IOP Publishing Phys. Med. Biol. 65 (2020) 015004 (19pp) https://doi.org/10.1088/1361-6560/ab554a

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

MR-MOTUS: model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal *k*-space data

Niek R F Huttinga[1], **Cornelis A T van den Berg, Peter R Luijten and Alessandro Sbrizzi**

Imaging Division of the University Medical Center, Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands

[1] Author to whom any correspondence should be addressed.

E-mail: n.r.f.huttinga@umcutrecht.nl

RECEIVED
12 August 2019

REVISED
5 November 2019

ACCEPTED FOR PUBLICATION
7 November 2019

PUBLISHED
10 January 2020

Keywords: MR-guided radiotherapy, non-rigid motion estimation, inverse problems, model-based reconstruction, high frame rate motion

Supplementary material for this article is available online

Abstract

Time-resolved motion estimation from MRI data has received an increasing amount of interest due to the advent of the MR-Linac. The combination of an MRI scanner and a linear accelerator enables radiation plan adaptation based on internal organ motion estimated from MRI data. However, time-resolved estimation of this motion from MRI data still remains a challenge. In light of this application, we propose MR-MOTUS, a framework to estimate non-rigid 3D motion from minimal *k*-space data. MR-MOTUS consists of two main components: (1) a signal model that explicitly relates the *k*-space signal of a deforming object to non-rigid motion-fields and a reference image, and (2) model-based reconstructions of the non-rigid motion-fields directly from *k*-space data. Using an *a priori* available reference image and the fact that internal body motion exhibits a high level of spatial correlation, we represent the motion-fields in a low-dimensional space and reconstruct them from minimal *k*-space data that can be acquired very rapidly. The signal model is validated through numerical experiments with a digital 3D phantom and motion-fields are reconstructed from retrospectively undersampled *in vivo* head and abdomen data using various undersampling strategies. A comparison is made with state-of-the-art image registration performed on images reconstructed from the same undersampled data. Results show that MR-MOTUS reconstructs *in vivo* 3D rigid head motion from 474-fold retrospectively downsampled *k*-space data, and *in vivo* non-rigid 3D respiratory motion from 63-fold retrospectively undersampled *k*-space data. Preliminary results on prospectively undersampled data acquired with a 2D golden angle acquisition during free-breathing demonstrate the practical feasibility of the method.

1. Introduction

In recent years, time-resolved motion estimation from MRI data has received an increasing amount of attention due to the advent of the MR-Linac (Raaymakers *et al* 2009). The combination of an MRI scanner and a linear accelerator allows for MR-guided radiotherapy (MRgRT): radiation plan adaptation based on tumor motion that is estimated from MRI data. Hence, for MRgRT, it is the motion itself that is of particular interest. Two types of motion are usually distinguished: inter and intrafraction motion. Interfraction motion is defined as the day-to-day motion of the target volume in between treatments, e.g. due to different patient positioning or volume differences in the stomach or bladder. Intrafraction motion is defined as the motion of the target volume during the treatment, e.g. due to respiration, cardiac contractions, bowel motion or abrupt patient movement. Until now, the MR-Linac has mostly been used to estimate and correct for interfraction motion. For this, a pre-treatment MRI is made to adjust the previously computed radiation plan to the day-to-day anatomy (Raaymakers *et al* 2017). Correcting for the intrafraction motion, however, is a much more challenging problem as it requires to estimate the non-rigid internal body motion during the actual radiotherapy treatment, i.e. in an online setting

with low latency (∼milliseconds). In this setting, only minimal *k*-space data will be available for the non-rigid motion reconstruction. In light of the MRgRT application, we focus in this work on the problem of estimating non-rigid motion from minimal *k*-space data. Several methods to estimate motion in MRI have been proposed that could fit in this online MRgRT setting. These can broadly be subdivided in three categories: image-based methods, surrogate signal methods and *k*-space methods.

Image-based methods estimate motion indirectly from *k*-space data by first reconstructing and subsequently co-registering images. A challenge for this type of methods is to obtain images from minimal data on which image registration still yields reasonable motion-fields. In Lee (2018) motion was estimated by co-registering corrupted images reconstructed using parallel imaging (Pruessmann *et al* 1999, Griswold *et al* 2002) and compressed sensing (Lustig *et al* 2008). Promising results were presented for rigid motion, but the application to non-rigid motion may be challenging as the undersampling artifacts can result in unrealistic motion-fields. Several methods have been proposed to infer 3D motion from 2D cine-MRI for radiotherapy guidance (Cervino *et al* 2011, Bjerre *et al* 2013, Tryggestad *et al* 2013, Brix *et al* 2014, Seregni *et al* 2016, Stemkens *et al* 2016). Another image-based method is PROMO (White *et al* 2010), which prospectively and in real-time estimates and corrects for rigid motion by utilizing three orthogonal two-dimensional spiral navigator acquisitions and an extended Kalman filter framework. In Glitzner *et al* (2015) 3D motion is estimated from low-resolution 3D images for the purpose of MRgRT, and it is reported that $5 \times 5 \times 5$ mm^3 spatial image resolution is sufficient for motion tracking. Other image-based methods estimate motion and reconstruct motion-compensated images from low resolution images obtained using volumetric navigators (vNAVs) (Usman *et al* 2013, Ingle *et al* 2014).

A different category of motion estimation methods aims at reconstructing motion-fields directly from surrogate signals such as a time series from a respiratory belt (McClelland *et al* 2013, 2017) or the time evolution of noise covariances of an RF coil-array (Andreychenko *et al* 2017). An application of these surrogate signal motion models is the GRICS framework (Odille *et al* 2008a, 2008b), in which non-rigid motion-fields are estimated from respiratory belt signals and used to recover motion-corrected images. Drawbacks of most surrogate signal methods are that the surrogate signals can be of poor quality and require additional hardware to be acquired, and that the methods can only reconstruct low-dimensional motion that is correlated with the input signals (Odille *et al* 2008a, 2008b).

Alternatively, *k*-space methods have been proposed to estimate motion directly from (highly undersampled) *k*-space data. Among these, several methods rely on the acquisition of a navigator signal for the motion estimation, see e.g. Fu *et al* (1995), Welch *et al* (2002), Van der Kouwe *et al* (2006), Pipe (1999), Stam *et al* (2012). These methods are mainly based on the explicit relation between a linear transformation and *k*-space data due to properties of the Fourier transform (Wisetphanichkij and Dejhan 2005) and are therefore limited to affine motion. In Prieto *et al* (2007) non-rigid motion-fields are reconstructed directly from *k*-space data, but the focus is on improving the reconstruction of a dynamic 2D image sequence.

Instead, for applications such as MRgRT, 3D motion-fields by themselves are of particular interest, and should be reconstructed from minimal *k*-space data to allow for online motion characterization. For this purpose, the application of image-based methods can be challenging given their difficulties to cope with minimal *k*-space data. The application of surrogate signal models to MRgRT can be complicated due to the requirement of additional hardware that may not be compatible with the MR-Linac setup. *K*-space methods are therefore most promising for this application. However, most previously proposed *k*-space methods are either limited to rigid motion or were designed to improve image reconstruction.

In this work, we introduce and demonstrate a new *k*-space method to reconstruct non-rigid 3D motion directly from minimal *k*-space data. The method will be referred to as MR-MOTUS, which stands for Model-based Reconstruction of MOTion from Undersampled Signals. In MR-MOTUS, a signal model explicitly relates the *k*-space signal of a deformed object to a reference image and a non-rigid 3D motion-field. Using an *a priori* available reference image, the motion-field can be reconstructed directly from *k*-space data by solving the inverse problem. The availability of a reference image is guaranteed in the MRgRT setting as pre-treatment MR-images are always reconstructed to assess the day-to-day variations in anatomy (Raaymakers *et al* 2017). Additionally, we observe that internal body motion exhibits a high level of spatial correlation; the connectivity and rigidity of tissue enforces similar motion locally. A key idea behind MR-MOTUS is to exploit this correlation by using a low-dimensional motion model to reduce the number of unknowns, such that motion-fields can be reconstructed from a snapshot of *k*-space data that can be acquired in the order of milliseconds.

The rest of this paper is organized as follows. First, the signal model that explicitly relates the snapshot *k*-space data to a motion-field is derived and subsequently validated using a numerical motion phantom. Second, the convergence and robustness of the reconstruction algorithm is assessed by comparison with the ground-truth motion-fields in several *in-silico* scenarios. Subsequently, MR-MOTUS is used to reconstruct *in vivo* 3D rigid head motion and *in vivo* 3D non-rigid respiratory motion from retrospectively highly undersampled *k*-space data using different undersampling strategies. To contextualize these results, MR-MOTUS reconstructions are compared with a reference method: state-of-the-art image registration applied to images reconstructed from the

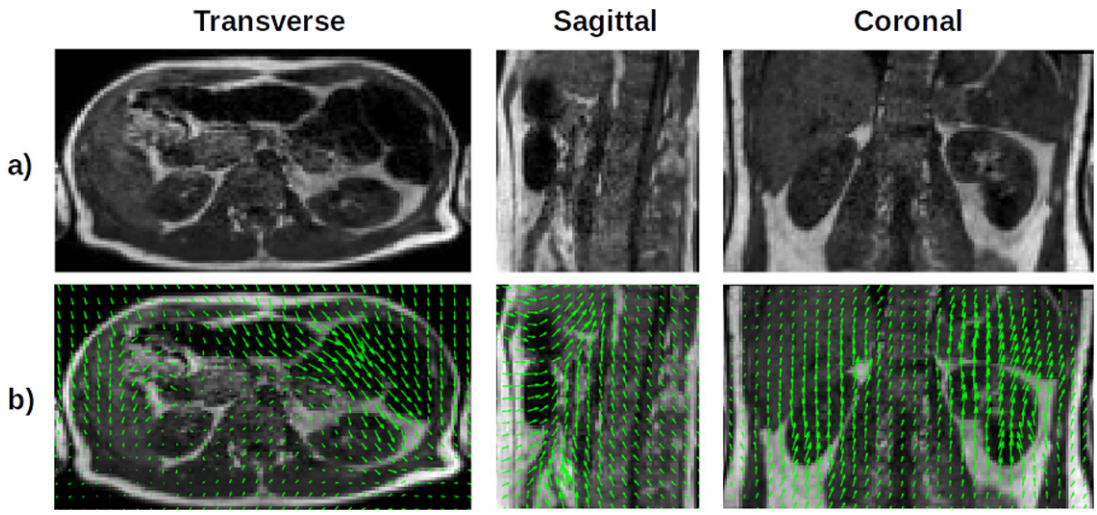

Figure 1. Visualization of the data used in section 2.1: (a) three slices of one of the images used for the registration, and (b) three in-plane projections of the motion-field obtained with optical flow.

same undersampled data. Finally, we demonstrate the potential of MR-MOTUS in practice by reconstructing motion-fields from prospectively undersampled snapshot data consisting of 15 spokes, acquired in 60 ms with a 2D golden angle acquisition during free-breathing.

2. Theory

2.1. Ansatz

Before we introduce the signal model, we illustrate the high compressibility of motion-fields. This is done by approximating the motion-fields with a gradually decreasing number of basis functions from a natural representation basis. Two 3D abdomen scans were acquired during breath-holds in different respiratory phases using a spoiled gradient echo sequence with TR/TE $= 2.30/1.15$ ms, a field of view (FOV) of $0.28 \times 0.34 \times 0.34$ m and a resolution of $3.0 \times 2.7 \times 2.7$ mm. To obtain a motion-field, the two images were registered using state-of-the-art optical flow software (Zachiu *et al* 2015a, 2015b). One of the images and the obtained motion-field are shown in figure 1.

