

Calibration Software for Quantitative PET/CT Imaging Using Pocket Phantoms

Dženan Zukić¹, Darrin W. Byrd², Paul E. Kinahan², and Andinet Enquobahrie¹

¹ Kitware, Inc., Carrboro, NC and ²Department of Radiology, University of Washington, Seattle, WA

Corresponding Author:

Dženan Zukić, PhD
Kitware Inc., 101 E. Weaver street, Suite G4, Carrboro, NC 27510;
E-mail: dzenan.zukic@kitware.com

Key Words: PET imaging, bias, correction, calibration, phantom

Abbreviations: Positron emission tomography (PET), computed tomography (CT), ¹⁸F-fluorodeoxyglucose (FDG), ¹⁸F-fluorothymidine (FLT), Software-as-a-Service (SaaS)

ABSTRACT

Multicenter clinical trials that use positron emission tomography (PET) imaging frequently rely on stable bias in imaging biomarkers to assess drug effectiveness. Many well-documented factors cause variability in PET intensity values. Two of the largest scanner-dependent errors are scanner calibration and reconstructed image resolution variations. For clinical trials, an increase in measurement error significantly increases the number of patient scans needed. We aim to provide a robust quality assurance system using portable PET/computed tomography “pocket” phantoms and automated image analysis algorithms with the goal of reducing PET measurement variability. A set of the “pocket” phantoms was scanned with patients, affixed to the underside of a patient bed. Our software analyzed the obtained images and estimated the image parameters. The analysis consisted of 2 steps, automated phantom detection and estimation of PET image resolution and global bias. Performance of the algorithm was tested under variations in image bias, resolution, noise, and errors in the expected sphere size. A web-based application was implemented to deploy the image analysis pipeline in a cloud-based infrastructure to support multicenter data acquisition, under Software-as-a-Service (SaaS) model. The automated detection algorithm localized the phantom reliably. Simulation results showed stable behavior when image properties and input parameters were varied. The PET “pocket” phantom has the potential to reduce and/or check for standardized uptake value measurement errors.

INTRODUCTION

In total, 1,688,780 new cancer cases and 600,920 deaths from cancer are estimated for the United States in 2017 (1). Positron emission tomography (PET) combined with x-ray computed tomography (CT) is a standard component of oncology diagnosis and staging (2-5). Quantitative PET/CT is a valuable tool for assessment of an individual’s response to therapy and for clinical trials of novel cancer therapies, because it can measure metabolic changes, which are a better indicator of response than anatomical size changes (6). Success with this approach has been shown using the glucose analogue ¹⁸F-fluorodeoxyglucose (FDG) for evaluation of therapy-induced changes in metabolic activity in several studies, including lung cancer (7) and gastrointestinal tumors (8). The thymidine analogue ¹⁸F-fluorothymidine (FLT) provides accurate measurements of tumor proliferative activity and can be used to monitor tumor responses to treatment (9). In these cases, PET imaging provides a reliable predictor for treatment responses and patient outcome. This suggests that quantitative PET imaging has an enormous potential to boost the efficiency of evaluating clinical trials of new therapies (10).

However, the use of quantitative PET imaging in clinical trials is hampered by the large degree of variability arising from

inconsistent and nonoptimized image acquisition, processing, and analysis (11-14). There are also sources of biological variability, but these are well characterized by test-retest studies to be ~10% (15-17). In contrast, the additional variability introduced by inconsistent and nonoptimized practice has ranged from 18% (18) to >40% (19-21). Two of the largest PET scanner-dependent errors are calibration (20-22) and variable resolution losses (23). For clinical trials using quantitative PET imaging, the effect of this additional variability is dramatic. For example, as standardized uptake value measurement error increases from 10% to 40%, the number of needed patient scans increases by >10-fold for an effect size of 20% and a study power of 0.8(21) (Table 1).

Despite this known variability of imaging parameters, there are currently no robust quality assurance techniques for clinical PET images. In this project, we use a PET/CT pocket phantom and associated analysis tools for quality assurance for 2 key image characteristics: the reconstructed image resolution and PET/CT scanner calibration. In our previous publication we introduced the pocket phantom (24) and briefly described web deployment (25).

The purpose of this work is to develop the technology infrastructure for enabling wider deployment and adoption of

Table 1. Impact of Measurement Error on Needed Sample Size for a Significance of $P = .05$ for a 2-tailed t-test^a

Trial Scenario	SUV Error (%)	Sample Size
Single Site (Good Calibration)	10	12
Multicenter (Good Calibration)	20	42
Multicenter (Poor Calibration)	40	158

^a Effect size is 20%; power is 0.9.

our calibration phantoms. We introduce an automated phantom detection module, an improved optimization module, and a web deployment module. We have rigorously tested this algorithm. The overarching goal of the project is to offer this automated analysis service as a Software-as-a-Service cloud application.

METHODOLOGY

The patient scans were approved by the Institutional Review Board at the University of Washington Medical Center. Informed consent was obtained from all patients.

The overall imaging workflow is as follows:

- (1) A patient or phantom scan is acquired with the pocket phantom(s) in the field of view, but not in the patient's pocket.
- (2) Images are uploaded to the analysis software.
- (3) Pocket phantoms are automatically detected and their location determined.
- (4) Algorithm is run to generate estimates of bias and resolution.

The pocket phantoms used in this workflow, shown in Figure 1, contain spherical radioactive regions. The spheres are 15 mm in diameter and varied in their activity concentrations. To conduct physical experiments, we acquired data with water-filled phantoms containing Fluorine-18 (¹⁸F, 110-minute half-life) as well as solid epoxy-filled phantoms that contained germanium-68/gallium-68 (⁶⁸Ge/⁶⁸Ga, 271-day half-life).

Phantom Detector Algorithm

We implemented an image analysis algorithm to automatically detect the pocket phantoms in PET scans. The algorithm flowchart is shown in Figure 2. First, we threshold the input image using a minimum activity level. We then reduce the number of disjoint objects using connected component analysis. We then calculate the volume and center of area of the connected components. Based on the known size of the phantom spheres, we filter the connected components based on minimum and maximum threshold radii. Using radius-filtered blobs (typically numbering in dozens, rarely in hundreds), all possible 3-blob combinations are constructed. Calculated centers enable pruning the 3-blob combinations by inappropriate sphere distances and usage of noncollinearity as a combination cost function. The lowest cost combinations are declared "positive detections," and other combinations that incorporate the same blobs (ie, conflicting combinations) are eliminated.

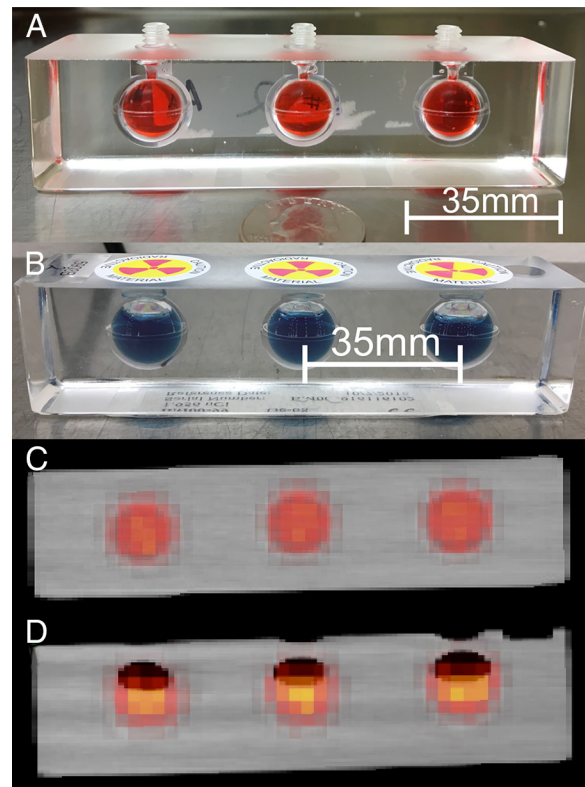


Figure 1. Pocket phantoms. Aqueous ¹⁸F phantom (A). Long-lived ⁶⁸Ge/⁶⁸Ga in epoxy phantom (B). Voids due to epoxy shrinkage are visible. Fused PET/CT image of the water-filled ¹⁸F phantom (C). Fused PET/CT image of the epoxy-filled ⁶⁸Ge/⁶⁸Ga phantom (D).

