A post-acquisition standardization method for positron emission tomography images

Mortazi, Aliasghar, Udupa, Jayaram, Tong, Yubing, Torigian, Drew

Aliasghar Mortazi, Jayaram K. Udupa, Yubing Tong, Drew A. Torigian, "A post-acquisition standardization method for positron emission tomography images," Proc. SPIE 11314, Medical Imaging 2020: Computer-Aided Diagnosis, 113143U (16 March 2020); doi: 10.1117/12.2550423

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

A Post-Acquisition Standardization Method for Positron Emission Tomography Images

Aliasghar Mortazi[1], Jayaram K. Udupa[1]*, Yubing Tong[1], Drew A. Torigian[1]

[1]Medical Image Processing Group, 602 Goddard building, 3710 Hamilton Walk, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, United States.

*Corresponding author

Abstract

We propose a new standardization method to mitigate the undesired effects of factors such as body weight, duration of radiotracer uptake period, partial volume effects, etc., that impede accurate quantitative analysis of PET images. We introduce new metrics for evaluation and comparison among different standardization methods. The proposed standardization method can improve disease quantification from PET images, as PET is one of the major clinical imaging tools used to evaluate patients with cancer and other disorders. It is based on the MRI standardization method proposed by Nyul et al. [8, 9], but modified for application to PET images. A new approach is proposed to remove the unwanted effect of the tail of image histograms by finding an optimal maximum percentile from the coefficient variations of metabolic activity of reference healthy organs. We show that after applying our standardization method, the coefficient of variation among the mean metabolic activities of healthy organs decreases compared to that obtained from non-standardized images.

Key words: Positron Emission Tomography, FDG-PET/CT, Standardizing image intensity, Image analysis.

1. Introduction

Cancer is the second most common cause of death in the United States and is a significant health problem worldwide. In 2019, about 1.8 million new cancer cases and about 0.6 million cancer deaths were reported in the United States [1]. Positron emission tomography (PET), a non-invasive molecular imaging modality, has become one of the major clinical tools for staging and response assessment in patients with cancer as well as in non-neoplastic conditions [2, 3]. For example, abnormal changes in tissue metabolic activity can be detected with 18F-fluorodeoxyglucose (FDG)-PET imaging before structural changes are detectable with computed tomography (CT) or magnetic resonance imaging (MRI). As such, metabolic activity measured from FDG-PET is an important biomarker that is clinically utilized for diagnostic, staging, prognostication, and treatment response assessment purposes in patients with cancer [4, 5].

The metabolic activity of tissues measured on FDG-PET images can be used as an early marker to distinguish between healthy and diseased tissues. However, the metabolic activity measured in PET images can be affected by a large number of factors such as body weight, body composition, duration of radiotracer uptake, partial-volume effects, etc. [6]. These factors make disease quantification challenging in PET images. Different standardization methods have been proposed during the last several decades to remove the effect of these unfavorable factors in order to decrease inter-reader variability and to improve diagnostic performance of study interpretation. The standardized uptake value (SUV) is the most commonly utilized measure of tissue metabolic activity in PET images in clinical practice, as it mitigates some effects of the above-described undesirable factors and is easy to use, although SUV does not compensate for all of these factors, potentially leading to errors in disease characterization and quantification.

Medical Imaging 2020: Computer-Aided Diagnosis, edited by Horst K. Hahn, Maciej A. Mazurowski, Proc. of SPIE Vol. 11314, 113143U · © 2020 SPIE
CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2550423

Disease quantification and diagnosis is easier and more accurate by having knowledge about the metabolic activity of normal healthy organs [7]. Several methods have been proposed for PET or SUV standardization. Some of these methods are performed at the scan acquisition level (e.g., by using a phantom, by modifying image reconstruction, by standardizing the parameters of scan image acquisition (slice thickness, table position time, etc.)). In some other methods, the standardization is performed at the patient level by controlling and correcting blood glucose levels, restricting the amount of radiotracer dose administered, restricting the allowable delay time for radiotracer uptake, etc. In yet other methods, standardization is performed at the image post-processing level by using various methods such as digital PET phantoms. Standardization methods developed for magnetic resonance imaging (MRI) such as the use of z-scores and various transformation methods can also be employed for PET standardization, although they have not been adapted to PET images. Yet, many of these methods have limitations or are impractical to use in daily clinical practice.

