33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Thoracic Image Matching with Appearance and Spatial Distribution

Yang Song, *Student Member IEEE*, Weidong Cai, *Member IEEE*, Stefan Eberl, *Member IEEE*,
Michael J Fulham, David Dagan Feng, *Fellow IEEE*

***Abstract*— Positron emission tomography – computed tomography (PET-CT) produces co-registered anatomical (CT) and functional (PET) patient information (3D image set) from a single scanning session, and 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). This paper presents a content-based image retrieval (CBIR) method for retrieving similar images as a reference dataset to potentially aid the physicians in PET-CT scan interpretation. We design a spatial distribution to describe the spatial information of each region-of-interest (ROI), and a pairwise ROI mapping scheme between images to compute the image matching level. Similar images are then retrieved based on the local and spatial information of the detected ROIs, and a learned weighted sum of ROI distances. Our evaluation on clinical data shows good image retrieval performance.**

I. INTRODUCTION

Digitized medical images are produced in ever increasing quantities and are an essential component of modern health care. It has been suggested that the physicians can gather valuable information from the large collection of images for decision support, by using CBIR to retrieve images similar to a given one [1]. Unlike the early CBIR systems, which perform image retrieval by extracting low-level visual features, such as texture and shape on the image content, more recent approaches extend the searching capability to high-level semantic concepts in predominantly two ways: (i) the low-level features are classified into a number of concept categories and represented as bag-of-features for the whole image [2], [3]; and (ii) the low-level features are first classified to detect regions of interest (ROIs) from the image of objects which are relevant to the classification of the image, and local features extracted for the ROIs are used to compute image dissimilarity [4], [5].

The bag-of-features model usually builds on image blocks, so the division of blocks can introduce problems when a block is at an object boundary. The ROI based methods can mitigate such problems by localizing the objects first. Both bag-of-features and ROI based feature representations normally ignore the spatial relationships between the image patches (blocks or ROIs), causing loss of structural information. Rahman et.al. addressed the problem by introducing a sliding-window based structural descriptor [2]; and Avni et.al. included the coordinates of the image patches in the feature vector [3]. More popular approaches for encoding the spatial information have been proposed in general computer vision, including spatial pyramid matching [6], directional spatial histogram [7], and visual words and phrases [8].

For PET-CT thoracic images with PET tracer ^{18}*F-fluoro-deoxy-glucose* (FDG), the primary lung tumor and metastatic disease in regional lymph nodes, which typically exhibit higher FDG uptake than surrounding normal structures, are usually considered as the main ROIs for diagnosis and therapy planning, motivating the work on tumor and lymph node segmentations [9], [10]. It will be beneficial, however, to also consider the spatial location of the ROIs relative to the lung and mediastinum, as it is important for staging NSCLC. In this context, a CBIR method for thoracic images should retrieve images that are not only similar in local ROI features, but also in the spatial features of the ROIs. For example, an image with a primary tumor near to the mediastinum should be matched to images with similar patterns.

The objective of our work is thus to design an image retrieval method for thoracic PET-CT images with NSCLC, based on the presence and spatial information of tumors and disease in lymph nodes. We approach the problem by first detecting the ROIs using support vector machines (SVM) on mean-shift clustered image regions; then extract the spatial location for each ROI relative to other structures; and compute the dissimilarity between the query image and the images in the database by creating a pairwise ROI mapping.

II. METHODS

A. Appearance-based ROI Detection

At the first stage, we detected the ROIs, such as tumor or disease in regional lymph nodes, from the thoracic 3D image set. The detection was designed to support the image matching, with the following criteria: (i) we preferred to detect only the tumors and abnormal lymph nodes, minimizing the inclusion of high uptake regions in the mediastinum; (ii) it was acceptable to include some spurious regions as ROI, rather than missing any true abnormal areas; and (iii) while considered beneficial, a good ROI delineation and precise segmentation are not required for the image matching.

