Joint Reconstruction of Image and Motion in Gated Positron Emission Tomography

Moritz Blume*, Axel Martinez-Möller, Andreas Keil, Nassir Navab, *Member, IEEE*, and Magdalena Rafecas, *Member, IEEE*

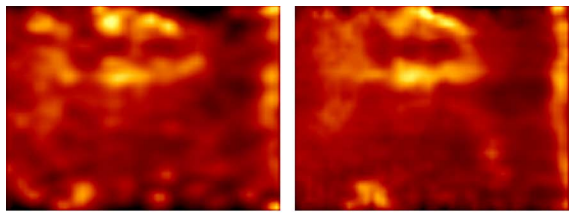
***Abstract*—We present a novel intrinsic method for joint reconstruction of both image and motion in positron emission tomography (PET). Intrinsic motion compensation methods exclusively work on the measured data, without any external motion measurements. Most of these methods separate image from motion estimation: They use deformable image registration/optical flow techniques in order to estimate the motion from individually reconstructed gates. Then, the image is estimated based on this motion information. With these methods, a main problem lies in the motion estimation step, which is based on the noisy gated frames. The more noise is present, the more inaccurate the image registration becomes. As we show both visually and quantitatively, joint reconstruction using a simple deformation field motion model can compete with state-of-the-art image registration methods which use robust multilevel *B*-spline motion models.**

***Index Terms*—Gating, motion compensation, positron emission tomography (PET), reconstruction.**

I. Introduction

With a physical resolution of less than 5 mm of modern positron emission tomography (PET) scanners, even small patient motion can significantly reduce the image quality and thus lead to image distortion or even false diagnosis. For example, in the case of cardiac studies, the reconstructed image is affected by both cardiac and respiratory motion. Coronary arteries have been shown to move 8–23 mm during the cardiac cycle [1] and heart movement during free breathing is reported to be around 4.9 and 9 mm [2]–[4].

In cardiac PET a typical way to deal with motion is to use both cardiac and respiratory gating (see e.g., [5]): each cardiac and respiratory cycle is subdivided into a predefined number of frames (usually four to eight). Corresponding frames of different cycles define a gate. Due to the periodicity of the movements, the heart is supposed to be situated in the same position for the same cardiac and respiratory phase. Accordingly, each gate is assumed to contain motion free data.



(a) (b)

Fig. 1. Sagittal views of a human's heart for a 2 min ^{18}FDG PET scan. (a) Reconstruction of one gate with a standard ML-EM algorithm. The image intensities are scaled by a factor of 24 in order to be comparable. (b) Reconstruction of the same gate with our proposed joint reconstruction method. The image is smoother and provides more details since all counts are used for reconstruction.

A severe disadvantage of reconstructing individual gates is the reduced statistical information. If, for example, 24 gates are used, the number of available counts for each gate is approximately only one divided by 24 of the total number of available counts. Accordingly, the images corresponding to the gates suffer from much more noise than the nongated reconstruction which is reconstructed based on the whole data. Fig. 1(a) shows one reconstructed gate.

Most current motion compensation methods consist of two independent steps: 1) motion estimation and 2) image estimation.

A popular way to estimate respiratory and/or cardiac motion is to 1) reconstruct each gate by a state-of-the-art PET reconstruction method (such as maximum-likelihood expectation-maximization, ML-EM, [6]), and 2) register each frame to a reference frame. For gated PET, this has been done with affine motion models [7], deformable optical flow models [8]–[11], or based on a previously established patient specific motion model [12]. Alternatively, gated CT images which correspond to the PET gates can be used for motion estimation [13], [14].

In the second step, the image is estimated based on the previously found motion. A very common approach is to deform reconstructed gates to the reference frame and then combine them (usually by taking the sum of these deformed frames) [7]–[10]. Instead, an adapted ML-EM algorithm which is able to deal with arbitrary nonrigid motion can be used [12]–[14].

There are some problems related to the motion estimation step. If it relies on individually reconstructed gates, the statistical properties of the gates are crucial. On the one hand, one

Manuscript received February 19, 2010; revised May 31, 2010; accepted June 03, 2010. Date of publication June 17, 2010; date of current version November 03, 2010. This work was supported in part by the Spanish Ministry of Science and Innovation under Grant TEC2007-61047 and CI09-252, in part by the CAMP Scholarship provided by the chair for Computer Aided Medical Procedures, Computer Science Department, Technische Universität München, and in part by the Deutscher Akademischer Austauschdienst (DAAD) under Grant D/09/50710. *Asterisk indicates corresponding author.*

*M. Blume is with the Instituto de Física Corpuscular (IFIC), Universidad de Valencia/CSIC, E-46071 Valencia, Spain, and also with Computer Aided Medical Procedures (CAMP), Department of Computer Science, Technische Universität München, D-85748 Munich, Germany (e-mail: moritz.blume@cs.tum.edu).

A. Martinez-Möller is withComputer Aided Medical Procedures (CAMP) and with the Nuklearmedizinische Klinik im Klinikum Rechts der Isar, Technische Universität München, D-81675 Munich, Germany.

A. Keil and N. Navab are with the Computer Aided Medical Procedures (CAMP), Technische Universität München, Munich, Germany.

M. Rafecas is with the Instituto de Física Corpuscular (IFIC), Universidad de Valencia/CSIC, Valencia, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2010.2053212

$\boldsymbol{f}^{(1)} \leftarrow \mathbf{1}$
$\boldsymbol{\varphi}^{(1)} \leftarrow \mathrm{id}$
for $k \leftarrow 1 \ldots$ #outer iterations **do**
  $\boldsymbol{f} \leftarrow \boldsymbol{f}^{(k)}$
  $\boldsymbol{\varphi} \leftarrow \boldsymbol{\varphi}^{(k)}$
  $\boldsymbol{f}^{(k+1)} = \boldsymbol{f} \frac{\sum_t j_t \boldsymbol{T}_{\boldsymbol{\varphi}_t^{-1}} \boldsymbol{H}^\top \frac{\boldsymbol{g}_t}{\bar{\boldsymbol{g}}_t}}{\sum_t j_t \boldsymbol{T}_{\boldsymbol{\varphi}_t^{-1}} \boldsymbol{H}^\top \mathbf{1}}$
  $\boldsymbol{f} \leftarrow \mathrm{SMOOTH}(\boldsymbol{f})$
  for $j \leftarrow 1 \ldots$ #inner iterations **do**
    for $i \leftarrow 1 \ldots 3, t \leftarrow 1 \ldots T$ **do**
      $\Delta_{it} \leftarrow \frac{1}{T} \boldsymbol{T}_{\boldsymbol{\varphi}_t} \nabla_i \boldsymbol{f} \boldsymbol{H}^\top \left(\mathbf{1} - \frac{\boldsymbol{g}_t}{\bar{\boldsymbol{g}}_t}\right)$
      $\Delta_{it} \leftarrow \Delta_{it} + \alpha \sum_j \nabla_j^2 \boldsymbol{\varphi}_{it}$
      $\boldsymbol{\varphi}_{it}^{\mathrm{new}} = \boldsymbol{\varphi}_{it} + \delta \Delta_{it}$
    end for
    if $J(\boldsymbol{f}^{(k+1)}, \boldsymbol{\varphi}^{\mathrm{new}}) < J(\boldsymbol{f}^{(k+1)}, \boldsymbol{\varphi})$ **then**
      $\delta \leftarrow 1.2\delta$
      $\boldsymbol{\varphi} \leftarrow \boldsymbol{\varphi}^{\mathrm{new}}$
    else
      $\delta \leftarrow \frac{1}{2}\delta$
    end if
  end for
  $\boldsymbol{\varphi}^{(k+1)} \leftarrow \boldsymbol{\varphi}$
end for

Fig. 2. Our proposed joint reconstruction algorithm.

wants to have as many gates as possible in order to accurately measure the motion. On the other hand, using more gates implies less photon counts in each gate and thus leads to noisier reconstructed images on which the motion estimation is based. So, with an increasing number of gates, the motion estimation step is less accurate or even fails completely. These statistical problems can be circumvented by using gated CT scans for motion estimation instead of individually reconstructed gates. However, this comes at the price of exposing the patient to several CT scans (depending on the number of gates) and a potential misalignment between the PET and CT gates.

Beyond the described methods which separate motion estimation from image estimation, there is another class of algorithms that jointly estimate image and motion. These methods work only on the acquired PET data and are based on a unified imaging model which explicitly includes the physical effects induced by motion. They can be broadly subclassified into two categories: those that use a different image estimate for each frame [15]–[20], and those that use a common image estimate for all frames [21]–[24].

Those that use a different image estimate for each frame model the image as a function of time. The motion function is not explicitly part of this model. The result of such methods is a 4-D (3-D $+t$) reconstruction of the tracer distribution.

The methods that use a common image estimate for each frame are also able to reconstruct a 4-D tracer distribution, but, additionally deliver the corresponding motion function which could be useful for further studies. All of the cited papers on a common image estimate work with a rigid motion model.

In this work, we present a novel method for joint reconstruction (JR) of image and motion. Our method is based on a motion-aware likelihood function and a smoothing term which regularizes the motion field in the space domain. The motion field is a 4-D deformation vector field which contains the deformation vector of each voxel at a specific point of time. We use a common image estimate and—in contrast to most current methods—a deformable motion model. Our objective function is very similar to the one used in [21], however, our optimization scheme is different. In contrast to [21], we observe a clear advantage for joint reconstruction methods with respect to reconstruction quality.

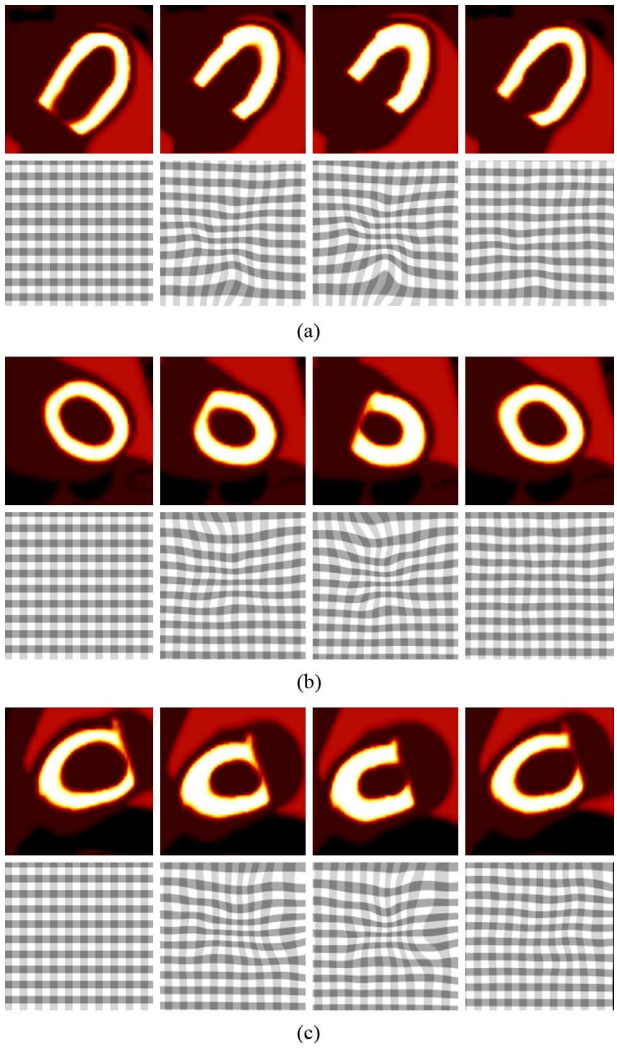


Fig. 3. Generated original frames from the XCAT phantom. Both respiratory motion (1 cm diaphragm movement) and cardiac motion are simulated. These frames are used for simulating measurement data for different statistical scenarios. The motion fields are visualized by a deformed grid (the first frame is the reference frame) (a) Transverse slices. (b) Coronal slices. (c) Sagittal slices.

