

# THORACIC IMAGE CASE RETRIEVAL WITH SPATIAL AND CONTEXTUAL INFORMATION

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

Positron emission tomography - computed tomography (PET-CT) is now accepted as the best imaging technique to accurately stage lung cancer. The consistent and accurate interpretation of PET-CT images, however, is not a trivial task. We propose a content-based image retrieval system for retrieving similar cases from an imaging database as a reference dataset to aid the physicians in PET-CT scan interpretation. Problematic areas in diagnosis are the abnormal FDG uptake in the parenchymal lung tumor and in the regional nodes in the pulmonary hilar regions and the mediastinum. The primary tumor and the nodal disease are detected from the scans of thorax with learning-based techniques and a voting method for 3D object localization. Similar cases are then retrieved based on the similarity measure between the feature vectors of the cases. Our preliminary evaluation with clinical data from lung cancer patients suggests our approach is accurate with high retrieval precision.

**Index Terms**— PET-CT, tumor, lymph nodes, detection, retrieval

## 1. INTRODUCTION

Lung cancer is among the most common malignancies in the Western world and PET-CT is now accepted as the best imaging technique to accurately stage the most common form of primary lung cancer, non-small cell lung cancer (NSCLC). PET-CT scanners combine a helical CT scanner and PET scanner and produce anatomical and functional patient information from a single scanning session. The PET tracer <sup>18</sup>F-fluoro-deoxy-glucose (FDG) is the most commonly used tracer for clinical PET-CT diagnosis. Tumors typically take up more FDG than surrounding normal structures, and hence

its value in detecting disease in the lung fields but also in regional lymph nodes, which may not be enlarged.

Content-based image retrieval (CBIR) techniques that retrieve similar images from the database have been suggested as an aid to diagnosis [1]. Several recent developments in CBIR include: a local feature classification based mechanism [2]; a method that uses interactive windowing for image enhancement and organ-specific gray level features [3]; and a learning-based technique for cyst bounding box detection and retrieval [4]. These developments, however, have not been previously directed at PET-CT lung images.

There are few publications on image retrieval for PET-CT lung images. Kim et al. used an approach where retrieval performance relied heavily on segmentation accuracy [5]. We proposed another retrieval system that required less accurate segmentation based on grid-based feature histograms [6]. We then added an improvement to include the mediastinum to extract features from regional lymph nodes [7]. However, all these approaches were restricted to 2D image retrieval measuring differences between pairs of slices. Since a case or a patient's PET-CT scan comprises hundreds of transaxial image slices, the complete case is unlikely to be represented by a single slice, and the similarities between cases can only be obtained by analyzing the entire set of PET-CT images.

Our aim was to develop a CBIR framework to retrieve cases with similar characteristics within the 3D thorax volume, for a given patient's PET-CT scan. Reading physicians could then refer to similar retrieved cases that have been previously diagnosed as an aid to interpretation. We limited our attention to PET-CT images of the thorax rather than the entire torso.

Our approach is a case retrieval method that incorporates spatial and contextual information at the image slice and then object/case level. The major features of our approach are: a) it is fully automated to detect a primary lung tumor and disease in regional lymph nodes; b) 3D rather than 2D image retrieval

utilizes characteristics of the entire thorax. We describe our method in section 2, and experimental results in section 3.

## 2. METHODS

### 2.1. Materials

In this study, a total 1134 transaxial PET-CT image slice pairs were selected from 40 patients with NSCLC. The non-thorax images were filtered out. The images were acquired using a Siemens TrueV 64 PET-CT scanner (Siemens, Hoffman Estates, IL) at the Royal Prince Alfred Hospital, Sydney. The locations of tumors and disease in regional lymph nodes were annotated manually as the ground truth. 615 and 303 image slices exhibited visible tumor or disease in lymph node. For each patient study, the other 39 patient studies were manually marked as similar or dissimilar as a benchmark for measuring the retrieval performance. The similarity of cases was determined based on the location and appearance of the tumor and abnormal lymph nodes. The number of similar cases for each case ranged from 1 to 11, with an average of 4.75.

### 2.2. Preprocessing

During preprocessing, the PET data were converted to *standardized uptake values* (SUV) [8]. The lung and mediastinum were extracted from each slice with morphological operations and connected component analysis [7]. A precise segmentation was not the aim, and the spinal cord and other bone and muscle structures remained in the resulting image.

