ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases

Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, Akio Kosaka

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

Alex M. Aisen

Department of Radiology, Indiana University Medical Center, Indianapolis, Indiana 46202
E-mail: aaisen@iupui.edu

and

Lynn S. Broderick

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

It is now recognized in many domains that content-based image retrieval from a database of images cannot be carried out by using completely automated approaches. One such domain is medical radiology for which the clinically useful information in an image typically consists of gray level variations in highly localized regions of the image. Currently, it is not possible to extract these regions by automatic image segmentation techniques. To address this problem, we have implemented a human-in-the-loop (a physician-in-the-loop, more specifically) approach in which the human delineates the pathology bearing regions (PBR) and a set of anatomical landmarks in the image when the image is entered into the database. To the regions thus marked, our approach applies low-level computer vision and image processing algorithms to extract attributes related to the variations in gray scale, texture, shape, etc. In addition, the system records attributes that capture relational information such as the position of a PBR with respect to certain anatomical landmarks. An overall multidimensional index is assigned to each image based on these attribute values. © **1999 Academic Press**

1. INTRODUCTION

Many content-based image retrieval (CBIR) systems have been developed during the last several years. Almost all of these systems are founded on the premise that images can be characterized by global signatures for the purpose of retrieval from a database [6, 14, 27, 18, 11, 22, 9]. For example, the CANDID system [14] computes histograms from normalized gray levels for image characterization and the QBIC system [6] characterizes images by global characteristics such as color histogram, texture values, and shape parameters of easily segmentable regions.

However, for many domains, global characterization alone cannot ensure satisfactory retrieval results. One such domain is medical radiology, for which the clinically useful information consists of gray level variations in highly localized regions of the image. For example, in a high-resolution computed tomographic (HRCT) image of the lung, a disease such as emphysema (shown in Fig. 1a) manifests itself in the form of low-attenuation regions that are textured differently from the rest of the lung. Local attributes are needed for such situations because the number of pathology bearing pixels in an image is small, relative to the number of pixels in the rest of the image, and any global signature would not be sufficiently impacted to serve as a useful attribute for image retrieval. This bodes ill for many of the previously developed methods for CBIR with regard to their use for medical radiology. Even those CBIR systems that are not explicitly global in their image characterization—such as those presented in [12, 4, 16]—would not work for the types of images of which Fig. 1a is an example. Indeed, the MR images in [12, 4] are characterized by just the shapes of the ventricular regions in the brain. The key to the success of the system of [12, 4] is that a ventricular region has an extremely distinct shape whose deformation is indicative of disease. In the system of [16], images in the database consist of a single tumor placed in the center without any background texture. In our application, the pathology bearing regions of the lung possess no such distinctive shapes.

In addition to lack of distinctive shapes, the pathology bearing regions (PBR) in our images cannot be segmented out by any of the state-of-the-art segmentation routines because these regions often do not possess sharp edges and contours. Our system, therefore, enlists the help of the physician for delineating the PBRs and any relevant anatomical landmarks. Using graphical
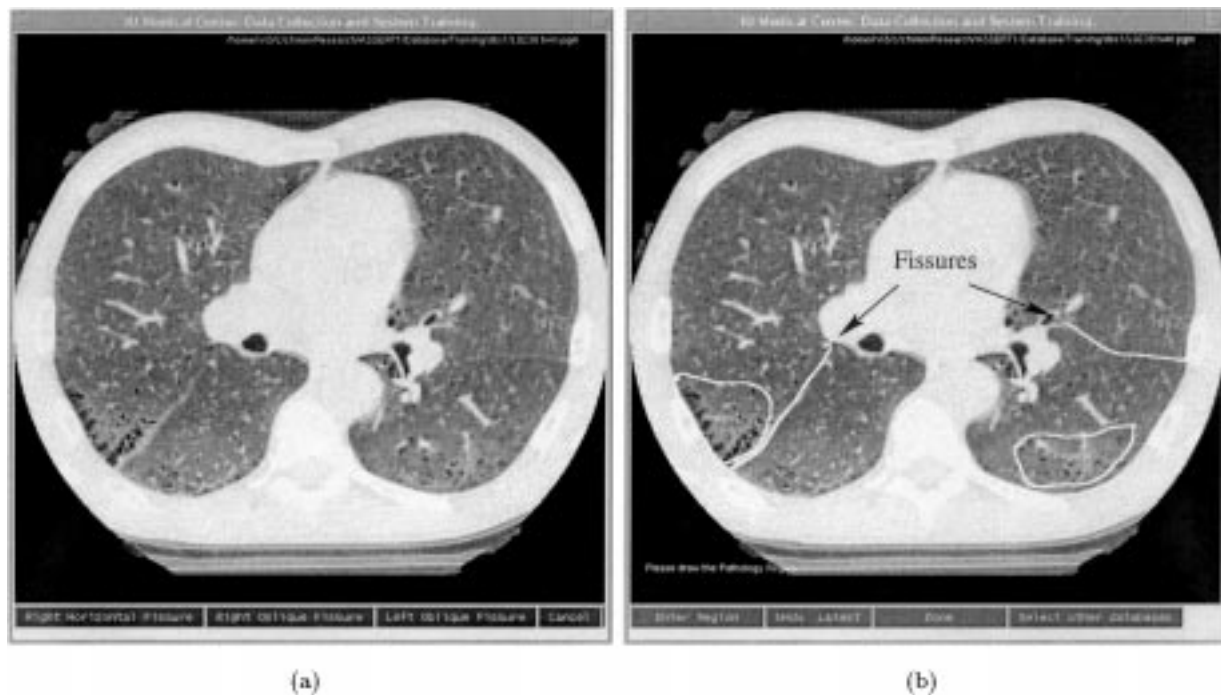
FIG. 1. Shown here are (a) an HRCT lung image, (b) the same image after a physician has delineated the PBRs and the fissures by dragging the screen pointer. Each PBR is delineated by an enclosing contour. Black arrows point to the fissures highlighted by the physician.

tools that we developed, it takes a physician only a few seconds to delineate the PBRs and any relevant anatomical landmarks.

The rest of this paper is organized as follows: In Section 2 we review the overall architecture of our system called ASSERT [26]. In Section 3, we first present the perceptual categories used by radiologists for detecting lung pathology in HRCT images. This serves as a background to our description of the image attributes used for characterizing each PBR. We then present a feature grouping concept in the form of lobular feature sets that allow for fast retrieval of images from the database. In Section 4, we describe how our image characterization scheme forms the basis for a multidimensional hashing approach to database indexing. Finally, in Section 5, we show three sets of experimental results. The first experiment shows that localized attributes are superior to global attributes for the domain of HRCT images of the lung. In the second experiment, we present empirical results showing that the subjective nature of a physician's delineation of a PBR has relatively small impact on retrieval results. The third experiment reports the results of a comparison of retrieval based on multidimensional hashing with retrieval based on the nearest-neighbor approach. The conclusion to be drawn from this experiment is that multidimensional hashing realizes a major saving in retrieval time while not incurring any significant loss in retrieval precision over the nearest-neighbor approach. This paper also includes an appendix where we have presented our algorithm for the extraction of lung regions in HRCT images.

2. SYSTEM OVERVIEW

The reasoning and control architecture of ASSERT is best explained with the help of the flow chart shown in Fig. 2. The figure shows two phases of the operation of the system: the image archiving phase is depicted using bold flow links and the image retrieval phase using thin flow links. The dashed bold links are shared by both phases. To apply our system to a different domain, only the shaded modules would need to be replaced.

To archive an image into the database, a physician delineates the PBRs and any relevant anatomical landmarks. This interaction takes only a few seconds for a trained domain expert (a radiologist). In the meantime, a lung region extraction algorithm is applied to the image to determine the boundary of the lungs. The system then executes a suite of image processing algorithms to create attribute vectors that characterize the PBRs individually and the portion of the image that consists of just the lung regions. These attributes are subject to a sequential forward selection algorithm (SFS) [15] to reduce the dimensionality of the attribute space while retaining the ability to accurately classify each image as belonging to its associated disease pattern. Figure 1b shows an HRCT image with the PBRs and the anatomical landmarks (in this case, the lung fissures) as delineated by a physician.

The information regarding the pathology of the lung resides as much in the location of each PBR as it does in the visual characteristics of the PBRs. Physicians characterize the location of a PBR as being either in the interior of a region that is referred
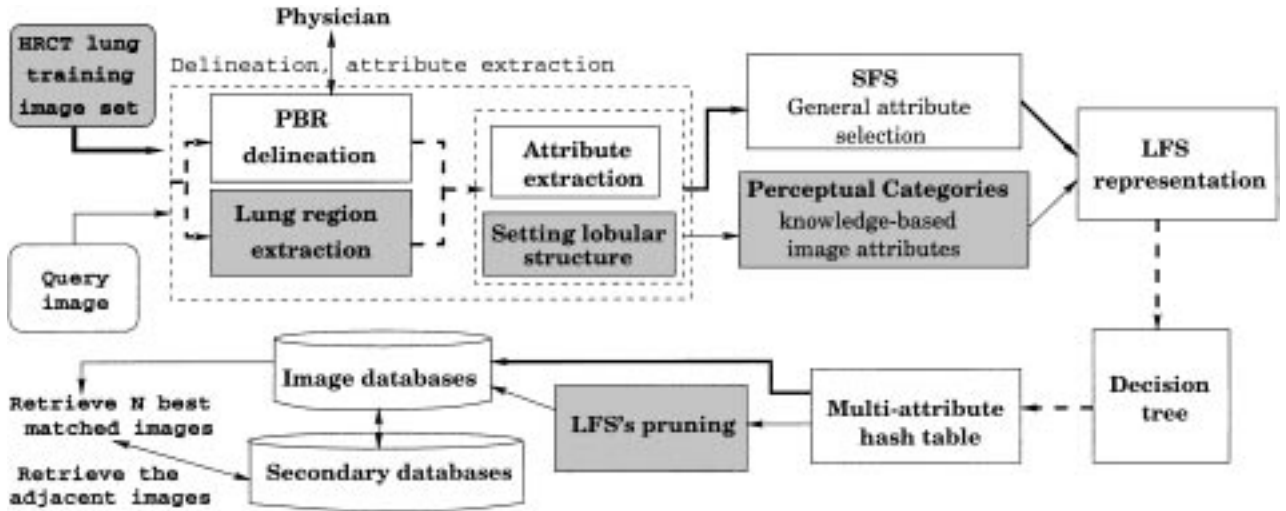
FIG. 2. The flow chart for archiving images and their retrieval.

to as a lobular region—a term that will be explained more fully shortly—or being adjacent to the boundary of a lobular region. In ASSERT, this fact is taken into account by constructing what we call lobular feature sets. A lobular feature set (LFS) is a grouping of a specific region of the lung, usually a lobe or a combination of adjacent lobes, and the PBRs found therein. In a majority of cases, each LFS will contain a single PBR that may either be in the interior of the lobular region or adjacent to its boundary. There are cases, however, in which there will be two or more PBRs in a given lobular region. To show how ASSERT takes into account the location of a PBR, let us assume that an LFS is allowed a maximum of two PBRs and, when both are present, one is adjacent to the boundary of the lobular region and the other is in the interior. If each PBR is represented by N attributes, a $2N$-dimensional attribute space may now be constructed for representing such an LFS, in accordance with the data structure shown in Fig. 3. Note that the first N dimensions are reserved for the attributes of a PBR that is adjacent to the boundary of the lobular region, and the last N dimensions for the attributes

of a PBR that is in the interior. Evidently, if a lobular region contains only a single PBR that is interior to the region, the first N attribute values will all be identically zero. By the same token, if a lobular region contains only a PBR that is adjacent to the boundary of the region, the last N attribute values will be identically zero. Very rarely, one does see a lobular region that contains two or more PBRs that are all either adjacent to the boundary or in the interior. In our current implementation, we retain only the largest of such PBRs for image archiving and retrieval. Further details on the construction of this LFS representation will be presented in Section 3.3.

The important point to note from an architectural standpoint is that each LFS gets represented by a single point in a $2N$-dimensional attribute space; this point also has associated with it the identity of the lobular region. Because of the noise phenomena associated with attribute measurement in images, clinically similar LFSs from different images in our database will occupy a statistical distribution in the $2N$-dimensional attribute space. All clinically similar LFSs will be considered to form an LFS
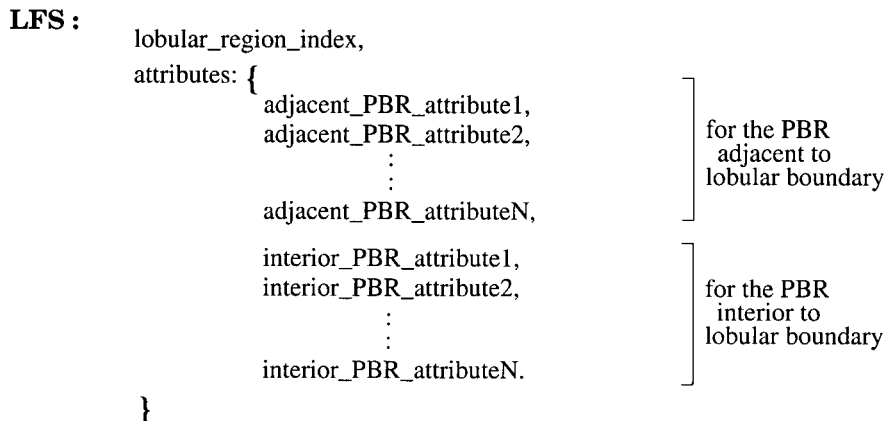

FIG. 3. The structure of a lobular region set.

class, denoted $\mathcal{LFS}_j$ for the jth such class. Our current database is represented by 46 $\mathcal{LFS}$ classes. The statistical distribution in the attribute space for each $\mathcal{LFS}$ class will be approximated as a Gaussian. As described later in Section 3.2, N equals 26 for the current implementation of ASSERT.