Imaging Parameter Estimation

We implemented an optimization algorithm to estimate the imaging parameters. The optimization step involves estimating parameters in a simple model of the PET scanner that is given in equation 1. We assumed that the scanner-generated image $I(x, y, z)$ differs from actual activity distribution $p(x, y, z)$ by a global scale factor g , by convolution with the 3-dimensional Gaussian kernel $k(\sigma_x, \sigma_y, \sigma_z)$, and additive noise $n(x, y, z)$. The constant g represents the global scanner sensitivity and is influenced by several physical factors, including the periodic scanner recalibrations. The function k represents the resolution, or point spread function (PSF), attributable to the acquisition and reconstruction, as well as any applied smoothing. It is assumed to be stationary and therefore fully characterized by its widths, namely, $\sigma_x, \sigma_y, \sigma_z$, which may vary independently.

$$I(x, y, z) = g \cdot p(x, y, z) * k(\sigma_x, \sigma_y, \sigma_z) + n(x, y, z) \quad (1)$$

The parameter estimation is performed on one pocket phantom at a time. The algorithm uses the detected centers to isolate regions of the input PET image that contain the phantoms. Using the sphere centers from the detector module and known phantom geometry, a synthetic, noise-free PET image is generated by applying initial guesses of the scale factor and blurring function

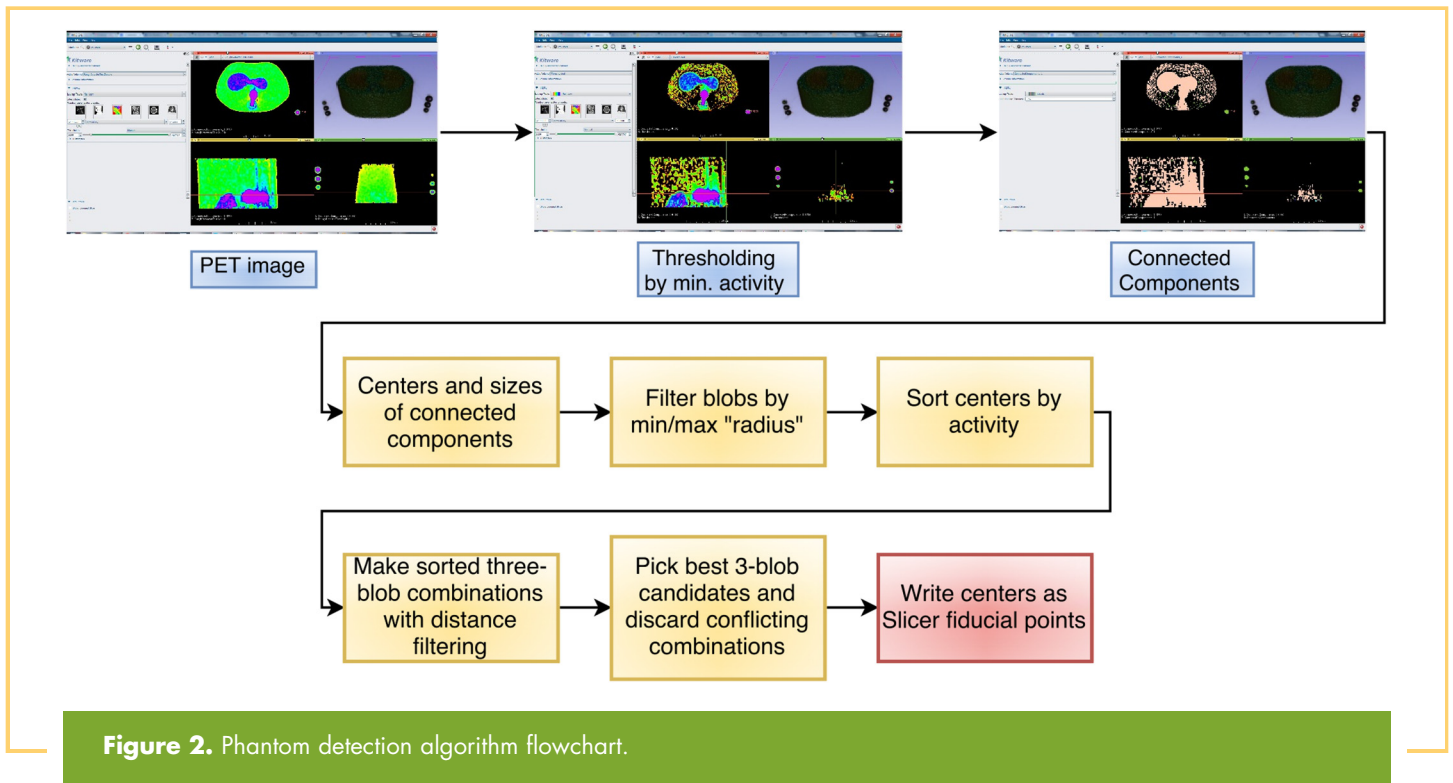


Figure 2. Phantom detection algorithm flowchart.

(analogous to g and k) to a model image of the phantom. These synthetic images are then iteratively compared with the measured image data (Figure 3). The mean-squared difference between the scanned phantom and the predicted phantom is computed as a cost function. This cost function is dependent on resolution estimates $\sigma_x, \sigma_y, \sigma_z$, global scaling coefficients, and noise present in the scanner image $\sigma_x, \sigma_y, \sigma_z$.

Our software allows determining $\sigma_x, \sigma_y, \sigma_z$ separately or a single σ for all axes. Similarly, global scaling can be estimated for each sphere independently, or as a single scaling (Figure 4, left). Internally, the software estimates the bias, which is expressed in kilobecquerels per milliliter instead of scaling factor expressed in percentages, simply because it requires less computation. All the parameters ($\sigma_x, \sigma_y, \sigma_z$, biases, sphere centers) that characterize the imaging process are estimated jointly by minimizing the cost function using Powell’s conjugate direction optimization method (26) with Brent’s line search (27) and golden section search (28). This method minimizes the cost function by doing a line search for each scalar parameter in turn. We made this modification because Powell’s method has shown better convergence properties than the Nelder-Mead simplex method (29), which we had used previously (24). The image analysis algorithm was implemented using the 3D Slicer platform (30, 31).

Software-as-a-Service

With the intention of supporting multicenter clinical trials, we deployed our image analysis pipeline as a Software-as-a-Service (SaaS) application. In the SaaS model, software is installed centrally and accessed as needed by (geographically) distributed users. This allows easy software improvements and bug fixes, because no intervention by the users is generally needed. Addi-

tional benefits include centralized data management, decoupled client-side software, and ease of access through a web browser.

To implement the SaaS application, we used Girder, an open source data management platform that is used in functional medical imaging, histology, and digital pathology research projects (32, 33).

Girder is a Python-based framework for building web applications that store, aggregate, and process scientific data. It is built on CherryPy and MongoDB and exposes data stored on a variety of backend storage engines (eg, Native file system, Amazon S3, GridFS, HDFS) through a unified RESTful API. Girder provides all essential data management functionality such as user/group authentication; fine-grained access control to data; custom metadata association; data provenance; intuitive UI to upload/browse/organize/download data and an extensible plugin framework for building web-based analytics applications. Girder consists of 2 key components, namely, the API layer and the single-page web application that serves as an example of that API’s usage. Applications can use and extend the single-page app, while others may simply use the API and write customized user-facing frontends. The Girder framework is tightly integrated with Kitware’s open-source tools for configuration and building [CTest, CDash, CMake (34-36)], visualization [VTK (37)] and image analysis [ITK (38, 39)]. Girder provides a unified interface to many distributed storage systems along with access control and extensible plugins.