A modified fuzzy c-means (FCM) approach was then used to delineate each PET-CT slice into regions of various sizes and shapes [7]. Based on anisotropic diffusion filtering (ADF) [9], Gabor and shape features, each region was represented by a 12-dimensional feature vector  $\mathbf{F}$ : mean and variance of ADF/Gabor filtered CT values; mean and variance of ADF/Gabor filtered PET SUVs; the size; the eccentricity; and the centroid x and y coordinates.

### 2.3. Tumor and lymph node detection

The purpose of tumor and lymph node detection was to automatically determine the 3D location of the primary tumor (T) and any disease in regional lymph node (N). Our objective was not to detect normal-appearing lymph nodes. We used a 3-step method, described in the following paragraphs.

#### 2.3.1. Region type identification

The regions delineated from a PET-CT slice pair were classified into 5 types: lung lobe, mediastinum, disease in regional lymph node (N), tumor, and border area surrounding the tumor. The last two types formulated the T type. The classification was based on feature  $\mathbf{F}$  that combined PET and CT information.

We used the *support vector machine* (SVM) [10] due to its robust performance. A one-versus-all multi-class SVM (5 binary SVMs) was trained with linear kernels on the feature vectors  $\mathbf{F}$  of the regions. The feature weights for each region type were also derived from the support vectors, resulting in a  $5 \times 12$  matrix of feature weights  $\mathbf{w}$ .

#### 2.3.2. Region type refinement

The region types identified in 2.3.1 were often misclassified, particularly in terms of correct T and N classifications. Since both types could be characterized by high CT values and high SUVs, the feature vector  $\mathbf{F}$  was not sufficient to differentiate the two types. A major difference between T and N, however, could be modeled by the spatial relationships between regions. For example, T was within lung lobes while N was in the mediastinum or the hilar area.

We thus refined the region classifications based on spatial information. During the preprocessing, the region delineation step normally formulated several near-concentric regions at the T/N area, and several large regions corresponding to the lung lobes and mediastinum surrounding the T/N area. Therefore, for each region,  $R_i$ , initially classified as T or N, its spatial information could be defined by the regions inside and outside of  $R_i$ . A pair of regions  $\langle R_j, R_k \rangle$  was thus identified, which were the regions inside and outside of  $R_i$ . When there were multiple inside or outside regions, the region having the biggest overlap with the boundary of  $R_i$  was selected.  $R_j$  could be null when there was no region inside of  $R_i$ . A 36-dimensional feature vector  $\mathbf{FS}$  was hence formulated by concatenating the feature vectors of  $R_i$ ,  $R_j$  and  $R_k$ .

A second multi-class SVM (5 binary SVMs) was then trained to classify the regions based on the new feature vector. The final region type was determined by combining the margins computed from both multi-class SVMs. The output with the maximum combined margin was chosen as the region type.

#### 2.3.3. 3D object localization

Given the T or N regions identified from each PET-CT slice pair, we then attempted to localize the 3D T or N objects within a case. Because not all T or N areas were correctly identified from the slices due to errors from the classification and the region delineation step, we designed a voting-based method for the localization.

Two scores  $S_I^T$  and  $S_I^N$  were assigned to each slice  $I$  in a case:

$$S_I^{T/N} = \exp[-\min(d(I, J), d(I, K))] \quad (1)$$

$J$  and  $K$  were the slices spatially nearest to  $I$  (above and below) with detected T/N regions; and  $d$  was the normalized weighted ( $\mathbf{w}$ ) Euclidean distance between the feature vectors of the T/N regions. When T/N regions were not present in  $I$ ,

the feature vector was calculated from the bounding box of T/N regions in  $J$  or  $K$  with a linear transformation.

For each case, the mean score value and median locations of the T/N regions were computed from its set of images. T/N regions with scores higher than the mean score value was marked as valid locations. The regions with scores lower than the mean score and with a spatial offset from the median location larger than the region width were marked as invalid. For the unmarked slices, eg.  $I$ , its validity was then determined based on the slices  $J$  and  $K$  (above and below  $I$ ) that had marked T/N regions. If  $J$  was valid and  $K$  was invalid (vice versa), the similarities between  $I$  and  $J$ , and  $I$  and  $K$  were calculated. If  $I$  was more similar to  $J$  (or  $K$ ),  $I$  was then marked the same type of validity as  $J$  (or  $K$ ).

