

**Towards real-time motion estimation
for MR-guided radiotherapy**

FROM MR-IMAGES TO MR-MOTUS

Niek Ricardo Ferdinand Huttinga

Cover description:

The lungs captured in a mathematical structure; it depicts our model that mathematically captures the motion of lungs in the body.

Towards real-time motion estimation for MR-guided radiotherapy

Ph.D. Thesis, Utrecht University, The Netherlands

Cover design: Joanan Blei
Layout: Niek Huttinga
Typesetting: L^AT_EX
Printing: ProefschriftMaken BV
ISBN: 978-90-393-7550-1

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means without the prior written permission from the author. The copyright of the papers that have been published or have been accepted for publication has been transferred to the respective journals.

Copyright

- © Niek Huttinga
- © IOP publishing Ltd (Chapter 2)
- © Wiley (Chapter 3)
- © IEEE (Chapter 4)

Financial support for publication and printing of this thesis was kindly provided by Modus QA, Elekta, and ChipSoft. The research was supported by NWO Grant #15115. Niek Huttinga gratefully acknowledges financial support by KNAW for his research internship at the University of California, Berkeley, through the Van Leersum grant.

Towards real-time motion estimation for MR-guided radiotherapy

FROM MR-IMAGES TO MR-MOTUS

Richting tijdsgetrouwe bewegingsschatting voor MRI-gestuurde radiotherapie

VAN MR-BEELDEN NAAR MR-MOTUS

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op
gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge
het besluit van het college voor promoties in het openbaar te verdedigen

op

woensdag 31 mei 2023 des middags te 4:15 uur

door

Niek Ricardo Ferdinand Huttinga

geboren op 18 februari 1993

te Arnhem

Promotoren:

Prof. dr. C.A.T. van den Berg

Prof. dr. P.R. Luijten

Copromotor:

Dr. A. Sbrizzi

Beoordelingscommissie:

Prof. dr. ir. J.J.W. Lagendijk

Prof. dr. B.W. Raaymakers (voorzitter)

Prof. dr. R.H. Bisseling

Dr. M.F. Fast

Prof. dr. M. Lustig

Contents

Contents	i
List of acronyms	iii
1 General introduction	1
2 MR-MOTUS: Model-based non-rigid motion estimation	17
3 Non-rigid 3D motion estimation at high temporal resolution using low-rank MR-MOTUS	57
4 Real-time non-rigid 3D respiratory motion estimation using MR-MOTUS	95
5 Gaussian Processes for real-time 3D motion & uncertainty estimation	129
6 General discussion	159
7 Appendices	173
Nederlandse samenvatting	175
References	179
Publications & Honors	189
Acknowledgements	193
Curriculum vitae	199

List of acronyms

2D+t	Two-dimensional space + time
3D+t	Three-dimensional space + time
aMRgRT	Adaptive MR-guided radiotherapy
AP	Anterior-posterior (Belly-Back)
ARD	Automatic Relevance Determination
ART	Adaptive radiotherapy
B₀-field	Static background magnetic field to polarize spins and create magnetization
B₁⁻-field	RF receive field; proportional to surface coil receive sensitivities
B₁⁺-field	RF transmit field used to tip the spins into the transverse plane
B-spline	Basis spline
BART	Berkeley Advanced Reconstruction Toolbox
CBCT	Cone beam computed tomography
CASPR	Cartesian Spiral Readouts
CG	Conjugate gradient
CG-SENSE	Conjugate gradient SENSE
CMR	Cardiac MR
COM	Center-of-mass
CS	Compressed Sensing
CS2Dt	Compressed Sensing 2D+t reconstruction
CT	Computed Tomography
CTV	Clinical target volume
DCE	Dynamic contrast enhanced
DVF	Deformation Vector Field
EBRT	External beam radiotherapy
EPE	End-point-error
FFD	Free-form deformations
FFT	Fast Fourier transform
FH	Feet-head
FOV	Field of view
GIRF	Gradient impulse response function
GTV	Gross tumor volume

