Computer Vision and Image Understanding **88,** 119–151 (2002)
doi:10.1006/cviu.2002.0972

Using Human Perceptual Categories for Content-Based Retrieval from a Medical Image Database

Chi-Ren Shyu, Christina Pavlopoulou, Avinash C. Kak, and Carla E. Brodley

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907
E-mail: chiren@purdue.edu, kak@purdue.edu, brodley@purdue.edu

and

Lynn S. Broderick

Department of Radiology, University of Wisconsin Hospital, Madison, Wisconsin 53792
E-mail: lsbroderick@facstaff.wisc.edu

Received August 29, 2000; accepted June 25, 2002

It is often difficult to come up with a well-principled approach to the selection of low-level features for characterizing images for content-based retrieval. This is particularly true for medical imagery, where gross characterizations on the basis of color and other global properties do not work. An alternative for medical imagery consists of the "scattershot" approach that first extracts a large number of features from an image and then reduces the dimensionality of the feature space by applying a feature selection algorithm such as the Sequential Forward Selection method.

This contribution presents a better alternative to initial feature extraction for medical imagery. The proposed new approach consists of (i) eliciting from the domain experts (physicians, in our case) the perceptual categories they use to recognize diseases in images; (ii) applying a suite of operators to the images to detect the presence or the absence of these perceptual categories; (iii) ascertaining the discriminatory power of the perceptual categories through statistical testing; and, finally, (iv) devising a retrieval algorithm using the perceptual categories. In this paper we will present our proposed approach for the domain of high-resolution computed tomography (HRCT) images of the lung. Our empirical evaluation shows that feature extraction based on physicians' perceptual categories achieves significantly higher retrieval precision than the traditional scattershot approach. Moreover, the use of perceptually based features gives the system the ability to provide an explanation for its retrieval decisions, thereby instilling more confidence in its users.

1077-3142/02 $35.00

Key Words: medical image databases; CBIR; feature extraction; feature design; human perception.

1. INTRODUCTION

Identifying what features to extract and devising algorithms for doing so is a critical step in the construction of any content-based image retrieval (CBIR) system. Important questions that arise during this phase include what part of the image the features should represent, and how one decides what features to extract in a disciplined way.

This problem of feature extraction and selection is not so acute for CBIR systems that have focussed on general purpose imagery of outdoor scenes, especially if retrieval is with respect to some global property of an image [1–8]. Such images tend to be rich in color and texture and can often be characterized by global signatures based on such properties. It should be mentioned that even in the domain of general purpose outdoor imagery, one may need to carry out retrieval with respect to some highly localized attribute, as when you are trying to find outdoor images with, say, a squirrel in them. To address this problem, more recently researchers have attempted to describe images with texture, color, and shape features extracted from local regions [9–14].

Medical CBIR systems are different from general purpose ones in several ways. For one, the retrieval has to take place with respect to pathology bearing regions (PBR) that tend to be highly localized. This means that retrieval on the basis of global signatures would make no sense at all for medical databases. Additionally, the PBRs cannot be segmented out automatically for many medical domains—which necessitates a physician-in-the-loop approach for both training the CBIR system and for its actual use. Another factor that makes medical CBIR very different from general purpose CBIR is that the ground truth is available—in the form of disease categories for the images—and can be used for coming up with performance numbers.

There have not been very many prior contributions to medical CBIR. A retrieval system for megnetic resonance images (MRI) of the brain has been reported in [15, 16]. The main image feature that is used for characterizing these images is the shape of the ventricular region. In another system reported in [17], the images in the database consist of a single tumor in the center without any background texture. The system presented in [18] aims at aiding physicians in the diagnosis of lymphoproliferative disorders of the blood. Shape and texture features are used to characterize the regions of interest delineated by the user. Cai *et al.* describe in [19] a CBIR system for positron emission tomographic (PET) images of the brain. In this case, a set of physiological features as well as text are used for retrieval. On the other hand, in [20] a retrieval system for volumetric images of the brain is introduced. The ASSERT system reported by us [21] is designed for high-resolution computed tomography (HRCT) images of the lung where a rich set of textural features derived from the disease-bearing regions are important for the characterization of the images. The physician is an integral part of ASSERT, in the sense that it is the physician who delineates the PBRs when an image is entered into the database and, also, in the query image.

Although the systems referenced above demonstrated experimentally the efficacy of the features devised to characterize the given images, identifying these features was the outcome of ingenuity, intuition and experimentation. This might not be a problem for the systems

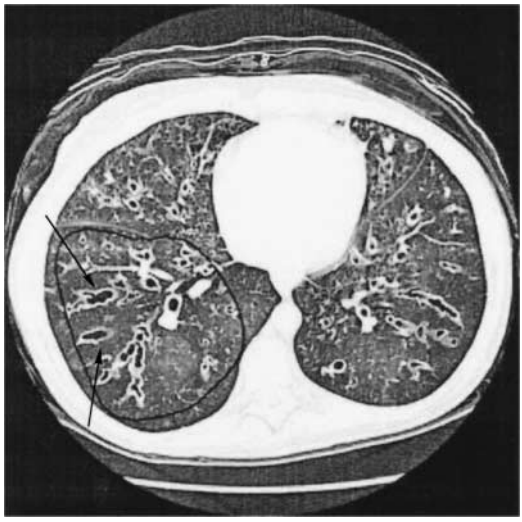

FIG. 1. An HRCT image of lung with the disease bronchiectasis which is highly localized. The two dark arrows point to the pathology.

reported in [15–17], where shape alone is sufficient for retrieval, but it is an issue for the more complex lung HRCT images (see, for example, Fig. 1). In this domain it is not clear at first glance what features are needed. The approach we followed in [21, 22] was what may be referred to as the *scattershot* approach. First an exhaustive set of low-level features to characterize image pathology was extracted. Next, the dimensionality of the feature space was reduced by searching for a representative subset using a greedy algorithm, such as the Sequential Forward Selection search [23], with the aim of retaining only those features that are maximally discriminatory with regard to the different diseases.

The work reported in this paper presents a better alternative to the scattershot approach. Our new approach is based on the rationale that medical images should be characterized on the basis of the visual patterns that the domain experts, in our case the expert physicians, rely upon for disease detection. We refer to these patterns as the domain expert's *perceptual categories*.

The question then becomes as to how to go about eliciting the perceptual categories from domain experts. Fortunately, in some domains, the domain of HRCT being one of them, there has already been considerable cogitation among the domain experts about what the relevant perceptual categories are. The scientific literature in these domains talks about the specific patterns the physicians should look for in order to declare the presence or the absence of various diseases.

To incorporate this domain knowledge in the feature extraction process, our first challenge was to come up with low-level features that would detect the presence or the absence of a perceptual category. We addressed this issue with a two-step approach: First, we guessed what low-level features would be good for detecting each perceptual category; and, second, we used the tools of MANOVA to ascertain the power of the chosen low-level features to discriminate between the different perceptual categories. After determining what perceptual categories are present in a PBR, we determine the disease of the PBR. Note that ours is essentially a *hierarchical* approach to feature design: low-level features are used to

describe the perceptual categories and then the perceptual categories are used to describe the diseases.

An advantage of the hierarchical approach is that it makes it easier to decide what features are needed to characterize the PBRs. The reason for this is that it is easier to come up with features that are suitable to describe well-defined entities such as the perceptual categories than to find features that solve the entire problem of disease classification. Furthermore, the use of perceptually based features gives the system the ability to provide an explanation for its retrieval decisions, thereby instilling more confidence in its users.

Taking into account the physician's perceptual categories, this paper presents a new approach to CBIR for medical image databases. We start by describing the various perceptual categories used by expert physicians who specialize in the detection of emphysema-like diseases in the HRCT images of the lung (Section 2). For each perceptual category, we list the low-level features that can be expected to indicate the presence of that perceptual category. This section is lengthy but necessary for a complete and accurate description of the system; it is not critical though for understanding the rest of the paper.

In Section 3 we describe how we test with MANOVA whether or not a chosen set of low-level features is actually measuring a perceptual category (Section 3). Subsequently in the same section, we describe our use of the Bonferroni method of multiple comparisons to give different weights to the low-level features in order to increase the measured "separation" between the various perceptual categories. These weights are then used to form linear classifiers for retrieval. Query and matching processes for retrieving images based on perceptual categories and disease categories are discussed in Section 4. Finally, in Section 5 we present retrieval results using the physicians' perceptual categories and compare them to the results using our earlier scattershot approach.

2. PHYSICIANS' PERCEPTUAL CATEGORIES FOR RECOGNIZING LUNG PATHOLOGY

Figure 2 shows the perceptual categories that physicians use for describing the visual structure of a pathology bearing region (PBR) in an HRCT image of the lungs. The four major categories are [24]: *linear and reticular opacities, nodular opacities, high-density areas*, and *low-density areas*. These categories are major in the sense that, in the physician's mind, they correlate strongly with the various lung diseases. The leaf nodes of the tree in Fig. 2 show the subcategories that the physicians actually use for labeling the PBRs. A PBR may exhibit a pathology corresponding to the major category "high-density areas," but the actual visual structure inside the PBR would either be "ground glass" or "calcification," corresponding to the two leaf nodes of this major category in Fig. 2. The rest of this section provides further details regarding the visual structures associated with these categories and describes the low-level features designed to detect each category.

2.1. Linear and Reticular Opacities

These patterns consist of line-like structures that can either be straight and elongated, web-like, or circular with dot-like protrusions. This major category encompasses the following six subcategories:

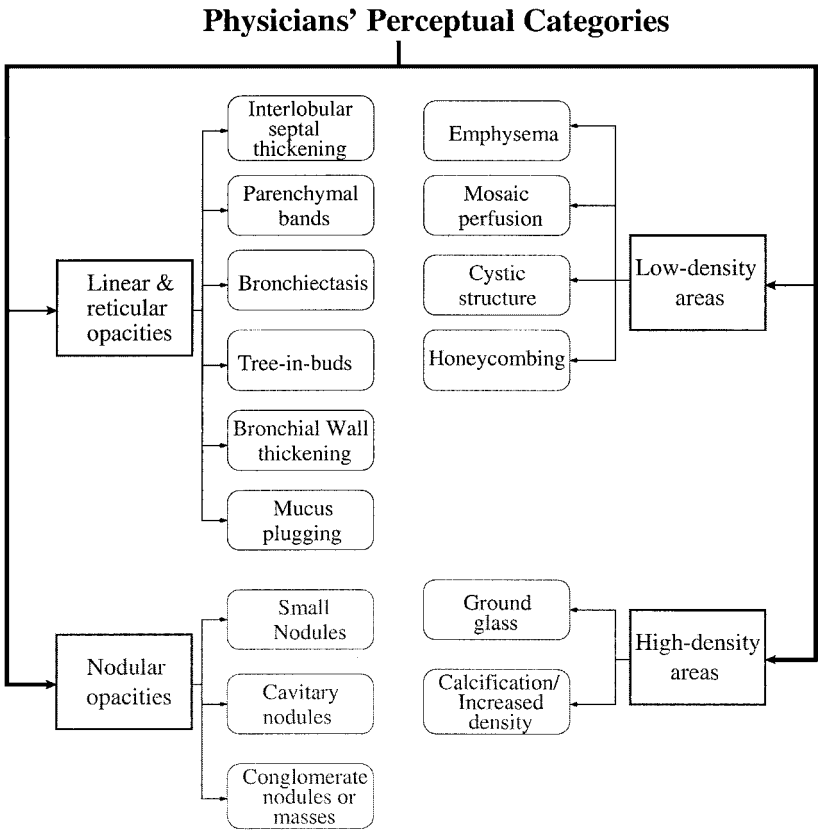

FIG. 2. Perceptual categories used by physicians for the domain of HRCT images of the lung.

