Optimizing the Frame Duration for Data-Driven Rigid Motion Estimation in Brain PET Imaging

Matthew G. Spangler-Bickell[1],
Department of Radiology, University of Wisconsin, Madison, WI, USA.

Samuel A. Hurley,
Department of Radiology, University of Wisconsin, Madison, WI, USA.

Timothy W. Deller,
PET/MR Engineering, GE Healthcare, Waukesha, WI, USA.

Floris Jansen,
PET/MR Engineering, GE Healthcare, Waukesha, WI, USA.

Valentino Bettinardi,
Nuclear Medicine Unit, IRCCS Ospedale San Raffaele, Milan, Italy.

Mackenzie Carlson,
Department of Radiology, Stanford University, Stanford, CA, USA.

Michael Zeineh,
Department of Radiology, Stanford University, Stanford, CA, USA.

Greg Zaharchuk,
Department of Radiology, Stanford University, Stanford, CA, USA.

Alan B. McMillan
Department of Radiology, University of Wisconsin, Madison, WI, USA.

Abstract

Purpose: Data-driven rigid motion estimation for PET brain imaging is usually performed using data frames sampled at low temporal resolution to reduce the overall computation time and to provide adequate signal-to-noise ratio in the frames. In recent work it has been demonstrated that list-mode reconstructions of ultra-short frames are sufficient for motion estimation and can be performed very quickly. In this work we take the approach of using image-based registration of reconstructions of very short frames for data-driven motion estimation, and optimize a number of reconstruction and registration parameters (frame duration, MLEM iterations, image pixel

Corresp. author: M. Spangler-Bickell (matthew.spangler-bickell@ge.com, address: 3200 North Grandview Boulevard, Waukesha, WI, 53188, USA).
[1]M. Spangler-Bickell is now with PET/MR Engineering, GE Healthcare, Waukesha, WI, USA.

size, post-smoothing filter, reference image creation, and registration metric) to ensure accurate registrations while maximizing temporal resolution and minimizing total computation time.

Methods: Data from ^{18}F-fluorodeoxyglucose (FDG) and ^{18}F-florbetaben (FBB) tracer studies with varying count rates are analysed, for PET/MR and PET/CT scanners. For framed reconstructions using various parameter combinations inter-frame motion is simulated and image-based registrations are performed to estimate that motion.

Results: For FDG and FBB tracers using 4×10^5 true and scattered coincidence events per frame ensures that 95% of the registrations will be accurate to within 1 mm of the ground truth. This corresponds to a frame duration of 0.5 –. 1 sec for typical clinical PET activity levels. Using 4 MLEM iterations with no subsets, a transaxial pixel size of 4 mm, a post-smoothing filter with 4–6 mm full-width at half-maximum, and averaging two or more frames to create the reference image provides an optimal set of parameters to produce accurate registrations while keeping the reconstruction and processing time low.

Conclusions: It is shown that very short frames (1 sec) can be used to provide accurate and quick data-driven rigid motion estimates for use in an event-by-event motion corrected reconstruction.

Keywords

I.   Introduction

Head motion in PET brain imaging is a substantial problem which causes loss in diagnostic value of the reconstructed images due to motion blur, or in the worst case renders them unusable. Even compliant patients struggle to remain completely motionless for the duration of most clinical scans which can range from several minutes to up to an hour. Devices such as head restraints are used to mitigate motion, but with the improved spatial resolution of modern scanners (3–4 mm) even motion on the order of millimetres can have a substantial effect.

Using an external tracking device to monitor head motion, followed by a full event-by-event rigid motion corrected list-mode reconstruction, has been shown to be a very successful solution to this problem[1,2]. However the external hardware and the subsequent impact on the clinical routine have hindered this solution's implementation in the clinic.

Data-driven techniques which estimate the rigid motion directly from the PET data are promising since they are completely software-based and have no impact on the clinical routine. They are also easily portable between scanners and can be applied retrospectively. The primary concern with such approaches is a low temporal resolution and poor motion accuracy, but, as will be shown, these issues are being solved.

Data-driven techniques generally fall into two groups, both of which split the data up into many frames. The first approach uses some format of the data, e.g. downsampled

sinograms[3], centroid-of-density calculations[4], or full reconstructions[5,6], to identify time points where motion occurred. The data are reframed according to these time points, reconstructions of each new frame are performed, and the subsequent reconstructions are aligned and combined. This is usually referred to as the multiple acquisition frame approach based on work by Picard *et al*.[7]. The second approach uses the framed data to directly estimate the 6 degrees-of-freedom describing the head position. These estimations can then be used to perform a single fully motion corrected (usually list-mode) reconstruction. This is preferable since superior noise properties are achieved when performing a single reconstruction of all the data rather than summing framed reconstructions, because of the non-negativity constraint of the maximum-likelihood expectation-maximization (MLEM) algorithm[8]. Feng *et al*.[9] used the first and second inertial moments of the list-mode data to derive the motion (and Rezaei *et al*.[10] extended this for time-of-flight data), and several groups have used the fully reconstructed frames to estimate the motion using image registration[11,12]. In the case of a largely static tracer distribution, registering the reconstructed frames to a reference image is the ideal approach since it can provide very accurate motion estimates. However, due to the combined factors of low count statistics per frame and typically long reconstruction times, it is common to use frame durations on the order of tens of seconds or longer. Thus intra-frame motion may not be detected or may bias the images leading to an inaccurate motion estimation.

Image registration, which forms the basis of the motion estimation in this work, is an important aspect of medical imaging. In the case of motion estimation for brain imaging, registration is generally straightforward since it is applied to the same patient in the same imaging session for a single modality. Various groups have studied how to optimize registration for brain imaging[13,14]. Usually a mean square difference or a cross correlation metric is used for static data where the pixel intensities are similar between the images, while the mutual information metric[15] is usually used for dynamic data where that assumption might not be true.

In previous work we have demonstrated the ability to take advantage of the improved sensitivity and timing resolution of modern scanners to very quickly produce reconstructions of ultra-short frames of list-mode data[16]. We now present an investigation into the optimal frame duration to provide high temporal sampling while maintaining a good image registration accuracy for the purpose of registration-based motion estimation (extending work in[17]). These estimates can then be used to perform an event-by-event motion corrected list-mode reconstruction of the data. List-mode reconstructions provide complete freedom to frame the data and thus static or dynamic reconstructions can be performed, although in the latter case, due to the tracer dynamics, the registration-based approach needs special considerations.

