KNOWLEDGE AUGMENTED MEDICAL IMAGE RETRIEVAL SYSTEM

Yao Shieh
Chang Gung University, Taiwan
yshieh@mail.cgu.edu.tw

Mengkai Shieh
Russian Medical Academy of Postgraduate Education, Russia
mks@alumni.stanford.edu

Chien-Hung Chang
Linkou Medical Center and College of Medicine, Chang Gung University, Taiwan
cva9514@gmail.com

Tsong-Hai Lee
Linkou Medical Center and College of Medicine, Chang Gung University, Taiwan
thlee@adm.cgmh.org.tw

Scott Goodwin
University of California at Irvine, Irvine, USA
sgoodwin@uci.edu

Abstract:

We are living in an era of information explosion. Medical images are generated at an accelerating rate. A more effective information technology to deal with storage and retrieval of such huge amount of medical image data is needed. The purpose of this paper is to demonstrate by presenting a concrete example that a knowledge augmented medical image retrieval system by means of automated feature extraction is possible. It provides not only decision support in the clinical setting but an education/research platform upon which issues regarding computer-aided diagnosis and inter-observer variations among radiologists can be addressed systematically and effectively. It inspires more productive man-computer collaboration by bringing computer intelligence to new heights through knowledge transfer to meet the challenge of information explosion.

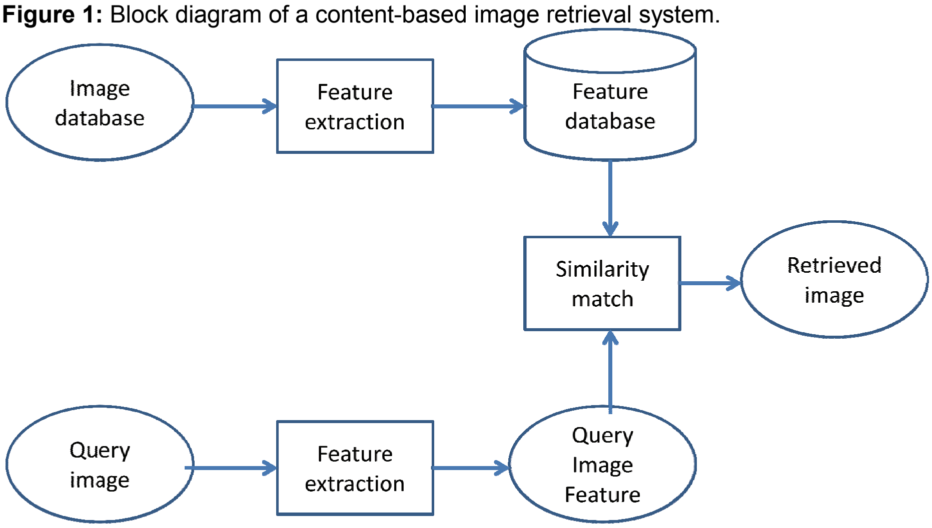
Keywords: *knowledge, information technology, content based image retrieval, feature extraction, database, education*

1. INTRODUCTION

We are living in an era of information explosion. New information has been continuously generated at an ever increasing speed. It has become an important skill to locate the relevant and up-to-date information/knowledge to help solve the problem that we are facing in a timely fashion.