In order to make retrieval computationally efficient, we use the hashing method presented in [7]. In this scheme, the $2N$-dimensional attribute space is divided into bins such that—at least in the ideal case—each bin would point to a single $\mathcal{LFS}$ class. (Of course, when the statistical distributions for the different $\mathcal{LFS}$ classes overlap, there will be bins that point to multiple $\mathcal{LFS}$ classes). When a new query image comes in, its LFS is translated into a hash index that directly points to the appropriate bin of the attribute space. This bin contains pointers to the relevant $\mathcal{LFS}$ classes whose distributions cover that bin. The class pointers lead us directly to a set of images that are similar to the query image on the basis of LFS similarity. Using a Euclidean metric, the database images thus retrieved are then tested more directly for similarity to the query image and ranked on that basis. As will be discussed later, the hash table itself is constructed with the help of a decision tree algorithm.

The manner in which we have used LFSs has a rich history in computer vision where such feature groupings have been used effectively to ameliorate the computational complexity of directly comparing every model to every possible object hypoth-

esis that can be constructed from a scene. Our LFSs are conceptually identical to local feature sets in [3, 7] and to local feature focus in the system devised by Bolles and Cain [2], the goal of these systems being fast and robust object recognition.

Not every image produced in an HRCT scan of a patient is entered into the database by the physician. Depending on the size of a diseased region inside the lung, it is not unusual for such regions to show up in many adjacent cross-sectional images. Usually, a physician will consider only one of these images as the most important exemplar of the pathology exhibited in that part of the lung. It is this image, to be called the *key frame* henceforth in analogy with such frames in video sequences, that a physician enters into the ASSERT database. Since the other contiguous images also carry important diagnostic information—information regarding the three-dimensional extent and shape of the pathology—these additional images are stored in secondary storage, as shown schematically in Fig. 2. A sequence of such contiguous cross-sectional images is indexed by its corresponding key frame in the main database. Subsequently, if a physician desires to view the entire sequence associated with a given retrieved image, the sequence is fetched from the secondary storage.

To whet the appetite of the reader for this aspect of ASSERT, we have shown in Fig. 4 an HRCT query image at the upper left
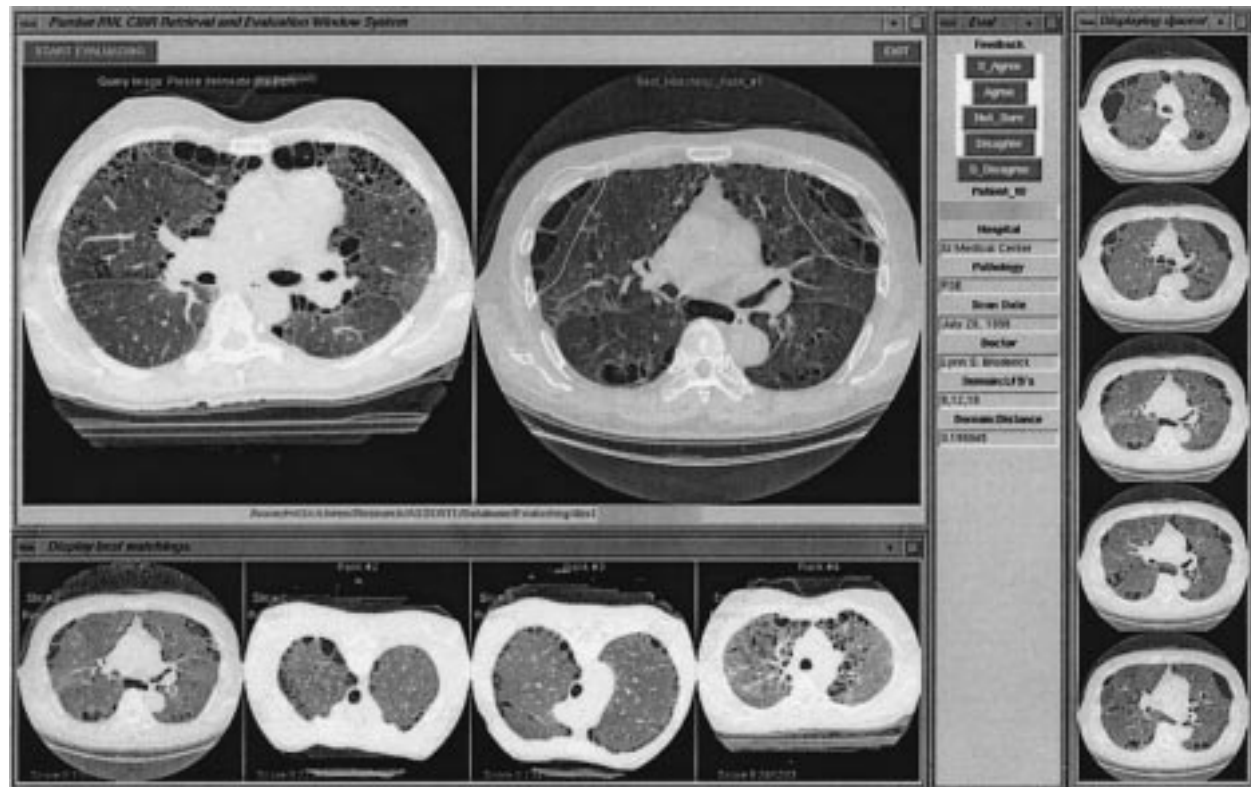

FIG. 4. The graphical interface for retrieval and for displaying adjacent images. The query image is on the left in the main window. The four best retrieved images are shown in the row below. An enlarged view of the retrieved image is displayed in the right side of the main window. HRCT images displayed in the right column are adjacent images to the retrieved image.
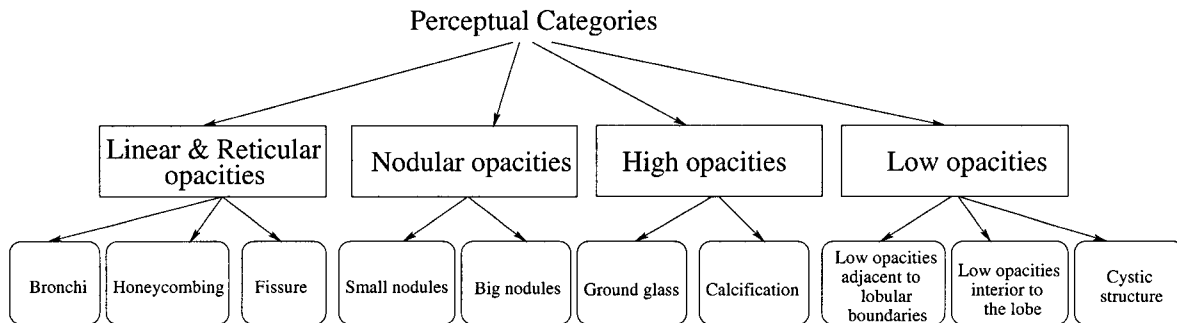
Perceptual Categories

[Linear & Reticular opacities] [Nodular opacities] [High opacities] [Low opacities]

Bronchi | Honeycombing | Fissure | Small nodules | Big nodules | Ground glass | Calcification | Low opacities adjacent to lobular boundaries | Low opacities interior to the lobe | Cystic structure

FIG. 5. Perceptual categories used by physicians.

in the main window. The best four retrieved images are shown in the row below. The physician clicks on one of these to get its enlarged view at the right side of the main window. In this case, the large image in the right part of the main window is the same as the leftmost retrieved image from the bottom row. In the right-most column, the system then displays the physically contiguous HRCT images that are from the same sequence as the enlarged retrieved image shown in the right side of the main window.

3. IMAGE CHARACTERIZATION

To facilitate accurate indexing and retrieval, it is of critical importance that the images in the database be characterized with relevant attributes. As might be expected, our attribute set includes all the usual attributes—such as those that measure texture, shape, and gray scale properties. But, more importantly, it also includes attributes that measure the perceptual categories used by physicians for identifying pathology in HRCT lung images. Because the first set of attributes is exhaustive, it can be too large and create statistical difficulties during classification. The dimensionality of this set is reduced by applying sequential forward selection search [15]. Together these two sets of attributes comprise an attribute vector that is used for database archiving and retrieval.

In the remainder of this section we first present the perceptual categories used by physicians, together with the attributes that we believe are most effective in measuring the presence or the absence of these categories. We then present the general purpose image attributes that are used to measure texture, shape, gray level properties, etc. Finally, we describe a higher level organization of these attributes into lobular feature sets (LFS); this organization captures spatial attributes such as the location of a PBR in relation to anatomical landmarks. It is this LFS representation that gets translated into a multidimensional index for archiving and retrieval.

3.1. Perceptual Categories of Lung Pathology

In this section we present the perceptual categories of the various diseases relevant to the domain of HRCT images of the lung. Figure 5 shows the perceptual categories that physicians

use for detecting pathology in HRCT images of the lung. The four major categories are [28]:

1. linear and reticular opacities,[1]
2. nodular opacities,
3. diffuse regions of high attenuation (high-opacities) and
4. diffuse regions of low attenuation (low-opacities).

These categories can be called major in the sense that, in the physician's mind, they capture what is perceptually common to the subcategories listed in the child nodes of the major categories. For example, thin and filament-like elements form the building blocks of the visual patterns corresponding to the leaf nodes "Bronchi," "Honeycombing," and "Fissure." For "Bronchi," these filament-like elements form circular rings; for "Honeycombing," they form criss-cross patterns; for "Fissure," they are simply just elongated.

Linear and reticular opacities.   As the name implies, the visual structures that correspond to this category consist of line or ribbon like formations that can either be straight and elongated, or web-like, or circular with a dot-like protrusion (the last is also referred to as a signet-ring pattern). These visual structures are caused by the thickening of the interstitial fiber network.[2] In our current collection of HRCT lung images, about 20% of the patient cases exhibit linear and reticular opacities.

Three diseases give rise to linear and reticular opacities in HRCT lung images. The first of these is called *bronchiectasis*, which results in the thickening of the walls of the bronchi. (Bronchi are air-filled passages in the lung that due to their low X-ray attenuation show up as dark regions.) In *bronchiectasis*, the lumen—the interior of the bronchi—also exhibit enlargement and their radii become larger than the radii of the adjacent pulmonary arteries. Figure 6a presents a simplified rendition of linear and reticular opacities. Figures 6b and c show this condition in HRCT images.

---

[1] Opacity refers to the extent of attenuation suffered by X-rays as they propagate through the tissue. High opacity is synonymous with high attenuation. Pixels corresponding to high attenuation regions will be whiter in HRCT images.

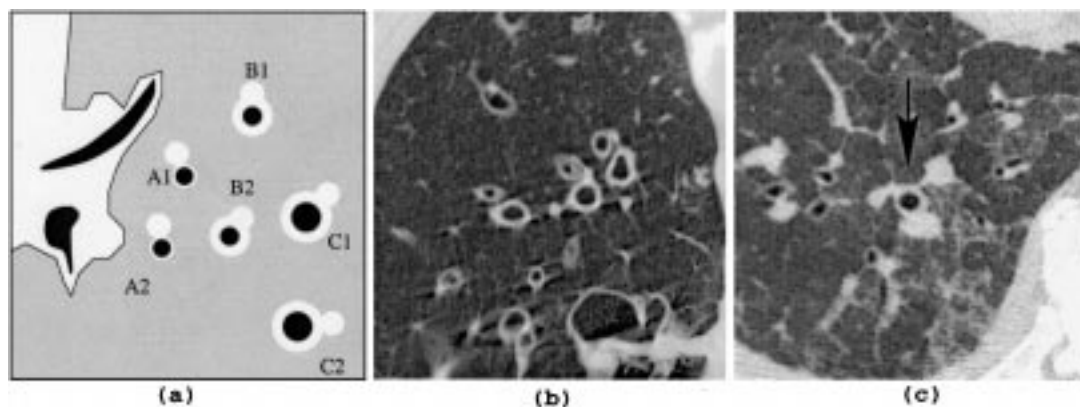[2] Interstitium is the fibrous supporting structure of the lung.

FIG. 6. (a) Normal bronchi labeled A1 and A2 show uniformly thin-walled bronchi with almost the same radii of adjacent pulmonary arteries. Thick-walled bronchi are labeled B1 and B2. Thick-walled bronchi with enlarged lumen are labeled C1 and C2. (b) A lung image exhibiting bronchiectasis. (c) A dark arrow points to bronchial wall thickening.

The image characteristics of the type shown in Fig. 6 respond well to a "dual thresholding" scheme in which a high threshold is first used to extract the white regions and then another high threshold is applied to the inverted image to extract the dark regions that are surrounded by the white regions. The resulting pairs of dark and white regions, corresponding to the walls (together with any adjoining arteries) and the lumen of bronchi, are accepted only if their centers of mass are colocated to within a tolerance value determined empirically. For each such pair, the following attributes are extracted: *the thickness of the bronchi walls*, and *the sizes of pulmonary arteries adjacent to the bronchi*.

The second disease pattern that exhibits linear and reticular opacities is *peripheral honeycombing*. As the name implies, *honeycombing* is a pattern formed by adjoining air-filled cysts, as shown in the region indicated by the arrows in Fig. 7a. Shown there are *peripheral honeycombing* regions adjacent to the boundaries of the lung. These patterns respond to the algorithm in which the system first applies a threshold to a PBR to segment out the reticular structure, followed by skeletonization. The cells formed by the skeletons are then collected to form a cluster cor-responding to the honeycomb region. The following attributes are extracted for such patterns: *the number of cells, the average cell size*, and *the number of cells adjacent to the lung boundaries or fissures*.

