REGION AND LEARNING BASED RETRIEVAL FOR MULTI-MODALITY MEDICAL IMAGES

Yang Song[1], Weidong Cai[1], Stefan Eberl[1,2], Michael J Fulham[1,2,3], Dagan Feng[1,4]
[1]Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia
[2]Department of PET and Nuclear Medicine, Royal Prince Alfred Hospital, Sydney, Australia
[3]Sydney Medical School, University of Sydney, Australia
[4]Center for Multimedia Signal Processing (CMSP), Department of Electronic &
Information Engineering, Hong Kong Polytechnic University, Hong Kong
[1]{ysong,tomc,feng}@it.usyd.edu.au,[2]stefan@nucmed.rpa.cs.nsw.gov.au,[3]mfulham@med.usyd.edu.au

ABSTRACT
We present a region-based image retrieval framework for multi-modality, positron emission tomography - computed tomography (PET-CT), images. An image retrieval system can be used to assist the diagnostic process, by providing reference cases that contain similar scans to the interpreting clinicians. PET-CT scans are essential tools for the accurate staging of lung cancer and provide co-registered functional (PET) and anatomical (CT) information from a single scan; the complexity of these data, however, place new challenges in computerized image analysis and retrieval. The choice of a region-based method was inspired by the objective of retrieving images with similar patterns of disease involvement, where there is a parenchymal lung tumor and disease in regional lymph nodes. Our results on clinical data from lung cancer patients show a higher retrieval precision over the usual techniques and the other non-region based methods.

KEY WORDS
PET-CT image, retrieval, region, weights, learning

1 Introduction

Advances in medical imaging instrumentation have resulted in a marked increase in the number of images that are available for each patient study. The patient imaging studies are generally archived in large databases and search methods of these databases, which rely on text-based approaches, are problematic. Hence new techniques that include content-based image retrieval (CBIR) have been proposed [1].

In this work our focus is PET-CT images. PET-CT scanners combine a helical CT and a PET scanner and produce anatomical and functional patient information from the single scan. For routine clinical diagnosis PET scanners use a PET tracer ^{18}*F-fluoro-deoxy-glucose* (FDG), which is taken up by tumor cells relative to the normal lung and normal lymph nodes. Lung cancer (non small cell lung cancer NSCLC) is among the most common malignancies in the developed and developing world and PET-CT is now accepted as the best imaging technique to accurately determine the extent (stage) of NSCLC. The extent of disease determines the best approach to therapy.

The consistent and accurate interpretation of PET-CT data, however, is not a trivial task. PET images have lower resolution that the CT images and do not delineate the precise location of a tumor or its boundaries, but PET can detect disease in non-enlarged lymph nodes. CT depicts anatomy but has poor soft tissue contrast resolution, has difficulty in separating tumors from normal surrounding tissues and disease in lymph nodes is determined by an increase in size of the node. Further, there is a learning curve associated with the accurate interpretation of the various findings in patients with NSCLC on PET-CT and can be problematic for an inexperienced reader.

Our aim was to develop a CBIR framework for PET-CT lung images. For a query image of the thorax, the system is expected to retrieve visually similar images from previously studied cases. Readers can then retrieve images that have been previously associated with diagnostic information to aid their interpretation. We have limited our attention to thoracic PET-CT images rather than the entire torso, which is usually the case in clinical practice for staging NSCLC.

There are few publications on image retrieval for thoracic PET-CT images. Kim et al. [2] used an approach where retrieval performance relied heavily on segmentation accuracy. Song et al. [3] proposed a retrieval that required less accurate segmentation of the lung fields and it searched similar images based on grid-based feature histograms with foreground and background differentiation. Both these methods applied only to the lung fields and pulmonary hilar regions and the mediastinum. The regional lymph nodes at the hilar and mediastinum are often the first sites of disease spread in NSCLC.

In CBIR research, local approaches such as the bag-of-features method are generally used. In this approach, a set of features are computed for every pixel, or a block of pixels, then the set of features are compiled as a histogram of feature occurrences. A shortcoming of such methods lies with the histogram representation, which ignores the loca-

tion of feature bins, and they fail to identify objects based on collective information from adjacent blocks [4]. Less commonly, region-based image retrieval has been used. As regions of a certain level of homogeneity of texture and shape are often considered to represent an object in an image, region-based approaches have gained favor in problem domains where images contain some distinctive objects amidst background [4, 5].

