

# A Content-based Image Retrieval Framework for Multi-Modality Lung Images

Yang Song<sup>1</sup>, Weidong Cai<sup>1</sup>, Stefan Eberl<sup>1,2</sup>, Michael J Fulham<sup>1,2,3</sup>, Dagan Feng<sup>1,4</sup>

<sup>1</sup>Biomedical and Multimedia Information Technology (BMIT) Research Group,  
School of Information Technologies, University of Sydney, Australia

<sup>2</sup>Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

<sup>3</sup>Sydney Medical School, University of Sydney, Australia

<sup>4</sup>Center for Multimedia Signal Processing (CMSP), Department of Electronic &  
Information Engineering, Hong Kong Polytechnic University, Hong Kong

<sup>1</sup>{ysong,tomc,feng}@it.usyd.edu.au,<sup>2</sup>stefan@nucmed.rpa.cs.nsw.gov.au,<sup>3</sup>mfulham@med.usyd.edu.au

## Abstract

*This paper presents a framework for effective and fast content-based image retrieval for multi-modality PET-CT lung scans. PET-CT scans present significant advantages in tumor staging, but also place new challenges in computerized image analysis and retrieval. Our framework comprises 5 major components: lung field estimation, texture feature extraction, feature categorization, refinement using SVM, and similarity measure. Clinical data from lung cancer patients are used as case studies, and effective retrieval performance is demonstrated.*

## 1. Introduction

Advances in medical imaging instrumentation and software have produced a large increase in the number of individual images for each patient study. In this context methods that rely on text-based searches of these image databases are problematic; this has led to new techniques such as content-based image retrieval (CBIR) [6, 13]. There are several recent developments, which include: a machine learning and relevance feedback based framework with database pre-filtering [16]; getting the semantic gap on X-ray image retrieval with a learning based feature selection and result ranking [7]; a local feature classification based mechanism [15]; and a CBIR approach that uses interactive windowing for image enhancement and organ-specific gray level features [3].

The focus for our work is on images produced by PET-CT scanners. These devices combine a helical CT scanner and PET scanner and produce anatomical and functional patient information from the single scan. In a clinical context PET-CT using the PET tracer, <sup>18</sup>F-fluoro-deoxy-

glucose (FDG), is widely used in the staging of lung cancer and a variety of other malignancies. However, the image datasets that are produced are large and complex.

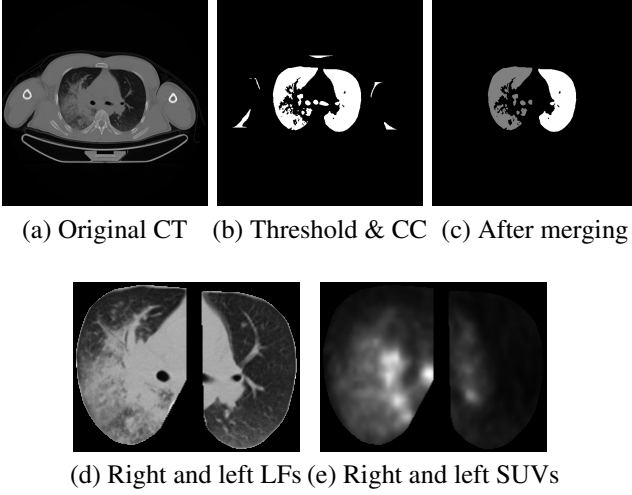
CT component has relatively poor soft tissue contrast resolution and thus there can be difficulty in separating lung tumors from normal surrounding tissues. FDG-PET images, however, do not define the exact tumor boundaries; and whilst most lung tumors display increased FDG uptake compared to normal tissue there can be high FDG uptake in non-cancerous tissue such as brown fat and inflammation. This image complexity of PET-CT renders the CBIR techniques listed above ineffective. Kim et al [11] reported on one approach but its retrieval performance relied heavily on segmentation accuracy and was limited to lung tumors only. Thus we propose a framework for PET-CT lung images retrieval using automatic lung field estimation, and grid-based texture features that are integrated with clustering and machine-learning techniques.

