

Classification of Thresholded Regions based on Selective Use of PET, CT and PET-CT Image Features

Lei Bi, *Student Member, IEEE*, Jinman Kim, *Member, IEEE*, David Dagan Feng, *Fellow, IEEE* and Michael Fulham

Abstract— Fluorodeoxyglucose positron emission tomography – computed tomography (FDG PET-CT) is the preferred image modality for lymphoma diagnosis. Sites of disease generally appear as foci of increased FDG uptake. Thresholding methods are often applied to robustly separate these regions. However, its main limitation is that it also includes sites of FDG excretion and physiological FDG uptake regions, which we define as FEPUs – sites of FEPUs include the bladder, renal, papillae, ureters, brain, heart and brown fat. FEPUs can make image interpretation problematic. The ability to identify and label FEPUs sites and separate them from abnormal regions is an important process that could improve image interpretation. We propose a new method to automatically separate and label FEPUs sites from the thresholded PET images. Our method is based on the selective use of features extracted from data types comprising of PET, CT and PET-CT. Our FEPUs classification of 43 clinical lymphoma patient studies revealed higher accuracy when compared to non-selective image features.

I. INTRODUCTION

Lymphoma accounts for about 8% of all cancer cases [1-2]. Fluorodeoxyglucose (FDG) positron emission tomography – computed tomography (FDG PET-CT) is considered the modality of choice for image-based staging and diagnosis of the lymphoma as sites of disease generally display increased FDG uptake [3-4].

A semiquantitative measure of FDG uptake, referred to as the standard uptake value (SUV), relates radiotracer concentration normalised by patient weight. A SUV is commonly used to describe sites of abnormal glucose metabolism when compared to normal structure [5]. On whole-body PET-CT studies, high SUVs are seen in the papillae of the ureters, the kidneys and the bladder due to FDG excretion. In addition, high values can be seen in the normal myocardium and the brain, which only includes glucose of energy metabolism. SUV thresholding is the most common method to detect sites of abnormal FDG uptake in lymphoma patients [4]. A number of investigators have proposed methods to calculate the threshold to separate normal from abnormal FDG uptake such as $50\%SUV_{max}$ and $SUV=2.5$ [6]. A consequence of these threshold methods is that they are applied globally to the entire image and so FDG

excretion and physiologic uptake, which we refer to as FEPUs, are identified alongside sites of disease (see Figure 1a). FEPUs makes image interpretation problematic and distracting for imaging specialists and clinicians. Further, regions of FEPUs can obscure sites of diseases in adjacent lymph nodes.

We propose a new classification method to automatically label sites of FEPUs from the thresholded whole-body PET images. Our method is based on region-level statistics and spatial features (RSS) and histogram of oriented gradients (HOG) features [7] extracted from PET, CT and combined data type PET-CT. Our method differs from traditional methods [8-10] in that it uses all the available features and there is an automated ‘selective’ use of the image features.

A. Related Work

Our work relates to image classification techniques that attempt to separate and label different structures based on automatically derived image features from CT and PET-CT data [8-10]. Venkatraghavan et al [8] used Gabor filtered CT slices with speeded up image feature extraction (SURF – a fast implementation of scale-invariant feature transform (SIFT)) for use in organ localisation. However, this investigation only considered ‘healthy’ structures thereby bypassing the complexity from abnormal structures. Pescia et al [9] used texture-based image features to detect liver tumours on CT studies via an AdaBoost classifier. Song et al [10] proposed a method to identify the lungs, mediastinum, lung tumours and lymph nodes with spatial and texture features on PET-CT studies. Both of these studies were applied to specific sections of the body i.e., the liver in [9] and the thorax in [10], which enables structure specific optimisation and semantic definition to aid in the classification. Further, all these methods relied on using all the available features rather than selecting only the relevant features to classify a structure from all the features.

II. METHODS AND MATERIALS

A. Materials and Ground Truth Construction

Our dataset consisted of 43 whole-body PET-CT studies from 11 patients diagnosed with lymphoma; each patient had multiple scans (1 patient with 6 scans; 7 patients with 4 scans; 3 patients with 3 scans) during the treatment. All studies were acquired using a Biograph TruePoint PET-CT scanner (Siemens Medical Solutions, Hoffman Estates, IL, USA) with a PET resolution of 168×168 pixels at 4.07mm^2 and CT resolution of 512×512 pixels at 0.98mm^2 and slice thickness of 3mm. The bed and linen were removed from CT by adaptive thresholding and image subtraction from a bed template [11].