We chose to employ region-based image retrieval due to its similarity to the normal visual perception of thoracic PET-CT images, whose characteristics are mainly described by the salient regions of the lung tumor and regional lymph nodes. However, region-based techniques are affected by segmentation accuracies. Our method was designed to reduce the effect of segmentation inaccuracies, while avoiding a complicated and time-consuming segmentation procedure. We have organized our paper as follows: a) Section 2 gives an overview of our image retrieval framework. b) Section 3 describes the clustering-based approach for region delineations. c) In section 4 we expand upon the details of similarity measure function with learning-based methods. d) In section 5 we present the experimental methods. e) The results are presented in section 6.

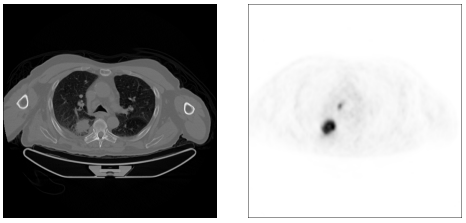
2 Overview of Our Approach

Our framework consists of: extraction of the lung and mediastinum from transaxial PET-CT image pairs, clustering-based region delineation, extraction of regional texture and structural features, and learning-based similarity measure for distance computations and image retrieval. A raw PET-CT image pair from a patient diagnosed with NSCLC is shown in Figure 1.

2.1 Lung and Mediastinum Extraction

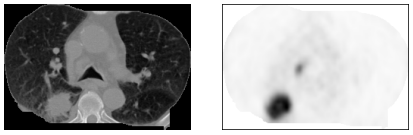
As a pre-processing step, the lung and mediastinum were extracted. The extraction process started with histogram equalization and Otsu thresholding to remove the patient bed and soft tissues outside the lung fields with simple subtraction and filling operations. Connected component analysis was then performed to approximate the contours of the lung fields. The area enclosed by the left and right lung fields was marked as the mediastinum. The resulting mask was mapped to the PET image, which was also normalized into standardized uptake values (SUV) by normalization of the FDG uptake with the injected dose and the body weight of the patient to reduce inter-patient variability [6].

Our retrieval method was insensitive to the inclusion of surrounding tissues, so precise segmentation was not necessary and the pre-processing step was simple and fast. Some components of vertebral column, ribs and muscles remained in the resulting image. Figure 2 shows the extraction output of Figure 1. We referred to the extracted areas as LM_{CT} and LM_{PET}.



(a) (b)

Figure 1. Transaxial PET-CT images from a patient with NSCLC. (a) CT (b) PET – The focal region of increase FDG uptake (large black region) reflects the primary lung tumor in the right lung; it is seen as a region of high density on CT when compared to normal lung. The smaller, central focal region of increased FDG uptake is disease in a mediastinal (right lower paratracheal) lymph node.



(a) (b)

Figure 2. Transaxial images with most of the extrapulmonary bone and soft tissues removed. (a) CT image - LM_{CT}. (b) SUV (PET) image - LM_{PET}.

2.2 Region Delineation

A modified *fuzzy c-means* (FCM) clustering approach incorporating spatial and textural information was applied to segment the LM_{CT} and LM_{PET} image pair into regions. The clustering process was also adaptive for each individual image pair, rather than adopting a set of fixed parameter settings. The method is presented in detail in Section 3.

2.3 Regional Feature Extraction

A 12-dimensional feature vector was extracted for each region and it contained the following information: (a) texture, described by mean and variance of the anisotropic diffusion filtering (ADF) filtered LM_{CT} and LM_{PET}, and the Gabor filtered ones (details of applying ADF and Gabor filters are described in section 3.1); (b) structure, given by the size and the eccentricity of the region; and (c) location, represented by the coordinates of the centroid of the region, normalized by the dimension of the image.

2.4 Similarity Measure

The similarity measure between two image pairs was equivalent to computing the distances between two sets of regional features. L1 or L2 distance was suboptimal because intuitively salient regions should be more important for matching different images. Each feature dimension of the feature vector might carry a different weight. Therefore, a learning-based similarity measure was designed,

and section 4 presents the details.

3 Clustering-based Region Delineation

To achieve an effective base for the similarity measure, the regions delineated from the lung and mediastinum had to fulfil the following criteria: (i) The tumor and disease in regional lymph nodes should form individual regions. (ii) Minor anatomical structures, such as segmental artery, need not form individual regions as they were not important for identifying similar images. (iii) The number of regions should be minimized so the possibility of mismatching regions was reduced.

Our choice of fuzzy clustering, in particular, FCM was because the boundaries of salient regions (e.g. tumors) were generally not well defined, thus rendering simple *k*-means clustering unsuitable. The expectation of having a fewer number of regions also meant that popular over-segmentation approaches, such as mean-shift, were not desired.