Since our method delivers both image and motion, it is possible to visualize the motion instances for every gate. Compared to a simple gated approach without motion compensation, our method delivers a smoother image, since for every gate all available counts of the whole acquisition are used [see Fig. 1(b)].

II. Joint Reconstruction

In the following, we design a cost functional which depends on both an image and a motion function and is minimal for a good fit of image and motion to the data. As a measure of fit, we use an adapted motion-aware likelihood function. Additionally, a regularization term which encourages spatially smooth motion is used.

Then, we derive an iterative algorithm for minimizing this cost functional. The algorithm basically consists of two update

TABLE I
REGISTRATION PARAMETERS USEDBY *DROP* FOR REGISTRATION AND FUSION OF RECONSTRUCTED FRAMES FOR ONE GRID LEVEL (RFRF-1G) AND THREE/FIVE GRID LEVELS (RFRF-3G/5G) FOR BOTH SYNTHETIC AND PATIENT DATA

	RFRF-1G Synthetic	RFRF-3G Synthetic	RFRF-1G Patient	RFRF-5G Patient
	1	1	1	1
Grid Levels	1	3	1	5
Starting Grid Size	49x49x49	5x5x5	65x65x19	5x5x3
Final Grid Size	49x49x49	17x17x17	65x65x19	65x65x19
Min. Dimension	0	0	0	0
Interpolation	Linear	Cubic B-Splines	Linear	Cubic B-Splines
Image Margin	0	0	0	0
Update Mode	Consecutive	Consecutive	Consecutive	Consecutive
Sampling	Sparse	Sparse	Sparse	Sparse
Link to Maximum Grid	yes	yes	yes	yes
Steps	5	5	5	5
Label Factor	1	0,67	1	0,67
Optimizer	FastPD	FastPD	FastPD	FastPD
Data Cost	SAD	SAD	SAD	SAD
Distance	Tr. Quad. Diff.	Tr. Quad. Diff.	Tr. Quad. Diff.	Tr. Quad. Diff.
Projection	Linear	Linear	Linear	Linear
Incremental Regularization	no	no	no	no
Iterations	100	10	200	20
Gamma	0	0	0	0
Lambda	10	10	10	10
Truncation	0	0	0	0
Histogram Bins	32	32	32	32

equations (one for the image and one for the motion function) which are repeated.

A. Cost Functional

1) Maximum-Likelihood Estimate: The number of counts $g(a)$ that is measured for a line-of-response (LOR) a is Poisson distributed

$$P(g(a)|f) = e^{-\hat{g}(a)} \cdot \frac{\hat{g}(a)^{g(a)}}{g(a)!}. \tag{1}$$

$\hat{g}(a) = 1/T \int H(a,\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$ is the expected number of counts given the tracer distribution f (which is subject to reconstruction) and the system response function H.

In the case of a moving anatomy, f can be modeled as $f(\boldsymbol{\varphi}(\boldsymbol{x},t))$, where $\boldsymbol{\varphi}$ is a deformation function $\boldsymbol{\varphi} : \mathbb{R}^4 \mapsto \mathbb{R}^3$ which maps a position $\boldsymbol{x}$ in space at time t to a new position $\boldsymbol{\varphi}(\boldsymbol{x},t)$ in the reference frame (throughout the paper, vector valued functions will be printed in bold letters). Accordingly, f corresponds to the (virtual) reference/reconstruction frame. The probability distribution has to be modified to

$$P(g(a,t)|f,\boldsymbol{\varphi}) = e^{-\hat{g}(a,t)} \cdot \frac{\hat{g}(a,t)^{g(a,t)}}{g(a,t)!} \tag{2}$$

where

$$\hat{g}(a,t) = \frac{1}{T} \int H(a,\boldsymbol{x}) f(\boldsymbol{\varphi}(\boldsymbol{x},t))\, d\boldsymbol{x} \tag{3}$$

is the expected number of counts in LOR a at time t given an image f and transformation $\boldsymbol{\varphi}$.

Note that t does not necessarily refer to a specific point of time. In our case, we think of t as a time frame in which no movement takes place.

The likelihood function for all measured events is

$$L(f,\boldsymbol{\varphi}) = \prod_{a,t} P(g(a,t)|f,\boldsymbol{\varphi}). \tag{4}$$

(a) (b) (c) (d)

Fig. 4. Different original frames registered to the first original frame. (a) Original frame one. (b) Original frame two registered to one. (c) Original frame three registered to frame one. (d) The original frame four registered to frame one. Three grid levels (starting with $5 \times 5 \times 5$ and ending with $17 \times 17 \times 17$) and 10 iterations were used. There is hardly any visible difference, which indicates that $17 \times 17 \times 17$ control points are sufficient in order to approximate the original motion.

We seek to find a pair of image f and motion $\boldsymbol{\varphi}$ that maximizes the likelihood function. This is equivalent to minimizing the negative log-likelihood function

$$\begin{aligned} &-\log(L(f,\boldsymbol{\varphi})) \\ &\quad = \sum_{a,t} \hat{g}(a,t) - g(a,t)\log(\hat{g}(a,t)) + \log(g(a,t)!). \end{aligned} \tag{5}$$

Here, $\sum_{t,a} \log(g(a,t)!)$ can be omitted since it does not affect the minimum. So, finally we arrive at

$$\mathcal{D}(f,\boldsymbol{\varphi}) := \sum_{a,t} \hat{g}(a,t) - g(a,t)\log(\hat{g}(a,t)) \tag{6}$$

which is subject to minimization.

2) Regularization: Both image and motion suffer from the highly noisy data and have to be regularized. In case of the image, we apply moderate Gaussian smoothing [3 mm full-width at half-maximum (FWHM)] after each image iteration.

In order to prevent extreme deformations we use homogeneous diffusion regularization [25]

$$\mathcal{S}(\boldsymbol{\varphi}) = \sum_t \sum_{i=1}^{3} \int \|\nabla_{\boldsymbol{x}} [\boldsymbol{\varphi}]_i(\boldsymbol{x},t)\|^2\, d\boldsymbol{x}. \tag{7}$$

$\|\boldsymbol{a}\|$ is the euclidean norm of vector $\boldsymbol{a}$ ($\|\boldsymbol{a}\| = \sqrt{\sum_i [\boldsymbol{a}]_i^2}$). We also implemented smoothing in the time domain; however, no differences in the reconstruction could be noted, so we will not discuss it further in the paper.

Finally, the complete cost functional we seek to minimize is

$$\mathcal{J}(f, \boldsymbol{\varphi}) = \mathcal{D}(f, \boldsymbol{\varphi}) + \alpha\, \mathcal{S}(\boldsymbol{\varphi}). \tag{8}$$

The regularization parameter α defines the smoothness of our sought deformation. It has to be carefully adjusted to the specific case. If it is too high, the resulting $\boldsymbol{\varphi}$ will represent no visible deformation and thus the resulting image f will still suffer from motion blur. Is it too low, $\mathcal{J}$ will be over-fitted and this will result in an unrealistic image and motion pair. Sections III-D and IV-D contain more information on the right choice of α.

B. Optimization

Our goal is to design an iterative scheme for f and $\boldsymbol{\varphi}$ that converges to values $f^\star$ and $\boldsymbol{\varphi}^\star$ which minimize $\mathcal{J}$. Since by definition $\mathcal{J}$ is minimal at $f^\star$ and $\boldsymbol{\varphi}^\star$, the variational derivative of $\mathcal{J}$ at $f^\star$ and $\boldsymbol{\varphi}^\star$ is zero for any arbitrary variation of $f^\star$ and $\boldsymbol{\varphi}^\star$. We use this necessary condition in order to derive two update equations

$$f^{(k+1)} = \tau_f(f^{(k)}, \boldsymbol{\varphi}^{(l)}) \tag{9}$$
$$\boldsymbol{\varphi}^{(l+1)} = \tau_{\boldsymbol{\varphi}}(f^{(k)}, \boldsymbol{\varphi}^{(l)}) \tag{10}$$

where (9) defines an image update and (10) defines a motion update.

1) Image Update: We start by setting the variational derivative with respect to f to zero

$$\delta_f \mathcal{J}(f, \boldsymbol{\varphi}) = 0. \tag{11}$$

The variational derivative is defined as

$$\delta_f \mathcal{J}(f, \boldsymbol{\varphi}) := \frac{\partial}{\partial \varepsilon} \mathcal{J}(f + \varepsilon\eta, \boldsymbol{\varphi})\bigg|_{\varepsilon=0} \tag{12}$$

(for details about calculus of variations, see e.g., [26]).

Since $\mathcal{S}$ does not depend on f, we get

$$\delta_f \mathcal{J}(f, \boldsymbol{\varphi}) \tag{13}$$
$$= \frac{\partial}{\partial \varepsilon} \mathcal{D}(f + \varepsilon\eta, \boldsymbol{\varphi})\bigg|_{\varepsilon=0} \tag{14}$$
$$= \sum_{t,a} \left[\delta_f \hat{g}(a,t,f) - g(a,t)\, \frac{\partial}{\partial \varepsilon} \log\left(\hat{g}(a,t,f+\varepsilon\eta)\right)\bigg|_{\varepsilon=0} \right] \tag{15}$$
$$= \sum_{t,a} \left[\delta_f \hat{g}(a,t,f) - g(a,t) \frac{\delta_f \hat{g}(a,t,f)}{\hat{g}(a,t,f)} \right] \tag{16}$$
$$= \sum_{t,a} \delta_f \hat{g}(a,t,f) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right). \tag{17}$$

From Appendix A we know that

$$\delta_f \hat{g}(a,t,f) = \int \nabla_f \hat{g}(a,t,f)\, \eta(\boldsymbol{x})\, d\boldsymbol{x} \tag{18}$$

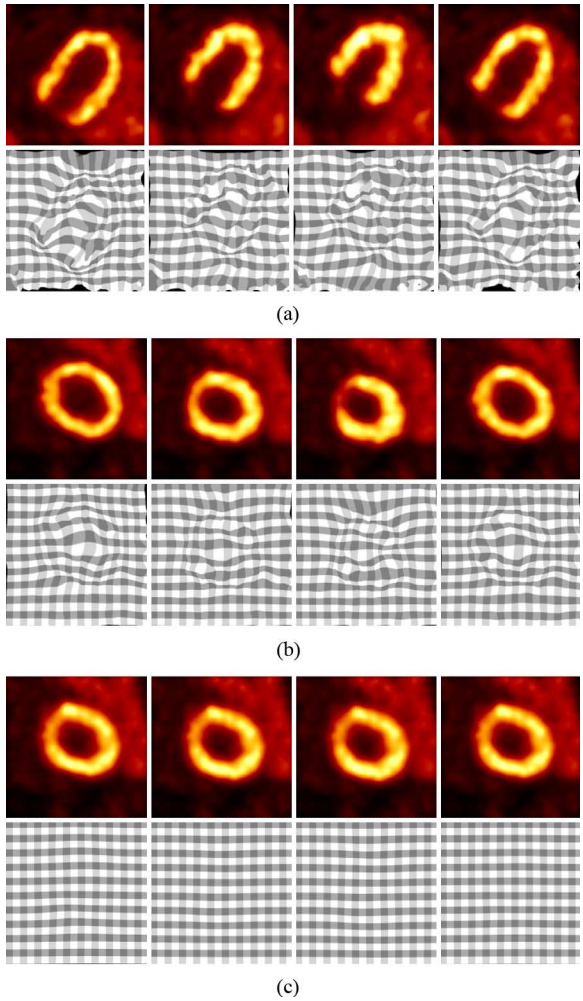


Fig. 5. Comparison of our JR method for different values of α for the simulation study. For a too low α, the reconstructed motion field contains extreme deformations which result in artifacts in the reconstructed image. For a too high α, deformations are almost completely suppressed which results in a motion blurred image, similar to as if no motion compensation had taken place. (a) $\alpha = 0.1$ (too low). (b) $\alpha = 2$ (good choice). (c) $\alpha = 100$ (too high).