1. Interlobular Septal Thickening

A lung consists of lobes, two in the left lung and three in the right lung, and each lobe contains smaller structures called lobules, typically between 1 and 2 cm in diameter. Interlobular septal thickening refers to the thickening of the spaces between the lobules. Shown at upper left in the cartoon representation in Fig. 3 is the pattern corresponding to interlobular septal thickening. The spaces enclosed by the white lines are the lobules. Shown in Fig. 4 is an HRCT image in which the dark arrows point to a region that was marked by an expert physician as exhibiting interlobular septal thickening. The visual pattern formed by the white "streaks" inside the physician-delineated PBR in Fig. 4, and shown more vividly at the upper left in Fig. 3, is also referred to as a reticular pattern in medical literature. For the purpose of characterization, such patterns respond to the skeletonization of a PBR, followed by the extraction of the following parameters associated with the lobules enclosed by the white contours in the upper left in Fig. 3:

- f_1^{SEP}: This is the number of lobules, which can be estimated by carrying out connectivity analysis of the PBR pixels after they are skeletonized. Figure 4b shows a skeletonization of the PBR outlined by a physician in the HRCT image on the left. In Fig. 4c, we show the lobule pixels enclosed by the skeleton branches in Fig. 4b. These lobule pixels are obtained by taking a complement of the skeletal image, followed by shrink and grow operations,

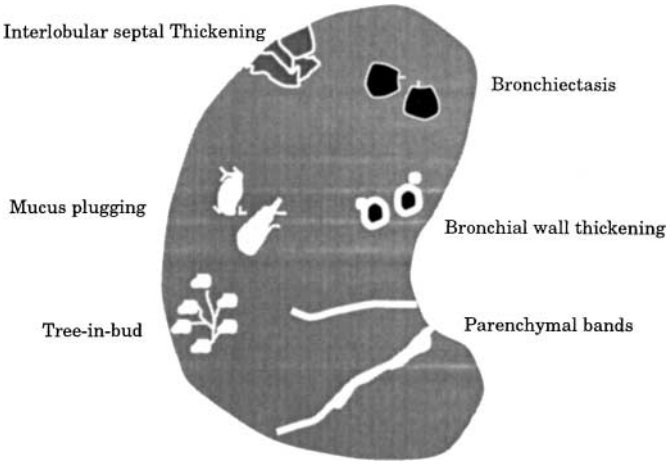

FIG. 3. Cartoon representation of the perceptual categories that fall under the label "linear and reticular opacities."

where the grow operation is performed with a constraint on the homogeneity of the pixel gray levels. A simple count of the regions shown in Fig. 4c yields f_1^{SEP}.

- f_2^{SEP}: This is the average area of a lobule. Extraction of this parameter is a simple extension of the procedure for extracting f_1^{SEP}. All that needs to be done is to count the number of pixels in each lobule and find the average value of this number for all the lobules.
- f_3^{SEP}: This is the average difference between the lobule gray levels and the gray levels of the white boundaries enclosing the lobules. The value of this parameter can be calculated by a straightforward extension of the algorithm for extracting f_1^{SEP}.

2. *Parenchymal Bands*

These are long and thick white lines in the images, caused by the presence of high-attenuation tissues in the lung that often touch the boundary of the lung. They are shown at the lower right in the simplified rendition of the perceptual categories in Fig. 3. An HRCT example of such patterns is shown in Fig. 5a where the two arrows point to parenchymal bands in a physician-outlined PBR. To extract such patterns, we first apply to the PBR

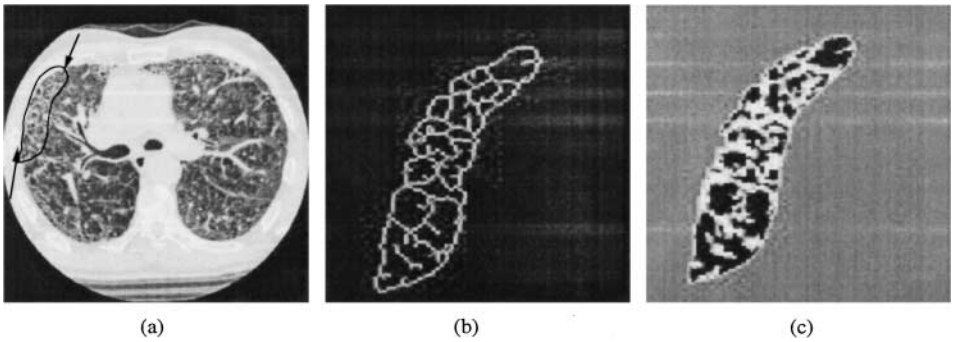

FIG. 4. The two arrows in (a) point to interlobular septal thickening inside a physician-delineated PBR. Shown in (b) is an expanded view of the PBR after it is skeletonized. Shown in (c) are the lobules extracted. The values of the characterizing parameters for the example shown here are $f_1^{SEP} = 28$, $f_2^{SEP} = 79.78$, and $f_3^{SEP} = 57.98$.

FIG. 5. The two arrows shown in (a) point to parenchymal bands inside a physician-delineated PBR, which, in this case, consists of an entire lung. Shown in (b) is the output obtained after lung-region extraction and an application of a high threshold. The bands, shown in (c), are obtained by rejecting those components in (b) that do not touch the lung boundary. The values of the characterizing parameters for the example shown here are $f_1^{PAR} = 3807.00$ and $f_2^{PAR} = 0.01$.

our lung boundary extraction algorithm described in [21].[1] Simple thresholding of the region interior to the bounding contour yields parenchymal bands and other artifact objects (see Fig. 5b). Nonband pixels are discarded by carrying out a connectivity analysis of the thresholded object and rejecting those that are not touching the lung boundary, yielding just the bands as shown in Fig. 5c. These bands are then characterized by

- f_1^{PAR}: The average area as obtained by simply counting the number of pixels in the bands shown in Fig. 5c.
- f_2^{PAR}: The average form factor given by

$$f_2^{PAR} = \frac{4\pi * Area}{Perimeter^2}, \tag{1}$$

where the *Area* and the *Perimeter* are calculated for each band separately.

3. *Bronchiectasis*

Bronchi are air-filled passages of the lung that, due to their low attenuation, show up as dark regions. Bronchiectasis means enlargement of the bronchi, as shown by the rendition at the upper right in Fig. 3. The arrows in Fig. 6a point to such structures. The dark region inside a bronchus is also referred to as a *lumen*. The enlargement of the bronchi is often accompanied by a thickening of the walls of the bronchi that show up as white contours surrounding the lumen. These patterns are best extracted by double thresholding [25] the physician-outlined PBRs. For the PBR of the example shown in Fig. 6a, the double thresholding yields the lumens and other nonlumen artifacts shown in Fig. 6b. The nonlumen artifacts are rejected by using the criterion that only the lumens are enclosed by the high-attenuation bronchial walls. These walls are detected by growing the boundaries of the lumens until no further high-valued pixels can be included. Shown in Fig. 6c are the bronchial walls extracted in this manner. The regions enclosed by these walls are the

[1] Note that since parenchymal bands usually touch the lung boundary, any PBR outlined by a physician that pertains to this perceptual category will also include at least a portion of the lung boundary. In many cases, it includes an entire lung, as shown by the example in Fig. 5a.

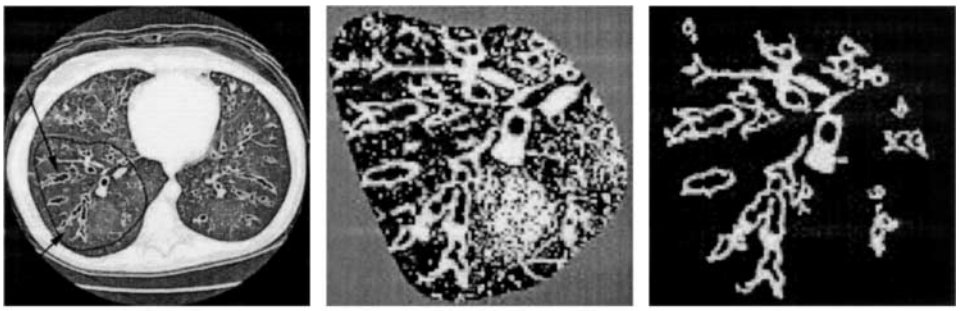

FIG. 6. The two arrows shown in (a) point to bronchiectasis inside a physician-delineated PBR. The output obtained when a dual threshold is applied to just the PBR is shown in (b). The nonlumen artifacts in (b) are rejected in the manner described in the text and the output obtained is as shown in (c). The values of the characterizing parameters for the example shown here are $f_1^{BRO} = 27.00$, $f_2^{BRO} = 3.82$, $f_3^{BRO} = 54.48$, $f_4^{BRO} = 0.334$, and $f_5^{BRO} = 131$.

lumens; for the example under consideration, these are shown in Fig. 6c. The lumens are characterized by:

- f_1^{BRO}: The number of lumens. This parameter is obtained simply by counting the completely-enclosed low-attenuation regions in an output such as shown in Fig. 6c.
- f_2^{BRO}: The average thickness of the bronchial walls surrounding the lumens. These are obtained trivially from the output like the one shown in Fig. 6c.
- f_3^{BRO}: The average area of the lumens obtained trivially from an output such as shown in Fig. 6c.
- f_4^{BRO}: The average ratio of the lumen radius to the thickness of the bronchial walls.
- f_5^{BRO}: The grey difference between the walls and the lumens.

4. Tree-in-Bud

In some patients, the small airways, also called bronchioles, may be dilated and filled with pus, mucus or inflammatory exudate, leading to the appearance of high X-ray attenuation (meaning, white) clusters of pixels that form "spotty" regions, as rendered at the lower left in Fig. 3. The arrows in the HRCT image of Fig. 7 point to such a region. The gray levels for those "spotty" tissues are high and close to the grey levels for the bone. Therefore, they can be extracted by applying a high threshold to a PBR. For the PBR of the example shown in Fig. 7a, shown in 7b is just the PBR after a threshold is applied to it.[2] Since the application of this threshold yields a single connected component for the spotty regions, this category can be characterized by

- f_1^{TIB}: The area of the connected component. This parameter is obtained simply by counting the number of pixels in an output such as shown in Fig. 7b.
- f_2^{TIB}: The average difference between the gray levels inside the connected component and those outside in the PBR.
- f_3^{TIB}: The form factor of the connected component as given by Eq. (1).

[2] The high threshold used for this purpose is chosen automatically for each image by making it equal to the average gray levels in the spinal-cord region of an HRCT image. This region, roughly in the same portion of each image, consists of the highest gray levels and therefore can be identified by a straightforward histogram analysis.

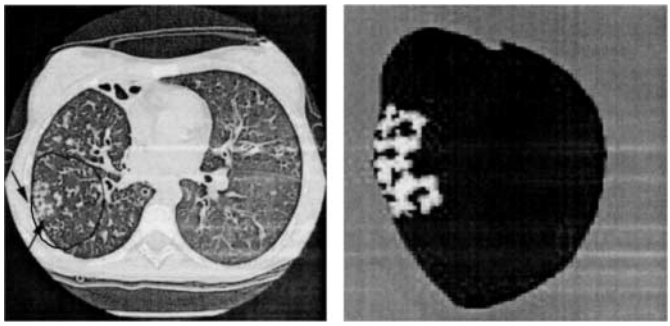

FIG. 7. The arrows shown in (a) point to the tree-in-buds perceptual category inside a physician-delineated PBR. Shown in (b) is the output obtained when the PBR is subject to a high threshold in the manner discussed in the text. The values of the characterizing parameters for the example shown here are $f_1^{TIB} = 756$, $f_2^{TIB} = 134$, and $f_3^{TIB} = 0.10$.

5. Bronchial Wall Thickening

As the name implies, this pattern corresponds to the thickening of the walls of the bronchi. A rendition of this perceptual category is shown at the middle right in Fig. 3. When a patient exhibits this condition, the artery that adjoins a bronchus shows up as a smaller structure, giving the appearance to the combination of the bronchus and the artery as a "signet ring," in which the ring is made of the dilated bronchus and the "diamond" of the adjoining artery. The patterns shown in the middle right of Fig. 3 are the signet-ring patterns. The two arrows in Fig. 8a point to bronchial wall thickening. This condition in a patient is also known as peribronchovascular interstitial thickening. These patterns are detected and characterized in a manner identical to what was used for bronchiectasis. Therefore, the same five parameters that were listed for bronchiectasis earlier are used for characterizing bronchial wall thickening. Of course, the values taken on by some of those parameters—for example, the parameters f_2^{BRO}, f_3^{BRO}, and f_3^{BRO}—will be different for the current case. Shown in Fig. 8b is an enlarged view of the PBR after it is taken through the processing steps mentioned previously for the case of bronchiectasis.