In the Girder instantiation for this project, we implemented the following 3 core modules (Figure 5):

- (1) *Data upload module*: Quantitative imaging and clinical researchers and other users of the platform will upload their data

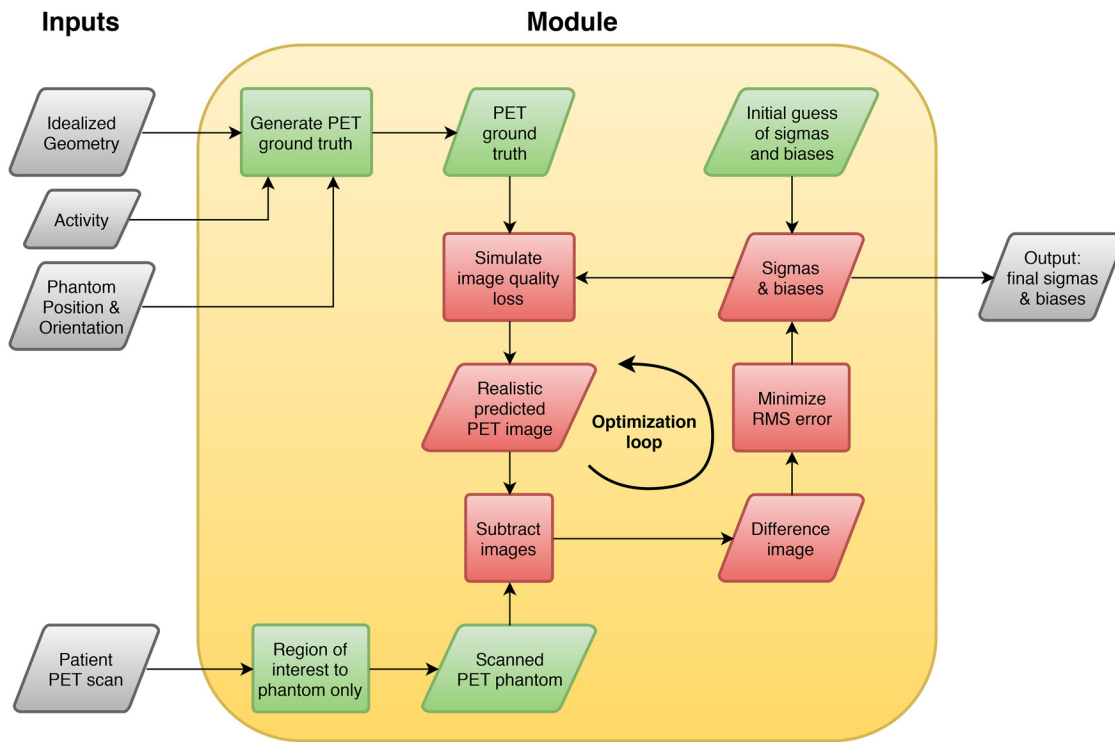


Figure 3. Positron emission tomography (PET) parameter estimation algorithm. Gray: inputs and outputs. Red: optimization loop. Green: other. Rectangular boxes: operations. Rhomboid boxes: images/data.

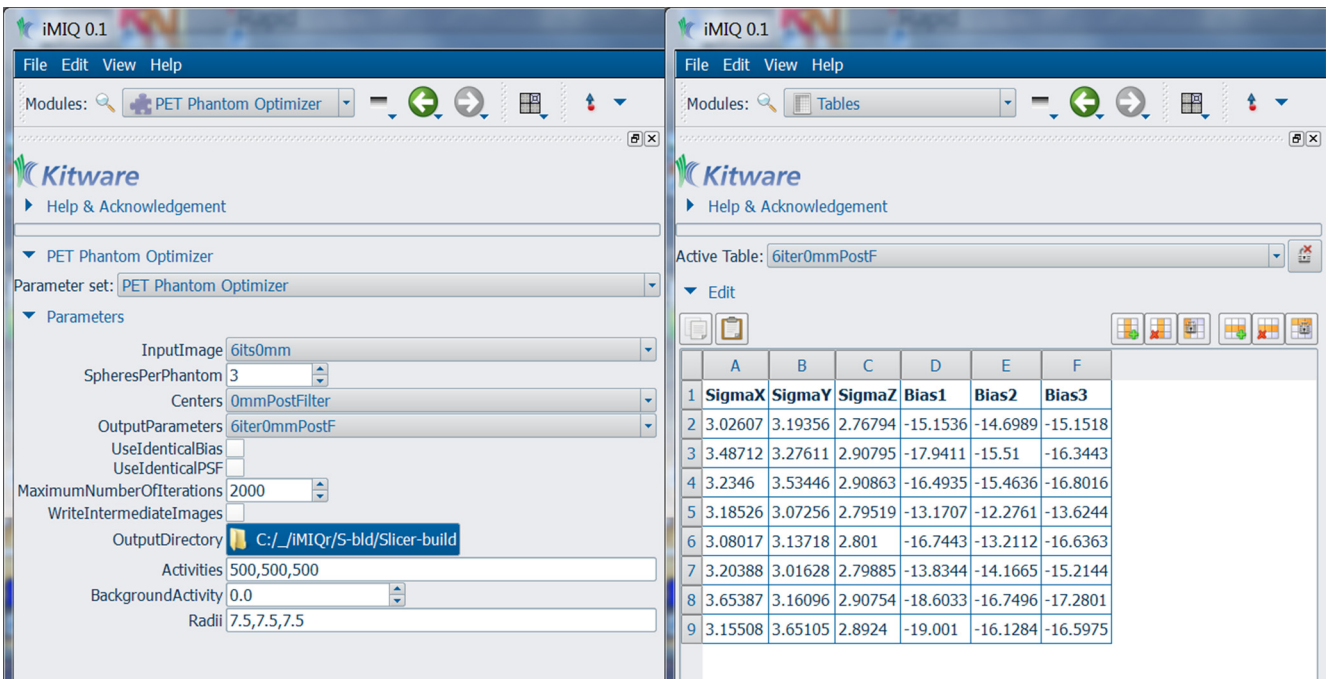


Figure 4. Parameter Optimizer Module: Left: The estimation algorithm can be run using sphere centers detected by the previous step. Right: the resulting parameters can be seen as a table. One row in the table corresponds to one phantom. The data analyzed in this case came from a realistically simulated image with 8 pocket phantoms.