with

$$\nabla_f \hat{g}(a,t,f) := \frac{1}{T} H(a, \varphi^{-1}(\boldsymbol{x},t))\, |\det(J_{\varphi^{-1}}(\boldsymbol{x},t))|. \tag{19}$$

Together with (17) we arrive at

$$\int \sum_{t,a} \nabla_f \hat{g}(a,t,f) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right) \eta(\boldsymbol{x})\, d\boldsymbol{x} = 0 \tag{20}$$

which is set to zero for all variations η. Applying the fundamental lemma of calculus of variations leads to

$$\sum_{t,a} \nabla_f \hat{g}(a,t,f) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right) = 0 \tag{21}$$
$$\Leftrightarrow \sum_{t,a} \nabla_f \hat{g}(a,t,f) - \sum_{t,a} \nabla_f \hat{g}(a,t,f) \frac{g(a,t)}{\hat{g}(a,t,f)} = 0 \tag{22}$$
$$\Leftrightarrow 1 - \frac{\sum_{t,a} \nabla_f \hat{g}(a,t,f) \frac{g(a,t)}{\hat{g}(a,t,f)}}{\sum_{t,a} \nabla_f \hat{g}(a,t,f)} = 0. \tag{23}$$

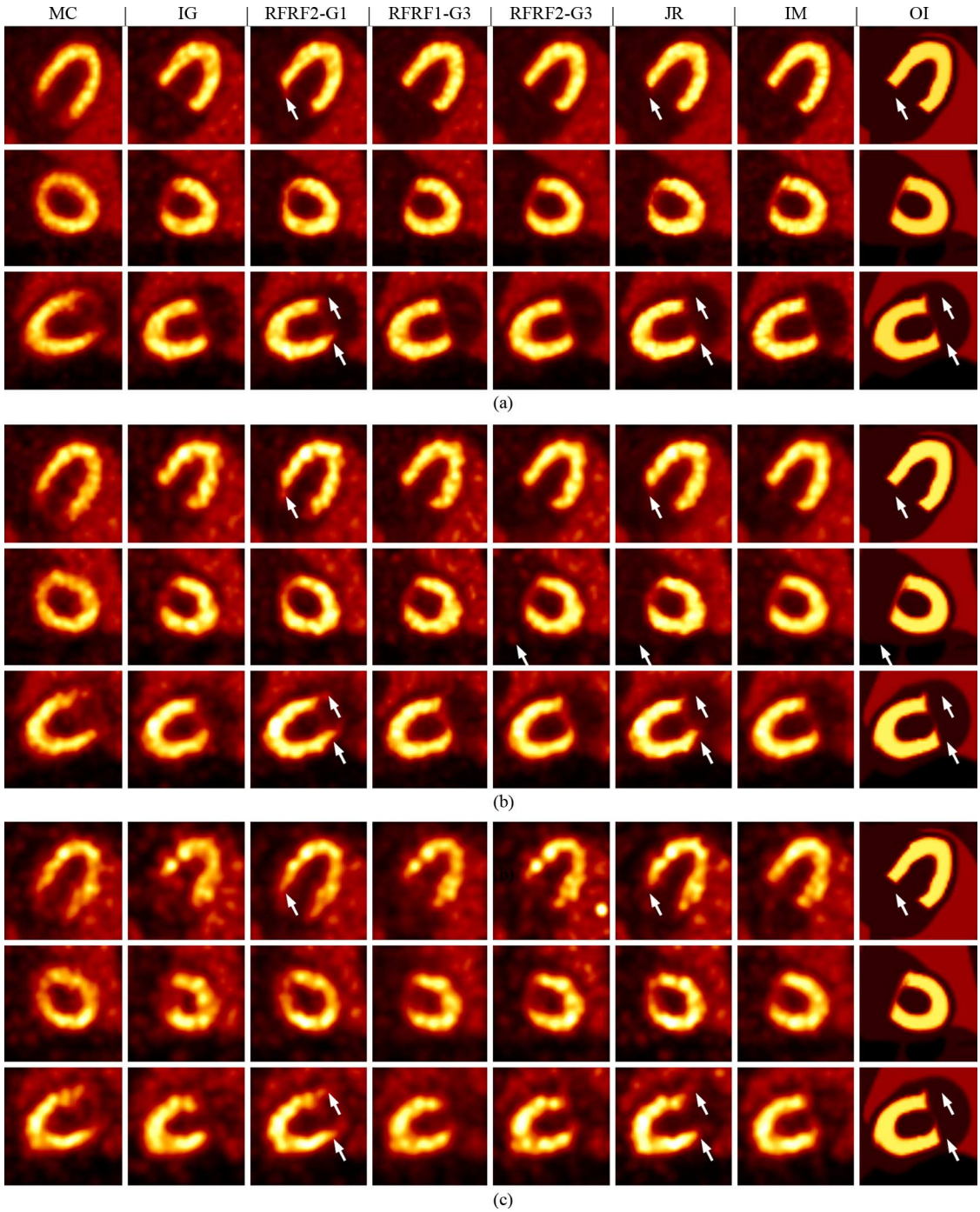


Fig. 6. Selected transverse, coronal and sagittal slices for different levels of noise and different reconstruction scenarios for the simulated data (from left to right): ML-EM reconstruction of motion-contaminated data (MC), ML-EM reconstruction of the individual gates (IG), RFRF2-1G (RFRF2 with one grid level, as described in Section III-C), RFRF1-3G, RFRF2-3G, our joint reconstruction (JR) and a motion compensating reconstruction based on the ideal motion (IM). For comparison, the original image (OI) is shown in the last column. The reference gate is the third gate. Since IG contains less statistics, it was scaled in order to be visibly comparable to the rest of the methods. (a) High number of counts: 1048576 annihilation events. (b) Moderate number of counts: 262144 annihilation events. (c) Low number of counts: 65536 annihilation events.

Multiplication of (23) with $f(\boldsymbol{x})$ finally yields the image update function

$$\tau_f(f^{(k)}, \boldsymbol{\varphi}^{(k)}) = f^{(k)}(\boldsymbol{x}) \frac{\sum_{t,a} \nabla_f \hat{g}(a,t,f) \frac{g(a,t)}{\hat{g}(a,t,f)}}{\sum_{t,a} \nabla_f \hat{g}(a,t,f)}. \qquad (24)$$

2) Motion Update: Analog to the image update, we start by setting the variational derivative with respect to the ith component ($i \in \{1,2,3\}$) of $\boldsymbol{\varphi}$ to zero

$$\delta_{[\boldsymbol{\varphi}]_i} \mathcal{J}(f, \boldsymbol{\varphi}) = 0. \qquad (25)$$

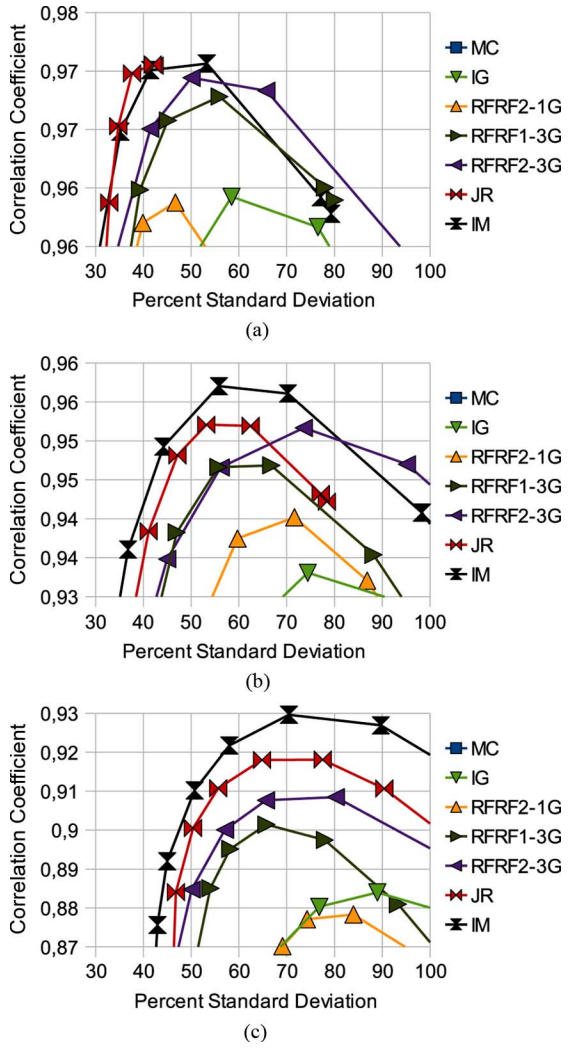


Fig. 7. Comparison of reconstructed images: background noise (standard deviation in relation to mean in percent) plotted versus mean (of all gates) correlation coefficients of the respective reconstruction methods and the original gates. Methods compared: ML-EM reconstruction of the motion-contaminated data (MC), an ML-EM reconstruction of the individual gates (IG), the registration and fusion of reconstructed frames (RFRF2-1G, RFRF1-3G, and RFRF2-3G) methods as described in the methods section, our joint reconstruction (JR) and a motion compensating reconstruction based on the ideal motion (IM). Note that the scale is different for each counts level. The standard deviation was evaluated for different smoothing levels, starting with the unsmoothed images and ending with 2 cm FWHM Gaussian smoothed images (the FWHM is increased by steps of 2 mm). (a) High number of counts: 1048576 annihilation events. (b) Moderate number of counts: 262144 annihilation events. (c) Low number of counts: 65536 annihilation events.