The last disease that gives rise to linear and reticular structures is *subpleural interstitial thickening*. For this disease, important evidence concerning lung pathology is contained in the thickening of the fissures that separate the different lobes of each of the two lungs, particularly if this thickening is irregular. In Fig. 7b, the arrow points to a normal fissure in the right lung. The width of the fissure is normal on the basis of its thickness in relation to the size of the vessels in the vicinity. In Fig. 7c, the arrow points to an abnormal fissure. As a part of the interaction between the physician and the system, for every image the physician is asked to drag the mouse along each of the fissures. The system then applies a snake algorithm [13] to optimally fit the delineation to the fissures. The one-pixel wide fissure contour thus obtained is then dilated to occupy all the pixels on the fissure if the evidence of fissure thickening exists. The attribute measurements for fissure thickening are *the average width of the fissure* and *the average curvature of the fissure contour*.
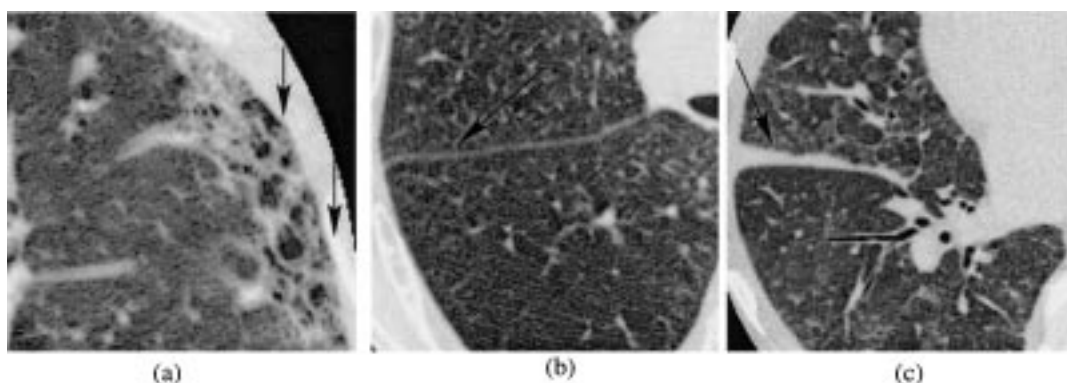

FIG. 7. (a) Dark arrows point to peripheral honeycombing regions. These regions are adjacent to the boundary of the lung. (b) A normal fissure in the lung. (c) Irregular thickening of fissure in the lung.
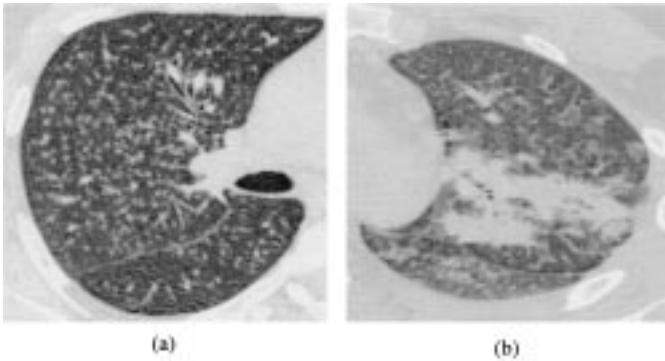
FIG. 8. (a) Small nodules; (b) conglomerate mass.

Nodular opacities. The gray values associated with nodules and 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 distinguished on the basis of the size and locational distributions associated with the nodular opacities. The disease *sarcoid* results in these types of images.

The nodular opacities appear typically in two different sizes: small nodules, which are roughly round and less than one centimeter in diameter, and large nodules of irregular shape, whose "diameter" exceeds one centimeter. Sometimes large nodules agglomerate into large masses, as shown in Fig. 8b. For the case of small nodules, their distribution carries diagnostic information. When the distribution is random, then the nodules appear widely and evenly throughout the lung as shown in Fig. 8a. Distributions become nonuniform when nodules attach themselves to the boundaries of the lungs or to the fissures.

Images with nodules respond to attribute extraction algorithms in which the system first applies a threshold to the lung regions, followed by the measurement of "roundness" property. (Note that this does not cause any confusion with the previously discussed method for the extraction of bronchi since, for the case of bronchi, we seek dark regions surrounded by concentric white regions.) The roundness property is particularly effective for extracting small nodules. The large 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. Effective attribute measurements for images with this type of pathology include *the average sizes of nodules, average roundness of nodules, histogram of nearest-neighbor distances among the nodule centers* [23], and the *gray-level mean of nodules.*

Diffuse regions of high attenuation (high-opacities). All lung pathologies manifest themselves as changes in the opacity of the tissue as seen in the images. For many of the pathologies, these opacities can be localized and appear as patterns with certain geometric characteristics; these cases were discussed above. For other pathologies, the entire lung may assume a different shade of gray in comparison to a normal lung shown in Fig. 9a. For example, shown in Fig. 9b is what is referred to as *ground-glass opacity*. Note that the underlying vessels are not obscured; that is, the vessels can be seen clearly in the lungs even though the tissue everywhere is characterized by a higher level of opacity. A different pattern in the same perceptual category corresponds to *calcification*, shown in Fig. 9c; the disease *metastatic calcification* gives rise to this pattern. For patients suffering from this disease, the overall visual effect on the HRCT image is that of a marked increase in density, similar to the density of bone. For this disease, there are also areas of ground-glass opacity.

Algorithms capable of separating the normal tissues from the ground-glass tissues make use of the fact that the gray-level histogram for the latter case is strongly bimodal, whereas it is primarily unimodal for the normal tissues. After the ground-glass tissues are extracted, the vascular structure is extracted by employing the well-known technique of co-occurrence matrices [10] with different values for the distance parameter. The computed measurements are *uniformity of energy, homogeneity, gray level mean of ground-glass regions*, and *the ratio of abnormal regions to lung regions.*

Diffuse regions of low attenuation (low opacities). All of the previous disease categories are marked by increased attenuation (meaning higher gray levels) associated with the pixels corresponding to the diseased tissues. The disease categories in
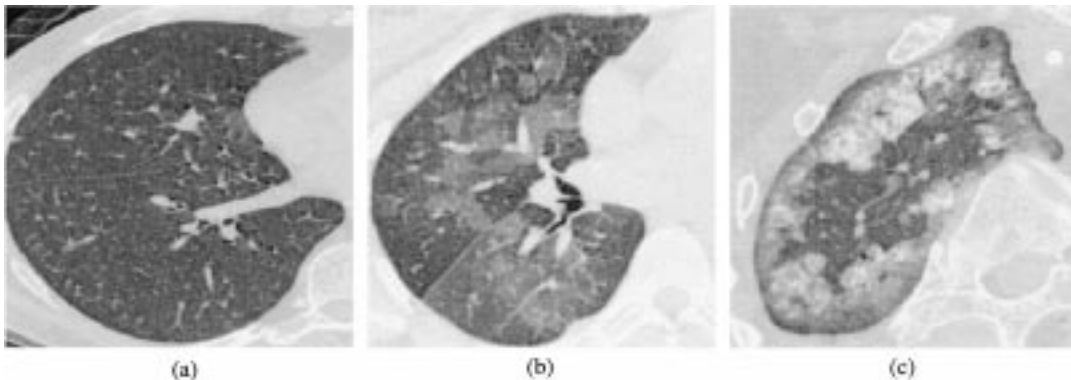

FIG. 9. (a) Normal lung. (b) Lung exhibiting ground-glass opacity with identifiable pulmonary vessels. (c) Lung exhibiting metastatic calcification.
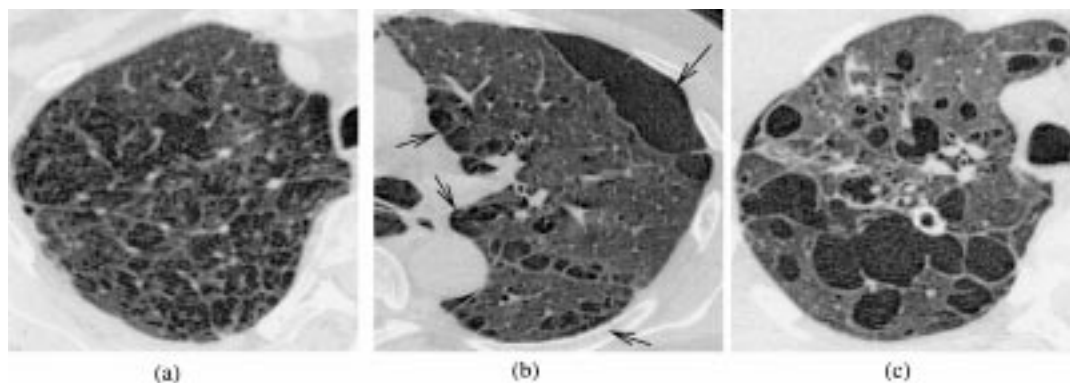
FIG. 10. (a) Centrilobular emphysema. (b) Paraseptal emphysema. (c) Lung cysts. Unlike centrilobular emphysema, lung cysts have clearly defined walls and larger low-attenuation regions.

this section are marked by *decreased* attenuation, meaning lower or darker gray levels.

For example, *centrilobular emphysema* shows up in HRCT images in the form of a large number of areas with significantly decreased density, as shown in Fig. 10a. These areas may occupy the entire lung region, but are likely to predominate in the upper lobes. When the disease becomes severe, these areas may join together to form a large region of low attenuation. *Paraseptal emphysema* also results in decreased density, except that the areas now occur adjacent to the boundaries of the lung or in the vicinity of the fissures, as shown in Fig. 10b. The adjacent areas shown are usually separated by thin walls. Diseases other than emphysema also exhibit decreased lung opacity. Figure 10c shows a lung with cystic structure. In this case the regions of low-attenuation are larger than those for the case of centrilobular emphysema and the walls of these regions possess better definition.

For attribute extraction, the lung regions extracted from these images respond to the following steps. First, the normal tissues and the low-attenuation tissues are separated by simple thresholding. (The gray-level histogram is strongly bimodal for all these diseases.) Next, the co-occurrence matrices are computed for the low pixels resulting from thresholding. Additionally, *the number of decreased density regions adjacent to the lung boundaries or fissures* is also computed, as it carries diagnostic information for the diseases mentioned in this section.

Selection of thresholds. The discussion so far has identified the following thresholds that are used to extract the attributes corresponding to the relevant perceptual categories:

1. High and low thresholds for extracting the bronchi structure,
2. Gray scale threshold to extract the nodules,
3. Gray scale thresholds for measuring high opacities,
4. Gray scale thresholds for measuring low opacities.

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 [17] 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 the between-class variance and the sum of the within-class variances. This approach allows each threshold to adapt to the image in question. In summary, Fig. 11 shows the perceptual categories and their corresponding image attributes.

3.2. General Purpose Image Attributes

While some of the general purpose attributes are local to the PBRs, others are global to the entire lung region.[3] Each PBR is characterized by a set of shape, texture, and other gray-level attributes. For characterizing texture within PBRs, we have implemented a statistical approach based on the notion of a gray-level co-occurrence matrix [10]. This matrix represents a spatial distribution of pairs of gray levels and has been shown to be effective for the characterization of random textures. In our implementation, the specific attributes we extract from this matrix are *energy, entropy, homogeneity, contrast, correlation*, and *cluster tendency*. In addition to the texture-related attributes, we compute three additional sets of attributes of the pixels within the PBR boundary. The first set computes gray-scale properties of the pixels within the PBR, specifically, the mean and standard deviation of the gray levels, a histogram of the local gray levels, and the attributes of its shape (longer axis, shorter axis, orientation, shape complexity measurement using both Fourier descriptors and moments). The second set computes the edginess of the PBR using the Sobel edge operator [20]. The extracted edges are used to obtain the distribution of the edges. We compute the ratio of the number of edge pixels to the total number of pixels in the region for different threshold channels, each channel corresponding to a different threshold for edge detection. Finally, to analyze the structure of gray level variations within

---

[3] Note that the sense in which we use the word "global" is different from how it is commonly used in the literature on CBIR. Our global attributes are global only to the extent that they are based on all the pixels in the extracted lung regions.
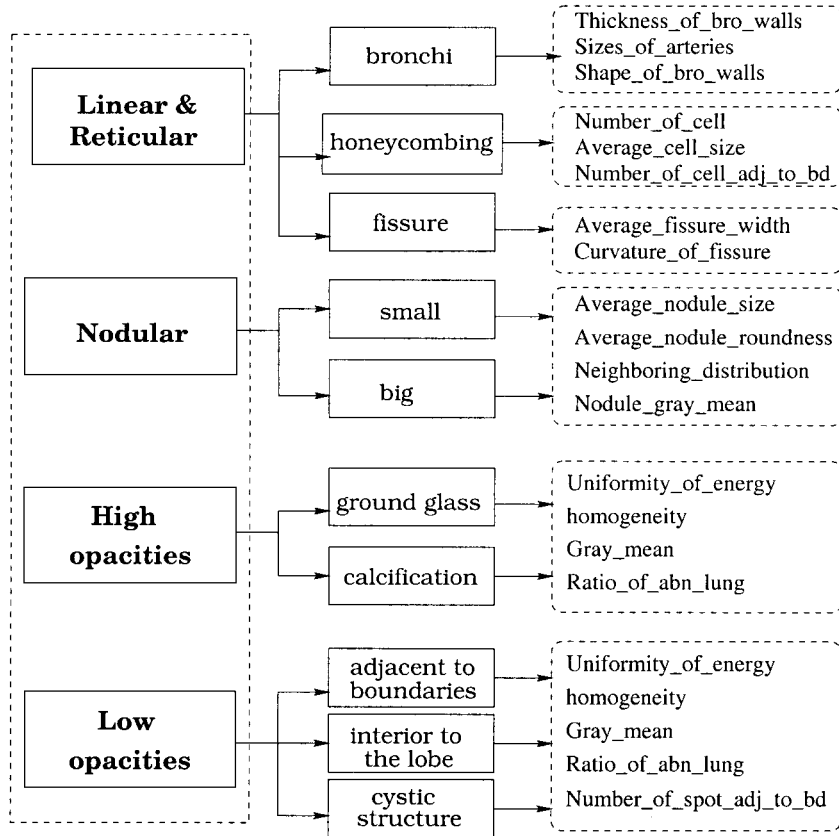
FIG. 11. The perceptual categories (in the first two columns) and the attributes that are extracted to measure the presence or the absence of the patterns corresponding to each category. These attributes are extracted from each PBR.

the PBR, we apply a region-based segmenter [19]. From the resulting segmentation, we compute the number of segmented regions in a PBR, a histogram of the sizes for all these regions, and gray-level statistics for each of the regions.