In this paper, we propose a standardization method for PET images similar to a previously described method developed for MRI image standardization [8, 9]. Our proposed method is independent from imaging, injection, and patient parameters and can remove undesired and adverse effects of those parameters. Also, we introduce new metrics enabling evaluation and comparison between different PET standardization strategies.

2. Methods

Data sets

This retrospective study was conducted following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act waiver. The following data sets were utilized for this study.

The first data set includes whole-body FDG-PET/CT scans with normal-appearing livers and spleens (as verified by a board-certified radiologist (co-author Torigian)) from 17 women (mean age 69, range 52-85 years) previously acquired on a Biograph mCT scanner (Siemens Healthcare, Erlangen, Germany). The second set includes PET/CT scans from 21 men (mean age 44, range 30-50 years) with normal-appearing livers and spleens (as verified by a board-certified radiologist (co-author Torigian)) previously acquired on a Gemini TF scanner (Philips Center, Amsterdam, The Netherlands). These 38 scans were acquired approximately 60 minutes after intravenous administration of approximately 15 mCi of FDG.

Standardizing PET Images

Let $\mathcal{I} = \{I_1, I_2, ..., I_N\}$ be a set of 3D PET images of a body region B. For any image I in $\mathcal{I}$, let its minimum and maximum intensities be denoted by min(I) and max(I), and let $p_{min}(I)$, med(I), and $p_{max}(I)$ denote a "low" percentile, 50^{th} percentile (median value), and a "high" percentile intensities of I, respectively. We denote I's SUV image by I_S. We will come back to the meaning of "low" and "high" later on. The proposed standardization method is based on first defining a standardized intensity scale. This involves finding landmark intensities $p_{min}(I)$, med(I), and $p_{max}(I)$ on the histograms of each image in a set $\mathcal{I}_c$ of PET images used for calibration, and finding the mean over the images in $\mathcal{I}_c$ of each of the three landmark intensities. Our intent is that the images in $\mathcal{I}_c$ should be normal. Subsequently, for any given image I in any set $\mathcal{I}_T$ (not necessarily normal) to be standardized, the same landmarks are determined in I, the mapping that results when the landmarks of I are matched to the mean landmarks estimated from the calibration images is computed, and I's intensities are transformed according to the mapping. The process is explained in the following sections. In our experiments, we use part (~60%) of the 38 healthy scans from data set $\mathcal{I}$ described above to constitute $\mathcal{I}_c$ and the remainder of $\mathcal{I}$ to constitute $\mathcal{I}_T$.

First, the three landmarks are estimated in all images in $\mathcal{I}_c$. Note that the median intensity med(I) is estimated for the foreground (body region) only after removing background voxels as described in [8]. Then, for the three landmark intensities, their mean values, denoted respectively by S_{min}, S_m, S_{max}, over the images in $\mathcal{I}_c$ are determined. These values define landmarks on the standard intensity scale. In the interval $[S_{min}, S_{max}]$, intensities in a test image $I \in \mathcal{I}_T$ get mapped in a non-linear manner as follows. The intensities of I in $[p_{min}(I), med(I)]$ are mapped linearly to $[S_{min}, S_m]$, and similarly intensities of I in $[med(I), p_{max}(I)]$ are mapped to $[S_m, S_{max}]$. Finally, I's intensities in $[min(I), p_{min}(I)]$ and $[p_{max}(I), max(I)]$ are mapped, respectively, following the mappings from $[p_{min}(I), med(I)]$ to $[S'_{min}, S_{min}]$

and [med(I), $p_{max}(I)$] to [S_{max}, S'_{max}] (see Figure 1). In our application, typically $p_{min}(I) = min(I)$ (after removing the background). However, $p_{max}(I) \neq max(I)$, and it plays a vital role in standardization since the maximum intensities in PET images are extremely variable and they pose serious challenges to intensity standardization for quantitative analysis as well as for visualization purposes such as 3D rendering via maximum intensity projection.