GM3DR	3D golden-mean radial trajectory
GP	Gaussian Processes
GPU	Graphics Processing Unit
IGRT	Image-guided radiotherapy
IMRT	Intensity modulated radiotherapy
IR	Image registration
LINAC	Linear accelerator
LR	Left-right
MLC	Multi-leaf collimator
MLE	Maximum Likelihood Estimation
MI	Mutual Information
MV	Mega Volt
MRI	Magnetic Resonance Imaging
MRgRT	MR-guided radiotherapy
MR-MOTUS	Model-based Reconstruction of MOTion from Undersampled Signal
NMI	Normalised mutual information
NRMSE	Normalised root mean squared error. In this work the NRMSE e between vector \mathbf{a} and target \mathbf{b} is defined as $e = \frac{\ \mathbf{a}-\mathbf{b}\ }{\ \mathbf{b}\ }$.
NUFFT	Non-uniform fast Fourier transform
NWO	Netherlands Organisation for Scientific Research
OAR	Organs-at-risk
OF	Optical Flow
PC	Principal component
PCA	Principal component analysis
PICS	Parallel Imaging & Compressed Sensing
PI	Parallel imaging
PET	Positron emission tomography
PTV	Planning target volume
rCASPR	Rewound Cartesian Spiral Readouts
RMSE	Root mean squared error
RF	Radio Frequency
RR-IR	Respiratory-resolved image reconstruction
SNR	Signal-to-noise ratio
SPGR	Spoiled gradient echo
STAR	Stereotactic arrhythmia radiotherapy
TE	Echo time
TR	Repetition time
TV	Total Variation
TSE	Turbo Spin Echo
UTE	Ultra-short echo time
vNAV	Volumetric navigator
VT	Ventricular Tachycardia

Chapter **1**

General introduction

1.1 Motivation

Cancer is still a leading cause of death today, with a staggering estimate of nearly 10 million deaths globally in 2020 [1]. Radiotherapy has turned out to be one of the most effective weapons in the ongoing battle against cancer; currently, about 40% of patients undergo radiotherapy, either for curative or for palliative treatments [2]. One of the major strengths of radiotherapy is that it allows for highly localized, non-invasive treatments.

Radiotherapy has significantly evolved over the last decade, resulting in accurate and effective treatments of many tumor types. This progress was largely driven by advancements in image-based guidance. This started with the advent of three-dimensional CT-guided radiotherapy in the 1990s, which enabled three-dimensional planning of radiation treatments on the basis of imaging [3]. During the last decade, magnetic resonance imaging [4, 5] is becoming increasingly more popular in radiotherapy workflows. MRI provides more flexibility and a better soft-tissue contrast [6], which allows for more accurate tumor delineations, and thereby allows for more precise treatments. However, imaging and radiation currently still frequently have to be performed in different hospital rooms, which results in uncertainties regarding the patient's position and tumor location.

Several years ago the MR-linac was introduced to reduce these uncertainties. This hybrid machine combines an MRI scanner and a linear accelerator (linac) in a single device [7–10]. A major benefit of radiotherapy with an MR-linac is the possibility for a see-what-you-treat workflow; in-room MR-imaging directly prior to the treatments, so-called *pre-beam imaging*, and imaging *during* radiotherapy treatments, so-called *beam-on imaging*. The latter allows for continuous monitoring of the tumor's position and is particularly beneficial for abdominal and thoracic tumors that are frequently subject to internal body motion. Possible sources of such motion are physiological processes such as respiration, cardiac contraction, bladder filling and bowel movements. Each of these processes can induce motion on a different timescale, and with a different amplitude; e.g. bladder filling: several centimeters in the order of hours, bowel motion: several centimeters in the order of minutes, respiratory motion: about a centimeter in 4 seconds, and cardiac motion: about 0.5 cm in a second. Since these types of motion can affect the tumor's location, they should be taken into account when designing the treatment plan. When beam-on imaging is not available, it is challenging to estimate the exact tumor position during the treatment. Consequently, a larger radiation margin is required to ensure the tumor is hit with the radiation at all times, regardless of its motion. However, the increased margins also result in more toxicity to the surrounding healthy tissue. To spare this healthy tissue the dose has to be lowered, which decreases the treatment's efficacy. The ultimate potential of the MR-linac is to exploit the beam-on MR-imaging to continuously monitor the tumor's location, and adapt the radiation plan in real-time accordingly. As a result, tight margins can be employed, without lowering the dose to the tumor, thereby only minimally sacrificing treatment efficiency. This is called real-time adaptive MR-guided radiotherapy (aMRgRT). With the advent of hybrid MR-scanners such as the MR-linac the field of radiotherapy will be pushed more and more towards aMRgRT in the coming years.

