Content-based image retrieval strategies for medical image libraries

Ahmed M. Ghanem[a,b], M. Emad M. Rasmy[a], and Yasser M. Kadah[a,1]

[a]Biomedical Engineering Department, Cairo University, Giza, Egypt
[b]Faculty of Information Systems, Suez Canal University, Ismailia, Egypt

ABSTRACT

The objective of Content-based image retrieval (CBIR) in medical field is to permit radiologists to retrieve images of similar features that lead to similar diagnosis as the input image. This is different from other fields where the objective is to find the nearest images from the same category or match a part of an image. Therefore, such techniques cannot be directly applied in the medical field. In this study, a modified wavelet-based matching technique is introduced that is more robust to motion, noise, and brightness changes within the image. Also, we propose a description-based technique in which a semantic net is built in which each node represents a specific region and its spatial relation to other regions in the image. This semantic net can be considered as a hierarchical relationship tree with links between the nodes in the same level to describe the geographic relations between these nodes. Nodes contain region-specific features such as the moments of region boundaries in addition to local textural features. In the matching phase, the semantic net is built for the input image then used in the matching process. The matching process starts from the highest level in the hierarchical relationship tree for fast convergence.

Keywords: Digital libraries, content-based image retrieval, image classification.

1. INTRODUCTION

In the past few years, the area of content-based image retrieval (CBIR) has received an increasingly higher interest from multimedia database researchers. The techniques of this area provide a flexible means of searching a digital image based on the description of the desired image or class of images. This description may take several forms such as comparison to a reference image or even a set of qualitative features representing quantitative pictorial features.

CBIR has been developed and implemented into several commercially available multimedia database systems. Examples of such systems are the IBM QBIC system (IBM Almaden Research Center), the Virage System (Virage, Inc.), and Photobook System (the MIT Media Lab). In these systems, a feature vector is generated for every image in the database. This feature vector must be smaller than the original image. In order to retrieve images from the database, the feature vector is calculated for the input image and compared to the previously generated feature vectors to determine the vectors most similar to the input vector. The images corresponding to these vectors are subsequently retrieved from the image database. Hence, a key factor in the success of a given CBIR system implementation is the selection of the feature vector. The most fundamental criteria for this selection is the ability to provide a good representation of the image while maintaining minimum size and minimum calculation time for vector generation and matching.

In this paper, we address the problem of CBIR as applied to medical imaging. We develop two CBIR strategies for medical images to be integrated with PACS for retrieving the images that contain similar features in common and lead to same diagnosis as the input image or at least similar diagnostic features. Once the system finds the related images, the radiological reports and medical records of these images are retrieved and provided to the radiologist. This property allows radiologists to make use of previous case diagnoses in order to increase their diagnostic accuracy. The feature is also valuable for training junior radiologists.

[1] E-mail: ymk@internetegypt.com

Medical Imaging 2001: Image Processing, Milan Sonka, Kenneth M. Hanson, Editors,
Proceedings of SPIE Vol. 4322 (2001) © 2001 SPIE · 1605-7422/01/$15.00

The paper is divided into three parts. The first part contains the problem definition. The second part describes wavelet-based CBIR strategy and how it can be modified for use in medical field. In the third part we propose a suitable strategy based on semantic net. The last section contains the result and the conclusions.

2. PROBLEM DEFINITION

In the design of a CBIR system for medical field we are faced with a number of unique constraints. The following is the definition and analysis of these constraints:

1. *Local versus Global Features*: Using global parameters such as color histogram and texture values cannot provide good results in medical imaging. This is because in medical images the local texture, shape, and/or brightness of some small objects or parts may have a high diagnostic value [1].
2. *Large number of different imaging modalities/modes*: The sources of medical images are many. Examples of such sources are diagnostic ultrasound, X-ray computed tomography (CT), magnetic resonance imaging (MRI), digital subtraction angiography (DSA), positron emission topography (PET), magnetic source imaging (MSI), conventional X-ray, and computed radiography (CR) [2]. Moreover, within each modality, there are several modes by which the image is acquired. Each mode differ significantly from the other modes, e.g., contrast modes in MRI. Due to the wide differences between the physics underlying these modalities, It is rather difficult to design CBIR system that works for all modalities using the simple feature vector approach.
3. *User interactivity*: The system must allow the user to determine the part(s) with high diagnostic value since no automatic system has been able to do that successfully.
4. *Huge size of image database*: The technique must offer fast performance since a full-scale picture archiving and communication systems (FS-PACS) of a mid-size hospital (around 600-800 beds) would easily require 1 TBytes of digital data per year in its image archive or library.

Based on the constraints above, we will address a particular aspect of the problem at hand defined as building a CBIR system dedicated for MRI, CT and X-ray images. This system must deal with local features of the images and allow the user to interactively locate the regions with high diagnostic values. The system matching strategy must be fast and the extracted features must be suitable for storage in a database.

