

## Local versus Global Features for Content-Based Image Retrieval\*

C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka  
{chiren, brodley, kak, kosaka}@ecn.purdue.edu  
School of Electrical and Computer Engineering  
Purdue University  
West Lafayette, IN 47907

A. Aisen, and L. Broderick  
{aaisen, lbrod}@iupui.edu  
Department of Radiology  
Indiana University Medical Center  
Indianapolis, IN 46202

### Abstract

*It is now recognized in many domains that content-based image retrieval (CBIR) 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 of the image at the time the image is entered into the database. From the regions thus marked, our approach applies low-level computer vision and image processing algorithms to extract features related to the variations of gray scale, texture, shape, etc. The extracted features create an index that characterizes the image. To form an image-based query the physician first marks the PBR's. The system then extracts the relevant image features, computes the distance of the query image to all image indices in the database, and retrieves the  $n$  most similar images. Our approach is based on the assumption that medical image characterization must contain features local to the PBR's. The focus of this paper is to assess the utility of localized versus global features for the domain of HRCT images of the lung, and to evaluate the system's sensitivity to physician subjectivity in delineating the PBR's.*

**Keywords:** Image characterization, medical images, human-computer interaction, evaluation.

### 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 [10, 7, 6, 4, 12]. For example, the QBIC system [2] characterizes images using global characteristics such as color histogram, texture values, shape parameters of easily segmentable regions, etc. For many databases, 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 an image, the localization being with respect to certain anatomical landmarks. For example, for high-resolution computed tomographic (HRCT) images of the lung, a disease such as emphysema (shown in Figure 3) manifests itself in the form of a low-attenuation region that is textured differently from the rest of the lung. Local features are needed for such situations because the number of pathology bearing pixels in an image is small relative to the rest of the pixels and any global signature would not be sufficiently impacted to serve as a useful attribute for image retrieval. Currently it is not possible to extract these regions by automatic segmentation routines – the regions of pathology in medical images often do not possess sharp edges and contours that can be extracted automatically. Our system, therefore, enlists the help of the physician for delineating the PBR's and any anatomical markers that might be relevant.

The focus of this paper is to assess the utility of localized versus global features for the domain of HRCT images. We will present an empirical evaluation of the implementation illustrating that local features significantly improve performance over using only global features. In addition, the evaluation will address a possible limitation of our approach: because PBR delineation is inherently subjective, the quality of the results could depend on a physician's ability to accurately circumscribe the PBR's. We present a sensitivity study that shows that physician subjectivity has little impact on retrieval performance. Before presenting the results of our study, we review the reasoning architecture of ASSERT (Automatic Search and Selection Engine with Retrieval Tools), the implementation

---

\*This work is supported by National Science Foundation under Grant NO. IRI9711535 and the Showalter Fund. Copyright 1998 IEEE. Published in the Proceedings of CBAIVL'98, 21 June 1998 in Santa Barbara, California. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: Manager, Copyrights and Permissions / IEEE Service Center / 445 Hoes Lane / P.O. Box 1331 / Piscataway, NJ 08855-1331, USA. Telephone: + Intl. 732-562-3966.

of our approach.

## 2 ASSERT

The reasoning and control architecture of our approach is best explained with the help of the flow chart shown in Figure 1. The figure shows two phases: the image archiving phase is depicted in bold flow links and the retrieval phase is depicted using thin flow links.

To archive an image in the database, a physician delineates the PBR's and anatomical landmarks. This interaction takes at most a minute for a well-trained domain expert (a radiologist). The system then executes the computer vision and image processing algorithms to create a feature vector that characterizes the image. To facilitate accurate indexing and retrieval it is of critical importance that the images in the database be characterized with relevant features. To this end we have collected a set of images from different disease patterns for which the diagnosis is known. After we have extracted the set of possibly relevant features (described in Section 3), we perform a sequential forward selection search [1] to reduce the dimensionality of the feature space while retaining the ability to accurately classify each image as belonging to its associated disease pattern. The resulting feature vector forms an index into the database. Currently, retrieval is done using a nearest neighbor approach in the reduced feature space. Our plan for the immediate future is to use a decision tree structure to create a multiple-attribute hash table [3]. As the database grows, we will periodically re-run the index archiving algorithms to ensure that the features and the decision tree selected are optimized for the current database.

The computations for retrieval proceed in the same manner as image archival. The physician brings up a query image on the screen and then requests the  $n$  most visually similar images from the database. After the physician delineates the PBR's in the query image, the system applies the lung region extraction algorithm and executes low-level computer vision and image processing procedures to extract image features from both PBR's and lung regions. The system then computes the distance of the query image to all images in the database and returns the  $n$  most similar images. The physician can then view the associated differential diagnoses of each returned image.

