

Structure-Adaptive Feature Extraction and Representation for Multi-Modality Lung Images Retrieval

Yang Song¹, Weidong Cai¹, Stefan Eberl^{1,2}, Michael J Fulham^{1,2,3}, Dagan Feng^{1,4}

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

²Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

³Sydney Medical School, University of Sydney, Australia

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

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

Abstract

Content-based image retrieval (CBIR) has been an active research area since mid 90's with major focus on feature extraction, due to its significant impact on image retrieval performance. When applying CBIR in the medical domain, different imaging modalities and anatomical regions require different feature extraction methods that integrate some domain-specific knowledge for effective image retrieval. This paper presents some new CBIR techniques for positron emission tomography - computed tomography (PET-CT) lung images, which exhibit special characteristics such as similar image intensities of lung tumors and soft tissues. Adaptive texture feature extraction and structural signature representation are proposed, and implemented based on our recently developed CBIR framework. Evaluation of the method on clinical data from lung cancer patients with various disease stages demonstrates its benefits.

1. Introduction

Advances in medical imaging instrumentation and software have produced a large increase in the number of digital images for patient studies. Automatic systems that retrieve cases similar to a given one are potentially beneficial for clinical diagnosis. CBIR techniques, which perform comparisons based on visual image features, can complement text-based searches and a variety of methods have been proposed for the medical domain [10].

A CBIR system is usually composed of feature extraction, similarity measure, storage access, indexing and retrieval, and user interaction components. The major fac-

tor determining the effectiveness of image retrieval is the feature extraction. There are several types of feature extractions including color or gray-level histograms, texture, shape, spatial relationships, and semantic features [10]. Abnormal lung tissues are often characterized by specific texture in CT images that help to identify a specific diagnosis [15]. Thus texture-based feature extractions are widely used in a variety of ways for lung images. Arzhaeva et al. proposed to derive texture features from Gaussian filtered images to estimate the progression of lung disease from CT images [1]. Kakar and Olsen used Gabor filters for feature extractions, and incorporated regional shape features together with *support vector machine* (SVM) for lung lesion recognition [6]. Korfiatis et al. performed a lung field extraction with SVM classifier and computed first order and co-occurrence matrix texture features for a further SVM classification for lung segmentation [8]. Saradhi et al. also employed co-occurrence matrix based texture features for tumor detection in PET lung images [12]. The feature vectors were high-dimensional computed at local level [8, 12]; or transformed into feature histograms or lower-dimensional signatures for more disease specific computation [1, 6].

Our work focuses on lung 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 session. PET-CT scans are widely used in the staging of lung cancer and a variety of other malignancies. The CT component has relatively poor soft tissue contrast resolution, making it challenging to separate lung tumors from normal surrounding tissues. PET images do not define the exact tumor boundaries. Whilst most lung tumors display increased standard uptake values (SUV) [4], non-cancerous tissue such as brown fat and

inflammation can also exhibit high SUV. This image complexity of PET-CT suggests that feature extractions tailored for PET-CT images are necessary.

Kim et al. [7] reported one approach for PET-CT lung image retrieval using a segmentation based technique. The lung regions segmented from PET and CT images were directly used as the features, i.e. gray-level intensities, without texture information. We have recently proposed a new framework that comprises an automatic lung field estimation, a three-step texture feature extraction, and a similarity measure. The system was evaluated on clinical data and demonstrated over 80% retrieval precision [14]. However, the work focused on the overall framework development, and lacked detailed optimization of the method. In particular, due to the local feature analysis approach, the features extracted included quantization errors for regions around the tumor boundaries. Furthermore, images were distinguished based on the distribution of signatures, without considering structural information. These introduced inaccuracies in image retrieval results. Thus we propose an adaptive feature extraction with structure information for CBIR of PET-CT lung images in this paper.

The remainder of the paper is structured as follows. Section 2 gives a brief summary of our previous work as a basis of the CBIR framework. Adaptive texture feature extracted is described in Section 3, and Section 4 presents structural signature representation. Experiments and results are discussed in Section 5 and 6. Section 7 concludes the paper.