3. WAVELET BASED CBIR SYSTEM

Compared to Histogram based CBIR systems, wavelet based CBIR systems are more reliable for medical images. In histogram based systems, feature vector neglects location information. This drawback in histogram based systems may lead to considering two images to be very similar, even though they represent very different diagnostic information. Similar to histogram based systems, Global texture based systems also neglect location information. On the other hand, based on the nature of the wavelet transform, which preserves the location information, wavelet based CBIR systems can be considered more suitable to representing medical images.

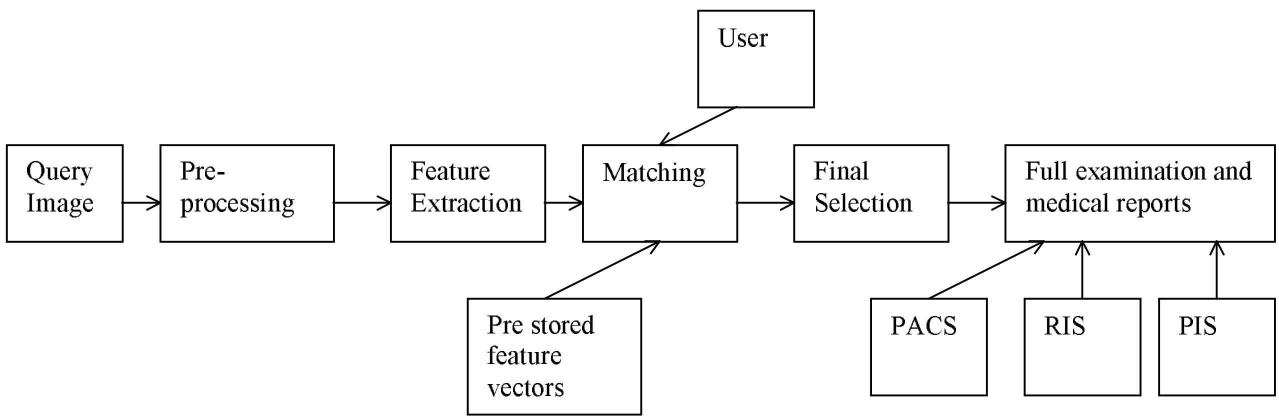


Figure 1. System block diagram

Searching the available CBIR literature, we observe a number of systems that stand out for their unique features. Examples of such systems are the system implemented by the University of Washington for multi-resolution image querying [4], and the system developed at Stanford University which was called WBIIS (Wavelet-Based Image Indexing and Searching) [3]. Both systems are considered wavelet-based CBIR systems. The system of the University of Washington is based on Haar wavelets, while Stanford University WBIIS is based on Daubechies wavelets. Both systems are designed for general-purpose image indexing and searching. As a result, their performance becomes very poor when applied on medical images since the constraints in the problem definition stated above are not properly addressed.

Here, we propose a wavelet-based system that is more suitable for medical CBIR systems. Figure 1 shows a block diagram of this system. As shown in this diagram, the system consists of the following components: preprocessing, feature extraction, matching, final selection, in addition to complete case images and medical report retrieval. Each one of these blocks will be described in more details in the following subsections.

3.1 Preprocessing

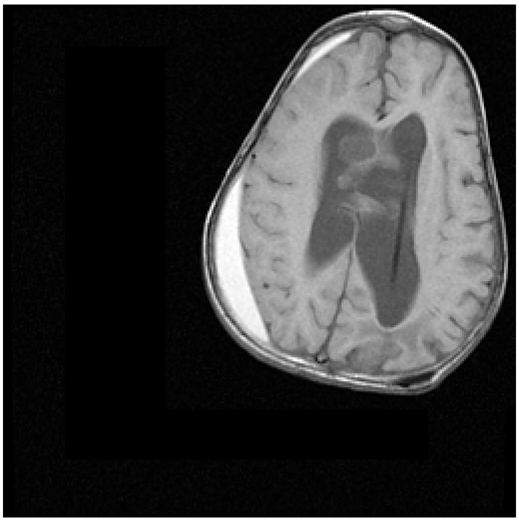
In our system, we work with CT/MRI/X-ray images stored in PACS. The size of the images used may be up to 1024×1024 with up to 12-bit gray scale. In the first step of preprocessing, the images are re-sampled to a smaller size of 256×256 using bilinear interpolation. The input image is overlaid with a grid of 256 × 256 points. The input image is then sampled at each grid point to determine the pixel color of the output image. When grid point lies between input pixel centers, the color value of the grid point is determined by linear interpolation between adjacent pixel gray levels both in the vertical and horizontal directions [5]. The gray levels are then re-quantized to have up to 256 gray levels (i.e., 8-bit gray scale). These normalized versions of the images will be used only for building the corresponding feature vectors and will not be stored in the PACS. Due to the huge size of the database, the original images are the only ones that are stored in the archiving system in order to reduce storage size. Moreover, the normalized images cannot replace the original images in the database since the latter may have better resolution and color quality and consequently of more diagnostic value.