Next, a cubic B-spline basis (Rueckert *et al* 1999) was chosen as the natural representation basis and all components of the motion-field were represented separately, i.e. left–right (LR), feet-head (FH) and anterior–posterior (AP). The maximum normalized root mean square error[2] (NRMSE) of approximation over all three components (LR, FH, AP) was computed at several compression ratios. Here we have defined the compression ratio as the ratio between the number of voxels in the motion-field and the number of basis functions. The compression curves for all components are shown in figure 2. Note that the LR component gave the highest representation error at all compression ratios. A maximum representation error of only 10% is made for all three components with 100 times as few approximation coefficients, which shows that the motion-fields are indeed very compressible.

2.2. Signal model derivation

2.2.1. *Outline of the derivation*

Let $q_t(\mathbf{r}) \in \mathbb{C}$ denote the transverse magnetization of a deforming object at time t and spatial coordinate $\mathbf{r} = (x, y, z)$. The k-space signal from q_t at coordinate $\mathbf{k} = (k_x, k_y, k_z)$ can then be modeled as

$$s_t(\mathbf{k}) = \int_\Omega q_t(\mathbf{r}) e^{-i2\pi \mathbf{k}\cdot\mathbf{r}} \, \mathrm{d}\mathbf{r}. \tag{1}$$

Here Ω denotes the spatially excited FOV. Let $\mathbf{U}_t : \mathbb{R}^3 \mapsto \mathbb{R}^3$ denote the motion-field that deforms q_0 to q_t:

$$\mathbf{U}_t(\mathbf{r}) = \mathbf{r} + \boldsymbol{\delta}_t(\mathbf{r}), \tag{2}$$

with displacement function $\boldsymbol{\delta}_t : \mathbb{R}^3 \mapsto \mathbb{R}^3$. We assume that q_t, can be approximated by warping a reference object q_0

[2] In this work the NRMSE e between a vector $\mathbf{a}$ and a target vector $\mathbf{b}$ is defined as $e = \frac{\|\mathbf{a}-\mathbf{b}\|}{\|\mathbf{b}\|}$.

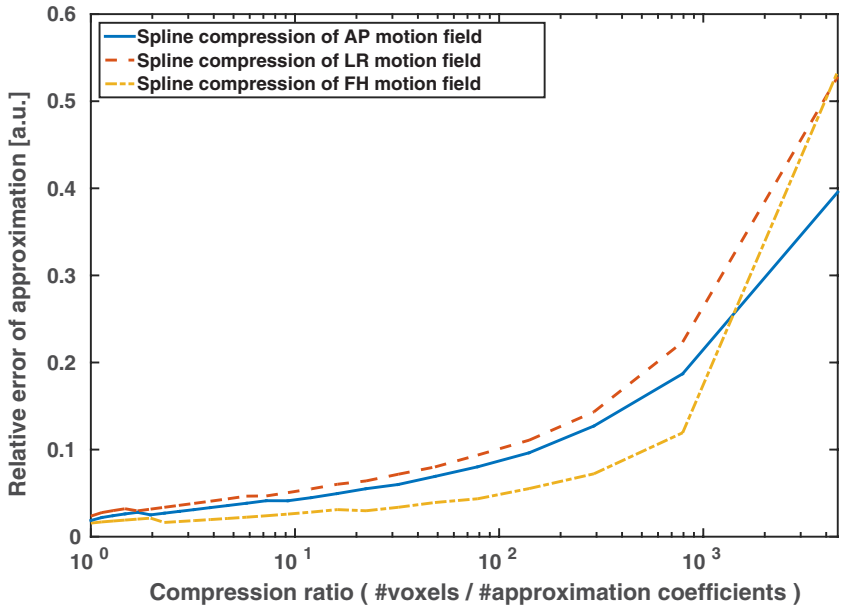

Figure 2. Compression ratios for the three components of the motion-field of respiratory motion in a cubic B-spline basis, as described in section 2.1. The curves show that the motion-fields are indeed very compressible: a maximum representation error of only 10% is made for all three components with 100 times as little approximation coefficients.

$$q_t(\mathbf{r})\mathrm{d}\mathbf{r} = q_0\left(\mathbf{U}_t(\mathbf{r})\right)\left|\det(\nabla\mathbf{U}_t)(\mathbf{r})\right|\mathrm{d}\mathbf{r}, \tag{3}$$

where $\nabla\mathbf{U}_t$ denotes the Jacobian of the motion-field $\mathbf{U}_t$. In the next two subsection we show under which assumptions this approximation is exact. In the rest of this work we will refer to q_0 as the *reference image* and to s_t as the *snapshot signal*. An explicit relation between the reference image, motion-fields, and the snapshot signal of the deforming object can be obtained by substituting (3) into (1), followed by a change of coordinates:

$$s_t(\mathbf{k}) = \int_\Omega q_0(\mathbf{r}_0)e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0)}\,\mathrm{d}\mathbf{r}_0. \tag{4}$$

Here $\mathbf{T}_t$ is defined as the inverse of $\mathbf{U}_t$, such that

$$\mathbf{T}_t(\mathbf{r}) = \mathbf{r} + \boldsymbol{\eta}_t(\mathbf{r}), \quad \mathbf{U}_t \circ \mathbf{T}_t = \mathbf{T}_t \circ \mathbf{U}_t = \mathbf{Id}, \tag{5}$$

where $\boldsymbol{\eta}_t(\mathbf{r})$ is the displacement of $\mathbf{r}$ due to $\mathbf{T}_t$ and $\mathbf{Id}$ denotes the identity operator. It is now evident that if the reference image is available then motion-fields can be estimated by inverting (4) with respect to motion-field.

To investigate under which assumptions (3) is valid we first separate the transverse magnetization q_t into the unit-length transverse magnetization $m_t : \mathbb{R}^3 \mapsto \mathbb{C}$ and spin density $\rho_t : \mathbb{R}^3 \mapsto \mathbb{R}^+$ as

$$q_t \equiv m_t \cdot \rho_t. \tag{6}$$

Next, we will derive temporal relations for m and ρ separately and combine them to obtain (3). Finally, we derive (4) by a substitution and a change of variables.

2.2.2. *Temporal relation for transverse magnetization*

We first derive the temporal relation for the transverse magnetization m. Suppose a steady-state sequence is employed to acquire signal from a static object deformed by dynamic motion-fields $\mathbf{U}_t$. We assume a sufficiently short read-out time, such that spin displacements and transverse and longitudinal decay effects have a noticeable effect only over one or several TR intervals. The transverse magnetization at time t of the spins at location $\mathbf{r}_t$ can then be written as the transverse magnetization at time 0 of the same spin before deformation by $\mathbf{U}_t$:

$$m_t(\mathbf{r}_t) = m_0\left(\mathbf{U}_t(\mathbf{r}_t)\right). \quad \text{(Steady-state condition)} \tag{7}$$

Note that for (7) to hold it must be assumed that the B_0 and B_1 fields are spatially slowly varying, which is a reasonable assumption at the targeted clinical field strength of 1.5 T. Formally, (7) is exactly fulfilled under a few other technical assumption, and for more mathematical details of the derivation we refer the reader to supplementary information (stacks.iop.org/PMB/65/015004/mmedia) section 1.

2.2.3. Temporal relation for spin density

Next, we derive the equation for the spin density. To be able to describe the complete dynamic image sequence of a deforming object in terms of a static reference image and dynamic motion-fields, it must be assumed that spins are conserved within the FOV. Under this assumption, the total number of spins during every TR remains constant. That is,

$$\int_{\mathbb{R}^3} \rho_t(\mathbf{r}_t)\ \mathrm{d}\mathbf{r}_t = C, \quad t = 0, 1, \ldots,$$

where $C \in \mathbb{R}$ is a constant. Hence, the deformations $\mathbf{U}_t$ must satisfy

$$\int_{\mathbf{U}_t(X)} \rho_0(\mathbf{r}_0)\ \mathrm{d}\mathbf{r}_0 = \int_X \rho_t(\mathbf{r}_t)\ \mathrm{d}\mathbf{r}_t, \tag{8}$$

for all sets $X \subseteq \mathbb{R}^3$. We assume all $\mathbf{U}_t$ are continuously differentiable everywhere and invertible, with inverse $\mathbf{T}_t$ as defined in (5). We can then rewrite the left-hand side of (8) using the change of variables $\mathbf{r}_0 \mapsto \mathbf{U}_t(\mathbf{r}_t)$:

$$\int_{\mathbf{U}_t(X)} \rho_0(\mathbf{r}_0)\ \mathrm{d}\mathbf{r}_0 = \int_X \rho_0\left(\mathbf{U}_t(\mathbf{r}_t)\right) \left|\det(\boldsymbol{\nabla}\mathbf{U}_t)(\mathbf{r}_t)\right|\ \mathrm{d}\mathbf{r}_t. \tag{9}$$

Combining (8) and (9) then yields

$$\int_X \rho_t(\mathbf{r}_t)\ \mathrm{d}\mathbf{r}_t = \int_X \rho_0\left(\mathbf{U}_t(\mathbf{r}_t)\right) \left|\det(\boldsymbol{\nabla}\mathbf{U}_t)(\mathbf{r}_t)\right|\ \mathrm{d}\mathbf{r}_t, \tag{10}$$

for all sets $X \subseteq \mathbb{R}^3$. We conclude that the following must hold for all $t = 0, 1, \ldots$

$$\begin{gathered}\rho_t(\mathbf{r}_t)\mathrm{d}\mathbf{r}_t = \rho_0\left(\mathbf{U}_t(\mathbf{r}_t)\right) \left|\det(\boldsymbol{\nabla}\mathbf{U}_t)(\mathbf{r}_t)\right|\mathrm{d}\mathbf{r}_t. \\ \text{(Local spin conservation)}\end{gathered} \tag{11}$$

2.2.4. Derivation of the signal model

Combining (7) and (11) yields the previously described temporal relation (3) between the reference object, deforming object and motion-fields:

$$q_t(\mathbf{r})\mathrm{d}\mathbf{r} = q_0\left(\mathbf{U}_t(\mathbf{r})\right) \left|\det(\boldsymbol{\nabla}\mathbf{U}_t)(\mathbf{r})\right|\mathrm{d}\mathbf{r}. \tag{12}$$

Substituting (12) into the signal model (1) then yields

$$s_t(\mathbf{k}) = \int_\Omega q_0\left(\mathbf{U}_t(\mathbf{r}_t)\right) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}_t} \left|\det(\boldsymbol{\nabla}\mathbf{U}_t)(\mathbf{r}_t)\right|\ \mathrm{d}\mathbf{r}_t.$$

By the inverse function theorem the determinant of the inverse is the inverse of the determinant, hence after the change of variables $\mathbf{r}_t \mapsto \mathbf{T}_t(\mathbf{r}_0)$ we obtain

$$s_t(\mathbf{k}) = \int_{\mathbf{U}_t(\Omega)} q_0(\mathbf{r}_0) e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0)}\ \mathrm{d}\mathbf{r}_0. \tag{13}$$

Note that here we have used the inverse property of $\mathbf{T}_t$, i.e. $\mathbf{U}_t \circ \mathbf{T}_t = \mathbf{Id}$. The domain of integration in (13) depends on the (unknown) motion-field $\mathbf{U}_t$, which is inconvenient in practice. If it is assumed that no signal-contributing spins flow across the boundary of the FOV during deformation, then the integration domain can be changed to Ω to obtain the final signal model:

$$s_t(\mathbf{k}) = \int_\Omega q_0(\mathbf{r}_0) e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0)}\ \mathrm{d}\mathbf{r}_0. \tag{14}$$

For a more formal derivation of the statements above we refer the reader to the supplementary information section 1.

