

# Improved frame-based estimation of head motion in PET brain imaging

J. M. Mukherjee<sup>a),b)</sup> and C. Lindsay<sup>a)</sup>

*Department of Radiology, University of Massachusetts Medical School, Worcester, Massachusetts 01655*

A. Mukherjee

*Aware, Inc., Bedford, Massachusetts 01730*

P. Olivier

*Philips Medical Systems, Cleveland, Ohio 44143*

L. Shao

*ViewRay, Oakwood Village, Ohio 44146*

M. A. King and R. Licho

*Department of Radiology, University of Massachusetts Medical School, Worcester, Massachusetts 01655*

(Received 27 July 2015; revised 29 March 2016; accepted for publication 2 April 2016; published 21 April 2016)

**Purpose:** Head motion during PET brain imaging can cause significant degradation of image quality. Several authors have proposed ways to compensate for PET brain motion to restore image quality and improve quantitation. Head restraints can reduce movement but are unreliable; thus the need for alternative strategies such as data-driven motion estimation or external motion tracking. Herein, the authors present a data-driven motion estimation method using a preprocessing technique that allows the usage of very short duration frames, thus reducing the intraframe motion problem commonly observed in the multiple frame acquisition method.

**Methods:** The list mode data for PET acquisition is uniformly divided into 5-s frames and images are reconstructed without attenuation correction. Interframe motion is estimated using a 3D multiresolution registration algorithm and subsequently compensated for. For this study, the authors used 8 PET brain studies that used F-18 FDG as the tracer and contained minor or no initial motion. After reconstruction and prior to motion estimation, known motion was introduced to each frame to simulate head motion during a PET acquisition. To investigate the trade-off in motion estimation and compensation with respect to frames of different length, the authors summed 5-s frames accordingly to produce 10 and 60 s frames. Summed images generated from the motion-compensated reconstructed frames were then compared to the original PET image reconstruction without motion compensation.

**Results:** The authors found that our method is able to compensate for both gradual and step-like motions using frame times as short as 5 s with a spatial accuracy of 0.2 mm on average. Complex volunteer motion involving all six degrees of freedom was estimated with lower accuracy (0.3 mm on average) than the other types investigated. Preprocessing of 5-s images was necessary for successful image registration. Since their method utilizes nonattenuation corrected frames, it is not susceptible to motion introduced between CT and PET acquisitions.

**Conclusions:** The authors have shown that they can estimate motion for frames with time intervals as short as 5 s using nonattenuation corrected reconstructed FDG PET brain images. Intraframe motion in 60-s frames causes degradation of accuracy to about 2 mm based on the motion type.

© 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4946814>]

Key words: PET, motion tracking, motion compensation, registration, reconstruction

## 1. INTRODUCTION

PET brain imaging is negatively impacted by head motion during PET imaging as well as between CT and PET imaging. The loss of resolution and misalignment of the CT attenuation map due to motion can cause distortions affecting the image quality and quantitation of PET images. In healthy volunteers undergoing simulated imaging and Parkinson's disease (PD) patients being PET imaged, continuous or drifting type motions, repetitive motions occurring about a mean position,

and large step motions that resulted in a sustained change in position were observed.<sup>1</sup> In this study long drift motions measured throughout healthy subject scans correlated to the subject falling asleep; whereas, for PD subjects these motions often reflected a commonly observed tendency to pull to one side. Drifting motions up to 13 mm were observed for one PD subject. Rotations of up to 3.0° were present in several PD data sets. In one case, brain drifts of up to 6 mm were measured in the striatum, and up to 15 mm in the occipital cortex region. In another study, about 15% of 500 2-h brain scans which were

divided into 24 5-min frames had more than 3 mm intraframe motion.<sup>2</sup> This may occur despite head-restraints to limit motion<sup>3–5</sup> where typical translations in the range of 2–20 mm and rotations of 1°–4° were observed, depending on the type of mask and the duration of scan. Please note that the actual impact of motion due to rotations depends on the choice of the rotation center.

A number of authors have proposed methods to estimate and compensate for such head motion during PET brain imaging.<sup>6–12</sup> Motion estimation can be broadly grouped into external-tracking based and data-driven methods. External tracking utilizing an electro-mechanical system was used in Green *et al.*;<sup>13</sup> however, by far the most commonly used method to track motion is with infrared stereo cameras by affixing passive reflective markers<sup>1,3,4,6,7,10,11,14–21</sup> or active markers that emit light<sup>22</sup> to the head of the patient. Recently, researchers have begun investigating the use of structured light cameras, such as the Microsoft Kinect<sup>23</sup> or other devices<sup>5,9,24</sup> which can be used to track the surface of the head without the need for markers. Data-driven methods which estimate motion from temporal frames of reconstructed PET data using registration to a reference frame<sup>2,22,25–28</sup> have also been reported. These methods have the advantage that no external equipment is necessary and motion compensation can be applied retrospectively. Nonattenuation corrected frames were used in Andersson *et al.*<sup>28</sup> to avoid registration errors due to motion that may have occurred between PET emission and corresponding attenuation scan. Fulton *et al.*<sup>14</sup> and Rahmim *et al.*<sup>29</sup> provide extensive reviews of the various methods of motion tracking and compensation in PET.

In this work we report on a data-driven motion estimation strategy for PET brain imaging which also uses nonattenuation corrected images reconstructed from clinical patient list-mode PET data to avoid errors associated with misaligned CT images.<sup>2</sup> The list-mode data are first divided into frames of 5 s, which are then reconstructed to produce a sequence of 3D nonattenuation corrected PET brain images. In this paper we will use the term “reconstructed frames” and images interchangeably. The images are then preprocessed (details discussed in Sec. 2.C) and spatially registered using an automated image registration algorithm<sup>30</sup> to the first image in the sequence to estimate the motion relative to it. In the absence of prior knowledge about motion, we chose fixed 5-s frames for data-driven motion estimation in PET. In prior work, the multiple acquisition frame (MAF) approach refers to the division of PET data to multiple frames based on a motion-threshold and compensation of motion by alignment and summation of reconstructed frames. Unlike MAF approaches,<sup>3,4,15,22,31,32</sup> we derive the motion estimates from the PET data itself using image registration. Similar methods as ours have been proposed previously.<sup>2,25,26</sup> However, none of the authors reported motion estimation using image registration from frames as short as 5 s. Jin *et al.*<sup>2</sup> used 5-min frames, Chen *et al.*<sup>25</sup> used frame sizes of 2 s to 5 min, but did not report registration error from first 22 s of data that consisted of 2-s frames. Costes *et al.*<sup>26</sup> used 2-min and longer frame sizes with intraframe motion occurring at least 30 s apart. One major argument against MAF has been the inability to compensate for intraframe motion due to the

use of frame sizes 30 s or longer.<sup>26</sup> Jin *et al.*<sup>32</sup> reported on a MAF-based method using minimum frame size of 2–4 s but the motion itself was estimated from external tracking system.