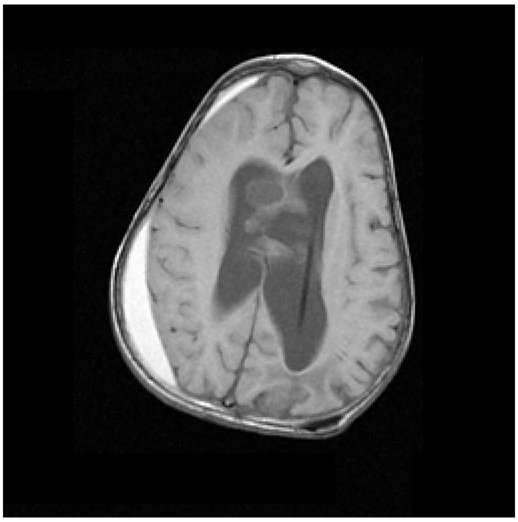
In the second step, we apply noise reduction filtering on the query images. Since the database images are already used for diagnosis, they are assumed to be of high quality and no filtering is needed for them in general. The denoising process is applied when the variance in background region of the query image is higher than a certain threshold. All CT, MRI and X-ray images have a common appearance of a dark background with one or more structures on top of it. The noise adds energy to the image specially to the background. Consequently, this energy causes a bigger error in the matching stage. Among the denoising techniques, Low-pass filtering or averaging techniques reduce the sharpness of the boundaries and edges and redistribute their energy. More complex denoising techniques such as Wiener filters or wavelet denoising cannot be used due to their relatively high computational cost, which slows down the speed of the entire procedure. Therefore, a 5×5 median filter has been selected and found to be sufficient for this application. It is more robust for boundary values in the area under the mask and performs well in reducing the noise energy especially in the background area.

One of the advantages of wavelet techniques over other global color indexing techniques, such as global histogram based techniques and texture-based techniques, is that the feature vector contains spatial information given that it is a color layout image indexing system. This feature, however, makes the matching process sensitive to translational motion. The special nature of MRI/CT X-ray images, which is black background with one or more gray object, allows us to design a system that is more robust against translation. The user is allowed to re-align the center of gravity of query image. First we calculate the center of gravity of the image. If the center of gravity is not identical to the center of the image, it is shifted to the center of the image, assuming that the image is extended in all directions with dark background (near zero gray scale value). This operation may truncate parts of the objects so the values in every removed raw or column are checked and if any value is greater than spatial threshold the alignment process is stopped and a warning massage is displayed. Figure 2 shows an example of using the alignment center of gravity.

3.2 Feature Extraction

After the image passes the preprocessing stage, the query image is ready for generating the feature vector that will be used in the matching process. This vector is generated from the wavelet transform of the image. The 5-layer Daubechies wavelet transform is applied to the image [6-7]. The output of the wavelet transformation is a 256×256 matrix. The upper-left 16×16 sub-matrix is the fifth layer in the Daubechies discrete wavelet transform. This 16×16 sub-matrix consists of four 8×8 sub-matrices. The feature vector is generated in two parts from these four 8×8 sub-matrices. The first part is the 8×8 sub-matrix in the upper-left corner of transformed image that contains the low frequency bands representing the color layout of the image. The second part is the three 8×8 sub-matrices nearby the 8×8 low frequency band sub-matrix. These three submatrices do not represent all image details. However, they contain sufficient information to represent important features within the image. Due to the similarity between the images from MRI, CT and X-ray, the variance of the low frequency 8×8 sub-matrix (upper-left corner sub-matrix) has no significance in distinguishing between these image types. Therefore, it cannot be utilized in the matching process.





(a) (b)

Figure 2. Illustration of the alignment procedure. (a) original image. (b) reference image.

3.3 Matching Process

The feature vector for a system such as WBIIS, which is a color layout indexing system, has to contain spatial information. In particular, the feature vector consists of three parts. The first part is the global features, which are the standard deviations of the three color-channels of 3×8×8 low frequency band sub-matrices. On the other hand, the second and the third parts are the wavelet coefficients of the 16×16 sub-matrix of the color channel. Given the nature of wavelets, this part contains information about spatial variations. The matching step of WBIIS system gives equal weights for all locations in the image. In medical images, however, small objects or image parts may have high diagnostic values. It is not reasonable to give the matching error in background pixels a weight that is equal to such high diagnostic value object within the image. On the other hand, all the pixels of the objects cannot be given the same weight. Therefore, some small parts may have diagnostic value higher than all other pixels. It is very difficult to design an expert system for automatic detection of the importance of each part of an image. Therefore, we allowed users to determine three levels of importance of image regions. First, the user selects the data regions. Then, he/she selects smaller regions inside the data regions with higher diagnostic value. Finally, he/she selects the most important regions. We use these regions to generate a weight function W(x,y) that holds the weight of every pixel in the image to be used during the matching step. From the average of this weight matrix we generate an 8×8 matrix or a 64-destination vector. In the matching phase, we calculate the weighted Euclidean distance based on the following equation for the first part of the feature vector containing gray layout information as a first estimation,