In addition to the texture and shape attributes, a PBR is also characterized by its average properties, such as gray scale mean and deviation, with respect to the pixels in the rest of the lung. Measurement of these properties requires that we be able to segment out the lung region (note that the lung region is also needed for the measurement of the global attributes). To extract the lung region, we apply a set of binary-image analysis routines that are described in Appendix A. In addition to the average-type attributes, the system also calculates the distance between the centroid of a marked PBR and the nearest lung boundary point. We include the anatomical information because physicians use this information to classify some pulmonary disease patterns.

The total number of general-purpose attributes, 255 in number, computed for a PBR is large. Since we did not know a priori which of these general purpose attributes would be effective for retrieval, we intentionally included all of them. For obvious reasons only a small subset of these attributes can be used for database indexing. The attributes actually used are found by applying the sequential forward search (SFS) algorithm [15] to all the 255 general attributes. The result is that we are left

with only 12 attributes, listed as a_1 through a_{12} in Table 1. Note that attributes a_{13} to a_{26} correspond to the perceptual categories described in the previous subsection.

3.3. Lobular Feature Sets

As mentioned before, the location of pathology plays an important role in disease diagnosis, and, as a consequence, it also plays an important role in image retrieval. The pathology bearing pixels must exist in certain locations with regard to a chosen set of anatomical landmarks in order to convey diagnostic information. For example, both centrilobular emphysema (CLE) and paraseptal emphysema (PSE) show low-attenuation areas in HRCT images. However, CLE is always interior to a lobe of the lung and PSE is always adjacent to either the lung boundary or one of the fissures. In this section we explain how such information is incorporated into the archival and retrieval process through a concept we have named lobular feature sets (LFS).

A *lobular region* is defined by the lung boundary and the fissures that are present in each lung. The left lung has one fissure, labeled LO for "left-oblique" in Fig. 12, and the right lung has two that in the same figure are labeled RO for "right-oblique" and RH for "right-horizontal." (Note that because of the geometry of imaging, the left lung shows up on the right

TABLE 1

The Attributes a_1 Through a_{12} Are the General-Purpose Attributes Selected by the SFS Algorithm. The Rest of the Attributes Correspond to the Various Perceptual Categories in Fig. 11

Index	Attribute
$a_1 - a_2$	Differences b/w global and the PBRs gray-level mean and STD
$a_3 - a_5$	Global histogram (bin 6, 9, 16 out of 16 bins)
a_6	Area of the PBR
a_7	Contrast of the PBR
a_8	Edginess histogram of the PBR
a_9	Distance from the PBR to lobular region boundary
a_{10}	Entropy after gamma
a_{11}	Homogeneity after gamma
a_{12}	Cluster tendency after gamma
a_{13}	Thickness of bronchial wall
a_{14}	Shape of bronchial wall
a_{15}	Adjacent artery size
a_{16}	Cell number
a_{17}	Average cell size
a_{18}	Adjacent cell number
a_{19}	Average width of adjacent fissure
a_{20}	Curvature of adjacent fissure
a_{21}	Average size of nodules
a_{22}	Average roundness of nodules
a_{23}	Histogram of nearest neighbor distances for nodules
a_{24}	Average gray mean of nodules
a_{25}	Uniformity of energy
a_{26}	Ratio of abnormal region to lung region.

and the right lung on the left in HRCT images.) The fissures divide each of the lungs into lobes. The left side of an HRCT lung image would ideally show two fissures and the right side one because the right lung has three lobes and the left lung two. Only a subset of these fissures is visible in most images in our database. A fissure may or may not be visible, depending on where exactly a lung cross section is taken. Those fissures that
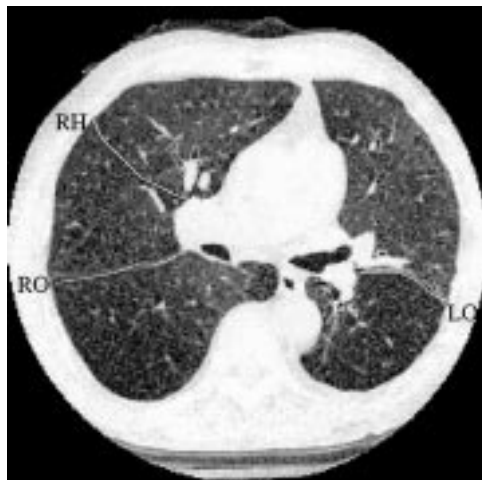

FIG. 12. Three fissures: RO, right oblique; RH, right horizontal; LO, left oblique.

are visually discernible are marked by the physician at the time he/she delineates the PBRs.

We use the term "lobular region" to refer to the divisions created within a lung by the physician-demarcated fissures. Therefore, depending on which, if any, of the fissures are visible, the right lung of an HRCT image can be divided into any of the six lobular-region categories as shown in Fig. 13. Each of these lobular regions thus generated carries a unique label as shown in Table 2. Similarly, the left lung can be divided into three lobular-region categories shown in Fig. 14. When comparing a query image with a database image, our first preference would, of course, be to ensure that the LFSs extracted from the two have identical lobular region labels. But, in the absence of lobular region labels being identical, we must allow for the possibility of, say, a query-image LFS with lobular region label RH_RO to be considered similar to a database image LFS whose lobular region label is RH_LL. This consideration gives rise to the two lobular region label trees shown in Fig. 15.

While the first component of an LFS is the lobular region label, the second component is the set of PBRs within the lobular region. Each lobular region and the PBRs contained therein constitute a lobular feature set. For the purpose of illustration, Fig. 16 shows two hand-drawn examples with partitioning of the lungs into different types of lobular regions. The LFSs for the example in Fig. 16a are

$$\text{LFS}_1 = \{\text{RH\_TL} \mid \text{PBR}_{\text{adjacent}} = \text{PBR}_1, \text{PBR}_{\text{interior}} = \text{PBR}_2\}$$

$$\text{LFS}_2 = \{\text{LL} \mid \text{PBR}_{\text{adjacent}} = \text{PBR}_3, \text{PBR}_{\text{interior}} = \text{PBR}_3\}.$$

The LFSs for the example on the right in Fig. 16 will be

$$\text{LFS}_1 = \{\text{RH\_TL} \mid \text{PBR}_{\text{adjacent}} = \text{null}, \text{PBR}_{\text{interior}} = \text{PBR}_3\}$$

$$\text{LFS}_2 = \{\text{RH\_RO} \mid \text{PBR}_{\text{adjacent}} = \text{PBR}_1, \text{PBR}_{\text{interior}} = \text{null}\}$$

$$\text{LFS}_3 = \{\text{RO\_LL} \mid \text{PBR}_{\text{adjacent}} = \text{null}, \text{PBR}_{\text{interior}} = \text{PBR}_2\}$$

$$\text{LFS}_4 = \{\text{LL} \mid \text{PBR}_{\text{adjacent}} = \text{PBR}_6, \text{PBR}_{\text{interior}} = \text{PBR}_4\}.$$

Note that in all cases, the first PBR is adjacent to the boundary of the lobular region and the second PBR is in the interior. This

TABLE 2
Lobular Region Labels

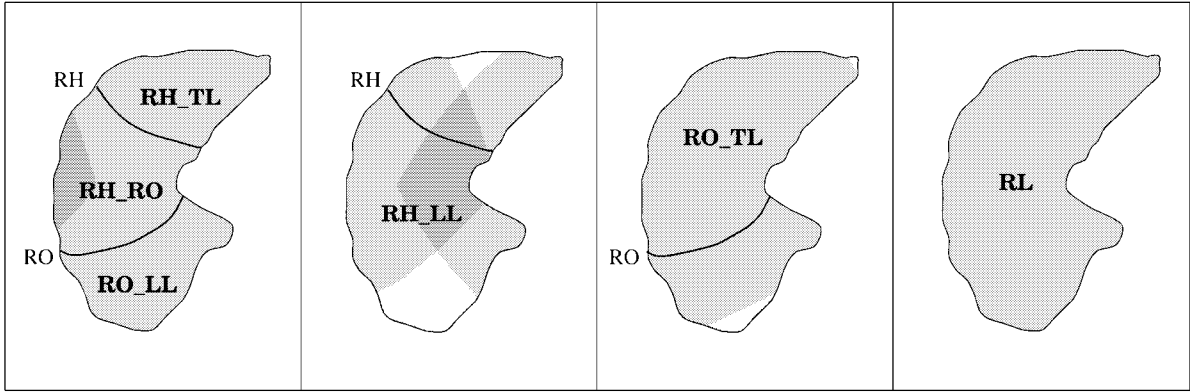| Region	Label
Right horizontal fissure with top lung boundary	RH_TL
Right horizontal fissure with oblique fissure	RH_RO
Right oblique fissure with bottom lung boundary	RO_LL
Right horizontal fissure with bottom lung boundary	RH_LL
Right oblique fissure with top lung boundary	RO_TL
Right lung boundary	RL
Left oblique fissure with top lung boundary	LO_TL
Left oblique fissure with bottom lung boundary	LO_LL
Left lung boundary	LL

FIG. 13. Six different lobular regions and their labels for the right lung.

relational design encodes the adjacency/interior attribute of the PBRs, in keeping with our discussion at the end of Section 2. Of course, if either of these PBRs is absent, then the respective entry in the LFS will be "null," as is the case with LFS$_3$. When a PBR covers the entire lobular region, for example PBR$_3$ in Fig. 16a, this PBR appears in both PBR$_\text{adjacent}$ and PBR$_\text{interior}$. When there is more than one PBR of the type "adjacent" or more than one PBR of the type "interior" inside a lobular region, we retain only the largest of each type. For example, the lobular region for LFS$_4$ contains three PBRs, two of them, PBR$_5$ and PBR$_6$, are adjacent to the boundary of the lobular region. Of these two, only PBR$_6$ is retained in LFS$_4$ because it is larger in the sense of its occupying a larger number of lung region pixels.

As mentioned previously, the notation LFS$_i$ represents a specific LFS. All LFSs extracted from the database are then divided into a discrete set of classes on the basis of equivalence of adjacency/interior relationships of the associated PBRs with respect to the lobular boundaries as well as the diagnosis made by the physician. For illustration, the LFSs extracted from the two images shown in Fig. 17 belong to the same $\mathcal{LFS}_j$ class because (1) the lobular regions are the same; (2) the adjacency/interior relationships of the PBR with respect to the lobular regions are the same; and (3) the disease diagnoses are the same. Each LFS class $\mathcal{LFS}_j$, $j = 0, 1, 2, \ldots$, will be modeled by a Gaussian distribution in the attribute space.

4. MULTIDIMENSIONAL HASHING FOR INDEXING IMAGES

In this section, we will explain how an LFS translates into an index for image archiving and retrieval. How an LFS is converted into a multidimensional index is based on the Multi-Hash method of [7] for fast 3D object recognition. From the standpoint of its application in ASSERT, a particularly noteworthy feature of Multi-Hash is that the bin boundaries in the attribute space are computed dynamically to suit the peculiarities of a database. To explain this point further, note that our goal is to carve the attribute space into bins such that each bin points to—at least in the ideal case—a single $\mathcal{LFS}$ class. Each $\mathcal{LFS}$ class will in turn point to the database images from which it was derived. The question then becomes how to partition the attribute space into bins. The Multi-Hash approach sets up the bin boundaries in such a way that an $\mathcal{LFS}$ class with large variability (because of measurement noise and other artifacts) occupies a large cell, compared with an $\mathcal{LFS}$ class with small variability. As was demonstrated in [7], this can be accomplished by minimizing the entropy associated with $\mathcal{LFS}$ class distributions.

To construct a multidimensional hash table for $\mathcal{LFS}$ classes, our system

1. Creates supervised training data by grouping those images together that yield identical LFSs. Two LFSs in two separate images belong to the same $\mathcal{LFS}$ class if the associated PBRs are in the same adjacency/interior relationship with respect to the lobular boundaries and if the PBRs carry the same diagnostic information.