II.   Method

II.A.   Data

Five PET data sets were used during this study, the use of which were approved by the relevant IRBs. The data were acquired on either a SIGNA PET/MR (GE Healthcare, Chicago, IL, USA) at UW Madison and Stanford University, or a D710 PET/CT (GE

Healthcare) at UW Madison. The SIGNA and D710 scanners have, respectively, a 4.4 mm and 4.52 mm radial resolution, an axial field-of-view of 25 cm and 15.7 cm, a sensitivity of 22.9 cps/kBq and 7.5 cps/kBq, a time-of-flight (TOF) resolution of 390 ps and 500 ps, and an energy resolution of 10.3% and 12.4%[18,19,20]. The data sets were:

A. An ^{18}F-fluorodeoxyglucose (FDG) research study on the SIGNA with a high injected activity (504 MBq, 27 min uptake time).

B. The same data as in (A) but with the events subsampled by a factor of 3.

C. An FDG clinical study on the SIGNA with a standard protocol (183 MBq injected, 70 min uptake time).

D. An ^{18}F-florbetaben (FBB) clinical study on the SIGNA with a standard protocol (245 MBq injected, 43 min uptake time).

E. An FDG clinical study on the D710 with a standard protocol (412 MBq injected, 49 min uptake time).

A summary of the data sets is given in Table 1. Data Sets A-D were acquired with an accompanying external motion tracker (HobbitView Inc., CA, USA)[21] which tracked the head motion to a high degree of accuracy and precision (0.15° rotations and 0.39 mm translations)[22] throughout the scan. This enabled us to identify a 200 sec segment of each data set where the motion of the patient's head was insignificant (< 1° rotations and < 1 mm translations). An example of the recorded motion over the chosen 200 sec segment for one data set is shown in Fig. 1. No external motion tracking was available for Data Set E, but the patient's head was restrained, and using the image registration techniques described in this paper it was verified as conclusively as possible that there was negligible motion during the chosen portion of data (< 0.5° rotations and < 0.5 mm translations were observed).

For all data sets 40 ultra-short frames were reconstructed with the frames separated by 5 sec, yielding 40 noise realizations spanning the 200 sec of data without overlap. A cohort of 40 noise realisations was used for each combination of parameter settings (e.g. frame duration, pixel size, etc.) as detailed in section II.E..

All reconstructions were performed using pure list-mode time-of-flight MLEM (with no subsets), as described in[16]. A low number of iterations was performed (between 2 and 8, as described in section II.E.) to ensure quick reconstructions. This was found to be sufficient for the purposes of motion estimation. An optimized ray-tracing projector was used without any point spread function modeling. The transaxial field-of-view was set to 300 mm. Attenuation and scatter correction were not performed since, if the attenuation map was misaligned with the PET data, these could bias the motion estimates towards the attenuation map. Randoms correction was performed using the singles count rate measured during the frame duration under consideration. All other corrections, i.e. normalization, deadtime, and decay, were performed. Example reconstructions for three frame durations are shown in Fig. 2 for Data Sets A and D.

II.B. Influence of Activity Level and Randoms

In order to make a general statement to guide the choice of frame duration it is necessary to know *a priori* what signal-to-noise ratio is expected for a given reconstruction. Due to variations in the injected tracer activity it is more appropriate to consider the number of coincidence events (or counts) per frame, rather than frame duration, in order to know what noise level to expect. Additionally, since the number of random coincidence events increases with the square of the activity, the number of true and scattered coincidence events is a more accurate indicator of the expected reconstruction noise level[23]. For example, if 1×10^5 true and scattered coincidence events are reconstructed from a low activity and a high activity data set then the resultant images will have similar signal-to-noise ratios but very different frame durations and randoms fractions (note that the high activity data set will have a slightly lower signal-to-noise ratio due to the increased presence of random events).

To verify this Data Sets A - D will be used, which all have different coincidence event rates.

II.C. Simulated motion

A diagram of the process of simulating motion is shown in Fig. 3. For a chosen set of parameters (e.g. frame duration, pixel size, etc.), the 40 ultra-short frames were reconstructed to produce the reference images for the registrations. Following this, for each of the 40 frames of list-mode data a single transformation was applied to the list-mode events (and similarly for the sensitivity image used in the MLEM reconstruction), which were then reconstructed again to produce the "moved" images. This was repeated for three transformations, given here as $[R_x, R_y, R_z, X, Y, Z]$ in degrees and millimetres:

Static [0, 0, 0, 0, 0, 0] (i.e. effectively no transformation was applied).

Motion A [4, −4, 4, −4, 4, 4]

Motion B [−8, 6, −12, 8, −8, −8]

These transformations represent motion which we have observed within about 50 clinical cases at our institution using data-driven motion estimation, with Motion B representing less common high motion cases. Similar values have been reported in, for example,[24,25]. These simulated motions represent only inter-frame motion; intra-frame motion is not considered. The aim of this work is to demonstrate that frames of 1 sec duration provide accurate registrations. In most clinical scenarios intra-frame motion within this time range would be very limited and have little impact on the registration accuracy.

Since it was verified that negligible patient motion was present in this data, when registering to the reference images the applied transformation can be treated as the ground truth when assessing the accuracy of the registration.

For our analyses we used a new reference image for each registration to ensure that we have a set of unique noise realizations. However in a real world scenario it is recommended to use the same reference frame for all registrations to avoid introducing any drift into the motion estimates.

II.D. Image Registration

Rigid registrations were performed using MATLAB's built-in functions (namely imregtform)[26]. A mean square difference metric was used with a gradient descent optimizer. The stopping condition was when the gradient magnitude fell below a small tolerance (10^{-9}). A mutual information metric was also investigated although it is not usually used for single-modal and static data registration. The registration sampled the image fully (i.e. no random pixel sampling was used) and used a 3 level multi-resolution approach where the images were downsampled by a factor of 4, then 2, then with no downsampling. Planes inferior to approximately the upper jaw region were cropped from the images since the neck does not move rigidly with the head. Several planes superior to and outside of the head were also cropped to remove possibly noisy edge planes. For the neck region cropping the image is preferable to masking since it avoids introducing false edges into the image. A mask was applied which encompassed the support of the head with a large margin. The mask was heavily smoothed to avoid any hard edges, although it was entirely outside of active regions.

