Assessment of Performance Improvement in Content-based Medical Image Retrieval Schemes Using Fractal Dimension[1]

Sang Cheol Park, PhD, Xiao-Hui Wang, MD, Bin Zheng, PhD

Rationale and Objectives. The aim of this study was to investigate whether using a fractal dimension as an objective index (quantitative measure) to assess and control the ''visual'' or ''texture'' similarity of reference-image regions selected by a content-based image retrieval (CBIR) scheme would (or would not) affect the performance of the scheme in classification between image regions depicting suspicious breast masses.

Materials and Methods. An image data set depicting 1500 verified mass regions and 1500 false-positive mass regions was used. Fourteen morphologic and intensity distribution features and a fractal dimension were computed. A CBIR scheme using a *k*-nearest neighbor classifier was applied, and two experiments were conducted. In the first experiment, the CBIR scheme was evaluated using all 15 features. In the second experiment, the fractal dimension was used as a prescreening feature to guide the CBIR scheme to search for the most similar reference images that had similar measures in the fractal dimension.

Results. The CBIR scheme achieved classification performance with areas under the receiver-operating characteristic curve of 0.857 (95% confidence interval [CI], 0.844–0.870) using 14 features and 0.866 (95% CI, 0.853–0.879) after adding the fractal dimension ($P = .005$ for both results). After using the fractal dimension as a prescreening feature, the CBIR scheme achieved an area under the receiver-operating characteristic curve of 0.851 (95% CI, 0.837–0.864), without a significant difference from the previous result using the original 14 features ($P = .120$). The difference of fractal dimension values between the selected similar reference images was reduced by 56.7%, indicating improvement in image texture similarity. In addition, more than half of references were discarded early, without similarity comparisons, indicating improvement in searching efficiency.

Conclusions. This study demonstrated the feasibility of applying a fractal dimension as an objective (quantitative) and efficient search index to assess and maintain the texture similarity of reference mass regions selected by a CBIR scheme without reducing the scheme's performance in classifying suspicious breast masses.

Key Words. Content-based image retrieval (CBIR); mammograms; fractal analysis; computer-aided diagnosis (CAD); visual similarity; receiver-operating characteristic (ROC) analysis.

Content-based image retrieval (CBIR) schemes have been developed to search for images similar to a queried image from large reference-image databases on the basis of features or image content inherently contained within the images (1,2). In particular, the CBIR method has been proposed to overcome the difficulties encountered in textual annotation or description by manual methods for large image databases (3). Although the use of purely visual image querying is unlikely to be able to completely replace text-based searching methods, CBIR has the potential to be a very useful complement to text-based searching methods because of unique image characteristics (4). In the field of computer vision, CBIR has been one of the most active research areas over the past 30 years (5).

Acad Radiol 2009; 16:1171–1178

[1] From the Department of Radiology, University of Pittsburgh, 3362 Fifth Avenue, Pittsburgh, PA 15213. This work is supported in part by grants 1 UL1 RR024153 from the National Center for Research Resources (Bethesda, MD) and CA101733 from the National Cancer Institute (Bethesda, MD) to the University of Pittsburgh. Received December 9, 2008; accepted April 21, 2009. **Address correspondence to:** S.C.P. e-mail: parks@upmc.edu

doi:10.1016/j.acra.2009.04.009

Currently, with the advances in digital technologies applied to medical imaging, a large number of diverse radiologic and pathologic images in digital format are rapidly produced in hospitals and medical centers using sophisticated image acquisition devices and digital scanners. These digital images have been routinely used for the purposes of diagnosis and therapy. The management of and access to these large image repositories has become increasingly complex and challenging (1). In reading and interpreting medical images in daily clinical practice, observers (ie, radiologists and pathologists) are often faced with new unknown and suspicious lesions, requiring them to search for and make comparisons to similar cases with previously verified results in their decision making in detection and diagnosis (6,7). This is a difficult and very time consuming task because of the rapid increase in the sizes of medical imaging databases. Therefore, developing and applying CBIR schemes to more effectively organize and retrieve images has attracted wide research interest in medical imaging and informatics (1,4,8). In particular, a number of studies have recently been conducted on how to develop and optimize CBIR schemes to search for similar breast lesions (including masses and microcalcification clusters) (3,6,9–11).