This research was funded by Australian Research Council (ARC) grants.

L. Bi, J. Kim D. Feng and M. Fulham are with the School of Information Technologies, University of Sydney, NSW, Australia. L. Bi (e-mail: lei.bi@sydney.edu.au); J. Kim (e-mail: jinman.kim@sydney.edu.au).

D. Feng is also with Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China (e-mail: dagan.feng@sydney.edu.au).

M. Fulham is also with Department of Molecular Imaging, Royal Prince Alfred Hospital, NSW, Australia and Sydney Medical School, University of Sydney, NSW, Australia (e-mail: michael.fulham@sydney.edu.au).

Training and ground truth data were constructed using PET Response Criteria in Solid Tumours (PERCIST) [6] thresholding method and refinements on each PET image (see Section II.B). The resulting binary mask comprising of the FEPUs, was then manually labelled as belonging to the brain, bladder, heart, left kidney, right kidney or other structures, as exemplified in Fig. 1(a). The “other” class contained potential abnormalities (based on patients’ clinical report), regions of physiological effects, brown fat and lymph node inflammations. A total of 730 thresholded regions were manually labelled which included 54 brain, 41 bladder, 40 heart, 112 left kidney, 123 right kidney and 360 other regions.

B. Automatic PERCIST Thresholding and Skeletal Bone Detection

PERCIST is a robust method for calculating SUV threshold based on the combined use of SUV normalised with lean body mass (SUV_{LBM}) together with a reference volume of interest (VOI) [12-14] – a 3cm diameter sphere placed on the right lobe of the liver to measure the average FDG. We considered the sites that are above this threshold value will be the FEPUs for this work. We adopted our previous method to automatically calculate PERCIST based threshold value [15]. The lungs were coarsely estimated from CT to localise the liver via adaptive thresholding [16]. Then, we used a multi-atlas based segmentation method on the CT to segment the liver. A multi-atlas method was used for its robustness on low-contrast CT images [17]. From the segmented structure, a VOI reference was placed on the right lobe of the liver, which was then mapped to the co-registered PET. The VOI was then used in the threshold calculation. Finally, a binary mask $T_{PERCIST}$ was generated after applying the threshold value on the PET. For some studies, skeletal structures are included in the thresholded results; these structures were removed by using the skeletal structures segmentation from the counterpart CT. A binary skeletal mask $T_{Skeletal}$ was generated using a threshold of >150 Hounsfield Units (HU) [18] on the CT image. $T_{Skeletal}$ was then subtracted from $T_{PERCIST}$. A morphological filter was then applied on the resulting binary mask to remove noise.

C. Contextual Features Extraction

Two types of features were extracted and combined to form the contextual features, which include region-level statistics and spatial (RSS) features and histogram of oriented gradients (HOG) features [7]. RSS features were used to describe the regions in a descriptive statistical manner representing a likelihood of a region at a location in the image. Mean, standard deviation, skewness and kurtosis on each region from both PET and CT images were calculated. We also measured the average location of each region in transverse, coronal and sagittal planes (represented in a percentage format) to add spatial context to our features. HOG features were used to describe the shape of the region. Recent studies have demonstrated that HOG features better differentiate structures on PET and CT images [19]. HOG features count the occurrences of gradient orientation in a cell. Felzenszwalb et al [20] set cell size equal to 8 and to have 9 directions in each cell and we used the same approach for this work. HOG features were sampled by using the cell over the regions and summarised it via a bag-of-words

(BoW) histogram. A bin size of 1500 was set for PET and CT, empirically. We normalised the FDG values into SUV with patient’s lean body mass (SUV_{LBM}) to reduce the variability of FDG values among patient.