Each "moved" image was registered to the reference image corresponding to the previous frame or mean of several frames, i.e. if X frames were averaged over to create the reference image, then frame N was registered to the mean of frames $(N-X)$ to $(N-1)$. Thus there were $(40-X)$ unique registrations for each parameter setting. Utilizing the preceding frames as the reference further ensures that there is negligible real motion between the frames. Observed mean residual displacement between frames was on the order of 0.08 mm.

The accuracy of a particular registration was quantified as follows (see Fig. 3): a 14-point ellipsoidal mesh was automatically generated with diameter of [150, 120, 100] mm in the anterior-posterior, left-right, inferior-superior directions, respectively, and was located to approximately coincide with the position of the brain under consideration. These mesh points were then transformed according to the ground truth transformation as well as the registration parameters. The pair-wise Euclidean distances between the points in these two sets were calculated and averaged yielding a "mesh displacement error". The mesh displacement error was averaged over all the noise realizations, and this "mean mesh error" was used as a metric for the accuracy of the registration. To report the spread of the mesh displacement errors the 95th percentile was calculated, i.e. the value below which 95% of the mesh displacement errors fell.

II.E. Investigated Parameters

The primary investigated parameter was the frame duration. As shown in Table 2 the frame duration was varied between 20 ms and 5 sec. In all cases the 40 frames were separated by 5 sec, thus spanning the 200 sec of data without overlap. Note that frame duration is used as a more relatable proxy for the number of coincidence events per frame, but the final conclusion will necessarily be for the number of coincidence events in order to be general, as discussed in Section II.B. For this analysis Data Sets A, D, and E were used to demonstrate the effect of different scanners (SIGNA and D710) and tracers (FDG and FBB). Note that for the FBB tracer, similar to FDG, the distribution of this tracer usually involves the whole brain, particularly for amyloid positive patients.

Additionally, a number of reconstruction and registration settings were investigated for their effect on registration accuracy: the number of MLEM iterations used for the reconstruction (more iterations yield less biased but noisier images, and take longer to execute); the transaxial pixel size of the reconstruction (smaller pixel sizes yield higher resolution images but with increased reconstruction time, the axial pixel size was kept fixed at 2.78 mm); the full-width-at-half-maximum (FWHM) of the image smoothing applied after reconstruction (more smoothing may improve registration accuracy to a point); the number of frames averaged together to create the reference image for the registration (more frames yield a less noisy reference); and the registration metric used (the mean square difference metric versus the mutual information metric). These parameters and their ranges are summarized in Table 2.

It was not feasible to optimize over the entire 6 dimensional parameter space. In preliminary work various combinations of parameters were tested, including with different frame durations, but for the final analysis when one parameter was being tested all of the other parameters were held at the bold-faced values shown in Table 2.

The 40 frames were reconstructed using a set of parameter values, and this was repeated for each of the three transformations; a diagram of the process is given in Fig. 3. Data Set A was used for all of these investigations. Except for the frame duration analysis all of the reconstructions used a frame duration of 0.5 sec which corresponded to approximately 3.15×10^5 true and scattered counts per frame.

III. Results

III.A. Influence of Activity and Randoms

Fig. 4 demonstrates the effect of the activity level and randoms fraction on the registration accuracy. The registration accuracy is plotted against both the frame duration and number of counts per frame. For a particular frame duration the registration accuracy across different data sets improves with increasing activity levels (Fig. 4, left), but when considering a particular number of true and scattered counts per frame the registration accuracy values are very similar across the data sets (Fig. 4, right). This indicates that even for very different tracer activity levels the registration accuracy can be predicted with some certainty using the number of true and scattered counts per frame. Data Set D has been included here even though it is a different tracer since it also follows the same behavior. Note that Data Set C, having the lowest randoms fraction and therefore less noisy images, does provide a slightly superior registration at the same count level particularly at very low count frames.

III.B. Investigated Parameters

Fig. 5 shows the results of the frame duration analysis for Data Set A, D, and E. The registration accuracy, as measured by the mean mesh error, improves with increasing number of counts per frame. For Data Set A, in all motion cases, for frames with 4.4×10^5 true and scattered counts, at least 95% of the noise realizations had registrations with a mesh error of less than 1 mm (the maximum error was 1.3 mm). For this data set this corresponds to a frame duration of 0.7 sec. For Data Set D this occurred for frames with at least 3.7×10^5

true and scattered counts (1.5 sec duration, 1.0 mm maximum error), and for Data Set E for frames with at least 3.4×10^5 true and scattered counts (1.2 sec duration, 1.1 mm maximum error). The mean mesh errors in the Static case are much smaller, which is not unexpected since in the Static case the registration initialization (i.e. the identity) is very close to the ground truth.

Fig. 6 shows the results from the investigation into the effect of various parameters on the registration accuracy. As before, the mesh errors for the Static case are smaller than those from the Motion cases, but this is not unexpected. The Motion cases are the most interesting since they more closely simulate real-world scenarios. The results do not indicate a very strong dependence of the registration accuracy on any of the parameters, but there are clear trends nonetheless. Very similar results were observed when using a frame duration of 1 sec (not shown here).

It was found that the mutual information metric was more susceptible to image noise than the mean square metric, as can be seen in Fig. 7 where the dependence of the registration accuracy on the number of counts per frame is shown for the two metrics. While we could not identify a source in the literature which investigated the performance of the mutual information metric in the presence of significant noise, it stands to reason that since it utilizes the joint entropy histogram increased noise in the images would disperse that histogram and possibly increase the likelihood of converging to a local maximum. It should be noted that the mutual information metric is usually used for multi-modality cases or where the tracer distribution varies between frames[27].

IV. Analysis

IV.A. Scanner Considerations

Certain scanner specifications might affect the results we report here: a different sensitivity (which is affected by the axial field-of-view, amongst other factors) as well as a different energy resolution would imply a different scatter fraction and thus require a different optimal number of true and scattered counts per frame, and a different TOF resolution would influence the quality and noise properties of the reconstructed images. Varying axial coverage also changes how much of the head is available for image registration. We have presented results from two different scanner geometries and demonstrated similarities in registration accuracy when frame duration is based on count rate. We expect that this investigation should approximately translate to most clinical scanners, but a new investigation may be required for systems with very different specifications.