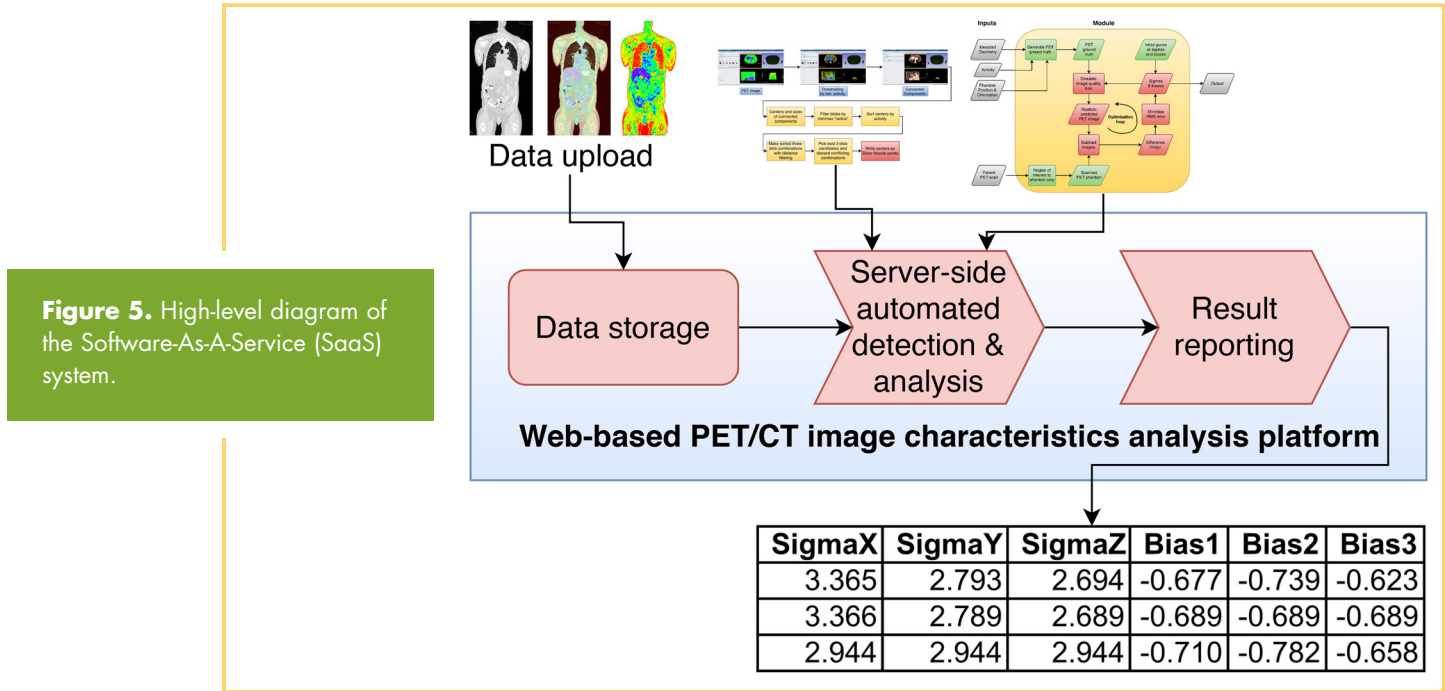


Figure 5. High-level diagram of the Software-As-A-Service (SaaS) system.

- sets using a DICOM transfer, and a review process will ensure that the data sets have been correctly uploaded (Figure 6).
- (2) *Server-side image analysis module:* The automated phantom detection and localization algorithms are encapsulated in this module.
- (3) *Result reporting module:* This module presents key PET image characteristic findings.

Girder Infrastructure

We separated our end-to-end architecture into several modules, each of which can scale independently of the others across any number of required physical machines. There are 2 major decoupled services in the deployment:

- (1) The web service for data management, user authentication and management, analysis setup, and displaying of results. This is provided by Girder, which itself can defer to scalable third-party data storage systems such as Amazon S3 or HDFS to securely persist the files it maintains.
- (2) The processing service that executes the analysis pipeline on the PET/CT data. Because this is often a computationally intensive task, it is critical that it can be run in parallel on other machines besides the ones serving the web front-end and communicating with the database.

Girder contains plugins that provide support for visualization, metadata extraction, and fast querying of the DICOM file format. The DICOM visualizer plugin allows users to navigate to a DICOM data set and view it section by section, including window and level controls, as well as showing the table of tags for each section. The DICOM metadata extractor plugin automatically inspects DICOM files as they are uploaded into the system and reads the DICOM tags from each file, recording them as structured metadata on the data set. Storing and

indexing these metadata fields in Girder’s database management system enables users to quickly search among a large collection of DICOM datasets based on the values of specific tags.

The processing service maintains a job queue which can be run in parallel across multiple machines to support load-balancing and minimize the response time of the jobs. The system is designed to be general enough to execute almost any task in a distributed environment by exposing several execution modes, including python scripting, R scripting, and even running arbitrary Docker containers. In our case, each job involves the execution of a Docker container with the analysis modules that can be maintained independently of the other parts of the system, so long as it conforms to a well-defined command-line interface. When finished, the resulting data is uploaded back into Girder for users to view.

Dockerized Slicer Modules

Docker is an open-source project that automates the deployment of applications inside software containers similar to virtual machines. That enables such a container to run on any Linux server, thus eliminating the issues of software libraries and their versioning. To build a dockerized container, one starts with one of the known Linux distributions and installs within it the required libraries. Then the custom software is added and compiled using the compiler suite contained within the container. For ease of use of the container, an entry point (a default program within the container to be invoked) can be specified. Our phantom detection and parameter estimation pipeline in Slicer were implemented in Docker containers.

Experimental Setup

Using image domain processing, we created test data with 5 combinations of $\sigma_x, \sigma_y, \sigma_z$, 5 levels of bias (−30% to +30%) for each of the 3 spheres, 5 levels of noise-to-signal ratio (0% to

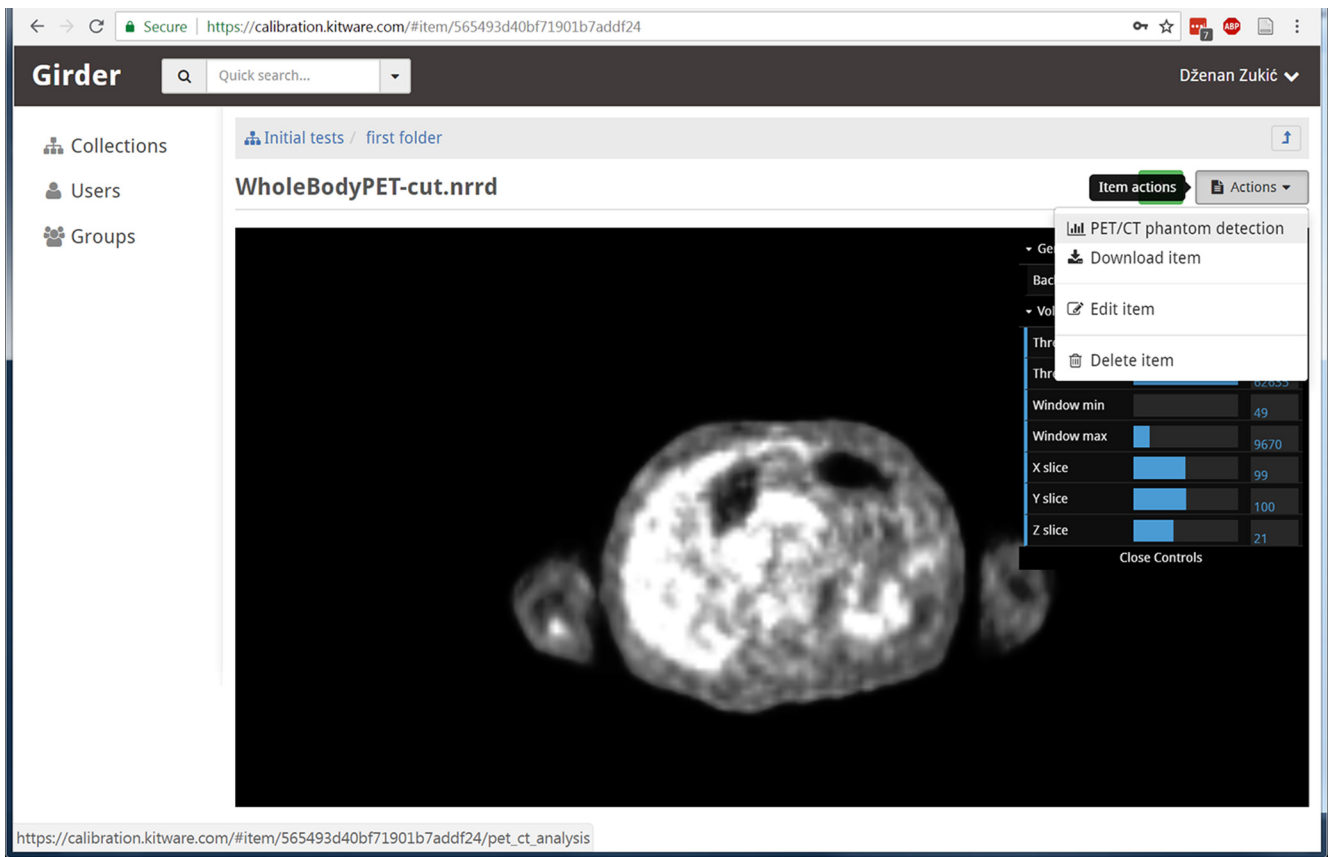


Figure 6. Web user interface for invoking PET analysis pipeline. In this review step, the user can view the rendering of the image to ensure that the right image is submitted for processing.