## 2. Methods

The overall framework comprises automatic lung field estimation, texture feature extractions, clustering-based feature space transformation, feature signature refinement with *support vector machines* (SVM), and a similarity measure for image retrieval. These elements are described below.

### 2.1. Automatic Lung Field Estimation

As a pre-processing step, we employed a fully-automatic fast estimation approach to extract left and right *lung fields* (LF). Rather than targeting precise segmentation [11], convex hulls for left and right lung were delineated. This design was to accommodate cases that lung pathologies could abut or extend into adjacent soft tissues (Figure 1a),



**Figure 1. Lung field estimation.**

where thresholding and boundary smoothing based techniques were inadequate [1, 8].

The estimation process started with histogram equalization of the CT image, to accentuate the edges between the lung fields and other tissues. Next, Otsu thresholding [14] was performed to obtain an initial mask, and patient bed and tissues outside of the lung were then removed with simple subtraction and filling operations. At this stage, the mask comprised a number of *connected components* (CC). The biggest two CCs depicted approximate contours of left and right LFs (Figure 1b). For cases when the left and right LFs were adjoined, they were separated by locating the linking junction.

Next, the adjacent small CCs were merged into the left or right LFs. Due to high density, the abnormal lung tissues caused disconnected small CCs during the thresholding step. A heuristic algorithm similar to  $k$  nearest neighbor ( $k$ NN), based on spatial relationships between the CCs, was utilized (Figure 1c). Finally, convex hulls of the left and right LF masks were created as a final safeguard to include any missing lung tissues (Figure 1d).

The above steps inevitably included some unnecessary surrounding tissues. However, all important areas such as the lung tumors, were encompassed in the LFs in this way. Such areas were the major determining factors for matching of similar images. The simplicity of computation ensured efficient execution as well. We also devised a solution for minimizing the disturbances caused by surrounding tissues (see 2.4 below).

Co-registered PET images were normalized to *standardized uptake values* (SUV) [9] and rescaled to match CT image size. The left and right LF masks were then mapped onto the SUV images to extract the corresponding region (Figure 1e), and the output was referred to as left and right SUVs.

## 2.2. Initial Grid-based Feature Extraction

In CT images, abnormal lung tissue is characterized by specific texture, and it is the occurrence of several such findings that helps to identify a specific diagnosis [20]. This suggests texture-based feature extraction is a reasonable choice, and it has been applied frequently [2, 10, 12]. Kakar and Olsen utilized Gabor filtering for its close resemblance to human’s visual processing [10]. PET images are traditionally analyzed on SUV values, with only a few studies on texture approaches [5, 9, 18].

We designed a *Gabor* filter based texture analysis to extract features from both LFs (CT) and SUVs (PET). A summary of the computation steps is as follows:

1. *Gabor* filters were applied to left and right LFs and SUVs. As Chen et al evaluated, the optimal parameters for *Gabor* filter were scale 6, orientation 4, filter mask size  $13 \times 13$ , totaling 24 filters [4]. This generated 24 filtered outputs for CT and 24 for PET.
2. Instead of keeping all 24 filtered outputs, we reduced the dimension by selecting only one  $I_{mv}$  that had the maximum variance. The visual distinctions among different patterns in the selected output were most prominent, and hence it enhanced the differentiability of features while preserving the integrity of the original image.

$$\{m_{mv}, n_{mv}\} = \underset{for\ m \in [0, \dots, 5],\ n \in [0, \dots, 3]}{\operatorname{argmax}}\ var(I_{Gabor}(m, n)),$$

$$I_{mv} = I_{Gabor}(m_{mv}, n_{mv})$$

3. The selected  $I_{mv}$  was evenly divided into 8-by-8 blocks, and texture features extracted for each block were:

$$F_{CT}(i, j) = \{mean, variance, entropy, location\}$$

$$F_{PET}(i, j) = \{mean, variance\}$$

where  $i$  and  $j$  were the row and column indices in the grid. The *mean*, *variance* and *entropy* assumed the standard definitions, and *location* was the percentage of pixels in the block falling within the left and right LF masks. We did not evaluate *entropy* and *location* for left and right SUVs, due to its low spatial resolution adding little value to such feature representation.