Image database and its associated retrieval are different from text-based document database. Researchers have investigated various techniques to build image databases and to retrieve relevant images thereafter more efficiently.[1,2,3,4] The term of "Content-based image retrieval" (CBIR) has been widely adopted to emphasize the non-text-based approach. Over the years, the generally accepted model of a content-based image retrieval system consists of three components: (1) image database, (2) feature extraction, and (3) similarity match, as shown in figure 1. The broad spectrum of techniques for CBIR that have been reported in the literature can be classified into two categories: the generic approach and the application specific approach. The former tries to solve a wide array of problems by appealing to common features shared by them. For instance, Rallabandi et al. reported a system that is targeted for a collection of professional photos, web images, animations, video and clips (Rallabandi, 2008). The strength of this approach is that it is highly automated. The drawback is that there are still many applications that do not fit into this one-size-fits-all mentality. The latter, primarily in medical applications, tries to solve a specific problem by enlisting domain expert's explicit guidance in the form of human intervention. Although it meets the specific requirements of a niche application, it is often very labor intensive and time consuming (Cai, et al., 2008). The argument of justification is that the clinically useful information in an image typically consists of image intensity variations in highly localized regions which is not possible to be extracted by the then state-of-the-art automatic image segmentation techniques (Shyu et al., 1999). What it implies is that to add a new image to the image database, a physician needs to manually delineate the "pathology bearing regions" (PBR) and relevant landmarks as part of the image database buildup effort. Likewise, a similar effort may incur when a query image is presented to the CBIR system.

Figure 1: Block diagram of a content-based image retrieval system.



However, such a human-in-the-loop approach proves inadequate to build a large scale medical image database given that new images are generated daily at an accelerated rate. Recall that in not-so-far-away past, full-time experts are hired to review newly published journal papers and then add relevant keywords and possibly additional comments to classify each individual document into meaningful categories. However, such an approach proves non-practical because the expert's review rate falls far behind the production rate of new documents. As a result, many automated data mining algorithms have been proposed to let computers to do the task of automatic keyword generation and categorization. By the same token, the task of feature extraction has to be automated if a large scale medical image retrieval system is to be implemented.

The purpose of this paper is to demonstrate by presenting a concrete example that a special-purpose, medical image retrieval system by means of automated feature extraction is possible. The advancement of computer hardware, software and image processing technology promises to narrow down the gap between human intelligence and computer intelligence. Presumably more and more intelligent tasks can be offloaded to computers in the future. The key is to transform the domain expert's knowledge from abstractive concepts into a form that is comprehensible and therefore executable by computers.

2. BASICS OF CONTENT-BASED IMAGE RETRIEVAL

In general, a content-based image retrieval system can be modeled to consist of three components: (1) image database, (2) feature extraction, and (3) similarity match. More sophisticate CBIR system may include an additional component of relevance feedback.

Typically, an image set contains a large number of data elements – commonly called pixels for 2D image and voxels for 3D images. A number of features are measured and calculated from the original image data set. The features are carefully chosen so as to greatly reduce the dimensionality of the data and at the same time minimize the information loss. When the user presents a query image, its corresponding feature vector, F_{IMG_Q}, is calculated, and compared with F_{IMG_K}, the feature vector of each individual image stored in the image database. A set of images that are most similar to the query image in accordance with a certain similarity criteria are output to the user. The user has the option of flagging each of the output images as either satisfactory or not as a feedback to the system which may revise its outputs. Such interaction continues until the user is satisfied or gives up.

3. KNOWLEDGE AUGMENTED IMAGE RETRIEVAL SYSTEM

In this paper, we present an acute stroke CT image retrieval system which is based on the automated Alberta Stroke Program Early CT score (ASPECTS) scoring system that we have developed within our medical center. The flow chart of the automated ASPECTS scoring system is shown in figure 2.

Spatial realignment is essential for brain image analysis involving multiple subjects. The goal of image registration/segmentation is to outline various regions of interest (ROI) on the image under investigation ("target image") accurately and automatically. We have relied on the anatomic knowledge to create the template for the purpose of segmentation. Our 2D template at the ganglionic level that serves as the "normalized reference image" was created by taking the average of brain images from patients who presented for workup of headache at our medical center. They are shown in figure 3. The image registration used in our algorithm is a two step process – initial global registration followed by local refinement. Initially, the Reference Image, along with its corresponding template of ROIs, is scaled by a factor of $W_{Target}/W_{Reference}$ and $H_{Target}/H_{Reference}$ in X and Y directions respectively so that it has the same size as the target image. $W_{Reference}$ and $H_{Reference}$ are the width and height of the rectangle that circumscribes the reference image, whereas W_{Target} and H_{Target} are their counterparts for the target image (i.e., the image under investigation).