40%), and 3 repetitions totaling 9375 experiments. Separately, to characterize the dependence of parameter estimates on mismatches between the algorithm’s model and the physical sphere sizes, systematic variations were introduced to the software’s expected sphere radii. These tests were conducted with the measured images of water-filled pocket phantoms and with a 20-cm flood phantom in the center of the field of view (Figure 7, left). The scan duration was 15 minutes, and the activity concentration was ~6 kBq/mL in the pocket phantoms and 4.3

kBq/mL in the flood phantom. The data were acquired on a General Electric Discovery STE PET/CT scanner (Waukesha, WI) and reconstructed in MATLAB (The MathWorks, Natick, MA) using reconstruction code equivalent to the manufacturer’s.

Two patients underwent scanning with long-lived prototype pocket phantoms in the GE PET/CT scanner. Four iterations with 28 subsets per iteration were performed during the reconstruction. An 8-mm postreconstruction filter was applied. Time-of-flight information was not used. The scans were acquired

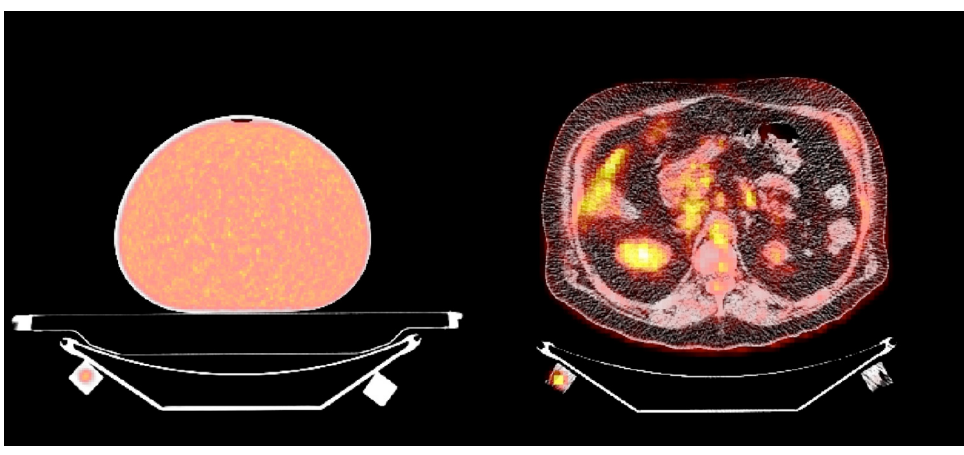


Figure 7. (Left) Fused PET/CT images of water-filled pocket phantoms and flood phantom. (Right) Patient image with epoxy-filled pocket phantoms.

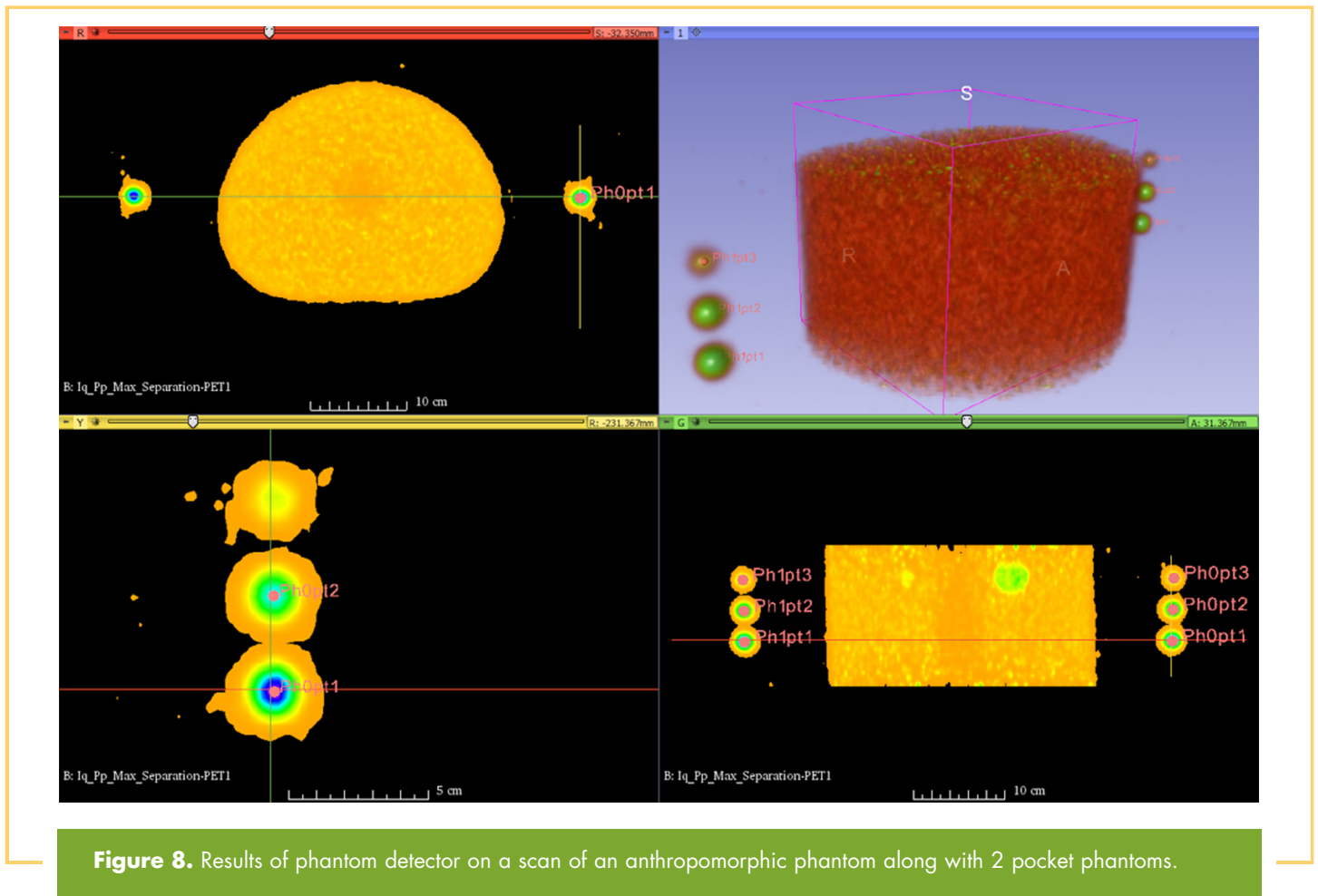


Figure 8. Results of phantom detector on a scan of an anthropomorphic phantom along with 2 pocket phantoms.

separately from the clinically indicated PET scans that the patients were undergoing. The pocket phantoms were placed on the underside of the patient bed and a clinical-duration scan was performed (Figure 7, right). The activity concentrations were ~13.0 kBq/mL and 3.1 kBq/mL in the 2 phantoms. Images were reconstructed with varying filter widths (3 mm to 9 mm in 1-mm increments). These images were then rescaled to have varying global scale factor between 0.85 and 1.15 in increments of 0.05. The equivalent bias is $\pm 15\%$ in 5% increments. This resulted in a 2-dimensional test-space of image parameters for the optimizer to estimate. The analysis of measured images was done with command-line invocations of detector and parameter optimizer modules generated automatically in MATLAB.