IV.B. Execution Time Considerations

The execution time of the reconstruction and registration is a significant factor which underlies all the other results already presented. The optimal parameter settings need to be balanced against their effect on the execution time. The most significant factors are the number of iterations used for the reconstruction (which has a linear relationship with the reconstruction time) and the transaxial pixel size (which has an inverse squared relationship with the reconstruction time). Note that the frame duration is not a concern

since computation time for list mode reconstruction is proportional to the number of counts, therefore the total time to reconstruct a series of frames covering a set amount of data is approximately constant regardless of how many frames are used (except for some overhead per frame).

Our reconstruction implementation is described in[16] and a more thorough discussion of the reconstruction time can be found there. The reconstruction of each frame is executed much faster than the duration of the data in the frame, depending on the number of events in each frame. Typically for frames of 1 sec duration (4 MLEM iterations, 4 mm transaxial pixel size) the reconstruction is executed in less than 0.5 sec on a standard 8-core 2.4 GHz Intel Core i9 CPU system, processing approximately 1.5×10^6 prompt events per second. Further acceleration can be achieved with a more powerful system or a GPU implementation. The registrations for the frame duration analyses shown in Figs. 4 and 5 using MATLAB's functions were executed in parallel with a mean processing time of approximately 0.25 sec per frame. The total time taken to estimate the motion for a typical clinical scan is on the order of minutes and is by no means computationally prohibitive.

IV.C.   Investigated Parameters

We chose a figure of merit (FOM) of 1 mm registration accuracy to identify parameters whose 95th percentile will be more accurate than that FOM.

IV.C.1.   Number of Counts per Frame—Fig. 5 demonstrates that using 4.4×10^5 true and scattered counts per frame satisfies the registration accuracy FOM of 1 mm for FDG and FBB studies, and for the SIGNA and D710 scanners. Fig. 4 shows that this conclusion holds across varying activity levels and randoms fractions since all data sets satisfy the accuracy FOM when using $3.0 - 4.4 \times 10^5$ true and scattered counts per frame. The fact that FDG and FBB show very similar results can be explained by the fact that both tracers have distribution patterns involving much of the brain (an example of which can be seen in Fig. 2) and therefore similar noise properties for a given number of true and scattered counts. Thus the conclusion of number of counts per frame should be applicable to almost any brain study where the brain uptake is distributed throughout the brain and is largely consistent across the patient population. In cases of severe pathologies where this assumption is not true (e.g. where a large brain tumor exists for FDG studies) it is likely that the distribution of activity would actually aid the registration, or at least not degrade it.

It can be concluded for FDG and FBB scans that by using at least 4×10^5 true and scattered counts per frame a registration accuracy of 1 mm can be achieved. The frame duration needed to achieve this count level will depend on the injected activity. By taking the difference between the prompt rate and the randoms event rate (calculated, for example, from the single events rate reported in the list-mode data), the frame duration can be set to ensure that each frame contains the desired number of true and scattered coincidence events (this may need to be corrected for decay over time). In most clinical cases the frame duration should be 0.5 − 1 sec.

IV.C.2.   Number of Reconstruction Iterations—Fig. 6(a) demonstrates that very similar results are achieved in the moving cases when using 4 iterations or more, with

6 and 8 satisfying the defined accuracy FOM. However when accounting for execution time considerations we conclude that using 4 iterations provides sufficiently accurate registrations.

IV.C.3. Pixel Size—Fig. 6(b) shows that using images with pixels in the 2 – 5 mm range produces very similar results in the moving case, with slightly higher spread in the registration accuracy for the larger pixels. Since the pixel size has the largest impact on the reconstruction time we conclude that a transaxial pixel size of 4 mm is appropriate to ensure accurate registrations.

IV.C.4. Post-smoothing FWHM—Fig. 6(c) indicates that the FWHM of the smoothing applied to the reconstructions before registration does not have a large impact on the registration accuracy. This could be due to the multiple resolution approach employed during the registration which behaves like a smoothing by downsampling the image. However, even though the impact of the smoothing is small, it is prudent to apply some smoothing to avoid any confounding effects due to noise, and therefore a FWHM of 4 – 6 mm is recommended.

IV.C.5. Number of Averaged Frames for Reference Image—Fig. 6(d) indicates that averaging over at least 2 frames improves image registration accuracy due to reduced noise. Care must be taken though in a real world case to ensure that there is no motion between the frames which are averaged over as this would produce a motion-blurred reference image.

V.   Discussion

In this work we have presented an investigation into the effect of various parameters on the accuracy of rigid registration of brain images for the purpose of data-driven head motion estimation.

While other groups have investigated how to optimally set up and execute a registration[13,14], we have focussed primarily on the frame duration and reconstruction and registration parameters which affect the noise properties of the images to establish how short the frame duration can be while ensuring an accurate registration.

We have shown that for registration based data-driven motion estimation in brain scans using either the FDG or FBB tracers, accurate motion estimates can be obtained by using 4×10^5 true and scattered coincidence events per frame, a TOF ray-tracing projector to perform pure MLEM reconstructions using 4 iterations, a transaxial image pixel size of 4 mm, a post-smoothing FWHM of 6 mm, and summing 2 frames together to produce the reference image. For most clinical scans this will result in frames of duration 0.5 – 1 sec, yielding sufficiently high temporal sampling of the motion to avoid intra-frame motion blurring.

The resulting estimated rigid motion parameters can then be used to perform a full event-by-event list-mode reconstruction, as described, for example, by Spangler-Bickell *et al.*[2], where each event is re-positioned according to the motion estimate at the corresponding time before being used in a reconstruction of the entire acquired data set.

V.A.   Limitations

This work focused on FDG and FBB and the results are largely specific to those tracers. Other tracers, particularly neuroreceptor tracers such as raclopride, may have a very different distribution in the brain which could affect the optimal frame count level for accurate registrations.