There are many factors to consider in the design of a CBIR system, such as the domain and purpose, the choice of appropriate features, the criteria for assessing image similarity, the indexing mechanism, and the query formulation technique (1,4). One of the most important factors in the design process is the choice of suitable visual features and the methods to extract them from raw images (1), because the query image is formulated and represented by exclusive features (9). Moreover, feature extraction affects all subsequent processes. In medical imaging applications, the performance and potential clinical utility of CBIR schemes are evaluated primarily by three factors: (1) clinical relevance (ie, classification performance evaluated using receiver-operating characteristic [ROC] analysis), (2) visual similarity between the queried image region and the selected reference regions (ie, finding the effective visually similar indices), and (3) search efficiency (ie, whether the computation task can be done in real time). Some previous studies have focused on improving CBIR schemes in distinguishing between positive and negative lesions (clinical relevance) (6,10), while others have assessed and compared the visual similarity of selected reference images using subjective ratings (11) and a two-alternative forced-choice observer preference study (12). Although visual similarity is very important in the application of CBIR schemes, previous studies have also found that it is a subjective concept, with high interobserver variability (11,12). Therefore, identifying and applying an objective index (a quantitative feature or feature set) to assess the visual similarity of reference-image regions selected by CBIR schemes is an important and technically challenging task. In addition, many similarity indices based on pixel value distribution (ie, mutual information and Pearson's correlation) used in previous CBIR schemes are computationally expensive and cannot be used to conduct efficient (''real-time'') image searches (13). Our goal is to guarantee that all reference-image regions selected by CBIR schemes have improved ''visual similarity'' assessed using an objective and computational efficient index (not the previously used subjective indices), without reducing the schemes' performance in the classification of suspicious lesions (diagnosis of medical images).

In computer vision, texture is defined by such terms as ''structure'' and ''randomness'' (5). The fractal model (or analysis) has been introduced to describe the ruggedness of natural objects (14). One of the advantages of fractal analysis is the ability to quantify and describe the irregularity and complexity of images with a measurable value, which is called the fractal dimension (15,16). Because the fractal dimension shows both self-similarity and overall roughness on multiple scales (17), it can be used to describe and interpret patterns of visual texture (18). For example, one previous study reported that because cancerous tumors exhibited a certain degree of randomness associated with their growth and were typically irregular and complex in shape, fractal analysis could provide a better measure of their complex patterns than conventional Euclidean geometry (19). Other studies have also suggested that the subjective evaluation of visual features was highly correlated with fractal dimensions of textile design images (20) and could help in determining the inherent complexity of visual images and serve as a tool for the estimation of visual complexity (21). Also, some researchers believed that visual similarity based on the fractal dimension might at least partially imply or correlate with semantic similarity (22). Therefore, the fractal dimension can be used as a feature for both visually recognizable image texture and semantic similarity in searching for irregular cancerous regions.

Because of its unique advantages, fractal analysis has been widely applied to many medical imaging research areas, including detection, segmentation, and classification, with varied success. For example, the fractal dimension has been used in the detection and segmentation of microcalcifications depicted on digital mammograms (14,23), distinguishing between benign and malignant breast masses (24), the classification and analysis of mammographic parenchyma patterns (17,25), and the analysis of trabecular bone structure (26,27). However, to the best of our knowledge, the fractal dimension has not been applied in any CBIR schemes to search for similar medical images (ie, those depicted breast mass regions) from reference databases. In this preliminary study, we investigated and tested whether using the fractal dimension as an objective index (quantitative measure) to assess and control the visual similarity of reference-image

regions selected by a CBIR scheme would (or would not) affect the performance of the scheme in the classification of image regions depicting suspicious breast masses.

MATERIALS AND METHODS

Reference-image Database

We have assembled a large and diverse image database of mammograms in our laboratory. The original digitized mammograms were generated using several film digitizers, with a pixel size of 50 × 50 μm and 12-bit gray-level resolution. To create a reference database used in CBIR-based computer-aided diagnosis (CAD) studies, our computer program first subsampled the images by a factor of two (increasing pixel size to 100 × 100 μm) and then extracted all selected regions of interest (ROIs) with a fixed size of 512 × 512 pixels. The center of each suspicious mass was also located in the center of the extracted ROI. Using the ROI center pixel as a mass region growth seed, a previously developed multilayer topographic region growth algorithm (28) used in our CAD scheme was applied to segment the mass region (define its boundary contour). For each true-positive mass, the automated segmentation result (its boundary contour) was visually examined. If a noticeable segmentation error was identified, the mass boundary contour was manually corrected (redrawn). Unlike some previously reported studies, in which negative ROIs were randomly selected and extracted from negative images, each negative ROI selected in our reference database actually contains a false-positive mass that is automatically segmented and cued by the CAD scheme.