3.1 Fuzzy Clustering with Spatial and Textural Information

A major drawback of FCM is that the clustering is performed at pixel level, and it ignores the spatial and textural context, which makes it sensitive to statistical noise and has a tendency to produce non-compact regions [7] (Figure 4a). This violates our requirement of having regions for major anatomic structures. Therefore, we enhanced the standard FCM algorithm by incorporating spatial and textural information into the clustering process.

3.1.1 FCM Clustering

The basic form of FCM clustering minimizes the objective function [8]:

$$J_{FCM} = \sum_{i}^{c}\sum_{j}^{n} \mu_{ij}^{m}||\mathbf{x}_j - \mathbf{v}_i||^2,\ s.t. \sum_{i}^{c}\mu_{ij} = 1 \quad (1)$$

where n is the size of the image, c the number of clusters, $\mathbf{x}_j$ the jth data vector, $\mathbf{v}_i$ the ith cluster center, μ_{ij} the membership of $\mathbf{x}_j$ belonging to the ith cluster, and m the index of fuzziness (usually 2). $\mathbf{x}_j$ and $\mathbf{v}_i$ are p-dimensional vectors, where p is 1 when only gray-level image intensity was used for clustering.

The optimization problem is solved by iteratively updating the cluster centers and the membership function using the following equations until the algorithm reaches convergence:

$$\mathbf{v}_i = \frac{\sum_j^n \mu_{ij}^m \mathbf{x}_j}{\sum_j^n \mu_{ij}^m},\ i = 1, ..., c \quad (2)$$

$$\mu_{ij} = \frac{||\mathbf{x}_j - \mathbf{v}_i||^{-2/(m-1)}}{\sum_k^c ||\mathbf{x}_j - \mathbf{v}_k||^{-2/(m-1)}},\ i = 1, ..., c,\ j = 1, ..., n \quad (3)$$

Several methods have been proposed to incorporate spatial information into FCM by modifying the objective and updating functions [7, 9]. We chose to expand $\mathbf{x}_j$ and $\mathbf{v}_i$ to include both spatial and textural information, keeping the above equations relatively unchanged.

3.1.2 Spatial Information

A mean or median filtered image exhibits the local information of the pixel at the center, hence it is often used as the spatial information term in the spatial FCM [9]. To reduce the blurring effect at the boundaries of the salient regions, an edge-preserving non-linear filter, anisotropic diffusion filter (ADF) [10] was employed. The gradient magnitude was computed based on eight neighboring pixels, and the original diffusivity function favoring high-contrast edges [10] was used. This helped to reduce the noise while keeping morphological details in the images.

3.1.3 Textural Information

Gabor filtering is often applied for texture feature extraction, due to its close resemblance to human visual processing. Chen et al. suggested using scale 6, orientation 4, filter mask size 13×13, and 24 filters for the best image retrieval performance [11]; we adopted the same settings.

The 24 filtered outputs $\mathbf{G}$ were then averaged to create a single feature:

$$\mathbf{gab} = \frac{\sum_k \mathbf{G}_k \sigma_k}{\sum_k \sigma_k},\ k = 1, ..., 24 \quad (4)$$

where σ_k was the standard deviation of $\mathbf{G}_k$. The averaging method assigned higher weightings on outputs having higher variance, since such outputs presented a higher distinction among different patterns.

3.1.4 Weighted Spatial and Textural FCM Clustering

The ADF and Gabor filters were applied to both LM_{CT} and LM_{PET}, generating two spatial features and two texture features for each pixel location:

$$\mathbf{x}_j = \{CT_j^{adf}, PET_j^{adf}, CT_j^{gab}, PET_j^{gab}\}, \quad (5)$$

Since each feature could carry different weights in representing a pixel, the distance between the data vector and cluster center $||\mathbf{x}_j - \mathbf{v}_i||^2$ was changed to a weighted one:

$$d_{ij} = (\sum_{k}^{4} w_k (x_{jk} - v_{ik})^2)^{\frac{1}{2}}, \quad (6)$$

The fuzzy clustering objective function for the proposed *weighted spatial and textural* FCM (WST-FCM) was thus:

$$J_{WST-FCM} = \sum_{i}^{c}\sum_{j}^{n} \mu_{ij}^m d_{ij}^2 \quad (7)$$

and the fuzzy membership updating function Eq. (3) was similarly modified with the weighted difference.

We also leveraged the idea of variance-based weight factor assignment, as proposed by Hung et al. [8]. However, their approach determined a fixed set of weights using bootstrapping training for the entire dataset. Our method was to compute the weights adaptively on each individual image pair at clustering time, avoiding the need to analyze the empirical distributions of the dataset and making the computation more flexible for new data. The weight factors were calculated using the following equation:

$$w_k = \frac{\sigma_k}{\mu_k},\ k = 1, ..., 4 \tag{8}$$

where σ_k and μ_k were the standard deviation and mean of the kth feature of the image data $\mathbf{x}$.

