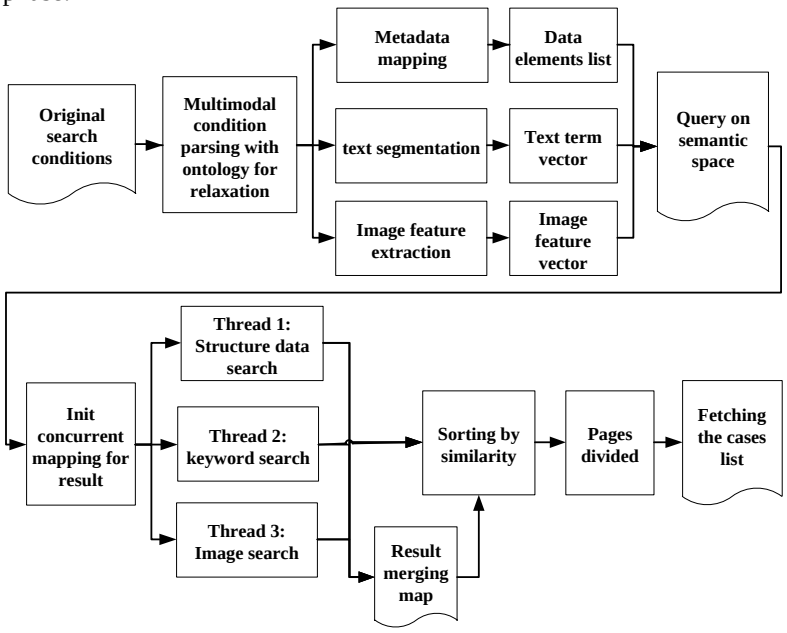


Fig. 5.2 Conditions analysis and query plan execution

- **Query condition phase:** Three types of input data on user interface are supported in our method, including structured table attributes, keyword and example image. Once the condition data is passed to the algorithm, the condition parsing module will make correspondent sub-procedure to parsing the input data. Medical ontology is used in the process to relax the query request, e.g. user wants to query "pulmonary disease, cough and inflammation", and the concept will be translated into all the real disease on pulmonary that has these symptoms. Furthermore, the algorithm will give the feedback to user and the user can make his/her query content more clearly.
- **Query plan generation and execution phases:** Once the client submits a request, the search condition may be one or more forms as table attribute, text term vector or image feature vector, then they are transferred into an executable query as SQL, high dimension search tree indexing and sorting procedures in

the semantic space. The query results are merged and sorted by multiple memory mappings.

- **Query results visualization phase:** This stage will focus on visualization of the selected search results as easy use style with text abstract, image and table. The implementation of user interface in this paper is represented by web technology.

5.3 Approach Evaluation

A medical image database system is implemented for the approach evaluation and has been deployed in hospital clinical process. The system user interfaces of search and result visualization are shown in Figure 5.3 and Figure 5.4.

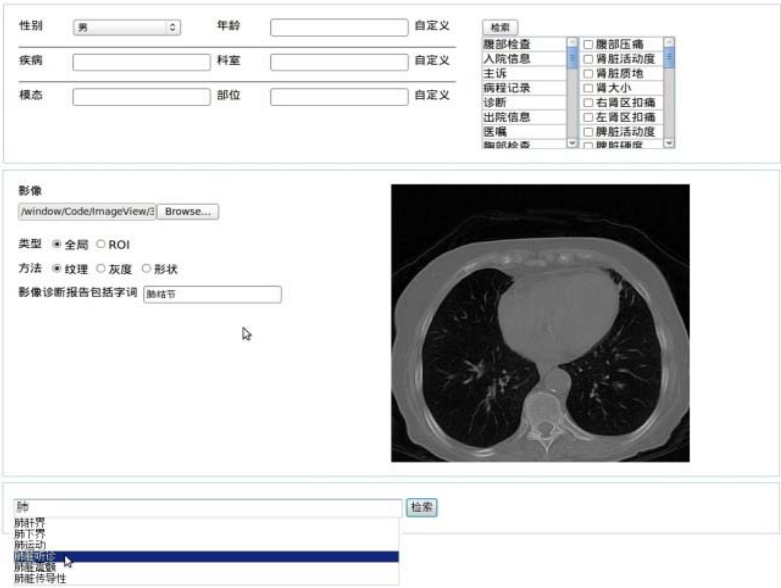


Fig. 5.3 The user interface for multi-mode condition

The medical image cases used in this paper are all obtained from the hospital about 4,000 medical cases, including patient image data, diagnostic reporting and clinical data. The total number of the DICOM images with 512*512 resolution stored in the database is 72,122.

According to the experimental results, it shows that the accuracy of content-based image retrieval only is 78%, but the accuracy can be promoted up more than 90% using multi-modal retrieval method with medical ontology.

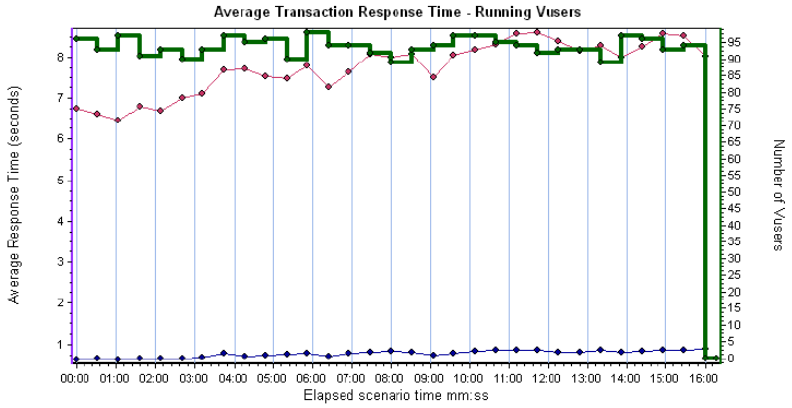
In this paper, LoadRunner tool is used to test the performance of the medical image resource database system. There are 100 concurrent virtual users to be set and the test results are shown in Figure 5.5. Figure 5.5 shows that the average response time for image retrieval is 0.77 second. Searching concurrency and response time are basically satisfying the real application goals. The time-consuming steps focus on the sample image feature extraction and searching on the high dimensional index database.