The reference database used in this study includes 3000 ROIs extracted from mammograms acquired from 1127 patients. In 336 cases, ROIs were extracted from two breasts, and in 791 cases, ROIs were extracted from one breast. Thus, among 1463 breasts (2×336 + 791), 843 depict verified mass regions (true-positive findings), and 620 do not. Among the 843 breasts depicting positive masses, 722 depict malignant masses, and 121 depict biopsy-proved benign masses. All these masses were originally rated in Breast Imaging Reporting and Data System categories 4 and 5 by radiologists. Similar to our previous study (12), the goal of the CBIR schemes compared and tested in this study is to detect suspicious mass regions (classification of whether the queried ROI depicted a suspicious breast mass). Thus, each true-positive ROI depicts one verified mass (either malignant or benign), and each false-positive ROI depicts a CAD-cued false-positive mass that is actually negative. In summary, among these 3000 ROIs, 1500 are true-positive regions, and the remaining 1500 are negative regions. The 1500 true-positive ROIs were extracted from 906 masses depicted on 843 positive breasts. Among them, 594 masses were extracted from both the craniocaudal and mediolateral oblique views, and 312 masses were extracted from only one view. The 1500 negative ROIs were extracted from CAD-cued false-positive mass regions depicted on 769 breasts (including 620 negative and 149 positive breasts). In addition, some of image characteristics of these selected true-positive mass regions have been reported elsewhere. Approximately half of these masses were rated subjectively as ''subtle'' to ''very subtle'' by radiologists (12).

On the basis of mass segmentation results, we used a computer scheme to compute 14 morphologic and intensity (pixel value) distribution features from each segmented mass region (including both true-positive and false-positive regions). These 14 features were selected from a large initial feature pool using genetic algorithm (GA), as reported in our previous study (12). The detailed definitions and computing methods of these features, including three global features computed from the whole breast area segmented from the image (average pixel value in the breast area, average local pixel value fluctuation in the breast area, and standard deviation of the local pixel value fluctuation in the breast area) and 11 local features computed from the segmented mass region and its surrounding background (region conspicuity, normalized mean radial length of a region, standard deviation of radial length, skew of radial length, shape factor ratio, standard deviation of pixel value inside the mass region, standard deviation of the gradient of boundary pixels, skew of the gradient of boundary pixels, standard deviation of pixel values in the surrounding background, average local pixel value fluctuation in the surround, and normalized central position shift), have been previously reported (12). These 14 computed image features were saved in a reference-feature data file that contains all extracted and selected ROIs in our reference database.

Fractal Dimension

We added a new feature, the fractal dimension, into the reference-feature data file for each selected ROI in this study. To compute the fractal dimension, the computer scheme first applied the fast Fourier transform to each ROI and produced a unique two-dimensional complex array called the power spectrum (27). The power spectrum is calculated as

$$P(u,v) = |F(u,v)^2| = R(u,v)^2 + I(u,v)^2,$$

where u and v represent the horizontal and vertical components of frequencies, respectively, and $R(u, v)$ and $I(u, v)$ are the real and imaginary parts of the Fourier transform $F(u, v)$, respectively. The power spectrum is displaced in polar form and shifted to locate zero frequency at the center. The scheme calculates $\log[f(u, v)]$, where $f(u,v) = \sqrt{u^2 + v^2}$, and average $\log[P(u, v)]$ in the condition of $f(u,v) \leq \sqrt{u^2 + v^2}$. Each

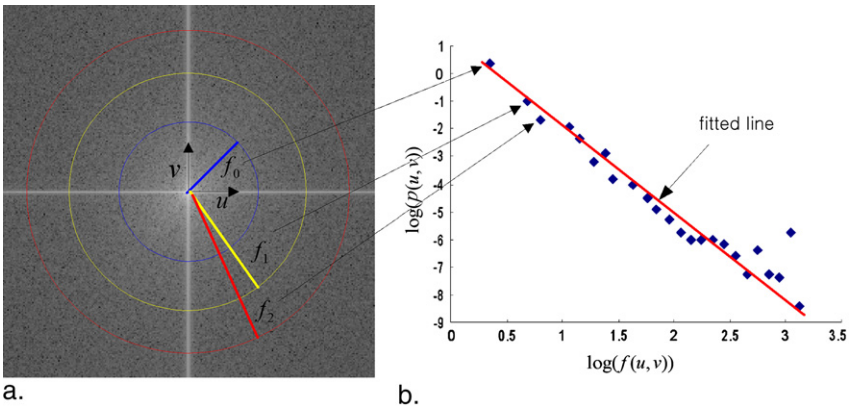

Figure 1. **(a)** Power spectrum using Fourier transform and **(b)** line fitting for the curve on average log[$P(u, v)$] versus log[$f(u, v)$]. The center of the image is the origin of the frequency coordinate system; f_0, f_1, and f_2 are the distances according to the frequencies, u and v, from the origin of the coordinate system.

set of u and v is defined by every 0.1 of log[$f(u, v)$] (Fig 1a). The slope of the curve on average log[$P(u, v)$] versus log[$f(u, v)$] is calculated by least-squares fitting (Fig 1b). Finally, the fractal dimension is calculated as

$$\text{fractal dimension} = (7 - \text{slope})/2.$$

Each fractal dimension computed from 3000 reference ROIs was then normalized to range from zero to one. The results were saved in the reference-feature data file together with the aforementioned optimal set of 14 features. Finally, in the study, the fractal dimension was either used as an individual (prescreening) feature or combined with the other 14 features to describe each ROI.