2. Our Previous Work

This section introduces briefly our CBIR framework proposed previously [14]. The lung fields (LFs) were first extracted from the CT image. This step was aimed at removing background, patient bed, and surrounding tissues from the image, and separating left and right LFs into individual parts. A fast, fully automatic approach was employed, to encompass all lung tissues especially the high intensity areas (tumors), with a side effect that some soft tissues could be included. The LF masks were mapped to the co-registered PET image generating SUV data for left and right LFs.

Texture features were then extracted from both CT and PET images. Both images were firstly convoluted with 24 Gabor filters, and the resultant images with the maximum variances were selected to subdivide into 8-by-8 blocks. Texture features were thereafter computed for each block, which were *mean*, *variance*, *entropy* and *location* for CT images, and *mean* and *variance* for PET images. This constructed a 6-dimensional *feature vector* for each corresponding CT and PET image block.

Next, we collected a set of training images, and a k -means clustering approach was used to cluster the feature vectors into K pattern categories. K was 7 determined us-

ing the *Silhouette* method. Each corresponding CT and PET block was then assigned to the nearest pattern category (signature) by computing its Euclidean distances from the mean value of each pattern categories. Thus the PET-CT image pair was transformed into a *signature grid*.

Because of their similar characteristics, surrounding tissues and abnormal lung tissues could be assigned to the same pattern category. Differentiation of the two types of tissues was, however, essential for a meaningful image retrieval. Two binary SVMs were hence trained and employed to identify incorrect labeling of signatures, and a k -nearest neighbor (k NN) method was used to reassign the signatures for such blocks.

Lastly, the grid of signatures was reformulated into a discrete distribution of K bins, representing the number of blocks belonging to each pattern category. Similarity between two images was then measured by computing the intersections of the two signature bin distributions. Each bin could carry a different weight in similarity measure and this was adjusted during runtime interactively.

3. Adaptive Texture Feature Extractions

In our previous work rectangular grids were used for simple processing and efficient execution. Such subdivision, however, could create inhomogeneous blocks especially at the boundaries between different tissues. Texture features should be extracted from homogeneous regions for effective representation. Dynamic region-subdivision methods have been reported in the literature including: a multi-scale hierarchical image representation with region growing [9]; and a clustering method dividing an image into homogeneous blocks of irregular shapes [13].

In the PET-CT domain, a subdivision needs to be aligned for both PET and CT images so that complementary features can be extracted from corresponding regions. Direct application of above techniques would only formulate homogeneous regions in PET or CT images, but not necessarily aligned in most cases. We proposed a different approach, whereby feature vectors were extracted from homogeneous regions at sub-block level using a uniform subdivision (Figure 1).

The first step of the method was to identify regions of interest (ROIs), which were areas with high intensities (close to soft tissues) in CT, and SUV higher than 2.5 in PET [4]. Different techniques were employed for PET and CT images:

- ROIs in PET images were delineated using a *region growing* method. The pixel with maximum SUV was identified as the seed point. Region growing was then applied and stopped when the SUV value of a neighboring pixel dropped to 40% of the maximum SUV value of the initial seed; this was a clinically meaningful threshold for

hot areas [4]. The procedure was repeated to cover all hot areas where SUV was higher than 2.5.

- ROIs in CT images were estimated using an *expectation-maximization* (EM) based segmentation approach, similar to Blobworld [2]. EM was used to determine the maximum likelihood parameters of a mixture of K Gaussians in the feature space. In our case, two classes ($K = 2$) were used, to separate an image into ROIs and background. The feature space was 1-dimensional comprising only image intensities. The intensities x were hence modeled as Gaussian distributions with parameters $\theta_i = (\mu_i, \sigma_i)$ and $i = 1, 2$:

$$f_i(x|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right) \quad (1)$$

The update equations for the parameter θ_i took the form:

$$\mu_i^{(new)} = \frac{\sum_j x_j p(i|x_j, \theta^{(old)})}{\sum_j p(i|x_j, \theta^{(old)})} \quad (2)$$

$$\sigma_i^{(new)} = \frac{\sum_j p(i|x_j, \theta^{(old)})(x_j - \mu_i^{(new)})^2}{\sum_j p(i|x_j, \theta^{(old)})} \quad (3)$$

where $p(i|x_j, \theta)$ was the probability of Gaussian i fitting the pixel x_j . The estimate-update process repeated until the log likelihood increased by less than 0.1 percent from the previous iteration.