We designed a three-step method for the detection. First, each PET-CT transaxial slice pair was clustered into regions using mean-shift clustering [11] by first clustering the CT and PET slices separately, then superimposing the clustering

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

Yang Song, Weidong Cai, Stefan Eberl, Michael J Fulham and David Dagan Feng are with BMIT Research Group, School of Information Technologies, University of Sydney, Australia

Stefan Eberl and Michael J Fulham are with Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia

Michael J Fulham is also with Sydney Medical School, University of Sydney, Australia

David Dagan Feng is also with CMSP, Department of EIE, Hong Kong Polytechnic University, Hong Kong, and Med-X Research Institute, Shanghai Jiao Tong University, China

outputs from both modalities into one set of regions. We chose to use the mean-shift clustering because it could better preserve the edges, and the resulting regions approximated the contours of the structures well; and it could determine the number of reqruied clusters for the image data dynamically.

Then, we classified each of the regions into one of three classes – the lung field, mediastinum and ROI. They could be distinguished based on the average CT density and standard uptake values (SUV) for PET, and we trained a three-class one-versus-all SVM for the classification. Note that we classified both tumors and abnormal lymph nodes (and if any non-pathological ROIs with relatively high uptake) as the same type – ROI, to avoid complicated procedures for discriminating between these types; and they were differentiated when computing the image dissimilarities (section *C*).

At the third step, a 3D connected component analysis was performed to group the spatially adjacent ROIs across transaxial slices into one volumetric ROI.

B. Spatial Distribution Description

We then described the spatial information of the detected ROIs in feature vectors, to facilitate the image matching. The spatial information was important, as we wanted to retrieve image sets with ROIs appearing at the similar position – near the chest wall or mediastinum, or well within the lung field (tumor) or mediastinum (lymph nodes). The commonly used bag-of-feature or bag-of-region representations, which neglect the relative spatial locations, were thus not sufficient.

To compute and encode the spatial relationships between the ROIs and the lung fields and mediastinum, a 4-directional circular spatial distribution was designed (Figure 1). Originating from the ROI boundary, N_o contours grew towards outside of the ROI and N_i contours towards inside of the ROI. The contours were of width d pixels, displacement from the ROI border or the previous contour. Each contour was divided into four bins, in the direction of anterior, posterior, lateral and medial of the thorax, relative to the ROI. Each bin counted the number of voxels belonging to class t – lung field (LF), mediastinum (MS) or ROI:

$$\mathbf{h}_{c,b}^{t} = \sum_{i} 1_{|y_i = t|} \tag{1}$$

where c and b indexed the contour and bin in the contour, and y_i was the class label of voxel i. The outside bins described the spatial information, e.g. an ROI (tumor as in Figure 1b) near to the mediastinum would have large $\mathbf{h}_{c,b}^{MS}$ in the medial bins but large $\mathbf{h}_{c,b}^{LF}$ in other bins. The inside bins represented roughly the shape of the ROI, depicted by the ratio among the four bins. The spatial distribution descriptor $\mathbf{H}$ was an ordered concatenation of all bins, accumulated over all image slices, weighted by the ratio between the size of ROI in the slice and the whole ROI volume, and computed for each volumetric ROI present in the 3D image set (Algorithm 1).

This spatial distribution was translation invariant, since each descriptor $\mathbf{H}$ was centered around the ROI. We made the descriptor scale and rotation variant, because they were important for discriminating the spatial information between

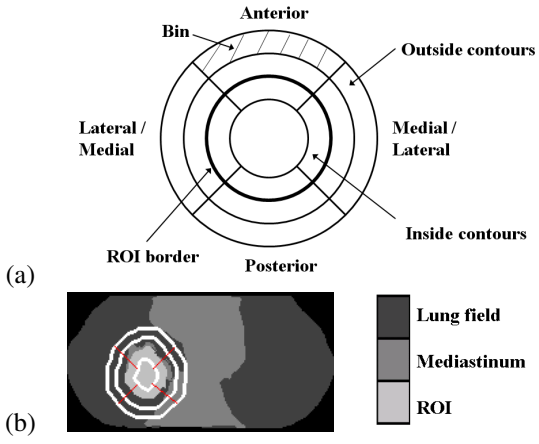

Fig. 1. Illustration of the spatial distribution bins. (a) Conceptual diagram (abstract view with ideal round circles). (b) Overlay of the borders of the contours (white lines) on the detection output (showing one transaxial slice only), with contours ($N_o = 2$ and $N_i = 1$) centered around the detected ROI, and the directional lines in red color.