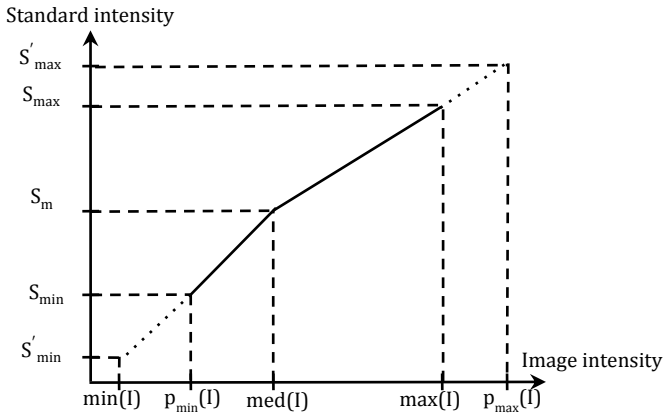

Figure 1: Mapping a given image to standard scale.

The reason for choosing $p_{max}(I) \neq max(I)$ is that the upper tail of the histogram of I is affected by artifacts, outlier intensities, and very high uptake values due to disease which cause significant variation among subjects and scanners. As we show in this paper, such variations in PET can lead to undesired SUV variations among healthy organs from different subjects and scanners. Following [8, 9] to solve this problem, we use $p_{min}(I)$ and $p_{max}(I)$ as landmarks such that only within the interval [$p_{min}(I)$, $p_{max}(I)$] do we seek to uniformize intensity meaning across subjects. This implies that normal tissues and organs are least affected by non-standardness after standardization, although outlier intensities in the histogram upper tail are transformed faithfully. It is important to note that S'_{min} and S'_{max} get calculated based on the obtained landmarks for the image to be transformed and not based on the calibration data set.

One of the main challenges in this standardization method is to find the appropriate maximum percentile $p_{max}(I)$ for removing the adverse effect from the tail of the histogram in the standardization procedure. In order to find the appropriate $p_{max}(I)$, deviating from [8], we utilize a different method as follows. We define the mean metabolic activity MMA(O) derived from the SUV image I_S of an object O as:

$$MMA(O) = \frac{TMA(O)}{\sum_v |v| \times FM_O(v)}, \quad (1)$$

where FM_O is the binary mask of O in I_S, |v| denotes the volume of voxel v in I_S (assuming all voxels are of the same size), and Total Metabolic Activity TMA(O) of O is defined as:

$$TMA(O) = \sum_v |v| \times FM_O(v). \quad (2)$$

Our argument is that any healthy organ (or tissue region) O should have very similar MMA(O) among different subjects. Thus, we expect that, after using any standardization method, the Coefficient of Variation (CV_{MMA}) of the computed MMA(O) across different subjects should decrease compared to the value before standardization. CV_{MMA} can then be used as a metric to investigate how to choose the proper maximum percentile for the standardization method. Also, it can be used as a metric to assess which standardization method is better. Along similar lines, instead of using the SUV image I_S, we define the mean activity MA(O) derived directly from the PET image I by

simply taking the mean image intensity value of I within the mask of O and the corresponding coefficient of variation CV_{MA} of MA(O) over different samples of O from different subjects.

Figures 2 and 3 illustrate the variation of CV_{MMA} and CV_{MA} as a function of p_{max} for performing calibration for standardization based separately on I_S and I, respectively, for O = Liver and O = Spleen over data set $\mathcal{I}_c$. The idea is to select p_{max} where CV_{MMA} or CV_{MA} is minimal. We denote the optimal value of p_{max} by p^*_{max}.

In our approach, we use part (~60%) of data set $\mathcal{I}_c$ (the healthy scan data set) to perform all calibration operations and estimate the parameters of the standardization mapping. Subsequently, the mapping arrived at can be applied to any other data sets such as $\mathcal{I}_T$.

3. Results

We employ CV_{MMA} and CV_{MA} determined from data set I_T to assess the effectiveness of the PET image standardization process. Our results by considering CV_{MMA} and CV_{MA} in data set I_T are summarized in Table 1 for two objects – liver and spleen. Here we assessed MMA and MA for the entire organ, which are confirmed radiologically to be healthy by a radiologist. The values of p^*_{max} found are as follows. For PET, p^*_{max} = 96.4%. For SUV, p^*_{max} = 95.6%. In other words, these values represent optimal values estimated via CV_{MA} and CV_{MMA}, respectively.