Herein, our objective is to investigate the tradeoffs in data-driven motion estimation using frame sizes as small as 5 s. Small frames have the potential to increase the accuracy of motion estimation but also create more noise affected images due to poor statistics compared to larger frame sizes. Furthermore, our method can be used to compensate for misalignment errors in CT-based attenuation correction by aligning all other PET reconstructed frames to a reference image that is closest to the CT image, followed by estimating the motion between the reference and the CT image.

## 2. MATERIALS AND METHODS

### 2.A. Patient data

We developed and tested our motion estimation method using eight anonymized patient list-mode datasets which we obtained with the approval of the Institutional Review Board (IRB). The selected studies had very little intrinsic motion as confirmed through visual assessment by an expert physician observer. All the studies used F-18 FDG as the PET tracer and the purpose of imaging was to assess regional glucose metabolism in the brain. The injected dose was between 200 and 555 MBq ( $328 \pm 126$  MBq) and the time of injection was between 0.9 and 1.5 h before acquisition. All patient studies used a CT based attenuation map which was acquired at the beginning of the study on a Philips Ingenuity TOF PET/CT system. Herein, it is assumed that these attenuation maps were aligned with the start of the 600 s duration of PET acquisition.

Our motion estimation and compensation methodology is schematically illustrated in Fig. 1. Each 10 min-long patient list-mode dataset was subdivided into 120 5-s time intervals (called frames) and reconstructed into 3D images. All processing (motion estimation and compensation) was performed on these images. Reconstruction was performed using PET/CT Ingenuity TF reconstruction software (version 4.0.1) provided by Philips Healthcare. The parameters of the protocols used for reconstructing these images were those provided to customers for use in the Ingenuity TOF PET/CT clinical systems. The reconstruction used the RAMLA 3D algorithm<sup>33</sup> (3 iterations, 33 subsets, and smoothing parameter set to normal) without attenuation correction producing reconstructed frames with  $128 \times 128 \times 90$  voxels of  $2 \times 2 \times 2$  mm<sup>3</sup> volume.

### 2.B. Motion simulation

To simulate head motion during acquisition, we artificially added motion to the patient data. This was done by applying six-degree-of-freedom (6-DOF) rigid-body transforms to the 5-s images of each patient as shown with a box with dashed outline in Fig. 1. Therefore, our motion simulation only accounts for interframe motion at a granularity of 5 s.

Four different motions were selected based on movements reported clinically.<sup>1–4</sup> The first was a *Step* motion which was simulated as sudden movements. This motion was applied with

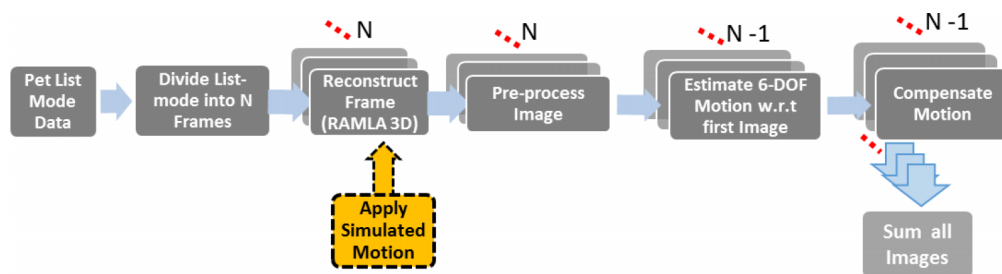


FIG. 1. Schematic diagram of our motion estimation and compensation pipeline. Our method divides PET list-mode acquisition data into a number of time frames ( $N$ ) which are reconstructed. In our test strategy we apply artificial motion (box with dashed outline) to the images reconstructed with and without attenuation. We then use preprocessing and image registration to estimate motion relative to the first image (reference). The resulting  $N - 1$  transforms are used to move  $N - 1$  images to the pose of the reference image and all  $N$  images are then summed to produce a single motion compensated image.

a maximum extent of  $\sim 13$  mm in the axial-direction (Z-axis), and about 4 mm in the lateral direction (X-axis). A small rotation of up to  $6^\circ$  about X-axis was also added. The rotation center was fixed at the center of the first 5-s reconstructed frame. The second was a *Gradual* motion which was applied as a slow drift along the lateral direction with a maximum extent of 10 mm. This transform is intended to simulate the motion of the head which can be associated with patients falling asleep and is similar to that in Olesen *et al.*<sup>24</sup> Gradual drifts, not necessarily associated with patients falling asleep, have been widely reported.<sup>1,5,7,27</sup> The third motion was that of an actual *Volunteer*. This motion was obtained by using a marker-based visual tracking system<sup>34</sup> tracking the head movement of a human volunteer who was coached to move while lying in the gantry of a SPECT/CT system where the motion tracking system was installed. This movement consisted of a complex set of small and large motions with a maximum extent of 3 cm. The fourth set called *Baseline* was without artificially applied motion and contained only the pre-existing motion in the patient datasets. Intraframe motion was also simulated by summing the moved 5-s frames accordingly to obtain frames with different time intervals, i.e., 10 and 60 s. Thus, for each of the eight patient datasets, we simulated three different frame sizes with four different motion transforms resulting in 96 motion datasets.

## 2.C. Preprocessing steps

As shown in Fig. 1, the images were preprocessed before image registration to suppress noise and enhance the outer edges of the brain. Our preprocessing steps consisted of Gaussian smoothing coupled with a median filter, followed by a nonlinear gamma correction to enhance the contrast of the edges. Together, these operations smoothed the reconstructed frame while preserving the edges needed for image registration. All the preprocessing filters were 2D and therefore applied to the volume slice-by-slice. Gaussian smoothing with  $\sigma = 2$  pixels was applied for noise reduction. A median filter with a wide square neighborhood ( $17 \times 17$  pixels) was applied for additional smoothing while preserving the edges of major structures. This smoothing degraded the contrast which was regained using a gamma power function<sup>35</sup> with an exponential of  $\gamma = 2.5$  ( $\gamma > 1$  signifies an expansion and  $\gamma < 1$  signifies a

compression). The gamma power function raises the intensity of each pixel  $P$  to the power of  $\gamma$  which is then scaled by the maximum voxel value  $M$ . Thus, the new pixel value was obtained from  $P' = M * (P/M)^\gamma$ .

An illustration of the alteration of the slices by our preprocessing method is shown in Fig. 2, where the two rows show the same slice from two different 5-s reconstructed frames of patient data. In each row, the first image on the left is the original reconstructed slice with pronounced noise, the second is after Gaussian filtering, the third after the additional median filtering, and the last image was obtained after gamma correction. Line profiles of all four images are plotted together on the right. The location of the outer edge of the brain in these profiles is indicated by red arrows. Note that for both slices the edge after preprocessing remains close to that in the unprocessed image, while other variations in the profile are suppressed.