Figure 1 shows an example of the right LF (CT) and SUV (PET), and the two ROI masks generated from both. The two masks were then simply OR'ed together to produce a single ROI mask.

Next, the images were divided evenly into 8-by-8 blocks; and for each block, a feature vector was derived in the following ways:

- If the block fell completely inside or outside of the ROI mask, the 6-dimensional texture features were computed in the same manner as our previous work [14].
- For a block partially inside the ROI mask, the 6-dimensional texture features were calculated for foreground (pixels in ROI mask) only. The background pixels were ignored to concentrate on features from high intensity areas that had major effect on image retrieval. In addition, a new feature dimension was created to represent the proportion of foreground in the block:

$$proportion = \frac{size\ of\ foreground}{size\ of\ block} \quad (4)$$

So, for each block, seven features were extracted from the PET and CT image pair, which were mean, variance, entropy, and location from CT (entire block or foreground only), and mean and variance from PET (entire block or foreground only), and the proportion of foreground pixels in the block.

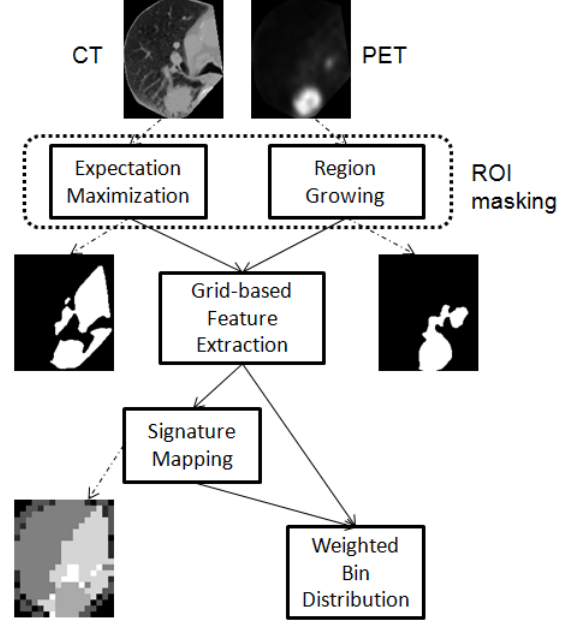


Figure 1. Flow of adaptive feature extraction.

Lastly, image signatures were generated using the clustering-based method and SVM refinement step that we reported in our previous work [14]. Each block was categorized into one of K feature patterns based on its feature vector (the first 6 dimensions); and the pattern category assigned was the signature sig . Feature *proportion* was used to consolidate the grid of signatures into a bin distribution \mathbf{H} of size K with X the total number of blocks in the grid:

$$H_k = \sum_{x=1}^X proportion_x v_{xk}, \quad v_{xk} = \begin{cases} 1 & \text{if } sig_x = k; \\ 0 & \text{otherwise} \end{cases} \quad \text{for } k = [1, \dots, K] \quad (5)$$

The bin distribution \mathbf{H} was K integers, each of which representing the number of blocks belonging to feature category k weighted by *proportion*.

4. Structural Signature Representations

A major limitation of the bin distribution representation \mathbf{H} was that the structural information of the image was lost. Because \mathbf{H} assembled only area information of the regions, it could not distinguish between two images where there were equivalent numbers of blocks in respective pattern categories while the region structures were different. Rahman et al. proposed a moving window technique creating overlapping blocks [11]. Kakar and Olsen incorporated shape descriptors in addition to texture features on a region basis [6]. We previously also experimented adding shape information of the lung fields at a global level, but concluded they provided no obvious benefits [14]. So, we designed

a different technique using a cluster-based structural representation, with the following three steps (Figure 2).