Figure 2 shows ASSERT's user interface. The large left frame at the top is used for displaying the query image. The large frame immediately to the right of the query image is reserved for displaying one of the retrieved images; it can also be used for displaying the image obtained after an image processing algorithm is applied to the query image. The best four retrieved images are displayed in a row at the bottom. Any of these images can be shown in a larger format immediately to the right of the query image. In Figure 2, the image shown to the right of the query image is the magnified version of the second best retrieved image (the second from the left in the bottom row of images). On the far right of the interface screen, a user can enter feedback about the query images; the feedback can range from strong agreement with re-

trieval results to strong disagreement. Currently, the user feedback is used only for system evaluation. In the future, we plan to use the feedback to improve the indexing scheme automatically.

## 3 Image Characterization

To characterize each image, the system computes features that are local to the PBR's and features that are global to the entire lung region.<sup>1</sup> The PBR's are characterized by a set of shape, texture and other gray-level attributes. For characterizing texture within PBR's, we have implemented a statistical approach based on the notion of a gray-level co-occurrence matrix [5]. 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 parameters we extract from this matrix are *energy*, *entropy*, *homogeneity*, *contrast*, *correlation*, and *cluster tendency*. In addition to the texture-related features, we compute three additional sets of features on the pixels within the PBR boundary. The first set computes measures of gray-scale of the pathological region, specifically, the mean and standard deviation of the region, a histogram of the local region, and 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 [9]. 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 the PBR, we apply a region-based segmenter [8]. From the results we compute the number of segmented regions per area and histograms of the area and gray-levels of the segmented regions.

In addition to the texture and shape features, a PBR is also characterized by its average properties, such as gray scale mean and deviation, with respect to the pixels corresponding to 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 features we mentioned earlier). To extract the lung region, we apply a set of binary-image analysis routines [11]. When applied to the current database, the algorithm was able to successfully extract the lung region from the HRCT image 93% of the time.<sup>2</sup> In addition to the average-type features, the system also calculates the distance between the centroid of a marked PBR and the nearest lung boundary point; physicians use

<sup>1</sup>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 features are global only to the extent that they are based on all the pixels in the entire lung region.

<sup>2</sup>For the 7% of the images for which the algorithm does not extract the lung region, currently we ask the physician to circumscribe it. We are working on improving the accuracy of this step.

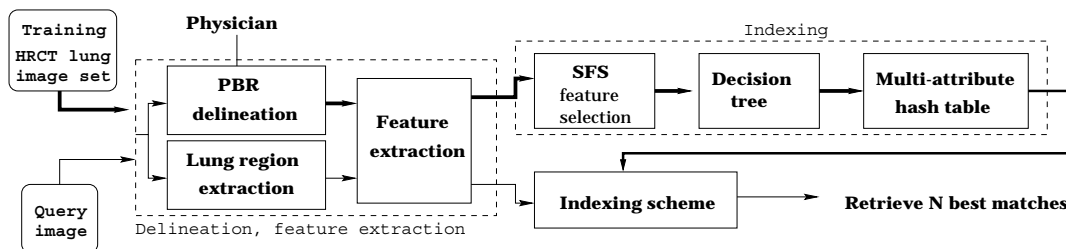


Figure 1: The flow chart for database archival and retrieval in ASSERT.



Figure 2: The user interface.

this information to classify some pulmonary disease patterns.

The total number of features, 255 in number, computed for a PBR is large (details of the set of features can be found in [11]). While this gives us an exhaustive characterization of a PBR (an intentional aspect of our design), for obvious reasons only a small subset of these features can be used for database indexing and retrieval. The features actually used are found by applying the sequential forward search (SFS) algorithm to all the 255 features.

#### 4 Empirical Evaluation

Ultimately the true test of a CBIR system is whether it is used by practitioners. To measure whether such a system would be useful, evaluation of an information retrieval system is done by measuring the recall and the precision of the queries. Recall is the proportion of relevant materials retrieved. Precision quantifies the proportion of the retrieved materials that is relevant to the query.

In our approach, the precision and recall are functions of 1) the feature vector used to characterize the

images, 2) the retrieval scheme and 3) the delineation of the PBR by the physician. In this evaluation we hold the second factor constant and investigate the impact on performance of the first and third factors. In addition to a clinical evaluation performed by a physician, we present off-line experiments designed to examine the utility of localized versus global features and the system's sensitivity to subjectivity in PBR delineation.