Figure 2: Block diagram of automated ASPECTS scoring method.

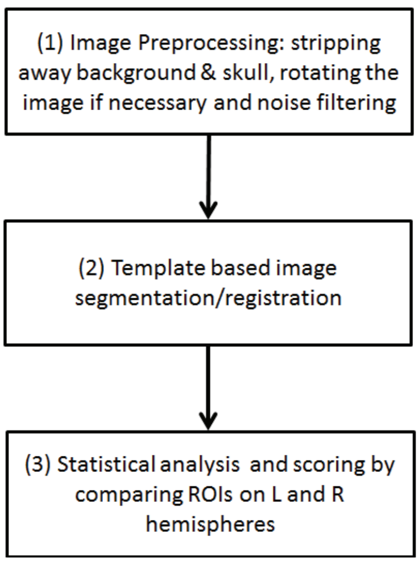
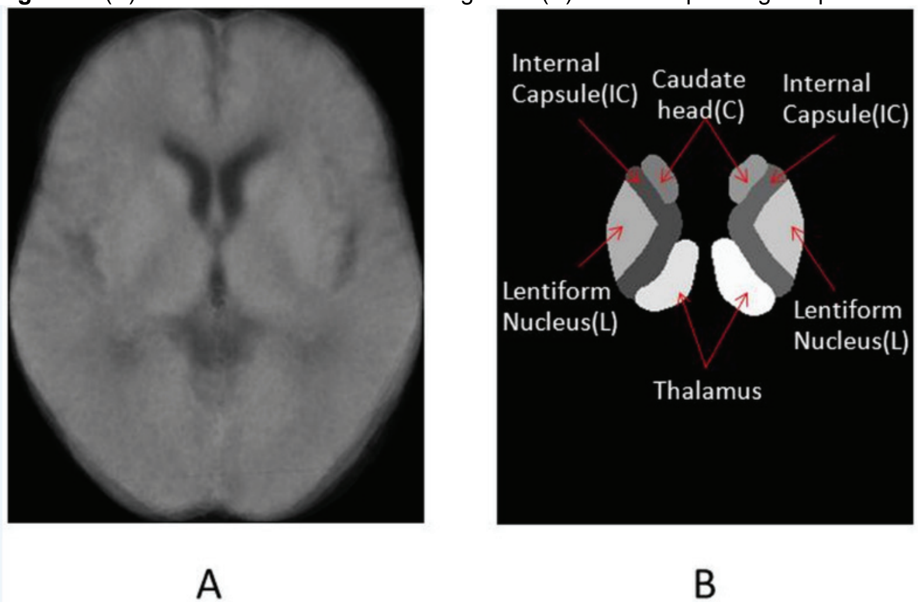


Figure 3: (A) The normalized reference image and (B) the corresponding template.



Next, individual ROIs of the template are allowed to move within a small range of displacement in both X and Y directions with respect to their initial positions to achieve maximum similarity between the template and target image. Automatic segmentation of the target images into ROIs is a prerequisite for subsequent automatic feature extraction. The primary purpose of enlisting human intervention by content based medical image retrieval systems that have been reported in the literature is to delineate ROIs manually due to the difficulty of automatic segmentation. The success of automatic segmentation in our particular application can be attributed to the following factors: (1) the ten ROIs are clearly defined by the ASPECTS method, and (2) ROIs maintain a relatively rigid geometric relationship with respect to each other – a property commonly found in neuroimages - and therefore, a template model of individual ROIs connected by elastic springs is suitable for automatic registration.