CBIR Scheme Using a *k*-Nearest Neighbor Classifier

In our previous study (12), we developed and tested a CBIR scheme using a multiple-feature-based k-nearest neighbor (kNN) classifier to search for similar breast masses depicted in the reference database. Our GA optimized the kNN-based classifier, which searches for and identifies 15 (K) of the most ''similar'' suspicious mass regions to the testing (queried) region from the preestablished reference-feature data file. The similarity is measured by the Euclidean distance (d) between a testing mass region (y_T) and each of the reference regions (x_i) in a multidimensional space with F_N selected image features (f_r):

$$d(y_T, x_i) = \sqrt{\sum_{r=1}^{F_N} [f_r(y_T) - f_r(x_i)]^2}.$$

A smaller distance indicates a higher degree of ''similarity'' between two compared regions. The kNN classifier then computes a detection score to indicate the likelihood of being a true mass to the queried region:

$$P_{TP} = \frac{\sum_{i=1}^{N} w_i^{TP}}{\sum_{i=1}^{N} w_i^{TP} + \sum_{j=1}^{M} w_j^{FP}}, N + M = K,$$

where $w_i = 1/[d(y_T, x_i)]^2$ (a distance weight), and w_i^{TP} and w_j^{FP} are the distance weights for the true-positive (i) and false-positive (j) mass regions, respectively. N is the number of verified true-positive mass regions, and M is the number of CAD-cued false-positive regions.

Performance Evaluation

When applying the CBIR scheme, a retrieved reference image (ROI) is considered to be clinically relevant if it belongs to the same class (ie, mass or nonmass in this study) of the query image (ROI). Because CBIR schemes use instance-based machine learning methods that depend on nearest neighbors and/or locally weighted regression to approximate real-valued or discrete-valued target functions, no pretraining process is needed to construct a general and explicit target function (28). In this study, we used a leave-one-case-out method to test and evaluate the performance of our CBIR scheme. In the experiment, each of 3000 ROIs in our reference database was separately used once as a testing (queried) ROI. In this iterative process of performance evaluation, once a testing ROI was selected, the CBIR scheme searched for the K ROIs through the remaining reference database (excluding itself and all other ROIs extracted from the same case or patient) that were considered the most similar to the testing ROI. As a result, a set of K similar reference ROIs and a corresponding detection score is generated for the testing ROI. On the basis of the detection scores for both true-positive and false-positive ROIs, we applied the ROC data-fitting

and analysis program ROCKIT (Charles E. Metz, University of Chicago, Chicago, IL) to compute ROC curves, including the areas under the curves (A_z) and 95% confidence intervals (CIs). The A_z values were used as indices to assess the performance of the CBIR scheme in selecting clinically relevant reference ROIs. Statistically significant differences (P values) were also used to compare the performance difference between two CBIR schemes.

In this study, we conducted two experiments to evaluate CBIR schemes with the fractal dimension. In the first experiment, we tested and evaluated our kNN-based CBIR scheme using 15 features (including the fractal dimension and the previously selected optimal feature set with 14 features). The primary purpose of this experiment was to test whether adding the fractal dimension was redundant or whether it could replace a number of the other existing features. We applied the GA with the same procedure as described in our previous study (12) to search for the optimal features among these 15 features (14 morphologic and intensity distribution features and the fractal dimension) and the number (K) of similar reference regions. In brief, a binary coding method was applied to create GA chromosomes. Each extracted feature corresponded to one gene of a chromosome, for which a value of 1 indicated that the feature was selected and a value of 0 indicated that the feature was discarded. Five additional genes were also appended to the chromosome to find an optimal reference K. For example, 01111 indicates that 15 neighbors are selected. Thus, each GA chromosome included 20 genes in this experiment. The GA iteratively performed crossover and mutation operations to find the compositions of genes improving the performance of the CBIR scheme. At each iteration, the A_z value was computed according to the combination of the features and K corresponding composition of genes selected by the GA. When there was no performance improvement in the new generation or the searching generation reacheed the predetermined maximum number (ie, 100 in our studies), the GA optimization terminated.