2.3. Inverse problem formulation

Note that (14) explicitly relates the k-space signal to a reference object through non-rigid/non-linear motion-fields $\mathbf{T}_t$. If a reference image q_0 and snapshot data $\mathbf{s}_t$ are available, then motion can be estimated by solving the inverse problem corresponding to (14). In order to exploit the compressibility of motion-fields, we represent them in a lower-dimensional basis using coefficients $\boldsymbol{\theta}_t \in \mathbb{R}^{N_c}$. We typically have $N_c \ll 3N$, where N is the number of voxels per motion-field. Equation (14) can then be rewritten in operator form as

$$\mathbf{s}_t = \mathbf{F}(\boldsymbol{\theta}_t | q_0), \tag{15}$$

where $\mathbf{F}(\boldsymbol{\theta}_t|q_0)$ is the vectorization over k-space coordinates of

$$F(\boldsymbol{\theta}_t|q_0)[\mathbf{k}] = \int_\Omega q_0(\mathbf{r}_0)e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0|\boldsymbol{\theta}_t)}\ \mathrm{d}\mathbf{r}_0,$$

and $\mathbf{s}_t$ is the vectorized k-space signal of the deforming object at time t. In the rest of this work we drop the dependency of $\mathbf{F}$ on q_0 for ease of notation, and because it is assumed to be known. Note that this forward model can be evaluated efficiently as a type-3 non-uniform fast Fourier transform (NUFFT) (Lee and Greengard 2005, Barnett *et al* 2018). To reconstruct motion-fields the following minimization problem is solved:

$$\min_{\boldsymbol{\theta}_t}\ \|\mathbf{F}(\boldsymbol{\theta}_t) - \mathbf{s}_t\|_2^2 + \lambda\mathcal{R}(\mathbf{T}_t(\cdot|\boldsymbol{\theta}_t)), \tag{16}$$

where $\mathcal{R}$ is a regularizer that models *a priori* knowledge on motion-fields, and $\lambda \in \mathbb{R}^+$ is the corresponding regularization coefficient that balances the objective function between a data-fit and being consistent with the *a priori* assumptions.

2.4. MR-MOTUS: model-based reconstruction of motion-fields from undersampled signals

2.4.1. Regularization functional

A natural choice for $\mathcal{R}$ in this setting, which was originally proposed in Fischer and Modersitzki (2003), is to assume smooth motion-fields by penalizing the spatial curvature of the motion-fields:

$$\mathcal{R}(\mathbf{T}_t(\cdot|\boldsymbol{\theta}_t)) := \sum_{p\in\{x,y,z\}}\int_\Omega |\Delta T_t^p(\mathbf{r}|\boldsymbol{\theta}_t^p)|^2\ \mathrm{d}\mathbf{r}. \tag{17}$$

Here Δ denotes the Laplace operator, and $T_t^x, T_t^y, T_t^z : \mathbb{R}^3 \mapsto \mathbb{R}$ denote the individual components of the motion-field $\mathbf{T}_t$. With this prior we obtain the following minimization problem to reconstruct the motion model parameters $\boldsymbol{\theta}_t = \{\boldsymbol{\theta}_t^x, \boldsymbol{\theta}_t^y, \boldsymbol{\theta}_t^z\}$:

$$\min_{\boldsymbol{\theta}_t}\ \|\mathbf{F}(\boldsymbol{\theta}_t) - \mathbf{s}_t\|_2^2\ +\ \lambda\sum_{p\in\{x,y,z\}}\int_\Omega |\Delta T_t^p(\mathbf{r}|\boldsymbol{\theta}_t^p)|^2\ \mathrm{d}\mathbf{r}. \tag{18}$$

2.4.2. Motion models

In this work two motion models are considered: 3D affine transformations and free-form deformations (FFD) parameterized using cubic B-splines (Rueckert *et al* 1999). The 3D affine transformation is defined as

$$\mathbf{T}^{\mathrm{aff}}(\mathbf{r}|\mathbf{A},\mathbf{v}) = \mathbf{A}\mathbf{r} + \mathbf{v}, \tag{19}$$

where $\mathbf{A} \in \mathbb{R}^{3\times3}$ is the affine matrix and $\mathbf{v} \in \mathbb{R}^{3\times1}$ is the shift vector. This results in $3 \times 3 + 3 = 12$ parameters. The free-form deformation is defined as

$$\mathbf{T}^{\mathrm{FFD}}(\mathbf{r}|\mathbf{c}^x,\mathbf{c}^y,\mathbf{c}^z) = \mathbf{r} + \begin{pmatrix}\mathbf{b}^x(\mathbf{r})\mathbf{c}^x\\ \mathbf{b}^y(\mathbf{r})\mathbf{c}^y\\ \mathbf{b}^z(\mathbf{r})\mathbf{c}^z\end{pmatrix}, \tag{20}$$

where $\mathbf{b}^p(\mathbf{r}) \in \mathbb{R}^{1\times N_c^p}$ are the row-vectors with N_c^p 3D B-spline basis functions, evaluated at the coordinate $\mathbf{r}$, and $\mathbf{c}^p \in \mathbb{R}^{N_c^p\times1}$ denote expansion coefficients. The 3D basis functions are constructed as a Kronecker product of three 1D bases with a spline order (i.e. the number of basis functions) of S each. If we use the same basis for all three components of the motion-field, then the total number of coefficients for the spline model is $N_c = \sum_p N_c^p = 3S^3$. In practice this usually implies $N_c \approx \mathcal{O}(10^4)$ for a $100 \times 100 \times 100$ motion-field. In contrast, reconstruction of a $100 \times 100 \times 100$ image has a total number of unknowns in $\mathcal{O}(10^6)$, which is two orders of magnitude higher than the spline model.

2.4.3. Optimization

Solving the optimization problem in (18) is challenging as it is both non-convex and non-linear. Nevertheless, various algorithms exist to tackle problems of this type. Most of these are based on Newton's method, where iterations of the form

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \left[\mathbf{HF}\left(\boldsymbol{\theta}^{(j)}\right)\right]^{-1}\boldsymbol{\nabla}\mathbf{F}\left(\boldsymbol{\theta}^{(j)}\right), \quad j \geqslant 0, \tag{21}$$

are performed. Here $\mathbf{H}$ denotes the Hessian, $\boldsymbol{\nabla}$ denotes the gradient, and the superscript (j) denotes the iteration index. In this work the interior-point method was combined with an L-BFGS (Liu and Nocedal 1989) Hessian approximation. The whole reconstruction pipeline was implemented in MATLAB 2015. Details on the gradients of the cost-function and other aspects of the optimization can be found in supplementary information

section 2. Several stopping criteria were tested, but fixing the number of iterations to 30 provided the most robust reconstructions in all experiments. The regularization parameter λ was optimized by grid-search for all experiments.

2.4.4. Inversion of the reconstructed motion-fields

Once the optimization algorithm has converged, the representation coefficients of the motion-fields $\mathbf{T}(\mathbf{r})$ are available. These are, however, the inverse motion-fields of $\mathbf{U}(\mathbf{r})$ (see (5)) which warp the reference image q_0 to the dynamic object through (12). Following (Chen *et al* 2008), the inverse property of $\mathbf{T}$ can be rewritten as a relation between the displacements:

$$\begin{aligned} \mathbf{r} &= \mathbf{T}(\mathbf{U}(\mathbf{r})) \\ &= \mathbf{U}(\mathbf{r}) + \boldsymbol{\eta}(\mathbf{U}(\mathbf{r})) \\ &= \mathbf{r} + \boldsymbol{\delta}(\mathbf{r}) + \boldsymbol{\eta}(\mathbf{r} + \boldsymbol{\delta}(\mathbf{r})) \\ \Rightarrow \quad \boldsymbol{\delta}(\mathbf{r}) &= -\boldsymbol{\eta}(\mathbf{r} + \boldsymbol{\delta}(\mathbf{r})). \end{aligned}$$

We follow (Chen *et al* 2008) again and perform fixed-point iterations to compute $\boldsymbol{\delta}$:

$$\boldsymbol{\delta}^{(0)}(\mathbf{r}) = 0, \tag{22}$$

$$\boldsymbol{\delta}^{(j)}(\mathbf{r}) = -\boldsymbol{\eta}\left(\mathbf{r} + \boldsymbol{\delta}^{(j-1)}(\mathbf{r})\right), \quad j \in \mathbb{N}. \tag{23}$$

Here the superscript (j) denotes the fixed-point iteration index. Note that one iteration results in the naive inversion $\boldsymbol{\delta}(\mathbf{r}) = -\boldsymbol{\eta}(\mathbf{r})$, which will only be reasonable for very small deformations. The required interpolation that is performed is based on cubic splines, as provided by the `interp3` function of MATLAB 2015. The fixed-point iterations (22) and (23) were terminated whenever the relative changes between two consecutive iterations, summed over all motion-fields (x, y, z), was lower than 0.1. This resulted in about 5–10 iterations in practice.

3. Methods

3.1. Signal model validation

A numerical phantom with known ground-truth forward and inverse motion-fields was used to validate the proposed signal model (14). The phantom consists of a main spherical compartment, filled with three smaller disjoint elliptical compartments. The forward ($\mathbf{U}_t^{\text{ph}}$) and inverse ($\mathbf{T}_t^{\text{ph}}$) analytical motion-fields were designed such that $\mathbf{U}_t^{\text{ph}} \circ \mathbf{T}_t^{\text{ph}} = \mathbf{T}_t^{\text{ph}} \circ \mathbf{U}_t^{\text{ph}} = \mathbf{Id}$. The ground-truth motion-fields were modeled as quadratic functions for the x and z directions, and linear for the y direction. A visualization of the phantom and a complete analytic description of the phantom and the motion-fields can be found in supplementary information section 3.

The phantom was discretized on a $120 \times 120 \times 120$ grid and the spatial FOV was set to $36 \times 36 \times 36\,\text{cm}^3$. This resulted in a spatial resolution of $3.3 \times 3.3 \times 3.3\,\text{mm}^3$.

To validate the signal model, the phantom is defined as the reference image q_0^{ph} and it is deformed to q_1^{ph} by applying the ground-truth motion-field $\mathbf{U}_1^{\text{ph}}$ using (12). The signal from the deformed phantom was computed from (1) and compared with the signal obtained from the forward model (14) using the ground-truth motion-fields $\mathbf{T}_1^{\text{ph}}$ and the phantom before deformation. As the k-space trajectory we have chosen the first 80 read-out samples from a single cone interleave taken from a 3D golden mean cone trajectory (Johnson 2017) (see figure 4 for a visualization). This trajectory will in practice be relatively insensitive to intra-acquisition motion (Block *et al* 2014) and is therefore a good candidate to acquire the snapshot data. For this reason, this trajectory will also be used later on to simulate snapshot data.