3.1 Automatic feature extraction

Our analysis/scoring method focuses on the detection of signs of ischemia-induced X-ray attenuation for each of the ten ROIs. We divide the ten ROIs defined by the ASPECTS scoring method into two groups. Group 1 contains caudate head (C), internal capsule (IC) and lentiform nucleus (L) regions. Group 2 contains insular ribbon region (I) and middle cerebral artery, M1 through M6 regions. The two ischemic signs that are used in group 1 are diffuse hypoattenuation and focal hypoattenuation. The high-level semantic feature that is used for the detection of diffuse hypoattenuation is the comparability of image intensity of an ROI versus its counterpart on the contralateral side. Mathematically, the two-sample Kolmogorov–Smirnov (KS) statistic is computed by comparing image intensities of all pixels that are contained in contralateral ROIs. The high-level semantic feature that is used for the detection of focal hypoattenuation is the size of the largest cluster consisting of pixels whose image intensity is lower than $\mu- n\sigma$, where μ and σ are the image intensity mean and standard deviation of the ROI on the unaffected side of the brain.

The two ischemic signs that are used in group 2 are diffuse hypoattenuation and degradation of contrast between gray matter (GM) and white matter (WM). The high-level semantic feature for diffuse hypoattenuation is the same as that for group 1.

The high-level semantic feature that is used for the detection of contrast degradation is the comparability of GM-WM contrast of an ROI versus its counterpart on the contralateral side.

The ASPECTS scoring scheme defines ten ROIs, each of which can be either 1 (normal) or 0(abnormal). Our automated scoring method utilizes two features, $<f_1, f_2>$, to characterize each ROI, and then assigns its corresponding score, $(s)_{Alg}$, based on their values. In addition, each ROI is given a neuroradiologist's score, $(s)_{Gold}$, which is regarded as the gold standard. Therefore, each ROI has four features: $<f_1, f_2>$, and derived score, $(s)_{Alg}$, as well as the neuroradiologist's score, $(s)_{Gold}$. Altogether, an image dataset, IMG_K, can be mapped to a feature point, F_{IMG_K} = (< f_{11}, f_{12}, $(s_1)_{Alg}$, $(s_1)_{Gold}$, ..., f_{i1}, f_{i2}, $(s_i)_{Alg}$, $(s_i)_{Gold}$, ..., $f_{10,1}$, $f_{10,2}$, $(s_{10})_{Alg}$, $(s_{10})_{Gold}$ >$)_{IMG\_K}$, in the 40-dimensional feature vector space.

3.2 Similarity match

When the user presents a query image, its feature vector, F_{IMG_Q}, is computed, with $(s_i)_{Gold}$ being left blank. To carry out similarity match, only the score features are used. Since there are two kinds of scores, two different outputs can be generated. But the principle of similarity measurement remains the same. Suppose that we choose to use gold standard scores as the reference, the similarity measurement, $(Similarity_{Q,K})_{Gold}$, between the query image, IMG_Q, and an arbitrary image stored in the image database, IMG_K, is defined below:

$$(Similarity_{Q,K})_{Gold} = \sum_{i=1}^{10}(((s_i)_{Alg})_Q - ((s_i)_{Gold})_K)$$

In essence, the similarity counts the number of identical regional scores between the query image and an arbitrary IMG_K stored in the image database. For instance, suppose that the algorithm calculated score of IMG_Q and the gold standard score of IMG_K are (0 1 1 1 0 1 1 1 1 1) and (1 1 1 1 1 1 1 1 1 0) respectively. Then the similarity between them is 7. Likewise, the similarity measurement, $(Similarity_{Q,K})_{Alg}$, using the algorithm-generated score as the reference can be defined as

$$(Similarity_{Q,K})_{Alg} = \sum_{i=1}^{10}(((s_i)_{Alg})_Q - ((s_i)_{Alg})_K)$$

4. PERFORMANCE EVALUATION

Recall and precision are two commonly used metrics to evaluate the performance of a information retrieval system. We have created a prototype image database consisting of 50 acute stroke cases, spanning from no stroke throughout severe stroke. We did an experiment by arbitrarily pulling out 5 cases from the image database to be used as query images. The preliminary results showed around 80% of recall as well as precision. Improvement on recall and precision actually relies on the performance improvement of our automatic feature extraction and subsequent scoring algorithms. The current rates of accuracy of our algorithms, when averaged over ten ROIs is 76% sensitivity and 84% specificity. We have been trying to improve the accuracy of our algorithms by incorporating additional features that have high power of differentiating between normal and acute stroke conditions.