So, for each co-registered PET and CT image pair, we obtained a 6-dimensional feature grid  $F$  with individual feature vector as:

$$F(i, j) = \{F_{CT}(i, j), F_{PET}(i, j)\}$$

Both PET and CT images were padded initially to multiples of 8, and each feature element was further normalized to range  $[0, 1]$ , giving each feature dimension an equal weighting. Two feature dimensions are illustrated in Figure 2.

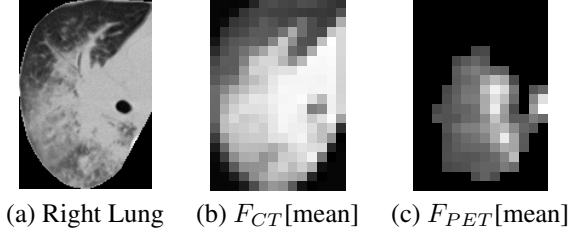


Figure 2. Sample feature grids.

### 2.3. Feature Space Transformation

Next, we transformed each texture feature vector into a *signature*. Physicians interpret images by identifying major patterns rather than using quantitative measures [20]. Collection of local homogeneous image patches can represent the perceptual information of an image, and Rahman et al used multi-class SVM to identify 30 classes of local patterns for a number of anatomical regions [15]. In our case, there was no prior knowledge about the distinctive patterns existing in the problem domain, so we employed an unsupervised learning approach –  $k$ -means clustering similar to the method used by Korfiatis et al [12].

A small training set of feature grids was selected for characterizing the clusters. We performed an iterative validation using the *Silhouette* method [17] to determine the number of feature pattern clusters. This led us to a cluster number  $K = 7$ . Euclidean distance was used during the clustering, and initial seeds were chosen randomly. To overcome the local minimum problem, the clustering process was executed multiple times to select the optimal result.

Once the centroids of clusters  $C(k)$  were trained, feature vectors  $\mathbf{F}$  were then mapped to certain cluster/feature categories via the minimum Euclidean distance classifier. The resultant category mapping represented the feature signature:

$$S(i, j) = \underset{k}{\operatorname{argmin}} \operatorname{dist}(F(i, j), C(k)), \quad k = [1, \dots, K]$$

The matrix  $\mathbf{S}$  was hence the signature for the PET-CT image pair.

Signatures for left and right lungs transformed from the feature grid for the example (Figure 1) are visualized in Figure 3.

### 2.4. Learning-based Signature Refinement

A closer examination of the signature data revealed that the unsupervised clustering procedure failed to separate abnormal lung areas from some body tissues. Specifically, as highlighted in Figure 3, pattern category 5 and 7 were intended for abnormal areas, but mistakenly covered some surrounding tissues as well. Such misclassification would

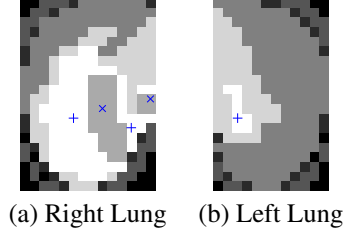


Figure 3. Signature mapping. “+” marks region with signature 7, “x” for signature 5.

adversely affect subsequent image comparison, and was caused by a lack of sufficient differentiating information for the two types of tissue in the feature vector  $\mathbf{F}$ .

To tackle this problem, we explored a supervised learning technique using SVM. SVM has been successfully used in the classification of lung CT images [10, 12].

At the training stage, a training dataset was selected manually. We collected eight signature grids for pattern category 5 and eight for pattern category 7, which amounted to 50+ data samples for each category. Each data sample was manually labeled with category 5, *non-5*, 7 or *non-7*. Manual selection was guided by limiting to those images with clear visual distinction between abnormal lung and surrounding tissues.