#### 2.4. Thorax case retrieval

The feature of a 3D T or N object  $\mathbf{F}_{\text{obj}}$  was computed as a weighted combination of the feature vectors  $\mathbf{F}_r$  of the comprising image slices.  $\mathbf{F}_r$  was computed from the bounding box of the T/N region in the image slice. The weight of an image slice was adaptively assigned as its mean ADF filtered CT intensity and PET SUV normalized by the maximum values present in the case, so that slices with more salient features would carry higher weights.

As a case could contain multiple Ts or Ns, the features of T or N objects were then combined, weighted according to the volume of the individual objects. A case was then represented by two feature vectors  $\mathbf{F}_T$  and  $\mathbf{F}_N$ , characterizing its tumor and lymph node appearances.

The distance between the query case  $X$  and the reference case  $Y$  was then defined as the weighted intersection differences:

$$D(X, Y) = \frac{v_T^Y D_T(X, Y) + v_N^Y D_N(X, Y)}{v_T^Y + v_N^Y} \quad (2)$$

$$D_{T/N}(X, Y) = \frac{|F_{T/N}^X - F_{T/N}^Y|}{F_{T/N}^X + F_{T/N}^Y} \quad (3)$$

where the weights were the volumes of the T or N objects  $v_T$  and  $v_N$ .

### 3. RESULTS

#### 3.1. Results on T/N detection

We first report the accuracy of the region type detection at the 2D image slice level. We performed the experiments by choosing 20 training samples per region category, and trained the first multi-class SVM. The training samples were selected incrementally by: starting from typical examples, and then adding samples that were wrongly classified from the initial trainings. The detection accuracy averaged from the classification results of all 1134 slice pairs is shown in Table 1a.

Ground Truth	Prediction (%)				
	L	M	N	T1	TB
Lung lobe	92.6	0.9	1.2	0.7	1.7
Mediastinum	0.6	93.5	0.9	2.3	3.0
Lymph node	1.7	9.9	74.6	8.3	5.6
Tumor	1.9	1.8	17.8	75.7	2.8
Tumor border	2.0	3.6	14.2	0.1	80.1

(a)

Ground Truth	Prediction (%)				
	L	M	N	T1	TB
Lung lobe	96.5	1.0	1.0	0.4	1.2
Mediastinum	0.5	95.4	0.5	2.0	1.8
Lymph node	1.3	9.9	83.8	3.3	1.7
Tumor	0.4	1.5	7.5	87.8	2.7
Tumor border	1.0	3.2	8.0	0.1	87.7

(b)

**Table 1.** The pairwise confusion matrix of the five region types tested on 1134 PET-CT image slice pairs. (a) Region type identification (section 2.3.1). (b) Region type refinement with spatial information (section 2.3.2)

	3D result		w/o voting	
	T	N	T	N
Sensitivity (%)	87.5	85.8	81.3	83.8
Specificity (%)	86.2	83.9	84.7	81.8

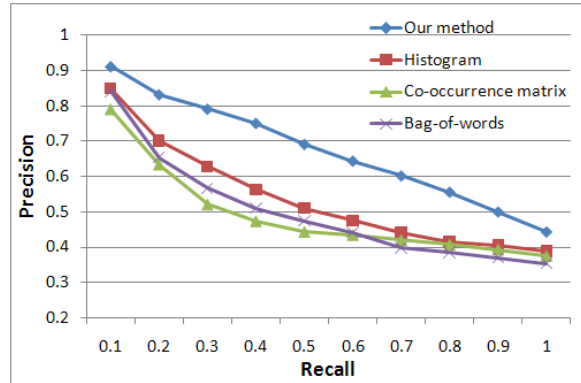
**Table 2.** The specificity and sensitivity of the T and N object localization with the voting procedure (section 2.3.3) tested on 40 cases. The result on 3D localization with region growing based on the region type detection output only without voting is also shown as a comparison.

We then trained the second multi-class SVM by incorporating the spatial information to refine the region types. The training data were based on the same set of samples used for the first SVM. The detection accuracy is shown in Table 1b. An example of region type detection is visualized in Figure 2.

We then assessed the localization of the T and N objects in a case. In testing, if the detected bounding box of the object overlapped with the ground truth box by at least 80%, it was considered as a correct detection. We computed the sensitivity and specificity of the object localization and the data were also compared with a region growing approach without the voting procedure. The results are shown in Table 2.