A potential method of improving the recall rate is to relax the constraint on similarity criterion. For instance, suppose that the algorithm-calculated score feature for the query image is <1 0 0 1 1 1 1 1 1 1>. Instead of retrieving images which have identical score, we may want to retrieve images that have similar but not identical scores, such as <0 0 0 1 1 1 1 1 1 1> or <1 1 0 1 1 1 1 1 1 1> and so on. Nevertheless, this needs to be done on a selective basis because there exist many possibilities. Furthermore, the precision rate may decrease while we try to increase recall rate, if it is not done properly. There exist one occasion, however, when we may want to relax the similarity constraint intentionally. Physicians constantly face the dilemma of differential diagnosis in routine clinical settings. It is very useful for physicians to view similar cases with subtle differences that end up with different diagnosis.

5. CONCLUSION

We have explored the idea of designing a high performance image retrieval system by incorporating the domain experts' knowledge into the automatic feature extraction algorithms. Specifically, we rely on the expert's knowledge to (1) create a template for reliable segmentation into meaningful ROIs, (2) locate potentially pathology bearing pixels by simulating human detection of their characteristic patterns in terms of gray-level variations, and (3) determine whether these suspicious pixels collectively form a plausible pathology signature by simulating human hypothesis testing that takes contextual information into consideration. A special purpose image retrieval system is built upon the automatic ASPECTS scoring system that we had developed before for the purpose of decision support in acute stroke care. Preliminary experimental results demonstrate its promising potential in fulfill this purpose.
Moreover, such an image retrieval system may play a very important role in setting up a platform upon which issues such as inter-observer variations among physicians and discrepancy between radiology

experts and state-of-the-art computer-aided diagnosis algorithms can be addressed in a very systematic and efficient manner. For instance, each image dataset has an algorithm generated score vector, S_{Alg}, and a radiologist's score vector, S_{Gold}. Image datasets that have discrepancy between these two sets of scores can be pulled out and examined closely so that the computer algorithms can be enhanced on a regular basis. Image datasets can also be presented to a number of board certified radiologists and then compare their readings to study inter-observer variations and/or revise the gold standard scores accordingly.

We are living in an age of information explosion. Medical images are generated at an accelerating rate. A more effective method of dealing with storage and retrieval of such huge amount of medical image data is needed. Knowledge augmented image retrieval paradigm proposed in this inspires more productive man-computer collaboration by bringing computer intelligence to new heights to meet the challenge of information explosion.

ACKNOWLEDGEMENT

This work was partially supported by grants from the National Science Council of Taiwan [101-2221-E-182-058].

REFERENCE LIST

1. L. H. Y. Tang, R. Hanka, and H. H. S. Ip. (1999). A review of intelligent content-based indexing and browsing of medical images. *Health Informatics J. 5*, 40-49.
2. H. Muller et al. (2004). A review of content-based image retrieval systems in medical applications—Clinical benefits and future directions. *Int. J. Med. Info. 73*,1-23.
3. H. D. Tagare, C. JaVe, and J. Duncan. (1997). Medical image databases: A content-based retrieval approach. *J. Am. Med. Inform. Assoc. 4*(3), 184-198.
4. Bankman (Ed.) (2000). *Handbook of Medical Imaging Processing and Analysis*. Academic Press.
5. 5.T.W.Cai, J. Kim and D.D. Feng. (2008). Content-based medical image retrieval. In D.D. Feng (Ed.) Biomedical information technology (pp. 83-113). Burlington, MA, USA, Academic Press.
6. V.P. S. Rallabandi and S.K. Sett. (2008). Knowledge-based image retrieval system. *Knowledge-based systems*, *21*, 89-100.
7. C. R. Shyu et al. ASSERT. (1999). A physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding, 75*(1/2), 111-132.