Both $\mathcal{D}$ and $\mathcal{S}$ depend on $\boldsymbol{\varphi}$ and therefore cannot be neglected when calculating the derivative

$$\delta_{[\boldsymbol{\varphi}]_i}\mathcal{J}(f,\boldsymbol{\varphi}) = \delta_{[\boldsymbol{\varphi}]_i}\mathcal{D}(f,\boldsymbol{\varphi}) + \alpha\,\delta_{[\boldsymbol{\varphi}]_i}\mathcal{S}(\boldsymbol{\varphi}). \tag{26}$$

The first steps towards the variational derivative of $\mathcal{D}$ are very similar to (14)–(17), and we arrive at

$$\delta_{[\boldsymbol{\varphi}]_i}\mathcal{D} = \sum_{t,a} \delta_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f)\left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right). \tag{27}$$

From Appendix B we get

$$\delta_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi}) = \frac{1}{T}\int \nabla_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi})\,\eta(\boldsymbol{x},t)\,d\boldsymbol{x} \tag{28}$$

where

$$\nabla_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi}) = H(a,\boldsymbol{x})\frac{\partial}{\partial_i} f(\boldsymbol{\varphi}(\boldsymbol{x},t)). \tag{29}$$

Putting (28) into (27) leads to

$$\delta_{[\boldsymbol{\varphi}]_i}\mathcal{D} = \frac{1}{T}\int \sum_{t,a} \nabla_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi}) \times \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right)\eta(\boldsymbol{x},t)\,d\boldsymbol{x}. \tag{30}$$

The variational derivative of $\mathcal{S}$ with respect to $[\boldsymbol{\varphi}]_i$ is (see Appendix C)

$$\delta_{[\boldsymbol{\varphi}]_i}\mathcal{S}(\boldsymbol{\varphi}) = -\int_\Omega \sum_t \sum_j \frac{\partial^2}{\partial^2 x_j}[\boldsymbol{\varphi}]_i(\boldsymbol{x},t)\,\eta(\boldsymbol{x},t)\,d\boldsymbol{x}. \tag{31}$$

Inserting (30) and (31) into (26) leads to (32)

$$\delta_{[\boldsymbol{\varphi}]_i}\mathcal{J}(f,\boldsymbol{\varphi}) = \int \sum_t \left\{ \frac{1}{T}\sum_a \nabla_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi}) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right) - \alpha \sum_j \frac{\partial^2}{\partial^2 x_j}[\boldsymbol{\varphi}]_i(\boldsymbol{x},t) \right\} \eta(\boldsymbol{x},t)\,d\boldsymbol{x}. \tag{32}$$

From (25) and the fundamental lemma of calculus of variations it follows that:

$$\frac{1}{T}\sum_a \nabla_{[\boldsymbol{\varphi}]_i}\hat{g}(a,t,f,\boldsymbol{\varphi}) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f)}\right) - \alpha \sum_j \frac{\partial^2}{\partial^2 x_j}[\boldsymbol{\varphi}]_i(\boldsymbol{x},t) = 0. \tag{33}$$

Based on this equation, we finally define the motion update function $\tau_{[\boldsymbol{\varphi}]_i}$

$$\tau_{[\boldsymbol{\varphi}]_i}(f^{(k)},\boldsymbol{\varphi}^{(l)}) = [\boldsymbol{\varphi}]_i^{(l)} + \delta \times \left\{ \frac{1}{T}\sum_a \nabla_{[\boldsymbol{\varphi}]_i^{(l)}}\hat{g}(a,t,f^{(k)},\boldsymbol{\varphi}^{(l)}) \left(1 - \frac{g(a,t)}{\hat{g}(a,t,f^{(k)},\varphi^{(l)})}\right) - \alpha \sum_j \frac{\partial^2}{\partial^2 x_j}[\boldsymbol{\varphi}]_i^{(l)}(\boldsymbol{x},t) \right\}. \tag{34}$$

3) Discretization: The discrete counterpart of functional $\mathcal{J}$ is

$$J(\boldsymbol{f},\boldsymbol{\varphi}) = D(\boldsymbol{f},\boldsymbol{\varphi}) + \alpha\,S(\boldsymbol{\varphi}) \tag{35}$$

with

$$D(\boldsymbol{f},\boldsymbol{\varphi}) = \sum_t \mathbf{1}\left(\hat{\boldsymbol{g}}_{\boldsymbol{t}} - \boldsymbol{g}_{\boldsymbol{t}}\log\left(\hat{\boldsymbol{g}}_{\boldsymbol{t}}\right)\right) \tag{36}$$

$$\hat{\boldsymbol{g}}_{\boldsymbol{t}} = s_x s_y s_z \frac{1}{T}\,\boldsymbol{H}\,\boldsymbol{T}_{\boldsymbol{\varphi}_i}\,\boldsymbol{f} \tag{37}$$

and

$$S(\boldsymbol{\varphi}) = s_x s_y s_z \sum_{t,i,j} \|\nabla_j \boldsymbol{\varphi}_{it}\|^2. \tag{38}$$

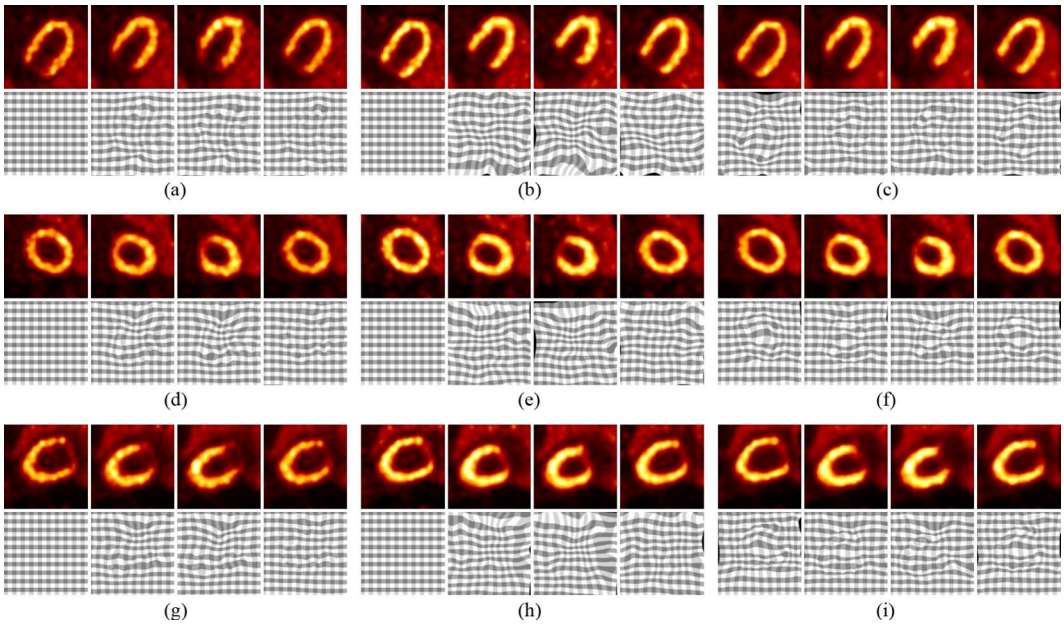


Fig. 8. Images reconstructed with JR, RFRF2-1G and RFRF2-3G for 262144 counts (medium level of statistical noise). Note that the motion fields of JR and the RFRF methods can not be directly compared since they refer to different reference frames (for JR this is a virtual reference frame, and for the RFRF methods it is the first frame). (a) Transverse slices, RFRF2-1G. (b) Transverse slices, RFRF2-3G. (c) Transverse slices, JR. (d) Coronal slices, RFRF2-1G. (e) Coronal slices, RFRF2-3G. (f) Coronal slices, JR. (g) Sagittal slices, RFRF2-1G. (h) Sagittal slices, RFRF2-3G. (i) Sagittal slices, JR.

$\boldsymbol{f}$ is a vector of size $n_x n_y n_z$ whose elements represent the tracer activities of the image function $f(\cdot)$. $\boldsymbol{T}_{\boldsymbol{\varphi}_t}$ is a $n_x n_y n_z \times n_x n_y n_z$ matrix that represents the deformation $\boldsymbol{\varphi}(\cdot, t)$ such that the matrix-vector multiplication $\boldsymbol{T}_{\boldsymbol{\varphi}_t}\boldsymbol{f}$ is the discrete representation of the deformed image $f(\boldsymbol{\varphi}(\cdot, t))$. Matrix-valued functions will be printed in bold and capital letters throughout the paper. $\boldsymbol{H}$ is the system matrix which is of size $n_l \times n_x n_y n_z$ (n_l is the number of LORs). $\mathbf{1}$ is a vector of size n_l with $[\mathbf{1}]_i = 1$ for every entry. s_x, s_y and s_z are the voxel spacings in all three spatial directions. Vector-vector products and vector-vector fractions are to be interpreted element-wise, so

$$\boldsymbol{a}\boldsymbol{b} := ([\boldsymbol{a}]_1[\boldsymbol{b}]_1, \ldots, [\boldsymbol{a}]_n[\boldsymbol{b}]_n)^\mathsf{T} \tag{39}$$

and

$$\frac{\boldsymbol{a}}{\boldsymbol{b}} := \left(\frac{[\boldsymbol{a}]_1}{[\boldsymbol{b}]_1}, \ldots, \frac{[\boldsymbol{a}]_n}{[\boldsymbol{b}]_n}\right)^\mathsf{T}. \tag{40}$$

$\boldsymbol{\varphi}_{it}$ is a vector of size $n_x n_y n_z$ representing the i-th motion function $[\boldsymbol{\varphi}]_i(\boldsymbol{x}, t)$ at time t. The complete motion vector is $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_{11}, \ldots, \boldsymbol{\varphi}_{1T}, \boldsymbol{\varphi}_{21} \ldots \boldsymbol{\varphi}_{2T}, \boldsymbol{\varphi}_{31} \ldots \boldsymbol{\varphi}_{3T})^\mathsf{T}$. ∇_i is a $n_x n_y n_z \times n_x n_y n_z$ matrix that represents the differential operator ∂/∂_i.

We use the following discretization for the image update:

$$\boldsymbol{f}^{(k+1)} = \boldsymbol{f}\frac{\sum_t \boldsymbol{j}_t \boldsymbol{T}_{\boldsymbol{\varphi}_t^{-1}} \boldsymbol{H}^\mathsf{T} \frac{\boldsymbol{g}_t}{\bar{\boldsymbol{g}}_t}}{\sum_t \boldsymbol{j}_t \boldsymbol{T}_{\boldsymbol{\varphi}_t^{-1}} \boldsymbol{H}^\mathsf{T} \mathbf{1}}. \tag{41}$$

$\boldsymbol{j}_t$ is a vector of size $n_x n_y n_z$ that represents $\left|\det(J_{\boldsymbol{\varphi}^{-1}(\cdot,t)})\right|$. We use Chen's method [27] (with five iterations) in order to calculate the inverse transformation which is needed for the image update. In theory, inversion may not be possible since it is not necessarily bijective. However, choosing a high enough parameter α prevents this problem, so no further regularization needs to be applied in practice (see also Section III-D and especially Fig. 13).

The motion update is discretized as follows:

$$\boldsymbol{\varphi}_{it}^{\text{new}} = \boldsymbol{\varphi}_{it} + \delta\left(\frac{1}{T}\boldsymbol{T}_{\boldsymbol{\varphi}_t}\nabla_i \boldsymbol{f}\boldsymbol{H}^\mathsf{T}\left(\mathbf{1} - \frac{\boldsymbol{g}_t}{\hat{\boldsymbol{g}}_t}\right) + \alpha\sum_j \nabla_j^2 \boldsymbol{\varphi}_{it}\right) \tag{42}$$

δ is a parameter that defines the stepsize. In the algorithm, it will be automatically updated (see Fig. 2). Similarly to ∇_i, ∇_i^2 is a matrix of the same dimensions that represents the differential operator ∂^2/∂_i^2.

Algorithm: The JR method is outlined in Algorithm B. It basically consists of alternating image and motion updates. For the current image, several motion updates are performed. The algorithm only accepts motion updates that improve the objective function, it is monotonic.

Since the objective function is nonconvex, the algorithm could get stuck in local minima. Also, in the motion update, it does not necessarily converge to a local minimum. However, we will see later in Section IV that for a well-chosen α in practice our algorithm seems to converge to the global optimum.

III. Evaluation

We test our algorithm for both simulated and patient data.