1. The *connected component* (CC) analysis was performed on the signature grid. A region of uniform signature (pattern category) was considered as a CC, from which the standard properties of a connected component, i.e. *eccentricity*, *ratio* and *complexity* were extracted [3]. The size of the CC was also calculated by summing the *proportion* values from its set of signatures. This transformed the signature grid into N 5-dimensional descriptors, where N was the number of CCs and a descriptor was $\{\text{sig}(\text{nature}), \text{s}(\text{ize}), \text{e}(\text{ccentricity}), \text{r}(\text{atio}), \text{c}(\text{omplexity})\}$.
2. The grid typically contained a large number of CCs, while only a small number exhibited the key characteristics for classifying the image. Keeping all CCs would only introduce noise in measuring image similarities. Thus insignificant CCs were removed from the descriptor matrix using *principal component analysis* (PCA) [5]. The matrix size was hence reduced to $M \times 5$, and M was set to 14 empirically. Considering CCs corresponding to different signatures should carry different significance/principal levels, the descriptors were first normalized by the mean values of pattern categories. This ensured the CCs representing high intensity regions would be prioritized in the PCA selection, filtering out only regions belonging to the background.
3. There could be multiple CCs of one pattern category, as they were not necessarily adjacent regions in an image. Rather than keeping individual CCs, they were grouped into clusters of identical signatures, to formulate a higher level representation of the image. When a cluster contained multiple CCs, the larger CCs would contribute more to the visual interpretation of the image. This suggested generating cluster descriptors by weighted averaging of the CCs in the cluster based on their sizes. The descriptor matrix was hence also reformulated into K numbers of 5-dimensional cluster descriptors.

$$\begin{aligned} D_k &= \{k, s_k, e_k, r_k, c_k\}, k = [1, \dots, K] \\ s_k &= \sum_{i=1}^M s_i v_{ik} & e_k &= \sum_{i=1}^M e_i v_{ik} s_i / s_k \\ r_k &= \sum_{i=1}^M r_i v_{ik} s_i / s_k & c_k &= \sum_{i=1}^M c_i v_{ik} s_i / s_k \\ v_{ik} &= 1 \text{ if } \text{sig}_i = k; 0 \text{ otherwise} \end{aligned} \quad (6)$$

As a result, the signature bin distribution \mathbf{H} was replaced by a descriptor matrix \mathbf{DM} comprising K cluster descriptors D . Its advantage over \mathbf{H} lay with the addition of region-based structural information $\{\text{eccentricity}, \text{ratio}, \text{complexity}\}$ on top of *size* for each pattern category k ; weighted averaging according to the sizes of homogeneous regions in the cluster; and inclusion of only visually significant regions into the representation.

Similarity measure between two PET-CT image pairs (a and b) were in turn computed based on the descriptor ma-

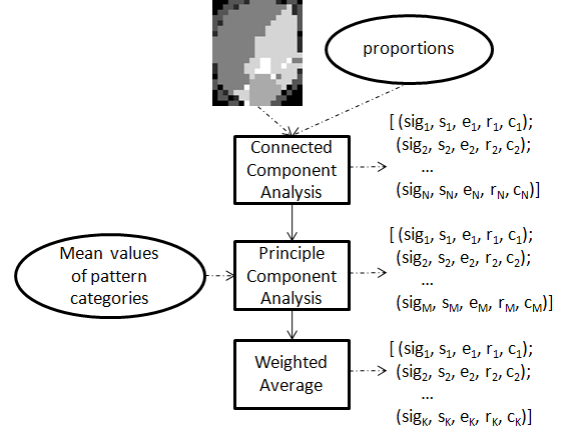


Figure 2. Flow of structural representation computation, where (sig, s, e, r, c) stands for (signature, size, eccentricity, ratio, complexity), respectively.

trix, using a weighted bin intersection approach:

$$d_{DM}(a, b) = \sum_{k=1}^K w_k d_D(a, b, k) \quad (7)$$

$$d_D(a, b, k) = \sum_{i=1}^5 \frac{DM_a(k, i) - DM_b(k, i)}{\max(DM_a(k, i), DM_b(k, i))} \quad (8)$$

where w_k was the mean value of each pattern category, effectively allocating bigger weightings to higher intensity areas.

5. Experimental Methods

We tested our approach 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 PET-CT image pairs were used as query images. Each one was compared against all images from the other 19 patients, ensuring results would not be skewed to adjacent slices of the same patient. Search was performed for left or right lung according to user command, but retrieval results could contain a mixture of matches from left and right lungs.

We conducted the experiments without and with the structural representation: (1) Adp-WH – adaptive feature extraction with weighted signature bin distribution; (2) Adp-DM – adaptive feature extraction plus structural representation DM . The results were also benchmarked with our testing results reported in the previous paper: Sig-WH (weighted signature bin distribution), Hist (gray-level 256-bin histogram), and GLCM (co-occurrence matrix) [14].