Figure 3 further illustrates the stability of the outer edge of the brain in 5-s PET images after preprocessing. The images shown are edge maps identified using a Sobel operator in the extraction method implemented in the image processing and analysis tool Fiji.<sup>36</sup> These images are meant for illustration purposes only and were not used for image registration. The top two rows show the overlaid edges in five slices of three different preprocessed 5-s images selected from a patient, and the bottom row shows the same images after adding simulated motion. In all the images, red, green, and blue color channels are used to overlay the three images. Therefore, the edges appear in white where they overlap in all three images. The top row shows the edges after the first stage of preprocessing, i.e., Gaussian filtering, and the middle row shows the edges after the full preprocessing sequence illustrated in Fig. 2. From the top row, it is evident that the outer boundary of the count distribution indicated by red arrows is consistent across images (due to its white color). Everywhere else, the edges do not perfectly overlap due to noise. By comparing the top and middle row, it is evident that our preprocessing sequence improved the continuity and consistency of the outer edges compared to plain Gaussian smoothing. This presents as improved sharpness of the edge and brightness of the white color in the middle row overlays. Additionally, the spurious edges in the region indicated by green arrows in the top row (background) are also suppressed in the middle row due to

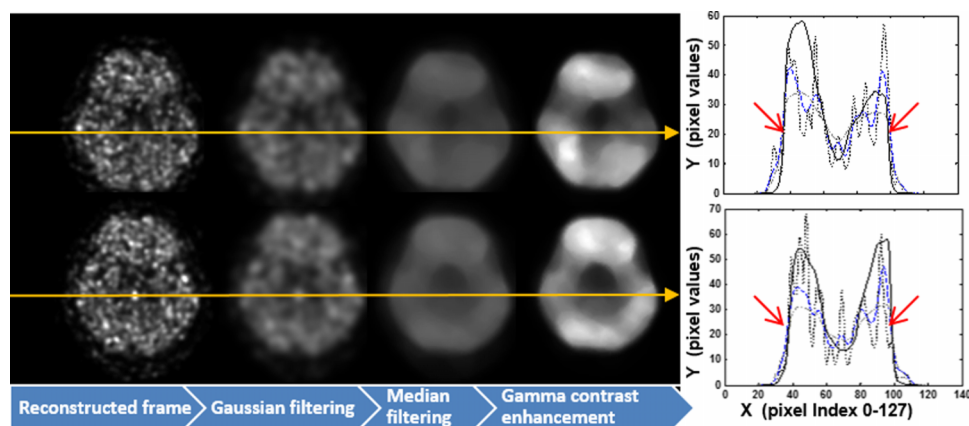


FIG. 2. Effect of our preprocessing method on 5-s reconstructed frames. The two rows show the same slice from two randomly selected 5-s images in a patient with small motion. In each row, the first image on the left is the reconstructed slice showing the level of noise in the frames, the second is after Gaussian filtering, the third after additional median filtering, and the last image is obtained after gamma correction. The line profiles of all four images are plotted together on the right for each row. The line profiles for the unprocessed slices are shown with dotted lines; dashed lines show the Gaussian filtered profiles, the dashed-dotted lines show the median filtered profiles, and the final gamma-corrected profiles are shown in bold black lines. The outer edge of the distribution is indicated with red arrows. Note that the preprocessing sequence keeps the outer edge intact.

nonlinear histogram adjustment by the gamma function. Thus, gamma correction increased the consistency of the distribution across images when motion was small facilitating image registration which is based on maximizing mutual information. In the middle row, the orange arrows indicate better delineation of the ventricles in the brain after preprocessing which are hard to identify in the top row. The bottom row shows the preprocessed slices from the same images after artificial motion was applied. The displacement of the outer edges is clearly seen as the red, green, and blue colors separate to a different extent in proportion to the amount of motion.

## 2.D. Rigid-transform estimation

After preprocessing, the images were registered in 3D using the ITK multiresolution registration algorithm.<sup>37</sup> In this algorithm, a multiresolution pyramid filter framework was used to progressively down-sample the image at successive levels. The transform parameters were estimated at the coarsest level by computing the best transform (in an optimization sense) that aligned one image onto the other and this estimate was then used to initialize the registration at the subsequent finer level. The registration method used a statistical mutual-information

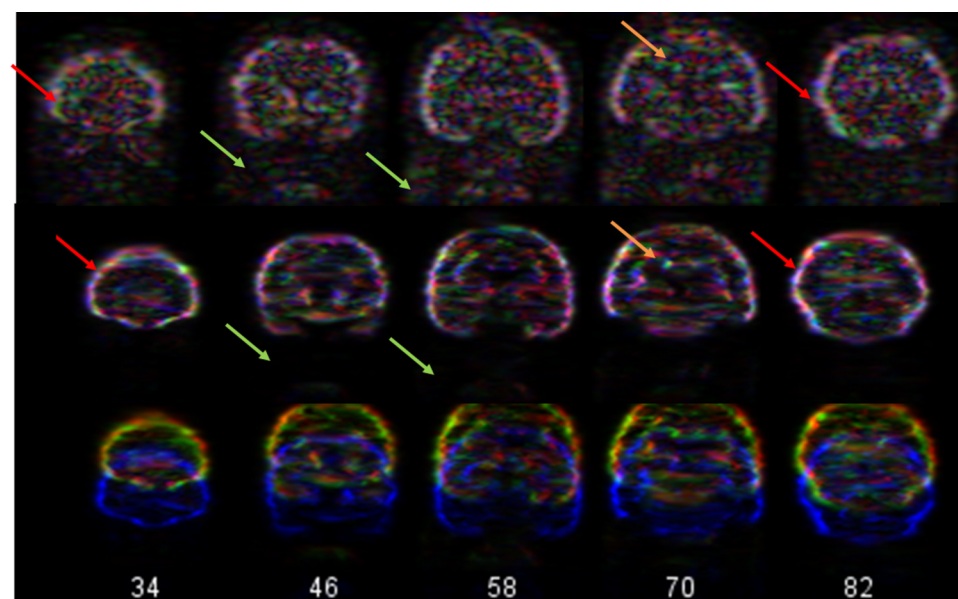


FIG. 3. Five different coronal slices (slices numbers are at the bottom) of the overlays of the edges from three 5-s images of a brain PET study. Edges were extracted using an edge enhancement operator. The top row shows edges extracted after Gaussian filtering only, middle row shows the same after our preprocessing, and the bottom row shows the edges after our preprocessing when artificial motion is applied to the images. The edges for each of the three images are displayed in a different color (red, blue, and green). Thus when there is overlap of all the images, the edge is displayed as white. Note that the outer boundary of the brain indicated by the red arrows is consistent across frames (in absence of artificial motion). The spurious edges in the background region indicated by green arrows in the top row are also suppressed by preprocessing in the middle row. When motion is applied, the displacement of the outer edges is clearly visible as the red, green, and blue colors separate to a different extent in proportion to the amount of motion. (See color online version.)



metric<sup>38</sup> to align the images. Mutual information based registration was chosen so that the differences in reconstructed voxel values between images did not affect the registration process. This method has been previously shown to be effective for interframe registration methods.<sup>25</sup> Image registration was performed using a regular step gradient-descent based optimization algorithm. The optimizer used 6-DOF rigid transform parameters, with 3-DOF for translations and 3-DOF for rotations. No scale parameters were applied or estimated as the images were from the same patient and a single modality. The center of rotation was fixed at the center of the reference image.