Algorithm 1: Computing the spatial distribution

Data: 3D image set $\mathbf{I}$ comprising S_I transaxial slices with class labels $\mathbf{Y}_I$, and containing R_I ROIs.
Result: spatial distribution $\mathbf{H}_{I,r}$ for each ROI r in $\mathbf{I}$.
initialization;
for $r = 1, ...R_I$ **do**
  for $s = 1, ...S_I$ **do**
    if *ROI* r *present in slice* s **then**
      Create N_o contours outside of ROI r, and N_i contours (of width d) inside of ROI r;
      for $c = 1, ...N_o$, *and* $1, ...N_i$ **do**
        Divide the contour c into four bins;
        Compute the feature vector $\mathbf{h}_{c,b}^{t}$;
        $\mathbf{H}_{I,r}^{s} = \mathbf{H}_{I,r}^{s} \mid \mathbf{h}_{c,b}^{t}$;
      end
      $\mathbf{H}_{I,r} = \mathbf{H}_{I,r} + \mathbf{H}_{I,r}^{s} \times area_{I,r}^{s}$;
    end
  end
  $\mathbf{H}_{I,r} = \mathbf{H}_{I,r} \,/\, volume_{I,r}$;
end

ROIs. N_o was chosen as 3 to describe the nearby areas of the ROI, N_i was 1 to outline the ROI contour, and d was set to 8 pixels empirically.

C. Discriminative Feature Matching

Given 3D image set $\mathbf{I}_1$, each volumetric ROI i of $\mathbf{I}_1$ was then described by the feature vector $\mathbf{f}_{1i}$, which comprised the spatial descriptor $\mathbf{H}_{I_1,i}$, size of the ROI, and its texture statistics (mean, standard deviation, skewness and kurtosis) for both CT and PET. The image dissimilarity D between two image sets $\mathbf{I}_1$ and $\mathbf{I}_2$ was computed as the distance between the two sets of ROI descriptors, $\mathbf{f}_1$ and $\mathbf{f}_2$:

$$D(\mathbf{I}_1, \mathbf{I}_2) = \sum_{\{\mathbf{f}_{1i}, \mathbf{f}_{2j}\} \in \mathbf{M}} u_{1i}\, \Delta(\mathbf{f}_{1i}, \mathbf{f}_{2j}) \tag{2}$$

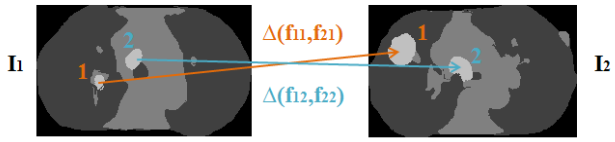

Fig. 2. Illustration of mapping between ROIs in $\mathbf{I}_1$ and $\mathbf{I}_2$ – ROI Set 1 (tumor) denoted with 1 and ROI Set 2 (abnormal lymph nodes) denoted with 2; and the pairwise ROI distances between the mapped ROIs denoted as $\Delta(\mathbf{f}_{11}, \mathbf{f}_{21})$ and $\Delta(\mathbf{f}_{12}, \mathbf{f}_{22})$. Simplifying the illustration with only one ROI in each set, and the 3D volumes of ROIs shown as a 2D transaxial slice view.

$$\Delta(\mathbf{f}_{1i}, \mathbf{f}_{2j}) = \langle \mathbf{w} \cdot \frac{|\mathbf{f}_{1i} - \mathbf{f}_{2j}|}{\mathbf{f}_{1i} + \mathbf{f}_{2j}} \rangle \quad (3)$$