D. Classification via Weighted Combination of PET, CT and PET-CT

We used multi-class support vector machines (SVM) [21] on RSS and HOG features to classify the regions. Before these features were placed into SVM, we trained these features with a radial basis function (RBF) kernel [22] to non-linearly map the data into a higher dimension space. This helps to make the training data more separable in a computationally efficient way, where a linear kernel usually has poor performance in a non-linear classification task while a polynomial kernel is computationally expensive [23, 24]. The RBF kernel is given by:

$$RBF(x, x') = e^{-\frac{\|x-x'\|_2^2}{2\sigma^2}}. \quad (1)$$

Here, x and x' represent two feature vectors and σ is a Gaussian value, which measures the standard deviation of the data. The probability score of a region r for each data type is calculated as the weighted combination of all the features, which is defined as:

$$\mathbf{P}_m(r) = \sum_{f \in F} w_f \cdot \rho_f(r) \quad (2)$$

$$m \in \{PET, CT, PET - CT\}, f \in F = \{f_{RSS}, f_{HOG}\}$$

where f denotes the feature and m denotes the data type. w_f represents feature associated weight. $\rho_f(r)$ is a probability matrix with input feature f for region r . It is calculated from the output of the SVM. The final labelling of r is based the weighted maximum probability score from all three data types (PET, CT and PET-CT), calculated as:

$$arg \max_l \{w_m(l) \cdot \mathbf{P}_m(r)\} \quad (3)$$

l denotes the label and w_m is a weight matrix, where the row and column represents the data types and the labels. In this study, the optimal σ , w_f and w_m were empirically determined from the training data via cross-validation that has the highest cross-validation accuracy.

III. RESULTS AND DISCUSSION

We considered a region to be classified correctly if it was labelled as the same as the ground truth. We randomly partitioned the data into two sets: 5 and 6 patient studies. Initially, the first set was used to construct the training set and the second set was used for evaluation. Then we reversed the role of the two sets. We compared our proposed method with the classification results from using PET, CT or PET-CT. Our results are shown in Table I.

The low classification accuracy among all the methods when classifying the heart was expected as the myocardium is usually separated into many regions (one primary and many small regions). The smaller regions usually interference with lymph node inflammations in mediastinum and were difficult to predict as these regions do not have consistent structure and location. As presented in the table, compared with the first three methods, our method had the best performance. We suggest that this is because our method is

able to identify the important image features from a particular data type and hence appoint a higher weight to it. This results in the selective usage of the best classification results from all methods. Compared to CT data, PET and PET-CT perform better in predicting high uptake and excretion regions such as the heart, bladder and brain, which relates to their relatively higher normal FDG uptake. Compared with PET and PET-CT, CT has higher classification accuracy in predicting “other” class. It is likely to be the “other” class involves many different structures and potential abnormalities in various locations; the use of CT data helps to localise anatomical information, which resulted in better classification accuracy. The left and right kidney regions are consistently performed well in all data types. This is attributed to the fact that the left and right kidneys have relatively unique spatial features compared with other structures. Figure 1 shows a classification result for a patient study. Here, PET and PET-CT data types misclassified kidney regions while CT data type alone misclassified bladder region, where our proposed method can make the best use of all three data types and produced better results.

As a further demonstration of the influence from using selective features from classification, we measured the distribution of each data types to be used for final labelling (selection) by our method. Table II presents these results, e.g., for all the regions are classified as brain, 70.37% is using PET while 7.41% and 22.22% are using CT and PET-CT.

TABLE II. THE DISTRIBUTION OF EACH DATA TYPE (PET, CT, PET-CT) TO USED IN LABELLING.

Label	Distribution (%)		
	PET	CT	PET-CT
Other	19.17	48.61	32.22
Brain	70.37	7.41	22.22
Bladder	12.20	2.44	85.37
Heart	40.00	12.50	47.50
L.Kidney	12.50	31.25	56.25
R.Kidney	69.92	20.33	9.76

IV. CONCLUSION AND FUTURE WORK

We proposed a new classification method to classify and label the FEPU regions automatically via selective use of image features derived from three data types of PET, CT and PET-CT. Our experiments with 43 clinical PET-CT lymphoma studies resulted in higher accuracy when compared to conventional methods from using only PET, CT or combined PET-CT image features. In our future studies, we will evaluate our method on greater number of clinical studies and disease types.

ACKNOWLEDGMENT

We would like to thank our collaborators at the Department of Molecular Imaging, at the Royal Prince Alfred (RPA) Hospital. This research was funded by Australian Research Council (ARC) grants.