3.2 Cluster Number Initialization

From our empirical study, a cluster number between 3 to 8 yielded the best clustering results. Hence, we designed an iterative pre-processing procedure to determine the number of clusters for each LM_{CT} and LM_{PET} pair adaptively as follows. The WST-FCM was executed on down-sized images iteratively with a cluster number from 3 to 8. The image was down-sized to a quarter to reduce the execution time. For each clustering output, a validity index was computed. There were a number of classical validity indices for FCM, and the index V_K [12] was found to comply closely with our criteria of region delineations. The iteration producing the minimum V_K was chosen as the cluster number.

4 Learning-based Similarity Measure

After the first three steps – lung and mediastinum extraction, region delineation and regional feature extraction – a PET-CT image pair was transformed into a set of regional features, denoted as $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N\}$, where N is the number of regions, and $\mathbf{f}_i$ was the 12-dimensional feature vector for region i (section 2.3).

Given two image pairs $\mathbf{F}^I$ and $\mathbf{F}^J$, because N^I and N^J were normally different, pair-wise similarity measure based on Euclidean distances or histogram intersection types became unsuitable. Methods to accommodate unbalanced data pairs with region-based similarity measures include: (1) The earth mover's distance (EMD) [13]. (2) Region weighted sum of minimum distances between regions (we name it RW-MRD) [14]:

$$D(I, J) = \sum_i w_i^I d_i^{IJ} = \langle \mathbf{w}^I, \mathbf{d}^{IJ} \rangle \tag{9}$$

$$d_i^{IJ} = \min_j d(\mathbf{f}_i^I, \mathbf{f}_j^J) \tag{10}$$

where w_i^I was the weight for region i, learned with a triplet learning approach [15]. (3) The integrated region matching (IRM) method [16], calculating the distances between two images by creating a multi-to-multi matching between regions:

$$D(I, J) = \sum_{i,j} s_{i,j} d_{i,j} \tag{11}$$

where $d_{i,j}$ was the distance between the two feature vectors, and $s_{i,j}$ was the *significance credit* indicating the importance of $d_{i,j}$ for determining the image similarities.

The slow performance of EMD made it less desirable for fast retrievals, but we used it as a benchmark for distance computations since it is a well-known technique for region-based similarity measure. The RW-MRD approach incorporated region weights, making salient regions contribute more to the similarity measure. However, the algorithm assumed that the region type was known to assign it a weight, which was not the case for our dataset without region identification. Secondly, the region matching between two images was one-to-one only, thus imposing a rigid requirement on segmentation accuracies. The IRM approach addressed the segmentation inaccuracy issue with its region matching scheme. However, the significance credit of region matching was area-based while important regions, such as tumors, were usually small.

Both RW-MRD and IRM omitted the design on the feature weights for distance computation between two feature vectors. Since a 12-dimensional feature vector was used, the weight of each feature dimension affected the distances between regions considerably. Therefore, we designed a learning-based similarity measure by integrating RW-MRD and IRM, and improving on the two problems mentioned above.

The similarity measure between two PET-CT image pairs $\mathbf{F}^I$ and $\mathbf{F}^J$ was thus formulated as:

$$D(I, J) = \sum_{i,j} \omega_{i,j} s_{i,j} d_{i,j} \tag{12}$$

where ω was the region weight factor, $\mathbf{s}$ was the significance credit of region matching, and $\mathbf{d}$ was the distance between two regions. The ways of computing $\mathbf{d}$, $\mathbf{s}$ and ω are discussed below.

4.1 Feature Weighted Region Distance

We describe the method to compute $d_{i,j}$, the distance between two regions $\mathbf{f}_i^I$ and $\mathbf{f}_j^J$ in this section:

$$d_{i,j} = \sum_m^M \frac{\upsilon_{im} f_{im} - \upsilon_{jm} f_{jm}}{\upsilon_{im} f_{im} + \upsilon_{jm} f_{jm}} \tag{13}$$

where m was the index into the feature vector, M was the feature dimension (12), and υ was the feature weight. The value υ was determined by both the region and feature types. For example, the texture features were more representative for a tumor, while location played an important role for background regions.

We chose to define 6 regions: background (RT1), lung lobes (RT2), mediastinum (RT3), disease in regional lymph

nodes (RT4), tumor areas with lower uptake (RT5), and tumor areas with high uptake and high density (RT6). To differentiate the region types and compute feature weights simultaneously, a multi-class *support vector machine* (SVM) problem was implied.