3.2. Analysis of spin flow assumption

To derive the final signal model (14) it is assumed that signal-contributing spins do not cross the boundary of the FOV during deformation. To analyze the impact of this assumption in a non-conservative spin system, we performed two motion reconstruction experiments on the analytical phantom for which ground-truth motion-fields are available (see section 3.1 and supplementary information section 3). The first experiment considered the scenario of approximately 15% spin out-flow, and the second considered 15% spin in-flow. Spin out-flow was realized by cropping the FOV such that the reference image is completely contained in the FOV but part of the phantom has left the FOV between reference and snapshot acquisition. The directions were reversed to realize spin in-flow. Motion-fields were modeled with a cubic B-spline model (20) with a spline order of 3 in all directions. The k-space data used for the reconstruction was computed on the Cartesian phantom grid and no undersampling was performed.

3.3. Motion estimation on retrospectively undersampled data

We have performed motion estimation experiments that build up in complexity of the reconstructions. We start in this section with *in-silico* motion estimation and then proceed to *in vivo* rigid and non-rigid motion estimation from retrospectively undersampled data. Finally, we perform in the next section *in vivo* non-rigid motion estimation from prospectively undersampled data acquired during free-breathing. The details of the experiments on the retrospectively undersampled data are discussed below.

3.3.1. *In-silico motion estimation*

In order to validate the proposed MR-MOTUS framework, motion-fields were reconstructed with snapshot data generated from the deforming analytical phantom as described in section 3.1 and compared with the analytic ground-truth motion-fields $\mathbf{T}_t^{\text{ph}}$. The maximum displacement in this case was 15 mm. In a practical setting, snapshot data required for motion estimation will be acquired as fast as possible using efficient non-Cartesian read-outs such as cones or spirals. To mimic this acquisition, a 3D golden mean cone trajectory (see figure 4 for a visualization) was projected on the Cartesian grid with a nearest neighbour interpolation, resulting in a pseudo-cone trajectory. The snapshot data was simulated on the Cartesian points. The cubic B-spline motion model (20) was employed with a spline order of 3 in all directions. Different undersampling factors were realized by varying the number of cones and the maximum k-space coordinate. The sensitivity to noise was assessed by comparing the reconstructions with noiseless data to reconstructions with noisy data. In the latter case, complex Gaussian noise was added to obtain a signal to noise ratio (SNR) of approximately 80.

3.3.2. *In vivo motion estimation*

Additionally, motion-fields were reconstructed from *in vivo* head (rigid) and *in vivo* abdomen data (non-rigid). In contrast with the *in-silico* setting, no ground-truth motion-fields are available. In order to assess the quality of the reconstructed motion-fields in this *in vivo* setting, reconstructions were performed on retrospectively undersampled snapshot data generated from fully-sampled reconstructions and the fully-sampled images were compared with the reference image warped using the reconstructed motion-fields. More specifically, we employed the following pipeline for these reconstructions:

1. Reconstruct the motion-fields $\mathbf{T}_t$;
2. Invert the motion-fields using fixed-point iterations (22) and (23) to obtain $\mathbf{U}_t$;
3. Warp the fully-sampled reference image $\mathbf{q}_0$ using (12) with the inverted motion-fields;
4. Calculate the NRMSE between the fully-sampled warped reference image and the fully-sampled ground-truth image $\mathbf{q}_t$.

See figure 3 for an overview of this workflow. The 3D affine motion model (19) was employed for the rigid head motion estimation, and the cubic B-spline motion model (20) with spline order 16 in all directions was employed for the non-rigid abdomen motion estimation. The regularization parameter was optimized in all experiments with a grid-search. The fixed-point iterations (22) and (23) were terminated whenever the relative changes between two consecutive iterations, summed over all motion-fields (x, y, z), was lower than 0.1. This resulted in about 5–10 iterations in practice. The warping of the reference image was performed using an interpolation based on a cubic kernel, as provided in MATLAB's `interp3` function. The NRMSE was computed on the complex images since both the warped reference image and the fully-sampled reconstruction are complex-valued.

To obtain the fully-sampled reconstructions, a healthy volunteer was scanned with a 3D spoiled gradient echo (SPGRE) sequence preceded by 200 dummy pulses to reach the required steady-state transverse magnetization. The volunteer was instructed to hold still during the scans to reduce effects of motion on the validation dataset. The volunteer gave written informed consent, all experiments were approved by the institutional review board of the University Medical Center Utrecht, and carried out in accordance with the relevant guidelines and regulations. Both the head and abdomen data were acquired with a 16-channel anterior coil that was elevated above the volunteer using a plastic bridge. The multi-channel data was reconstructed to a single-channel image using the coil sensitivities. Sequence parameters and other details of both *in vivo* experiments can be found in table 1.

3.3.3. *Quality assessment on the in vivo reconstructions*

Robustness of reconstruction

To assess the robustness of the reconstruction, multiple motion-field reconstructions were performed using the same reference image but different snapshot data acquired at different motion states. A total of seven independent reconstructions were performed for the head data, and four independent reconstructions for the abdomen data. The volunteer was instructed to hold still during the scans to reduce the effects of motion on the validation dataset, and move to the next motion state in between the scans. One fully-sampled reconstruction

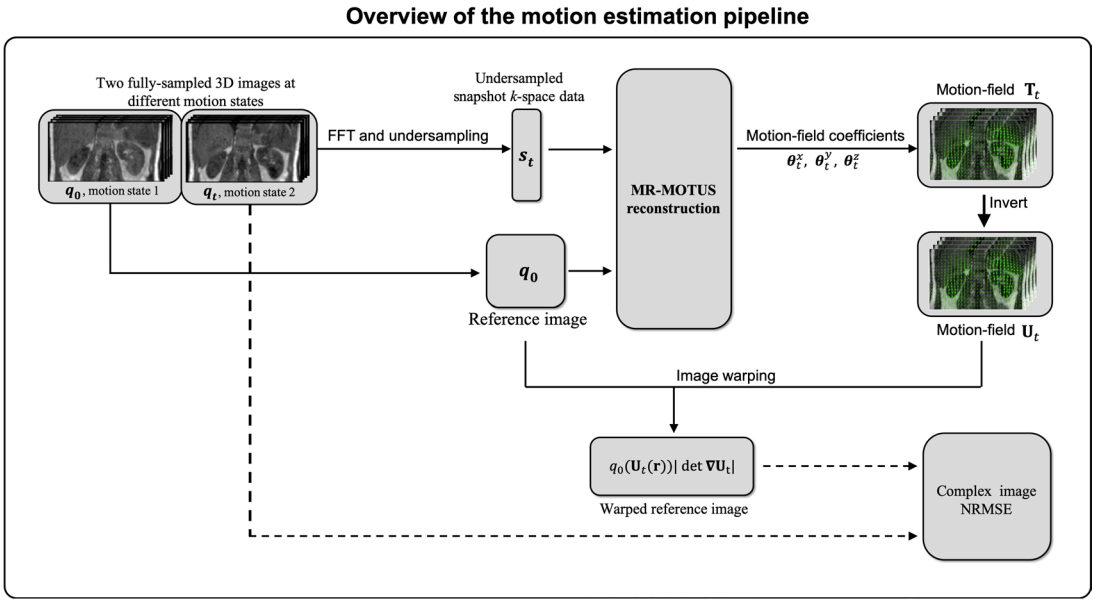

Figure 3. An overview of the motion estimation pipeline described in section 3.3.2. One of two fully-sampled images, each acquired at a different motion state (e.g. different position in the respiratory cycle), is transformed to k-space and retrospectively undersampled to simulate the snapshot k-space data. The other is used as reference image. The reference image and the snapshot k-space data are input into the MR-MOTUS reconstruction which returns motion-field coefficients that are in turn transformed into a motion-field $\mathbf{T}_t$. This motion-field is inverted to $\mathbf{U}_t$, which is then used to warp the reference image. Finally, the complex image NRMSE is computed between the warped reference image and the fully-sampled q_t to assess the quality of the reconstructed motion-fields.

Table 1. Details of the *in vivo* experiments as described in section 3.3.2.

	Retrospective *in vivo*		Prospective *in vivo*
Parameter	Head	Abdomen	Abdomen
FOV (m)	0.25 × 0.25 × 0.13	0.28 × 0.34 × 0.34	0.48 × 0.48 × 0.010
Acquisition size	144 × 144 × 74	94 × 128 × 128	160 × 160 × 1
Spatial resolution (mm)	1.74 × 1.74 × 1.81	3.00 × 2.70 × 2.70	3.00 × 3.00 × 10.00
Repetition time (ms)	8.00	2.30	4.00
Echo time (ms)	3.00	1.20	1.85
Flip angle (°)	16	20	20
Trajectory	Cartesian	Cartesian	Radial
Dummy pulses	200	200	200
Pulse sequence	3D SPGRE	3D SPGRE	2D golden angle
Scanner	Philips Ingenia 1.5 T	Philips Ingenia 1.5 T	Philips Ingenia 1.5 T
Motion model	Affine	Cubic B-Splines	Cubic B-Splines

was used as reference image, and the snapshot k-space were retrospectively downsampled from different fully-sampled images using the 3D Cartesian approach described above with different undersampling factors. All reconstructions for the different motion states were performed independently, i.e. no correlation in time was exploited. The corresponding warped reference images were stacked into a dynamic image sequence to assess the robustness and quality of the reconstructions.

Comparison of reconstructed motion-fields with a reference method

We contextualize the MR-MOTUS motion reconstructions by comparison with a reference method. For this reference method, images were reconstructed from the undersampled snapshot data and registered with optical flow image registration software (Zachiu *et al* 2015a, 2015b). We refer to Zachiu *et al* 2015a, 2015b for more details on the optical flow method. To obtain the images, zero-filling reconstructions were performed on the 3D Cartesian downsampled data, and compressed sensing reconstruction were performed for all other undersampling patterns (see section 3.3.4) using the BART toolbox (Uecker *et al* 2019). L1-wavelet regularization was applied in the BART toolbox, and the regularization parameter was optimized with respect to the final image NRMSE. We will refer to this reference method as image reconstruction + optical flow, i.e. IR +OF. Besides the snapshot images, the reference image was also downsampled and reconstructed in the same way as the snapshot

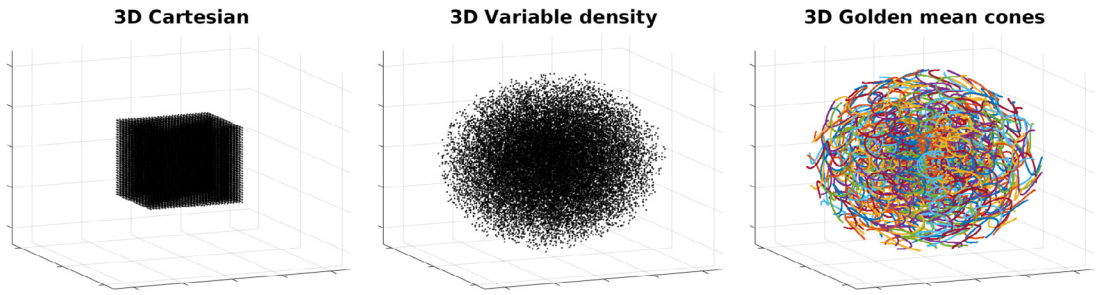

Figure 4. Visualization of the undersampling trajectories with undersampling factor 63 for the three strategies discussed in section 3.3.4 and reported in table 4.

images for the image registration. The same pipeline as described above was employed for both MR-MOTUS and IR +OF to compute the final image NRMSE. Although the reference image was downsampled for image registration, the fully-sampled reference image was used to compute final image NRMSE. It has to be noted that additional errors could arise from the interpolation that is required for the warping, but since we use the same pipeline for IR +OF and MR-MOTUS these errors will be similar and not favor one method over the other.