In the second experiment, we added the fractal dimension as a prescreening feature (condition) to our previously optimized CBIR scheme using 14 features. The purpose of this experiment was to force the CBIR scheme to search only for the reference ROIs that had similar fractal dimension values (texture similarity) to the testing ROI. In the experiment, once a testing ROI was queried, the fractal dimension was used as a criterion to discard early all the reference ROIs if the difference (d_{FD}) of fractal dimensions between the testing ROI and these reference ROIs was larger than a predetermined threshold (α). The criterion was $-\alpha \leq d_{FD} \leq \alpha$. The reference ROIs beyond the condition were discarded in advance before the previously developed CBIR scheme (12) was applied to search for similar ROIs from the remaining images in the reference database. Hence, computational complexity was reduced because the number of reference regions removed by the condition resulted in further improvement in search efficiency. We systematically tested and evaluated the performance of the CBIR scheme and computational complexity as a function of the threshold in fractal dimension difference (α). Finally, we compared the statistically significant difference between performance of the CBIR schemes using the fractal dimension as one additional feature (experiment 1) or a prescreening feature (experiment 2) and the CBIR scheme using only the previously optimized 14 features using the CORROC program in ROCKIT.

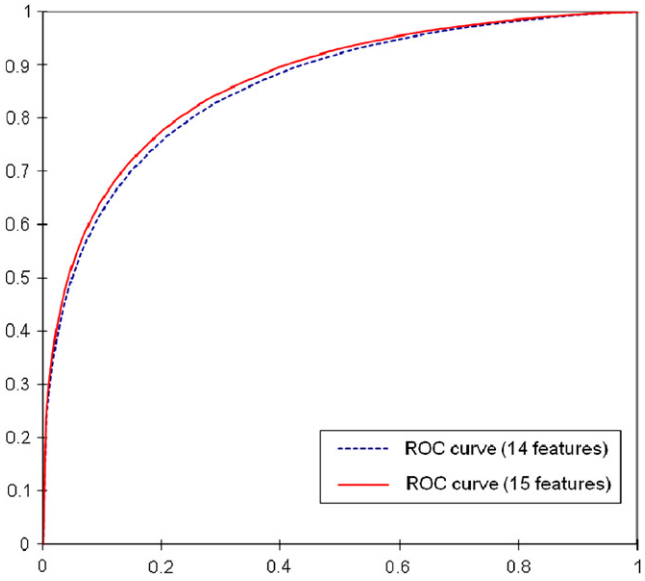

Figure 2. Comparison of the two receiver-operating characteristic (ROC) performance curves generated using two content-based image retrieval schemes with the original 14 features and adding the Fourier spectrum–based fractal dimension as the 15th feature. The areas under the ROC curves are 0.857 (95% confidence interval, 0.844–0.870) and 0.866 (95% CI, 0.853–0.879), respectively.

RESULTS

In the first experiment, the GA generated an optimal kNN including all 15 features and 26 neighbors (K = 26). The CBIR scheme achieved the best performance (A_z = 0.866; 95% CI, 0.853–0.879). Compared to the previously optimized CBIR scheme, which achieved an A_z value of 0.857 (95% CI, 0.844–0.870) using 14 features and the 15 nearest neighbors (K = 15), we found that the fractal dimension was not a redundant feature, and it could make the contribution to improving the CBIR scheme's performance. Figure 2 shows two ROC curves plotted using classification results generated by the previous scheme using 14 features and the new scheme adding the fractal dimension. The difference between the two A_z values was assessed with a two-tailed P value computed in

Table 1
Performance of the Content-Based Image Retrieval Scheme and the Average Number of Early Discarded (Removed) Reference ROIs as a Function of the Comparison Threshold of the Fractal Dimension

α (Half of Difference Between Two Fractal Dimensions)	A_z	95% CI	Average Number of Removed Reference ROIs
0.06	0.843	0.844–0.870	2140
0.07	0.843	0.829–0.857	1967
0.08	0.845	0.931–0.859	1835
0.09	0.850	0.836–0.862	1709
0.10	0.851	0.837–0.864	1588
0.20	0.853	0.839–0.866	695
0.30	0.853	0.839–0.865	272

A_z, area under the receiver-operating characteristic curve; CI, confidence interval; ROI, region of interest.