To avoid an accumulating error between successive inter-frame registrations, the transform was estimated using the first image in the sequence as a reference for registering subsequent images. We performed this operation for the three frame sets of different temporal duration (5, 10, and 60 s) and for every patient.

## 2.E. Motion compensation

For motion compensation, the rigid-transform estimates determined in Sec. 2.D were used to spatially transform each image in 3D. The transforms estimated by the ITK registration were of the following form:

$$\vec{P}_{ic} - \vec{C} = \mathbf{R}_i * (\vec{P}_i - \vec{C}) + \vec{T}_i, \quad (1)$$

where  $\vec{P}_i$  is the position of a point in the  $i$ th frame and  $\vec{P}_{ic}$  is its compensated position, i.e., transformed to align with the reference image; the rotation is specified by the  $3 \times 3$  matrix  $\mathbf{R}_i$  and the translation given by  $3 \times 1$  vector  $\vec{T}_i$ . The

transform was estimated relative to the fixed center  $\vec{C}$ , i.e.,  $\vec{C} = [63.5, 63.5, 44.5]$  voxels. Therefore,  $\vec{P}_{ic}$  was obtained as

$$\vec{P}_{ic} = \mathbf{R}_i * \vec{P}_i + (\vec{T}_i + \vec{C} - \mathbf{R}_i * \vec{C}), \quad (2)$$

where  $\mathbf{R}_i$  and  $\vec{T}_i + \vec{C} - \mathbf{R}_i * \vec{C}$  constitute the matrix and offset, respectively, that is used to specify the ITK transform for motion compensation. We resampled each image using a linear interpolation scheme in ITK. Once all the images were spatially aligned, we summed them to generate a single motion compensated image.

## 2.F. Assessment of accuracy of motion estimation and compensation

To assess the accuracy of our data-driven motion estimation method, we compared the estimated motion to the “ground truth.” This ground truth was in fact an approximation since the motion between images was a combination of the applied motion and the motion already present in the patient data, for which we only had an estimate. Registration of the no-motion-added baseline images provided this estimate. We compared the ground truth and estimated transforms by applying them both to a fixed point in 3D space and plotting its displacement along  $X$ ,  $Y$ , and  $Z$  axes over time. This point was selected as the centroid of the first 5-s baseline image. The ground truth transform was applied as

$$\vec{P}_{igt} = \mathbf{R}_{igt} * (\vec{P}_0 - \vec{C}) + \vec{T}_{igt} + \vec{C}, \quad (3)$$

where  $\vec{P}_{igt}$  is the position of a point  $\vec{P}_0$  in the  $i$ th image when moved by the ground truth transform. The ground truth

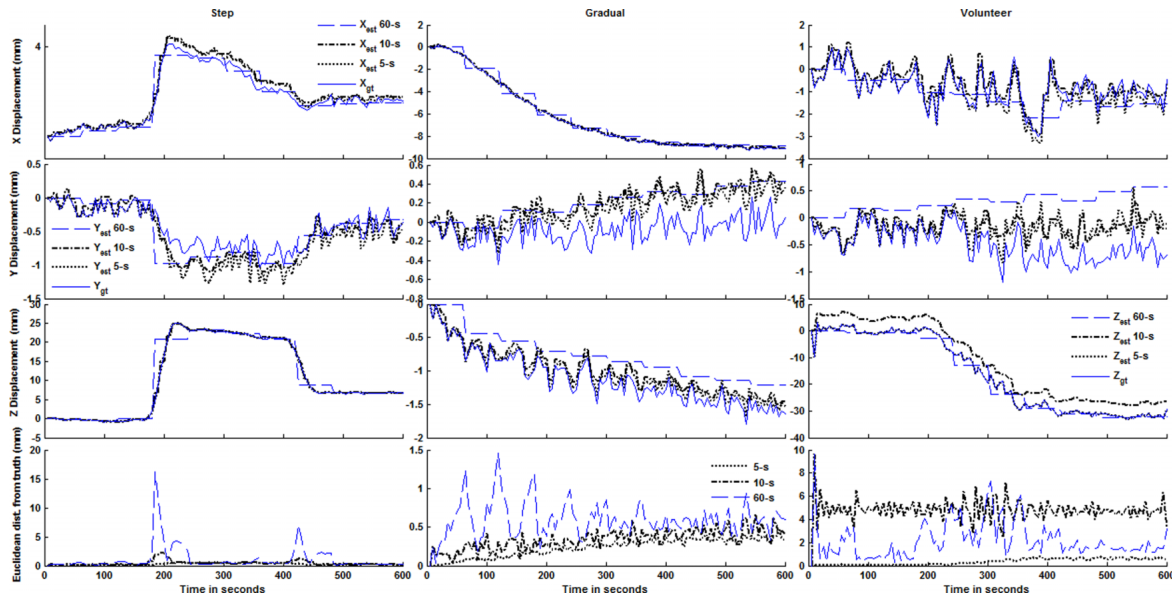


FIG. 4. Change in position of the centroid of the 5-s reference image for estimated versus ground truth transforms for patient 1 in Table I. The first row at the top shows the displacement along  $X$  (lateral), second row along  $Y$  (vertical), and the third row along  $Z$  (axial) direction as estimated by registration of 5-, 10-, 60-s images, and as per the true displacement (ground truth). The columns show the same for each motion type simulated, i.e., Step, Gradual, and Volunteer motion. The Euclidean distance from ground truth is shown in the last row at the bottom. Registration of 5-s images produces the smallest distance error for all motion types.





the summed reconstructed (SR) frames and the full reconstructed (FR) image. We determined that the SR from the 5-s list-mode subsets were on average 3% lower in the total voxel value than the FR.

We also examined the impact of using larger frame sizes on motion estimation. The 10 and 60 s frames were created by summing the 5-s frames. Baseline motion estimates for the 5-s images are summarized in Table I with  $X$  being the lateral direction,  $Y$  the anterior/posterior direction, and  $Z$  being the superior/inferior direction. The pre-existing motion in these eight patients is small as seen from the values of the maximum Euclidean distance from the start of the acquisition (all <2 mm).