Secondly, for each data sample, we compiled a feature vector:

$$F_{ref}(i, j) = \{F(i, j), N(i, j)\}$$

where  $N(i, j)$  was  $K$ -dimensional vector as the number of its neighbors belonging to feature category  $k$ :

$$N_k(i, j) = \frac{\sum_{x, y \in \operatorname{neigh}(S(i, j))} \mathbb{1}_{S(x, y) = k}}{\sum_{x, y \in \operatorname{neigh}(S(i, j))} 1}, \quad \text{for } k = [1, \dots, K]$$

Other than using 4 or 8 connected neighbors, *neigh* computed the smallest circle of neighbors surrounding  $S(i, j)$  that have different signatures from itself.

Then, with feature vector  $\mathbf{F}_{ref}$ , two binary SVM’s were trained: one for category 5 (classifying 5 and *non-5*), and the other for category 7 (classifying 7 and *non-7*). To keep the SVM simple, we used a linear kernel.

During on-line processing, feature  $\mathbf{N}$  was calculated for each signature to formulate  $\mathbf{F}_{ref}$ , and the two SVMs were applied to re-assign a feature category. If it was categorized as *non-5* or *non-7*, a  $k$ NN based algorithm was used to assign a valid signature, through majority voting of its neighboring signatures. The final refined signature was named as  $\mathbf{S}_{ref}$ .

Figure 4 shows the refined signature grid of the previous example (Figure 3), presenting an obvious improvement over the original results.

We did not guarantee complete differentiation between

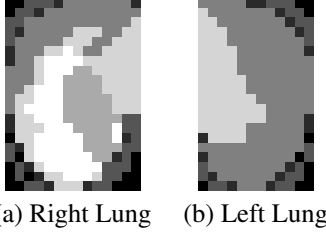


Figure 4. Signature refinement.

abnormal lung tissue and other soft tissues via this refinement step. This was partly due to our small training set, as it was sometimes impractical to determine the actual area type without expert clinical knowledge. Different sets of training data or choice of SVM kernel and parameters affected the performance as well. Furthermore, the feature vector  $\mathbf{F}_{\text{ref}}$  might not be comprehensive enough for all cases.

## 2.5. Similarity Measure

The image signature  $\mathbf{S}_{\text{ref}}$  was in a grid form, preserving the spatial relationships among the blocks. However, spatial relationships were not important when performing lung image retrieval – images with abnormal nodules at different locations should be considered similar. So, we reformed the signature into discrete bin distributions:

$$H(k) = \sum_{i,j} \text{equal}(S_{\text{ref}}(i,j), k), \quad \text{for } k = [1, \dots, K]$$

Each signature grid was then abbreviated into  $K$  integers, each of which representing the total number of blocks in the left and right LFs and SUVs belonging to feature category  $k$ .

Three types of similarity measures were implemented:

- Images ( $a$  and  $b$ ) were compared based on intersection of bin distributions:

$$D(a, b) = \sum_k w_k \frac{|H_a(k) - H_b(k)|}{\max(H_a(k), H_b(k))} \quad (1)$$

The four most similar images retrieved for the same example (Figure 1) are shown in Figure 5.

- Shape features were used to describe lung regions in addition to texture features by Bugatti et al [3] and Kakar et al [10]. We also extracted a set of shape features from left and right LFs:  $F_{\text{shape}} = \{\text{eccentricity}, \text{ratio}, \text{complexity}, \text{solidity}, \text{orientation}, \text{area}\}$ , where the parameters followed the standard property definitions of a connected component [10]. Formula (1) was employed to compute the distance between two features. It was then combined with the signature distance as the total difference value.
- Considering  $\mathbf{H}$  being a vector of  $K$  elements, rather than having each element play an equal part in determining the distance, different weighting could be applied to each element. This could help to highlight certain patterns, especially the abnormal areas. The difference computation