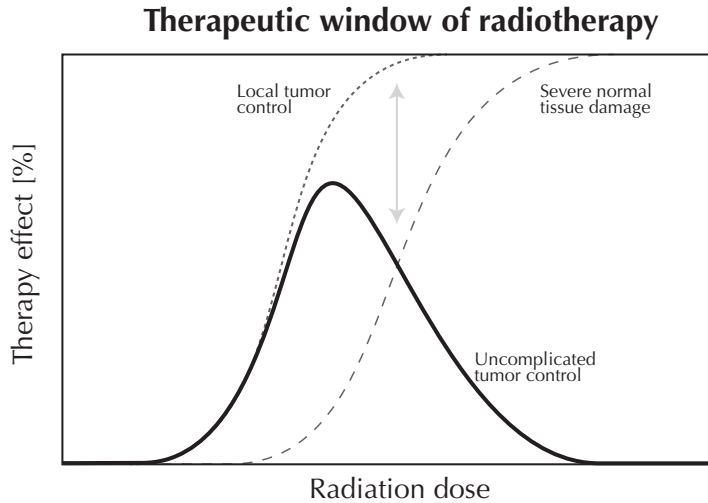


Fig. 1.1: Holthusen [12] curves that visualize the therapeutic treatment window. In this window cancerous tissue is killed, and healthy tissue survives. Figure based on [12].

One of the major hurdles towards the implementation of real-time aMRgRT on the MR-linac is the 3D imaging speed of MRI, which largely influences the treatment efficiency. The present-day MR-imaging speed allows aMRgRT treatments of slowly moving tumors such as prostate tumors [11]. Unfortunately, 3D imaging is currently too slow for treatments of abdominal or thoracic tumors, which are subject to faster motion, such as respiratory motion.

In this thesis, we develop new methods that enable high-speed 3D motion estimation from MR-data with the aim to close the gap towards real-time aMRgRT with the MR-linac. A key observation that motivated these methods is that MR-imaging is an intermediate step in the acquisition of the actual quantity of interest, namely the motion of the tumor and organs-at-risk (OAR), while this step does currently form the bottleneck for the speed of motion estimation. Our main method called MR-MOTUS allows to omit the intermediate image registrations steps in the motion estimation pipeline altogether, and allows to reconstruct motion-fields directly from raw MR-signal. We envision applications to aMRgRT with the MR-linac, where MR-MOTUS is used instead of, or in conjunction with MR-imaging.

1.2 Radiotherapy

An introduction to radiotherapy

Radiotherapy uses high-energy ionizing radiation such as X-rays, gamma rays, electron beams, or proton beams, to damage malignant tumor cells. The radiation is aimed at the tumor and damages the DNA of cancerous tissue. This impedes future tumor cell divisions and increases the probability of cell death [13–15].

A frequently used form of radiotherapy is external beam radiotherapy (EBRT). In EBRT, a focused X-ray beam is generated with a LINAC, and applied from outside the body according to a pre-specified radiation plan. Unfortunately, it is inevitable that an external X-ray beam, focused on an internal target, also hits surrounding healthy tissue. To spare the healthy tissue, a trade-off has to be made between treatment efficacy and risk of toxicity.

Fortunately, cancerous tissue is more sensitive to radiation than healthy tissue. This results in a range of radiation dose for which healthy tissue survives, and cancerous tissue dies: the so-called treatment window. This is visualized by Holthusen curves [12] in Figure 1.1. There are two main strategies to exploit the differences in the effect of radiotherapy between healthy and cancerous tissue. Firstly, the external radiation beam can be focused on the tumor from different inclination angles. Secondly, the complete treatment is usually split in so-called fractions. The first approach reduces the maximum dose to healthy tissue by spreading it over a larger volume, while the second allows for a recovery of the healthy tissue in between radiotherapy fractions.