A multi-class SVM was trained using 20 PET-CT image pairs (about 2.3% of the dataset), which were selected as they represented the typical examples of the regions. A linear kernel was used, and a one-versus-all multi-class realization was implemented based on the binary SVM [17]. Hence, 6 linear binary SVMs were trained, and the feature weights for each region type were derived from the support vectors as follows:

$$\mathbf{v}_k = \sum_s^S \alpha_s c_s \mathbf{x}_s,\ k = 1, ..., 6 \tag{14}$$

where α, c, and $\mathbf{x}$ were the Lagrange multipliers, class labels, and M-dimensional support vectors for the kth SVM, resulting in a $6 \times M$ matrix of feature weights $\mathbf{v}$. Therefore, $v_{im} = v_{t(i)m}$ where $t(i)$ was the region type of i determined based on the SVMs trained.

4.2 Region Matching with IRM

We describe the region-matching method based on IRM [16] for computing $s_{i,j}$, the *significance credit* of the matching between two regions $\mathbf{f}_i^I$ and $\mathbf{f}_j^J$ in this section.

The essence of IRM was that rather than a one-to-one region matching between two images, a region could be matched to multiple regions in the other image with each match contributing differently to the overall image similarity value. The degree of contribution was thus the significance credit. This method was useful when the clustering approach segmented similar objects in different images into different number of regions with varying sizes. The value of significance credit was bounded by the following constraints:

$$\sum_j^{N^J} s_{i,j} = p_i,\ \sum_i^{N^I} p_i = 1 \tag{15}$$

where p_i was the percentage of image covered by region i.

To compute $s_{i,j}$, IRM employed a greedy-type algorithm with a *most similar highest priority* principle, by starting from the best matching region pair (i', j'), allocating them with a significance credit $s_{i',j'} = min(p_{i'}, p_{j'})$, and updating $p_{i'}$ and $p_{j'}$ by taking away $s_{i',j'}$, then iteratively moving through the less matching pairs until all p_i or p_j went to zero, to create a multi-to-multi region matching map. The region matching for the tumor area is illustrated in Figure 3.

4.3 Discriminative Region Weight Learning

We describe the way of generating region weights ω_i in this section. Note that region weights ω_i had entirely different

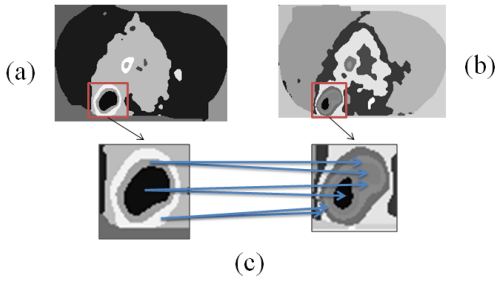


Figure 3. Region matching example for the tumor area (marked inside red box). (a) Regions of the example Figure 1. (b) Regions of another PET-CT image pair in the dataset. (c) Region matching for the tumor area shown as zoomed views from (a) and (b).

meanings from the significance credit $s_{i,j}$ in that ω_i represented the degree of importance of a region for discriminating two images, while $s_{i,j}$ indicated the importance of the match between two regions for determining similarity between images.

The value of $s_{i,j}$ was related to the sizes of the regions; however, larger regions were not necessarily more important for differentiating images. For example, when matching two images, the presence of tumors, which were usually much smaller than the normal lung lobes area, was a more important factor. Therefore, we incorporated region weights ω_i into the similarity measure.

We adapted the triplet learning framework [15] to learn the region weights ω_i. Triplet learning was based on selecting three images $\mathbf{F}^I, \mathbf{F}^J, \mathbf{F}^K$ where $\mathbf{F}^I$ and $\mathbf{F}^J$ were similar, and $\mathbf{F}^I$ and $\mathbf{F}^K$ were dissimilar. The learning algorithm thus enforced the condition: $D(I,K) > D(I,J)$. Based on the distance functions defined in Eq. (9) and (10), the condition became:

$$\begin{array}{rcl} \langle \mathbf{w}^I, \mathbf{d}^{IK} \rangle & > & \langle \mathbf{w}^I, \mathbf{d}^{IJ} \rangle \\ \langle \mathbf{w}^I, \mathbf{x}^{IJK} \rangle & > & 0 \end{array} \tag{16}$$

where $\mathbf{x}^{IJK} = \mathbf{d}^{IK} - \mathbf{d}^{IJ}$.