A. Simulation

1) Data Generation: As shown in Fig. 3, we generate four frames using the XCAT phantom [28]. One complete respiratory cycle of a length of five seconds is simulated, together with cardiac motion. Two experiments are performed: in the first one,

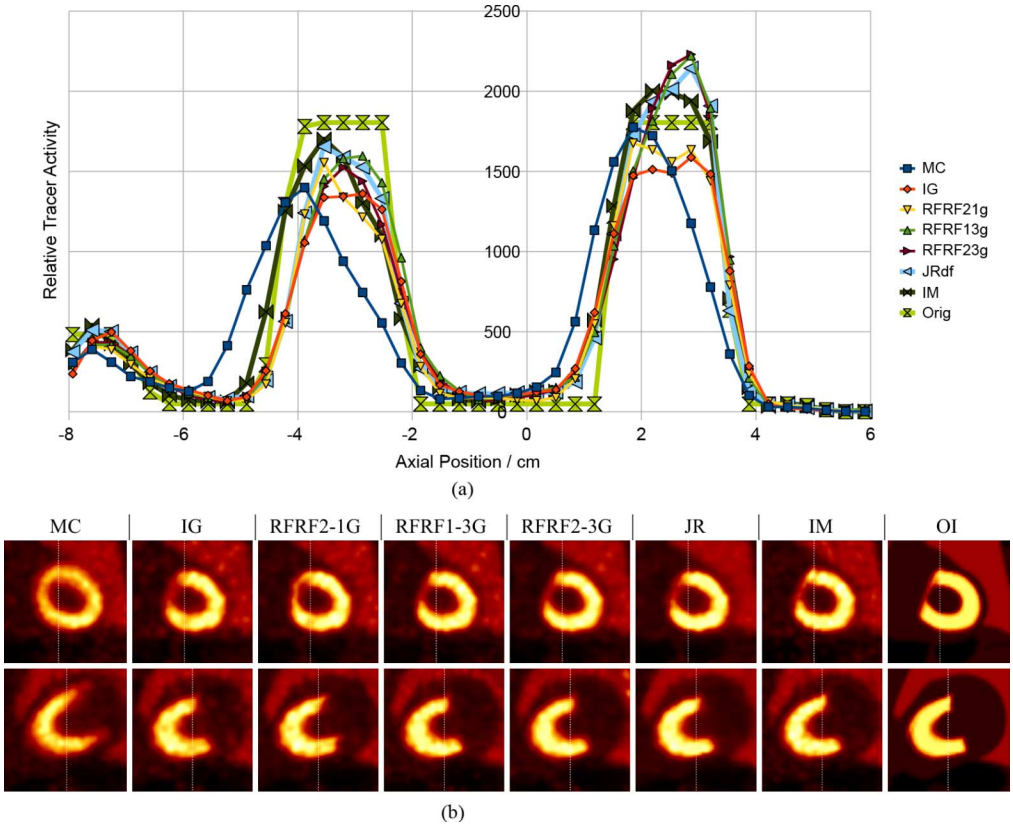


Fig. 9. Simulation: line profiles and their corresponding coronal and sagittal views at 65536 counts. (a) Axial profiles of the third gate drawn for different reconstruction methods from simulated data for a medium level of noise (262144 counts). (b) Coronal and sagittal views corresponding to the profiles. The dots mark the cutting line of the profile.

the extent of diaphragm motion is set to 1 cm, and in the second one to 2 cms. The cardiac frames were uniformly distributed over a cardiac cycle. For each frame, a volume of $48 \times 48 \times 48$ voxels containing the heart is cropped.

The expected number of counts for each LOR is calculated by projecting each gate to measurement space. The simulated measurements are finally generated from the expected number of counts by a Poisson random generator. In this way we take into account the acquisition time and activity. Several levels of statistical noise, representing long to extremely short acquisition times, are simulated. Since we want to focus on image degradations induced by motion, we did not make use of Monte-Carlo simulation packages which could model effects like scattering, random coincidences, etc. We simulate a Siemens Biograph Sensation 16 PET/CT scanner and use Scheins's algorithm to generate the system matrix [29].

2) Comparison: We compare our joint reconstruction approach (JR) to the following methods.

- An ML-EM reconstruction of motion-contaminated data (MC).
- An ML-EM reconstruction of the individual gates (IG).
- A registration and fusion of reconstructed frames method with image re-reconstruction and one grid level used for registration (RFRF2-1G, as described in Section III-C).
- A registration and fusion of reconstructed frames method with image summation and three grid levels used for registration (RFRF1-3G, as described in Section III-C).
- A registration and fusion of reconstructed frames method with image re-reconstruction and three grid levels used for registration (RFRF2-3G, as described in Section III-C).
- A motion compensating reconstruction based on the ideal motion (IM, as described below).
- The original images (OI).

The IM method is basically the same as the RFRF2 method, with the only difference that the motion is not estimated from the reconstructed noisy gates but from the original gates that were used for the simulation. In a clinical setting, instead of the original gates, gated CT images could be used for motion estimation.

As reference frame for the visual comparison we choose the third gate, since it is most different from the first gate which is used as the reference gate for the RFRF methods. Note that in the case of MC we take the complete reconstruction from all gates since this method is not able to deliver results for specific gates.

Comparison of Reconstructed Images: We compare the resulting images from the respective reconstruction methods to the original images (OI) by calculating the correlation coefficient. The correlation coefficient between two images $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$\mathcal{CC}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}}{||\boldsymbol{x}||||\boldsymbol{y}||}. \tag{43}$$

Both $\boldsymbol{x}$ and $\boldsymbol{y}$ are shifted such that their mean value is zero. The correlation coefficient ranges between -1 for totally anticorrelated images and 1 for perfectly correlated images.

In the case of reconstruction methods which deliver different images for each gate (IG, RFRF1, RFRF2, and JR), we calculate the correlation coefficient for each reconstructed gate with the respective original gate and take the average value (for the RFRF methods and JR this involves a deformation of the reconstructed image to the respective motion gates)

$$\mathcal{CC}^{\mathrm{IG}}_{\mathrm{image}} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{CC}(\boldsymbol{f}^{\mathrm{IG}}_t, \boldsymbol{f}^{\mathrm{OI}}_t) \tag{44}$$

$$\mathcal{CC}^{\mathrm{RFRF}}_{\mathrm{image}} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{CC}(\boldsymbol{T}_{\boldsymbol{\varphi}^{\mathrm{RFRF}-1}_t}\boldsymbol{f}^{\mathrm{RFRF}}, \boldsymbol{f}^{\mathrm{OI}}_t) \tag{45}$$

$$\mathcal{CC}^{\mathrm{JR}}_{\mathrm{image}} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{CC}(\boldsymbol{T}_{\boldsymbol{\varphi}^{\mathrm{JR}}_t}\boldsymbol{f}^{\mathrm{JR}}, \boldsymbol{f}^{\mathrm{OI}}_t). \tag{46}$$

Since the correlation coefficient depends heavily on the smoothness of the image, we have to make sure that we compare the images for the same level of background noise. We measure the noise by calculating the percent standard deviation (the standard deviation related to the mean value in percent) for a specific background region-of-interest (ROI). The reconstructed images are postsmoothed with differently sized Gaussian kernels, and the correlation coefficient calculated for each instance.

Two issues have to be taken care of when defining the background ROI. Firstly, since the heart is in a different position in the different frames, the ROI has to be chosen in such a manner that it contains background values for *all* image frames. Secondly, for a large smoothing, nonbackground values must not enter into the background ROI (if they did they would be falsely interpreted as background noise).

The results are discussed in Section IV-A.

Comparison of Reconstructed Motion: We visually compare the reconstructed motion fields one to another for different views (transverse, coronal, and sagittal) by extracting the 2-D in-plane components of the respective motion field and applying them to a 2-D checkerboard-like pattern. Note that for JR these motion fields are shown relative to a virtual reconstruction frame, and therefore they are not directly comparable to the RFRF motion fields.

B. Patient Data

1) Data Acquisition and Preprocessing: The data from a patient referred for an $^{18}\mathrm{F}-\mathrm{FDG}$ PET-CT examination for assessment of the myocardial viability was used. The patient was injected 400 MBq of $^{18}\mathrm{F}-\mathrm{FDG}$ and, 60 min after injection, a 10 min list-mode acquisition was performed using a Siemens Biograph Sensation 16 PET-CT scanner.

A novel image based gating procedure is used in order to divide the data into 24 gates: 1-s frames of data are reconstructed at a low resolution ($32 \times 32 \times 48$ voxels) and with only five ML-EM iterations. The heart is extracted from these reconstructed frames. Then, the heart frames are grouped into 24 gates, such that each gate contains the reconstructed frames which are most similar to each other. In this way, both cardiac and respiratory movements are taken into account.

Like for the simulation, we also use Scheins's algorithm for calculating the system matrix. Since we reconstruct a larger volume than in the simulation, we use a finer space voxel grid of $128 \times 128 \times 48$ voxels. Corrections for attenuation, scatter or randoms were not considered since the necessary data was not available.

2) Comparison: Since we do not know the original tracer distribution as in the simulation study, it is hard to quantitatively determine which method is performing best. We will thus rely on a visual comparison and the analysis of the pronunciation of motion by means of profiles. The results are shown in Section IV-B.

C. Registration and Fusion of Reconstructed Frames

As mentioned in the introduction, most motion compensation methods separate motion estimation from image estimation. "Registration and fusion of reconstructed frames" (RFRF) is a class of methods that extract the motion from the data by image registration. For comparison with our joint reconstruction algorithm, we implemented two types of RFRF which are commonly used in practice. They are based on the same motion estimation procedure and differ only in the image estimation part.

1) Motion Estimation: Motion estimation consists in 1) reconstructing each gate individually, and 2) registration of each of the reconstructed gates to a reference gate (in our case the first gate).

Reconstruction is done using the ML-EM algorithm with thirty iterations. The images are then postsmoothed with a 5 mm FWHM Gaussian kernel, a procedure which is also employed by the reconstruction software of the Siemens Biograph 16.

The goal of the image registration step is to obtain a motion function $\boldsymbol{\varphi}_{\mathrm{RFRF}}(\boldsymbol{x}, t)$ which defines the deformation of a reference frame (we use the first frame) to an arbitrary frame t. In order to estimate this motion function, we use *drop*, a software for deformable image registration using discrete optimization [30]–[32]. In the following, we will shortly describe the relevant parameters of *drop*.

The exact parameters used for registration with *drop* are given in Table I.

Multiple Grid Levels: The most important parameter is the number of grid levels and the initial grid size. We obtained the best registration results by setting the the initial grid size to $5 \times 5 \times 5$ and the number of grid levels to three in case of the simulations and four in case of patient data. The grid is refined in each level (actually the number of patches, which is one less than the grid size, is doubled), so in the third level we arrive at a grid size of $17 \times 17 \times 17$. In general it is advisable to not use more control points than necessary in order to reproduce the motion. Few control points lead to a more robust registration. Fig. 4 shows a justification that a grid of $17 \times 17 \times 17$ control points is enough in order to model the original motion.

The number of iterations was set to 10 (per grid level). The objective function did not further improve after 10 iterations, which indicates that more iterations were not necessary.