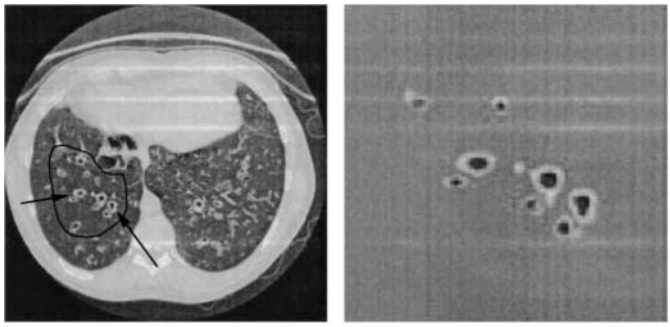

FIG. 8. The arrows shown in (a) point to bronchial wall thickening inside a physician-delineated PBR. Shown in (b) are the visual structures corresponding to this perceptual category. These are obtained in a manner similar to that for the bronchiectasis perceptual category. The values of the characterizing parameters for the example shown here are $f_1^{BRO} = 9.00$, $f_2^{BRO} = 3.82$, $f_3^{BRO} = 36.44$, $f_4^{BRO} = 1.51$, and $f_5^{BRO} = 146$.

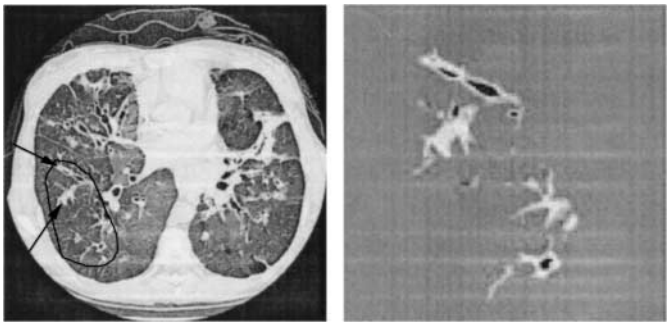

FIG. 9. The arrows shown in (a) point to mucus plugging inside a physician-delineated PBR. Shown in (b) are the visual structures corresponding to this category. These are obtained in a manner that is similar to that for tree-in-buds. The values of the characterizing parameters for the example shown are $f_1^{MUP} = 598$, $f_2^{MUP} = 0.18$, and $f_3^{MUP} = 5.0$.

6. Mucus Plugging

When dilated bronchi become filled with mucus, pus or inflammatory exudate, they show up as large clusters of white pixels (on a scale larger than is the case with the individual clusters in the tree-in-bud pattern). Also, sometimes these clusters exhibit linear and/or branching structures, as shown in the middle left in Fig. 3. It is almost always the case that one finds bronchial structures in the vicinity of such patterns. Therefore, the presence of bronchial structures can be used as supporting evidence for this pattern. These patterns are detected in a manner identical to that for tree-in-bud, except for the difference that the output obtained after applying a high threshold to the PBR is not now a single connected component. Also, the multiple components obtained are analyzed for the presence of a bronchial structure in the vicinity of each connected blob of high gray levels. For the example of Fig. 9a in which the two arrows point to mucus plugging, the output obtained by applying a high threshold[3] to the PBR is shown in 9b. The high pixels that surround the dark holes constitute the bronchial structures. Therefore, in this case, the rest of the connected components are taken as constituting mucus plugging. This perceptual category is characterized by the following parameters:

- f_1^{MUP}: Average area of the mucus-plugging components.
- f_2^{MUP}: Average form factor of the mucus-plugging components, defined as before for parenchymal bands.
- f_3^{MUP}: Number of bronchial structures in the PBR.

2.2. Nodular Opacities

The patterns that fall under this major category consist of nodules of different shapes, sizes, spatial distributions, and at different states of agglomeration, the differences captured by the following three subcategories:

[3] The value of this threshold is chosen in a manner identical to that for the tree-in-bud pattern.

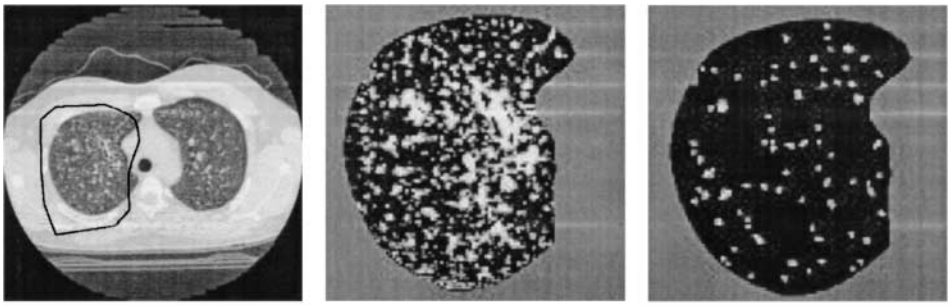

FIG. 10. Shown in (a) is a physician-delineated PBR containing small nodules. The PBR after it is thresholded is shown in (b). The nonnodule artifacts are removed on the basis of the roundness property and the resulting nodules are shown in (c). The values of the characterizing parameters for this PBR are $f_1^{SNO} = 75$, $f_2^{SNO} = 1.04$, $f_3^{SNO} = 177.36$, $f_4^{SNO} = 8.90$, $f_5^{SNO} = 3.60$, and $f_6^{SNO} = [48, 25, 1, 0, 0, 0]$.

1. Small Nodules

These are roughly round and less than one centimeter in diameter. Their distribution carries diagnostic information. When the distribution is random, then the nodules appear widely and evenly throughout the lung as shown in Fig. 10. Distributions become nonuniform when nodules attach themselves to the boundaries of the lungs or to the fissures. The gray values associated with nodular opacities carry important information with regard to whether the tissue is benign or malignant. HRCT images that show this type of evidence can be further categorized on the basis of the size and locational distributions associated with the nodular opacities. Patterns corresponding to this perceptual category can be extracted by first applying a high threshold to the PBR, followed by the measurement of "roundness" property using the formula shown below, and by discarding nonnodule objects on the basis of roundness. For each PBR, the threshold is selected automatically by using Otsu's threshold selection algorithm [26], which in the present context was discussed in some detail in [21]. Shown in Fig. 10a is a physician-delineated PBR for this example. The output obtained by applying a threshold to the PBR is shown in Fig. 10b. The nonnodule objects in this output are rejected on the basis of the value of "roundness" parameter defined below. Only those objects are accepted whose "roundness" parameter value is between 0.9 and 1.1. Shown in Fig. 10c are the nodule pixels extracted in this manner for the PBR of Fig. 10a. This perceptual category is characterized by the following parameters:

- f_1^{SNO}: Number of small nodules. This parameter is obtained by counting the number of labeled regions from Fig. 10c.
- f_2^{SNO}: Average roundness of small nodules. The roundness is given by

$$f_2^{SNO} = \frac{4 * Area}{\pi * Diameter^2}, \tag{2}$$

where the *Area* and the *Diameter* are calculated for each extracted small nodule.

- f_3^{SNO}: Average gray level of small nodules.
- f_4^{SNO}: Average nearest-neighbor (NN) distance between the nodule centers [27].
- f_5^{SNO}: Standard deviation of NN distance.
- f_6^{SNO}: Histogram of NN distances. We use six bins for this histogram—a number arrived at by trial and error. Each bin spans a distance of five pixels.

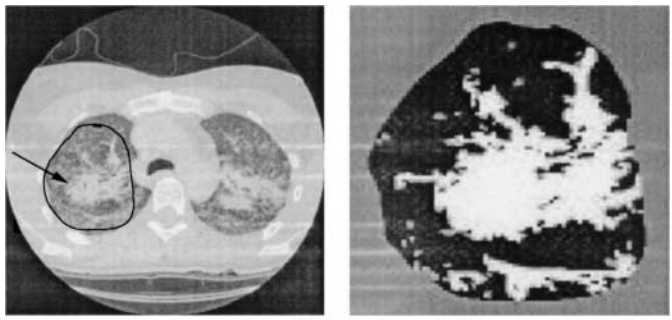

FIG. 11. The arrows in (a) point to conglomerate nodules inside a physician-delineated PBR. The PBR is thresholded and the nonconglomerate nodules are rejected on the basis of roundness. (b) The values of the characterizing parameters for this example are $f_1^{CON} = 4.0$, $f_2^{CON} = 0.34$.

2. *Conglomerate Nodules*

Large nodules usually have irregular shape, whose "diameter" exceeds 1 cm. Sometimes large nodules agglomerate into large masses, as shown in Fig. 11. The conglomerate nodules are extracted with a lower threshold on the roundness parameter. In other words, the value of the roundness threshold is keyed to the size of the object extracted after thresholding. Shown in Fig. 11a is a physician-delineated PBR containing this perceptual category. With processing similar to that for the case of small nodules but with relaxed conditions on the roundness property, the output obtained for the PBR is as shown in Fig. 11b. The regions thus extracted are characterized by the following parameters:

- f_1^{CON}: The number of large nodules. Physicians consider a nodule large if its diameter exceeds six pixels.
- f_2^{CON}: The fraction of PBR occupied by the conglomerate nodules.

3. *Cavitary Nodules*

For patients suffering from pneumonia, the large nodules can exhibit holes inside them. Those holes correspond to the dead lung tissue. Figure 12a shows this perceptual category

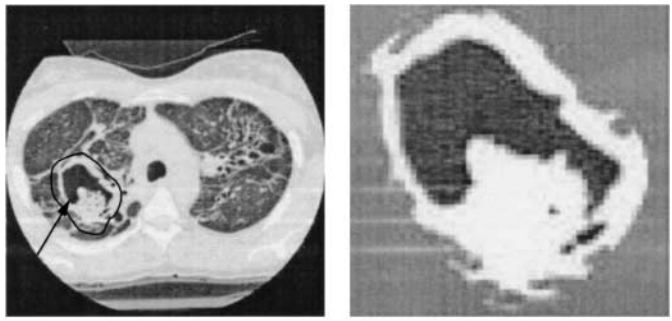

FIG. 12. Shown in (a) is a physician-delineated PBR containing a cavitary nodule. The PBR after it is subject to the extraction of this nodule is shown in (b). The values of the characterizing parameters for this example are $f_1^{CAV} = 419.57$, $f_2^{CAV} = 0.81$, $f_3^{CAV} = 123$.

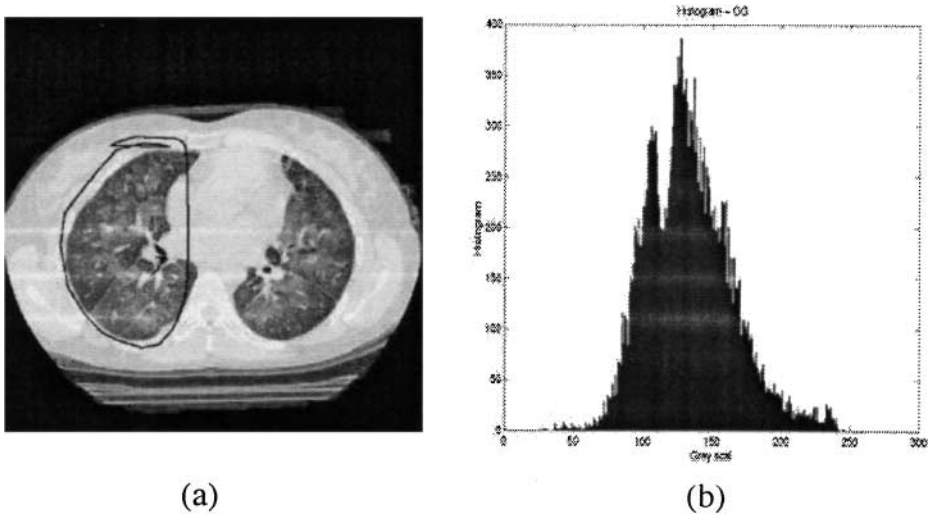

FIG. 13. Shown in (a) is a physician-delineated PBR for the case of ground glass. Shown in (b) is the histogram for the pixels in the PBR. The values of the characterizing parameters for this example are $f_1^{GG} = 0.62$, $f_2^{GG} = 49$, $f_3^{GG} = 162.98$, and $f_4^{GG} = 113.64$.

inside a physician-delineated PBR. Since such patterns can be extracted in a manner identical to that for bronchial structures, we do not describe the feature extraction methods here. For the example here, Fig. 12b shows the PBR after the extraction of the pattern corresponding to this perceptual category. Such patterns are characterized by:

- f_1^{CAV}: Average area of cavities.
- f_2^{CAV}: Fraction of the PBR occupied by the cavities.
- f_3^{CAV}: Average gray level difference between the walls surrounding the cavities and the cavities themselves.