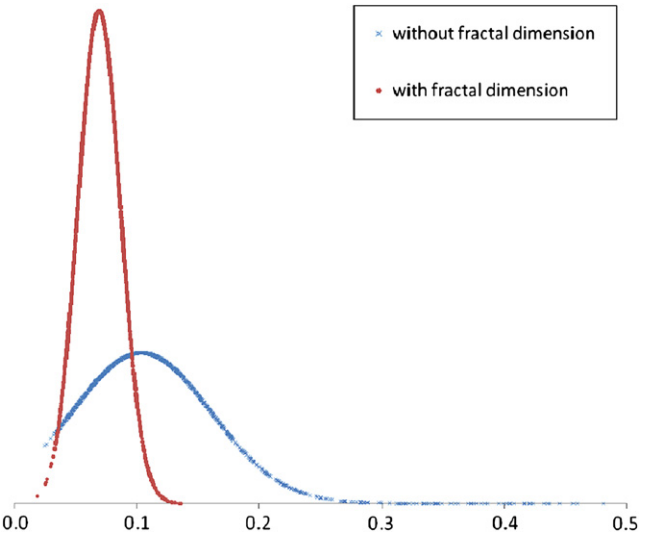

Figure 3. Distributions of the mean of fractal dimension differences between the testing region and each of 15 similar regions selected by two content-based image retrieval schemes with and without using the fractal dimension as a prescreening condition.

ROCKIT (P = .005), with a 95% CI for the difference of −0.0136 to −0.0025.

For the second experiment, Table 1 shows the computed A_z values and their 95% CIs of the CBIR scheme as the variation of the thresholds (α) on the allowed difference in fractal dimension values between the queried ROI and the selected reference ROIs. For example, the results show that for α = 0.10, our CBIR scheme using 14 features and 15 nearest neighbors achieved the best performance (A_z = 0.851; 95% CI, 0.837–0.864). There was no statistically significant difference between the performance of the CBIR scheme using the fractal dimension as a prescreening condition (ie, α = 0.10) and the original CBIR scheme using only the previously optimized 14 features (P = .120). At this threshold level, 1588 reference ROIs (approximately 53%) were discarded early (Table 1). The mean ± standard deviation of fractal dimension differences between one test region and each of the 15 most similar reference regions selected by the CBIR scheme using the fractal dimension as a prescreening condition was 0.045 ± 0.028. On the other hand, the mean ± standard deviation of those regions selected by the CBIR scheme without the fractal dimension was 0.104 ± 0.057 (Fig 3). The results indicated that when using the new CBIR scheme, the difference of selected reference ROIs in the fractal dimension was substantially reduced by 56.7% without reducing performance in the classification of suspicious breast masses.

DISCUSSION

Although a large number of CBIR schemes have been developed and tested, there is no universally applicable CBIR scheme for medical imaging applications. All CBIR schemes are domain knowledge dependent, because image characteristics or features vary widely in different applications. Medical imaging is an ideal application for CBIR schemes because of the limited definition of image classes (eg, digital mammograms) and because the meaning and interpretation of medical images are better understood and characterized (3). Among the components of CBIR schemes, feature selection to describe image properties is among the most important. As a ''visual aid'' tool, our previously pilot study demonstrated that the low performance level of a CBIR scheme in the classification of lesions (clinical relevance) might mislead radiologists and reduce their diagnostic performance, while the poor visual similarity between lesions would also result in radiologists' ignoring the CBIR-selected reference ROIs in their decision making (29). Therefore, finding an optimal feature set to improve the performance of CBIR schemes in both clinical relevance and visual similarity is a significant issue at present. However, the most previous studies of developing CBIR schemes for medical images (in particular using mammograms) separately focused on improving either clinical relevance or visual similarity. The unique characteristic of this study is that we developed and assessed a CBIR scheme whose aim is to achieve high performance in both clinical relevance and visual similarity.

In general, the use of multiple features leads to more accurate pattern classification than the use of a single feature because the weaknesses of one feature could be compensated by the strengths of the other features (22). On the other hand, a coarse feature set without refining for the specified purpose does not always promise an improvement in performance.

Therefore, even though any feature is well known as a good feature to describe an image, to combine with a previously well organized feature set, it needs to be evaluated again with the previously developed feature set. In the first experiment, the GA optimization resulted in a new *k*NN-based CBIR scheme that contained all 15 features, including the previously optimized 14 morphologic and intensity distribution features and the fractal dimension, which indicates that the fractal dimension is not a redundant feature or highly correlated to any of the previously selected 14 morphologic and intensity distribution features. As an additional feature used in a *k*NN algorithm, the fractal dimension contributes to improvements in CBIR scheme performance.

Because there are two ways to define the relevance (or performance) of CBIR results, visual or semantic (30), in which visual similarity means that two images ''look visually similar'' regardless of their content, the selected reference-image set must also be considered by observers as actually visually ''similar'' and ''relevant'' to the case being diagnosed. Otherwise, observers will largely ignore the CBIR results (12,29). However, using subjective rating methods to assess visual similarity is very difficult and often unreliable because of the large interobserver variation. Studies have shown that the difference between computerized selection results and the average visual selection results of a panel of radiologists was often smaller than the interobserver selection results (11,31). Thus, developing an objective (or quantitative) index to assess visual similarity may be an important and practical alternative in developing and evaluating CBIR schemes. Previous studies have shown that the fractal dimension can be used as a visually similar texture feature (19–22). Thus, in this study, we selected the fractal dimension as an objective index to increase visual texture similarity in the retrieved reference regions. The potentially clinical utility of using the fractal dimension as an objective index to improve visual similarity depends on whether this will not significantly affect (in particular reduce) the performance of CBIR schemes in selecting clinically relevant images. Our second experiment clearly demonstrated that the use of the fractal dimension as a criterion for visual texture similarity did not significantly affect the performance of the CBIR scheme in the classification of suspicious breast masses depicted on digital mammograms. However, it increased the visual texture similarity by substantially reducing the difference of the fractal dimension between the queried (testing) ROI and each of the similar reference regions selected by the CBIR scheme.