RESULTS

We conducted experiments using synthetic phantom data, physical phantom data and patient data with pocket phantoms. The results are described below.

Detection Algorithm Performance

In 9375 synthetic experiments, the phantom detector had no failures. The parameter estimator had an average relative error below 1% for each parameter. In reality, the spheres would all have the same level of bias and the noise would not approach 40%.

The detector module was tested on real images. All spheres were correctly detected in the test images, including patient images. The phantom detector takes <1 second to run on a typical PET image (matrix size, $256 \times 256 \times 47$).

The phantom detector module can be invoked via Girder's web interface (Figure 6) or from within iMIQ (a customized 3D Slicer application), as shown in Figure 8. Figure 4 shows the parameter optimizer module that implements the algorithm shown in Figure 3. The estimation algorithm can be run individually for each detected sphere, and the corresponding imaging parameters exported and presented to the user in iMIQ.

Imaging Parameter Estimation Algorithm Performance

The parameter estimator takes some 10–30 seconds per 3-sphere phantom. These tests were conducted on an Intel Xeon E3-1220 processor (4 cores @ 3.1 GHz).

Figure 9 plots the cost function on y-axes with varying individual parameters in synthetic and real images. For a given subplot, 1 parameter is varied, while all the others are kept fixed at their minimum value. The y-axes express average voxel difference, in kilobecquerels per milliliter. This average is per-voxel root-mean square error between a realistic predicted image and a scanned PET phantom (Figure 3).

Figure 10 (top) shows that the iterative updates of bias, resolution, and sphere locations cause fluctuations in the

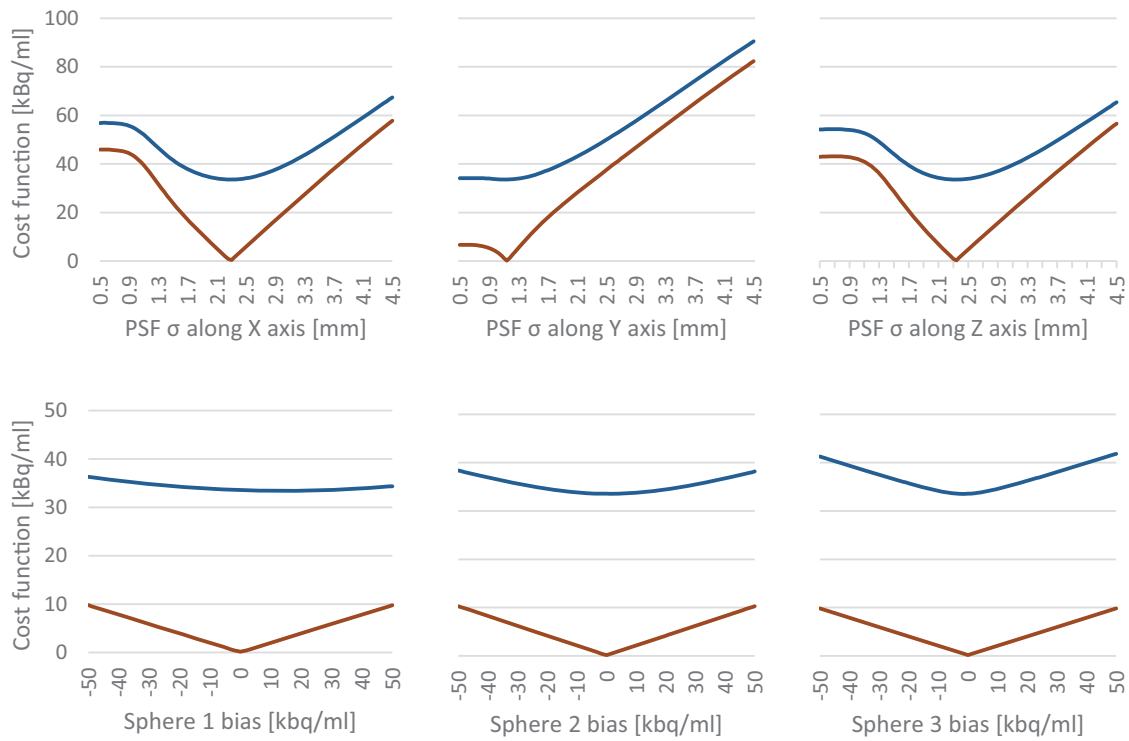


Figure 9. 1D plots of the cost function on a real image of the phantom (blue) and closely matching synthetic noiseless image (red). The minimum is at [2.28, 1.14, 2.33, 0, 0, 0].

value of the cost function. However, the software retains the lowest value achieved in the entire optimization (Figure 10 bottom), which shows good stability well before the algorithm stops.

Figure 11 shows that the mismatches spanning $\pm 5\%$ of the input sphere diameter led to modest changes in the bias and resolution estimates. Resolution estimates varied by <0.5 mm for any single postreconstruction smoothing filter size. Scale

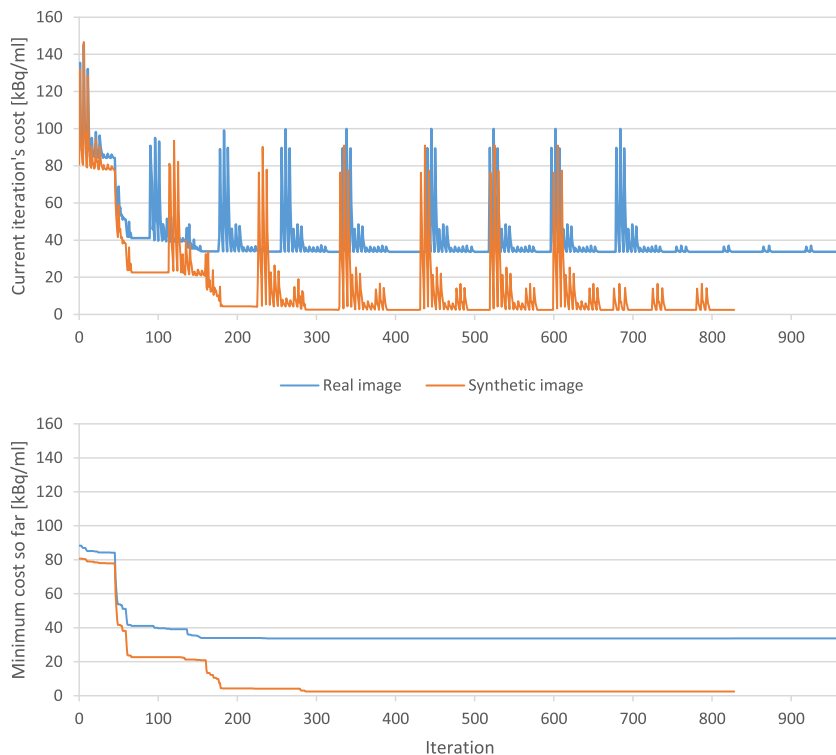
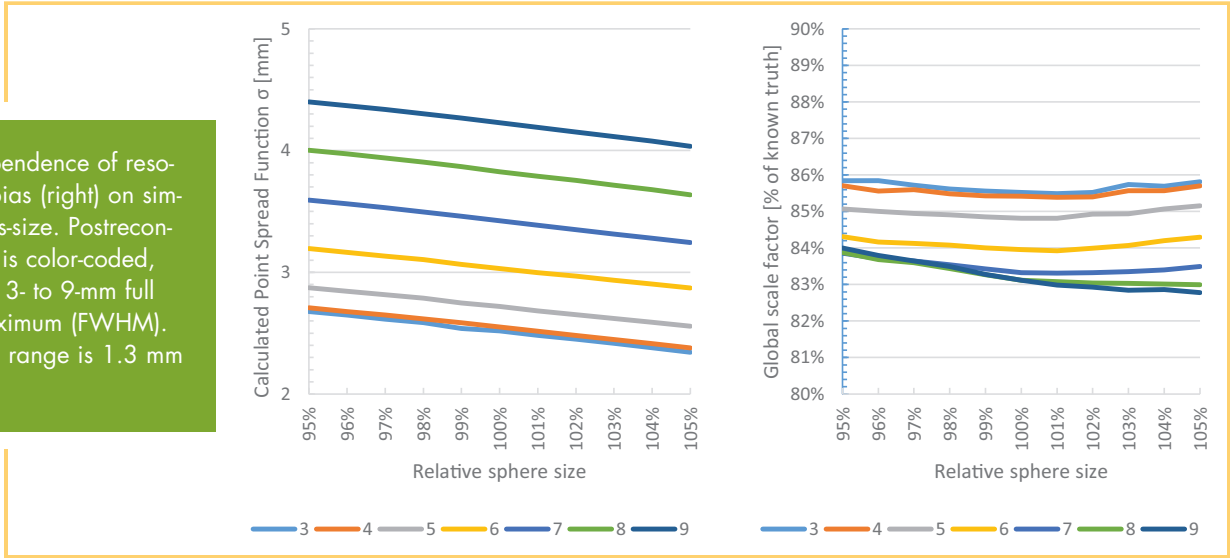