2.3. High-Density Areas

For some lung diseases, an entire lung may exhibit a generally elevated brightness level in comparison to a normal lung. When that happens, the elevated gray level in itself becomes a visual characterization of the disease, a characterization that goes under the label "high-density areas." There are two subcategories to consider for this case:

1. Ground-Glass Opacities

Fig. 13a shows a PBR that exhibits ground-glass opacity. Note that the generally elevated brightness of the lung does not obscure the underlying vessels. The vessels can be seen clearly in the lungs even though the tissues everywhere are characterized by a higher level of attenuation. Algorithms capable of separating the normal tissues from the ground-glass tissues make use of the fact that gray-level histogram for the latter case is bimodal, whereas it is primarily unimodal for the normal tissues.[4] The threshold corresponding to the dip

[4] For some images, this histogram may be multi-modal, but with a marked and easily identifiable dip that separates the normal-tissue pixels from the diseased pixels. This threshold corresponding to this dip can again be extracted by Otsu's algorithm.

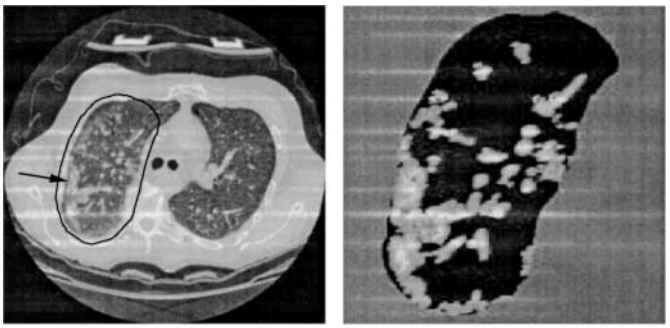

FIG. 14. Shown in (a) is a physician-delineated PBR exhibiting calcification. The PBR after it is subject to thresholding operations is shown in (b). The values of the characterizing parameters for this example are $f_1^{CAL} = 11$, $f_2^{CAL} = 557.72$, $f_3^{CAL} = 170.16$, and $f_4^{CAL} = 0.28$.

between the two humps of the histogram is detected by applying Otsu's algorithm [26] to the histogram. Figure 13b shows the histogram for the example here. The following parameters are extracted from such histograms:

- f_1^{GG}: Ratio of the number of pixels beyond the threshold that separates the two major humps of the histogram to the total number of pixels in the PBR.
- f_2^{GG}: Average gray-level difference between the pixels corresponding to the two major humps of the histogram.
- f_3^{GG}: Average gray level for the pixels corresponding to the "higher" hump of the histogram.
- f_4^{GG}: Average gray level of the pixels in the "lower" hump of the histogram.

2. Calcification

The overall visual effect in an HRCT image with calcification is that of marked increase in density, similar to bone. The dark arrow in Fig. 14a points to calcified patterns. These patterns are detected in a manner identical to that for extracting regions corresponding to the mucus plugging perceptual category. Figure 14(b) shows the PBR after those processing steps. These patterns can be characterized by the following parameters:

- f_1^{CAL}: Number of connected components.
- f_2^{CAL}: Average area of the connected components.
- f_3^{CAL}: Average grey level of the connected components.
- f_4^{CAL}: The fraction of the PBR area occupied by the connected components.

2.4. Low-Density Areas

Most of the previously mentioned perceptual categories consist predominantly of increased attenuation pixels (meaning pixels of higher gray levels). The defining characteristics of the category we will describe in this section are set by pixels whose gray levels are darker than the average—that is, pixels for tissues with low attenuation. There are four subcategories to consider under this category:

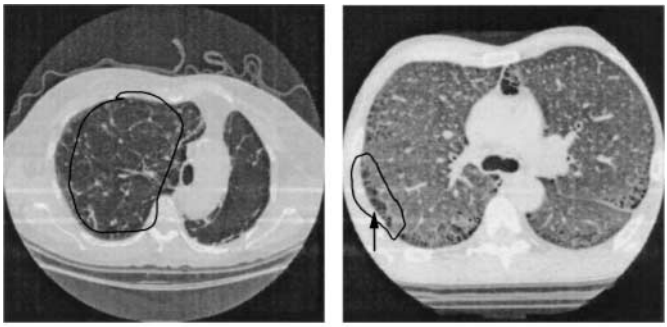

FIG. 15. Shown in (a) is a physician-delineated PBR for the case of centrilobular emphysema. The values of the characterizing parameters for this example are: $f_1^{EMP} = 65.38$, $f_2^{EMP} = 1.32$, $f_3^{EMP} = 0$. Shown in (b) is a physician-delineated PBR for the case of paraseptal emphysema. The values of the characterizing parameters for this example are: $f_1^{EMP} = 103.02$, $f_2^{EMP} = 0.76$, $f_3^{EMP} = 3$.

1. Emphysema

A PBR exhibits the pattern emphysema if the gray levels are significantly lower compared to what a physician expects to see in a normal healthy lung. These reduced gray level areas may occupy a part of a lung or an entire lung region, but are likely to be found more frequently in the upper lobes of a lung. Also, when the disease becomes severe, these areas may join together to form a large region of low attenuation. The physician-delineated PBR in Fig. 15a shows centrilobular emphysema and the one in Fig. 15b paraseptal emphysema. While centrilobular emphysema manifests itself in the form of a large number of areas with significantly low gray levels inside the lung regions; for paraseptal emphysema the low gray-level regions occur adjacent to the boundaries of the lung or in the vicinity of the fissures. Since there are no special visual patterns associated with centrilobular emphysema except the low-gray levels and the homogeneous texture properties, we characterize such a PBR by the following measurements:

- f_1^{EMP}: Average gray level of the PBR.
- f_2^{EMP}: Homogeneity from the cooccurrence matrices [28].
- f_3^{EMP}: The number of low-gray-level regions adjacent to the lung boundary.

2. Lung Cysts

These are thin-walled, well-defined, and circumscribed lesions containing air. The PBR shown in Fig. 16a exhibits the visual form corresponding to lung cysts. These forms are differentiated from emphysema by their discernible walls. These patterns are detected in a manner identical to what was used for bronchiectasis since both categories have patterns consisting of dark regions surrounded by white walls. These patterns are characterized by:

- f_1^{CYS}: Number of individual low gray level regions.
- f_2^{CYS}: Average area of the low gray level regions.
- f_3^{CYS}: Average grey level of the low gray level regions.
- f_4^{CYS}: Average grey level for the walls.
- f_5^{CYS}: The fraction of the PBR area occupied by the low gray level regions.

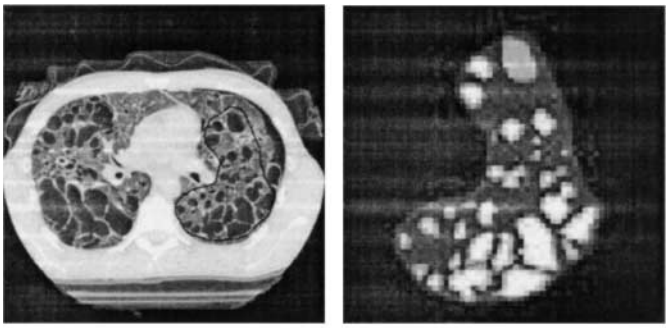

FIG. 16. Shown in (a) is a physician-delineated PBR for the case of cysts. The output is shown in (b). The values of the characterizing parameters for this example are $f_1^{CYS} = 28$, $f_2^{CYS} = 207.64$, $f_3^{CYS} = 55.65$, $f_4^{CYS} = 139.0$, and $f_5^{CYS} = 0.31$.

3. *Mosaic Perfusion*

The visual form here is very similar to that for the case of ground glass, except that the relatively elevated pixels in this case do not occupy a whole lung, as shown by the physician-delineated PBR in Fig. 17. Such patterns are extracted by the same histogram based technique that was described for the case of ground glass and characterized by the same set of parameters listed there.

4. *Honeycombing*

As the name implies, the visual form for the honeycombing pattern consists of small cells, corresponding to air-filled regions in the lung, separated by shared walls, as shown by the example in Fig. 18. The shared walls differentiate this pattern from lung cysts.

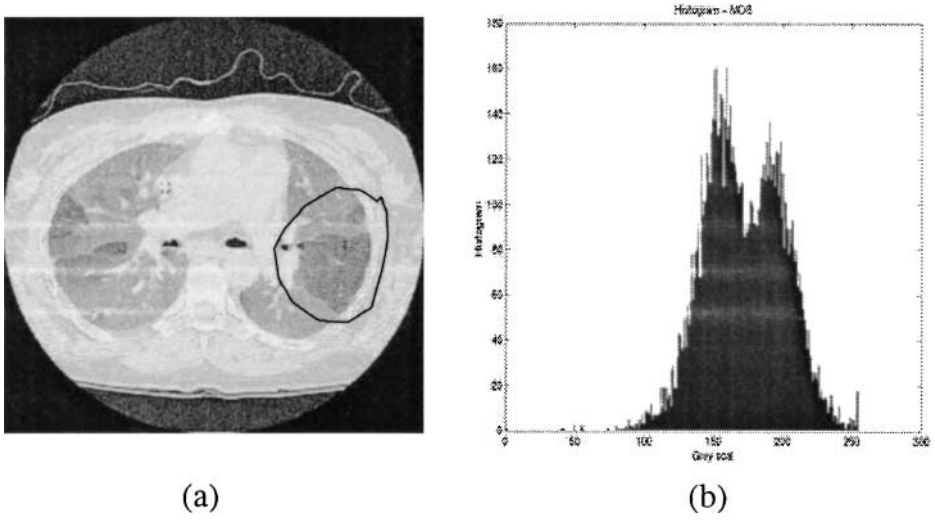

FIG. 17. Shown in (a) is a physician-delineated PBR for the case of mosaic perfusion. Shown in (b) is the histogram for the pixels in the PBR. The values of the characterizing parameters for this example are $f_1^{MOS} = 0.31$, $f_2^{MOS} = 42$, $f_3^{MOS} = 195.30$, and $f_4^{MOS} = 153.83$.

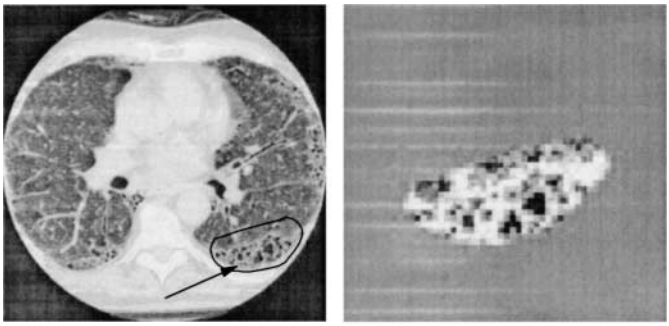

FIG. 18. Shown in (a) is a physician-delineated PBR for the case of honeycombing. The output is shown in (b). The values of the characterizing parameters for this example are $f_1^{HON} = 19$, $f_2^{HON} = 197.53$, $f_3^{HON} = 87.33$, $f_4^{HON} = 171.73$.