Fig. 5.4 Result rendering for medical image retrieval

6 Conclusions

This paper presents an approach of combining the text and content based image retrieval on medical image database. At the same time, a unified semantic space model is constructed by multi-modal data content with medical ontology. In this study, a medical image database system is designed and implemented for approach evaluation. According to the requirement of practical application and research, the retrieval system provides three kinds of searching ways in this paper including structure data query, keyword based full-text retrieval, and content-based image retrieval. User can give keywords related to diagnosis reports and get the medical

image cases about the keywords quickly by using the full-text retrieval technology. The system also provides a kind of retrieval that combining diagnostic semantic and image content. The image retrieval results in this method contain both the text and image semantic, and we will improve the performance in much more big data in the future.



Color	Graph	Scale	Measurement	Graph's Minimum	Graph's Average	Graph's Maximum	Graph's Median	Graph's Std. Deviation
	Average Transaction Response Time	1	get_img	6.451	7.742	8.597	7.987	0.628
	Average Transaction Response Time	1	query_img	0.625	0.77	0.886	0.786	0.079
	Running Vusers	1	Run	0	87.969	98	94	22.837

Fig. 5.5 Performance test of the approach

Acknowledgements This research is supported in part by Chinese NSF grants (61172002 and 61001047) and the Fundamental Research Funds for the Central Universities under grant (N120518001 and N110718001) and Liaoning Natural Science Foundation under grant 2013020021

References

[1] Belongie S., Malik J., and Puzicha J.. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4):509-522, 2002.

[2] Bo L., Wei L., Cao P., and Dazhe Z.. Multi-feature based medical image classi_cation and evaluation. In the 3th International Conference on Intelligent Computing and Intelligent Systems, pages 669-673, 2011.

[3] Dy J. G., Kak C. E, A., Broderick L. S., and Aisen A. M.. Unsupervised feature selection applied to content-based retrieval of lung images. IEEE Transactions on Pattern analysis and Machine Intelligence, 25(3):373-378, 2003.

[4] Flickner M. and Sawhney N. H.. Query by image and video content: the qbic system. IEEE Computer, 28(9):23-32, 1995.
[5] Haralick R., Shanmugam K., and Dinstein I.. Textural features for image classi_cation. IEEE Transactions on Systems Man and Cybernetics, 3:610-621, 1973.
[6] Lam M., Disney T., Raicu D., Furst J., and Channin D.. Brisc: An open source pulmonary nodule image retrieval framework. Journal of Digital Imaging, 20(Supplement 1):63-71, 2007.
[7] Lehmann T., Guld M., and Thies C.. Content-based image retrieval in medical applications. Methods Inf. Med., 43(4):354-361, 2004.
[8] Mori G., Belongie S., and Malik J.. Shape contexts enable e_cient retrieval of similar shapes. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, pages 723-730, 2001.
[9] Mori G., Belongie S., and Malik J.. E_cient shape matching using shape contexts. IEEE trans. Pattern Analysis and Machine Intelligence, 27:1832-1837, 2005.
[10] A. Pentland, R. Picard, and Sclaro_s. Photobook: Content-based manip ulation of image databases. International Journal of Computer Vision, 18(3):233-254, 1995.
[11] L. Prevedello, K. Andriole, and Khorasani R.. Informatics in radiology: integration of the medical imaging resource center into a teaching hospital network to allow single sign-on access. Radiographics, 29(4):973-979, 2009.
[12] Shyu C., Brodley C., Kak A., Kosaka A., Aisen A., and Broderick L.. Assert: a physician-in-the-loop content-based retrieval system for hrct image databases. Computer Vision and Image Understanding, 75(1-2):111-132, 1999.
[13] Siegel E., Channin D., and Perry J.. Medical image resource center 2002: an update on the rsnas medical image resource center. Journal of Digital Imaging, 15(1):2-4, 2002.
[14] Smith J. and Chang S.. Tools and techniques for color image retrieval. In Storage & Retrieval for Image and Video Database IV, IS&TSPIE Proceedings, page 426437. In: Sethi, I.K., Jain, R.C., 1996.
[15] Smith J. and Chang S.. Visualseek: a fully automated content based image query system. In Proceedings ACM Multimedia, page 8798, 1996.
[16] Tachphetpiboon S. and Amornraksa T.. Applying _t features for. _ngerprint matching. In IEEE Wireless Pervasive Computing, pages 1-5. 2006 1st International Symposium on, 2006.
[17] Wei L.. Construction of extensible markup language-based chinese mr brain imaging database. Chinese Journal of Medical Imaging Technology, 26(9):1640-1643, 2010.
[18] Wei L., Bo L., Dazhe Z., Jinzhu Y., and Qin M.. A method of metadata based medical resource integration. In CARS 2011: Computer Assisted Radiology and Surgery, 2011.
[19] Yang Y.. The established of database about medical imaging based on dicom. Changchun University of Science and Technology, pages 2-4, 2008.