Table 1: CV values for liver and spleen in data set I_T, before and after standardization for the proposed method as well as for the z-score method.

	PET-CV_{MA}	PET_z-CV_{MA}	sSUV-CV_{MMA}	sPET-CV_{MA}
Liver	42.28%	21.62%	11.78%	**11.56%**
Spleen	37.50%	17.73%	12.40%	**12.24%**

We also compared our method with the z-score method [10] that has been used in MRI intensity standardization but adapted for PET standardization. The CV_{MMA} and CV_{MA} values resulting from this method are also listed in Table 1. In Table 1, the meaning of the different column labels is as follows. PET-CV_{MA}: CV_{MA} values from the original PET images; PET_z-CV_{MA}: CV_{MA} values from the PET images after standardization by the z-score method; SUV-CV_{MMA}: CV_{MMA} values obtained from original SUV images; sPET-CV_{MA}: CV_{MA} values after standardizing PET images.

Figure 2 displays $CV_{MA}(p_{max})$ and $CV_{MMA}(p_{max})$ as a function of p_{max} derived from data set $\mathcal{I}_c$ for liver. Similarly Figure 3 demonstrates these functions for the spleen.

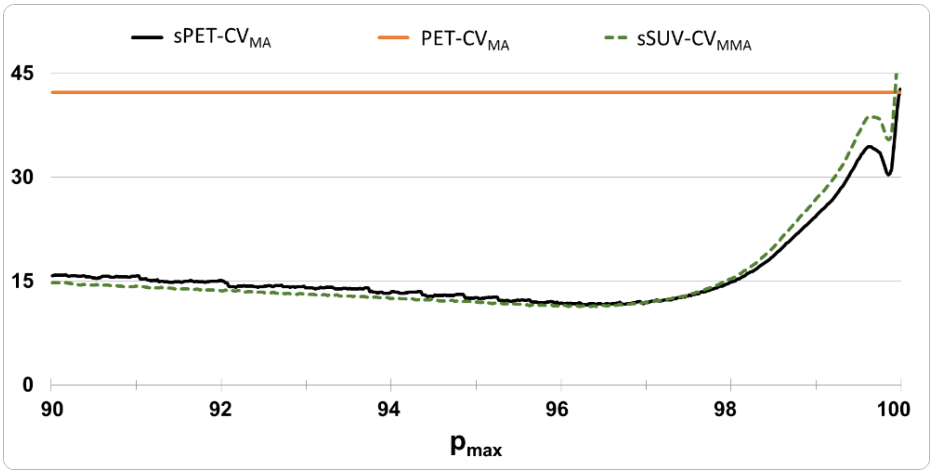

Figure 2: $CV_{MA}(p_{max})$ and $CV_{MMA}(p_{max})$ as a function derived from data set $\mathcal{I}_c$ for liver.

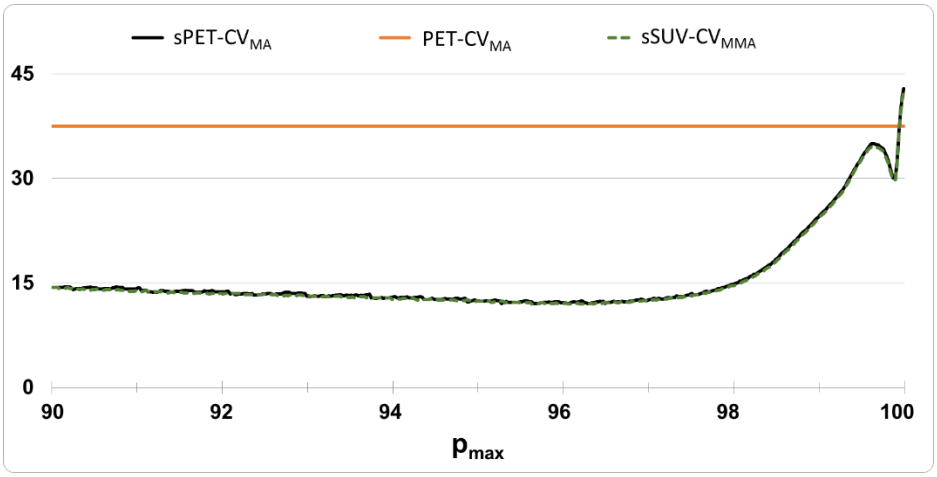

Figure 3: $CV_{MA}(p_{max})$ and $CV_{MMA}(p_{max})$ as a function derived from data set $\mathcal{I}_c$ for spleen.