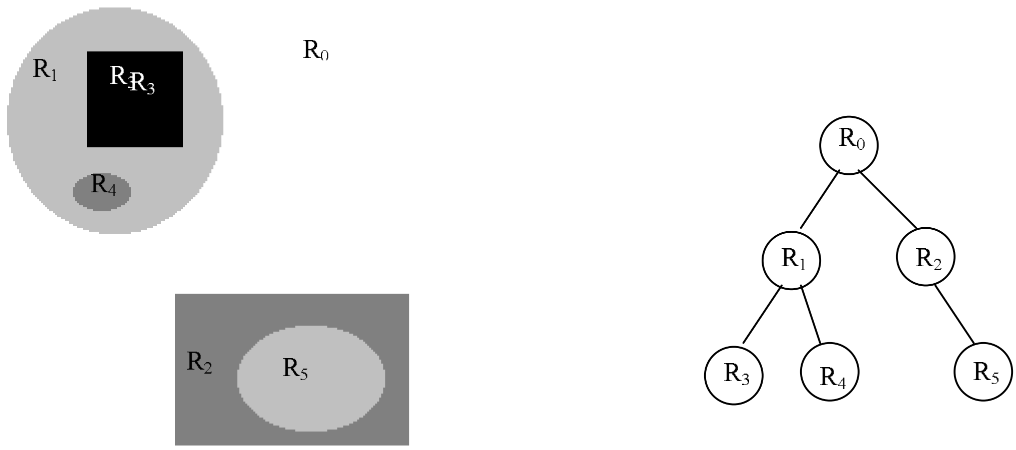
$$d(I,I')=\sum_{i=0}^{3}\|W\ .*\ D_i-W\ .*\ D'_i\|$$

where d(I,I') is the distance between image *I* and image *I'* ; *W* is an 8×8 weight matrix ; D_0 , D_1, D_2 , and D_3 are the four 8×8 sub-matrices representing the feature vector, and the '.*' operator denotes component-wise product .

In the final matching, we calculate the distance depending on the second part of the matching vector, which contains detail information. Then, we retrieve 30 images with the smallest weight Euclidean distance from the PACS and resample them to 128×128 and display them for the user and allow him/her to make the final decision by selecting the nearest image to the query image. The system displays the selected images in their actual size and retrieves the diagnostic report of these images from the RIS. Also, the system allows the user to retrieve all images of the same examination and the patient record from the PIS.

4. SEMANTIC NET BASED SYSTEM

Most of CBIR strategies such as histogram-based, wavelet-based, texture-based, and color model based techniques are based on the signal field of view or the statistical field of view. Since morphological field of view may be more suitable for some applications especially in the medical field, semantic description of the objects in the image and the relations between them can perform well in CBIR systems. We can fully describe the different objects in an image by describing their external boundaries, their color parameters, as well as the relations between them. The algorithm starts by generating a semantic tree representation of all images. This tree stores the relations between all the parent and child regions and the relations between the children regions. This semantic description can be used later in the CBIR. The key factor here is how this semantic description can be generated and formed in suitable form for matching and sorting in database. In this section we develop a morphology based CBIR system for medical images. First to explain the idea of semantic description, we will give an example:



(a)

(b)

Figure 3. (a) the original image and (b) is a semantic net which describes the morphological relations between the objects and the specific parameters of them in the image.

As shown in Fig.3, the image is composed of a region R_0 that is the parent of two regions, R_1 and R_2. These are separate regions that are not intersected. Inside region R_1, we find R_3 and R_4 which are separated. While region R_2 contains region R_5. Our system for CBIR, which builds and uses semantic net, will be described in details in the following subsections.

4.1 Preprocessing

First, before the semantic net being generated the image must be normalized. The image is resampled to be 256×256 pixel and the gray level re-quantized to be 256 gray levels. There is no need for noise removing stage since our segmentation technique is more robust for noise as we will describe later. Also there is no need for alignment the center of gravity since the matching based on the semantic net is affected only by the relative relation between the objects, object shape prompters and objects color parameters [8].

4.2 Segmentation

In order to generate the semantic net we must have the image segmented into regions. There are several segmentation algorithms suitable here. We used region-growing techniques because they are better for segmentation of noisy image [5]. In general, region-based segmentation techniques split the input image into N regions according to the following equation,

$$I = \bigcup_{i=1}^{n} R_i \qquad R_i \cap R_j = \Phi \qquad i \neq j$$

Here, I is the input image which is a finite set of regions R1….Rn ,and every region Ri must satisfy the following criteria

$$H(R_i) = TRUE \qquad i = 1,2,\ldots\ldots\ldots,n$$
$$H(R_i \cup R_j) = FALSE \qquad i \neq j \qquad R_i \, adjacent \quad to \quad R_j$$

where H(Ri) is a binary homogeneity evaluation of the region Ri . In our case, we consider a region homogeneous if the color distance between any two-neighbor pixels is less than a spatial threshold (local threshold) and the color distance between any pixel in the region and the mean of the region is less than another spatial threshold (global threshold). This definition can be formed as follows:

$$abs\left(g(P_i), g(P_j)\right) < T_L,$$
$$abs\left(g(P_i), \bar{g}(R)\right) < T_G.$$