If constructing T such triplets as the training set, to solve for $\mathbf{w}$, a large-margin optimization with the slack variables ξ_i was used:

$$\begin{array}{ll} \min_{\mathbf{w},\xi} & \frac{1}{2}\mathbf{w}'\mathbf{w} + C\sum_i^T \xi_i \\ s.t. & \mathbf{w}'\mathbf{x}_i \geq 1 - \xi_i, \xi_i \geq 0, \mathbf{w} \geq 0 \end{array} \tag{17}$$

where C was a trade-off constant between the empirical loss ξ and regularization $\mathbf{w}$. The formulation was similar to SVM, but without the bias term, thus $\mathbf{w}$ could be efficiently solved using a dual variable iterative updating of the following two steps for the ith triplet:

$$\mathbf{w} = \max\{\sum_i^T \alpha_i \mathbf{x}_i, 0\} \tag{18}$$

$$\alpha_i = \min\{\max\{\frac{1 - \mathbf{w}'\mathbf{x}_i}{||\mathbf{x}_i||^2} + \alpha_i, 0\}, C\} \tag{19}$$

In our similarity measure function, $\omega_{i,j}$ took two possible values: $\omega_{i,j} = \omega_i$, if region i and j were of the same type (ω_i was equal to ω_j); and $\omega_{i,j} = \omega_i + \omega_j$, if region i and j are of different types, so penalties were given when images comprised apparently different regions.

To formulate our problem into a triplet learning format, the similarity measure function was rewritten into:

$$\begin{aligned} D(I,J) &= \textstyle\sum_{i,j} \omega_{i,j} s_{i,j} d_{i,j} \\ &= \textstyle\sum_{k}^{6} w_k \sum_{i,j} s_{i,j} d_{i,j} t_{i,j}^k \\ &= \langle \mathbf{w}, \mathbf{sd} \rangle \end{aligned} \tag{20}$$

where $t_{i,j}^k = 1$ if region i or j was of type k, and 0 otherwise; and $\mathbf{w}$ was associated with the region type, hence $\omega_i = w_{t(i)}$ where $t(i)$ was the region type of i. And a triplet $\mathbf{x}^{IJK}$ was thus $\mathbf{sd}^{IK} - \mathbf{sd}^{IJ}$.

Selection of training triplets followed an incremental approach, starting from a set of positive examples, then adding in triplets that did not fulfil $\mathbf{sd}^{IK} > \mathbf{sd}^{IJ}$, so that empirical errors were reduced by focusing on triplets lying close to the separation hyperplane while reducing the training time required. In total, 250 triplets were selected from our dataset (out of $870 \times 869 \times 868$ possible triplets).

5 Experimental Setup

5.1 Materials

There were 870 image pairs selected from PET-CT scans of 20 cases with NSCLC. The co-registered DICOM images were acquired using a Siemens TrueV 64-slice PET-CT scanner (Siemens, Hoffman Estates, IL).

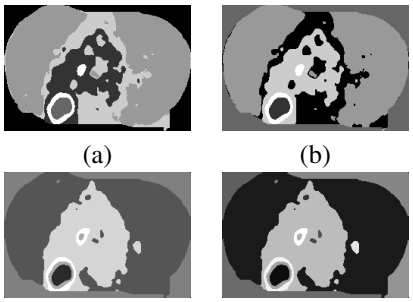
5.2 Method for Retrieval Evaluation

Given a query image, the most similar M images were retrieved. The *average precision* (AP) was taken as the performance metric for measuring the retrieval accuracy [1]:

$$AP = \frac{1}{N} \sum_{i}^{N} \frac{s_i}{M} \tag{21}$$

where s_i was the number of actual similar images returned, and N was the number of tests. All PET-CT image pairs were used as query images ($N = 870$). Each one was compared against all images from the other 19 cases, ensuring results were not skewed to adjacent slices of the same case. The average precision was evaluated for retrieving 4 and 8 images ($M = 4$ and $M = 8$).

The criteria of determining the similarity of a retrieved image were: (a) If the query image contained a tumor, the return result should contain a tumor of similar texture and structure. (b) If the query image contained disease in lymph nodes, the return result should contain similar node appearances. (c) If (a) and (b) above were met, images showing similar lung field structures should get higher ranking.



(a) (b) (c) (d)

Figure 4. Clustering and regions formulated. (a) Clustering output of standard FCM. (b) Clustering output of non-weighted FCM. (c) Clustering output of WST-FCM. Number of clusters is 6. (d) Regions formulated. Number of regions is 19.

To evaluate the effectiveness of our proposed method (i.e. RFW-IRM – region weighted similarity measure based on IRM with feature weights for each region type), we compared it to the following region-based approaches: FW-IRM – feature weighted IRM; IRM – the IRM technique (Eq. (11)); RW-MRD – region weighted minimum region distances (Eq. (9) and (10)); RFW-MRD – feature-weight enhanced RW-MRD; and EMD – direct earth mover's distance. These approaches were the base techniques that we designed our method upon.