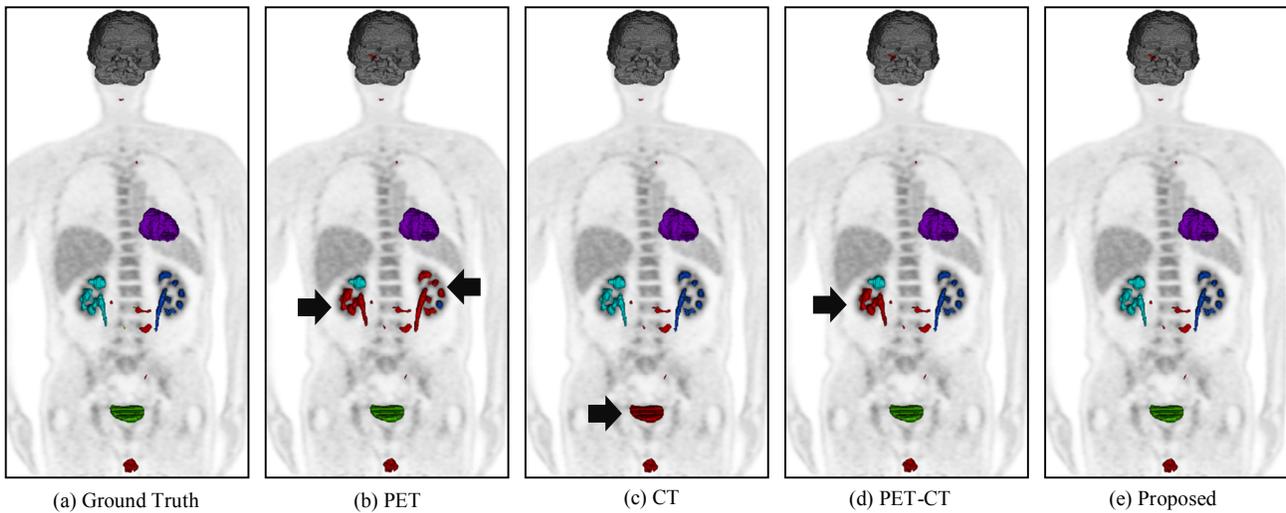


Figure 1. Whole-body FDG PET-CT scan in a patient with lymphoma. (a) is the ground truth and (b) – (e) are the results of using image features derived from PET, CT, combined PET-CT and our method that selected the best results from (b) to (d). Grey, green, purple, dark blue, light blue and red labels represent brain, bladder, heart, left kidney, right kidney and “other” classes. Black arrows indicate the misclassified regions and labels are visualised using direct volume rendering on PET.

TABLE I. CLASSIFICATION RESULTS OF OUR PROPOSED METHOD AND COMPARISON RESULTS OF USING PET, CT AND PET-CT.

Methods (Overall Prediction)	Ground Truth	Prediction (%)					
		Other	Brain	Bladder	Heart	Left Kidney	Right Kidney
PET (87.26%)	Other	88.06	9.44	0.28	-	1.39	0.83
	Brain	22.22	77.78	-	-	-	-
	Bladder	9.76	-	90.24	-	-	-
	Heart	27.50	-	-	57.50	15.00	-
	L.Kidney	8.93	-	0.89	-	90.18	-
	R.Kidney	4.88	-	-	-	-	95.12
CT (85.89%)	Other	92.78	2.50	1.11	0.28	1.39	1.94
	Brain	22.22	55.56	22.22	-	-	-
	Bladder	26.83	-	73.17	-	-	-
	Heart	30.00	-	35.00	30.00	5.00	-
	L.Kidney	6.25	-	-	-	93.75	-
	R.Kidney	4.07	-	1.63	-	-	94.31
PET-CT (88.36%)	Other	90.83	4.17	1.39	0.56	2.50	0.56
	Brain	20.37	79.63	-	-	-	-
	Bladder	4.88	-	95.12	-	-	-
	Heart	37.50	-	10.00	52.50	-	-
	L.Kidney	8.04	0.89	-	-	91.07	-
	R.Kidney	6.50	-	0.81	-	0.81	91.87
Proposed (90.55%)	Other	93.06	3.89	0.83	-	1.11	1.11
	Brain	20.37	79.63	-	-	-	-
	Bladder	4.88	-	95.12	-	-	-
	Heart	30.00	-	7.50	57.50	5.00	-
	L.Kidney	6.25	-	0.89	-	92.86	-
	R.Kidney	4.07	-	0.81	-	-	95.12