Here pi & pj are two neighbor pixels in the region, g(pi) is the gray level of pixel pj and $\bar{g}(R)$ the average gray level of the region R where,

$$\bar{g}(R) = \frac{\sum_{i=1}^{N} g(P_i)}{N}.$$

Here P_I is a pixel in region R and N is number of pixel in the region. In order to segment the image into regions we start with the upper left pixel to be the first pixel in the region R_0 this region is growing in the neighbor pixels as long as the homogeneity condition is satisfied. Just when the growing of region R_0 is stopped, we consider the first pixel outside the region R_0 boundary as the first pixel in region R_1 and thus, starting region growing for region R_1 and so on. These procedure run until every pixel in the image is assigned to region .We merge any two regions if the absolute value of the difference between their average gray values is less than the spatial threshold or if the size of any of them is less than a certain threshold [8-10].

4.3 Building Semantic Net

Building region hierarchy is the first step in building our semantic net. The output of this module is the region relation hierarchy. Region relation hierarchy can be fully described if the parent and the level of every region are determined. First, we set the dark background as the root region of the hierarchy. Then, we start a loop to assign every region in the image to its correct location in the hierarchy. The loop continues over all the successive levels and for every region within every level. For every region, the neighbor region is checked. If the neighbor region is not assigned to a location in the hierarchy, it assigned as a child to the current region. On the contrary, if the neighbor region is already assigned to a location, it is not the parent of the current region, and the parent of this region is not the parent of the current region or the current region itself, then this region is reassigned as a child of the current parent region [11].

The following example will clarify the idea of the algorithm (see Figure 4). We start with the R_0 as the root. For level zero and region zero the current region is region zero the neighbor of region zero is region R_1 is not assigned to location in the hierarchy then assign it as a child for region zero. In level1, when the current region is R_1 and the neighbors of $R_1$are R_0, R_2, R_4, R_5, and R_6. R_0 is assigned to a location and it is the current region parent. R_4, R_5, R_6, and R_2 have not been assigned yet, then they become assigned as children of R_1. In level 2, when the current region is R_2, the neighbor regions are R_1, R_5, R_6, and R_3. R_3 is not assigned then it becomes assigned as a child of R_2. R_1 is assigned to location and it is the current region parent. R_5 and R_6 are assigned to location and their parent is the parent of the current region (which is R_1). In level2, when the current region is R_4, the neighbor regions are R_1, R_3, R_5, and R_6. R_1 is assigned to location and it is the current region parent. R_5 and R_6 are assigned to some locations and their parent is the parent of the current region (which is R_1). R_3 is assigned to a location but it is not the current region parent and its parent is not the current region parent and it is not a child of the current region. Then, R_3 is assigned to error location and its current location as a child of current region parent (which is R_1) [12-14].