We also compared our method to three other well-known non-region based techniques: HIST – a 256 level histogram generated from both LM_{CT} and LM_{PET}, and Euclidean distances between the histograms as the similarity measure; GLCM – 4-directional co-occurrence matrix and Haralick features in place of histograms; and BOF – bag-of-features generated by classifying LM_{CT} and LM_{PET} into 6 feature categories using *k*-means clustering.

6 Results

6.1 Illustration of Intermediate Results

An example comparing the clustering performance of (a) standard FCM, (b) FCM with spatial and textural information but no weightings, and (c) our proposed clustering approach WST-FCM, for the example image given in Figure 2, is shown in Figure 4. The output from the WST-FCM algorithm depicted a closer approximation of actual salient regions; and was closer to our expectation of the region delineation. Figure 4d shows the final region delineation result after each connected component was identified as an individual region. Figure 5a shows the region types identified for Figure 4b. The feature weights learned based on region type differentiation are illustrated in Figure 5b. Figure 6 shows the weights of the 6 types of regions from our training results.

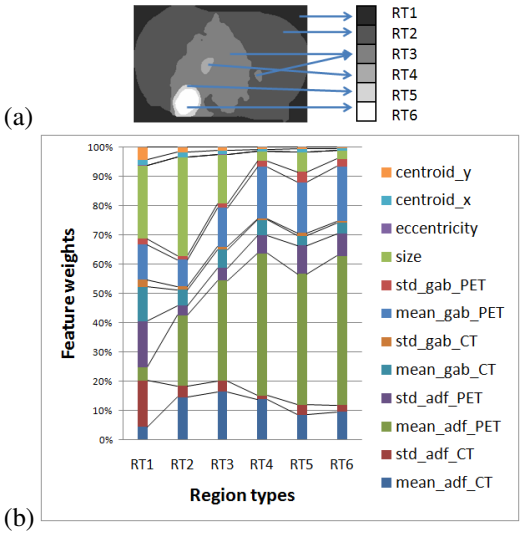


Figure 5. Region types and feature weights. (a) Region types determined for Figure 4d. (b) Feature weights learned for the 6 types of regions – **v**.

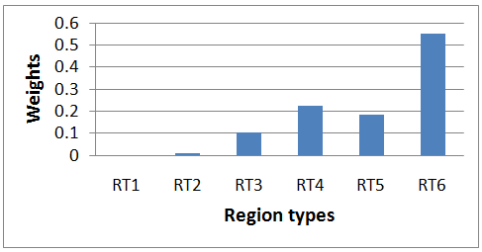


Figure 6. Region weights – **w**.

6.2 Results of Retrievals

The average precision results are illustrated in Table 1 and Figure 7. Our method (RFW-IRM) achieved the highest accuracy, which was 88.5% when retrieving 4 most similar images, and 83.6% for 8 similar images. This was on average about 8% and 12% improvement over the RW-MRD and IRM methods that we used as the basis algorithms, illustrating the advantage of incorporating IRM and region weights into the similarity measure. By comparing the results of our method with RFW-MRD, the advantage of using an IRM type of region distance computations was shown. Performance enhancement from incorporating region and feature weights was clearly indicated by comparing the results of RFW-IRM, FW-IRM and IRM. All region-based methods, except EMD, obtained higher average precisions over the other global or local image retrieval techniques – HIST, GLCM and BOF.

RFW-IRM	FW-IRM	IRM
88.5%/83.6%	82.9%/79.1%	76.3%/72.9%
RW-MRD	RFW-MRD	EMD
80.6%/75.2%	84.3%/80.2%	65.2%/64.4%
HIST	GLCM	BOF
68.4%/67.1%	42.9%/45.2%	58.6%/57.1%

Table 1. Average precision when returning 4 or 8 similar images – (AP of top-4)/(AP of top-8)

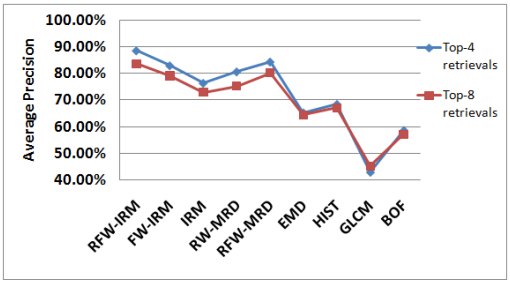


Figure 7. Average precision when returning 4 or 8 similar images.