The effect of frame size on motion estimation is also shown in Table I, where the standard deviation about the mean position is shown for 5-, 10-, and 60-s frames. The standard deviation is slightly higher for the 5-s frame-based estimation in part due to increased noise in the estimates, but may also indicate pre-existing motion that is estimated more accurately using short duration frames. In patient 3 the actual motion is very small, so most of the variance in the estimate using 5-s frame size is due to noise. However in patient 1 along the  $Z$ -direction, the higher standard deviation is due to the presence of a monotonic trend. This trend can be visualized in Fig. 4 (third row, middle column) as a displacement along  $Z$ -axis for the case of Gradual motion that was simulated along  $X$ -axis only. From Table I we can see that motion estimation using 5-s temporal frame size was not disadvantaged by an increase in statistical variation compared to the longer frames. We also assessed the impact of larger frame size in Sec. 3.B using simulated motion datasets, with the ground truth transformation computed as in Sec. 2.F.

### 3.B. Accuracy of motion estimation versus time interval

Figure 4 shows plots of the spatial displacement of the centroid of the first 5-s baseline image in patient 1 when moved by the ground truth transformation (solid lines) as defined in Eq. (3); and the estimated transformation obtained by registration of 5-s (dotted lines), 10-s (dashed-dotted lines), and 60-s images (dashed lines).

As is evident, the motion estimated from 60-s images is rather coarse and has larger bias probably due to intraframe motion. The motion estimated from 5-s frames (which matches the frequency at which the motion was simulated) is the closest to the ground truth motion. Table II quantitatively summarizes the motion estimation accuracy for various motion types shown in Fig. 4 using the average Euclidean distance (AED) between the position of the centroid when moved by the ground truth and the estimated transforms.

Larger AED implies lower accuracy. The standard deviation of the Euclidean distance is also shown in Table II, with larger values indicating larger deviations in the estimates from ground truth during the imaging time. The AED was the largest for the 60-s frame size due to intraframe motion for all motion types except the Volunteer motion. For the Volunteer motion type, 10-s frame size had an AED of about 5 mm for all patients. This was caused by a reference image containing uncompensated large motion (~10 mm). The reference 10-s image for Volunteer motion type was formed by adding two 5-s images displaced by about 15 mm along the  $Z$ -direction. On one hand, this caused the centroid of the 10-s reference image to be displaced with respect to the 5-s reference image and is seen in Fig. 4 as a bias in the  $Z$ -axis plot (right column,

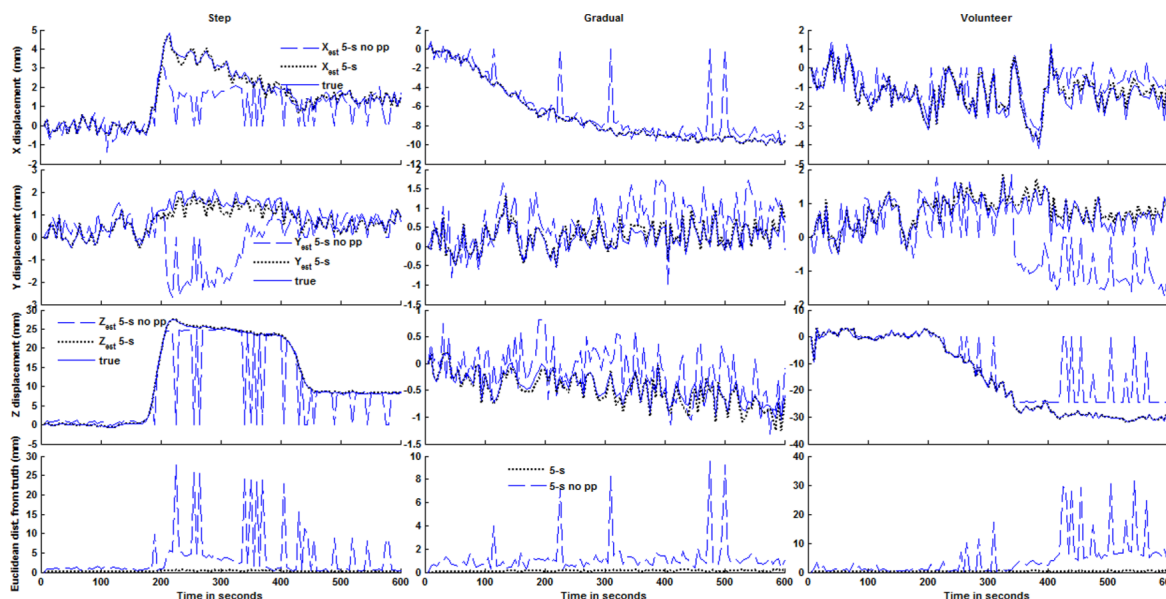


FIG. 5. Change in position of the centroid of the 5-s reference image for motion estimated with and without preprocessing versus the ground truth transform for patient 5 in Table I. The first row at the top shows the displacement along  $X$  (lateral), second row along  $Y$  (vertical), and the third row along  $Z$  (axial) direction as estimated by registration of 5-s images with preprocessing (dotted lines), without preprocessing (dashed lines), and as per the ground truth (solid lines). The columns show the same for each motion type simulated, i.e., Step, Gradual, and Volunteer motion. The Euclidean distance from ground truth is shown in the last row at the bottom. Registration of 5-s images without preprocessing fails completely due to noise for some frames as seen by the large error.