2. Computes statistical parameters of the observed data for each attribute within each class. At this time we assume that the statistical distributions associated with the attributes are Gaussian for each of the attributes.
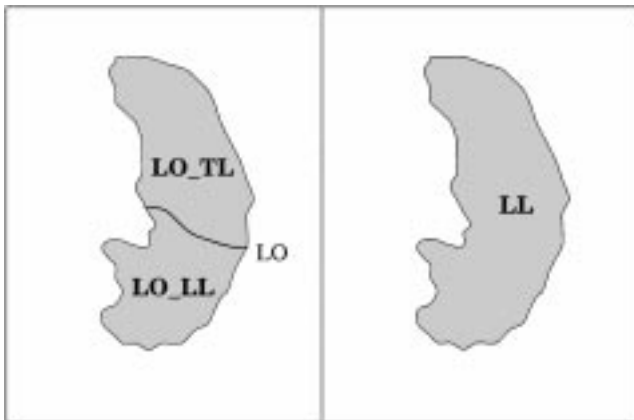

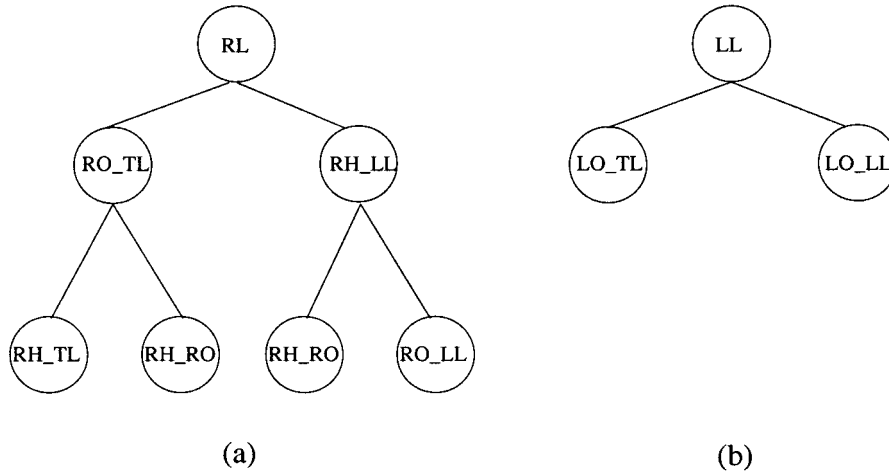**FIG. 14.** Three lobular regions and their labels for the left lung.

FIG. 15. Lobular region label tree (a) for the right lung and (b) for the left lung.

3. Creates a decision tree that minimizes the entropy over all the distributions associated with all $\mathcal{LFS}$ classes. In the ideal case, the minimum entropy will be zero and, in that case, each bin in the attribute space will point to a single $\mathcal{LFS}$ class. In practice, due to the overlapping nature of $\mathcal{LFS}$ class distributions, the minimum will be nonzero. Each test node in the tree splits the data with a test of the form $a_i > t_i$. Each leaf node is labeled with the $\mathcal{LFS}$ classes whose distributions overlap the bin corresponding to that node.

4. Maps the decision tree into a multidimensional hash table.

As a simple illustration for a two-dimensional attribute space, a decision tree and the corresponding hash table are shown in Fig. 18. The label (a_i, t_i) under each node means that attribute a_i

was subject to a decision threshold t_i. The solid lines in Fig. 18b show the decision tree thresholds in the attribute space. When the solid lines are extended to span the entire attribute space, we end up with a hash table.

To form a query to the database, the physician marks the relevant PBRs and anatomical landmarks in the query image. The same process as is used for archiving is now applied to form the LFSs for the query image. As we will explain below, the point in the attribute space that corresponds to the query LFS yields the address of the bin that points to the most similar database $\mathcal{LFS}$ classes. For our current database, on average this step returns 2.63 $\mathcal{LFS}$ classes for each hash table access. Pooling together the images corresponding to the retrieved $\mathcal{LFS}$ classes, this process, on average, results in the system retrieving 25 images
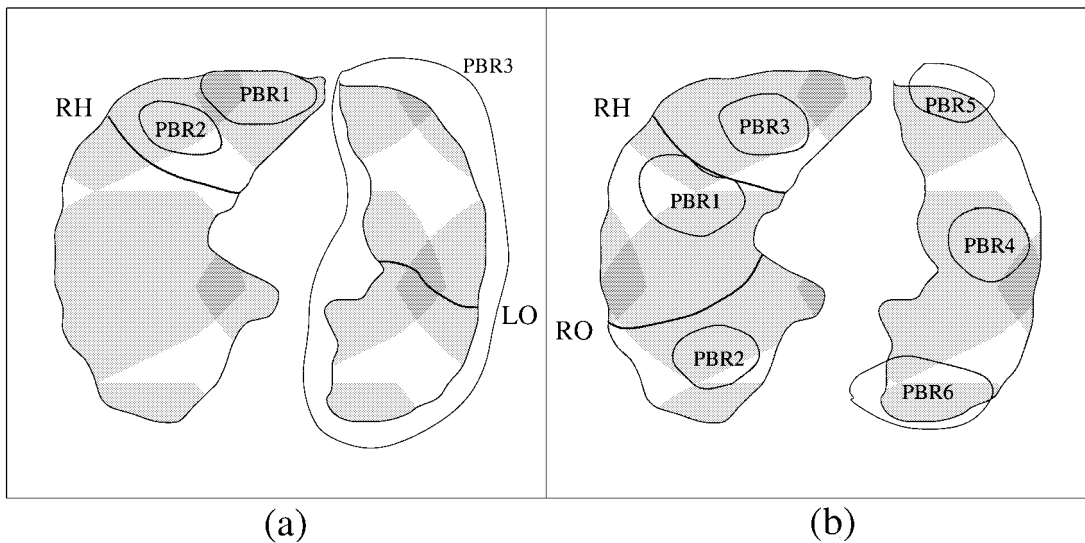

FIG. 16. (a) In the right lung, right horizontal fissure exists with top and bottom lung boundaries. In the left lung, left oblique fissure exists with top and bottom lung boundaries. (b) Right oblique and horizontal fissures exist but no fissure exists in the left lung.
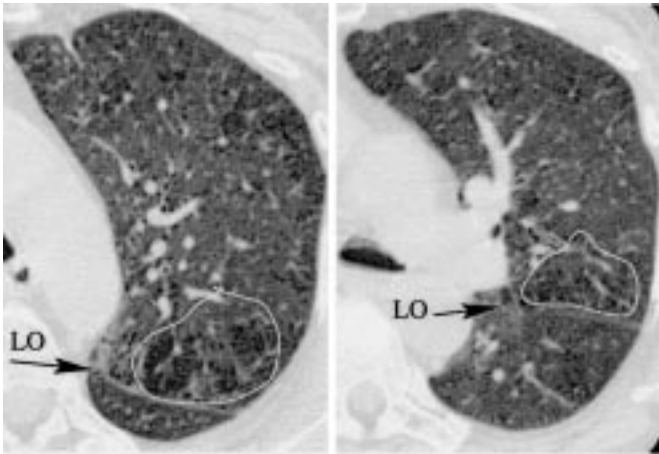
FIG. 17. LFSs in these two images belong to the same class.

from the database. To determine which of these images to display, we perform a filtering step described at the end of this section. Briefly, we apply a nearest-neighbor match between the query image's PBRs and the PBRs of the images corresponding to the $\mathcal{LFS}$ classes returned by Multi-Hash. In this manner, we reduce the computational burden that would otherwise be associated with comparing the query image attributes with those of every image in the database using, say, the nearest-neighbor algorithm.[4]

4.1. Growing a Decision Tree

We will now describe how a decision-tree-based algorithm is used to construct a minimum-entropy partition of the attribute space in which the $\mathcal{LFS}$ classes reside.[5] To determine which attribute to place at a node of the decision tree, the algorithm examines each attribute separately. For each attribute, the algorithm examines a set of candidate thresholds, selecting one that affects the largest reduction in the entropy of all the distributions in the attribute space. What follows is somewhat terse; the reader is referred to [7] for further details.

As mentioned previously, a Gaussian model is used to describe the distribution associated with each $\mathcal{LFS}$ class in the attribute space. We will use η_i to represent the mean value of the attribute a_i and σ_i^2 its variance for class $\mathcal{LFS}_j$. These two parameters are estimated from the observed instances. The sample mean and variance define the following marginal density function for attribute a_i:

$$f_j^i(a_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(a_i - \eta_i)^2/2\sigma_i^2}. \qquad (1)$$

Since we consider each attribute to be independently Gaussian, we are in effect saying that the 26 attributes (a_1 to a_{26} in Table 1) are statistically independent, an assumption that remains to be seriously tested.[6] For practical reasons, a truncated version of the above function is used in constructing the decision tree. The extent of the a_i axis that is used in the hash table is set to $\eta_i \pm 2\sigma_i$.

To determine whether a decision threshold t_i for attribute a_i reduces the entropy, we need to compare the entropy of the original bin (keeping all $\mathcal{LFS}$ classes together in that bin) to the entropy that would result from partitioning the bin into two child bins using a decision threshold. Assume that we have a total of L $\mathcal{LFS}$ class distributions occupying the original bin. The entropy of all these distributions in this one bin is given by

$$H(\text{bin}) = -\sum_{j=1}^{L} P(\mathcal{LFS}_j \mid \text{bin}) \log(\mathcal{LFS}_j \mid \text{bin}). \qquad (2)$$

The entropy of the partition is obtained by the weighted average of the entropy of each child bin resulting from partitioning the data at the point $t_i = a_i$:

$$H_{\text{Average}} = \sum_{i=1}^{2} P(\text{bin}_i) H(\text{bin}_i). \qquad (3)$$

If this average entropy is smaller than the entropy in Eq. (2), then dividing up the original bin with a decision threshold t_i along the axis a_i would be a good idea. Therefore, our goal is to select a decision threshold t_i for attribute a_i that minimizes the value of Eq. (3).

The search for the decision threshold is implemented by dividing the attribute axis uniformly into a large number of divisions, as shown in Fig. 19. Each of the quantization levels thus created is used as a candidate decision threshold. That candidate threshold which yields the smallest value for the average entropy in Eq. (3) is retained.

While the above approach to growing a decision tree is straightforward, it is made somewhat complex by the fact that the probability $P(\mathcal{LFS}_j \mid \text{bin}_i)$ is not available directly. We can, however, calculate this probability from the attribute distributions by using Bayes rule. We estimate from the attribute distributions

$$P(\text{bin}_i \mid \mathcal{LFS}_j) = \prod_{m=1}^{k} \int_{\max(a_T^m, C_{j,L}^m)}^{\min(a_T^m, C_{j,R}^m)} f_j^m(a_m) \, da_m, \qquad (4)$$

where a_T^m is the candidate threshold along attribute axis a_m. The integrals of Gaussians can be obtained from standard published

---

[4] Other possible approaches for indexing images that would also be computationally efficient could be based on B-trees [5], R-trees [8], and their variants [1, 21, 24].

[5] Finding a partition that would correspond to a global minimum for entropy is an NP-hard problem and cannot be guaranteed by the decision-tree-based algorithm used in Multi-Hash. The Multi-Hash partitions are generated by using locally optimal decisions.

[6] We believe that any correlations between the attributes are important only for getting an accurate measure of the entropy associated with a given bin structure. If most bins in the final hash table are pure, in the sense of pointing to single classes, the correlations between the attributes should not have much effect on the precision of retrieval.
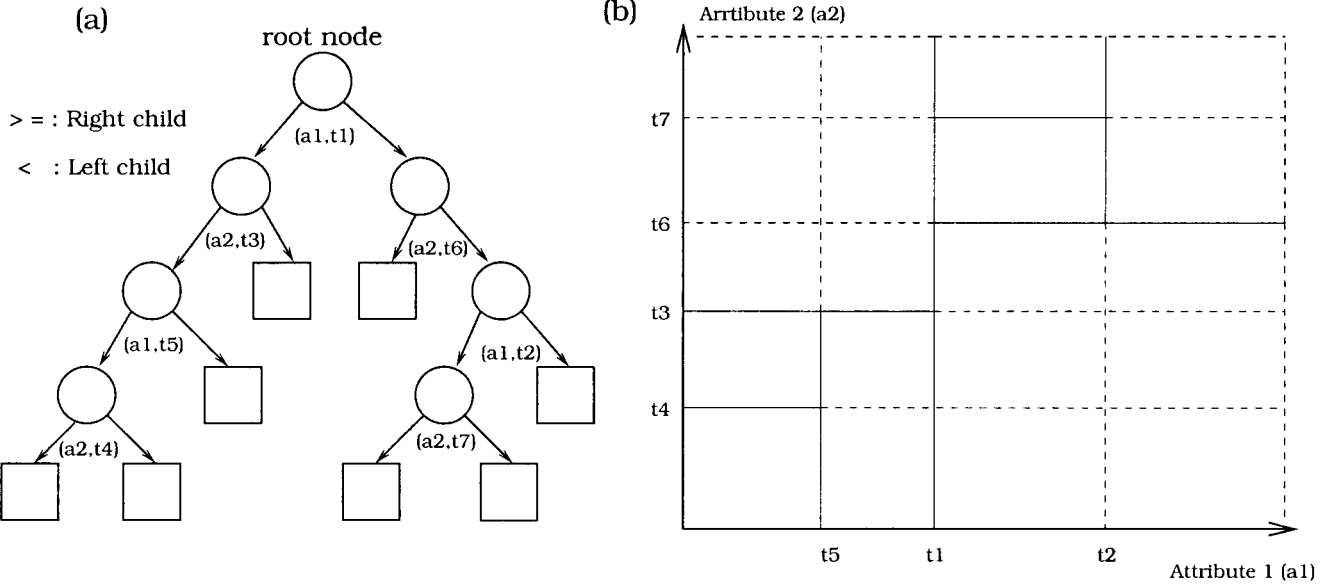
FIG. 18. (a) An example decision tree for the case of two attributes. Test nodes are depicted as circles and leaf nodes as rectangular boxes. (b) The corresponding hash table. The solid lines correspond to the decision thresholds in the decision tree in (a). The dashed lines then extend these solid lines so as to span the entire attribute space. The resulting bin structure forms the hash table.