Here the matrix $\mathbf{M}$ indicated the pairwise mapping between ROIs in $\mathbf{I}_1$ and $\mathbf{I}_2$, so that the image dissimilarity was computed as a weighted sum of the pairwise ROI differences between the mapped ROIs. The weight u_{1i} was the ratio between the size of ROI i and all ROIs in $\mathbf{I}_1$. The mapping $\mathbf{M}$ was supposed to observe a contextual map, i.e. ROIs in $\mathbf{I}_1$ representing tumors should be mapped to tumor ROIs in $\mathbf{I}_2$, and similarly for ROIs representing abnormal lymph nodes (Figure 2). To create such mapping, the ROIs in each 3D image set were first divided into two sets: Set 1 – likely to be tumor; and Set 2 – likely to be abnormal lymph nodes. Each set could contain multiple ROIs, and the division was based on the location of the ROI relative to the lung and mediastinum, computed by the proportion of voxels labeled as lung field in the spatial distribution of the ROI, using k-means clustering (2 clusters). For Set 1 of $\mathbf{I}_1$ and $\mathbf{I}_2$, a one-to-one mapping was generated by first sorting the ROIs based on their sizes separately for Set 1 of $\mathbf{I}_1$ and $\mathbf{I}_2$, then aligning the ROIs between $\mathbf{I}_1$ and $\mathbf{I}_2$ in the sorted order. For Set 2, each ROI in $\mathbf{I}_1$ was mapped to the ROI in $\mathbf{I}_2$ having the minimum distance, thus the mapping could be many-to-one. If there were fewer ROIs in Set 1 of $\mathbf{I}_2$ than Set 1 of $\mathbf{I}_1$, the one-to-one mapping for Set 1 implied that some ROIs in Set 1 of $\mathbf{I}_1$ would be left unmapped, which were then mapped to Set 2 of $\mathbf{I}_2$ based on the minimum distance.

The pairwise ROI difference $\Delta(\mathbf{f}_{1i}, \mathbf{f}_{2j})$ was weighted by vector $\mathbf{w}$, which controlled the level of contribution of each feature dimension. We derived $\mathbf{w}$ with the triplet-learning method. A triplet $\langle \mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3 \rangle$ denoted that $\mathbf{I}_1$ was similar to $\mathbf{I}_2$ and dissimilar to $\mathbf{I}_3$; and it was thus expected to satisfy $D(\mathbf{I}_1, \mathbf{I}_3) > D(\mathbf{I}_1, \mathbf{I}_2)$, which could be rewritten as the following:

$$\langle \mathbf{w} \cdot \sum u_{1i} \frac{|\mathbf{f}_{1i} - \mathbf{f}_{3j}|}{\mathbf{f}_{1i} + \mathbf{f}_{3j}} \rangle > \langle \mathbf{w} \cdot \sum u_{1i} \frac{|\mathbf{f}_{1i} - \mathbf{f}_{2j}|}{\mathbf{f}_{1i} + \mathbf{f}_{2j}} \rangle \quad (4)$$

With a set of training triplets, $\mathbf{w}$ was then solved using the large-margin optimization method [12].

D. Materials and Evaluation Study

A total 1134 transaxial thoracic PET-CT slice pairs were selected manually from 40 studies (3D image sets) to filter out non-thoracic image slices. The images were acquired using a Siemens TrueV 64 PET-CT scanner (Siemens, Hoffman Estates, IL) at the Royal Prince Alfred Hospital, Sydney. For each study, the other 39 image sets were manually marked as similar or dissimilar as the ground truth for retrieval. The similarity of image sets was determined based on the location and appearance of the tumor and abnormal lymph nodes. The number of similar image sets for each set ranged from 1 to 11, with an average of 4.75.

Fully-automatic preprocessing was performed on each CT slice to remove the patient bed and soft tissues outside of the lung and mediastinum, based on simple thresholding, morphological operations and connected component analysis. The resulting mask was then mapped to the co-registered PET slice.