4. Conclusion

We propose a new method for standardizing PET images in order to mitigate the effect of undesired factors that impede accurate quantitative analysis of PET images. The strength of the proposed method is shown through different experiments and with different metrics. The proposed method is an adaptation of an established method that introduced image standardization into MR imaging. The results indicate that substantial improvement in the uniformity of numerical meaning is achieved for both PET and SUV after standardization, and more so for the former. Note that although the subjects included here are considered as healthy based on a radiological review of the scans, it is likely that there are differences in the exact health state of individual subjects. This is perhaps one reason, among others, for the residual non-standardness that is left over after standardization has been applied.

The proposed method is superior to other methods in that it is not dependent upon subject and image acquisition related parameters. The sPET method was directly applied to PET images without use of image acquisition-related or patient-related parameters and outperformed other popular methods described in the literature including SUV and z-score normalization developed for MR images. CV was used in healthy scan data sets to measure how similar the metabolic activities of healthy organs were. We have shown that our proposed method achieved better results in comparison with other popular methods for removing non-standardness in PET images.

5. Acknowledgement

This study utilized some PET/CT scans previously acquired in normal volunteers via a pilot study from the Center of Excellence in Environmental Toxicology, and Grant IP30 ES013508-05 from the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH) to Dr. Drew Torigian (PI) and Dr. Judith Green-McKenzie at the University of Pennsylvania.

References

[1] Siegel, R. L., Miller, K. D., and Jemal, A., "Cancer statistics, 2019," *CA: a cancer journal for clinicians* **69**(1), 7–34 (2019).

[2] Kwee, T., Basu, S., Saboury, B., Alavi, A., and Torigian, D., "Functional oncoimaging techniques with potential clinical applications.," *Frontiers in bioscience (Elite edition)* **4**, 1081–1096 (2012).

[3] Torigian, D. A., Zaidi, H., Kwee, T. C., Saboury, B., Udupa, J. K., Cho, Z.-H., and Alavi, A., "PET/MR imaging: technical aspects and potential clinical applications," *Radiology* **267**(1), 26–44 (2013).

[4] T. Kwee, S. Basu, B. Saboury, A. Alavi and D. Torigian, "Functional oncoimaging techniques with potential clinical applications," *Frontiers in bioscience (Elite edition),* pp. 1081-1096, 2012.

[5] T. Kwee, D. Torigian and A. Alavi, "Oncological applications of positron emission tomography for evaluation of the thorax," *Journal of thoracic imaging 28,* pp. 11-24, 2013

[6] Larson, S. M., Erdi, Y., Akhurst, T., Mazumdar, M., Macapinlac, H. A., Finn, R. D., Casilla, C., Fazzari, M., Srivastava, N., Yeung, H. W., et al., "Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging: the visual response score and the change in total lesion glycolysis," *Clinical Positron Imaging* **2**(3), 159–171 (1999).

[7] H. Engel, H. Steinert, A. Buck, T. Berthold, R. Böni and S. G.K., "Whole-body PET: physiological and artifactual fluorodeoxyglucose accumulations," *Journal of Nuclear Medicine 37,* pp. 441-445, 1996.

[8] Nyul, L. G. and Udupa, J. K., "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **42**(6), 1072– 1081 (1999).

[9] Nyul, L. G., Udupa, J. K., and Zhang, X., "New variants of a method of MRI scale standardization," *IEEE transactions on medical imaging* **19**(2), 143–150 (2000).

[10] B. Ellingson, T. Zaw, T. Cloughesy, K. Naeini, S. Lalezari, S. Mong, A. Lai, P. Nghiemphu and W. Pope, "Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas," *Journal of Magnetic Resonance Imaging,* vol. 35.6, pp. 1472-1477, 2012