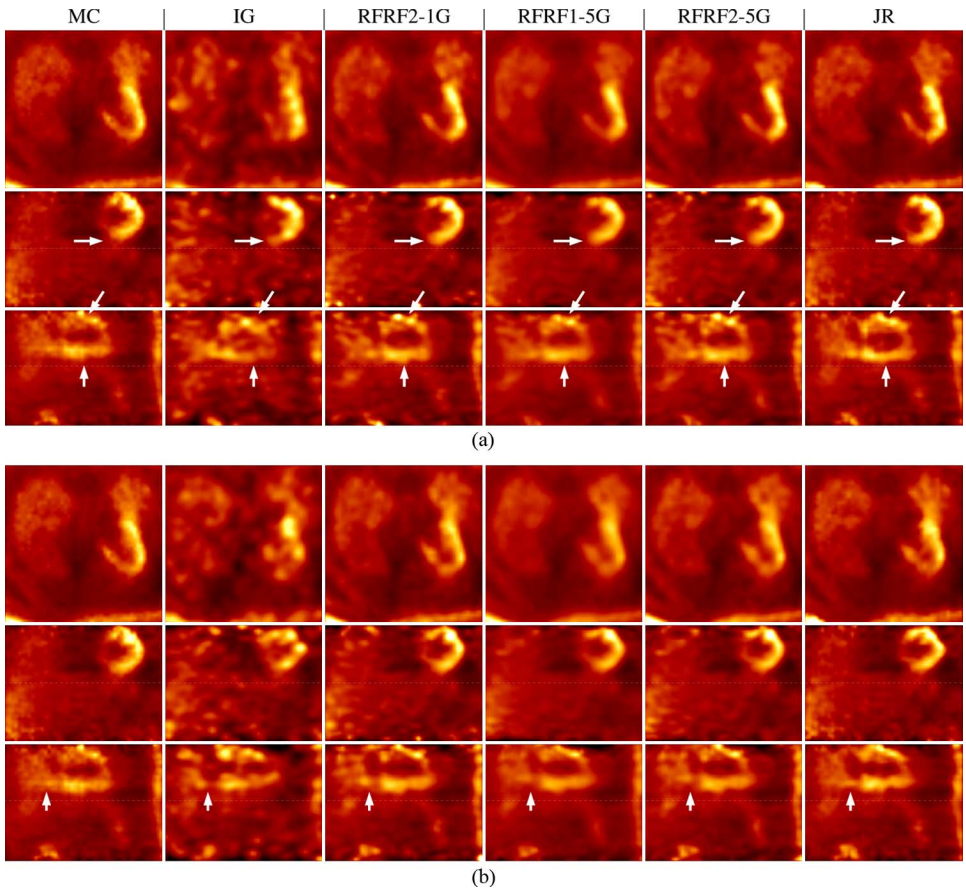


Fig. 10. Transverse, coronal and sagittal slices for patient data (from left to right): ML-EM reconstruction of motion-contaminated data (MC), ML-EM reconstruction of an individual gate (IG), RFRF2-1G (as described in section III-C), RFRF1-5G, RFRF2-5G and our joint reconstruction (JR). There are dotted line which is located at the same position in the coronal and sagittal slices facilitates the visual assessment of the motion. (a) Inhalation. (b) Exhalation.

There is a parameter λ to be set which defines the stiffness of the displacement field. We choose λ in the following manner: we start with $\lambda = 1$. If the result is an overlapping displacement field, which means that it is noninvertible, we increase λ until no overlapping can be noticed anymore. If the displacement field is not overlapping, than we reduce λ until we get as close as possible to an overlapping displacement field.

In theory, λ should depend on the image intensity ranges used in the source and target image. However, we did not observe any differences for different images (maybe λ is not very sensitive to the latter), and so the same λ was used for all count levels.

One Grid Level: Since our joint reconstruction approach does not make use of multiple grid levels but rather works on only one grid level (with a grid size of $49 \times 49 \times 49$ for the case of synthetic data), it is interesting to see how well the registration approach performs for a similar setting. With only one grid level, one has to increase the label factor to one. The label factor defines the increase of the search area. We set the grid size to $49 \times 49 \times 49$. The correct stiffness parameter λ is found in the same manner as in the case of three grid levels. Again, different λ's were not necessary for the different count levels.

For patient data, in theory we would have to apply a $129 \times 129 \times 49$ in order to be consistent with taking one patch per voxel. However, this is a vast number of grid points and dramatically increases the computation time and memory requirements. We therefore used a $65 \times 65 \times 19$ grid size.

Like in the multiple grid levels case, the number of iterations was chosen as high that no further improvements of the objective function were noted (at the precision given by "drop").

2) Image Estimation:

RFRF1: RFRF1 computes the image as the summation of all registered gates (similar to [7]–[10], [33]): $f_{\mathrm{RFRF}}(\boldsymbol{x}) = \sum_t f_t(\boldsymbol{\varphi}^{-1}_{\mathrm{RFRF}}(\boldsymbol{x}, t))$.

RFRF2: The image is completely re-reconstructed based on the whole data. We do this in a similar manner as in [12]–[14]. Note that, in contrast to [12]–[14], we derive the motion model from the individually reconstructed gates (as explained previously). In this way, we achieve a fair comparison since all methods work with the same data (except for IM, of course).

Technically, the re-reconstruction is done by omitting the complete motion estimation part in our joint reconstruction algorithm. Instead, the deformation field that results from the previously explained registration is used in JR as a constant value for all iterations.

D. Choice of Regularization Parameter α

As mentioned before, the correct choice of the parameter α is crucial for successful motion compensation. Fig. 5 shows three JR reconstructions with different values for α. When α is too low, extreme deformations in the motion field and image

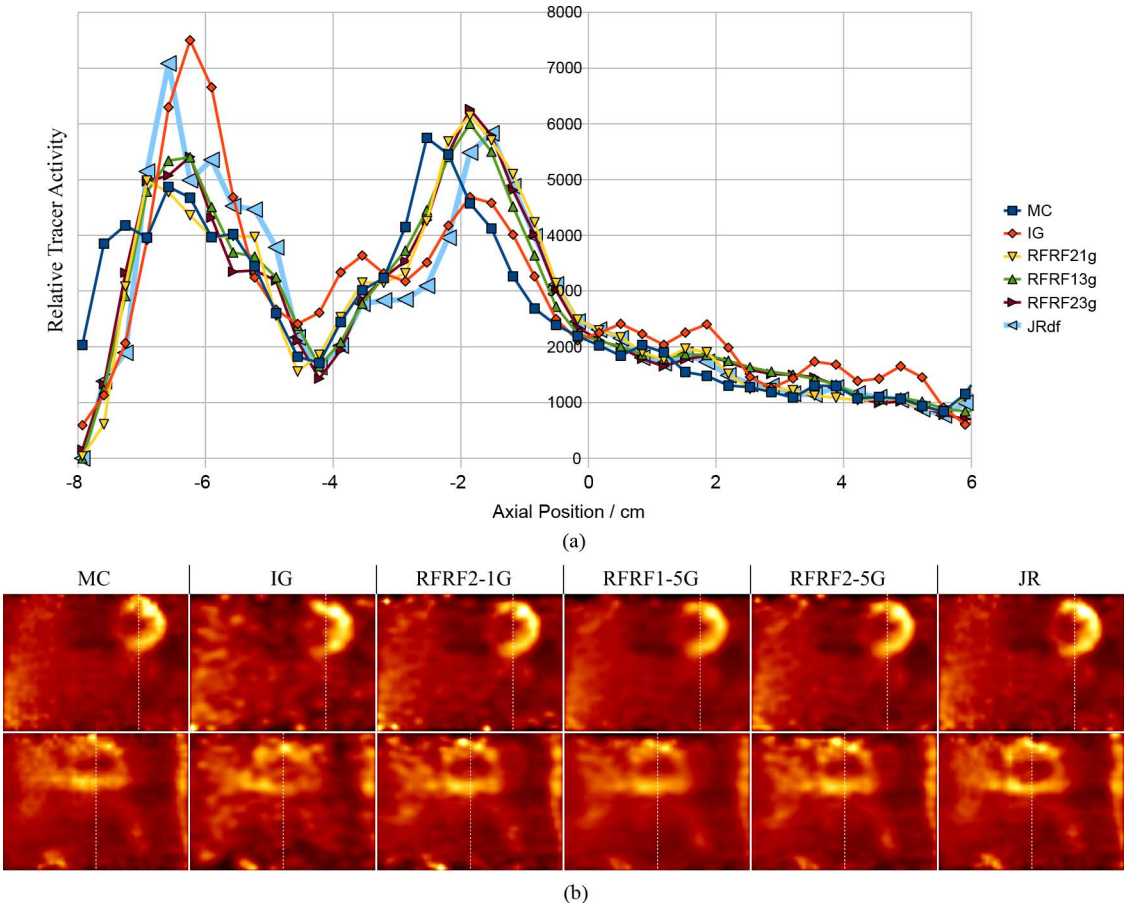


Fig. 11. Patient data: line profiles and their corresponding coronal and sagittal views for an inhalation gate. The dotted line indicates the position at which the profiles are drawn from (a) axial profiles of the third gate drawn for different reconstruction methods from simulated data for a medium level of noise (262144 counts), (b) coronal and sagittal views corresponding to the profiles. The dots mark the cutting line of the profile.

artifacts are present. Also, a too small α may lead to noninvertible deformation fields (for a discussion on regularization specifically dedicated to preventing noninvertible deformation fields, see [34]). When α is too high, almost no deformations are present and the image looks like as if no motion compensation had taken place (it is blurred).

The correct choice of such regularization parameters is still an unsolved problem in general. For some specific problems, an analysis of the L-curve has been successful [35], [36]. Also, generalized cross-validation is a potential candidate for finding the correct regularization parameter [37].

In any case, finding the correct α is very time consuming. L-curve methods and generalized cross-validation are aimed at scenarios where no ground truth data are available. In our case, we have done a simulation of the system and then made a quantitative comparison with the ground truth data. The α that gave the best results was chosen for any subsequent reconstruction using the same scanner.

In order to do this, we will normalize α in such manner that it is invariant to the number of counts (that is, acquisition time and activity) and not too sensitive to the present motion. Our goal is to be able to use the same α for different patients, acquisition times and tracer activities. α will then be calculated just once for each system.

The rationale behind the normalization is the following: the first motion update should be ideally the same for every noise level, in relation to the smoothing term $\mathcal{S}$. We, therefore, calculate the average displacement vector (for each gate individually, then averaged over all gates). α is then divided by the magnitude of this average displacement vector.

In Fig. 13 we plot the correlation coefficients for α for different noise scenarios and reconstruction methods (the same normalization has been applied to the RFRF methods). We have also assessed whether the same parameter can be used for different motion. The results are discussed in Section IV.

E. Convergence Properties and Postsmoothing

All compared algorithms were executed until convergence.

For ML-EM it is well known that the maximum-likelihood result is not necessarily the closest to the ideal tracer distribution, since the image is adapted too much to the actually measured data (overfitting). The most common solutions to this problem are to stop the ML-EM algorithm prematurely (that is, before convergence), to use regularization methods like Gibbs priors in order to encourage smooth images, or to postsmooth the resulting image. We apply postsmoothing after convergence of the respective ML-EM algorithm.

For simulated data, the smoothing kernel which leads to the maximal correlation of the respective method with the original data is chosen. For patient data, we decided on using a Gaussian postsmoothing with 2 mm FWHM, since this is the recommended setting in the clinical reconstruction software of the Siemens Biograph 16.

For the RFRF methods, the smoothed IG images are taken for motion estimation. For image estimation, in case of RFRF1 the smoothed IG images are transformed according to the registered transformation and summed. No further smoothing is then performed. For RFRF2 smoothing is also applied after image estimation.

In Section IV-E the convergence curves of the RFRF motion estimation step and the JR algorithm are discussed for different values of α.

IV. Results and Discussion

In the following we quantitatively and visually inspect the reconstructed image and motion for both simulation and real patient data.