Radiotherapy in practice

Radiotherapy treatments are often carried out in three phases: 1) pre-treatment phase; 2) pre-beam phase; 3) beam-on phase. Each of the three steps will briefly be described below.

Pre-treatment phase

The first step in the pre-treatment phase is the acquisition of a CT scan. This CT allows to visualize the internal anatomy and to determine the local electron density of the patient, a quantity that is needed to simulate the radiation dose in a later stage. Next, a pre-treatment MRI is acquired, and the tumor and OAR surrounding the tumor are delineated on the pre-treatment MRI. The delineation of all visible tumor tissue is called the gross tumor volume (GTV). The OAR are sensitive organs to which the radiation dose should be minimized to prevent complications [16]. To account for microscopic tumor tissue that is invisible on the MRI, the GTV is slightly expanded to a clinical target volume (CTV).

After pre-treatment imaging and delineation, a dose is prescribed to the tumor and OAR. In the subsequent inverse planning step the actual radiation plan is computed: a set of angles and shapes that can be processed by the LINAC to deliver the

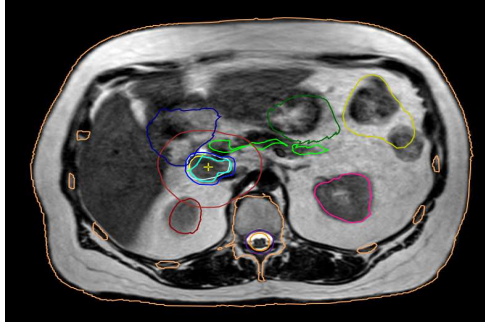


Fig. 1.2: A practical example of radiation treatment margins for a pancreas cancer patient. Different margins are denoted by different colors overlaid on the pre-beam MRI.

prescribed dose map. These steps are usually referred to as the treatment *simulation* phase, since it simulates what will actually happen during the treatment.

In conventional radiotherapy, the actual radiation is spread out over multiple fractions. The design of the radiation plan described above, however, is performed only once - in the pre-treatment phase which is performed prior to any fraction - and the resulting radiation plan is subsequently delivered in all fractions. Frequently, the fractions are delivered over several weeks, and in a different hospital room than the pre-treatment phase was performed. Evidently, this workflow may result in undesirable differences between delivered and prescribed dose, for example due to a small difference in the patient's position between pre-treatment and radiation delivery, a different organization of the internal anatomy, or tumor motion due to physiological processes such as respiratory motion, cardiac contractions, or bowel movements. In general, all of these factors contribute to an increased uncertainty in the tumor's location during dose delivery. This uncertainty may be reduced by laser-based patient positioning, or by using so-called *pre-beam imaging* that visualizes the daily internal anatomy and/or bony landmarks (see next section). The higher the quality of the pre-beam imaging, the more the uncertainty can be reduced, and the higher the efficiency of the treatment. To account for residual uncertainty in the tumor's location, the CTV margins are expanded to a planning target volume (PTV) [17]; the larger the uncertainties, the larger the PTV. Figure 1.2 shows an example of margins used pre-treatment phase.

Finally, the radiation treatment plan is calculated using all available images. In EBRT the radiation is typically applied with several shapes from different angles, with a focus point at the tumor. This is referred to as intensity modulated radiotherapy (IMRT) [18]. IMRT results in a high dose on the tumor, and minimizes the dose to healthy tissue and OAR by spreading the radiation over neighboring tissues in all directions.

Pre-beam phase

The pre-beam phase is performed in the same room as the dose delivery. The goal of this phase is to reproduce the patient's position in the pre-treatment phase as closely as possible. This ensures that the simulated treatment plan will actually be delivered to the correct locations. There are several ways to achieve this. For example, for head-and-neck radiotherapy an immobilization mask created during the simulation phase could be used, and in abdominal radiotherapy the patient could be aligned with lasers and skin marks. Alternatively, in-room imaging may be performed, for example with cone beam CT (CBCT), MegaVolt (MV), or MRI in case of an MR-linac. Based on the in-room images, the patient's position may be altered to resemble the positioning in the pre-treatment phase as closely as possible. Eventually the green light is given by the attending physician to proceed to the next phase: radiation delivery.