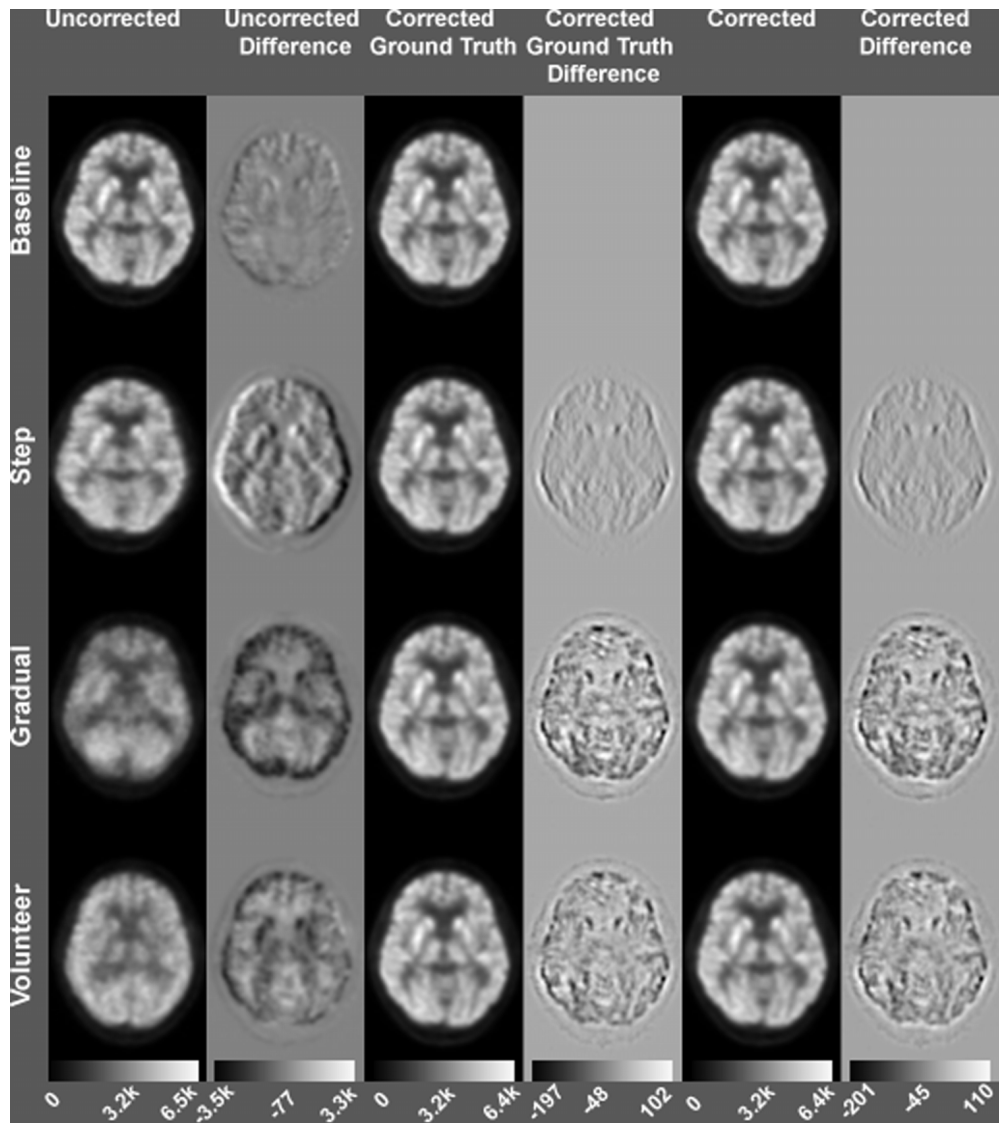


FIG. 6. Motion compensation in patient 1 using 5-s reconstructed frames. Colorbar at the bottom of the figure shows the range of the voxel values in each column.

truth (middle), and compensated with estimated motion (right) slices for the *Step* motion. The third row shows the same for *Gradual* motion, and the last row for the *Volunteer* motion as described in Sec. 2.B. The uncompensated slices for all motion types have the largest difference as expected. The difference image for each uncompensated slice illustrates where the variations are a result of motion. The compensated images with motion estimates (RTC, right column) are very similar to the corrected images with ground truth motion (GTC, middle column, rows 2–4). Note that the RMSDs in Table IV for the RTC images are only slightly higher than the GTC images, confirming the visual impression of how close the two images appear in Fig. 6. This is also evident in the smaller difference image. The *Volunteer* motion type was estimated with larger error compared to the others. The complexity of this motion was also higher compared to the other two types involving all 6-DOFs. The RTC for all motion types was qualitatively restored to the BTC image indicating motion compensation efficacy.

4. DISCUSSION

As can be seen in Table II, we were able to obtain motion estimates within 1 mm accuracy using the 5-s images. The result is that, as can be seen in Fig. 6 where there is very little visual difference in the images using our estimate and the ground truth, there are obvious differences with the uncompensated images. The RMSD values further support the utility of our motion estimation strategy. To appreciate the difficulty of the task of motion estimation, please note the high noise level of the 5-s transverse reconstructions illustrated at the left in Fig. 2. The 5 s duration of these frames represents 1/120th of the data a physician would use for clinical diagnosis. To the best of our knowledge, data-driven motion estimation from static PET frames as small as 5 s has not been reported before. Using frame sizes of 5 s allowed us to estimate gradual monotonic drifts with sub-mm accuracy. Motion estimation using registration allows full 6-DOF motion compensation, which cannot be achieved using sinogram-

based motion estimation as performed in data-driven gating methods.<sup>40</sup> Sinogram-based motion estimation can however be used to identify when the motion occurred and improve the efficiency of our motion estimation scheme by matching the frames to instances of motion occurrence. Alternatively, automatic frame division using tracking methods<sup>41</sup> can be used to divide frames based on motion occurrence. Recent work in marker-less tracking<sup>42</sup> using inexpensive consumer grade depth-sensing cameras shows that motion occurrence can be reliably detected. Further, if an external motion tracking device is available, the registration-based estimates can be used to augment the motion capture accuracy. This is especially useful when acquisition conditions cause a loss of motion capture accuracy, such as obstruction of line of sight, tracking device malfunction, or loss of time-synchronization with the PET list-mode, and insufficient signal to noise ratio due to patient features, garment, etc.

In Jin *et al.*<sup>2</sup> it was noted that image registration may work as well as having true motion information for high count frames, but accuracy may be less for noisy frames. Herein, we have demonstrated a preprocessing method to suppress noise effectively and use image registration for low count PET frames ( $2\text{--}4 \times 10^6$  prompts in the 5-s frames). Table III demonstrates the efficacy of our preprocessing method for reducing the impact of the high noise level in the 5-s frames. Registration of the gross shape of the count distribution using mutual information is not affected by any bias in the reconstructed voxel values which may occur for OSEM-based reconstruction in low count data.<sup>43</sup> However, registering to a reference image with motion artifacts as in the case of the 10-s frames for the *Volunteer* motion causes degradation of accuracy. As the registration used reconstructions without attenuation and scatter correction similar to the methods described in Jin *et al.*,<sup>2</sup> the motion estimates were not affected by motion between CT and PET images. Although we used fixed 5-s frame sizes for registration which is computationally expensive due to the increased number of reconstructions, future parallelization using GPUs on clinical Philips PET systems is expected to reduce the time taken to obtain the motion estimates. Our preprocessing and registration scheme takes about 8 s on a system with 2.7 GHz Intel processor and 8.00 GB RAM for registering two reconstructed frames with  $128 \times 128 \times 90$  voxels irrespective of the frame duration. On the same system, reconstruction of 5-s frames takes 4–5 min using a non-TOF protocol. However, migrating from a CPU to GPU implementation has improved the reconstruction times by at least a factor 10.

A limitation of data-driven motion estimation has been that it is dependent on the specific radiotracer and may not work equally well in other types of radiotracer images. In this work we used brain perfusion images using F-18 FDG and observed that in all the studies, the brain boundary of the reconstructed voxel values was consistent between images when no motion occurred. The success of current data-driven estimation scheme in other radiotracer distributions will rely on the presence of similar consistent features. The underlying assumption was that in practice the motion of a patient's head is also fairly rigid. However, if a change in voxel intensity distribution was not associated with movement such

as a change due to tracer kinetics (e.g., in dynamic studies), or if the movement did not present itself as a rigid pose change, then the accuracy of data-driven estimation will be reduced. In dynamic studies, motion may be estimated using our scheme by registering adjacent frames where the shape of the voxel intensity distribution appears fairly similar.