We employed the *precision* criteria for a quantitative evaluation of the retrieval performance:

$$\text{precision} = \frac{\text{number of relevant items retrieved}}{\text{number of items retrieved}} \quad (9)$$

The choice was based on the following:

1. As a CBIR system meant for practical clinical applications, our system returned only a small number (e.g. four) of most similar results. This suggested irrelevance of the *recall* criteria that was also frequently used in CBIR system evaluation [10].
2. Our dataset was not pre-labeled with disease categories or special regions of interest, and assessment of retrieval accuracy was entirely based on visual appearances. This prevented an objective measure for recall.

The system was implemented in Matlab version 2009b, running on a 2.66 GHz Pentium processor with a MS Windows XP operating system.

6. Results

The precision evaluation is summarized in Figure 3, for retrieving both four and eight most similar images. When texture features were extracted in the adaptive way (Adp-WH), 91.2% precision was achieved for returning top-four matches, and 85.7% for returning top-eight matches. This on average exhibited about 5% enhancement over our previous approach Sig-WH, 28% advantage over Hist and 40% over GLCM. The better performances were expected as by extracting texture features at homogeneous sub-block level, quantization errors caused by low-intensity areas were reduced, resulting in more accurate signature representation of the image.

By incorporating the structural descriptor matrix (Adp-DM), the precision was further increased to 94.1% and 92.4% for retrieving four and eight top matches respectively, which was about another 5% improvement on average. This enhancement was attributed to the inclusion of structural information of each visually-significant region with homogeneous signature, in addition to the size measure. Mismatches of images with similar distributions of signatures but different region structures were reduced. Using mean values of the pattern categories as the weightings for similarity measure was also one of the contributing factors.

An example of query results using Adp-DM is given in Figure 4. The search was conducted for the right lung containing a non-small-cell lung cancer (NSCLC). The eight most similar results were all positive retrievals in this case, with five matches in the right lung and three in the left lung. The matching lung fields were highlighted by rectangle boxes for illustration only.

Execution speed was also measured for both: (1) time taken for feature extraction and signature representation of

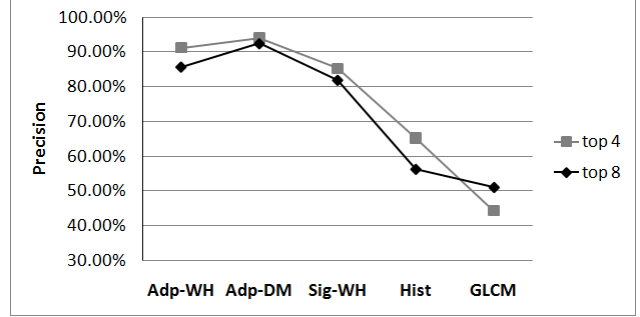


Figure 3. Average precision when retrieving top-4 or top-8 best matching images.

one PET-CT image pair, which also included the time for pre-processing; (2) time taken for similarity measure and retrieval of results. Time (1) was about 2.3 seconds, and time (2) was about 0.8 seconds in our experimental environment.

7. Conclusion

In this paper, we proposed new techniques for improving precision of medical image retrieval by representing image features adaptive to region structures. The intermediate level of image representation was a texture feature extraction at homogeneous sub-blocks which were aligned with ROIs identified from PET and CT images. With feature clustering and refinement using SVM, the PET-CT image pair was transformed into a grid of 1-dimensional signatures. The final image representation was K 5-dimensional vectors, which depicted structural information of signature-wise homogeneous regions, derived from connected component analysis, principal component analysis and weighted averaging methods. The experiment results validated the design and demonstrated improved overall retrieval precision over our previous approaches.

8. Acknowledgment

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

References

- [1] 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.
- [2] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.
- [3] R. C. Gozalez and R. E. Woods. *Digital image processing*. Addison-Wesley Publishing Company, MA, 1993.