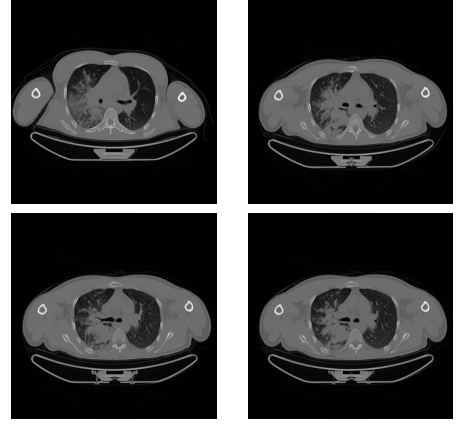


Figure 5. Retrieval results for Figure 1a.

became:

$$D(a, b) = \sum_k w_k \frac{|H_a(k) - H_b(k)|}{\max(H_a(k), H_b(k))} \quad (2)$$

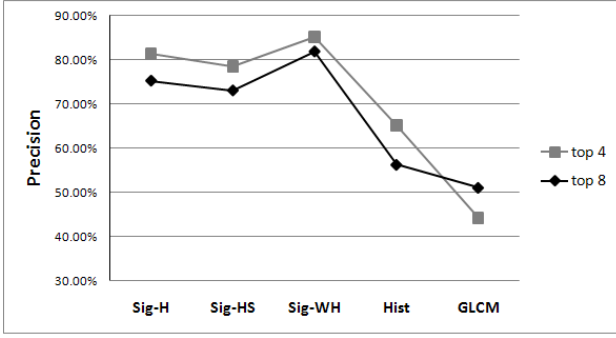
## 3. Experimental Setup

We tested our experiments on 20 patients of various stages of lung cancers, provided by Royal Prince Alfred Hospital of Sydney. Whole body co-registered PET-CT images were acquired in DICOM format on a Siemens TrueV 64 PET-CT scanner (Siemens, Hoffman Estates, IL). Image slices depicting anatomical regions other than lungs were filtered out manually, formulating 870 PET-CT image pairs.

All data were employed as query images. Each PET-CT image pair was compared against all images belonging to the other 19 patients. This restriction was to boost the diversity in the database during searching, effectively imposing more difficulties for retrieval. Search was performed for left or right lung, according to user command. Cross-matching, i.e. comparison between left and right lung images, was also supported to find common patterns residing at different locations.

We experimented three types of similarity measures in our framework: (1) Sig-H – comparison based on signature distribution  $\mathbf{H}$ ; (2) Sig-HS – including shape features in addition to  $\mathbf{H}$ ; (3) Sig-WH – using weighted  $\mathbf{H}$ . We only supported two levels of weightings: equal and double proportion; and the tuning was an interactive process during retrieval.

As suggested by Bugatti et al [3], we also compared our algorithm with traditional gray-level histogram (Hist) and co-occurrence matrix (GLCM). To compute the Hist and GLCM features, PET and CT images were firstly pre-processed to extract an approximate lung contour. Gray-scale histogram was then computed for 256 bins. Co-occurrence matrix of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$  were constructed, and 4 of Haralick features – homogeneity, energy,



**Figure 6. Average precision when retrieving top-4 or top-8 best matching images.**

contrast and correlation were calculated for both PET and CT images.

The *precision* and *recall* are typically used to evaluate CBIR systems [13]. We chose to evaluate the *precision* criteria only:

$$precision = \frac{\text{number of relevant items retrieved}}{\text{number of items retrieved}}$$