3.3.4. *Retrospective undersampling trajectories*

For the head motion reconstruction we have considered 3D Cartesian downsampling as a first test scenario. Since respiratory motion estimation is of particular interest for MRgRT, three different types of undersampling were considered for the respiratory motion snapshot data: 3D Cartesian, 3D random variable density, 3D golden mean cones (Johnson 2017) (see figure 4 for a visualization).

For 3D Cartesian downsampling the k-space was symmetrically truncated in all directions to realize the downsampling. The same downsampling factor was applied along each dimension, e.g. for downsampling of 64x in 3D, every direction is downsampled by a factor of 4 by symmetric truncation. This effectively reduces the spatial resolution of the data. Different downsampling factors were realized by varying the size of the k-space truncation window. Note that downsampling was also performed in the read-out direction, please see section 3.3.5 for the motivation.

A 3D variable density undersampling scheme was included to facilitate an undersampling that is ideal for compressed sensing (Chauffert *et al* 2013). The power in the density decay was optimized w.r.t. the final NRMSE in the IR +OF approach, resulting in a probability density of

$$P(k_x, k_y, k_z) = \left(1 - \sqrt{k_x^2 + k_y^2 + k_z^2}\right)^2.$$

The density was scaled w.r.t. the $k_{\max}$ values of the original fully sampled k-space. Different undersampling factors were realized by reducing the number of samples.

We have also considered a 3D golden mean cone trajectory (Johnson 2017) that can efficiently acquire variable density undersampled data by sampling non-Cartesian cone read-outs. Additionally, this trajectory is most interesting in practice, since it is relatively insensitive to intra-acquisition motion (Block *et al* 2014). The number of samples on a cone read-out was fixed to 1000 and the different undersampling factors were realized by varying the number of cones and the maximum k-space coordinate. Since only Cartesian data was available, we have projected the non-Cartesian trajectory onto the fully-sampled Cartesian grid with a nearest neighbour interpolation. Overlapping points after the projection were only used once.

3.3.5. *Computation of the undersampling factor*

In order to compare different undersampling strategies and validate the reconstructions on fully-sampled images, the undersampled snapshot k-space data used for the reconstructions described in this section was retrospectively generated from fully-sampled Cartesian acquisitions. To compare the potential performance of the undersampling strategies in practice two aspects should be considered: the quality of the reconstructed motion-fields and the acquisition efficiency. The reconstruction quality is reported with the image NRMSE between the reference image warped with the reconstructed motion-field and the ground-truth image. In practice the snapshot data will be acquired as fast as possible to minimize the latency of the motion estimation. We therefore envision this framework with fast and dedicated non-Cartesian acquisitions, where reducing the acquired data in any direction could directly reduce the acquisition time. To reflect the differences in acquisition efficiency of the different undersampling strategies, the undersampling factor reported in this work is chosen as the total undersampling factor of the data, including the read-out direction. The undersampling factor thus relates to the amount of data used for the reconstructions and it is calculated as the number of k-space points on

the undersampled Cartesian grid divided by the number of samples on the fully-sampled Cartesian grid. This definition of the undersampling factor allows for the fairest comparison of the performance of the proposed method on the different undersampling strategies considered in this work. Note that *acceleration factor* is different than *undersampling factor* for Cartesian acquisitions, since the read-out undersampling does not lead to accelerated acquisitions. However, for the envisioned non-Cartesian acquisition, the undersampling is more closely related to the acceleration.

3.4. Motion estimation on prospectively undersampled *in vivo* data

To demonstrate the practical feasibility of MR-MOTUS, respiratory motion-fields were additionally reconstructed from prospectively undersampled data acquired during free-breathing. Data was continuously acquired from a volunteer using a 2D golden angle radial acquisition with a total scan time of 15.84 s. Sequence parameters and other details can be found in table 1. The volunteer gave written informed consent, all experiments were approved by the institutional review board of the University Medical Center Utrecht, and carried out in accordance with the relevant guidelines and regulations. Only the body coil was used to obtain single-channel data with nearly homogeneous sensitivities. The scan consisted of two phases: it started with a breath-hold phase of 3.17 s and ended with a free-breathing phase of 12.67 s. In the breath-hold phase 792 spokes were acquired and these were used to reconstruct a reference image by NUFFT. From the free-breathing phase, 15 consecutive spokes were selected after 8.67 s of free-breathing and these were used as snapshot data. The cubic B-spline model (20) with order 40 in every direction was employed, and motion-fields were reconstructed from the reference image and snapshot data by solving (18) and subsequently applying (22) and (23). The number of reconstruction iterations was fixed to 30 and the total reconstruction time was about 6 s. Similar to the other reconstructions presented in this work, we have employed the curvature regularization (17). The effect of this regularization was analyzed by comparing reconstructions with different degrees of regularization. The best parameter was selected by visual inspection of the reconstructed motion-fields.

Finally, the quality of the reconstruction was assessed by warping the reference image with the reconstructed motion-field and comparing the result with the ground-truth image at the snapshot acquisition time. This ground-truth image was obtained from a 2D + t compressed sensing reconstruction using the data acquired in the free-breathing phase. Every 31 consecutive spokes were binned as a dynamic (without overlap), which resulted in a temporal resolution of 124 ms/frame. The image reconstruction was performed with the BART toolbox (Uecker *et al* 2019) using spatial L1-wavelet and temporal total variation regularization. The data was organized such that the central spoke of the data for one of the dynamics coincides with the central spoke of the snapshot data used for motion reconstruction. This dynamic was finally selected for the comparison with the warped reference image.

4. Results

4.1. Signal model validation

The results of the validation are shown in figure 5. Note that the model's prediction of the signal after deformation is indistinguishable from the true signal after deformation. The minor deviations are likely caused by discretization of the continuous analytical phantom and motion-fields.

4.2. Analysis of spin flow assumption

To analyze the effect of the violation of the spin flow assumption described earlier we have reconstructed motion-fields in case this assumption is not satisfied. In figure 6 the motion-fields reconstructed in this case are qualitatively compared with the ground-truth motion-fields. It can be observed that the reconstructions in the spin out-flow scenario are slightly underestimated by MR-MOTUS, most notably at the boundaries of the FOV. In the spin in-flow scenario there seems little effect on the motion-field overall. Quantitative motion-field error maps for these experiments are presented in the supplementary information, figures 3–8. It should be noted that the scenarios considered in this experiment, where 15% of the total spins flows across the boundary, are extreme and will likely not occur in practice. Nevertheless, the quality of the reconstructed motion-fields is minimally affected. The spin flow assumption is therefore not too restrictive and can be relaxed in realistic scenarios where spins may flow across the boundary.

4.3. Motion estimation on retrospectively undersampled data

4.3.1. In-silico motion estimation

The results on *in-silico* motion reconstruction are presented in table 2. The RMSE is computed for motion-fields in each direction and the NRMSE is computed between the warped reference image and the ground-truth deformed phantom. The reconstructions show that the RMSE for the highest undersampling factors is only slightly larger than the voxel size of 3 mm^3. It can also be observed that the reconstruction are very robust

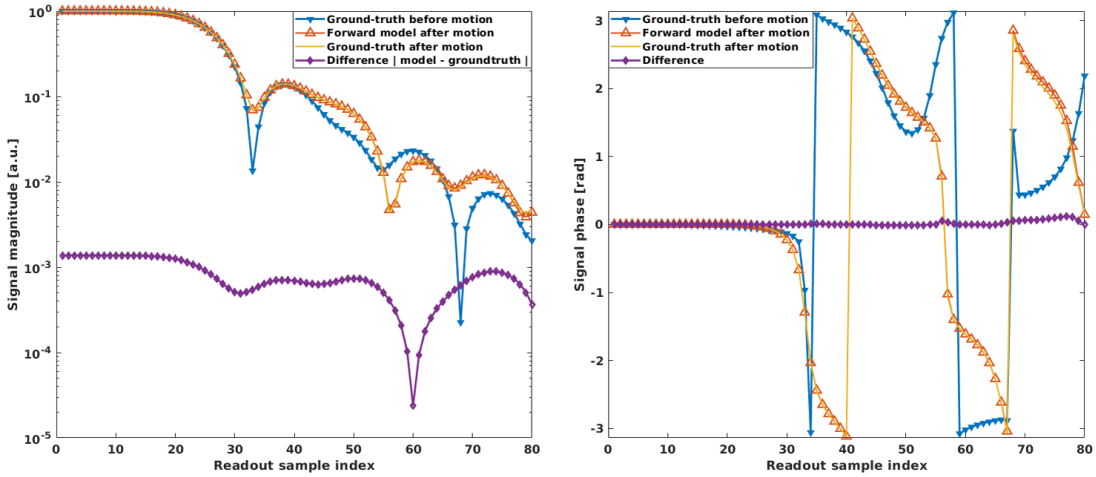

Figure 5. Results of the signal model validation in sections 3.1 and 4.3. Plots of the magnitude (left) and phase (right) of the signal before deformation (blue), true signal after deformation (yellow), the signal calculated with the proposed signal model (14) (red), and the difference between the model's signal and true signal (purple). Note that the difference between the model's signal and the ground-truth signal is about two orders of magnitude lower than the signal. The small deviations between the model and ground-truth are likely caused by discretization errors.