tables. Applying Bayes rule, we have

$$P(\text{bin}_i, \mathcal{LFS}_j) = P(\text{bin}_i \mid \mathcal{LFS}_j)P(\mathcal{LFS}_j), \qquad (5)$$

where $P(\mathcal{LFS}_j)$ is the relative frequency with which the $\mathcal{LFS}_j$ appears in a training image. We also need the bin probabilities

$$P(\text{bin}_i) = \sum_{j=1}^{L} P(\text{bin}_i, \mathcal{LFS}_j). \qquad (6)$$
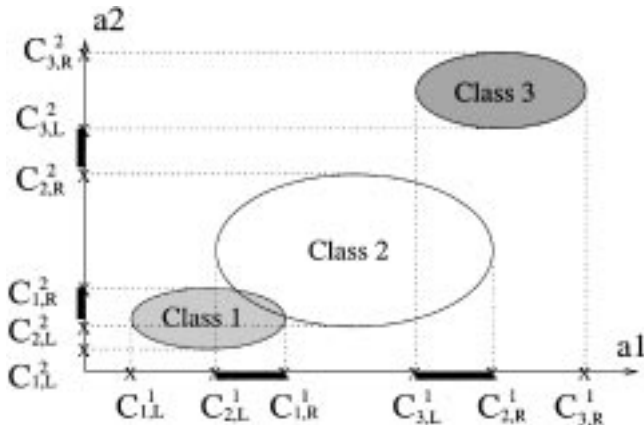
Now we can write down the expression for the probabilities needed in Eq. (2):

$$P(\mathcal{LFS}_j \mid \text{bin}_i) = \frac{P(\text{bin}_i, \mathcal{LFS}_j)}{P(\text{bin}_i)}. \qquad (7)$$

Therefore, we can find the entropy associated with child node bin_i,

$$H(\text{bin}_i) = -\sum_{j=1}^{L} P(\mathcal{LFS}_j \mid \text{bin}_i) \log(P(\mathcal{LFS}_j \mid \text{bin}_i)). \qquad (8)$$

These entropies are weighted by the probability $P(\text{bin}_i)$, which is given by Eq. (6). The summation of these weighted entropies forms an average entropy for all child nodes of node N,

$$E(N, \text{test for } a_k) = \sum_{i=1}^{\#\text{children}} P(\text{bin}_i)H(\text{bin}_i). \qquad (9)$$

We continue growing the tree until one of the following stopping criteria is met:

1. The reduction in entropy is below a preset level (i.e., if $E(n, \text{test}) - H(n) < \text{Threshold}_{\text{Zero difference}}$). In our work, when the reduction in entropy is less than 5%, we do not split the node.

2. The distributions of all $\mathcal{LFS}$ classes occupying the node bin are completely overlapping. This means that this leaf node will point to more than one class.

FIG. 19. $C_{j,L}^i$ stands for the left truncated point for $\mathcal{LFS}_j$ along attribute axis a_i, $C_{j,R}^i$ is the right truncated point. Candidate thresholds are in the intervals $[C_{2,L}^1, C_{1,R}^1]$ and $[C_{3,L}^1, C_{2,R}^1]$ along a_1 axis and $[C_{2,L}^2, C_{1,R}^2]$ and $[C_{2,R}^2, C_{3,L}^2]$ along a_2 axis.
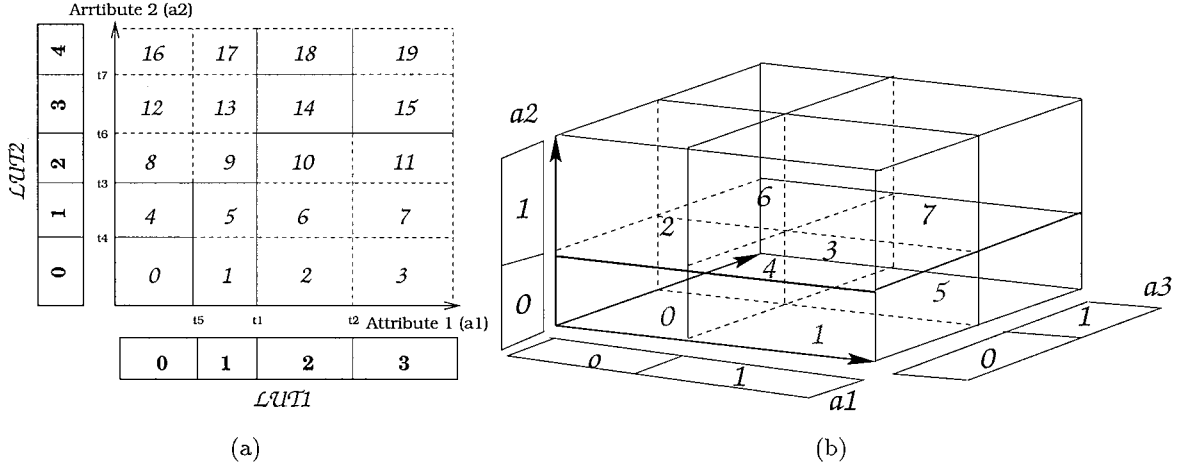
FIG. 20. (a) 2D hash table and the corresponding LUT. (b) 3D hash table and the corresponding LUT.

4.2. Translating a Decision Tree to a Hash Table

As was illustrated in Fig. 18b, the partitions created by the decision tree do not span the entire attribute space. In that figure, only those bin boundaries that correspond to the solid lines were created by the decision tree. However, it is trivial to make a hash table from these bin partitions by simply extending them to the end of the attribute space, i.e., to the maximum values observed in the data.

Speaking more specifically, we associate an LUT—a lookup table—with each attribute axis, which is simply a list of all decision thresholds along that axis. For example, the LUT for the attribute a_1 for the example in Fig. 20a is the ordered list (t5, t1, t2) and the LUT for the attribute a_2 is the ordered list (t4, t3, t6, t7). An LUT for the attribute a_i helps us determine in constant time into which bin an arbitrary value of a_i would fall.

LUTs allow us to associate a hash index with each bin in the attribute space. For example, for the two-dimensional case shown in Fig. 20a, we can write

$$\text{HashIndex}(a_1, a_2)$$
$$= \text{LUT}_2(a_2) \times \text{Number\_of\_bins\_along\_}a_1 + \text{LUT}_1(a_1), \quad (10)$$

where $\text{LUT}_i(a_i)$ returns the bin index for that value of a_i. For example, if the value of a_1 were to fall between the thresholds t5 and t1 along a_1 axis in Fig. 20a, the $\text{LUT}_1(a_1)$ for that a_1 would equal 1.

Shown in Fig. 20b is a three-attribute hash table and its indices. Along the same lines as above, the general formula for indexing the hash table is given by

$$\text{HashIndex}(a_1, \ldots, a_k)$$
$$= \sum_{i=k}^{2} \left(\text{LUT}_i(a_i) \times \prod_{j=1}^{i-1} \text{Number\_of\_bins\_along\_}a_j \right) + \text{LUT}_1(a_1). \quad (11)$$

After the structure of the hash table has been set up, the bins are filled with pointers to the $\mathcal{LFS}$ classes whose distributions occupy them.

4.3. Pruning the Set of Retrieved $\mathcal{LFS}$ Classes and Candidate Images

During retrieval, the measured attribute values are converted into a hash index in the manner described above. This index is used to access the corresponding bin and thus to retrieve the $\mathcal{LFS}$ classes that populate that bin. As was mentioned in the introduction to Section 4, on average this process results in the retrieval of 2.63 $\mathcal{LFS}$ classes for each hash table access. This list of retrieved $\mathcal{LFS}$ classes is pruned by comparing the lobular region labels of the query LFS with those of the retrieved $\mathcal{LFS}$ classes. Comparison of the lobular region labels pays due regard to the tree structures shown in Fig. 15. For example, the query image shown on the left in Fig. 21 yields the LFS's:

$$\text{LFS}_1 = \{\text{RO\_TL} \mid \text{PBR}_{\text{adjacent}} = \text{null}, \text{PBR}_{\text{interior}} = \text{YES}\}$$

$$\text{LFS}_2 = \{\text{RO\_LL} \mid \text{PBR}_{\text{adjacent}} = \text{null}, \text{PBR}_{\text{interior}} = \text{YES}\}$$

$$\text{LFS}_3 = \{\text{LO\_LL} \mid \text{PBR}_{\text{adjacent}} = \text{null}, \text{PBR}_{\text{interior}} = \text{YES}\}.$$

In Fig. 15, the lobular region label RO_TL of LFS$_1$ has the label RL as its parent. This means that only those retrieved $\mathcal{LFS}$ classes whose lobular region lables belong to the set {RO_TL, RL} can be matched with this query image LFS. Similar considerations would be applied to the other two LFS's in the query image. This type of pruning constitutes Filter 1 in Fig. 21.

After the above-described pruning by Filter 1, the remaining retrieved $\mathcal{LFS}$ classes for this query point to 25 database images on the average. We refer to these images as the candidate image set from which the four best images are selected for display. For this selection, we carry out a simple Euclidean comparison of the query PBR to its corresponding PBR in the candidate image set. This process is depicted as Filter 2 in Fig. 21.
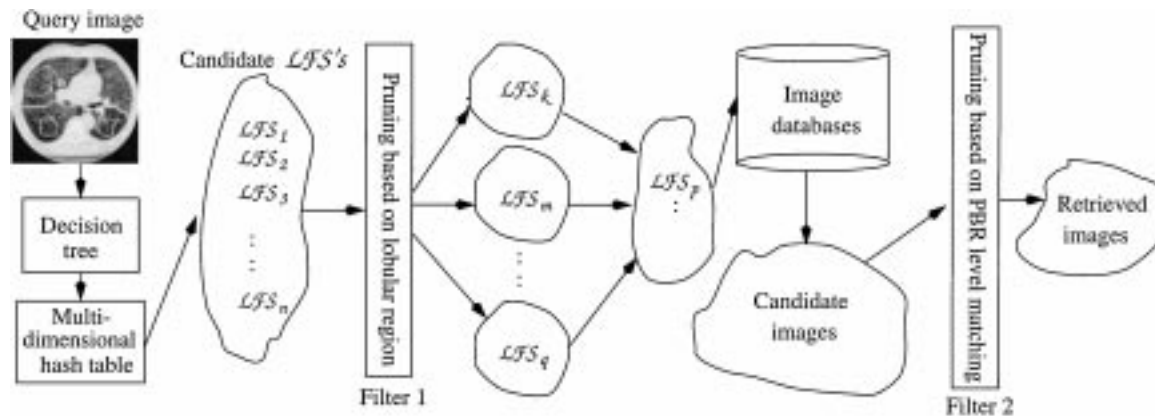
FIG. 21. Pruning of the retrieved $\mathcal{LFS}$ classes on the basis of lobular region labels constitutes Filter 1 shown on the left. This is followed by the application of Filter 2 on the right in which a Euclidean metric is used to select the final four images for display.

5. EXPERIMENTAL RESULTS

Recall and precision are two criteria that are frequently used to evaluate CBIR systems. For a query image of a given label, the recall criterion measures the proportion of all of the images in the database possessing the same label that are returned in response to the query. The precision criterion, on the other hand, measures the proportion of the returned images that have the same label as the query image. As is clear from the formulas given by Smith [25], both the recall and the precision depend on the number of the database images that the retrieval system must return in response to a query image. When a CBIR system is required to return a small, fixed number of images—exactly four images in our case—we believe that the recall criterion ceases to be an important measure of the effectiveness of retrieval.

In the rest of this section, therefore, we will only use the precision criterion to measure the effectiveness of retrieval in ASSERT. In particular, we will show how precision depends on whether we use localized or global features for image characterization. We will report results on the sensitivity of precision to subjectivity in PBR delineation. We will end this section by discussing the computational efficiency achieved with Multi-Hash indexing.

The database used in the experiments described here contains 302 images from 78 patients. These images were gathered by the physicians during routine medical care at Indiana University Medical Center. As would be expected, the disease labels assigned to each image by the physician play an important role in the evaluation of the system. The image labels themselves play no part during retrieval since the retrieval is carried out solely on the basis of image content. However, these labels are critical for measuring the precision rates. The distribution of images over disease patterns of the current database is shown in first column of Table 3. Currently, the diseases in the database are centrilobular emphysema (CLE), paraseptal emphysema (PSE), bronchiolitis obliterans (BO), hemorrhage (HE), metastatic cal-

cification (MC), panacinar (PA), sarcoid (SA), and pneumocystis carinii pneumonia (PCP).

To ensure a situation that would mirror its use in a clinical setting, in our evaluation we omit all of the query-image patient's other images from the database search. Our statistics were generated from the four highest ranking images returned by the system for each query. Moreover, a retrieval is considered successful only when the retrieved images have disease labels identical to those of the query image. (Note that an image can have more than one type of pathology). For example, if a query image has CLE and PSE, the retrieved images are classified as correct retrievals only when they bear both the CLE and the PSE labels. Although it might still be useful to a physician to retrieve images with only one of the pathologies, for the purpose of system evaluation we would like to make our evaluation as stringent as possible.