REFERENCES

- [1] Yan, J., et al. "Marker-controlled watershed for lymphoma segmentation in sequential CT images." *Med. Phys.*, 2452-60, 2006.
- [2] Xu, J., et al. "Automated temporal tracking and segmentation of lymphoma on serial CT examinations." *Med. Phys.*, 5879-86, 2011.
- [3] Freudenberg, L. S., et al. "FDG-PET/CT in re-staging of patients with lymphoma." *Eur. J. Nucl. Med. Mol. I.*, 325-9, 2004.
- [4] Moule, R. N., et al. "Adaptive^{18F} Fluoro-2-deoxyglucose Positron Emission Tomography/Computed Tomography-based Target Volume Delineation in Radiotherapy Planning of Head and Neck Cancer." *Clin. Oncol.*, 364-71, 2011.
- [5] Graham MM, et al., "Comparison of simplified quantitative analyses of FDG uptake." *Nucl. Med. Biol.*, 647-55, 2000.
- [6] Wahl, RL., et al. "From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors." *J. Nucl. Med.*, 122S-50S, 2009.
- [7] Dalal, N., et al. "Histograms of oriented gradients for human detection." In: *CVPR*, 2005.
- [8] Venkatraghavan, V., et al. "Semantic analysis of 3D anatomical medical images for sub-image retrieval." In: *MBCR-CDS*, 139-51, 2012.
- [9] Pescia, D., et al. "Automatic detection of liver tumors." In: *ISBI*, 2008.
- [10] Song, Y., et al. "Discriminative pathological context detection in thoracic images based on multi-level inference." In: *MICCAI*, 191-8, 2011.
- [11] Kim, J., et al., "A fully automatic bed/linen segmentation for fused PET/CT MIP rendering." *J. Nucl. Med.*, 49(Suppl 1), 387P, 2008.
- [12] Bi, L., et al. "Cellular Automata and Anisotropic Diffusion Filter based Interactive Tumor Segmentation for Positron Emission Tomography," In: *EMBC*, 2013.
- [13] Bi, L., et al. "Automatic Descending Aorta Segmentation in Whole-Body PET-CT Studies for PERCIST-Based Thresholding," *DICTA*, vol., no., pp.1-6, 3-5, 2012.
- [14] Niyazi, M., et al. "Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT." *Radiat. Oncol.*, 180, 2013.
- [15] Bi, L., et al. "Automated and Robust PERCIST-based Thresholding framework for whole body PET-CT studies," In: *EMBC*, 2012.
- [16] Hu, S., et al. "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images." *IEEE. T. Med. Imaging.*, 490-8, 2001.
- [17] Li, C., et al., "Automated PET-guided liver segmentation from low-contrast CT volumes using probabilistic atlas," *Comput. Meth. Prog. Bio.*, 2012.
- [18] Manniesing, R., et al. "Level set based cerebral vasculature segmentation and diameter quantification in CT angiography." *Med. Image. Anal.*, 200-14, 2006.
- [19] Song, Y., et al. "A multistage discriminative model for tumor and lymph node detection in thoracic images." *IEEE. T. Med. Imaging.*, 1061-75, 2012.
- [20] Felzenszwalb, PF., et al. "Object detection with discriminatively trained part-based models." *IEEE. T. Pattern. Anal.*, 1627-45, 2010.
- [21] Chang, C.C., et al., "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, 2011.
- [22] Burges, CJC., "A tutorial on support vector machines for pattern recognition," *Data. Min. Knowl. Discov.*, 2(2):121-67, 1998.
- [23] Kakar, M., et al., "Automatic segmentation and recognition of lungs and lesion from CT scans of thorax," *Comput. Med. Imag. Grap.*, 33(1): 72-82, 2009.
- [24] Bi, L., et al. "Multi-stage Thresholded Region Classification for Whole-Body PET-CT Lymphoma Studies," In: *MICCAI*, pp569-576, 2014.