In addition, using the fractal dimension as an objective index of visual similarity assessment has a number of other advantages. First, as the size of a reference database increases, computational efficiency becomes an important issue in the real-time application of CBIR schemes (5). Using the fractal dimension as a prescreening tool, CBIR schemes can discard early a large fraction of ROIs and reduce searching space in the reference database before more computationally complex methods (or algorithms) are used. Second, unlike several other potentially visual similarity comparison indices (ie, mutual information and Pearson's correlation coefficient [13]), the fractal dimensions of all reference ROIs in the database can be precomputed (offline), similar to all other morphologic and intensity distribution features used in our *k*NN algorithm. Thus, the fractal dimension is a unique feature suitable to be implemented in CBIR schemes with real-time computation and comparison capabilities. Third, fractal dimension or analysis is also scale independent (15). Although shape features are a good cue in searching for the same objects in an image, they rely on accurate segmentation results and are not robust to various scales. Therefore, the fractal dimension can serve as a reliable feature in searching for or matching mass regions of the same patient depicted on images acquired during sequential examinations.

We recognize that although the results are encouraging, this was a very preliminary study that addressed a difficult but very important technical challenge in developing and evaluating CBIR schemes applied for medical images. This study had a number of limitations. First, although the fractal dimension and analysis is related to visual similarity in image texture features, only when it is applied to an ideal fractal surface (continuous and truly self-similar) will make the computed fractal dimension values be the same (16). Therefore, in a future study, we will compare other computational methods of the fractal dimension to determine the optimal solution that enables the fractal dimension to have the best capability for classifying suspicious breast masses and searching for texturally similar mass regions depicted on mammography. In addition, although improving texture similarity is an important step toward improving overall visual similarity in selecting medical images using CBIR schemes, whether using the similarity of the fractal dimension can actually produce acceptable visual similarity required in clinical practice by radiologists still needs to be investigated using different observer preference studies. Second, the ultimate clinical utility of CBIR schemes depends on many factors, including (1) scheme performance level, (2) observers' confidence levels in accepting the CBIR-generated results, (3) observers' experience in medical image diagnosis, and (4) the subtlety of the queried cases. This study focused only on improving scheme performance. Many application-related issues need to be investigated further before any CBIR schemes can be optimally used in clinical practice.

In summary, because of the growing sizes of image repositories, their complexity, and the need for reference images to compare similar cases with previously verified results, CBIR has recently been attracting research interest in

medical imaging and informatics. Because of the large difference between human and computer vision, developing an efficient CBIR scheme that can achieve high performance in both clinical relevance and visual similarity remains a difficultly technical challenge. In this preliminary study, we adopted and tested the fractal dimension, a well-recognized texture feature that somehow correlates with visual similarity, as a texture measure in a CBIR scheme to describe the roughness of mass regions. The results of this study indicate that (1) combining the fractal dimension with other morphologic and intensity distribution features is not redundant and may increase the performance of a CBIR scheme, and (2) using the fractal dimension as a prescreening tool to improve visual textural similarity of selected reference ROIs does not significantly affect the performance of a CBIR scheme in the classification of suspicious breast mass regions. Therefore, the fractal dimension can be selected as a ''visually'' and semantically promising feature used in CBIR schemes to improve their performance and computational efficiency.