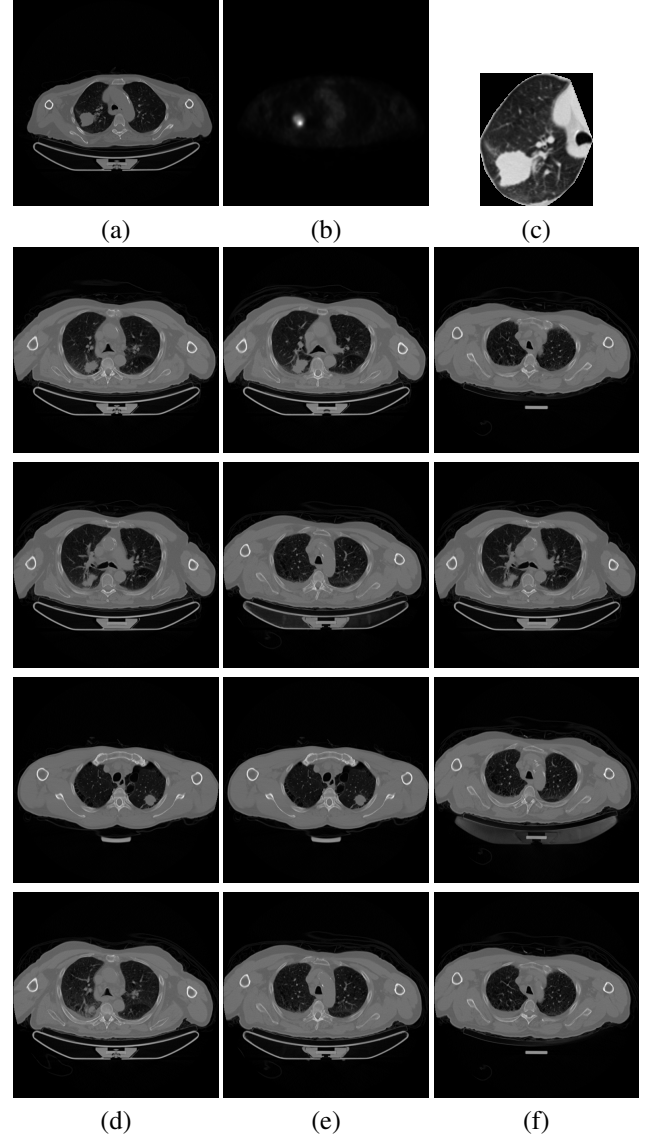
The reasons were two folded. Firstly, as a CBIR system meant for practical clinical applications, we followed the convention of ASSERT [19], which retrieved only the four most similar images for a query image and recall ceased to be an important factor. Secondly, our dataset was not pre-labeled with a disease category or any special region of interest (ROI), and assessment of retrieval accuracy was solely based on visual appearance; hence, preventing an objective measure for recall.

#### 4. Results

Figure 6 summarizes the precision measure when retrieving four and eight most similar images. About 81.4% of the returned images exhibited the same visual texture patterns as the query images with Sig-H. Using Sig-HS resulted in a slightly lower precision of 78.5%. By tuning the weighting of each element of **H**, a more optimal precision of 85.2% was achieved with Sig-WH. The precision readings for Hist and GLCM approaches were 65.2% and 44.3%.

When eight images were returned for a query image, precision readings for Sig-H, Sig-HS, Sig-WH and Hist lowered to 75.3%, 73.1%, 81.9% and 56.3%, respectively. For GLCM the performance improved to 51.1%. While decrease in precision was expected with inclusion of additional less similar images, improvement for GLCM was because the less-similar results deemed by the algorithm actually comprised more positive matches.

It was evident that our framework with Sig-WH for similarity measure demonstrated the highest precision. It outperformed the other two variations particularly in images with small but conspicuous problematic areas. Further enhancement could be possible by incorporating adaptive weight adjustment. It was also observed that for images with abnormal lung tissues, additional shape features (Sig-HS) often degraded the retrieval accuracy; and they provided added advantages only when the high FDG uptake areas in PET were relatively less obvious.



**Figure 7. Retrieval example. (a) Query image (CT). (b) Query image (PET). (c) Right LF extracted. (d) Sig-WH. (e) Hist. (f) GLCM.**

An example comparing our approach with the histogram

and co-occurrence matrix features is illustrated in Figure 7. Column (d) is the retrieval results in our framework with Sig-WH, showing all four images perceptually accurate. Histogram technique managed to retrieve two positive matches while co-occurrence matrix of only one.

The system was implemented in Matlab version 2009b. Running on a 2.66 GHz Pentium processor with a MS Windows XP operating system, one PET-CT image pair took about 2 seconds from pre-processing to signature finalization. Retrieval for one query image from the dataset took less than one second to execute.