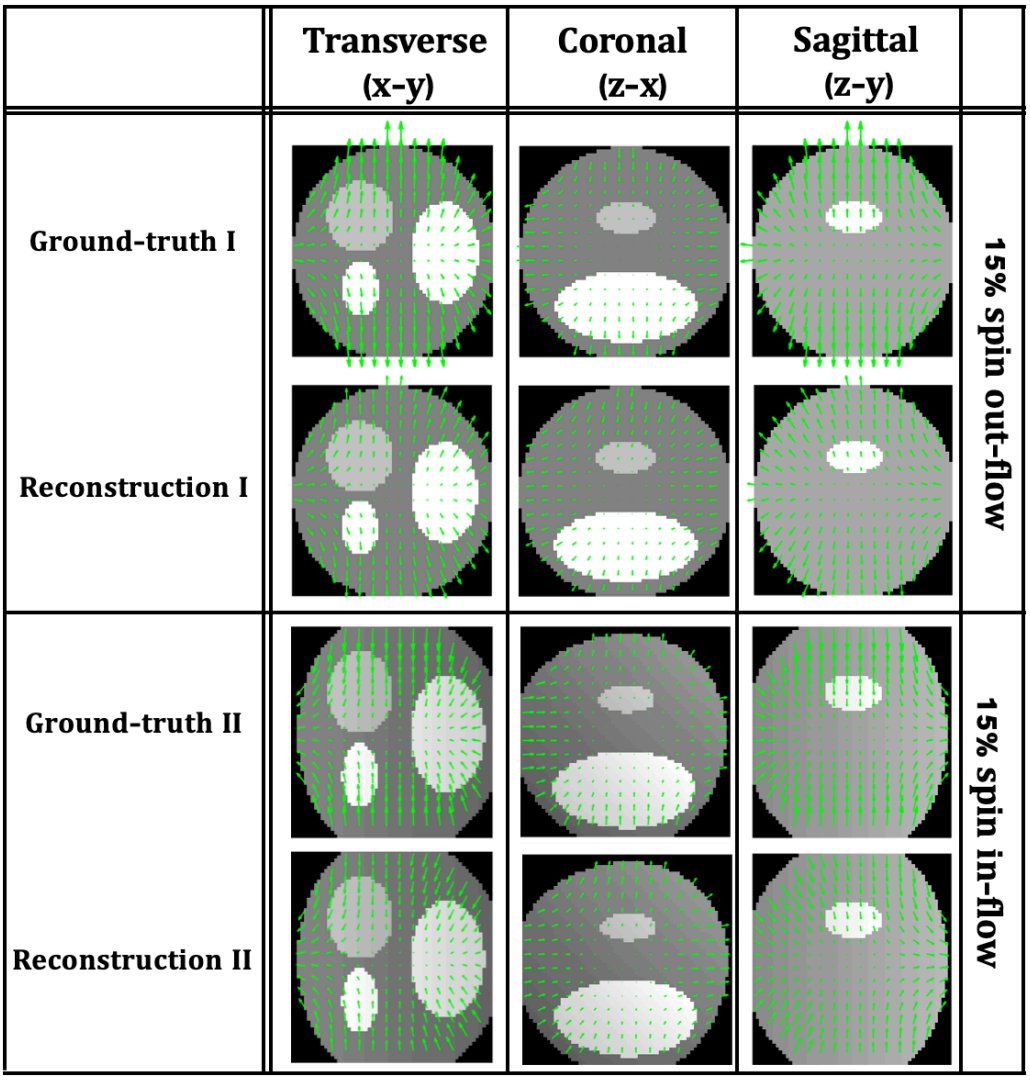

Figure 6. Motion-field reconstructions in case the spin flow assumption is violated, as described in sections 3.2 and 4.2. Two cases are considered: 15% spin out-flow (top) and 15% spin in-flow (bottom).

Table 2. Convergence analysis with analytic motion-fields as discussed in sections 3.3.1 and 4.3.1. The reported numbers are the RMSE between the reconstructed motion-fields and ground-truth motion-fields. The phantom resolution is 3 mm^3 and the maximum ground-truth displacement is 15 mm.

	No noise				Gaussian noise, SNR ≈ 80			
Undersampling factor	1	10	82	558	1	10	82	558
RMSE in x (mm)	1.32	2.65	3.24	3.53	1.34	2.66	3.25	3.54
RMSE in y (mm)	0.75	1.38	1.72	1.84	0.74	1.45	1.74	2.00
RMSE in z (mm)	1.80	2.80	3.21	3.36	1.78	2.77	3.22	3.57
NRMSE images (%)	10.47	12.43	15.02	17.55	10.52	12.66	15.05	17.47

Table 3. Results of the 3D rigid head motion estimation as discussed in sections 3.3.2 and 4.3.2. The reported numbers are the normalized root mean square complex image errors (in percentage) between ground-truth images and reference images warped with motion-fields reconstructed from the retrospectively downsampled data. The downsampling factors 1, 8, 66, 474 and 2551 correspond to 1534 464, 186 624, 3240 and 600 *k*-space points respectively.

	3D Cartesian downsampling				
Downsampling factor	1	8	66	474	2551
MR-MOTUS	**27.33**	28.14	**28.42**	**29.28**	**31.15**
IR + OF	28.39	**27.38**	29.92	37.41	45.20

against noise; there is hardly any increase in RMSE and image NRMSE, and for the highest undersampling factor there is even a decrease in image NRMSE. The latter observation shows the ill-posedness of the motion estimation problem: the image error or objective function residual may decrease while deviating more from the ground-truth motion-fields. For the reconstructions we have applied curvature regularization as described in section 2.4.1 which penalizes second order spatial derivatives. The ground-truth motion-fields in x and z directions are of quadratic nature, so we expect that results may be improved by penalizing higher order spatial derivatives instead. This is also justified by the lower errors in the y directions, for which the ground-truth motion-field is of linear nature.

4.3.2. *In vivo motion estimation*

In vivo head motion estimation

As a first *in vivo* test we applied MR-MOTUS to estimate rigid head motion. A total of seven independent motion-field reconstructions were performed using the same reference image but different snapshot data retrospectively downsampled from fully-sampled reconstructions at the different head positions. The reconstruction time for each reconstruction was about 5 min. In table 3 the image NRMSE between the warped reference images and the ground-truth fully-sampled reconstructions are reported for one of the reconstructions. The results are compared with IR + OF to place them into context.

Additionally, robustness and quality of reconstruction was qualitatively assessed by stacking the warped reference images into a dynamic sequence and comparing this with the ground-truth fully-sampled reconstructions. Note that this dynamic sequence constitutes images obtained with independent reconstructions from independent snapshot data, and are not simultaneously reconstructed. The results of this qualitative comparison are shown in AnimatedFigures 1–3[3].

The quantitative results in table 3 show that MR-MOTUS with 3D Cartesian downsampling with factors of up to 474 performs similarly to optical flow image registration applied to fully-sampled images. It can also be observed that the NRMSE increase due to downsampling is worse for IR + OF than for MR-MOTUS. These results reflect the compressibility of affine motion, which requires only 12 parameters in 3D, that is exploited by MR-MOTUS to reconstruct the motion from minimal *k*-space points. Interestingly, IR + OF outperforms MR-MOTUS for 8-fold downsampling.

The quantitative results in AnimatedFigures 1–3 show that the reconstruction using the same reference image is robust to reconstruct motion between the reference head position and all other head positions. For high downsampling factors of 2551 (AnimatedFigure 3) the reconstructed motion-fields show a shearing, which is unrealistic for this rigid head motion. We expect that results may be improved for higher undersampling factors by adding additional rigidity regularization.

[3] The AnimatedFigures can be found online. See section 4.3.3 for a description of all the animated figures.

Table 4. Results of the 3D respiratory motion estimation as discussed in sections 3.3.2 and 4.3.2. The reported numbers are the normalized root mean square complex image errors (in percentage) between ground-truth images and reference images warped with motion-fields reconstructed from the retrospectively undersampled data.

	3D Cartesian				3D variable density			3D golden mean cones		
Undersampling factor	1	8	63	501	8	63	501	8	63	501
MR-MOTUS	**28.79**	**30.24**	**33.22**	**39.48**	30.70	**32.86**	**36.82**	**30.47**	**33.77**	**37.71**
IR+OF	29.70	30.51	34.26	40.17	**30.33**	35.02	42.71	30.97	37.65	42.97

In vivo respiratory motion estimation

As a second *in vivo* test we applied MR-MOTUS to estimate respiratory motion. A total of four *k*-spaces were acquired at different states in the respiratory cycle. The subject was instructed to move to a different respiratory state in between the acquisitions, and to hold the breath during actual acquisition. Four independent reconstructions were performed on the retrospectively undersampled data, and the fully-sampled reconstructions were used for validation. The reconstruction time for each reconstruction was about 5 min. The reference image was kept fixed but different snapshot data from the different motion states was used for every reconstruction.

Robustness and quality of the reconstructions was assessed by stacking the warped reference images into a dynamic sequence. Note that this dynamic sequence constitutes images obtained with independent reconstructions. The dynamic sequence is compared with the fully-sampled ground-truth reconstructions in the AnimatedFigures 4 and 5[4]. Moreover, the reconstructions for several undersampling factors and undersampling strategies are quantitatively compared with IR+OF in table 4. Additionally, a visual comparison between MR-MOTUS and IR+OF reconstructions is provided in figure 7 for 3D variable density undersampling with an undersampling factor of 63.

Results in table 4 show that in this experiment MR-MOTUS and IR+OF perform similar for undersampling factors 1 and 8, and MR-MOTUS outperforms IR+OF for undersampling factors larger than 8. For higher undersampling factors MR-MOTUS benefits from the compressibility of the motion-fields and the *a priori* information that is available in the form of the reference image. What can also be observed is that MR-MOTUS and IR+OF perform very similar for the 3D Cartesian downsampling, which effectively lowers the resolution of the images. In this case the image reconstruction still yields reasonable images such that motion estimation is feasible. MR-MOTUS performs best with the 3D variable density undersampling with undersampling factors 63 and 501, and best with the 3D Cartesian downsampling with undersampling factor 8. The 3D golden mean cone trajectory can be used to efficiently sample the 3D variable density points in practice, and shows analogous results.

Figure 7 shows a comparison between warped reference images obtained with MR-MOTUS and IR+OF versus the ground-truth. The quality of the IR+OF result for this experiment is low, while the MR-MOTUS result shows better overlap with the ground-truth. The reconstructions with the high undersampling factors may be improved by using different regularization or by using a different undersampling pattern that contains more information about the motion for high undersampling factors.

4.3.3. *Description of the animated figures*

The animated figures are part of the main body of this work and can be found online.

AnimatedFigure1.gif

Affine motion reconstruction from retrospectively downsampled *in vivo* head data using MR-MOTUS, compared with the ground-truth image reconstructions. The 3D Cartesian technique was employed for the downsampling. A downsampling factor of 2.0 was applied in AP and LR direction and 2.1 in FH. This resulted in a total number of 186 624 *k*-space points per snapshot and total downsampling factor of about 8.2.

AnimatedFigure2.gif

Affine motion reconstruction from retrospectively downsampled *in vivo* head data using MR-MOTUS, compared with the ground-truth image reconstructions. The 3D Cartesian technique was employed for the downsampling. A downsampling factor of 8.0 was applied in AP and LR directions and 7.4 in FH. This resulted in a total number of 3240 *k*-space points per snapshot and total downsampling factor of about 474.

AnimatedFigure3.gif

Affine motion reconstruction from retrospectively downsampled *in vivo* head data using MR-MOTUS, compared with the ground-truth image reconstructions. The 3D Cartesian technique was employed for the

[4] See footnote 3.

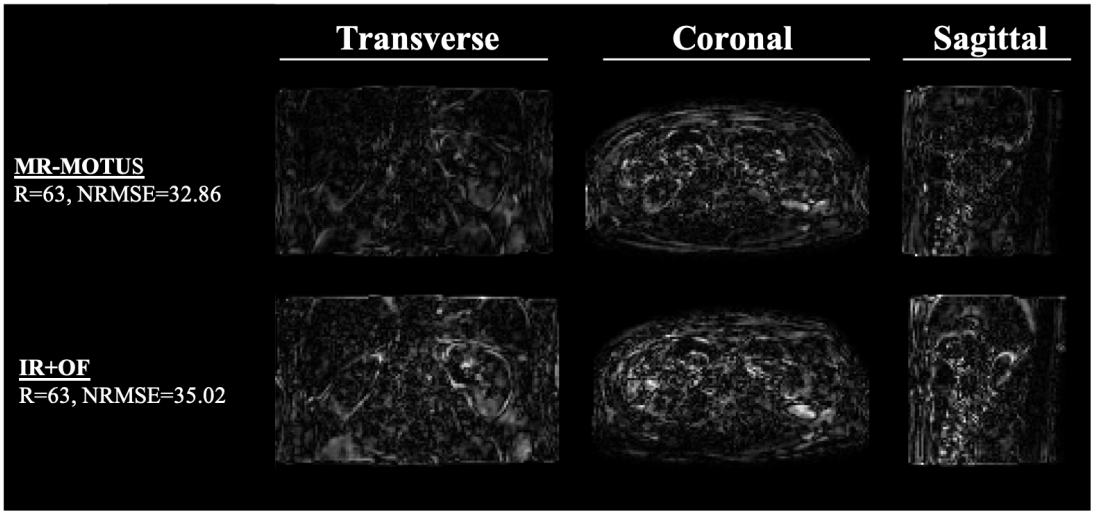

Figure 7. Comparison between MR-MOTUS (top) and IR +OF (bottom) reconstruction quality for a specific case reported in table 4. In this case 3D variable density with an undersampling factor of 63 was applied. The IR +OF clearly shows more quality degrading in comparison with MR-MOTUS.

downsampling. A downsampling factor of 14.4 was applied in AP and LR directions and 12.3 in FH. This resulted in a total number of 600 *k*-space points per snapshot and total downsampling factor of 2551.