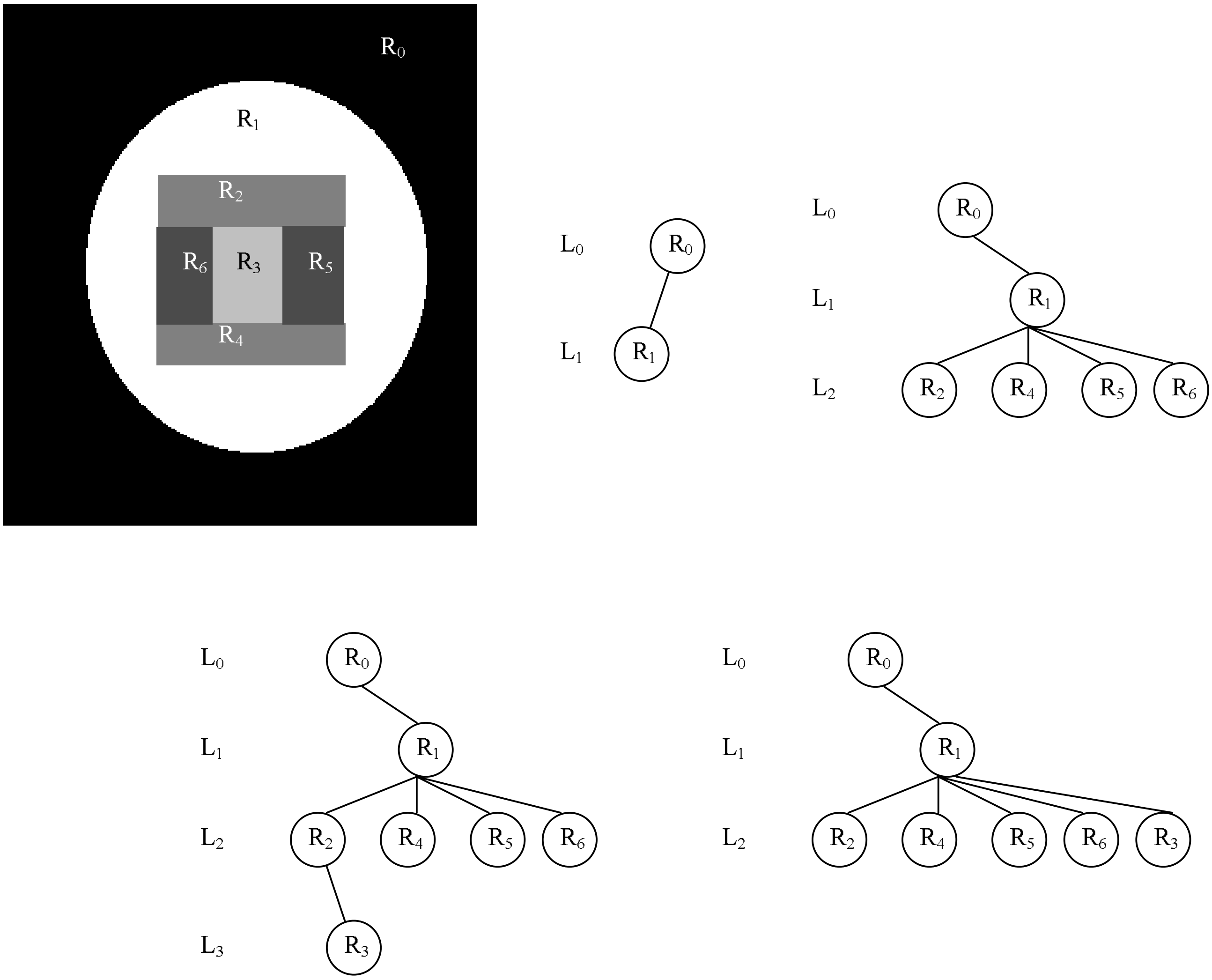


Figure 4. Illustration of the steps of the proposed algorithm.

4.4 Region description and inter-level links

The second step in the building the semantic net is to describe the regions of every node and the relation between the nodes in the same level. The average gray level is used to describe the region color. The region moments match our need for describing the region shape more than other region and boundary descriptors, since the moments are translation-, rotation-, and scale-invariant. Compared to chain code, signature and segment sequence moments are easier in feature vector construction, storing and matching process because moments give a fixed length feature vector of seven elements for area–based moments, four elements for contour sequence moment. Moments also facilitate the matching process which can be based on pure Euclidean distance only while other descriptors require a complicated correlation process. Compared to Fourier descriptors, moments are also independent from the starting points.

4.5 Matching

One of the difficulties of representing the semantic net in the feature vector is that it is of variable length since the number of regions is different from image to another. Also, there is a problem in storing the semantic net relations. It is a good idea to build a matching process based on multi resolution description and divide the feature vector to be suitable for matching. The first part of the feature vector consists of image ID, number of nodes in first three levels. The second part describes the nodes of level one and consists of image ID, objects ID and a vector contains the object parameters (average gray levels and moments). The objects are arranged based on image ID. The third and the forth parts consist of region ID, object ID and object feature vector for the objects in levels two and three.

The first cut from the matching process is based on the number of objects in level1. The second cut is based on the features of these objects. The third cut is based on the number of objects in level2 and the fourth cut is based on the features of objects in level2.

5. EVALUATION OF CBIR PERFORMANCE

One of the problems in CBIR system design is how the performance of the system can be assessed. The accuracy of CBIR systems is commonly evaluated by test-running the system on known data. Different query images are applied to the system and the output images are checked. Even though this is sufficient for preliminary testing, it is not adequate to provide an overall evaluation for two reasons. The first is that Some CBIR systems give good results for special cases of database images or query images. This becomes more complicated if we know that there is currently no standard data set that is designed for testing CBIR system accuracy. The second reason is that the evaluation of output images may not be accurate. It is usually a difficult task to determine that the output images are the most similar images in the database for any given query image according to human assessment.

In our case, besides the above simple evaluation, we apply an additional quantitative test. We select sample images representing most of cases and create a small data set. The distances between the query images and the other images within the data set are calculated. Different kinds of degradations and transformations with different degrees are applied on the query images then the distances between them and data set images are recalculated. These results show the sensitivity of the system for the transformations and degradations. Applied transformations include translation, rotation, and scaling. On the other hand, the applied degradation schemes include adding noise and local white spots with different sizes representing small objects with high diagnostic values. Results are compared with that of WBIIS system.

6. RESULTS

Compared to WBIIS, the proposed wavelet-based and semantic-based techniques have been found to be more robust against translation and noise. On the other hand, WBIIS and the wavelet-based techniques have been demonstrated to be sensitive for rotation, especially if the rotation angle is more than 20 degrees. Meanwhile, semantic-based technique has been shown to be rotation-, scaling-, and translation-insensitive technique. The proposed wavelet-based and semantic-based techniques discriminate between similar images based on small objects, which are determined by the user to be of high diagnostic values. Figures 5 and 6 represent two samples of our testing process. Figure 5 shows how the proposed wavelet-based technique is robust to translation. Figure 6 illustrates the effect of small object with high diagnostic value on the matching. In Figure 5, image (b) is an image in the data set and image (a) is the translated version of image (b). In order to measure the effect of translation, we computed the distance (i.e., the Frobenius norm of the difference image) between images (a) and (b) using the outputs from WBIIS and the proposed wavelet-based technique. The distance in WIIBS case is 17291 units while in the proposed Wavelet-based case the distance is zero.