1

Beam-on phase

In the beam-on phase the dose is actually delivered with the LINAC according to the pre-computed radiation plan. The LINAC uses multi-leaf collimators (MLC) to shape the radiation beam according to the radiation plan. The MLC are made up of small rectangular panels, usually made of tungsten to block radiation. These panels can be positioned very accurately with little latency to shape the beams according to position measurements. The delivery of the radiation plan is typically spread out over fractions. This allows healthy tissue that is only mildly exposed to the radiation to recover in between fractions, while cancer cells are forced into a form of programmed cell death referred to as apoptosis.

Image guidance during radiotherapy

The radiation delivery may be guided by in-room imaging, both during the beam-on and pre-beam phase. This is called image-guided radiotherapy (IGRT) [19]. This is especially useful for moving tumors, such as abdominal or thoracic tumors subject to respiratory motion. Furthermore, the in-room imaging may in theory be used to adapt the treatment plan inter-fraction [19] or intra-fraction [20]. This is referred to as adaptive radiotherapy (ART). It should be noted that especially intra-fraction ART is currently still challenging to realize in practice. The recently introduced MR-linac is a big step forward towards the practical implementation of ART. Additionally, it has the ultimate potential to combine MR-guided radiotherapy with ART into adaptive MR-guided radiotherapy (aMRgRT).

The MR-linac for adaptive MR-guided radiotherapy (aMRgRT)

As mentioned previously, several geometric uncertainties arise in a typical radiotherapy workflow, which are accounted for in the PTV. The main advantage of the MR-linac is that it has the potential to reduce these uncertainties using in-room MR-imaging; the MRI provides excellent soft-tissue contrast to simultaneously visualize patient contours, tumors, and OAR. Guided by the in-room MR-imaging, aMRgRT can be realized by subsequently adapting the radiation plan based on the difference between the current positions of the organs and the positions during the pre-treatment phase in which the radiation plan was designed. Using accurate estimates of tumor locations from the MRI, and adapting the radiation plan accordingly effectively allows to increase the radiation dose to the tumor, while preserving low radiation dose to OAR. This increases the efficiency of the treatment, and eventually allows for the same treatment results with fewer radiation fractions, i.e. so-called hypo-fractionated radiotherapy.

The geometric uncertainties in the tumor's location can in general be attributed to motion with respect to positioning in the pre-treatment radiation phase. This motion can be divided into inter-fraction motion that occurs *in between* fractions, and intra-fraction motion that occurs *during* a fraction. Inter-fraction motion includes all motion with large magnitude that occurs on a relatively large timescale (days), e.g. bulk motion due to differences in patient positioning, and daily differences in the internal anatomy. Intra-fraction motion includes all motion that occurs within a fraction on a relatively small timescale (seconds-minutes), and mainly caused by physiological processes such as respiration, cardiac contractions, bowel movements, organ drifts due to muscle relaxation, and bladder filling.

Inter-fraction motion is typically the largest component, and can be resolved with the MR-linac using pre-beam MR-imaging. Intra-fraction motion is typically a smaller component, and resolving this type of motion is of particular interest for tumors with large motion relative to their size. To make this more intuitive, consider a spherical tumor of 1 cm diameter, subject to motion with a magnitude of 5 cm in feet-head. Without any knowledge on the exact location, but just the magnitude and direction of the motion, the PTV extent in feet-head will be 5 times larger than the actual tumor's size in feet-head. As a consequence, the amount of radiated healthy tissue is 4 times as large as the amount of radiated cancerous tissue. In conclusion, the better and the faster we know the motion of the internal anatomy, the higher the potential efficiency of an adaptive treatment.