Honeycombing is detected in the same manner as interlobular septal thickening. The difference between the two is the grey levels of the lobules; they are darker for honeycombing. Figure 18b shows the extracted honeycombing structure from the PBR delineated in Fig. 18a. The parameters used for characterizing this perceptual category are the same as for interlobular septal thickening.

2.5. Optimal Threshold Determination

The discussion so far has identified a set of gray-level thresholds that are used to extract the features corresponding to the relevant perceptual categories. How to set these thresholds is obviously an important issue in the design of a CBIR system. Each threshold is chosen by applying Otsu's algorithm [26] to the relevant histograms. This algorithm is based on the assumption that a histogram is a mixture of two Gaussian classes and that the optimum threshold that separates them is the ratio of between-class variance and the sum of within-class variances. This approach allows each threshold to adapt to each image separately.

3. ARE THE LOW-LEVEL FEATURES MEASURING THE PHYSICIANS' PERCEPTUAL CATEGORIES?

We have used multivariate analysis of variance (MANOVA) [29] to determine whether or not the low-level features we use for determining the presence or the absence of the perceptual categories are doing their job. MANOVA is used to compute the means of the low-level features separately for the different perceptual categories; the between-category differences of these means; and a measure of the power of the low-level features to discriminate between the different perceptual categories. To assess the normality assumption for the multivariate feature vectors, we applied the chi-squared plot test which is discussed in detail in Appendix A.

3.1. Data Collection and Sample Grouping

To collect the data, we asked an expert physician participating in our research program to mark our HRCT database images with regard to the presence or the absence of the

TABLE 1
Distributions of Lung Diseases and Perceptual Categories in Our Database

	Lung Diseases													
PC	ASP	BOOP	BRO	CLE	DIP	EG	IPF	MET	PAN	PAR	PCP	POL	SAR	Total
SEP	4	64	0	12	13	0	336	0	0	0	0	1	30	478
PAR*	0	1	0	4	0	0	8	0	1	0	0	0	0	14
BRO	1	0	232	6	0	0	1	0	0	0	0	0	0	234
TIB	1	0	32	0	0	0	2	0	0	0	0	13	31	79
BWT	3	1	121	17	0	0	5	0	0	0	0	0	73	220
MUP	0	0	23	0	0	0	0	0	0	0	0	0	0	23
SNO	5	54	6	7	3	0	44	0	0	0	16	22	62	231
CAV*	2	0	0	0	0	0	0	0	0	0	0	0	0	2
CON	33	26	0	1	0	0	5	0	0	0	0	0	14	78
GG	26	71	27	1	37	0	320	0	0	0	59	29	105	728
CAL*	0	0	0	0	0	0	0	8	0	0	0	0	0	8
EMP	0	0	0	654	4	0	10	0	60	63	0	0	0	788
MOS	0	0	20	0	0	0	0	0	0	0	0	0	0	25
CYS	0	0	0	1	0	56	0	0	0	0	0	0	0	57
HON	0	1	0	0	22	0	177	0	0	0	0	0	1	199
Total	34	71	233	656	37	56	383	8	60	63	60	29	143	1873

* Perceptual category with small sample size. (Unit: number of marked pathology-bearing regions.)

various perceptual categories. A special graphical interface tool was devised for this purpose. Using this tool, for each database image the physician could check as many of perceptual categories as applicable to the PBR's in that image. Table 1 shows the distribution of the PBR's having both a particular perceptual category and a particular disease. The top row lists the lung diseases and the left column lists the perceptual categories. Tables 6 and 7 of Appendix C show full definitions of the abbreviations used in Table 1. It is noteworthy that while different diseases give rise to different perceptual categories, the same perceptual category can be seen in the PBR's for different diseases. For example, the perceptual category SEP (interlobular septal thickening) is exhibited by the diseases CLE (centrilobular emphysema), BOOP (bronchiolitis obliterans organizing pneumonia), DIP (desquamitive interstitial pneumonitis), IPF (idiopathic pulmonary fibrosis), and SAR (scleroderma).

Excluding those perceptual categories which have small population sizes (marked with a star in Table 1), the PBRs labeled by a physician are grouped into twelve perceptual categories, corresponding to the terminal leaves of the tree shown in Fig. 2. We will use the following symbols to refer to these 12 categories: (G_{SEP}), (G_{BRO}), (G_{TIB}), (G_{BWT}), (G_{MUP}), (G_{SNO}), (G_{CON}), (G_{GG}), (G_{EMP}), (G_{MOS}), (G_{CYS}), and (G_{HON}). To keep the MANOVA part of the discussion general, we will use N_G to denote the number of perceptual categories.

For the purpose of applying the tools of MANOVA, each observation consists of a vector of p low-level feature measurements from a PBR. Note that the p low-level features for category A will, in general, be different from the p low-level features for category B. Additionally, the value of p for category A is allowed to be different from the value of p for category B. This point is important because the categories do not reside in the same p-dimensional feature space. The presence or absence of a particular perceptual category is decided in its own p-dimensional feature space.

3.2. Sample Grouping and Hypothesis Testing

Although MANOVA could be used to analyze the data for all the categories simultaneously (in order to determine whether or not sufficient discrimination is provided by the features), we chose to perform *pairwise* hypothesis tests to assess whether the low-level features are able to discriminate each perceptual category against each of the remaining categories. The reason for this stems from the fact that multiple perceptual categories characterize each disease and one needs to find all the categories present in a PBR before determining its disease. Furthermore, fewer features are needed to discriminate among two classes than many classes [30]. In what follows we describe the methodology followed for the statistical testing in more detail.

Suppose that we want to ascertain whether or not the features designated to characterize category g are capable of discriminating between perceptual categories g and r. Let N_g and N_r be the number of observations, or sample vectors, for categories g and r respectively and let p_g be the number of the low-level features for category g. For this two-class problem, we can then test the hypothesis that the p_g features, all considered equally important at this stage of analysis, are able to differentiate between categories g and r. This hypothesis test would, of course, need to be carried out separately for each category. For the remaining discussion here, we will use $\mathbf{X}_{g,k}$ to denote the kth observation in category g, and $\mathbf{X}_{r,k}$ for the kth observation in category r.[5]

The mean sample vector for category g is denoted $\overline{\mathbf{X}_g}$. We will use $\bar{\mathbf{X}}$ to denote the mean of the samples of both categories g and r. Both the means $\overline{\mathbf{X}_g}$ and $\bar{\mathbf{X}}$ are defined in the p-dimensional space corresponding to the perceptual category g.

In the p_g-dimensional space used for category g, it is possible to express an observation vector $\mathbf{X}_{g,k}$ by

$$\mathbf{X}_{g,k} = \bar{\mathbf{X}} + (\overline{\mathbf{X}_g} - \bar{\mathbf{X}}) + (\mathbf{X}_{g,k} - \overline{\mathbf{X}_g}). \tag{3}$$

This decomposition highlights the contribution made by the deviation of the observation vector from its own category mean and the difference between a category mean and the entire population mean. The latter will be denoted by $\boldsymbol{\tau}_g = (\overline{\mathbf{X}_g} - \bar{\mathbf{X}})$. In the same p_g-dimensional space, the expression for the overall covariance of the data can now be expressed as

$$\begin{aligned}
\mathbf{T} &= \sum_{i\in\{g,r\}} \sum_{k=1}^{N_i} (\mathbf{X}_{i,k} - \bar{\mathbf{X}})(\mathbf{X}_{i,k} - \bar{\mathbf{X}})^T \\
&= \sum_{i\in\{g,r\}} N_i (\overline{\mathbf{X}_i} - \bar{\mathbf{X}})(\overline{\mathbf{X}_i} - \bar{\mathbf{X}})^T + \sum_{i\in\{g,r\}} \sum_{k=1}^{N_i} (\mathbf{X}_{i,k} - \overline{\mathbf{X}_i})(\mathbf{X}_{i,k} - \overline{\mathbf{X}_i})^T \\
\mathbf{T} &= \mathbf{B} + \mathbf{W}.
\end{aligned}$$

This shows that the overall data variance $\mathbf{T}$ consists of two parts: $\mathbf{B}$, the between category variance, which has $d_B = 1$ degree of freedom for the two-class problem we are analyzing here; and $\mathbf{W}$, the within-category residual variance with $d_W = \sum_{i\in\{g,r\}} N_i - 2$ degrees of freedom.

To determine whether or not there exists category discrimination information in the low-level features used to measure the presence or absence of a category in a PBR, we can

[5] We use bold symbols for vectors and matrices.

perform the following likelihood ratio test. We construct a hypothesis $H_0 : \tau_g = \tau_r$, meaning that the mean for category g is the same as the mean for all other categories lumped together within a chosen confidence interval in the p-dimensional space specific to category g. τ_r denotes $\overline{\mathbf{X}_r} - \bar{\mathbf{X}}$. To test the H_0 hypothesis, we first compute *Wilks's lambda* Λ^*,

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}, \tag{4}$$

where $|\cdot|$ is the determinant of the argument matrix. The exact distribution of Λ^* can be obtained from any standard published table if the size of the category vector is known. A criterion derived from the applicable distribution can then be compared against a threshold for either accepting or rejecting the hypothesis H_0 at a chosen confidence level. For example, when each observation vector consists of two low-level features, meaning $p = 2$, the following F-test criterion obtained from the applicable distribution,

$$F = \left(\frac{d_W - 1}{d_B}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right), \tag{5}$$

can be compared to a threshold,

$$F > F_{d_B, d_W}(\alpha), \tag{6}$$

in order to reject hypothesis H_0 at confidence level $(1 - \alpha)$. $F_{d_B, d_W}(\alpha)$ is the upper $100\alpha\%$ of the F-distribution with d_B and d_W degrees of freedom.

In this manner, we can determine whether or not a given p-dimensional feature set can discriminate a category vector from the rest of the data. This pairwise hypothesis test is carried out separately for each category.

A problem with our methodology is that discriminating between two perceptual categories is not a symmetric procedure, since different features will be considered when discriminating category A from B and B from A. A way to resolve this issue would be to consider the low level features for *all the categories simultaneously* and from these to select the ones that give us the best discrimination between A and B (using a feature selection algorithm such as sequential forward selection).

3.3. Weighting the Mean Differences

If the inequality of Eq. (6) holds for the aforementioned pairwise hypothesis testing for each of the categories, we can conclude that, within the confidence interval used, the chosen low-level features are able to detect the respective perceptual categories. But the following questions remain: What is the relative contribution of each of the low-level features to the differences in the means of the different categories? Could knowledge of these relative contributions be used to weight the image features differently? This section addresses these two questions.

To assess the relative weights to be assigned to the individual low-level features, we used the Bonferroni method of multiple comparisons [29]. For the sake of explanation, assume that we have two perceptual categories: g and r. Moreover, both the categories g and r are characterized by the features designated for category g. We will use p_g to represent the number of image features designated for the category g. Assume that this feature set accepts the hypothesis H_0 at confidence level $1 - \alpha$.

To ascertain the relative importance to be assigned to each g feature, we compute the differences in the means of the feature values for the pair (g, r). For each such pair, we also calculate the uncertainty associated with the mean difference. It goes without saying that the larger the uncertainty in relation to the mean difference, the poorer the feature. These mean differences will then be utilized to set a weight vector for the feature.

For pairwise comparisons, the Bonferroni approach can be used to construct uncertainty intervals for the individual feature components of the difference vector $\overline{\mathbf{X}_g} - \overline{\mathbf{X}_r}$. Let $N_t = N_g + N_r$ be the total number of sample vectors available. Let X_j^i denote the i^{th} element of the feature vector for category j. Under the condition that the confidence level is at least $(1-\alpha)$, we can obtain the following interval for the uncertainty in the difference of the mean values of the ith feature,

$$(L_i, R_i) = \overline{X_g^i} - \overline{X_r^i} \pm t_{N_t-2}(\alpha') \sqrt{\frac{w_{i,i}}{N_t - 2}\left(\frac{1}{N_g} + \frac{1}{N_r}\right)}, \tag{7}$$

where $\alpha' = \frac{\alpha}{2p_g}$ and $w_{i,i}$ is the ith diagonal element of $\mathbf{W}$ (defined in Section 3.2) and $t_{N_t-2}(\alpha')$ is the student t-distribution with $N_t - 2$ degrees of freedom. The size of this uncertainty interval is given by $R_i - L_i$. Evidently, when the second term in Eq. (7) is zero, there is no uncertainty in the difference of the mean values for feature i since L_i is equal to R_i. By the same token, when the second term in Eq. (7) is greater than the first, the uncertainty dominates, making such a feature unreliable. The weight given to such a feature is one. We only assign more weight to a feature if the second term of Eq. (7) is less than the first term for that feature.