The database used in this experiment contains 345 pathology regions in 200 images from 60 patients. These images were identified by physicians during routine medical care at Indiana University Medical Center. To evaluate our CBIR system we use the differential diagnosis associated with each image. *Note that this information is not used during retrieval.* The distribution of images over disease patterns of the current database is shown in Table 1. Currently, the diseases in the database are centrallobular emphysema (CLE), paraseptal emphysema (PSE), invasive aspergillosis (ASP), broncheitis (BR), eosinophilic granuloma (EG), and idiopathic pulmonary fibrosis (IPF).

Table 1: Comparison of localized versus global features.

Diagnosis	PBR Queries	Correct Retrievals				Percent of Total			
		$R_1(P) + C$	$R_1(G)$	$R_2(P)$	$R_2(G)$	$R_1(P) + C$	$R_1(G)$	$R_2(P)$	$R_2(G)$
CLE:	256	$3.04 \pm 0.31$	$1.72 \pm 0.66$	$2.60 \pm 0.36$	$2.04 \pm 0.19$	76	43	65	51
PSE:	26	$3.28 \pm 0.12$	$2.00 \pm 0.40$	$1.68 \pm 1.08$	$1.20 \pm 0.63$	82	50	42	30
ASP:	5	$3.00 \pm 1.20$	$2.40 \pm 1.18$	$0.00 \pm 0.00$	$2.40 \pm 0.20$	75	60	0	60
BR:	12	$2.84 \pm 0.22$	$2.32 \pm 0.35$	$1.88 \pm 0.07$	$2.52 \pm 0.23$	71	58	47	63
EG:	37	$2.80 \pm 1.05$	$2.48 \pm 0.87$	$1.60 \pm 0.15$	$2.48 \pm 0.17$	70	62	40	62
IPF:	9	$2.88 \pm 1.00$	$1.76 \pm 0.21$	$1.32 \pm 0.11$	$2.20 \pm 0.09$	72	44	33	55
Total DB	345	$3.02 \pm 0.40$	$1.85 \pm 0.65$	$2.33 \pm 0.37$	$2.05 \pm 0.22$	75	46	58	51

#### 4.1 Localized versus Global Image Characterization

Our first experiment is designed to test the utility of localized features. To ensure a situation that would mirror its use in a clinical setting, 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. For each disease category in our database, we show the total number of queries for the category, the mean and standard deviation of the number of the four highest ranking images that shared the same diagnosis as the query image, and percentage of the four retrieved images that have the same diagnosis as the query image. Note that in these experiments we consider each PBR as a query rather than each image. Although our ultimate goal is to allow multiple PBR’s and relevant anatomical markers to form a query, at present our implementation restricts a query to a single PBR. Table 1 shows results for four different sets of features. The first is a combination of features extracted from the PBR region ( $R_1(P)$ ) and features contrasting the PBR to the rest of the lung region ( $C$ ). The second uses the same features as the first, but applied to the entire lung region ( $R_1(G)$ ). Because the entire lung is used, the contrast features are not included. The third set of features was customized to a global approach to image characterization. Because  $R_1$  features were selected to maximize performance when PBR’s were used, we ran the SFS algorithm using features computed from the entire lung region, producing set  $R_2$ . The table shows results for this feature set on the entire lung ( $R_2(G)$ ) and on PBR’s only ( $R_2(P)$ ).

The features 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 features in set  $C$  contrasting the PBR to the entire lung are: the area of the PBR, the Mahalanobis distance from the centroid of PBR to the nearest lung boundary point, the difference of gray-scale mean of the PBR and the entire lung, and the difference of gray-scale deviation of the PBR and the entire lung. The features in set  $R_2$  are: gray scale mean and deviation, histogram distribution, histogram distribution after gamma correction,<sup>3</sup> and

<sup>3</sup>Typically, one can observe non-linear distortion of gray in-

Table 2: Sensitivity of results to PBR delineation – Precision

Diagnosis	Number of Queries	Percent of Total				
		0.5	0.75	1.0	1.25	1.5
CLE	256	69	75	<b>76</b>	71	73
PSE	26	69	73	<b>82</b>	82	84
ASP	5	40	55	<b>75</b>	65	60
BR	12	62	67	<b>71</b>	73	73
EG	37	62	68	<b>70</b>	74	75
IPF	9	72	70	<b>72</b>	72	78
Total DB	345	68	73	<b>75</b>	72	74

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 features of the PBR and contrast features. A comparison of  $R_2(P)$  to  $R_2(G)$  (2.33 versus 2.05) illustrates that *even when the features are customized to a global approach* it is better to compute them for the PBR. The results show that local features significantly improve performance over global features.