5.1. Local versus Global Attributes

This experiment is designed to test the utility of localized attributes. For each disease category in our database, in Table 3 we show the total number of queries for the category, the mean

TABLE 3
Comparison of Localized versus Global Attributes

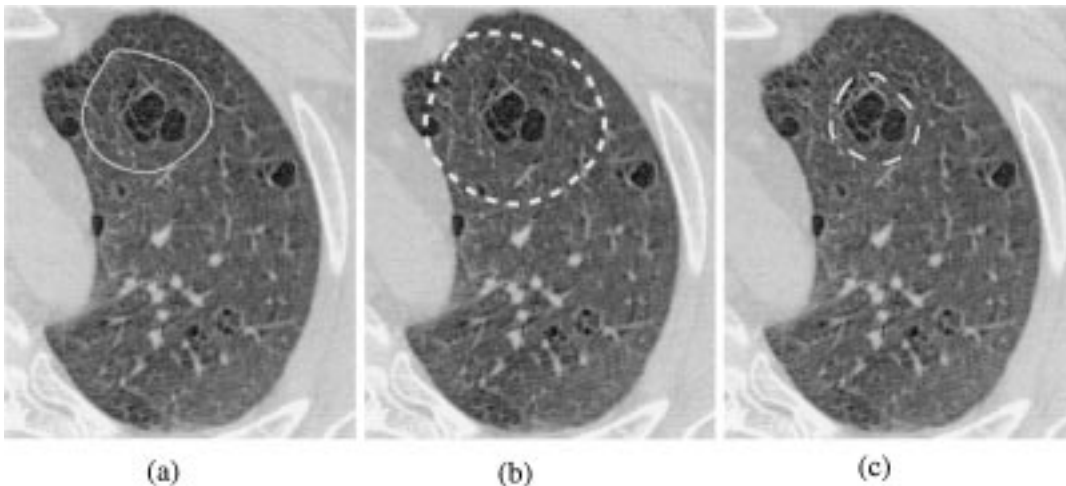| Diagnosis	Query images	Correct retrievals		Percentage of total	
		$R_1(P) + C$	$R_2(G)$	$R_1(P) + C$	$R_2(G)$
CLE	168	2.92 ± 0.18	2.12 ± 0.85	73	53
PSE	29	3.04 ± 0.27	1.68 ± 1.07	76	42
BO	28	3.00 ± 0.32	2.32 ± 0.55	75	58
HE	18	2.92 ± 0.53	2.40 ± 0.33	73	60
MC	12	2.64 ± 1.02	2.48 ± 1.21	66	62
PA	16	2.80 ± 1.08	2.32 ± 1.32	70	58
PCP	15	2.48 ± 0.85	2.40 ± 0.23	62	60
SA	16	2.76 ± 0.71	1.96 ± 0.75	69	49
Total DB	302	2.89 ± 0.36	2.14 ± 0.82	72.3	53.6

FIG. 22. (a) The PBR delineated by a physician; (b) $1.5 \times$ PBR; (c) $0.5 \times$ PBR.

and standard deviation of the number of the four highest ranking images that shared the same diagnoses as the query image, and the percentage of the four retrieved images that have the same diagnoses as the query image. Note that only when the diagnoses of the retrieved images are identical to the query image's, do we consider it a correct retrieval. Table 3 shows results for two different sets of attributes. The first is a combination of attributes extracted from the PBR region ($R_1(P)$) and attributes contrasting the PBR to the rest of the lung region (C). The second set of attributes was customized to a global approach to image characterization. The R_2 attributes were chosen by the SFS algorithm when optimizing performance for the entire lung region. To ensure a fair comparison, we omit from both R_1 and R_2 the attributes selected via the perceptual categories, discussed in Section 3.1, as these were developed to optimize performance of our localized approach. For this experiment we used the nearest-neighbor retrieval method after removing from the database the query-patient images.

The attributes in R_1 are: the gray scale deviation inside the region, gray-level histogram values inside the region, and four texture measurements (homogeneity, contrast, correlation, and cluster). The attributes in set C contrasting the PBR to the entire lung are: the area of the PBR, the Mahalanobis distance from the centroid of the PBR to the nearest lung boundary point, the difference of gray-scale mean of the PBR and that of the entire lung, and the difference of gray-scale deviation of the PBR and that of the entire lung. The attributes in set R_2 are: gray scale mean and deviation, histogram distribution, histogram distribution after gamma correction, and four texture measures (cluster, contrast after gamma, cluster after gamma, and edginess of strength after gamma). The last row of the table gives a summary across all diseases. The best method ($R_1(P) + C$) combines attributes of the PBR and contrast attributes. This empirical evaluation illustrates that local attributes significantly improve retrieval performance in the domain of HRCT of the lung.

5.2. Sensitivity to Physician Subjectivity

The second experiment addresses the concern that the precision might be a function of PBR delineation. Using the same experimental method as described in the previous section, we compared the retrieval results of the physician-marked PBRs to larger and smaller PBRs. Specifically, Table 4 reports results for 0.5, 0.75, 1.0, 1.25, and 1.5 times the size of the physician-entered PBRs. Figure 22 shows the original PBR region as marked by a physician, the same PBR but with an enlarged circumscribing circle so it is now occupying an area that is 50% larger, and the same PBR when its area is shrunk by 50%.

The results show that PBR size does not significantly impact retrieval results. Shrinking the PBR region has a slightly larger negative impact on performance than increasing the size of the PBR, the extent of this effect however seems to depend on the disease.

TABLE 4
Sensitivity of Results to PBR Delineation—Nearest Neighbor Approach (with Location Attribute)

Diagnosis	Number of queries	Percentage—scaling				
		0.5	0.75	1.0	1.25	1.5
CLE	168	69	72	**76**	74	72
PSE	29	72	72	**77**	76	75
BO	28	69	71	**79**	75	74
HE	18	65	68	**71**	73	71
MC	12	66	68	**68**	68	67
PA	16	72	74	**72**	79	66
PCP	15	59	62	**57**	63	66
SA	16	57	63	**72**	72	69
Total DB	302	67.9	70.6	**74.4**	73.6	71.4

TABLE 5
Filters 1, 2 vs Filter 2 and Nearest-Neighbor vs Nearest-Neighbor with Locatioon Attributes

Diagnosis	Image number	Nearest-Neighbor		Multi-Hash + nearest-neighbor			Multi-hash + nearest-neighbor + Filters		
		P_{Loc}	$P_{\text{no\_Loc}}$	P_{Loc}	$P_{\text{no\_Loc}}$	SPEEDUP	P_{Loc}	$P_{\text{no\_Loc}}$	SPEEDUP
CLE	168	46%	78%	37%	76%	6.56	71%	73%	9.15
PSE	29	40%	81%	45%	79%	6.29	74%	74%	16.77
BO	28	62%	82%	73%	86%	8.38	86%	84%	14.38
HE	18	41%	74%	27%	75%	9.43	69%	69%	25.16
MC	12	39%	63%	58%	72%	10.78	67%	67%	13.72
PA	16	44%	78%	38%	61%	13.13	63%	66%	11.62
PCP	15	51%	57%	67%	67%	17.76	73%	73%	37.75
SA	16	61%	69%	38%	58%	13.13	66%	66%	27.45
Total	302	47.3%	76.3%	42.9%	74.8%	7.58	71.8%	72.9%	11.61

5.3. Retrieval Accuracy and Efficiency

We will next demonstrate the speedup in retrieval achieved by using the multidimensional hashing of Section 4. The speedup is over the nearest-neighbor approach. The following measurements are relevant to estimating the retrieval speedup: $P_{\text{no\_Loc}}$ is the precision achieved with the retrieval of four best images whose PBRs have the same disease labels as that of the query image, regardless of the adjacency/interior attributes of these PBRs. P_{Loc} is the precision achieved for the four best matching images whose PBRs have the same disease labels and the same adjacency/interior attributes as the query PBRs. SPEEDUP is the speedup factor achieved and is defined as

$$\text{SPEEDUP} = \frac{\text{Total number of database images}}{\text{Number of candidate images}}, \quad (12)$$

where by "Number of candidate images" we mean the number of images returned by the Multi-Hash scheme of Section 4. Recall that the final four images selected for retrieval are chosen from this set of candidate images.

The third and the fourth columns of Table 5 show the retrieval accuracy obtained by applying the nearest-neighbor search. This evaluation is based on the matching of PBRs directly. The measurement P_{Loc} is obtained by first matching the lobular location of PBRs and then computing the similarity distances between the query PBR and the PBRs contained in the database. The distances contributed by the different PBRs in the query image are accumulated and ranked. The measurement $P_{\text{no\_Loc}}$ omits the step that matches the lobular location. P_{Loc} yields a 47.3% accuracy rate and $P_{\text{no\_Loc}}$ has 76.3%. Since nearest-neighbour searches through all images in the database, SPEEDUP is 1 and is not listed in Table 5

In the second set of results on retrieval speedup, for a given query image the system first generates a set of candidate database images using the hashing scheme of Section 4. The final four retrieved images are obtained by applying the nearest-neighbor algorithm to this candidate set. The measurement $P_{\text{no\_Loc}}$ corresponds to the retrieved four images vis-a-vis the candidate set

returned by the hashing table. We obtained a value of 74.8% for $P_{\text{no\_Loc}}$. However, if the measurement of retrieval precision includes the lobular location of the PBRs, the value of the precision, now denoted by P_{Loc} goes down to 42.9%. The SPEEDUP in this case was 7.58.

The third set of results that we will now report on corresponds to how ASSERT is supposed to work in a clinical setting. Now all the candidate images for a given query image are obtained by using the multidimensional hashing scheme of Section 4 and this list of candidate images is pruned by using the two filters described in the same section. The values of $P_{\text{no\_Loc}}$ and P_{Loc} are now 72.9% and 71.8%, respectively. These numbers are slightly less than the numbers for the purely nearest-neighbor based approach, but the decrease in precision is not appreciable at all, especially considering that the SPEEDUP achieved in this case was 11.61.

6. CONCLUSION

This paper described a physician-in-the-loop system for medical images. We believe that our system combines the best of what can be gleaned from a physician, without burdening him or her unduly, and what can be accomplished by a computer. We also demonstrated unequivocally that localized attributes significantly improve retrieval performance in the domain of HRCT images of the lung over a purely global approach. A sensitivity study showed that physician subjectivity in PBR delineation impacts performance by only a negligible amount.

While retrieval accuracy is of paramount importance in CBIR, not to be ignored is the retrieval efficiency, meaning the number of database images actually compared to a given query image. One of the important results in this paper is that retrieval efficiency—as measured by a speedup factor, SPEEDUP—improves considerably if we base the indexing of the database images on a hashing function derived from feature groupings called lobular feature sets.

Although the work reported here deals specifically with HRCT images of the lung, it could be extended easily to other domains, where image characterization can only be carried out

by using localized features. We have presented ASSERT as a stand-alone visual information retrieval system. However, it could also be integrated easily with most of the existing PACS (picture archiving and communication systems), HIS (hospital information systems), and RIS (radiology information systems). Other application possibilities for our system include telemedicine for long-distance diagnosis and education in medical schools.

APPENDIX A: LUNG EXTRACTION ALGORITHM

Because of the diagnostic information in measurements such as "the number of reticular cells adjacent to the lung boundary," and "the number of low opacity regions next to the lung boundary," it is important to be able to extract the lung regions from HRCT images. In this appendix, we present such an algorithm that successfully extracts the lungs for over 93% of the images in our database. For the small number of cases in which our algorithm fails, we plan to develop an approach that elicits the help of the physician by asking him/her to quickly mark a few points on the boundary of the lung. This information will then be fed into one of the energy-minimization algorithms known in the literature for the extraction of the lung boundary.

Before we describe the algorithm, the reader should note the following property of HRCT images. These images can be trivially divided into two regions: the main image that is inside a circle called the CT circle, and the remainder that is outside the circle. Since the pixels outside the circle do not get any values from the sensors that are used for the measurement of tomographic projections, their gray levels are equal to 0, a fact that can be used (in conjunction with connectivity properties) for quick elimination of all such pixels. On the other hand, inside the CT circle, where the main image is located, there are three types of regions: (a) the region composed of pixels exterior to the body; (b) the lung region; and (c) the region composed of pixels interior to the body but exterior to the lungs.

For extracting the lung regions from inside the CT circle, we employ an algorithm whose flow of control is depicted in Fig. 23. We will now explain the various steps of this algorithm. These steps correspond to the numbered circles attached to the various boxes in the figure. Each step will be illustrated by showing the resulting image in Fig. 24.