REFERENCES

1. Muller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefit and future direction. Int J Med Inform 2004; 73:1–23.
2. Lehmann TM, Guld MO, Deselaers T, et al. Automatic categorization of medical images for content-based retrieval and data mining. Comput Med Imaging Graph 2005; 29:143–155.
3. El-Naqa I, Yang Y, Galatsanos NP, Wernick M. A similarity learning approach to content based image retrieval: application to digital mammography. IEEE Trans Med Imaging 2004; 23:1233–1244.
4. Lam MO, Disney T, Raicu DS, Furst J, Channin DS. BRISC—an open source pulmonary nodule image retrieval framework. J Digit Imaging 2007; 20:63–77.
5. Smeulders AMW, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Trans Patt Anal Mach Intell 2000; 22:1349–1380.
6. Park SC, Sukthankar R, Mummert L, Satyanarayanan M, Zheng B. Optimization of reference library used in content-based medical image retrieval scheme. Med Phys 2007; 34:4331–4339.
7. Muramatsu C, Li Q, Suzuki K, et al. Investigation of psychophysical measures for evaluation of similar images for mammographic masses: preliminary results. Med Phys 2005; 32:2295–2304.
8. Karam OH, Hamad AM, Ghoniemy S, Rady S. Enhancement of wavelet-based medical image retrieval through feature evaluation using an information gain measure. In: Proceedings of the Eighteenth Annual ACM Symposium on Applied Computing. New York: Association for Computing Machinery; 2003:9–12.
9. Alto H, Rangayyan RM, Desautels JEL. Content-based retrieval and analysis of mammographic masses. J. Electron Imaging 2005; 14:1–17.
10. Wei C, Li C, Wilson R. A general framework for content-based medical image retrieval with its application to mammograms. Proc SPIE 2005; 5748:134–143.
11. Muramatsu C, Li Q, Schmidt RA, et al. Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings. Med Phys 2007; 34:2890–2895.
12. Zheng B, Lu A, Hardesty LA, Gur D. A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment. Med Phys 2006; 33:111–117.
13. Wang XH, Park SC, Zheng B. Improving performance of content-based image retrieval schemes in searching for similar breast mass regions: an assessment. Phys Med Biol 2009; 54:949–961.
14. Lefebvre F, Benali H, Gilles R, Kahn E, Paola RD. A Fractal approach to the segmentation of microcalcifications in digital mammograms. Med Phys 1995; 22:381–390.
15. Sedivy R, Windischberger C, Svozil K, Moser E, Breitenecker G. Fractal analysis: an objective method for identifying atypical nuclei in dysplastic lesions of the cervix uteri. Gynecol Oncol 1999; 75:78–83.
16. Geraets WG, van der Stelt PF. Fractal properties of bone. Dentomaxillofac Radiol 2000; 29:144–153.
17. Li H, Giger ML, Olopade OI, Lanm L. Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment. Acad Radiol 2007; 4:513–521.
18. Hendee WR. Cognitive interpretation of visual signals. In: Hendee WR, Wells PNT, eds. The perception of visual information. 2nd ed. New York: Springer-Verlag, 1997; 149–176.
19. Rangayyan RM, Nguyen TM. Fractal analysis of contours of breast masses in mammograms. J Digit Imaging 2007; 20:223–237.
20. Xu C, Zhuang TG, Hua YQ. Discrimination of ground glass opacity on lung HRCT images using visual complexity measurements. In: Proceeding of the Second Joint EMBS/BMES Conference. New York: IEEE; 2002:1116–1117.
21. Peng F, Yu X, Xu G, Xia Q. Fuzzy classification based on fractal features for undersea image. Int J Inform Technol 2005; 11:133–142.
22. Chevallet JP, Maillot N, Lim JH. Concept propagation based on visual similarity application to medical image annotation. Third Asia Inform Retriev Symp 2006; 4182:514–521.
23. Soares F, Andruszkiewic P, Freire MM, Cruz P, Pereira M. Self-similarity analysis applied to 2D breast cancer imaging. Proc Int Conf Syst Network Commun 2007; 1:1–6.
24. Velanovich V. Fractal analysis of mammographic lesions: a feasibility study quantifying the difference between benign and malignant masses. Am J Med Sci 1996; 311:211–214.
25. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. Automated analysis of mammographic densities. Phys Med Biol 1996; 41:909–923.
26. Chen J, Zheng B, Chang YH, Shaw CC, Towers JD, Gur D. Fractal analysis of trabecular patterns in projection radiographs. An assessment. Invest Radiol 1994; 29:624–629.
27. Millard J, Augat P, Link TM, et al. Power spectral analysis of vertebral trabecular bone structure from radiographs: orientation dependence and correlation with bone mineral density and mechanical properties. Calcifield Tissue Int 1998; 63:482–489.
28. Mitchell TM. Machine learning. Boston, MA: WCB/McGraw-Hill, 1997.
29. Zheng B, Abrams G, Britton CA, et al. Evaluation of an interactive computer-aided diagnosis scheme for mammography: a pilot study. Proc SPIE 2007; 6515:65151M-1–65151M-8.
30. Felipe JC, Traina C, Traina AJ. A new family of distance functions for perceptual similarity retrieval of medical images. J Digit Imaging 2009; 22:183–201.
31. Zheng B, Abrams G, Leader JK, et al. Agreement between ratings of mass spiculations by observers and a computer scheme. Proc SPIE 2007; 6514:65141P-1–65141P-8.