The quality of the ith feature for discriminating between the categories g and r can now be measured by the following h factor:

$$h_{g,r}^i = \begin{cases} \left| \dfrac{\overline{X_g^i} - \overline{X_r^i}}{\frac{R_i - L_i}{2}} \right| & \text{if } 0 < \dfrac{R_i - L_i}{2} < \left|\overline{X_g^i} - \overline{X_r^i}\right| \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

These quality factors can be computed for every feature for the pair of g and r. Subsequently, the quality factors are grouped to form a quality vector of mean differences for the features that can discriminate g from r:

$$\mathbf{h}_{g,r} = \left[h_{g,r}^1, h_{g,r}^2, \ldots, h_{g,r}^{p_g}\right]^t. \tag{9}$$

This quality vector is then used to weight the difference of the mean vectors from two populations, g and r. The weighted mean difference vector can be obtained,

$$\hat{\mathbf{d}}_{g,r} = \mathbf{h}_{g,r}^t \, \mathbf{I}\left(\overline{\mathbf{X}_g} - \overline{\mathbf{X}_r}\right), \tag{10}$$

where $\mathbf{I}$ is an identity matrix with dimension $p_g \times p_g$. In general, for each perceptual category j, a set of $N_G - 1$ weighted mean difference vectors $\hat{\mathbf{d}}_{j,k}$ are computed ($1 \le j, k \le N_G$, $j \ne k$).

4. QUERY AND MATCHING

With regard to its use by a physician, we want our system to retrieve the most visually similar database images that have the same disease label that an expert physician would assign to the query image. For reasons that will be explained in the next section, the retrieval requires a combination of a classifier and a voting scheme.

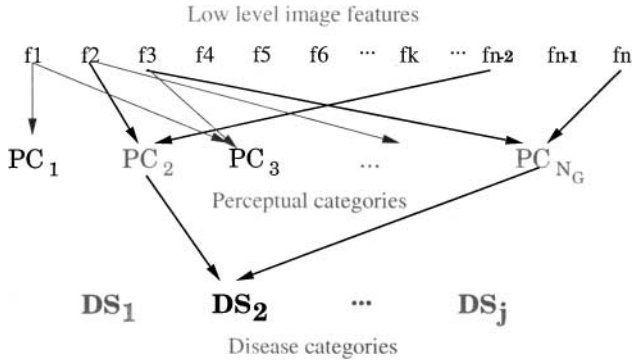

FIG. 19. To ascertain the disease label of a new PBR, the low-level features are first mapped to perceptual categories. It is possible for a PBR to be associated with multiple perceptual categories. The perceptual categories for a given PBR are then mapped to the disease label.

An issue of singular importance in the design of our retrieval engine is that a PBR can fall into multiple perceptual categories at the same time. In Appendix B Table 4 shows the co-occurrence frequencies of the perceptual categories. Each entry in the table, denoted $s_{j,k}$ for the entry in jth row and kth column, shows the number of PBRs that exhibited the row and column perceptual category labels at the same time. For example, $s_{2,4} = 122$. This means that 122 PBRs simultaneously exhibited the perceptual categories BRO and BWT. For obvious reasons, $s_{j,k} = s_{k,j}$. Since each PBR has only one disease diagnosis, the cooccurrence frequency $s_{j,k}$ also tells us how often the two perceptual categories result in one disease diagnosis.

The fact that a PBR can possess multiple perceptual categories dictates the approach shown in Fig. 19 for ascertaining the disease label associated with a PBR. From all the feature measurements on a PBR, we must first determine all the perceptual category labels that apply to the PBR. We must then determine how the perceptual category labels map into a disease label.

In the rest of this section, we will first discuss how we deal with the problem of determining all the perceptual category labels that apply to a PBR.

4.1. A Recognizer for Determining the Perceptual Categories of a Query PBR

Since a PBR can possess multiple perceptual categories simultaneously, one cannot directly apply the traditional notion of classification to a PBR with respect to perceptual category labels. Instead, we will use the notion of a *recognizer*. While a classifier attempts to assign a single label to an unknown object, a *recognizer* seeks to come up with all possible labels that could be used to characterize the same object. Of course, within the framework of a *recognizer*, the assignment of each perceptual category label can still be based on a simple classifier that tries to distinguish between the applicability of a given perceptual category label.

The schema used for the *recognizer* is as shown in Fig. 20. Let $\mathcal{PC} = \{PC_1, PC_2, \ldots, PC_{N_G}\}$ be the set of all perceptual category labels. To determine all the labels that apply to a given PBR, we test for the applicability of each label separately. Given a PBR P_q of unknown perceptual category, a feature vector $\mathbf{X}_{P_q,j}$ that contains image features relevant to a specific perceptual category label j is extracted. Let $\mathbf{X}_{P_q} = \{\mathbf{X}_{P_q,1}, \mathbf{X}_{P_q,2}, \ldots, \mathbf{X}_{P_q,N_G}\}$ be the collection of feature vectors of P_q for all the perceptual category labels. For the category label j, a set of classifiers, $\{L_{j,k}(\cdot) | 1 \le k \le N_G,\ j \ne k\}$, is built off-line to test the

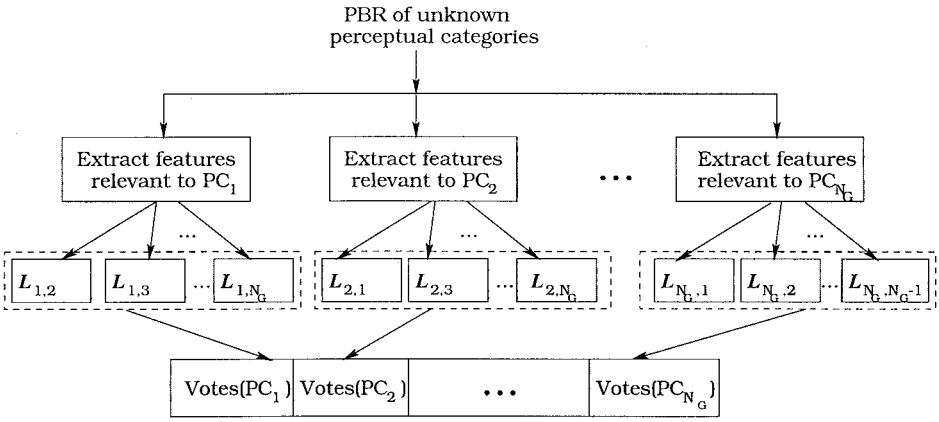

FIG. 20. A flow chart to recognize all possible perceptual categories.

applicability of label j. Each classifier $L_{j,k}$ consists of a set of decision thresholds that when applied to the features $\mathbf{X}_{P_q,j}$ tell us whether the PBR belongs to the perceptual category j or to the perceptual category k. There are $N_G - 1$ classifiers for each label and $N_G(N_G - 1)$ classifiers for all labels in the system. The decision threshold in each classifier $L_{j,k}$ is based on the minimization of the estimated cost of misclassification (ECM) [30]. The classifier design is discussed in greater detail in Appendix B. Each perceptual classifier $L_{j,k}$ casts one vote for the category j if the decision threshold corresponding to that classifier is satisfied.

For a query PBR P_q, the *recognizer* is then able to provide the system the votes $\mathcal{V}_{P_q}$ for all possible perceptual category labels. Let $\mathcal{V}_{P_q}(j)$ be the vote for the jth perceptual category. A pseudo-code description of this schema follows:

Pseudo code for the assignment of each category label

```
01  for (j = 0; j < N_G; j ++) {
02    for (k = 0; k < N_G; k ++) {
03      if (j ≠ k) {
04        // extracting features for category j
05        X_{P_q,j} ← feature Extraction(P_{q,j})
06        if (l_{j,k}(X_{P_q,j}) ≥ 0)
07          V_{P_q}(j)++;
08      }
09    }
10  }
```

where $l_{j,k}$ is the likelihood ratio test for the classifier $L_{j,k}$.

What comes out of the recognizer is a *vote vector* for a given PBR. Ideally, we would like to reject votes that are too few in number and accept the rest as legitimate perceptual categories for the PBR. But for lack of suitable logical considerations that could be used for vote rejection, we have used the approach discussed below.

4.2. Retrieval Based on Vote Vectors

The *recognizer* of the preceding section yields an N_G-dimensional vote vector, $\mathcal{V}_{P_q}$, for a query PBR. This also applies to each new image as it is entered into the database.

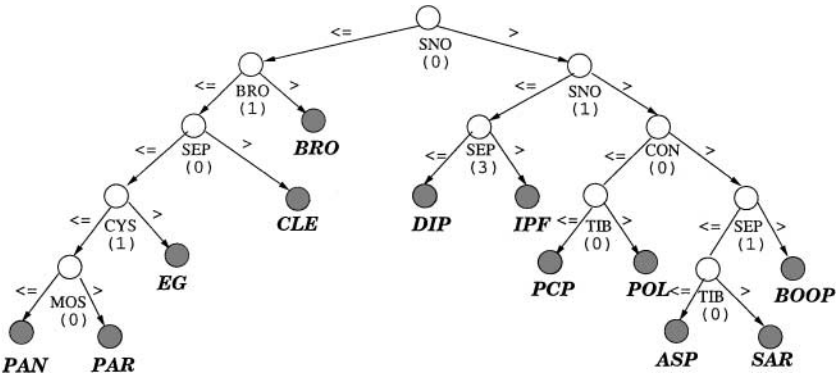

FIG. 21. A decision tree created to index PBRs based on the vote vector. The label under each nonleaf node means that the vote for a certain perceptual category was subject to a decision threshold.

Therefore, each PBR in each database image has its own vote vector $\mathcal{V}$. Each element of $\mathcal{V}$ has value from zero to N_G. Retrieval on the basis of disease categories can then be carried out on the basis of similarity of vote vectors. How this is done will be described next.

We create supervised training data by grouping those PBR vote vectors together that yield identical diagnoses (disease labels). A decision tree created by C4.5 [31] is grown to index these PBR vote vectors. The decision tree is shown in Fig. 21.

To retrieve database PBRs, the vote vector $\mathcal{V}_{P_q}$ of a query PBR is used to locate a leaf node of the decision tree. At least one representative disease category will be retrieved from that leaf node and all PBR's associated with that node will be pooled and denoted by P_{pooled}.

For example, if $\mathcal{V}_{P_q} = [SEP = 5,\ BRO = 0,\ TIB = 0,\ \mathrm{BWT} = 0,\ MUP = 0,\ SNO = 1,\ CON = 0,\ GG = 4,\ EMP = 0,\ MOS = 0,\ CYS = 0,\ HON = 2]$, the system would retrieve PBRs with disease label IPF. Let P^k_{pooled} be the kth candidate PBR from P_{pooled}. If the number of retrieved PBRs is less than what a query requested, we can relax the query by including the leaf nodes that share the same parent node with the selected leaf nodes. On the other hand, if there are more candidate database PBRs than what we expect after query relaxation, we rank the PBRs in P_{pooled} based on the following measurement:

$$\mathcal{S}\left(P_q, P^k_{pooled}\right) = \frac{\sum_{j=1}^{N_G} min\left(\mathcal{V}_{P_q}(j), \mathcal{V}_{P^k_{pooled}}(j)\right)}{max\left(\left|\mathcal{V}_{P_q}\right|, \left|\mathcal{V}_{P^k_{pooled}}\right|\right)}. \tag{11}$$

The value of $\mathcal{S}$ ranges from zero to one. A perfect retrieval is expected to have $\mathcal{S} = 1$; a totally irrelevant retrieval $\mathcal{S} = 0$.