AnimatedFigure4.gif
Non-rigid motion reconstruction from *in vivo* abdomen data using MR-MOTUS with a cubic B-spline motion model, compared with ground-truth image reconstructions acquired during breath-hold. No downsampling was applied on these snapshots.

AnimatedFigure5.gif
Non-rigid motion reconstruction from retrospectively downsampled *in vivo* abdomen data using MR-MOTUS with a cubic B-spline motion model, compared with ground-truth image reconstructions acquired during breath-hold. The 3D Cartesian technique was employed for the downsampling. A downsampling factor of about 4 was applied in all direction, resulting in a total downsampling factor of 63.

4.4. Motion estimation on prospectively undersampled *in vivo* data

Figure 8 shows the result of motion-field reconstructions on prospectively undersampled data acquired during free-breathing with three different degrees of regularization: $\lambda = 7 \cdot 10^{-3}, 7 \cdot 10^{-2}, 7 \cdot 10^{-1}$. Little regularization ($\lambda = 7 \cdot 10^{-2}$) results in an unrealistic irregular motion-field, whereas high regularization ($\lambda = 7 \cdot 10^{-1}$) results in a very smooth motion-field. The oversmoothed motion-field does not resolve the natural discontinuous in motion that are present between moving and stationary areas, e.g. between the kidneys and spine.

The quality of the motion-field reconstructed with $\lambda = 7 \cdot 10^{-2}$ is assessed in figure 9, which shows a comparison between the reference image warped with the reconstructed motion-field and the ground-truth image selected from a dynamic CS reconstruction. A high level of overlap can be observed between the warped reference and ground-truth image. This is especially visible at the contours of the liver dome, kidney and lower frontal region of the abdomen, which all move with respiratory motion.

5. Discussion

In this work we introduced MR-MOTUS: a framework for Model-based Reconstructions of MOTion-fields from Undersampled Signals. A signal model is derived that explicitly relates *k*-space data to non-rigid and/or non-linear motion-fields and a reference image. Non-rigid 3D motion-fields are reconstructed from minimal *k*-space data and an *a priori* available reference image by leveraging on the compressibility of internal body motion and by solving the corresponding non-linear inverse problem with respect to the coefficients of a low-dimensional motion model.

The signal model that explicitly relates *k*-space data to non-rigid and/or non-linear motion-fields and a reference image is derived by assuming that the reference image warped by the unknown motion-fields can be used as an approximation to the deforming object, similarly to Odille *et al* (2008a), Prieto *et al* (2007). We formally show

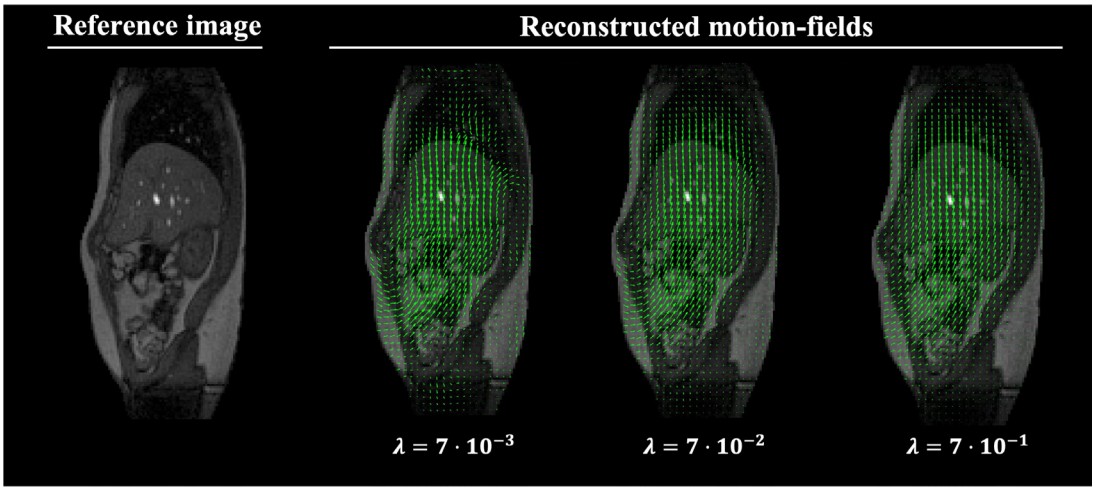

Figure 8. Motion-fields reconstructed from prospectively undersampled data acquired during free breathing, as described in section 3.4. Three different degrees of regularization are considered: $\lambda = 7 \cdot 10^{-3}, 7 \cdot 10^{-2}, 7 \cdot 10^{-1}$. See also figure 9 for a qualitative analysis of the error in the reconstruction.

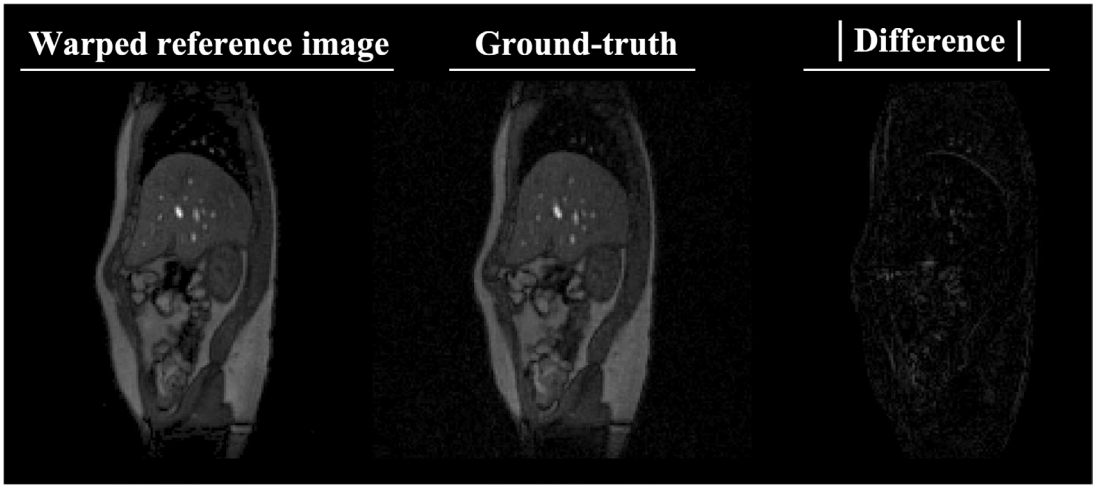

Figure 9. A visual comparison between the reference image warped the MR-MOTUS motion-field reconstructed with $\lambda = 7 \cdot 10^{-2}$, and the ground-truth image selected from a dynamic CS reconstruction, as described in section 3.4.

that this approximation is exact when two conditions are met: (1) no signal-contributing spins flow across the boundary of the FOV, and (2) the transverse magnetization is in a steady-state.

First, the assumption that no signal-contributing spins flow across the boundary of the FOV is made to change the integration domain in (13) to the excited FOV Ω. This assumption also appears in other works for similar reasons (Odille *et al* 2008a). In the *in vivo* 3D cases considered in this work, i.e. head motion and respiratory motion, we have empirically observed that a violation of the spin flow assumption hardly degrades the quality of the motion-fields (see e.g. figure 7). This observation is also supported by the presented *in-silico* phantom experiments in the extreme scenario of 15% spin flow. Although these are extreme cases, which will likely not occur in practice, the quality of the reconstructed motion-fields hardly degraded. This indicates that this assumption can in principle be relaxed.

Second, the assumption of a steady-state transverse magnetization may be partly invalid due to B_0 and B_1 inhomogeneity or temporal B_0 drift. The inhomogeneities and temporal drift may change the equilibrium magnetization—and thereby the steady-state—of moving spins. However, in this proof-of-concept work these effects are assumed to be negligible for the combination of small spin displacements and the targeted field strengths of up to 1.5 T.

For MRgRT, both the acquisition of data and reconstruction of organ motion need to be performed online in the order of milliseconds. Regarding the acquisition, we have investigated Cartesian, variable density and golden mean cones undersampling schemes by means of retrospective undersampling, and we have employed a 2D golden angle radial acquisition for preliminary results with prospective undersampling. The prospectively undersampled snapshot data consisting of 15 spokes were acquired in 60 ms, which shows the potential of the

framework in practice. In future work we plan to extend the prospective acquisitions to 3D by using the golden mean cone acquisition that was considered in this work; it performed similarly to the other schemes, while being time-efficient in acquisition and insensitive to intra-acquisition motion (Block *et al* 2014). Based on the retrospective results shown in this work, we can give an estimate of the required acquisition time for a prospective acquisition with the 3D golden mean cone trajectory and undersampling factor 63. The undersampling factor of 63 corresponds to approximately $(128 \times 128 \times 94)/63 \approx 24\,446$ k-space points with a variable density distribution. The golden mean cone trajectory is designed such that it results in a variable density sampling distribution after any number of read-outs (Johnson 2017). Hence, with a TR of 4ms and 1000 points per read-out, the acquisition time of the snapshot data can be approximated as $24\,446/1000 \cdot 4 = 98$ ms. Future research will address motion reconstructions from prospectively undersampled data using the non-Cartesian 3D golden mean cone trajectory.

Besides acquisition of data, application to MRgRT also requires the actual reconstruction of motion-fields to be in the order of milliseconds. The reconstruction time for the 3D motion estimations considered in this paper is currently about 5 min, but there is still room for improvements. For example, to compute the required type-3 NUFFTs we have used the efficient implementation from (Barnett *et al* 2018). However, these computations still constitute about 80% of the total reconstruction time. Ongoing investigations suggest that downsampling the reference image by a factor of two may result in a significant decrease in computation time and memory requirement without degrading the reconstruction quality. Moreover, although the implementation from Barnett *et al* (2018) supports parallel computing on the CPU, a GPU implementation is not yet available but could significantly reduce the reconstruction time. Additionally, the reconstruction can be accelerated by further reducing the number of unknowns in the motion-fields by exploiting correlations not only in space, but also in time. For example, a patient-specific motion model could be trained in an offline MR-MOTUS reconstruction, and subsequently only time-varying coefficients could be reconstructed in a very fast online MR-MOTUS reconstruction (McClelland *et al* 2017, Sbrizzi *et al* 2019).