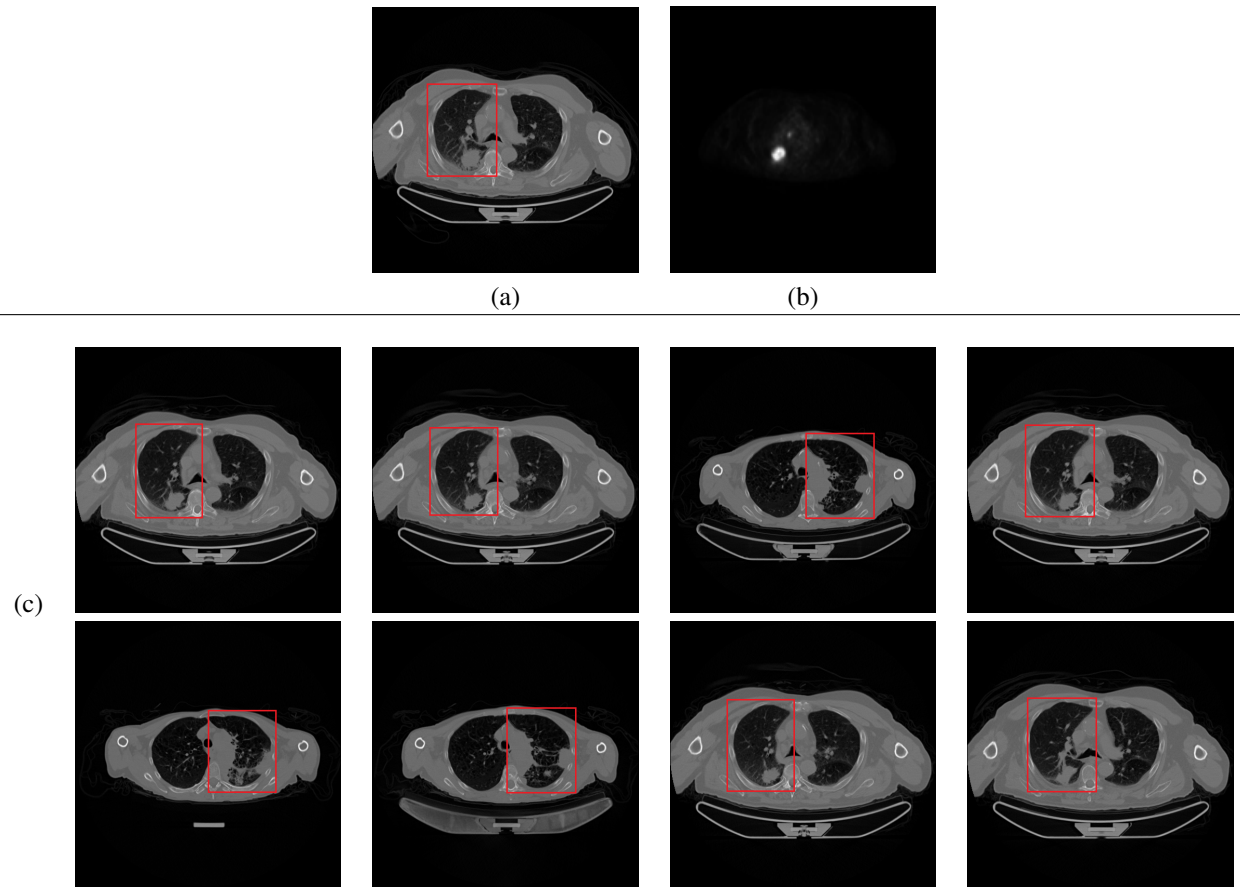


Figure 4. Retrieval example. (a) Query image (CT). (b) Query image (PET). (c) Retrieval results - 8 most similar images, showing only the CT components due to space constraints. Rectangle boxes highlighting matching lung fields for illustration only.

- [4] H. Guan, T. Kubota, X. Huang, X. S. Zhou, and M. Turk. Automatic hot spot detection and segmentation in whole body FDG-PET images. *Proceedings ICIP*, pages 85–88, 2006.
- [5] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal. Mach. Intell.*, 22:4–37, 2000.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. B. Wein. Content-based image retrieval in medical applications. *Methods Inf.Med.*, 43:354–261, 2004.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] I. C. Sluimer, M. Prokop, I. Hartmann, and B. Ginneken. Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution ct of the lung. *Medical Physics*, 33(7):2610–2620, 2006.
- [14] Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng. A content-based image retrieval framework for multi-modality lung images. *CBMS*, 2010.
- [15] W. Webb, N. Muller, and D. Naidich. *High resolution CT of the lung*. Lippincott Williams Wilkins, Philadelphia, PA, 2001.