5. EXPERIMENTAL RESULTS

Our database, created in the manner described in Section 3, contains 1873 PBRs from 998 HRCT lung images. Table 1 shows the database distribution with respect to the different diseases and with respect to the different perceptual categories. Note that the distribution of our data is skewed and the majority class—the disease emphysema—comprises around 60% of all the PBRs. This is a fact of life about medical databases in general since diseases occur with widely varying frequencies in the population.

TABLE 2
Retrieval Precision Based on Disease Categories

Diagnosis	Number of queries	Percentage		
		Scatter-shot	Perceptual-category	SFS-on-perceptual
ASP	10	58	**72**	61
BOOP	10	62	**82**	70
BRO	20	72	**84**	75
CLE	20	80	**82**	82
DIP	10	62	**75**	68
EG	20	75	**83**	73
IPF	20	73	**78**	76
PAN	20	37	**68**	58
PAR	10	85	**83**	82
PCP	10	52	**81**	57
POL	10	48	**78**	49
SAR	20	67	**71**	67
Total DB	180	65.27	**77.39**	69.38

We have evaluated our approach by measuring the retrieval precision based on disease class. 818 images were used to train the system and the remaining 180 images to evaluate the performance of the system. The experiment consists of the following steps: (1) Select an image from the 180 images (testing database) as a query image; (2) Ask the system to retrieve the four most similar images from the database, taking into account the modified mean differences and linear classifiers discussed in Section 3.3 for the different perceptual categories; and (3) Compare the disease class of the PBRs in the query image with the disease class of the PBRs in the retrieved images from Eq. (11).

The precision of the retrieval results is shown in Table 2. On the average, using perceptual categories for retrieval in the manner described here resulted in improving the precision rates from 65.27% to 77.39% over the traditional methods. The first column in this table corresponds to using the "traditional" approach described in [21] in which we start with an exhaustive list of low-level image features that are subsequently pruned by employing the sequential forward selection (SFS) method [23]. Table 2 also shows the retrieval results obtained when sequential forward selection is applied to only those low-level features that we use for detecting the perceptual categories (see Section 2). Retrieval using these low-level features (but not including the perceptual categories themselves) is accomplished with the k-NN (k nearest neighbors) algorithm. As the table shows, on the average the precision rate with this algorithm is 69.38%, as compared with the 77.39% obtained when perceptual categories are used as described in this paper.

6. CONCLUSION

What specific features to use for content-based retrieval is more a function of the level of ingenuity of researchers than a result of some precise scientific analysis. In the past, we used the scattershot approach, which characterized the pathology bearing regions with an exhaustive set of features and then used a standard dimensionality reduction tool to pull out the feature set that was maximally discriminatory with respect to the disease categories.

In the work reported in this paper, we have eschewed this previous approach. We now extract only those low-level features that measure the presence or the absence of the various perceptual categories that the physicians use for disease diagnosis. MANOVA is then used to test the discriminatory power of these features and the Bonferroni method used to determine how much weight to assign to each feature. According to our experimental results, this new approach to feature extraction and image characterization has yielded retrieval performance that is superior for most of the disease classes.

This paper has focused primarily on the computer vision aspects of designing an image retrieval system using the perceptual categories of the domain experts. But, before closing, we should also mention that such as system is more likely to be accepted by the end users. A physician is more likely to identify with and accept a system such as the one described here because the decision processes involved bear some resemblance to those of the physician. If an expert physician disagreed with the disease labels assigned by our system to a new image, the physician could question the system about the perceptual categories detected in the image and ascertain the appropriateness of those categories. In that sense, the system described here possess superior explanatory powers for a richer interaction with the physician.

APPENDIX A

Assessing the Normality Assumption

Before MANOVA can be applied, the data must satisfy certain assumptions. The most notable of these are: (1) each observation $\mathbf{X}_{g,k}$ is a random sample from perceptual category g; (2) the random samples from different categories are independent; and (3) the distribution corresponding to each category is multivariate normal. We believe that our data does indeed satisfy the first two assumptions.[6] With regard to the third assumption, we have performed normality tests which will be discussed in the following subsection.

1. Quantile Plots for Establishing Normality

We will now briefly present the quantile plot technique we use to establish the normality of the feature space distributions associated with the various perceptual categories. For the multivariate case, these plots are also referred to as the *Chi-squared plots* [29].

To explain the basic idea underlying this approach to testing for normality, let x_i's be univariate samples ($1 \leq i \leq n$) and let z_i's be their standardized versions,

$$z_i = \frac{x_i - \bar{x}}{s}, \tag{12}$$

where s is an estimate of sample variance. Let $x_{(j)}$'s be the ordered versions of x_i's;

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}, \tag{13}$$

where $x_{(1)}$ is the smallest value among the x_i's and $x_{(n)}$ the largest. Their corresponding

[6] The standard formulas of MANOVA also require that the covariance matrices of the populations be the same. For now we have ignored this requirement. Hypothesis testing without this assumption will be addressed by us in the future.

$z_{(j)}$'s also preserve the order:

$$z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(n-1)} \leq z_{(n)}. \tag{14}$$

Assuming the ordered samples are continuous, the standardized version $z_{(j)}$ is the $(\frac{j}{n})$th quantile of the standard normal distribution. We denote $q_{(j)}$ as the jth quantile, which is defined by the relation

$$Pr\big[Z \leq q_{(j)}\big] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} = p_{(j)} = \frac{j-1}{n}. \tag{15}$$

For the purpose of implementation convenience, as suggested by Looney *et al.* [32], we use $\frac{j-\frac{3}{8}}{n+\frac{1}{4}}$ instead of $\frac{j-1}{n}$. The expected value of $z_{(j)}$ can be estimated as follows:

$$E\big[z_{(j)}\big] \cong \Phi^{-1}\left(\frac{j-\frac{3}{8}}{n+\frac{1}{4}}\right) = q_{(j)}. \tag{16}$$

What we are interested in knowing is how to utilize the properties of the above standard normal distribution to test the normality of the collected samples x_i's. From Eqs. (12) and (16), the relationship between the collected samples $x_{(j)}$ and their corresponding quantile values $q_{(j)}$ can be expressed in the form

$$x_{(j)} \cong \bar{x} + sq_{(j)} + \epsilon, \tag{17}$$

where, as argued by [29], the deviation ϵ is small, particularly when the sample size n is large. Therefore, as shown in [29], the straightness of a plot of $x_{(j)}$ versus $q_{(j)}$ can be used to establish normality when n is as small as 20. Such plots are known as Q–Q plots, for quantile–quantile, since one plots the quantiles obtained from the actual data versus the quantiles from a true normal distribution. Figure 22a shows a Q–Q plot for $n = 20$ for a data set drawn by a random number generator from a normal distribution. As suggested by Eq. (17), the

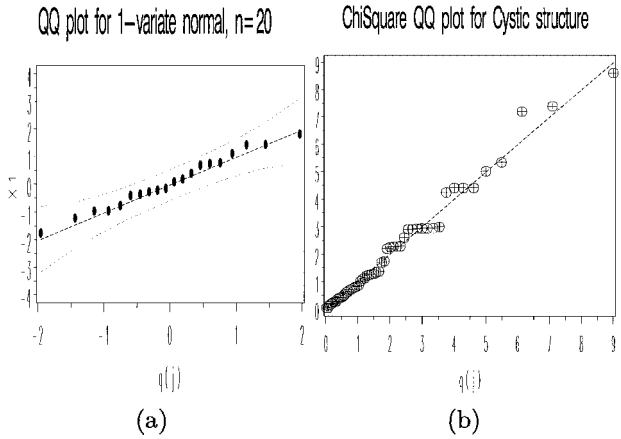

FIG. 22. (a) A QQ plot for a univariate normal distribution. (b) A chi-squared QQ plot for cystic structure features: the number of cystic structures, the average size of cystic structures, the grey mean for the lumens of the cystic structures, the grey mean for the walls, and the coverage of cystic structures in the marked PBR. The measured correlation coefficient is 0.993 with sample size 56.

TABLE 3
Critical Points for the Q–Q Plot Correlation Coefficient Test for Normality

Sample size	0.01	0.05	0.10
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
15	0.9126	0.9389	0.9503
20	0.9269	0.9508	0.9604
25	0.9410	0.9591	0.9665
30	0.9479	0.9652	0.9715
35	0.9538	0.9682	0.9740
40	0.9599	0.9726	0.9771
45	0.9632	0.9749	0.9792
50	0.9671	0.9768	0.9809
55	0.9695	0.9787	0.9822
60	0.9720	0.9801	0.9836
75	0.9771	0.9838	0.9866
100	0.9822	0.9873	0.9895
150	0.9879	0.9913	0.9828
200	0.9905	0.9931	0.9942
300	0.9935	0.9953	0.9960

slope of the plot is the estimated sample variance s and the offset (or intercept) of the vertical axis is the estimated sample mean. In addition to visually examining the "straightness" of the paired $(x_{(j)}, q_{(j)})$'s values in the plot, we can use the following correlation coefficient to provide a quantitative measure of the test:

$$r_Q = \frac{\sum_{j=1}^{n}\left(x_{(j)} - \bar{x}\right)\left(q_{(j)} - \bar{q}\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{(j)} - \bar{x}\right)^2}\sqrt{\sum_{j=1}^{n}\left(q_{(j)} - \bar{q}\right)^2}}. \tag{18}$$

Table 3 lists what statisticians refer to as critical points for accepting or rejecting the hypothesis that a sample set was drawn from a normal distribution on the basis of the value of r_Q within a chosen confidence interval and for a given sample size. So, if $n = 100$ and we want to test for normality at a confidence level of 99%, we would want r_Q to be at least 0.9822.

For the multivariate case, let $\mathbf{X}_j$ be the jth sample vector, $j = 1, \ldots, n$, of dimension p and let $\bar{\mathbf{X}}$ and $\mathbf{S}$ be the estimated mean vector and the estimated convariance matrix, respectively. To test for multivariant normality, first we compute the squared Mahalanobis distances:

$$d_j^2 = (\mathbf{X}_j - \bar{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}). \tag{19}$$

When the sample size n and $n - p$ are greater than 30 [29], each d_j should behave like a chi-squared random variable. Similar to what was done for the univariate case, we order the squared Mahalanobis distances for all samples:

$$d_{(1)}^2 \le d_{(2)}^2 \le \cdots \le d_{(n)}^2. \tag{20}$$

The quantile corresponding to the sorted $\mathbf{X}_j$ is then obtained as

$$q_{c,p}\left(\frac{j-\frac{3}{8}}{n+\frac{1}{4}}\right) = \chi^2_p\left(\frac{n-j+\frac{3}{8}}{n+\frac{1}{4}}\right), \tag{21}$$

where $q_{c,p}(\frac{j-\frac{3}{8}}{n+\frac{1}{4}})$ is the $100_{q_{c,p}}(\frac{j-\frac{3}{8}}{n+\frac{1}{4}})$ quantile of the chi-squared distribution with p degrees of freedom. The paired $(q_{c,p}(\frac{j-\frac{3}{8}}{n+\frac{1}{4}}),\ d^2_{(j)})$'s should remain along a straight line with slope equals to one and passing through the origin. We then use the correlation coefficients to test the normality of the observed multivariate samples. Figure 22b shows a Q–Q plot for testing the normality of three image attributes used to detect the presence of the cystic structure.

APPENDIX B

Design of the Recognizer and Misclassification Costs

As mentioned earlier, within the framework of a *recognizer*, the assignment of a perceptual category label to a PBR is based on a set of linear classifiers. Taking into consideration the category population distributions, these classifiers seek to minimize the expected cost of misclassification (ECM).