The conclusions from this investigation are primarily for static data where changes in the tracer distribution over time are negligible. For dynamic data there are other considerations which fall outside of the scope of this work. Usually for static data a single reference frame is used throughout the motion estimation, while for dynamic data it may be necessary to continuously update the reference image to account for the changing distribution. This may introduce drift into the motion estimates and cause an inadequate motion correction. Alternatively the mutual information metric could be utilized, perhaps using an anatomical scan from another modality (computed tomography (CT) or magnetic resonance (MR) imaging) as the reference image. The best strategy to handle this requires further research.

As mentioned earlier this analysis does not account for intra-frame motion in the ultra-short frames used for the registration. Our assumption has been that with such short frame durations (1 sec) intra-frame motion will be very limited and will likely not have a significant impact on the registration. In our experience with clinical data, if sufficiently rapid motion occurs which causes substantial intra-frame motion within a 1 sec frame (e.g. sneezing or coughing) that motion is short-lived and has negligible impact on the reconstruction. Nonetheless, the conclusions above represent an ideal case where there is essentially no intra-frame motion.

The approach described in this manuscript leverages the spatial domain for motion estimation. Other data-driven motion estimation techniques, e.g. using center-of-mass analysis, inertial tensors, principal component analysis (PCA), make use of the raw sonogram or list-mode data to avoid performing lengthy image reconstructions. But provided that the reconstructions can be easily generated (as has been demonstrated in[16]), using image registration is expected to provide more accurate and precise estimates.

We identified parameter settings which ensured that the 95th percentile of the accuracy of the registrations fell below a certain figure of merit. The small number of registrations which fall outside of this threshold should have only a very minor impact on the final motion corrected reconstruction.

VI.   Conclusion

In this work we have presented an investigation into the optimal parameter set to perform an accurate data-driven registration based motion estimation for brain scanning. For FDG and FBB tracers, with the SIGNA PET/MR or D710 PET/CT scanners, using 4×10^5 true and scattered coincidence events per frame will provide registrations which are accurate to at least 1 mm (95th percentile), defined as the error in the resultant displacement of the brain according to the registration parameters. In most clinical cases this will translate to a frame duration of 0.5–1 sec. Other reconstruction parameters were also optimized, such as

the number of MLEM iterations, pixel size, post-smoothing, and how to create the reference frame.

With these results one can reliably determine motion estimates directly from the data with a high temporal resolution, and then perform an event-by-event motion corrected reconstruction.

Acknowledgments

References

1. Carson R, Barker W, Liow J-S and Johnson CA, "Design of a motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction for the HRRT," 2003 IEEE Nuclear Science Symposium and Medical Imaging Conference, vol. 20892, pp. 3281–3285, 2003.

2. Spangler-Bickell MG, Khalighi MM, Hoo C et al. , "Rigid Motion Correction for Brain PET / MR Imaging Using Optical Tracking," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 4, pp. 498–503, 2019. [PubMed: 31396580]

3. Schleyer PJ, Dunn JT, Reeves S, Brownings S, Marsden PK and Thielemans K, "Detecting and estimating head motion in brain PET acquisitions using raw time-of-flight PET data.," Physics in Medicine and Biology, vol. 60, no. 16, pp. 6441–58, 2015. [PubMed: 26248198]

4. Lu Y, Naganawa M, Toyonaga T et al. , "Data-driven motion detection and event-by-event correction for brain PET: Comparison with Vicra," Journal of Nuclear Medicine, p. 10.2967/jnumed.119.235515, 2020.

5. Chen K, Smilovici Q, Lee W et al., "Inter-Frame Co-Registration of Dynamically Acquired Fluoro-Deoxyglucose Positron Emission Tomography Human Brain Data," in 2007 IEEE/ICME International Conference on Complex Medical Engineering, pp. 901–906, 2007.

6. Costes N, Dagher A, Larcher K, Evans AC, Collins DL and Reilhac A, "Motion correction of multi-frame PET data in neuroreceptor mapping: Simulation based validation," NeuroImage, vol. 47, pp. 1496–1505, 2009. [PubMed: 19481154]

7. Picard Y and Thompson CJ, "Motion correction of PET images using multiple acquisition frames.," IEEE Transactions on Medical Imaging, vol. 16, no. 2, pp. 137–144, 1997. [PubMed: 9101323]

8. Boellaard R, Van Lingen A and Lammertsma AA, "Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: Quantitative characteristics and effects on kinetic modeling," Journal of Nuclear Medicine, vol. 42, no. 5, pp. 808–817, 2001. [PubMed: 11337581]

9. Feng T, Yang D, Zhu W, Dong Y and Li H, "Real - Time Data - Driven Rigid Motion Detection and Correction for Brain Scan with Listmode PET," in 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 1–4, 2016.

10. Rezaei A, Spangler-Bickell M, Shramm G, Van Laere K, Nuyts J, "Rigid Motion Tracking using Moments of Inertia in TOF-PET Brain Studies," in 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp. 1–2, 2020.

11. Jin X, Mulnix T, Gallezot J-D and Carson RE, "Evaluation of motion correction methods in human brain PET imaging–a simulation study based on human motion data.," Medical Physics, vol. 40, no. 10, p. 102503, 2013. [PubMed: 24089924]

12. Mukherjee JM, Lindsay C, Mukherjee A et al. , "Improved frame-based estimation of head motion in PET brain imaging," Medical Physics, vol. 43, no. 5, pp. 2443–2454, 2016. [PubMed: 27147355]

13. Andersson JL, "How to obtain high-accuracy image registration: Application to movement correction of dynamic positron emission tomography data," European Journal of Nuclear Medicine, vol. 25, no. 6, pp. 575–586, 1998. [PubMed: 9618571]

14. Jenkinson M and Smith S, "A global optimisation method for robust affine registration of brain images," Medical Image Analysis, vol. 5, pp. 143–156, 2001. [PubMed: 11516708]

15. Maes F, Collignon A, Vandermeulen D, Marchal G and Suetens P, "Multi-modality image registration by maximization of mutual information," Proceedings of the Workship on Mathematical Methods in Biomedical Image Analysis, no. December 2013, pp. 14–22, 1996.

16. Spangler-Bickell MG, Deller TW, Bettinardi V and Jansen F, "Ultra-Fast List-Mode Reconstruction of Short PET Frames and Example Applications," Journal of Nuclear Medicine, p. 10.2967/jnumed.120.245597, 2020.

17. Spangler-Bickell MG, Deller T, Hurley SA, McMillan AB, Bettinardi V and Jansen F, "Effect of Image Noise on Registration in PET Brain Imaging," 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC, pp. 1–3, 2019.