The image retrieval performance was evaluated by using each 3D image set as a query image, and the other 39 sets were ranked according to their similarity level with the query image. The retrieval results were then compared with the ground truth to calculate the precision and recall. Since we expected that normally physicians would be interested to look at only the first few retrieved images, we focused on assessing the top one, three and five retrievals.

Besides our proposed method, we also evaluated the retrieval performance with equal weights (all ones) in the distance function (3); replacing our pairwise ROI mapping scheme to a minimum sum distance based approach; and computing the image dissimilarity based on the local ROI features only, i.e. its texture, size and location (coordinates), without the spatial distribution descriptor.

We then compared our method with three standard approaches: (i) global histogram (HIST), with 256 bins for CT and 256 bins for PET; (ii) bag-of-features (BOF), classifying 8 by 8 image patches into 7 clusters (the sizes were chosen to have the resulting BOF approximate the original image appearances closely); and (iii) spatial pyramid matching (SPM), with 3 levels (as suggested in the original paper [6]) based on the BOF representation.

III. RESULTS

Our method achieved 100% true positive rate (TPR) for ROI detection that all abnormal areas, i.e. tumors and disease in regional lymph nodes, were correctly detected. The detection precision was 66.4% with some non-pathological areas with high uptake values (e.g. myocardium) in the mediastinum detected as ROIs as well. Because our image dissimilarity was computed based on localized ROIs, exclusion of any important ROI would result in incomplete feature representation of the image. So, the high detection TPR was desired. The low precision was a compromise to achieve the high TPR; the extra ROIs detected were normally small in size thus had low impact on the dissimilarity computation. Examples of the ROI detection are shown in Figure 3.

The average retrieval precision and recall of our method are listed in Table I, column *Ours*. With equal weights in the distance function, lower precision and recall was observed (Table I, column *EqW*), which suggested the benefit of using learning-based feature weights. The lower results from the minimum sum distance based approach (column *MinD*)

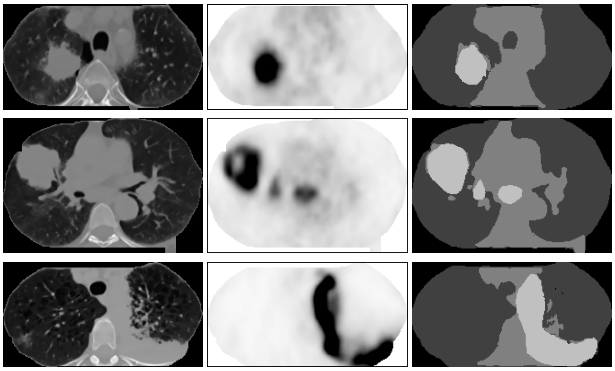

Fig. 3. Three examples of ROI detection, showing one transaxial slice pair per patient study. The left column is the CT image slice (after preprocessing); the middle column is the co-registered PET slice; and the right column is the classification output.

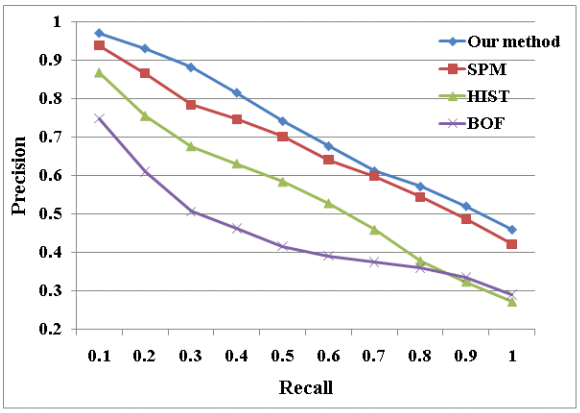

Fig. 4. The precision-recall curve of the retrieval results.

TABLE I

THE PRECISION (P) AND RECALL (R) MEASURE OF THE RETRIEVAL RESULTS OF THE TOP ONE, THREE OR FIVE MOST SIMILAR MATCHES.