Figure 10. Plot of cost function values over iterations of the optimizer. Plotted are the values of the cost function which the optimizer explored and retained minimums, for a real image as well as a noiseless synthetic image, both from Figure 9. The synthetic image's cost function does not reach 0 due to approximate calculation of smoothing, which makes the computation much faster and has negligible effect on the result.

Figure 11. Dependence of resolution (left) and bias (right) on simulated sphere mis-size. Postreconstruction filtering is color-coded, and ranges from 3- to 9-mm full width at half maximum (FWHM). Equivalent sigma range is 1.3 mm to 3.8 mm.



factor estimates varied by <0.015 (1.5%) for a single filter and by <0.031 (3.1%) across all filters.

Feasibility Study on Patients

Figure 12 (left) shows that the algorithm successfully ran and accurately estimated the bias in our patient data. Estimates of global scale factor were divided by the known scale factor to show stability (i.e., flatness) of the algorithm’s performance as the applied bias and image resolution varied. Data were similar for the second set of patient images (data not shown). Figure 12 (right) shows the estimated resolution over the same test space of images. In contrast to the bias estimates, there is not a simple normalization that can make the plot more easily interpreted visually, but we note that the data show the expected independence of bias factor (with some wobble) and a smooth dependence on filter width as expected.

DISCUSSION

We have created a working software platform for the pocket phantom system and further tested its use in real-world conditions. By implementing our algorithm with the Girder infrastructure and dockerized containers, we have extended our previous work to create a server-ready automated analysis pipeline that makes the pocket phantom system available to geographically distributed users. The optimizer showed good tolerance of the real-world conditions under which it was tested.

We note that the resolution measurements discussed here may differ from those obtained by other measurement methods that use different phantom geometries. In particular, we expect that our resolution estimates depend on the source positioning in the field of view, the energy of positrons emitted by ^{68}Ge , and the lack of background activity around our spherical sources. These constraints have been previously noted for methods in-

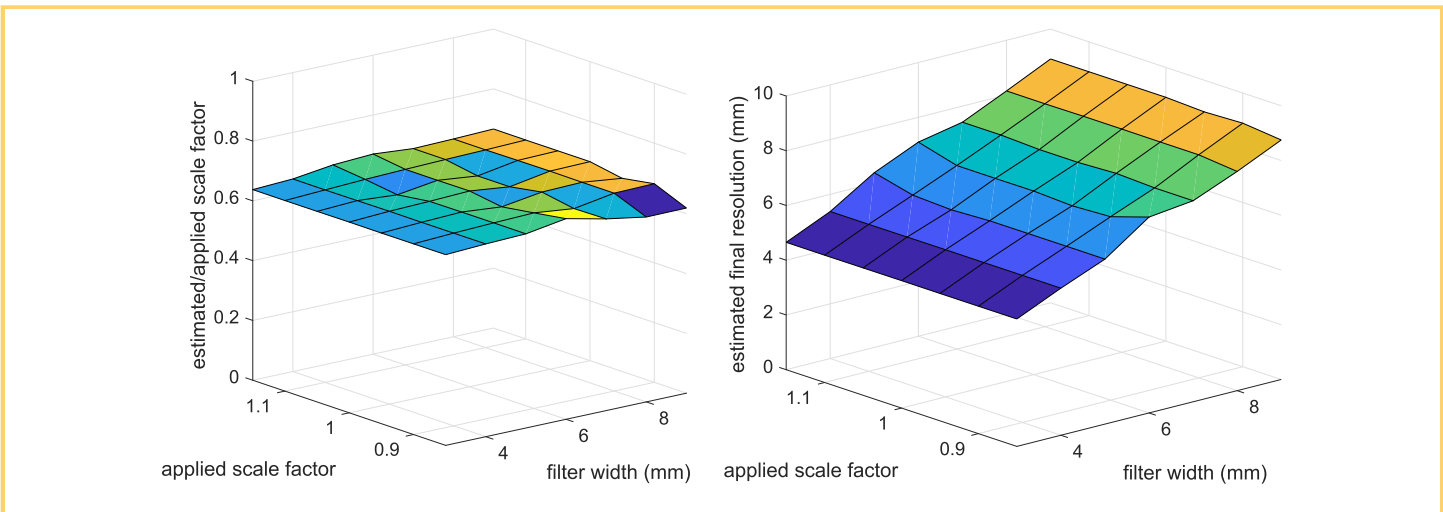


Figure 12. (Left) Normalized global scaling estimates from a patient scan with the pocket phantoms that was rescaled and filtered with systematic variations. Here, global scale factor was simulated by multiplying the image by a series of constant factors, and the accuracy of the algorithm results in relatively flat surfaces after dividing the scale estimate by the same factor. (Right) Estimates of image resolution for the same patient scan.

volving solid uniform sources (40). However, we feel that for quality assurance, our measurements may still have value in their ability to detect changes in image resolution, which may indicate unstable biases in the image-based metrics.

The current phantom detection module differs from the previous version in that it operates on the PET image instead of the CT image. During early research, the detection sometimes required user intervention to successfully locate the phantoms. This should not be an undue burden on users, as the effort required was minimal in our experience. We note that because the optimization step in the algorithm refines the estimates of sphere centers, the detection step will not affect the overall effectiveness of the pocket phantom system except in cases of complete failure, which the user can detect visually in the interface.

Figure 9 shows that the cost function has global minima as a function of each variable returned by the algorithm. While this does not guarantee that the algorithm is convergent in the full multidimensional space, Figure 10 suggests that updates to the parameters estimates are small well before the algorithm terminates. The final value of the cost function is dependent upon several factors. Noise in the image means that even for a perfect physical model, the cost function will have some finite value for optimal parameters. In addition, the assumption that the PSF is isotropic in the transaxial plane may not hold due to detector parallax (41), meaning that even a noise-free image would not perfectly match the algorithm's optimal model image.

The algorithm performed with reasonable stability as sphere size mismatch varied as far as $\pm 5\%$ of the known radius. We note that while this test was intended to gauge the algorithm's tolerance

to manufacturing variability with fixed "known" sphere size in the algorithm, we have instead used a fixed physical sphere size and varied the value in the software to avoid the need to manufacture 11 phantoms in a precisely wrong fashion. The modest dependence of the estimates on sphere mismatch suggests that the algorithm performs with acceptable accuracy if the spheres are manufactured to within 5% of their nominal radius.