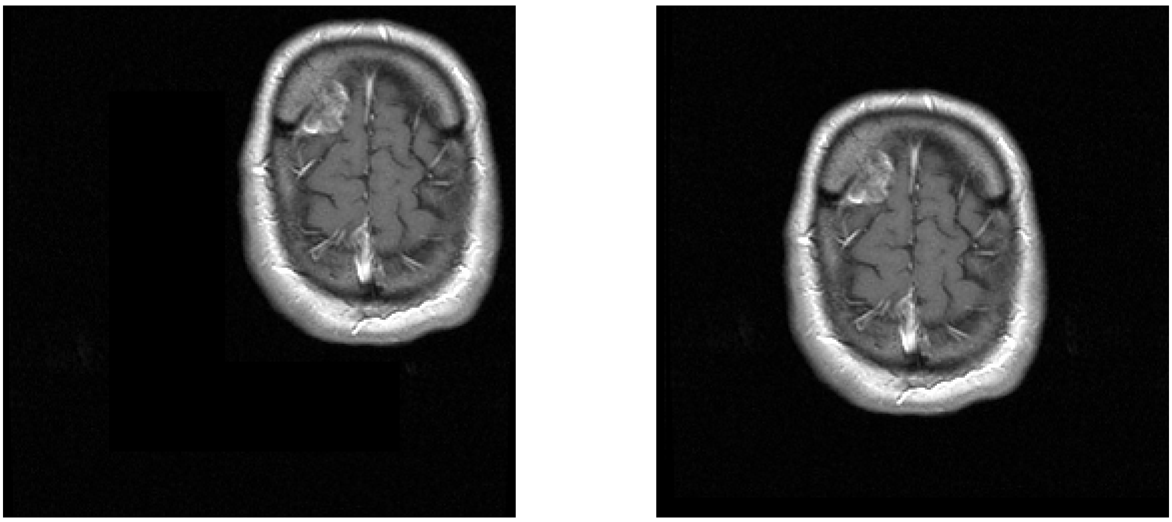
In Figure 6, image (c) is an image in the data set and image (a) is the image in (c) with a superimposed small white spot representing a small object with high diagnostic value. Image (b) is the same image as in (a) after the user determined the region with highest priority. Images in (a) and (c) are similar images but do not lead to the same diagnosis. In order to measure how the system can differentiate between these images, we calculate the distance between image (a) and image (c) according to WBIIS and our wavelet-based technique. The distance in WIIBS case is 853 units while in our Wavelet-based case the distance is 17051 units when the weight is selected to be 20.

7. CONCLUSIONS

A CBIR system for medical imaging has been developed for use with PACS. Two approaches based on the wavelet transform and semantic net have been implemented and shown to be useful for practical implementation. Future work includes quantization and coding of the wavelet coefficient to integer values in order to reduce the time required for computations. Also, extension of this work to include ultrasound imaging and other modalities should be addressed.

REFERENCES

1. C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. Aisen, and L. Broderick, "Local versus global features for content-based image retrieval," *Proc. of the workshop on Content-Based Access of Image/Video Library, Santa Barbara, CA,* June. 21, 1998.
2. S.T. C. Wong, and D.A. Tjandra, "A digital library for biomedical imaging on the Internet," *IEEE Communication Magazine,* **37**, no. 1, pp 84-91, January 1999.
3. J.Z. Wang, G. Wiederhold, O. Firschein, and S.X. Wei, "Content-based image indexing and searching using Daubechies' wavelets," *Inter. J. Digital Libraries*, pp 311-328, 1997.
4. C.E. Jacobs, A. Finkelstein, D.H. Salesin, "Fast multiresalution image querying," *Proc. of SIGGAPH 95 , in computer graphics proceedings Annual conference series,* pp. 277-286, Augast 1995.
5. M. Sonka, V. Hlavac, R. Boyle , *Image processing , Analysis and machine vision, 2nd ed.*, PWS, 1995.
6. Ingrid Daubechies, *Ten lectures on Wavelets, (BMS-NSF) region conference series in Applied Mathematics*, 1992.
7. S.G. Mallat , "A theory for multiresalution signal Decomposition : the wavelet representation," *IEEE transaction on Pattern Analysis and Machine Intelligence,* **11**, no. 77, pp. 674-692, July 1989.
8. C.-S. Fuh, S. Wencho, and K. Essig, "Hierarchical color image region segmentation for content-based image retrieval system," *IEEE transaction on Image Processing,* **9**, no. 1, pp 156-162, January 2000.
9. A. Tremeau and N. Borel, "A region growing and merging algorithms to color segmentation," *Pattern Recognition,* **30**, pp 1191-1203, 1997.
10. J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition,* **26**, pp 167-174, 1993.
11. P.H. Winston, *Artificial Intelligence, 3rd ed.*, Addison wesley, 1993.
12. T.-P. Redich, M. Suzuki, and S. Weinstein, "Distributed object technology for networking," *IEEE Communication Magazine,* **36**, no. 10, pp. 100-111, October 1998
13. H.S. Stone, "Image libraries and the internet", *IEEE Communication Magazine,* **37**, no.1, pp. 99-106, Jan. 1999.
14. A. Paepcke, M.Q. Wang Baldonado, C.-C.K. Chang, S. Cousins, and E. Garcia-Molina, "Using distributed objects to build the Stanford digital library infobus," *IEEE Computer Magazine,* pp. 80-87, February 1999.