18. Grant AM, Deller TW, Khalighi MM, Maramraju SH, Delso G and Levin CS, "NEMA NU 2–2012 performance studies for the SiPM-based ToF-PET component of the GE SIGNA PET/MR system," Medical Physics, vol. 43, pp. 2334–2343, 2016. [PubMed: 27147345]

19. Levin CS, Maramraju SH, M Khalighi M et al. , "Design Features and Mutual Compatibility Studies of the Time-of-Flight PET Capable GE SIGNA PET/MR System," IEEE Transactions on Medical Imaging, vol. 35, no. 8, pp. 1907–1914, 2016. [PubMed: 26978664]

20. Bettinardi V, Presotto L, Rapisarda E et al. , "Physical Performance of the new hybrid PETCT Discovery-690," Medical Physics, vol. 35, no. 38, pp. 5394–5411, 2011.

21. Aksoy M, Forman C, Straka M et al. , "Real-time optical motion correction for diffusion tensor imaging," Magnetic Resonance in Medicine, vol. 66, no. 2, pp. 366–378, 2011. [PubMed: 21432898]

22. Forman C, Aksoy M, Hornegger J and Bammer R, "Self-encoded marker for optical prospective head motion correction in MRI," Medical Image Analysis, vol. 15, no. 5, pp. 708–719, 2011. [PubMed: 21708477]

23. Strother SC, Casey ME and Hoffman EJ, "Measuring PET Scanner Sensitivity: Relating Countrates to Image Signal-to-Noise Ratios using Noise Equivalent Counts," IEEE Transactions on Nuclear Science, vol. 37, no. 2, pp. 783–788, 1990.

24. Lopresti B, Russo A, Jones W, et al. , "Implementation and performance of an optical motion tracking system for high resolution brain PET imaging," IEEE Transactions on Nuclear Science, vol. 46, no. 6, pp. 2059–2067, 1999.

25. Fulton R, Meikle S, and Eberl S, "Correction for head movements in positron emission tomography using an optical motion-tracking system," IEEE Transactions on Nuclear Science, vol. 49, no. 1, pp. 116–123, 2002.