A. Simulation

Fig. 6 shows representative transverse, coronal and sagittal slices for three levels of noise. While JR and RFRF1/2-3G provide pretty similar results (for the moderate and low count levels one can note the higher background noise present in RFRF1/2-3G compared to JR), it can be clearly noted that the myocardial muscle reconstructed with RFRF2-1G is thinner, more elongated and less contracted than in JR, IM, and OI for the moderate and low count levels. MC is obviously worse for all count levels, and IG at least in the moderate and low count levels appears to be very noisy.

Fig. 7 summarizes the results for the correlation coefficient plotted versus the noise level between reconstructed images and original images as described in Section III-AII. As expected, the maximum of the correlation curve corresponding to IM is higher than the maximum of any other method and MC is lowest, in all three statistical scenarios. For all count levels, JR has the second highest maximal correlation coefficient, followed by RFRF2-3G, RFRF1-3G, RFRF2-1G, and IG. In the high and moderate count level, JR's maximal correlation coefficient is only slightly higher than RFRF2-3G and could also be considered as equal. In case of moderate and high count numbers, JR's correlation curve together with its maximum is shifted to a lower noise level: for the high count numbers case, the maximum of RFRF2-3G lies at about 50%, and the maximum of JR at about 40%; for the moderate count numbers case, the maximum of RFRF2-3G lies at about 72% and the maximum of JR at about 52%.

Fig. 8 shows exemplarily the reconstructed gates and their corresponding motion fields for JR and RFRF2-3G at 262144 counts. The images are compared at the noise level corresponding to their maximal correlation coefficient with the original image. It can be clearly noted that RFRF2-1G is most noisy, both in the background and in the foreground.

RFRF2-3G is noisier than JR in the background. Note that JR has a virtual reference frame, and accordingly the deformation field refers to this virtual reference frame and can not be directly compared to their RFRF counterparts.

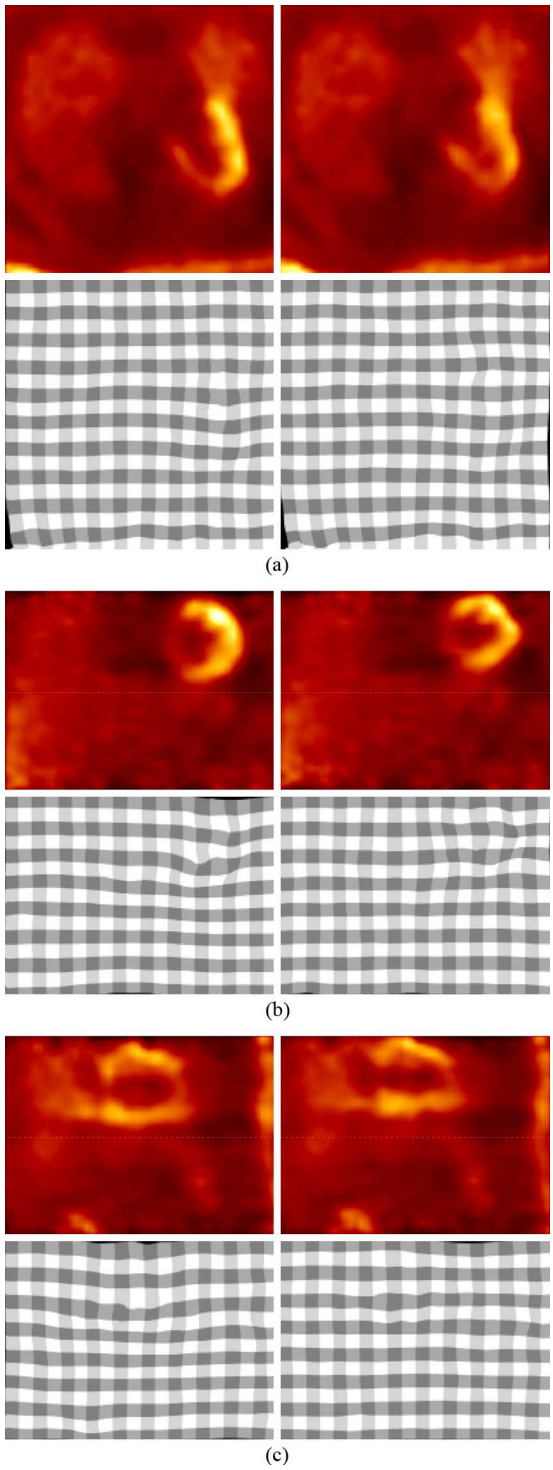


Fig. 12. Patient data: transverse, coronal and sagittal slices of an inhalation gate (left) and an exhalation gate (right) together with the corresponding motion fields, reconstructed with JR. The dotted line is drawn for reference in order to facilitate the visual assessment of the movement. (a) Transverse view. (b) Coronal view. (c) Sagittal view.

Profiles for the 1048576 counts case are drawn in order to assess the capability of reconstructing a clear peak which represents the myocardial wall (see Fig. 9). The effect of deblurring of the motion compensated methods compared to MC can be

clearly seen. Note that due to the high level of noise, we do not recommend to take the profile plots as a reliable basis for judgement between IG, RFRF, JR, and IG.

B. Patient Data

Fig. 10 compares JR to MC, IG, and RFRF. Approximately 45 million counts were acquired in a 2-min time window used for reconstruction. However, it is not possible to directly compare the patient data case to a simulation case, partly because these 45 million counts originate from the whole body, not only from the heart.

Note that the quality may seem inferior to the simulated data, which can be attributed to physical degradation phenomena like attenuation, randoms, and scatter which we do not correct for in this paper.

Two comparison figures are shown, one for maximal inhalation and one for maximal exhalation. The dotted line marks the lower limit of the myocardial wall in the maximal exhalation state for JR.

The coronal and sagittal slices in the inhalation state of JR show that the myocardial wall is less blurred than in the rest of the methods, especially compared to MC. In the exhalation state, the left atrial myocardium is clearer visible for JR.

The previous observations with respect to the blurring are confirmed by an axial profile plot (see Fig. 11). JR's myocardial wall is thinner due to reduced blurring. Note that the profile plot should be inspected with care due to the high amount of noise.

Fig. 12 shows coronal slices of all gates and their corresponding motion fields for JR. The dotted line again marks the lower limit of the myocardial wall.

C. Registration and Fusion of Reconstructed Frames

As shown in the preceding Sections IV-A and IV-B, JR is visually and quantitatively clearly better than RFRF2-1G. However, compared to RFRF1/2-3G/5G, differences are more subtle. In our opinion, JR can be compared in a fair way only to RFRF2-1G and not to RFRF1/2-3G/5G, because RFRF1/2-3G/5G work with multiple levels of deformation grids, starting with a coarse deformation grid which is refined in later iterations. Especially for noisy data, a coarse deformation grid is much more robust than a fine grid. Since JR is working with only one fine grid level (a displacement field where each image voxel corresponds to a displacement vector), it has difficulties at adopting for large movements.

D. Choice of Regularization Parameter α

Fig. 13 shows the dependence of the quality of the reconstructed images of α. The values given are the values for α before normalization (as described in Section III-D). For all methods, $\alpha = 2$ corresponds to the highest correlation coefficient. For JR this is the case consistently for all levels of noise.

Fig. 13 also relates the choice of α to the maximum correlation coefficient of the reconstructed images for two different kinds of motions. The value of α that maximizes the correlation is the same in both cases, which indicates that α does not have to be adjusted for different patients.

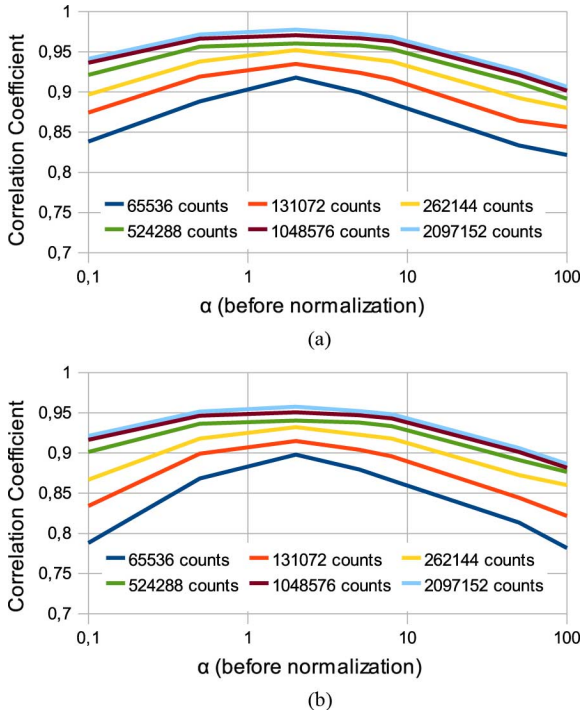


Fig. 13. The maximal (chosen from different postsmoothing scenarios) mean (of all four gates) correlation coefficient of JR plotted for various values of α. The α corresponding to the highest correlation coefficient is selected as the best α for further comparisons. Note that α does not depend on the number of measured counts nor on the breathing motion. This is an important practical result, because once the parameter α is found for a specific system, it can be reused for different patient studies with varying activity/duration. (a) 1 cm diaphragm motion. (b) 2 cm diaphragm motion.

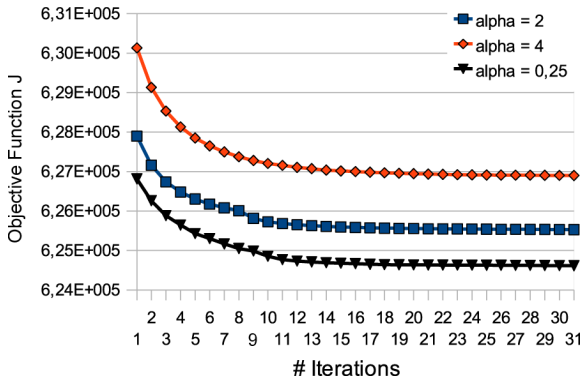


Fig. 14. Value of objective function of JR for the number of outer iterations.

E. Convergence Properties and Postsmoothing

As mentioned in Section III-E, the convergence curves for JR depend on the parameter α. Fig. 14 shows three representative curves.

V. Conclusion

We present a novel reconstruction algorithm for gated positron emission tomography which jointly reconstructs both image and motion. No external motion information is needed, gating is achieved by a novel image based procedure. In a simulation study, both quantitative and visual comparison show that

our method can compete with state-of-the-art registration algorithms for all simulated count conditions and particularly in low counts scenarios. We show the feasibility of the method with patient data, indicating the potential of the presented approach for reducing motion blur. Further validation on patient data should help determine the impact of the improvement achieved in the diagnostic accuracy of positron emission tomography. The results for patient data can be expected to improve by including physical degradation effects like attenuation, scattering and randoms in the system model.