P (%)	Ours	MinD	EqW	ROI	HIST	BOF	SPM
Top-1	80.7	74.2	54.8	67.7	48.4	19.4	67.7
Top-3	69.4	64.5	58.6	59.7	45.2	28.0	59.7
Top-5	61.1	59.0	49.4	57.3	41.9	34.0	55.2

R (%)	Ours	MinD	EqW	ROI	HIST	BOF	SPM
Top-1	25.1	23.2	18.2	23.8	14.5	8.3	19.4
Top-3	48.6	45.5	39.4	43.1	26.9	19.1	39.7
Top-5	59.1	56.4	46.0	55.2	36.6	35.6	52.8

indicated the advantage of creating a customized pairwise ROI mapping between two images. The image dissimilarity based on the local ROI features only also resulted in lower precision and recall (column *ROI*), which were expected due to omission of the spatial information.

The precision and recall values of the three standard approaches are listed in Table I (column *HIST*, *BOF* and *SPM*) for top retrievals, and the average precisions for all levels of recall are shown in Figure 4. All three approaches incorporated the same feature weight training for the distance function as our method. If without the optimized feature weights, the average precisions of the three compared methods were 5-8% lower than the listed values. HIST performed better than BOF, although we tried to choose the best settings for BOF; and we believe it was mainly due to the subdivision of image patches around the object boundaries. As expected, SPM improved considerably over BOF and HIST, since it modeled the spatial information. However, the hierarchical structure of SPM was based on the image dimension only with even sub-level divisions, and not centralized on ROIs; hence the descriptor was not translation invariant, and this would result in large difference between images due to translation only. Our method effectively shifted the center of the hierarchy to the ROIs, and assigned higher weights to the nearby image patches with our spatial distribution design; and thus achieved higher results.

IV. CONCLUSIONS

We proposed a new method for retrieving PET-CT images with similar location and appearance of the lung tumor and abnormal lymph nodes. Three major components contributed to the high retrieval performance: clustering and classification based ROI detection, ROI feature description with spatial distribution, and an image dissimilarity measure based on pairwise ROI mapping and optimized feature weights. Based on the evaluation results on 40 NSCLC patient studies, our system showed high retrieval precision and recall, and outperformed other proposed techniques.

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, 2004, pp 1-23.

[2] M.M. Rahman, S.K. Antani and G.R. Thoma, "A medical image retrieval framework in correlation enhanced visual concept feature space", *in IEEE CBMS*, 2009, pp 1-4.

[3] U. Avni, H. Greenspan, M. Sharon, E. Konen and J. Goldberger, "X-ray image categorization and retrieval using patch-based visual words representation", *in ISBI*, 2009, pp 350-353.

[4] Y. Song, W. Cai, S. Eberl, M.J. Fulham and D. Feng, "Thoracic image case retrieval with spatial and contextual information", *in ISBI*, 2011, pp 1885-1888.

[5] L. Sorensen, M. Loog, P. Lo, H. Ashraf, A. Dirksen, R.P.W. Duin and M.D. Bruijne, "Image dissimilarity-based quantification of lung disease from CT", *in MICCAI*, 2010, pp 37-44.

[6] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories", *in CVPR*, 2006, pp 2169-2178.

[7] D. Liu, G. Hua, P. Viola and T. Chen, "Integrated feature selection and higher-order spatial feature extraction for object categorization", *in CVPR*, 2008, pp 1-8.

[8] S. Zhang, Q. Tian, G. Hua, Q. Huang and S. Li, "Descriptive visual words and visual phrases for image applications", *in ACM MM*, 2009, pp 75-84.

[9] V. Potesil, X. Huang and X.S. Zhou, "Automated tumour delineation using joint PET/CT information", *in SPIE*, 2007, pp 65142Y.

[10] J. Wojak, E.D. Angelini and I. Bloch, "Joint variational segmentation of CT-PET data for tumoral lesions", *in ISBI*, 2010, pp 217-220.

[11] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", *PAMI*, vol. 24, no. 5, 2002, pp 603-619.

[12] A. Frome, Y. Singer, F. Sha and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification" *in ICCV*, 2007, pp 1-8.