(a) (b)

Figure 5. Illustration of performance with translation. (a) query image (b) obtained most similar image from the data set.

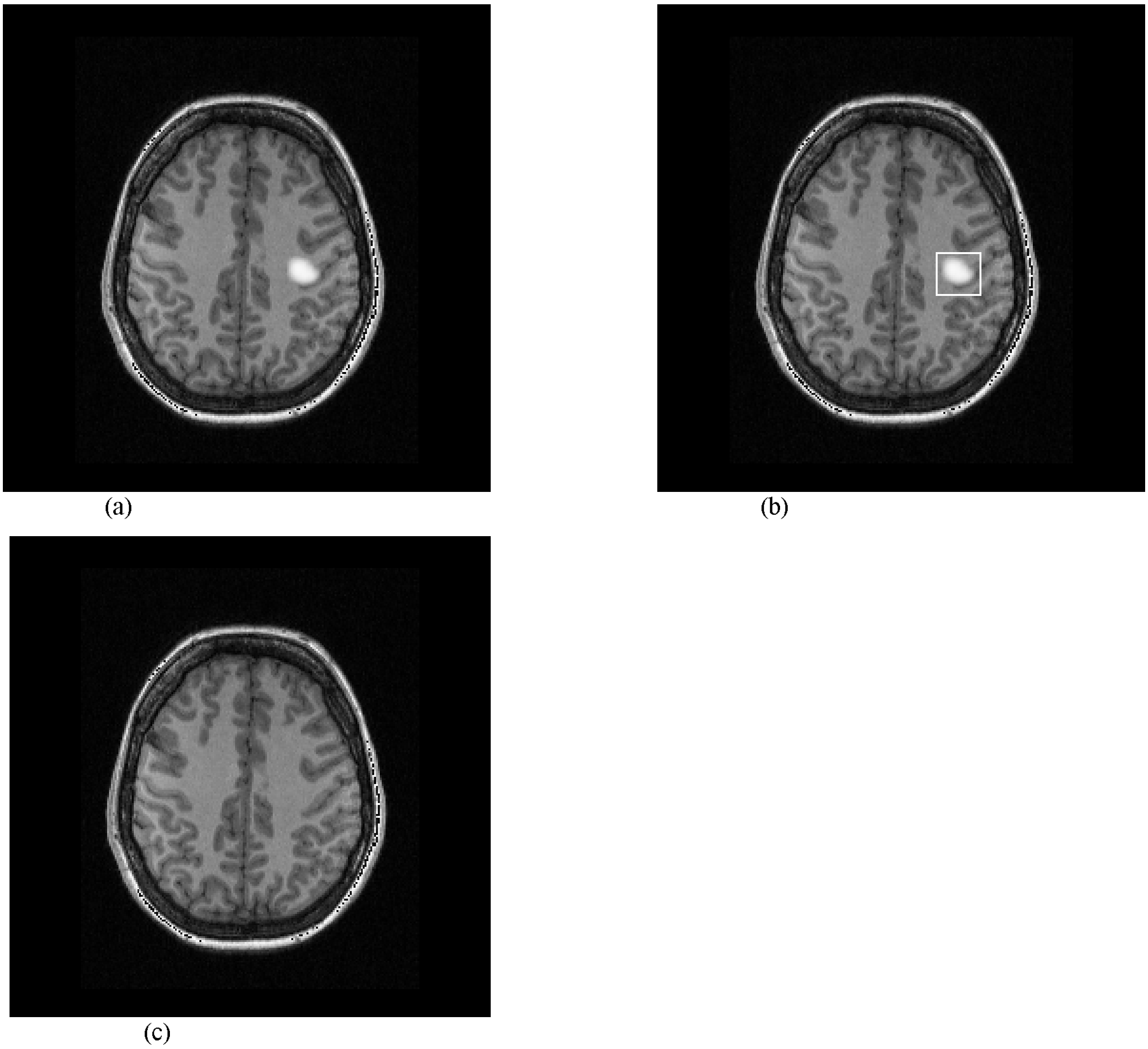


Figure 6. (a) a query image (b) a query image after the user determined the image part with high diagnostic value (c) most similar image to query image in the data set.