1. Apply a 5×5 median filter to remove noise inside the lung regions. Then apply a threshold to separate out the high-attenuation regions in the image. High gray levels correspond to the parts of the body with high X-ray attenuation. The threshold
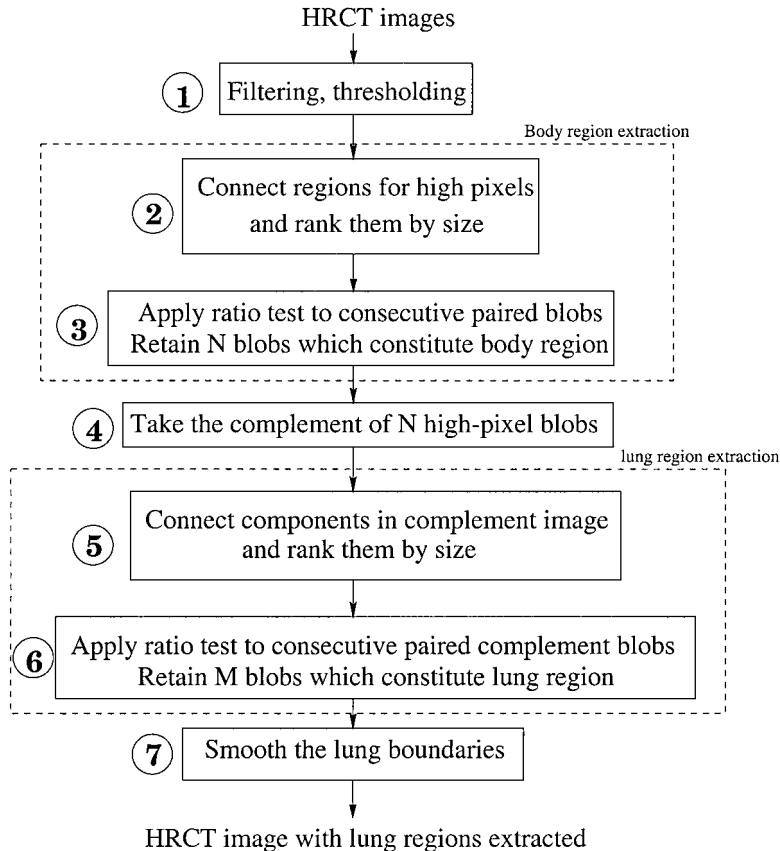
HRCT images

| 1 | Filtering, thresholding |

Body region extraction

| 2 | Connect regions for high pixels and rank them by size |

| 3 | Apply ratio test to consecutive paired blobs Retain N blobs which constitute body region |

| 4 | Take the complement of N high-pixel blobs |

lung region extraction

| 5 | Connect components in complement image and rank them by size |

| 6 | Apply ratio test to consecutive paired complement blobs Retain M blobs which constitute lung region |

| 7 | Smooth the lung boundaries |

HRCT image with lung regions extracted

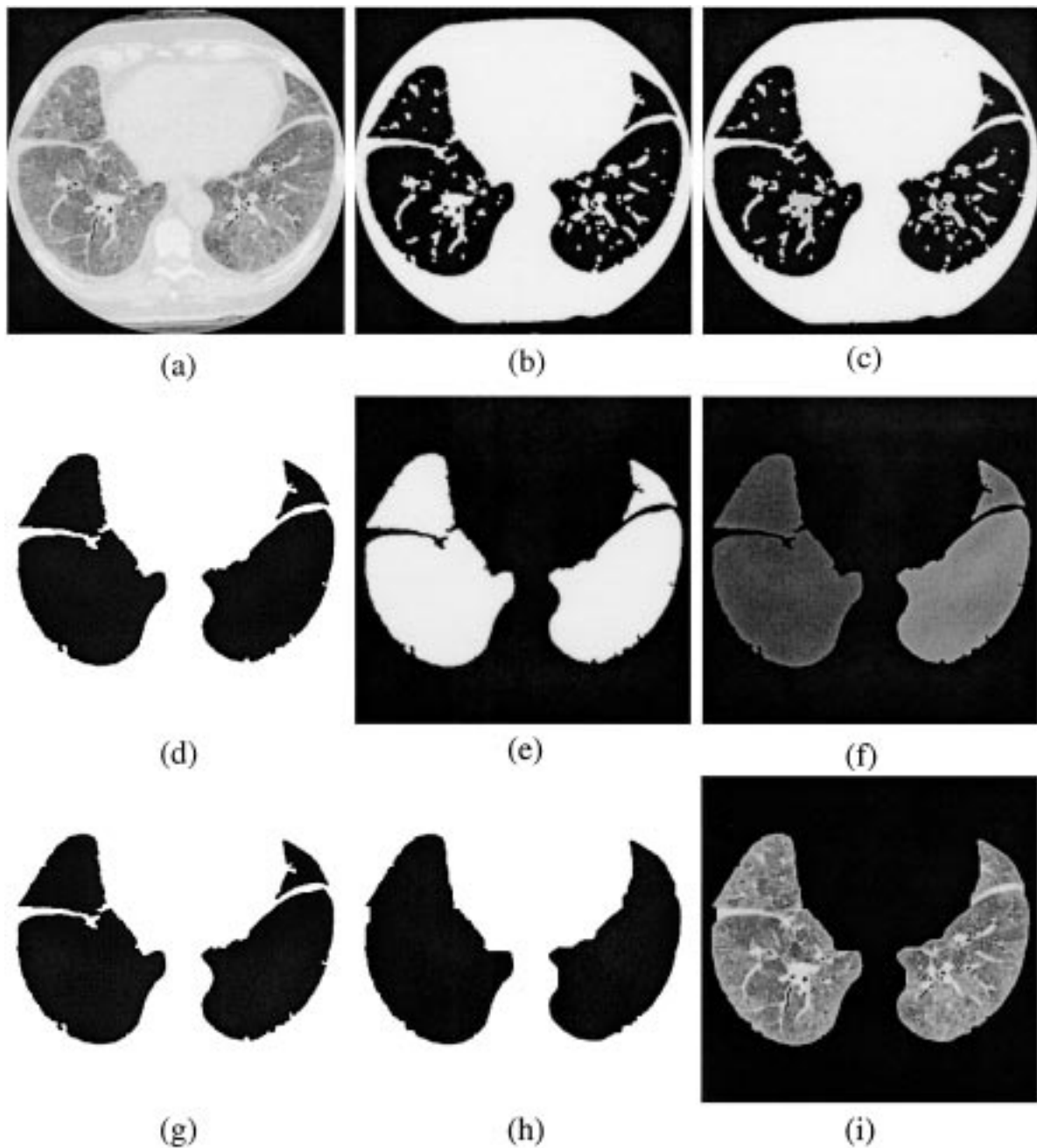**FIG. 23.** The flow chart for extracting the lung regions.

FIG. 24. The intermediate results from lung region extraction algorithm: (a) an HRCT image, (b) after thresholding to separate out high-attenuation pixels, (c) after component labeling, (d) after applying the ratio test, (e) complement of image in (d), (f) component labeling of the complement image, (g) after applying the ratio test to the complement image, (h) after the application of dilation and erosion operators (h) the final result which has only the lung regions.

is determined empirically (its value in our case is 190) and is based on our experience with over 300 images. All pixels whose values exceed the threshold are remapped to a value of 1 and the pixels that fall in the part of the histogram below the threshold are remapped to 0. This step creates a binary image, or, more precisely, a binary mask for the nonlung regions of the image inside the CT circle. For the HRCT image of Fig. 24a, the result obtained after thresholding is shown in Fig. 24b.

2. Apply a component labeling algorithm to the pixels labeled as 1 in the previous step. This step results in the identification of the separate connected regions of high-attenuation pixels. These connected regions correspond to the body pixels excluding the lung regions. In many but not all cases, the largest of these connected regions will surround the lung region. Using different gray levels, Fig. 24c shows the connected components of the high pixels shown in Fig. 24b. Some of these connected regions

may not be discernible, either because they are too small in size or because the gray levels that are automatically assigned to them are not sufficiently different from those of neighboring regions.

3. The body regions identified by the previous step are now ranked by size. Let these regions be denoted REG_1, REG_2, ..., REG_M in order of decreasing size. The algorithm now applies a ratio test to the consecutive pairs of these regions by testing the following inequality:

$$\frac{\text{size}(REG_i)}{\text{size}(REG_{i+1})} < T, \quad (13)$$

where T is a predetermined threshold. Let us say that there are N regions that satisfy this ratio test. These N regions are retained and the rest discarded, in the sense that, in the binary mask for the image, their pixels are changed to low. For the case of Fig. 24, this step retains only the component shown in Fig. 24d.

4. Next, the complement of the N body regions retained by the previous step is taken. This complement contains, but is not exclusively made of, the lung regions. For the example of Fig. 24, the result of this step is shown in Fig. 24e. To separate the lung regions from this complement, we proceed as follows.

5. Apply component labeling to the complement image. The separate components for the example case are shown in Fig. 24f using different shades of gray.

6. The separate regions produced by the previous step are ranked in the order of decreasing size and labeled reg_1, reg_2, ..., reg_J by ranking the sizes of those labeled regions in decreasing order. The algorithm now applies a ratio test to these consecutive pairs of regions:

$$\frac{\text{size}(reg_i)}{\text{size}(reg_{i+1})} < t, \quad (14)$$

where t is a predetermined threshold. The first M blobs that pass the ratio test are retained as lung regions and the rest are discarded. Shown in (g) are the regions retained. These M blobs constitute the lung regions.

7. Merge the M regions produced by the previous step using the dilation and erosion operators. These operators are needed because the regions corresponding to the different lobes of a lung may come out with gaps between them; these gaps are produced by the fissures. Figure 24h shows the merged lungs thus obtained.

8. The previous step produces a binary mask whose 1's correspond to the lung regions. This mask can be used to separate the original image—such as the image of Fig. 24a—into the lung regions and nonlung regions, as shown in Fig. 24i.

The 7% of the images for which this algorithm does not work had one or both of the following characteristics: (1) The images had only one of the two lungs and the size of the lung was not significantly larger than the sizes of the other low-attenuation artifacts in the image. (2) One of the lungs was very fragmented. As mentioned before, the lung regions in all such images will be extracted with the help of the physician.

APPENDIX B: LIST OF ABBREVIATIONS USED

CBIR	Content-based image retrieval
CT	Computed tomography
HRCT	High resolution computed tomography
HIS	Hospital information system
LFS	Instance of a lobular feature set
$\mathcal{LFS}$	Lobular feature set class
LO	Left oblique
LUT	Lookup table
MR	Magnetic resonance
PACS	Picture archiving and communication system
PBR	Pathology bearing regions
RO	Right oblique
RH	Right horizontal
SFS	Sequential forward selection.

ACKNOWLEDGMENTS

REFERENCES

1. J. L. Bentley and J. H. Friedman, Data structures for range searching, *ACM Comput. Surveys* **11**(4), 1979, 397–409.

2. R. C. Bolles and R. A. Cain, Recognizing and locating partially visible objects: The local-feature-focus method, *Intl. J. Robot. Res.* **1**(3), 1982, 57–82.

3. C. Chen and A. Kak, A robot vision system for recognizing 3-D objects in lower-order polynomial time, *IEEE Trans. Systems Man. Cybern.* **19**(6), 1989, 1535–1563.

4. 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.

5. D. Comer, The ubiquitous B-tree, *ACM Comput. Surveys* **11**(2), 1979.

6. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, 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.* September 1995, 23–32.

7. L. Grewe and A. C. Kak, Interactive learning of a multi-attribute hash table classifier for fast object recognition, *Comput. Vision Image Understand.* **61**(3), 1995, 387–416.

8. A. Guttman, R-Trees: A dynamic index structure for spatial searching, in *Proc. of ACM SIGMOD, June 1984*, pp. 47–57.

9. Y. Hara, K. Hirata, H. Takano, and S. Kawasaki, Hypermedia navigation and content-based retrieval for distributed multimedia databases, in *Proc. of the 6th NEC Research Symposium on Multimedia Computing, 1995*.

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

11. T. Y. Hou, A. Hsu, P. Liu, and M. Y. Chiu, A content-based indexing technique using relative geometry features, in *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases, 1992*, pp. 59–68.

12. 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.

13. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, *Int. J. Comput. Vision* **1**(4), 1998, 321–331.

14. P. M. Kelly, T. M. Cannon, and D. R. Hush, Query by image example: The CANDID approach, in *SPIE Vol. 2420 Storage and Retrieval for Image and Video Databases III, 1995*, pp. 238–248.

15. J. Kittler, Feature set search algorithms, in *Pattern Recognition and Signal Processing* (C. H. Chen, Ed.), pp. 41–60, Sijthoff & Noordhoff, Alphen ann den Rijn, The Netherlands, 1978.

16. 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.

17. N. Otsu, A threshold selection method from gray-level histogram, *IEEE Trans. Systems, Man, Cybernetics* **SMC-9**(1), 1979, 62–66.

18. A. Pentland, R. W. Picard, and S. Sclaroff, Photobook: Tools for content-based manipulation of image databases, in *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases, 1994*, pp. 34–47.

19. K. Rahardja and A. Kosaka, Vision-based bin-picking: Recognition and localization of multiple complex objects using simple visual cues, in *1996 IEEE/RSJ International Conference on Intelligent Robots and Systems, Osaka, Japan, November, 1996*, pp. 1448–1457.

20. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

21. N. Roussopoulos and D. Leifker, Direct spatial search on pictorial databases using packed R-trees, in *Proc. ACM SIGMOD, May 1985*.

22. R. Samadani, C. Han, and L. K. Katragadda, Content-based event selection from satellite image of the aurora, in *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases, 1993*, pp. 50–59.

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

24. T. Sellis, N. Roussopoulos, and C. Faloutsos. The R^+-tree: A dynamic index for multidimensional objects, in *Proc. of 13th Internal Conference on VLDB, England, September 1987*, pp. 507–518.

25. J. R. Smith, Image retrieval evaluation, in *Proc. IEEE Workshop of Content-Based Access of Image and Video Databases, Santa Barbara, CA, June 1998*, pp. 112–113.

26. C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. Aisen, and L. Broderick, Local versus global features for content-based image retrieval, in *Proc. IEEE Workshop of Content-Based Access of Image and Video Databases, Santa Barbara, CA, June 1998*, pp. 30–34.

27. H. S. Stone and C. S. Li, Image matching by means of intensity and texture matching in the fourier domain, in *Proc. SPIE Conf. in Image and Video Databases, San Jose, CA, Jan. 1996*, pp. 337–349.

28. W. R. Webb, N. L. Muller, and D. P. Naidich, *High-Resolution CT of The Lung*, 2nd ed., Lippincott-Raven, Philadelphia, 1996.