#### 4.2 Sensitivity of Retrieval to PBR Delineation

The second experiment addresses the concern that precision is a function of PBR delineation. Using the same experimental setup as the previous section, we compared the retrieval results of the physician marked PBR’s to larger and smaller PBR’s. Specifically, Table 2 reports results for 0.5, 0.75, 1.0, 1.25 and 1.5 times the size of the physician entered PBR’s. Figure 3 shows the original PBR region, and the PBR’s when increased by 50% and shrunk by 50%.

The results show that PBR size does not significantly impact retrieval results on the database as a whole. Shrinking the PBR region has a slightly larger negative impact on performance than increasing the

tensity from slice to slice on the same patient and images taken from different CT scanners will have different distributions of gray scale. Applying the gamma correction allows us to mitigate these problems and facilitates the use of gray-scale histogram features for retrieval.

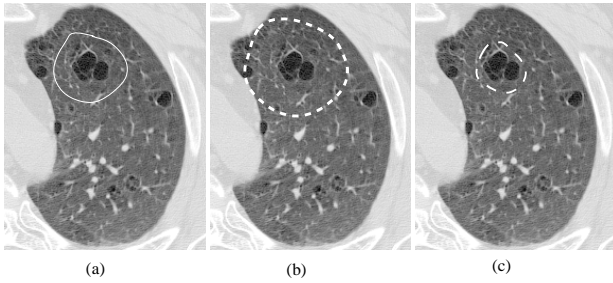


Figure 3: (a) The PBR delineated by a physician. (b)  $1.5 \times$  PBR (c)  $0.5 \times$  PBR.

Table 3: Retrieval results from clinical experiments (SA = strongly agree; A = agree; N = not sure; D = disagree; and SD= strongly disagree)

Disease	Number of Queries	Percent of Total				
		SA	A	N	D	SD
CLE	67	68	20	2	7	3
PSE	14	84	9	2	0	5
BR	8	46	22	12	3	17
IPF	4	12	25	25	25	13

size of the PBR, the extent to which appears to depend on the particular disease.

### 4.3 Clinical Experimental Results

In Table 3 we show results from a clinical trial of the system based on the best method ( $R_1(P)+C$ ). We collected results from an expert in pulmonary disease (lung expert). The retrieval results for CLE, PSE, BR and IPF<sup>4</sup> are classified into five categories specifying the amount to which the expert agreed with the results. For CLE, PSE and BR the majority of the retrieved queries (88%, 93% and 68%, respectively) are in the strongly agree to agree category. The results for IPF are not as compelling, but are based on a much smaller sample.

## 5 Conclusions and Future Work

In this paper, we have 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. An empirical evaluation of the current implementation illustrates that local features significantly improve retrieval performance in the domain of HRCT of the lung. A sensitivity study shows that subjectivity in PBR delineation impacts performance by a negligible amount.

Our plans for the future include adding more images to the database both from IUMC and from other sources, incorporating user feedback into the design

<sup>4</sup>At the time of the evaluation, we did not have any ASP or EG images in the database.

of the indexing scheme, increasing the speed of retrieval by implementing the multihash decision tree algorithm, and allowing multiple PBR's in a single query.

## References

- [1] R. O. Duda & P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley & Sons, NY, 1973.
- [2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, & P. Yanker, "Query by image and video content: The QBIC system", *IEEE Computer*, pp. 23-32, September 1995.
- [3] Lynne Grewe and A. C. Kak, "Interactive learning of a multi-attribute hash table classifier for fast object recognition", *Computer Vision and Image Understanding*, 61(3), pp. 387-416, 1995.
- [4] Y. Hara, K. Hirata, H. Takano, and S. Kawasaki, "Hypermedia navigation and content-based retrieval for distributed multimedia databases", *Proc. of the 6th NEC Research Symposium on Multimedia Computing*, 1995.
- [5] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley, 1992.
- [6] 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*, pp. 238-248, 1995.
- [7] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Tools for content-based manipulation of image databases", *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, pp. 34-47, 1994.
- [8] 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.
- [9] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, 1982.
- [10] R. Samadani, C. Han, and L. K. Katragadda, "Content-based event selection from satellite image of the aurora", *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, pp. 50-59, 1993.
- [11] C. R. Shyu, "Human-in-the-loop Content-based Image Retrieval from Large Scale Image Databases", *Technical Report, School of Electrical and Computer Engineering, Purdue University*, 1997.
- [12] H. S. Stone and C. S. Li, "Image matching by means of intensity and texture matching in the fourier domain", *Proc. SPIE Conf. in Image and Video Databases*, San Jose, CA, Jan. 1996.