## 5. Conclusion

In this paper, we proposed a novel framework for CBIR of PET-CT lung images. The method did not require a precise lung segmentation thus improved the speed of execution. The visual patterns were represented using texture features extracted with Gabor filtering, and further transformed into discrete pattern categories. Some incorrect categorizations resulted from inclusion of surrounding tissues during the lung field estimation, and a lack of sufficient differentiating information in the feature vectors. They were then refined with two binary SVMs to reduce the occurrences of mis-matching. Similarity measures were performed on signature distribution bins with tunable weightings. The experiment results demonstrated better performances comparing to traditional histogram and co-occurrence matrix based approaches. The speed of processing indicated another advantage of this framework.

## 6. Acknowledgment

This work was supported in part by ARC and PolyU grants.

## References

- [1] S. G. Armato and W. F. Sensakovic. Automated lung segmentation for thoracic CT. *Academic Radiology*, 11(9):1011–1021, 2004.
- [2] Y. Arzhaeva, M. Prokop, K. Murphy, and E. M. Rikxoort. Automated estimation of progression of interstitial lung disease in CT images. *Medical Physics*, 37(1):63–73, 2010.
- [3] P. H. Bugatti, M. P. Silva, A. J. M. Traina, C. Traina, and P. M. A. Marques. Content-based retrieval of medical images: from context to perception. *Proceedings CBMS*, 2009.
- [4] L. Chen, G. Lu, and D. Zhang. Effects of different Gabor filters parameters on image retrieval by texture. *Proceedings International Multimedia Modelling Conference*, pages 273–278, 2004.
- [5] I. ElNaqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad, and J. O. Deasy. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognition*, 42(6):1162–1171, 2009.
- [6] H. Greenspan and T. M. Deserno. Content-based image retrieval for biomedical image archives: achievements, problems, and prospects. *Proceedings MICCAI 2007 Workshop*, pages 1–6, 2007.
- [7] H. Guan, S. Antani, L. R. Long, and G. R. Thoma. Minimizing the semantic gap in biomedical content-based image retrieval. *Proceedings SPIE*, pages 45–52, 2010.
- [8] S. Hu, E. Hoffman, and J. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*, 20:490–498, 2001.
- [9] Y. Huan, C. Caldwell, K. Mah, and D. Mozeg. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging*, 28(3):374–383, 2009.
- [10] M. Kakar and D. R. Olsen. Automatic segmentation and recognition of lungs and lesion from CT scans of thorax. *Computerized Medical Imaging and Graphics*, 33(1):72–82, 2009.
- [11] J. Kim, L. Constantinescu, W. Cai, and D. Feng. Content-based dual-modality biomedical data retrieval using co-aligned functional and anatomical features. *Proceedings MICCAI 2007 Workshop*, pages 45–52, 2007.
- [12] P. Korfiatis, C. Kalogeropoulou, A. Karahaliou, and A. Kazantzi. Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. *Medical Physics*, 35(12):5290–5302, 2008.
- [13] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.
- [14] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber*, 9:62–66, 1979.
- [15] M. M. Rahman, S. K. Antani, and G. R. Thoma. A medical image retrieval framework in correlation enhanced visual concept feature space. *Proceedings CBMS*, 2009.
- [16] M. M. Rahman, P. Bhattacharya, and B. C. Desai. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. *IEEE Transactions on Information Technology in Biomedicine*, 11(1):58–69, 2007.
- [17] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, 20:53–65, 1987.
- [18] G. V. Saradhi, G. Gopalakrishnan, A. S. Roy, R. Mullick, R. Manjeshwar, K. Thielemans, and U. Patil. A framework for automated tumor detection in thoracic FDG PET images using texture-based features. *Proceedings ISBI*, pages 97–100, 2009.
- [19] C. R. Shyu, C. E. Brodley, A. C. Kak, and A. Kosaka. AS-SERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding*, 75(1/2):111–132, 1999.
- [20] W. Webb, N. Muller, and D. Naidich. *High resolution CT of the lung*. Lippincott Williams Wilkins, Philadelphia, PA, 2001.