If Perceptual Category A does not share any PBRs with Perceptual Category B, we assign the highest cost to the misclassification from A to B and vice versa. Not sharing any PBRs is tantamount to not sharing any disease diagnoses. On the other hand, when two perceptual categories share common PBRs, we would like to find the costs of misclassification from A to B, as well as from B to A. The reason why these costs may be different can be explained with the following example. In Table 4, *MUP* shares 23 out of 25 PBRs with *BRO*. This means *MUP* shares the same diagnoses as *BRO* in 92% of the cases in our training sample set. Therefore, we assign a "small" penalty to misclassifying a query PBR that has the perceptual category *MUP* as the perceptual category *BRO*. However, *BRO* shares only 23 out of 234 PBRs with *MUP*, meaning that only 10% of the PBRs with the perceptual category *BRO* share the same diagnosis as the perceptual category *MUP*. We would, therefore, want to assign a "bigger" misclassification penalty in this case.

TABLE 4
A Matrix of Cooccurrence Frequencies among Perceptual Categories

Perceptual	SEP	BRO	TIB	BWT	MUP	SNO	CON	GG	EMP	MOS	CYS	HON
SEP	478	1	3	31	0	114	41	413	22	0	0	177
BRO	1	234	33	122	23	8	0	26	6	20	0	1
TIB	3	33	79	24	2	21	3	32	0	2	0	0
BWT	31	122	24	220	10	37	12	84	24	19	0	4
MUP	0	23	2	10	25	2	0	2	0	0	0	0
SNO	114	8	21	37	2	231	36	193	8	0	3	14
CON	41	0	3	12	0	36	78	72	0	0	0	4
GG	413	26	32	84	2	193	72	728	13	0	0	158
EMP	22	6	0	24	0	8	0	13	788	0	0	0
MOS	0	20	2	19	0	0	0	0	0	23	0	0
CYS	0	0	0	0	0	3	0	0	0	0	57	1
HON	177	1	0	4	0	14	4	158	0	0	1	199

TABLE 5
Costs of Misclassification between Perceptual Categories

Perceptual	SEP	BRO	TIB	BWT	MUP	SNO	CON	GG	EMP	MOS	CYS	HON
SEP	0	0.99	0.96	0.86	1.00	0.51	0.47	0.43	0.97	1.00	1.00	0.11
BRO	0.99	0	0.58	0.45	0.08	0.97	1.00	0.96	0.99	0.13	1.00	0.99
TIB	0.99	0.86	0	0.89	0.92	0.91	0.96	0.95	1.00	0.91	1.00	1.00
BWT	0.94	0.48	0.70	0	0.60	0.84	0.85	0.88	0.96	0.17	1.00	0.98
MUP	1.00	0.90	0.97	0.95	0	0.99	1.00	0.99	1.00	1.00	1.00	1.00
SNO	0.76	0.96	0.73	0.83	0.92	0	0.54	0.73	0.99	1.00	0.95	0.93
CON	0.91	1.00	0.96	0.95	1.00	0.84	0	0.90	1.00	1.00	1.00	0.98
GG	0.14	0.89	0.59	0.62	0.92	0.16	0.08	0	0.98	1.00	1.00	0.21
EMP	0.95	0.97	1.00	0.89	1.00	0.97	1.00	0.98	0	1.00	1.00	1.00
MOS	1.00	0.91	0.97	0.91	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00
CYS	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0	0.99
HON	0.63	0.99	1.00	0.98	1.00	0.94	0.95	0.78	1.00	1.00	0.98	0

Let $C(pc_k|pc_j)$ be the misclassification cost when an observation comes from perceptual category j, but is misclassified as perceptual category k. We use the following formula for this cost,

$$C(pc_k|pc_j) = \frac{s_{j,j} - s_{k,j}}{s_{j,j}}, \tag{22}$$

where $s_{j,j}$ is the total number of PBRs with perceptual category j and $s_{j,k}$ is the number of PBRs common to the perceptual categories j and k. For example, from Table 4 we have $s_{SEP,SEP} = 478$ and $s_{SEP,SNO} = 114$. Table 5 lists the misclassification costs obtained in this manner.

Let p_j and p_k be the prior probabilities of the perceptual categories j and k. The prior p_j is given by $s_{j,j}/s_T$, where s_T is the total number of PBRs in a database. In our case, $s_T = 1873$. Taking into account these priors, the misclassification costs $C(pc_j|pc_k)$ and $C(pc_k|pc_j)$, and the weighted mean difference $\hat{\mathbf{d}}_{\mathbf{j,k}}$ as given by Eq. (10), the likelihood ratio test for pairwise classification between any two perceptual categories j and k is given by

$$l_{j,k}(\mathbf{X}) = \hat{\mathbf{d}}_{j,k}\mathbf{S}^{-1}_{pooled}\mathbf{X} - \frac{1}{2}\hat{\mathbf{d}}_{j,k}\mathbf{S}^{-1}_{pooled}(\overline{\mathbf{X_j}} + \overline{\mathbf{X_k}}) - \ln\left[\left(\frac{C(pc_j|pc_k)}{C(pc_k|pc_j)}\right)\left(\frac{p_k}{p_j}\right)\right], \tag{23}$$

where the overall covariance matrix $\mathbf{S}_{pooled}$ is defined as

$$\mathbf{S}_{pooled} = \left[\frac{N_{pc_j} - 1}{(N_{pc_j} - 1) + (N_{pc_k} - 1)}\right]\mathbf{S}_j + \left[\frac{N_{pc_k} - 1}{(N_{pc_j} - 1) + (N_{pc_k} - 1)}\right]\mathbf{S}_k. \tag{24}$$

In the above equation, $\mathbf{S}_j$ and $\mathbf{S}_k$ are the covariance matrices associated with the populations for the perceptual categories j and k, respectively. Let $\hat{\mathbf{d}}_{j,k}\mathbf{S}^{-1}_{pooled} = \mathbf{a}'$ and $\mathbf{y} = \mathbf{a}'\mathbf{X}$. The likelihood ratio test in Eq. (23) can then be reexpressed as

$$l_{j,k}(\mathbf{X}) = \mathbf{a}'\mathbf{x} - \frac{1}{2}(\mathbf{y}_1 + \mathbf{y}_2) - \ln\left[\left(\frac{C(pc_j|pc_k)}{C(pc_k|pc_j)}\right)\left(\frac{p_k}{p_j}\right)\right]. \tag{25}$$

APPENDIX C

Abbreviations for Medical Terms Used in This Paper

TABLE 6
List of Abbreviations Used for Disease Categories

Abbreviations	Meaning
ASP	Aspergillus
BOOP	Bronchiolitis obliterans organizing pneumonia
BRO	Bronchiectasis
CLE	Centrilobular emphysema
DIP	Desquamitive interstitial pneumonitis
EG	Eosinophilic granuloma
IPF	Idiopathic pulmonary fibrosis
MET	Metastatic calcification
PAN	Panacinar
PAR	Paraseptal emphysema
PCP	Pneumocystis carinii pneumonia
POL	Polymyositis
SAR	Scleroderma

TABLE 7
List of Abbreviations Used for Perceptual Categories

Abbreviations	Meaning
SEP	Interlobular septal thickening
PAR	Parenchymal bands
BRO	Bronchiectasis
TIB	Tree-in-bud
BWT	Bronchial wall thickening
MUP	Mucus plugging
SNO	Small nodules
CAV	Cavity
CON	Conglomerate nodules
CAL	Calcification
EMP	Emphysema
MOS	Mosaic perfusion
CYS	Cysts
HON	Honeycombing

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant IRI9711535 and the National Institute of Health under Grant 1 R01 LM06543-01A1. We also knowledge fruitful discussions with Professor George McCabe and Leming Qu of the Department of Statistics at Purdue University.

REFERENCES

1. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang bs B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, Query by image and video content: The QBIC system, *IEEE Comput.*, 1995, 23–32.

2. P. M. Kelly, T. M. Cannon, and D. R. Hush, Query by image example: The CANDID approach, in *Storage and Retrieval for Image and Video Databases III*, SPIE, Vol. 2420, pp. 238–248, 1995.
3. V. Ogle and M. Stonebraker, Chatbot: Retrieval from a relational database of images, *IEEE Comput.* **28**(9), 1995, 40–48.
4. M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. Huang, Supporting similarity queries in MARS, in *5th Int. Multimedia Conference*, *ACM, 1997*.
5. A. Pentland, R. W. Picard, and S. Sclaroff, Photobook: Tools for content-based manipulation of image databases, in *Storage and Retrieval for Image and Video Databases*, pp. 34–47, SPIE, 1994.
6. S. Sclaroff, L. Taycher, and M La Cascia, Imagerover: A content-based image browser for the World Wide Web, in *Workshop on Content-based Access of Image and Video Libraries*, *IEEE, 1997*.
7. H. Stone and C. Li, Image matching by means of intensity and texture matching in the fourier domain, in *Workshop of Content-Based Access of Image and Video Databases*, *IEEE, 1996*.
8. A. Turtur *et al.*, IDB: An image database system. *IBM J. Res. Dev.* **35**(1/2), 1991, 88–96.
9. S. Berretti, A. Del Bimbo, and E. Vicario, Efficient matching and indexing of graph models in content-based image retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(10), 2001, 1089–1105.
10. C. Carson, M. Thomas, S. Belongie, J. Hellerstein, and J. Malik, Blobworld: A system for region-based image indexing and retrieval, in *Proc. International Conf. on Visual Information Systems, 1999*.
11. W. Ma and B. Manjunath, Netra: A toolbox for navigating large image databases, *Multimedia Systems* **7**, 1999, 184–198.
12. T. Minka and R. Picard, Interactive learning using a society of models, *Pattern Recognition*, 1997, 565–581.
13. J. Smith and C. Li, Image classification and querying using composite region templates, *Comput. Vision Image Understanding* **75**(1/2), 1999, 165–174.
14. J. Wang, J. Li, and G. Wiederhold, SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9), 2001, 1–17.
15. W. W. Chu, C. C. Hsu, A. F. Cardenas, and R. K. Taira, A knowledge-based image retrieval with spatial and temporal constructs, *IEEE Trans. Knowledge Data Eng.* **10**(6), 1998, 872–888.
16. C. C. Hsu, W. W. Chu, and R. K. Taira, A knowledge-based approach for retrieving images by content, *IEEE Trans. Knowledge Data Eng.* **8**(4), 1996, 522–532.
17. F. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas, Fast and effective retrieval of medical tumor shapes, *IEEE Trans. Knowledge Data Eng.* **10**(6), 1998, 889–904.
18. D. Comaniciu, P. Meer, D. Foran, and A. Medl, Bimodal system for interactive indexing and retrieval of pathology images, in *Workshop on Applications of Computer Vision*, *IEEE, 1998*.
19. W. Cai, D. Feng, and R. Fulton, Content-based retrieval of dynamic pet functional images, *IEEE Trans. Inform. Technol. Biomed.* **4**(2), 2000, 152–158.
20. Y. Liu, F. Dellaert, W. Rothfus, A. Moore, J. Schneider, and T. Kanade, Classification-driven pathological neuroimage retrieval using statistical assymetry measures, in *MICCAI*, 2001.
21. C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. Aisen, and L. Broderick, ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases, *Comput. Vision Image Understanding* **75**(1/2), 1999, 111–132.
22. C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. Aisen, and L. Broderick, Local versus global features for content-based image retrieval, in *Workshop of Content-Based Access of Image and Video Databases*, pp. 30–34, IEEE, 1998.
23. J. Kittler, Feature set search algorithms, in *Pattern Recognition and Signal Processing*, pp. 41–60, 1978.
24. W. R. Webb, N. L. Muller, and D. P. Naidich, *High-Resolution CT of the Lung*, Lippincott–Raven, Philadelphia, 1996.
25. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, San Diego, 1982.
26. N. Otsu. A threshold selection method from gray-level histogram, *IEEE Trans. Systems Man Cybernet.* **9**(1), 1979, 62–66.

27. H. Schwarz and H. Exner, The characterization of the arrangement of feature centroids in plans and volumes, *J. Microsc.*, 1983, 129–155.

28. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, Reading, MA, 1992.

29. R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood cliffs, NJ, 1998.

30. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, 1990.

31. J. R. Quinlan, *Programs for Machine Learning*, Morgan Kaufman, San Mateo, CA, 1993.

32. S. Looney and T. Gulledge, Use of the correlation coefficient with normal probability plots, *Am. Stat.* **39**(1), 1985, 75–79.