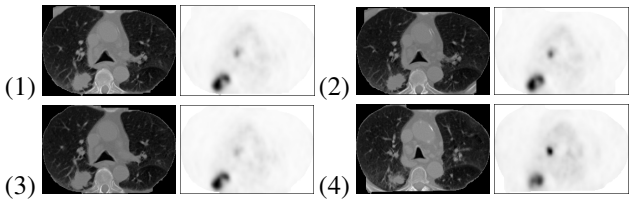


Figure 8. The top 4 matches from other patients, when compared with the query image pair from Figure 1.

Figure 8 displays the retrieval results for the example image in Figure 1. The query image was compared against all the images from the other 19 cases (818 PET-CT image pairs), and the four best matching image pairs were retrieved. The returned images exhibit similar tumor, lymph node, and lung field characteristics in decreasing order of similarity.

We performed another test by retrieving one best matching image from each case (excluding the query image from the case it belongs to), thus 20 retrieval results for one query image; then ranking the results according to the similarity level. Figure 9 shows the 4 top ranked images retrieved when compared to the case in Figure 1. The 4th image pair was the least similar and had different tumor characteristics; but this was a better match than the remaining 16 results. This test shows the effectiveness of ranking among retrieval results, when the retrieved images exhibited diverse patterns.

7 Conclusion

We have presented a region-based image retrieval framework for thoracic PET-CT images. The framework can retrieve images with similar characteristics to a primary lung tumor and with disease in regional lymph nodes. Our method does not require precise segmentation. We incorporated spatial and textural information with variance-based weights into the standard FCM clustering. The lung and

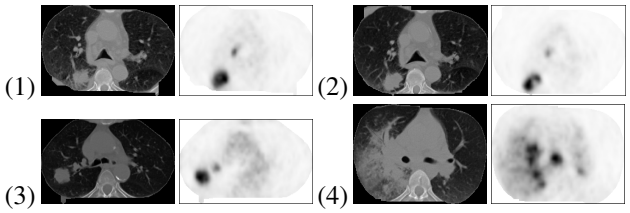


Figure 9. The top 4 matches from 4 cases, when compared with the query image pair from Figure 1.

mediastinum were divided into regions of various sizes, and each region was represented by a 12-dimensional feature vector comprising texture, structure and location features. We show that feature weights and region weights improve the retrieval accuracy; a multi-to-multi region matching provides higher average precision; and region-based methods perform better than the histogram, co-occurrence matrix and bag-of-features methods.

8 Acknowledgment

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

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*, 73, 2004, 1-23.

[2] J. Kim, L. Constantinescu, W. Cai, and D. Feng, Content-based dual-modality biomedical data retrieval using co-aligned functional and anatomical features, *Proc. MICCAI 2007 Workshop*, Brisbane, Australia, 2007, 45-52.

[3] 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*, Sydney, Australia, 2010.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang, Image retrieval: ideas, influences, and trends of the new age, *ACM Computing Surveys*, 40(2), 2008.

[5] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, 40, 2007, 262-282.

[6] 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*, 2010.

[7] S. Belhassen, and H. Zaidi, A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET, *Medical Physics*, 37(3), 2010, 1309-1324.

[8] W. L. Hung, M. S. Yang, and D. H. Chen, Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation, *Pattern Recognition Letters*, 29, 2008, 1317-1325.

[9] W. Cai, S. Chen, and D. Zhang, Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation, *Pattern Recognition*, 40, 2007, 825-838.

[10] P. Perona, and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans Pattern Anal Mach Intell*, 12, 1990, 629-639.

[11] L. Chen, G. Lu, and D. Zhang, Effects of different Gabor filter parameters on image retrieval by texture, *Proc. Int Multimedia Modeling Conference*, Brisbane, Australia, 2004, 273-278.

[12] S. H. Kwon, Clustering validity index for fuzzy clustering, *Electronics letters*, 34(22), 1998, 2176-2177.

[13] H. Greenspan, G. Dvir, and Y. Rubner, Region correspondence for image matching via EMD flow, *Proc. Workshop on content-based access of image and video libraries*, Hilton Head Island, USA, 2000, 27-31.

[14] C. Gu, J. J. Lim, P. Arbelaez, and J. Malik, Recognition using regions, *Proc. CVPR*, Miami, USA, 2009, 1030-1037.

[15] A. Frome, Y. Singer, and J. Malik, Image retrieval and classification using local distance functions, *Proc. NIPS*, Vancouver, Canada, 2006, 417-424.

[16] J. Li, J. Z. Wang, and G. Wiederhold, IRM: Integrated region matching for image retrieval, *Proc. 8th ACM Conf. on Multimedia*, Los Angeles, USA, 2000, 147-156.

[17] C. Cortes, and V. Vapnik, Support-vector networks, *Machine learning*, 20(3), 1995, 273-297.