## 5. CONCLUSION

In conclusion, we have shown that we can estimate and compensate for motion for time frames with time intervals as short as 5 s with an accuracy of 0.3 mm using nonattenuation corrected F-18 FDG PET brain images. Intraframe motion in 60-s frames results in degradation of accuracy to 2 mm or more based on the motion type. Using a reference image with motion artifacts causes lower accuracy as in the case of the estimation of complex *Volunteer* motion from 10-s frames. Finally, appropriate preprocessing is necessary for successful image registration of short 5-s frames.

In future work, we will use patient data with motion during emission and move the CT to the position of the frames using registration-based motion estimates for attenuation and scatter corrected reconstruction. We will also test the effect of varied count distributions of other radioactive tracers (such as FDA approved tracers for imaging of amyloid) on data-driven motion estimation. We will investigate even shorter frames, for example, 1-s, for registration-based motion estimation to determine the threshold where image registration fails due to insufficient count statistics, and integrate the motion compensation transformations into the reconstruction source code or into the list mode events.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) Grant No. R01 EB001457 and a research grant from Philips Medical systems. The contents are solely the responsibility of the authors and do not represent the official views of the NIBIB or Philips Medical Systems. All patient data used were existing and anonymized. It was obtained under IRB approval. A preliminary version of this study was presented at SNMMI 2014 and published in the conference proceedings in abstract form. We would also like to thank Xiyun (Steven) Song from Philips Healthcare for his help in providing PET brain studies.

<sup>a)</sup>J. M. Mukherjee and C. Lindsay contributed equally to this work.

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: joyeeta.mitra@umassmed.edu

<sup>1</sup>K. Dinelle, S. Blinder, J.-C. Cheng, S. Lidstone, K. Buckley, T. J. Ruth, and V. Sossi, "Investigation of subject motion encountered during a typical positron emission tomography scan," in *Nuclear Science Symposium Conference Record* (IEEE, San Diego, CA, 2006), pp. 3283–3287.

<sup>2</sup>X. Jin, T. Mulnix, J. D. Gallezot, and R. E. Carson, "Evaluation of motion correction methods in human brain PET imaging—A simulation study based on human motion data," *Med. Phys.* **40**, 102503 (12pp.) (2013).

<sup>3</sup>B. J. Lopresti, A. Russo, W. F. Jones, T. Fisher, D. G. Crouch, D. E. Altenburger, and D. W. Townsend, "Implementation and performance of