APPENDIX A

$$\delta_f \hat{g}(a,t,f) \tag{47}$$

$$= \frac{\partial}{\partial \varepsilon} \hat{g}(a,t,f+\varepsilon\eta)\Big|_{\varepsilon=0} \tag{48}$$

$$= \frac{\partial}{\partial \varepsilon} \frac{1}{T} \int H(a,\boldsymbol{x})[f+\varepsilon\eta](\boldsymbol{\varphi}(\boldsymbol{x},t))\, d\boldsymbol{x}\Big|_{\varepsilon=0} \tag{49}$$

$$= \frac{1}{T} \int H(a,\boldsymbol{x})\, \frac{\partial}{\partial \varepsilon}[f+\varepsilon\eta](\boldsymbol{\varphi}(\boldsymbol{x},t))\Big|_{\varepsilon=0} d\boldsymbol{x} \tag{50}$$

$$= \frac{1}{T} \int H(a,\boldsymbol{x})\, \eta(\boldsymbol{\varphi}(\boldsymbol{x},t))\, d\boldsymbol{x}. \tag{51}$$

With integration by substitution for multiple variables, (51) can be rewritten as

$$\delta_f \hat{g}(a,t,f) = \int \nabla_f \hat{g}(a,t,f)\, \eta(\boldsymbol{x})\, d\boldsymbol{x} \tag{52}$$

with

$$\nabla_f \hat{g}(a,t,f) := \frac{1}{T} H(a,\varphi^{-1}(\boldsymbol{x},t))\, |\det(J_{\varphi^{-1}}(\boldsymbol{x},t))|. \tag{53}$$

APPENDIX B

$$\delta_{[\boldsymbol{\varphi}]_i} \hat{g}(a,t,f) \tag{54}$$

$$= \frac{\partial}{\partial \varepsilon} \hat{g}(a,t,[\boldsymbol{\varphi}]_i+\varepsilon\eta)\Big|_{\varepsilon=0} \tag{55}$$

$$= \frac{1}{T} \int H(a,\boldsymbol{x})\ \frac{\partial}{\partial \varepsilon} f(\ldots[[\boldsymbol{\varphi}]_i+\varepsilon\eta](\boldsymbol{x},t)\ldots)\Big|_{\varepsilon=0} d\boldsymbol{x} \tag{56}$$

$$= \frac{1}{T} \int H(a,\boldsymbol{x})\, \frac{\partial}{\partial_i} f(\boldsymbol{\varphi}(\boldsymbol{x},t))\, \eta(\boldsymbol{x})\, d\boldsymbol{x}. \tag{57}$$

We rewrite (57) as

$$\delta_{[\boldsymbol{\varphi}]_i} \hat{g}(a,t,f,\boldsymbol{\varphi}) = \frac{1}{T} \int \nabla_{[\boldsymbol{\varphi}]_i} \hat{g}(a,t,f,\boldsymbol{\varphi})\, \eta(\boldsymbol{x},t)\, d\boldsymbol{x} \tag{58}$$

where

$$\nabla_{[\boldsymbol{\varphi}]_i} \hat{g}(a,t,f,\boldsymbol{\varphi}) = H(a,\boldsymbol{x}) \frac{\partial}{\partial_i} f(\boldsymbol{\varphi}(\boldsymbol{x},t)). \tag{59}$$

APPENDIX C

$$\delta_{[\boldsymbol{\varphi}]_i} \mathcal{S}(\boldsymbol{\varphi}) \tag{60}$$

$$= \frac{\partial}{\partial \varepsilon} \mathcal{S}(\ldots[\boldsymbol{\varphi}]_i+\varepsilon\eta\ldots)\Big|_{\varepsilon=0} \tag{61}$$

$$= \sum_t \sum_j \int_\Omega \frac{\partial}{\partial \varepsilon} \left(\frac{\partial}{\partial x_j} ([\boldsymbol{\varphi}]_i+\varepsilon_i\eta)(\boldsymbol{x},t) \right)^2 \Big|_{\varepsilon=0} d\boldsymbol{x} \tag{62}$$

$$= \sum_t \sum_j \int_\Omega \frac{\partial}{\partial x_j} [\boldsymbol{\varphi}]_i(\boldsymbol{x},t) \frac{\partial}{\partial x_j} \eta(\boldsymbol{x},t)\, d\boldsymbol{x} \tag{63}$$

$$= \sum_t \sum_j \left[\frac{\partial}{\partial x_j} [\boldsymbol{\varphi}]_i(\boldsymbol{x},t) \eta(\boldsymbol{x},t) \right]_\Omega - \int_\Omega \frac{\partial^2}{\partial^2 x_j} [\boldsymbol{\varphi}]_i(\boldsymbol{x},t) \eta(\boldsymbol{x},t)\, d\boldsymbol{x} \tag{64}$$

$$= -\int_\Omega \sum_t \sum_j \frac{\partial^2}{\partial^2 x_j} [\boldsymbol{\varphi}]_i(\boldsymbol{x},t)\, \eta(\boldsymbol{x},t)\, d\boldsymbol{x}. \tag{65}$$

ACKNOWLEDGMENT

The authors would like to thank K. Brzezinski for proofreading.

REFERENCES

[1] Y. Wang, E. Vidan, and G. W. Bergman, "Cardiac motion of coronary arteries: Variability in the rest period and implications for coronary MR angiography," *Radiology*, vol. 213, no. 3, pp. 751–758, 1999.

[2] G. Shechter, C. Ozturk, J. Resar, and E. McVeigh, "Respiratory motion of the heart from free breathing coronary angiograms," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 1046–1056, Aug. 2004.

[3] L. Boucher, S. Rodrigue, R. Lecomte, and F. Bénard, "Respiratory gating for 3-dimensional PET of the thorax: Feasibility and initial results," *J. Nucl. Med.*, vol. 45, no. 2, pp. 214–219, Feb. 2004.

[4] P. Danias, M. Stuber, R. Botnar, K. Kissinger, R. Edelman, and W. Manning, "Relationship between motion of coronary arteries and diaphragm during free breathing: Lessons from real-time MR imaging," *Am. J. Roentgenol.*, vol. 172, no. 4, pp. 1061–1065, 1999.

[5] A. Martínez-Möller, D. Zikic, R. Botnar, R. Bundschuh, W. Howe, S. Ziegler, N. Navab, M. Schwaiger, and S. Nekolla, "Dual cardiac-respiratory gated PET: Implementation and results from a feasibility study," *Eur. J. Nucl. Med. Molecular Imag.*, vol. 34, no. 9, pp. 1447–1454, 2007.

[6] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Image Process.*, vol. 1, no. 2, pp. 113–122, Oct. 1982.

[7] G. J. Klein, B. W. Reutter, and R. H. Huesman, "4-D affine registration models for respiratory-gated PET," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, Lyon, France, 2000, vol. 2, pp. 41–15.

[8] G. J. Klein, B. W. Reutter, and R. H. Huesman, "Non-rigid summing of gated PET via optical flow," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, Anaheim, CA, USA, Nov. 1996, vol. 2, pp. 1339–1342.

[9] G. J. Klein, "Four-dimensional processing of deformable cardiac PET data," in *Proc. IEEE Workshok Math. Methods in Biomed. Image Anal.*, Hilton Head Island, SC, 2000, pp. 127–134.

[10] M. Dawood, N. Lang, X. Jiang, and K. P. Schäfers, "Lung motion correction on respiratory gated 3-D PET/CT images," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 476–485, Apr. 2006.

[11] D. R. Gilland, B. A. Mair, and J. G. Parker, "Motion estimation for cardiac emission tomography by optical flow methods," *Phys. Med. Biol.*, vol. 53, no. 11, pp. 2991–2991, 2008.

[12] M. Reyes, G. Malandain, P. M. Koulibaly, M. A. González-Ballester, and J. Darcourt, "Model-based respiratory motion compensation for emission tomography image reconstruction," *Phys. Med. Biol.*, vol. 52, no. 12, pp. 3579–3600, 2007.

[13] F. Qiao, T. Pan, J. W. C. , Jr., and O. R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies," *Phys. Medi. Biol.*, vol. 51, no. 15, pp. 3769–3783, 2006.

[14] F. Lamare, M. J. L. Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C. Cheze, L. Rest, A. J. Reader, and D. Visvikis, "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," *Phys. Med. Biol.*, vol. 52, no. 17, pp. 5187–5204, 2007.

[15] D. Gilland, B. Mair, J. Bowsher, and R. Jaszczak, "Simultaneous reconstruction and motion estimation for gated cardiac ECT," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 5, pp. 2344–2349, Oct. 2002.
[16] Z. Cao, D. Gilland, B. Mair, and R. Jaszczak, "Three-dimensional motion estimation with image reconstruction for gated cardiac ECT," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 3, pp. 384–388, Jun. 2003.
[17] B. Mair, D. Gilland, and J. Sun, "Estimation of images and nonrigid deformations in gated emission CT," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1130–1144, Sep. 2006.
[18] E. Gravier, Y. Yang, M. A. King, and M. Jin, "Fully 4-D motion-compensated reconstruction of cardiac SPECT images," *Phys. Med. Biol.*, vol. 51, no. 18, pp. 4603–4619, 2006.
[19] A. J. Reader, F. C. Sureau, C. C. R. Trébossen, and I. Buvat, "Joint estimation of dynamic PET images and temporal basis functions using fully 4-D ML-EM," *Phys. Med. Biol.*, vol. 51, no. 21, pp. 5455–5474, 2006.
[20] J. Parker and D. Gilland, "Wall motion estimation for gated cardiac emission tomography: Physical phantom evaluation," *IEEE Trans. Nucl. Sci.*, vol. 55, no. 1, pp. 531–536, Feb. 2008.
[21] M. Jacobson and J. Fessler, "Joint estimation of image and deformation parameters in motion-corrected PET," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct. 2003, vol. 5, pp. 3290–3294.
[22] M. Blume, M. Rafecas, S. Ziegler, and N. Navab, "Combined motion compensation and reconstruction for PET," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct. 2008, pp. 5485–5487.
[23] M. Blume, A. Keil, N. Navab, and M. Rafecas, "Blind motion compensation for positron-emission-tomography," *Proc. SPIE*, vol. 7258, p. 72 580T–72 580T–8, Feb. 2009.
[24] H. Schumacher, J. Modersitzki, and B. Fischer, "Combined reconstruction and motion correction in SPECT imaging," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 1, pp. 73–80, Feb. 2009.
[25] J. Modersitzki, *Numerical Methods for Image Registration*. New York: Oxford Univ. Press, 2004.
[26] D. R. Smith, *Variational Methods in Optimization*. Mineola, NY: Dover, 1998.
[27] M. Chen, W. Lu, Q. Chen, K. J. Ruchala, and G. H. Olivera, "A simple fixed-point approach to invert a deformation field," *Med. Phys.*, vol. 35, no. 1, pp. 81–88, 2008.
[28] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4-D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, 2008.
[29] J. J. Scheins, F. Boschen, and H. Herzog, "Analytical calculation of volumes-of-intersection for iterative, fully 3-D PET reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1363–1369, Oct. 2006.
[30] B. Glocker and N. Komodakis, Drop Oct. 2009 [Online]. Available: http://www.mrf-registration.net/
[31] B. Glocker, N. Komodakisa, G. Tziritasc, N. Navab, and N. Paragios, "Dense image registration through MRFs and efficient linear programming," *Med. Image Anal.*, vol. 12, pp. 731–741, 2008.
[32] N. Komodakis, G. Tziritas, and N. Paragios, "Fast, approximately optimal solutions for single and dynamic MRFs," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007.
[33] G. J. Klein and R. H. Huesman, "A 3D optical flow approach to addition of deformable PET volumes," in *Proc. IEEE Nonrigid Articulated Motion Workshop*, San Juan, Puerto Rico, Jun. 1997, pp. 136–143.
[34] S. Y. Chun and J. A. Fessler, "Spatial resolution and noise properties of regularized motion-compensated image reconstruction," in *Proc. 6th IEEE Int. Conf. Symp. Biomed. Imag.*, 2009, pp. 863–866.
[35] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
[36] N. Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using L-curve," in *IEEE Int. Symp. Circuits Syst.*, May 2001, vol. 2, pp. 433–436.
[37] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.