#### 3.2. Results on case retrieval

Given a case with detected T or N objects, all the other 39 cases were sorted based on the similarity measurement (section 2.4), and top matching cases were retrieved. Figure 2 shows a retrieval example, and we just show two image slice



**Fig. 1.** The precision-recall curve of the retrieval results on 40 cases. Our method is compared with the histogram, co-occurrence matrix and bag-of-words techniques.

pairs per retrieved case for easy visualization.

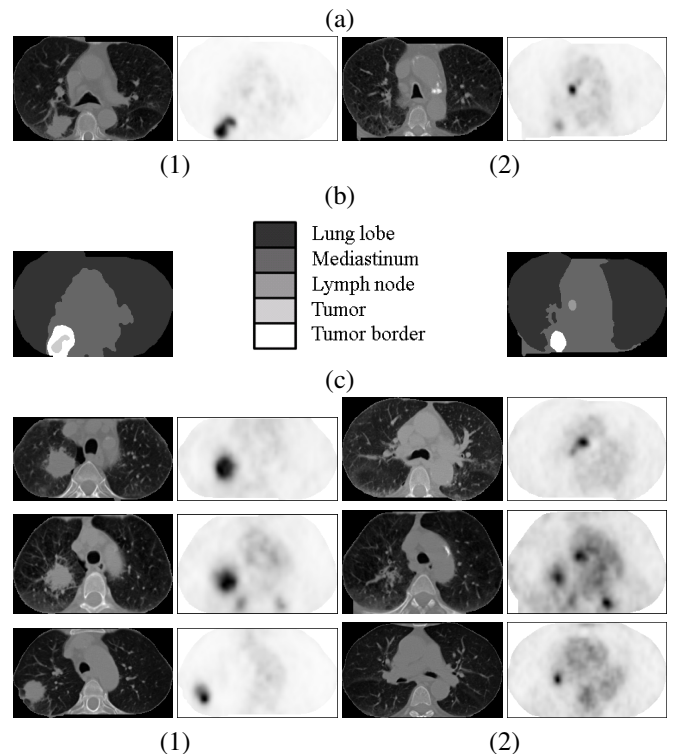
We also measured the retrieval performance by using each case as the query case, and calculating the average precision and recall of all 40 retrieval tests. We compared our method (T/N detection and similarity measure) with conventional histogram, co-occurrence matrix and bag-of-words methods (in 3D space). The results are shown in Figure 1.

#### 4. CONCLUSIONS

We propose a CBIR system for retrieving cases with similar tumor and lymph node appearances on PET-CT scans. Our system is based on automatic detection of tumor and disease in regional lymph nodes with learning and voting techniques. By extending the image retrieval techniques to case-level processing, the system can extract the overall feature from the thorax scans and retrieve cases with similar characteristics.

#### 5. REFERENCES

- [1] 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," *Int J. Medical Informatics*, vol. 73, pp. 1–23, 2004.
- [2] Md. M. Rahman, S. K. Antani, and G. R. Thoma, "A medical image retrieval framework in correlation enhanced visual concept feature space," *Proc CBMS*, 2009.
- [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," *Proc CBMS*, 2009.
- [4] J. Zhang, S. K. Zhou, S. Brunke, C. Lowery, and D. Comaniciu, "Detection and retrieval of cysts in joint ultrasound b-mode and elasticity breast images," *Proc ISBI*, pp. 173–176, 2010.
- [5] J. Kim, L. Constantinescu, W. Cai, and D. Feng, "Content-based dual-modality biomedical data retrieval using co-aligned



**Fig. 2.** Examples. (a) Two PET-CT image slice pairs (only the thorax area extracted) from case # 21 - (1) the PET-CT image showing T (2) the PET-CT image showing N. (b) Results of region type detection, where different gray-scale values indicate different region types. (c) Retrieval results of the three most matching cases with each row displaying two image pairs of one case - (1) the PET-CT image showing T (2) the PET-CT image showing N.

- functional and anatomical features," *Proc. MICCAI 2007 Workshop*, pp. 45–52, 2007.
- [6] Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, "Structure-adaptive feature extraction and representation for multi-modality lung image retrieval," *Int Conf Digital Image Computing: Techniques and Applications*, 2010.
- [7] Y. Song, W. Cai, S. Eberl, M. J. Fulham, and D. Feng, "Region and learning based retrieval for multi-modality medical images," *Biomed*, 2011.
- [8] H. Zaidi and I. E. Naqa, "PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *Eur J Nucl Med Mol Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010.
- [9] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans Pattern Anal Mach Intell*, vol. 12, pp. 629–639, 1990.
- [10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.