26. MATLAB, 9.8.0.1323502 (R2020a). Natick, Massachusetts: The MathWorks Inc., 2020.

27. Pluim JP, Maintz JB and Viergever MA, "Mutual-information-based registration of medical images: A survey," IEEE Transactions on Medical Imaging, vol. 22, no. 8, pp. 986–1004, 2003. [PubMed: 12906253]
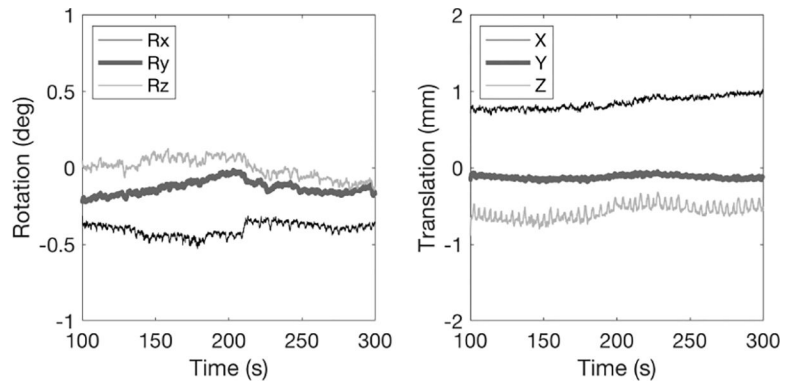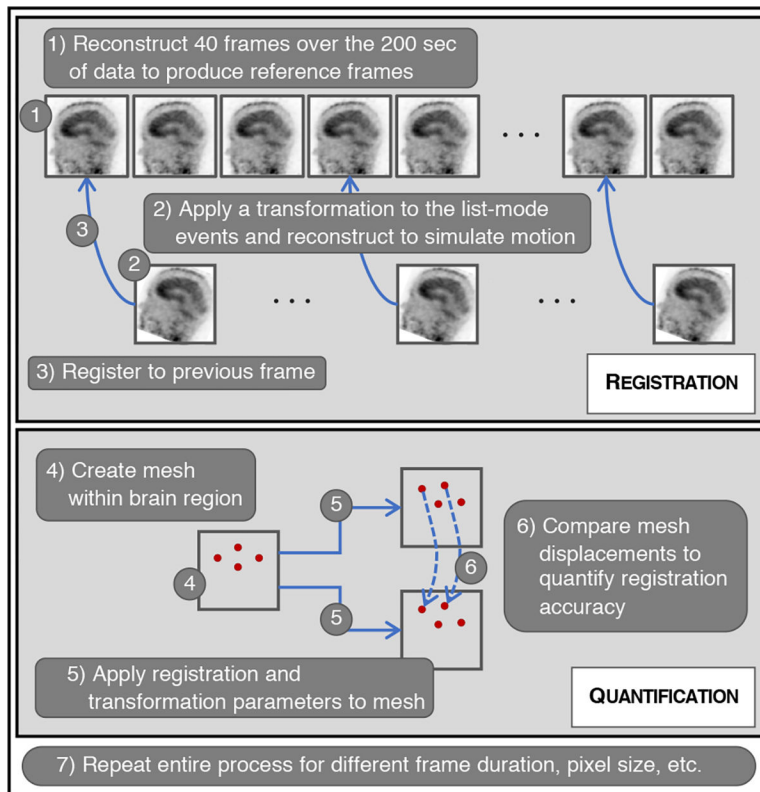
Figure 1:
Motion plots as recorded by the HobbitView external tracker for Data Set A. Relative to the first position the rotations change by less than 0.2° and the translations by less than 0.4 mm. The small high frequency oscillations which can be seen are due to respiration.
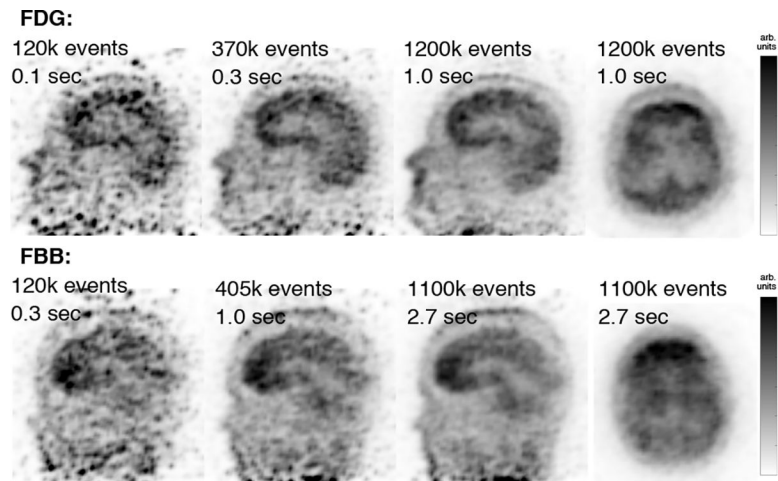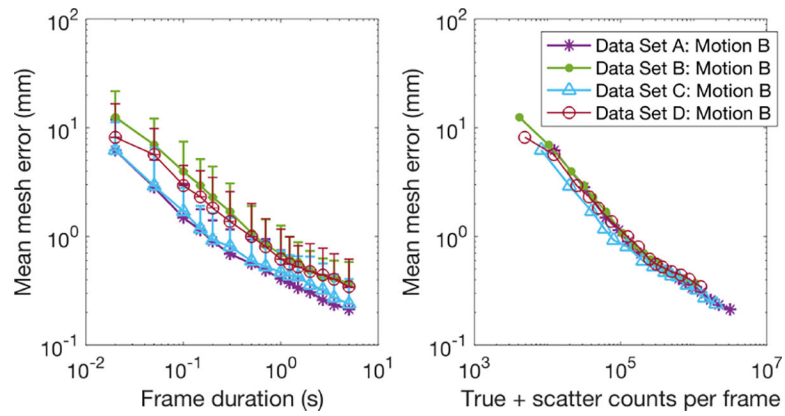
Figure 2:
Examples of reconstructions for Data Set A (FDG, top row) and Data Set D (FBB, bottom row) for three different frame durations and number of counts to demonstrate the image noise level. A transaxial slice is shown for the last reconstruction. These have been smoothed with an isotropic Gaussian kernel with 6 mm FWHM. No attenuation or scatter correction was applied.
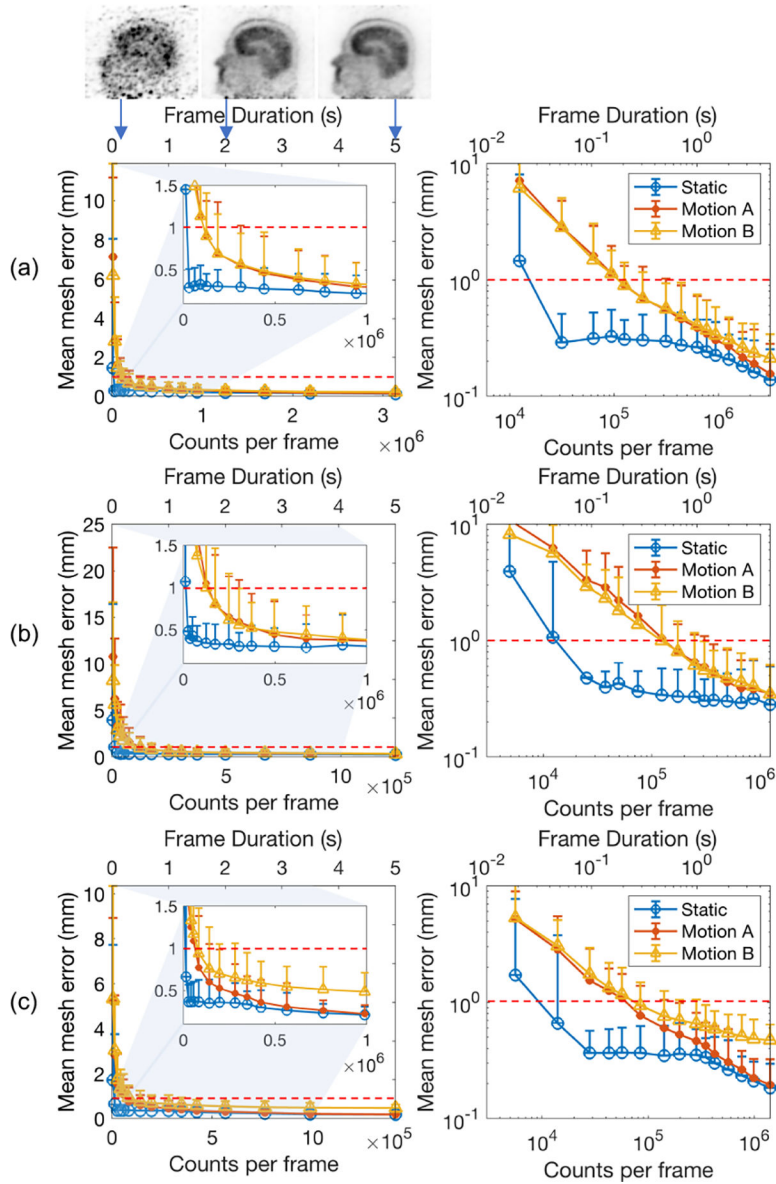
Figure 3:
Diagram illustrating the motion simulation, registration, and quantification process.
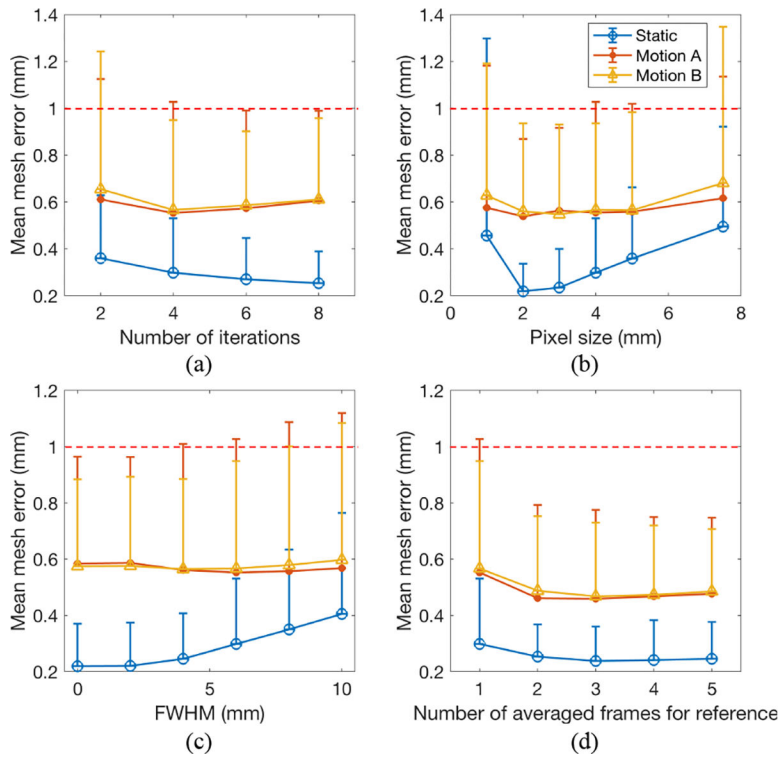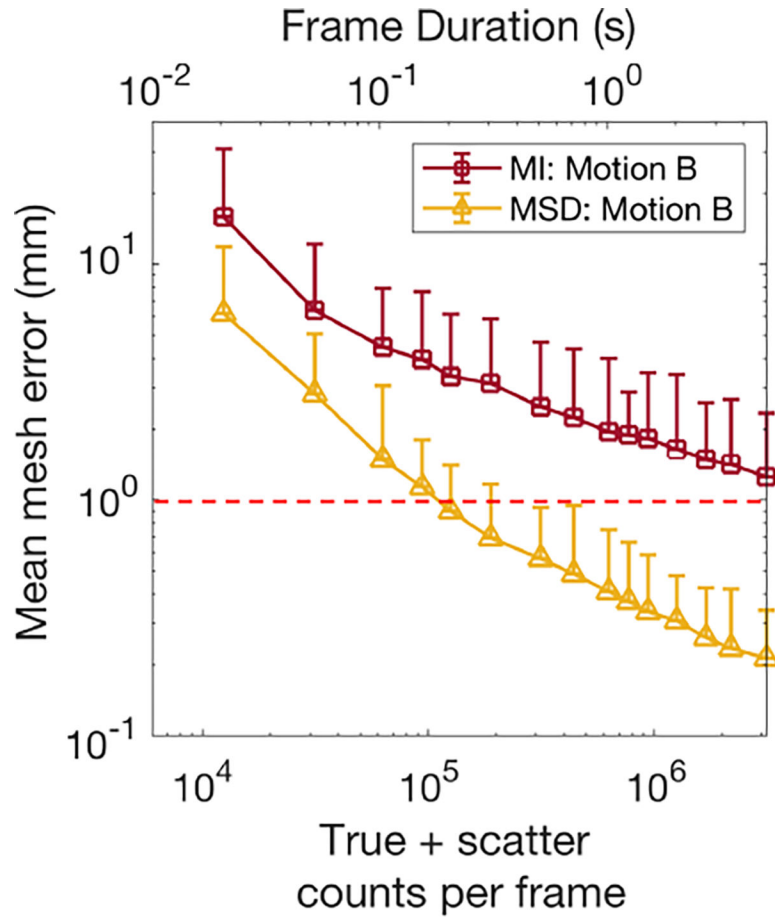
Figure 4:

Left: The log-log scale plots of the registration accuracy against the frame duration for Data Sets A - D. See Table 1 for count rates and randoms fractions. The one-sided error bars indicate the value of the 95th percentile. Right: When plotting the same data against the number of true and scattered counts per frame the plots all lie very close together, indicating that the number of true and scattered counts per frame is a good indicator of the expected registration accuracy even for very different activity levels. The errors bars have been removed in this plot for clarity.

Figure 5:
Left: The mean accuracy of the registration of the 40 frames for (a) Data Set A, (b) Data Set D, and (c) Data Set E, plotted against both the number of true and scatter counts per frame (bottom *x*-axis) and the frame duration (top *x*-axis), with a zoomed in inset, showing that the registration accuracy improves with increasing counts per frame.The one-sided error bars indicate the value of the 95th percentile. The dashed line is for reference and indicates where the chosen figure of merit of 1 mm lies. Above figure (a) are shown some representative reconstructed frames for Data Set A with arrows indicating their frame duration. Right: The same plots on a log-log scale.

Figure 6:
Results of the investigations into the effect of various parameters on the registration accuracy using Data Set A and 0.5 sec frames. The one-sided error bars indicate the value of the 95th percentile. The dashed line indicates where the chosen figure of merit of 1 mm lies.

Figure 7:
Comparing the influence of the registration metric on the accuracy with respect to the number of counts per frame for Data Set A on a log-log scale. Only one motion case is shown. The mutual information (MI) metric results in less accurate registrations than the mean square difference (MSD) metric. The one-sided error bars indicate the value of the 95th percentile. The dashed line is for reference and indicates where the chosen figure of merit of 1 mm lies.

Table 1:

Summary of Data Sets

	Set A	Set B[*]	Sec C	Set D	Set E
Tracer	FDG	FDG	FDG	FBB	FDG
Scanner	SIGNA	SIGNA	SIGNA	SIGNA	D710
Injected Activity (MBq)	504		183	245	412
Uptake Time (min)	27		70	43	49
Prompts per sec ($\times 10^3$)	1215	405	541	422	575
True & Scatter per sec ($\times 10^3$)	636	212	402	258	402
Randoms per sec ($\times 10^3$)	579	193	139	164	173
Randoms fraction	0.477	0.477	0.257	0.389	0.302

[*] This data set was created by subsampling Set A by a factor of 3.

Table 2:

Investigated Parameters

Parameter	Range
Frame Duration	[0.02, …, **0.5**, …, 5.0] sec
MLEM Iterations	[2, **4**, 6, 8]
Transaxial Pixel Size	[1.0, 2.0, 3.0, **4.0**, 5.0, 7.5] mm
Post-smoothing FWHM	[0, 2, 4, **6**, 8, 10] mm
Number of averaged frames	[**1**, 2, 3, 4, 5]
Metric	[**Mean Square Diff.**, Mutual Info.]

When one parameter was being tested all of the other parameters were held at the bold-faced values shown here.