The primary purpose of the patient measurements was to show that the detection and optimization could still run successfully with the additional constraints and challenges that clinical imaging presents. The system performed well on the clinical data, showing stability of global scaling (bias) estimates and smooth dependence on filter width as expected (Figure 12).

The pocket phantom software can also be run locally without the cloud-based architecture described above. This is advantageous where it is not desirable to send patient data off-site, for example, due to legal regulations.

Concluding Remarks and Future Work

The PET pocket phantom system has the potential to reduce PET measurement errors. The phantom and estimation software performed adequately in the tested scenarios and can be made available to sites by our software.

We are currently concluding an investigation into the numerous mathematical and physical effects of the imaging process that may affect the pocket phantom system, and further investigating how the physical phantoms may be reliably manufactured, which is a requisite to the deployment of our system.

ACKNOWLEDGEMENTS

We would like to thank Rebecca Christofel from Department of Radiology, University of Washington for help with recruiting the patients for this study. The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U01CA148131, R01CA042593, and R42CA167907. The

content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosures: No disclosures to report.

Conflict of Interest: The authors have no conflict of interest to declare.

REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin*. 2017; 67:7–30.
- Facey K, Bradbury I, Laking G, Payne E. Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers. *Health Technol Assess*. 2017;11:iii–iv, xi–267.
- Lardinois D, Weder W, Hany TF, Kamel EM, Korom S, Seifert B, von Schulthess GK, Steinert HC. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med*. 2003;348:2500–2507.
- von Schulthess GK, Steinert HC, Hany TF. Integrated PET/CT: current applications and future directions. *Radiology*. 2006;238:405–422.
- Wahl RL. Why nearly all PET of abdominal and pelvic cancers will be performed as PET/CT. *J Nucl Med*. 2004;45(1 suppl):82S–95S.
- Weber WA. Assessing tumor response to therapy. *J Nucl Med*. 2009;50(Suppl 1):1S–10S.
- Weber WA, Petersen V, Schmidt B, Tyndale-Hines L, Link T, Peschel C, Schwaiger M. Positron emission tomography in non-small-cell lung cancer: prediction of response to chemotherapy by quantitative assessment of glucose use. *J Clin Oncol*. 2003;21:2651–2657.
- Stroobants S, Goeminne J, Seegers M, Dimitrijevic S, Dupont P, Nuyts J, Martens M, van den Borne B, Cole P, Sciort R, Dumez H, Silberman S, Mortelmans L, van Oosterom A. 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). *Eur J Cancer*. 2003;39:2012–2020.
- Pio BS, Park CK, Pietras R, Hsueh W-A, Satyamurthy N, Pegram MD, Czernin J, Phelps ME, Silverman DH. Usefulness of 3'-[F-18]fluoro-3'-deoxythymidine with positron emission tomography in predicting breast cancer response to therapy. *Mol Imaging Biol*. 2006;8:36–42.
- Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov*. 2003;2:566.
- Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med*. 45:1519–1527.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50(Suppl 1):11S–20S.
- Beyer T, Czernin J, Freudenberger LS. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. *J Nucl Med*. 2011; 52:303–310.
- Graham MM, Badawi RD, Wahl RL. Variations in PET/CT methodology for oncologic imaging at U.S. academic medical centers: an imaging response assessment team survey. *J Nucl Med*. 2011;52:311–317.
- Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]fluoro-2-deoxy-D-glucose uptake at PET. *Radiology*. 1995;196:167–173.
- Weber WA, Ziegler SI, Thodtman R, Hanauske A-R, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med*. 1999;40:1771–1777.

17. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med.* 2008;49:1804–1808.
18. Velasquez LM, Boellaard R, Kolia G, Hayes W, Hoekstra OS, Lammertsma AA, Galbraith SM. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med.* 2009;50:1646–1654.
19. Fahey FH, Kinahan PE, Doot RK, Kocak M, Thurston H, Poussaint TY. Variability in PET quantitation within a multicenter consortium. *Med Phys.* 2010;137(7Part1):3660–3666.
20. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. 2010;31:496–505.
21. Doot R, Allberg K, Kinahan P. Errors in serial PET SUV measurements. *J Nucl Med.* 2010;51(supplement 2):126–126.
22. Lockhart CM, MacDonald LR, Alessio AM, McDougald WA, Doot RK, Kinahan PE. Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. *J Nucl Med.* 2011; 52:218–224.
23. Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys.* 2010;37:6035–6046.
24. Kinahan PE, Byrd D, Helba B, Wangerin KA, Liu X, Levy JR, Allberg KC, Krishnan K, Avila RS. Simultaneous estimation of bias and resolution in PET images with a long-lived “pocket” phantom system. *Tomography.* 2018;4:33–41.
25. Zukić D, Mullen Z, Byrd D, Kinahan P, Enquobahrie A. A web-based platform for a high throughput calibration of PET scans. In: *Computer Assisted Radiology and Surgery [Internet].* Heidelberg, Germany: Springer; 2016. p. S29–30. Available from: <http://link.springer.com/journal/11548/11/1/suppl/page/1>.
26. Powell MJ. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput J.* 1964;7:155–162.
27. Brent RP. An algorithm with guaranteed convergence for finding a zero of a function. *Comput J.* 1971;14:422–425.
28. Kiefer J. Sequential minimax search for a maximum. *Proc Am Math Soc.* 1953;4:502–506.
29. Nelder JA, Mead R. A simplex method for function minimization. *Comput J.* 1965;7:308–313.
30. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* 2012;30:1323–1341.
31. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. In: *Intraoperative imaging and image-guided therapy [Internet].* Springer; 2014 [cited 2017 Oct 10]. p. 277–289. Available from: http://link.springer.com/chapter/10.1007/978-1-4614-7657-3_19.
32. Jomier J, Aylward SR, Marion C, Lee J, Styner M. A digital archiving system and distributed server-side processing of large datasets. In: Siddiqui KM, Liu BJ, editors. *Proc SPIE [Internet].* 2009 [cited 2017 Oct 10]. p. 726413. Available from: https://www.researchgate.net/profile/Martin_Styner/publication/253893215_A_Digital_archiving_system_and_distributed_server-side_processing_of_large_datasets/links/00b49538e7d9de1bd8000000.pdf.
33. Gutman DA, Khalilia M, Lee S, Nalisnik M, Mullen Z, Beezley J, Chittajallu DR, Manthey D, Cooper LAD. The Digital Slide Archive: a software platform for management, integration, and analysis of histology for cancer research. *Cancer Res.* 2017;77:e75–78.
34. Martin K, Hoffman B. *Mastering CMake: a cross-platform build system: version 3.1.* Kitware; 2015.
35. Martin K, Hoffman B. *Mastering CMake: a cross-platform build system.* Kitware; 2010.
36. *The Architecture of Open Source Applications: CMake [Internet].* 2014 [cited 2016 Feb 15]. Available from: <http://www.aosabook.org/en/cmake.html>.
37. Schroeder WJ, Lorensen B, Martin K. *The visualization toolkit: an object-oriented approach to 3D graphics.* Kitware; 2004.
38. Yoo TS, Ackerman MJ, Lorensen WE, Schroeder W, Chalana V, Aylward S, Metaxas D, Whitaker R. Engineering and algorithm design for an image processing API: a technical report on ITK—the Insight Toolkit. *Stud Health Technol Inform.* 2002;85:586–592.
39. Johnson HJ, McCormick MM, Ibáñez L. *The ITK Software Guide.* Kitware, Inc.; 2013.
40. Lodge MA, Rahmim A, Wahl RL. A practical, automated quality assurance method for measuring spatial resolution in PET. *J Nucl Med.* 2009;50(8):1307.
41. Alessio AM, Stearns CW, Tong S, Ross SG, Kohlmyer S, Ganin A, Kinahan PE. Application and evaluation of a measured spatially variant system model for PET image reconstruction. *IEEE Trans Med Imaging.* 2010;29:938–949.