- an optical motion tracking system for high resolution brain PET imaging," *IEEE Trans. Nucl. Sci.* **46**, 2059–2067 (1999).
- <sup>4</sup>R. R. Fulton, S. R. Meikle, S. Eberl, J. Pfeiffer, and C. J. Constable, "Correction for head movements in positron emission tomography using an optical motion-tracking system," *IEEE Trans. Nucl. Sci.* **49**, 116–123 (2002).
  - <sup>5</sup>R. D. Wiersma, S. L. Tomarken, Z. Grelewicz, A. H. Belcher, and H. Kang, "Spatial and temporal performance of 3D optical surface imaging for real-time head position tracking," *Med. Phys.* **40**, 111712 (8pp.) (2013).
  - <sup>6</sup>S. R. Goldstein, M. E. Daube-Witherspoon, M. V. Green, and A. Eidsath, "A head motion measurement system suitable for emission computed tomography," *IEEE Trans. Med. Imaging* **16**, 17–27 (1997).
  - <sup>7</sup>P. M. Bloomfield, T. J. Spinks, J. Reed, L. Schnorr, A. M. Westrip, L. Livieratos, R. Fulton, and T. Jones, "The design and implementation of a motion correction scheme for neurological PET," *Phys. Med. Biol.* **48**, 959–978 (2003).
  - <sup>8</sup>V. Zhou, A. Kyme, S. Meikle, and R. Fulton, "An event-driven motion correction method for neurological PET studies of awake laboratory animals," *Mol. Imaging Biol.* **10**, 315–324 (2008).
  - <sup>9</sup>O. V. Olesen, J. M. Sullivan, T. Mulnix, R. R. Paulsen, L. Hojgaard, B. Roed, R. E. Carson, E. D. Morris, and R. Larsen, "List-mode PET motion correction using markerless head tracking: Proof-of-concept with scans of human subject," *IEEE Trans. Med. Imaging* **32**, 200–209 (2013).
  - <sup>10</sup>P. Bühler, U. Just, E. Will, J. Kotzerke, and J. Van den Hoff, "An accurate method for correction of head movement in PET," *IEEE Trans. Med. Imaging* **23**(9), 1176–1185 (2004).
  - <sup>11</sup>J. Langner, "Event-driven motion compensation in positron emission tomography: Development of a clinically applicable method," Ph.D. dissertation, University of Technology, Dresden, 2008.
  - <sup>12</sup>M. G. Ullisch, J. J. Scheins, C. Weirich, E. Rota Kops, A. Celik, L. Tellmann, T. Stocker, H. Herzog, and N. J. Shah, "MR-based PET motion correction procedure for simultaneous MR-PET neuroimaging of human brain," *PLoS One* **7**(11), e48149 (13pp.) (2012).
  - <sup>13</sup>M. V. Green, J. Seidel, S. D. Stein, T. E. Tedder, K. M. Kempner, C. Kertzman, and T. A. Zeffiro, "Head movement in normal subjects during simulated PET brain imaging with and without head restraint," *J. Nucl. Med.* **35**, 1538–1546 (1994).
  - <sup>14</sup>R. Fulton, L. Tellmann, U. Pietrzyk, O. Winz, I. Stangier, I. Nickel, A. Schmid, S. Meikle, and H. Herzog, "Accuracy of motion correction methods for PET brain imaging," in *Nuclear Science Symposium Conference Record* (IEEE, Rome, Italy, 2004), pp. 4226–4230.
  - <sup>15</sup>H. Herzog, L. Tellman, R. Fulton, and U. Pietrzyk, "Motion correction in PET brain studies," in *The Fourth International Workshop on Multidimensional Systems* (IEEE, Wuppertal, Germany, 2005), pp. 178–181.
  - <sup>16</sup>L. Tellmann, R. R. Fulton, K. Bente, I. Stangier, O. Winz, U. Just, H. Herzog, and U. K. Pietrzyk, "Motion correction of head movements in PET: Realisation for routine usage," in *Nuclear Science Symposium Conference Record* (IEEE, Portland, OR, 2003), pp. 3105–3107.
  - <sup>17</sup>J. Verhaeghe, P. Gravel, R. Mio, R. Fukasawa, P. Rosa-Neto, J. P. Soucy, C. J. Thompson, and A. J. Reader, "Motion-compensated fully 4D PET reconstruction using PET data supersets," in *Nuclear Science Symposium Conference Record (NSS/MIC)* (IEEE, Orlando, Florida, 2009), pp. 3000–3004.
  - <sup>18</sup>S. H. Keller, M. Sibomana, O. V. Olesen, C. Svarer, S. Holm, F. L. Andersen, and L. Hojgaard, "Methods for motion correction evaluation using 18F-FDG human brain scans on a high-resolution PET scanner," *J. Nucl. Med.* **53**, 495–504 (2012).
  - <sup>19</sup>H. Mohy-ud-Din, N. A. Karakatsanis, J. S. Goddard, J. Baba, W. Wills, A. K. Tahari, D. F. Wong, and A. Rahmim, "Generalized dynamic PET inter-frame and intraframe motion correction: Phantom and human validation studies," in *Nuclear Science Symposium Conference Record (NSS/MIC)* (IEEE, Anaheim, CA, 2012), pp. 3067–3078.
  - <sup>20</sup>H. Mohy-ud-Din, N. A. Karakatsanis, W. Willis, A. K. Tahari, D. F. Wong, and A. Rahmim, "Intraframe motion compensation in multi-frame brain PET imaging," *Front. Biomed. Technol.* **2**(2), 366–380 (2015).
  - <sup>21</sup>T. L. Faber, N. Raghunath, D. Tudorascu, and J. R. Votaw, "Motion correction of pet brain images through deconvolution. I. Theoretical development and analysis in software simulations," *Phys. Med. Biol.* **54**(3), 797–811 (2009).
  - <sup>22</sup>Y. Picard and C. J. Thompson, "Motion correction of PET images using multiple acquisition frames," *IEEE Trans. Med. Imaging* **16**, 137–144 (1997).
  - <sup>23</sup>P. J. Noonan, J. Howard, T. F. Cootes, W. A. Hallett, and R. Hinz, "Realtime markerless rigid body head motion tracking using the Microsoft Kinect," in *Nuclear Science Symposium Conference Record (NSS/MIC)* (IEEE, Anaheim, CA, 2012), pp. 2241–2246.
  - <sup>24</sup>O. Olesen, R. Paulsen, R. Jensen, S. Keller, M. Sibomana, L. Hojgaard, B. Roed, and R. Larsen, "3D surface realignment tracking for medical imaging: A phantom study with PET motion correction," in *Image-Based Geometric Modeling and Mesh Generation*, edited by Y. Zhang (Springer, Netherlands, 2013), Vol. 3, pp. 11–19.
  - <sup>25</sup>K. Chen, O. Smilovici, W. Lee, C. Reschke, Z. Zhu, D. Bandy, and E. Reiman, "Inter-frame co-registration of dynamically acquired fluorodeoxyglucose positron emission tomography human brain data," in *IEEE International Conference on Complex Medical Engineering* (IEEE, Beijing, China, 2007), pp. 901–906.
  - <sup>26</sup>N. Costes, A. Dagher, K. Larcher, A. C. Evans, D. L. Collins, and A. Reilhac, "Motion correction of multi-frame PET data in neuroreceptor mapping: Simulation based validation," *NeuroImage* **47**, 1496–1505 (2009).
  - <sup>27</sup>A. J. Montgomery, K. Thielemans, M. A. Mehta, F. Turkheimer, S. Mustafaovic, and P. M. Grasby, "Correction of head movement on PET studies: Comparison of methods," *J. Nucl. Med.* **47**, 1936–1944 (2006).
  - <sup>28</sup>J. L. Andersson, B. E. Vagnhammar, and H. Schneider, "Accurate attenuation correction despite movement during PET imaging," *J. Nucl. Med.* **36**, 670–678 (1995).
  - <sup>29</sup>A. Rahmim, O. Rousset, and H. Zaidi, "Strategies for motion tracking and correction in PET," *PET Clin.* **2**, 251–266 (2007).
  - <sup>30</sup>J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imaging* **22**, 986–1004 (2003).
  - <sup>31</sup>H. Herzog, L. Tellmann, R. Fulton, I. Stangier, E. Rota Kops, K. Bente, C. Boy, R. Hurlmann, and U. Pietrzyk, "Motion artifact reduction on parametric PET images of neuroreceptor binding," *J. Nucl. Med.* **46**, 1059–1065 (2005).
  - <sup>32</sup>X. Jin, T. Mulnix, C. M. Sandiego, and R. E. Carson, "Evaluation of frame-based and event-by-event motion-correction methods for awake monkey brain PET imaging," *J. Nucl. Med.* **55**, 287–293 (2014).
  - <sup>33</sup>J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans. Med. Imaging* **15**, 687–699 (1996).
  - <sup>34</sup>J. E. McNamara, P. H. Pretorius, K. Johnson, J. M. Mukherjee, J. Dey, M. A. Gennert, and M. A. King, "A flexible multicamera visual-tracking system for detecting and correcting motion-induced artifacts in cardiac SPECT slices," *Med. Phys.* **36**, 1913–1923 (2009).
  - <sup>35</sup>M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 3rd ed. (Thompson Learning, Toronto, 2008).
  - <sup>36</sup>J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: An open-source platform for biological-image analysis," *Nat. Methods* **9**, 676–682 (2012).
  - <sup>37</sup>T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing Api: A technical report on ITK—the Insight Toolkit," *Stud. Health Technol. Inf.* **85**, 586–592 (2002).
  - <sup>38</sup>D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imaging* **22**, 120–128 (2003).
  - <sup>39</sup>D. Gagnon, O. Patrick, and K. K. Parmeshwar, U.S. patent 8,515,148 (20 August 2013).
  - <sup>40</sup>P. J. Schleyer, K. Thielemans, and P. K. Marsden, "Extracting a respiratory signal from raw dynamic PET data that contain tracer kinetics," *Phys. Med. Biol.* **59**, 4345–4356 (2014).
  - <sup>41</sup>O. V. Olesen, S. H. Keller, M. Sibomana, R. Larsen, B. Roed, and L. Hojgaard, "Automatic thresholding for frame-repositioning using external tracking in PET brain imaging," in *Nuclear Science Symposium Conference Record (NSS/MIC)* (IEEE, Knoxville, TN, 2010), pp. 2669–2675.
  - <sup>42</sup>C. Lindsay, J. M. Mukherjee, K. Johnson, P. Olivier, X. Song, L. Shao, and M. A. King, "Marker-less multi-frame motion tracking and compensation in PET-brain imaging," *Proc. SPIE* **9417**, 94170J (7pp.) (2015).
  - <sup>43</sup>A. Reilhac, S. Tomei, I. Buvat, C. Michel, F. Keheren, and N. Costes, "Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies," *NeuroImage* **39**, 359–368 (2008).