The quality of the reconstructed motion-fields may be improved in several ways. The small amount of snapshot data used for the reconstructions with MR-MOTUS implies that additional information must be provided in terms of *a priori* knowledge on the motion-fields. In this work this in done through the motion model (cubic B-splines and affine) and the curvature regularization. Both penalize irregular motion-fields and promote smoothness. Smoothness in the motion-fields is a good approximation in most of the cases, but may be suboptimal along sliding organ surfaces that naturally introduce discontinuities in the motion-fields (see figure 8). Different regularization and motion models that promote e.g. piece-wise smoothness may improve upon the results in this work.

In order to assess the quality of the reconstructed motion-fields, fully-sampled Cartesian data were acquired in discrete motion states and snapshot data was retrospectively undersampled from the corresponding fully-sampled k-spaces. The fully-sampled data allowed for image reconstructions that were used to contextualize the MR-MOTUS results by making a comparison with IR+OF. MR-MOTUS outperforms IR+OF in most experiments in this work, but since only two subjects were included in the comparison no general conclusions can be drawn regarding the performance of MR-MOTUS versus IR+OF. Additionally, snapshot data will in practice be acquired as fast as possible with prospective undersampling, non-Cartesian acquisitions and during continuous motion. Preliminary results on prospectively undersampled data acquired with a 2D golden angle radial trajectory during free-breathing demonstrate the feasibility of MR-MOTUS in practice. A thorough validation of MR-MOTUS in a practical setting, including more subjects and 3D prospective undersampling, is required and will be the subject of a future work.

The requirement of an *a priori* available reference image can be perceived as a limitation of this work. However, a reference image is always available in an MRgRT setting through the pre-treatment MRI that is acquired to adjust the radiation plan to the day-to-day variations in anatomy (Raaymakers *et al* 2017). Besides MRgRT, MR-MOTUS can also be applied to other settings where a reference image is available. Examples include MR-guided interventions, but also dynamic contrast-enhanced (DCE) imaging. The latter application does require the extension of the signal model to incorporate the DCE dynamics, which could be done by following the approach suggested in Lam *et al* (2011). Applying MR-MOTUS in these settings will be considered for future research.

6. Conclusion

We have demonstrated the MR-MOTUS concept: reconstruct non-rigid, non-linear, 3D motion-fields directly from minimal k-space by exploiting the availability of a reference image and the compressibility of motion-fields. The signal model behind MR-MOTUS is validated on a numerical motion phantom. The reconstructions are validated both *in-silico* and *in vivo*. Results show that MR-MOTUS reconstructs *in vivo* 3D rigid head motion from 474-fold retrospectively downsampled k-space data, and *in vivo* non-rigid 3D respiratory motion from 63-fold retrospectively undersampled k-space data. This data could potentially be acquired in few milliseconds

using a time-efficient and dedicated non-Cartesian trajectory, and could therefore result in high frame rate non-rigid motion-fields that can be valuable for MR-guided radiotherapy and MR-guided interventions. Preliminary results on prospectively undersampled 2D data confirm the feasibility of the method in a practical setting. Future research will focus on reducing the reconstruction time and prospective 3D acquisition of the snapshot *k*-space data.

Acknowledgments

This work was supported in part by the Netherlands Organisation for Scientific Research (NWO) under Grant 15115. The authors would like to thank Tom Bruijnen, Stefano Mandija and Oscar van der Heide for the fruitful discussions, and Tom Bruijnen and Stefano Mandija for proofreading the manuscript.

ORCID iDs

Niek R F Huttinga https://orcid.org/0000-0002-1558-1827

References

Andreychenko A, Raaijmakers A, Sbrizzi A, Crijns S, Lagendijk J J, Luijten P R and van den Berg C A T 2017 Thermal noise variance of a receive radiofrequency coil as a respiratory motion sensor *Magn. Reson. Med.* **77** 221–8

Barnett A H, Magland J F and af Klinteberg L 2018 A parallel non-uniform fast Fourier transform library based on an 'exponential of semicircle' kernel *SIAM J. Sci. Comput.* **41** C479–C504

Bjerre T, Crijns S, af Rosenschöld P M, Aznar M, Specht L, Larsen R and Keall P 2013 Three-dimensional MRI-Linac intra-fraction guidance using multiple orthogonal cine-MRI planes *Phys. Med. Biol.* **58** 4943

Block K T *et al* 2014 Towards routine clinical use of radial stack-of-stars 3D gradient-echo sequences for reducing motion sensitivity *J. Korean Soc. Magn. Reson. Med.* **18** 87–106

Brix L, Ringgaard S, Sørensen T S and Poulsen P R 2014 Three-dimensional liver motion tracking using real-time two-dimensional MRI *Med. Phys.* **41** 042302

Cervino L I, Du J and Jiang S B 2011 MRI-guided tumor tracking in lung cancer radiotherapy *Phys. Med. Biol.* **56** 3773

Chauffert N, Ciuciu P and Weiss P 2013 Variable density compressed sensing in MRI. Theoretical versus heuristic sampling strategies *IEEE 10th Int. Symp. on Biomedical Imaging* (IEEE) pp 298–301

Chen M, Lu W, Chen Q, Ruchala K J and Olivera G H 2008 A simple fixed-point approach to invert a deformation field *Med. Phys.* **35** 81–8

Fischer B and Modersitzki J 2003 Curvature based image registration *J. Math. Imaging Vis.* **18** 81–5

Fu Z W, Wang Y, Grimm R C, Rossman P J, Felmlee J P, Riederer S J and Ehman R L 1995 Orbital navigator echoes for motion measurements in magnetic resonance imaging *Magn. Reson. Med.* **34** 746–53

Glitzner M, de Senneville B, Lagendijk J, Raaymakers B and Crijns S 2015 On-line 3D motion estimation using low resolution MRI *Phys. Med. Biol.* **60** N301

Griswold M A, Jakob P M, Heidemann R M, Nittka M, Jellus V, Wang J, Kiefer B and Haase A 2002 Generalized autocalibrating partially parallel acquisitions (GRAPPA) *Magn. Reson. Med.* **47** 1202–10

Ingle R R, Wu H H, Addy N O, Cheng J Y, Yang P C, Hu B S and Nishimura D G 2014 Nonrigid autofocus motion correction for coronary MR angiography with a 3D cones trajectory *Magn. Reson. Med.* **72** 347–61

Johnson K M 2017 Hybrid radial-cones trajectory for accelerated MRI *Magn. Reson. Med.* **77** 1068–81

Lam F, Haldar J P and Liang Z P 2011 Motion compensation for reference-constrained image reconstruction from limited data *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro* (IEEE) pp 73–6

Lee G R 2018 Robust rigid-body motion estimation from extremely short subsets of 3D Cartesian scans (Ann. Meeting ISMRM 2018 Paris) *Proc. Intl. Soc. Mag. Reson. Med.* **26** 1171

Lee J Y and Greengard L 2005 The type 3 nonuniform FFT and its applications *J. Comput. Phys.* **206** 1–5

Liu D C and Nocedal J 1989 On the limited memory bf GS method for large scale optimization *Math. Program.* **45** 503–28

Lustig M, Donoho D L, Santos J M and Pauly J M 2008 Compressed sensing MRI *IEEE Signal Process. Mag.* **25** 72–82

McClelland J R, Hawkes D J, Schaeffter T and King A P 2013 Respiratory motion models: a review *Med. Image Anal.* **17** 19–42

McClelland J R *et al* 2017 A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images *Phys. Med. Biol.* **62** 4273

Odille F, Cîndea N, Mandry D, Pasquier C, Vuissoz P A and Felblinger J 2008a Generalized MRI reconstruction including elastic physiological motion and coil sensitivity encoding *Magn. Reson. Med.* **59** 1401–11

Odille F, Vuissoz P A, Marie P Y and Felblinger J 2008b Generalized reconstruction by inversion of coupled systems (GRICS) applied to free-breathing MRI *Magn. Reson. Med.* **60** 146–57

Pipe J G 1999 Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging *Magn. Reson. Med.* **42** 963–9

Prieto C, Batchelor P G, Hill D, Hajnal J V, Guarini M and Irarrazaval P 2007 Reconstruction of undersampled dynamic images by modeling the motion of object elements *Magn. Reson. Med.* **57** 939–49

Pruessmann K P, Weiger M, Scheidegger M B and Boesiger P 1999 SENSE: sensitivity encoding for fast MRI *Magn. Reson. Med.* **42** 952–62

Raaymakers B *et al* 2017 First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment *Phys. Med. Biol.* **62** L41

Raaymakers B *et al* 2009 Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept *Phys. Med. Biol.* **54** N229

Rueckert D, Sonoda L I, Hayes C, Hill D L, Leach M O and Hawkes D J 1999 Nonrigid registration using free-form deformations: application to breast MR images *IEEE Trans. Med. Imaging* **18** 712–21

Sbrizzi A, Huttinga N R F and van den Berg C A T 2019 Acquisition, reconstruction and uncertainty quantification of 3D non-rigid motion fields directly from *k*-space data at 100 Hz frame rate (Ann. Meeting ISMRM 2019 Montreal) *Proc. Intl. Soc. Mag. Reson. Med.* **27** 0795

Seregni M, Paganelli C, Lee D, Greer P, Baroni G, Keall P and Riboldi M 2016 Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI *Phys. Med. Biol.* **61** 872

Stam M K, Crijns S P, Zonnenberg B A, Barendrecht M M, van Vulpen M, Lagendijk J J and Raaymakers B W 2012 Navigators for motion detection during real-time MRI-guided radiotherapy *Phys. Med. Biol.* **57** 6797

Stemkens B, Tijssen R H, de Senneville B D, Lagendijk J J and van den Berg C A 2016 Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy *Phys. Med. Biol.* **61** 5335

Tryggestad E, Flammang A, Hales R, Herman J, Lee J, McNutt T, Roland T, Shea S M and Wong J 2013 4D tumor centroid tracking using orthogonal 2D dynamic MRI: implications for radiotherapy planning *Med. Phys.* **40** 091712

Uecker M and Tamir J 2019 mrirecon/bart: version 0.5.00 (Version v0.5.00) *Zenodo* (https://doi.org/10.5281/zenodo.3376744)

Usman M, Atkinson D, Odille F, Kolbitsch C, Vaillant G, Schaeffter T, Batchelor P G and Prieto C 2013 Motion corrected compressed sensing for free-breathing dynamic cardiac MRI *Magn. Reson. Med.* **70** 504–16

Van der Kouwe A J, Benner T and Dale A M 2006 Real-time rigid body motion correction and shimming using cloverleaf navigators *Magn. Reson. Med.* **56** 1019–32

Welch E B, Manduca A, Grimm R C, Ward H A and Jack C R Jr 2002 Spherical navigator echoes for full 3D rigid body motion measurement in MRI *Magn. Reson. Med.* **47** 32–41

White N, Roddey C, Shankaranarayanan A, Han E, Rettmann D, Santos J, Kuperman J and Dale A 2010 PROMO: real-time prospective motion correction in MRI using image-based tracking *Magn. Reson. Med.* **63** 91–105

Wisetphanichkij S and Dejhan K 2005 Fast Fourier transform technique and affine transform estimation-based high precision image registration method *GESTS Int'l Trans. Comput. Sci. Eng.* **20** 179–91

Zachiu C, de Senneville B, Moonen C and Ries M 2015a A framework for the correction of slow physiological drifts during MR-guided HIFU therapies: proof of concept *Med. Phys.* **42** 4137–48

Zachiu C, Papadakis N, Ries M, Moonen C and de Senneville B D 2015b An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs *Phys. Med. Biol.* **60** 9003