Intra-fraction motion can be addressed with two strategies: gating and tracking. With gating, a repetitive motion is assumed, and the radiation is only applied when the tumor location is within a pre-determined range [21]. Under the assumption of periodic motion, this requires beam-on monitoring of the phase of the motion. Evidently, the treatment's time-efficiency is inversely proportional to both the period and amplitude of the tumor motion. Another strategy to address intra-fraction motion is real-time aMRgRT. that is, reconstruct the motion in real-time, and adjust the radiation targets accordingly. A real-time aMRgRT workflow for the MR-linac is visualized in Figure 1.3. For abdominothoracic tumors - largely displaying respiratory motion - a successful implementation of this workflow requires motion-field

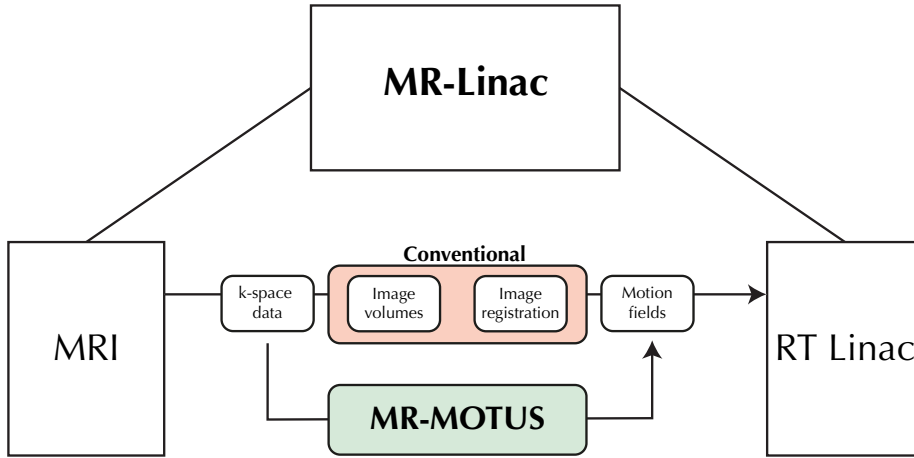


Fig. 1.3: A schematic overview that visualizes MR-MOTUS in a real-time adaptive MR-guided radiotherapy workflow with the MR-linac.

estimates at 5 frames per second [22, 23]. Since both OAR and tumors move in 3D, a 3D motion-field is preferred over a 2D motion-field. Although high-speed motion estimation is currently feasible in 2D, it is still a technical challenge in 3D due to the relatively slow imaging speed of MRI.

1.3 From MR-images to MR-MOTUS

Motivation

From the discussion above it is clear that high-speed 3D motion estimation is important to further improve radiotherapy treatments with an MR-linac. Before giving an overview of conventional methods in this context, and introducing our own proposed method MR-MOTUS, it is important to further refine the notion of 'high-speed' in this context. For real-time ART, the speed of motion estimation should be measured in terms of total latency of the motion estimation procedure, including both the time required for the actual estimation of the motion from the data, as well as the time required to acquire this data. After all, the motion estimates are made to reduce uncertainty in the tumor's location, but their reliability is inversely proportional to the total latency of the motion estimation process. Hence, for real-time MR-guided radiotherapy, a real-time 3D motion estimation method should be able to estimate 3D motion in real-time, from a limited amount of MR-data that can be acquired in real-time.

A straightforward strategy for MR-based motion estimation in this context follows three steps. In the first step, MR-data is acquired in Fourier space, referred to as k -space data. Next, an MR-image is reconstructed from consecutively acquired data. In the third step, two consecutive MR-images are registered, resulting in the motion-

field that can spatially transform one image to the other. For real-time aMRgRT these three steps should be repeated as fast as possible. The bottleneck in terms of speed is the second step, which requires the acquisition of a minimal amount of data; for one 3D image with sufficient SNR and spatial resolution the required k -space data can be acquired in about 1 second [24, 25]. This would result in a frame-rate of 1 Hz. Although this is sufficient for slowly moving tumors such as prostate tumors [11], it is about a factor 5 too slow for tumors subject to respiratory motion [22, 23]. One strategy to resolve this, is to reduce the amount of data required for step 2. This can be done using techniques like Parallel Imaging [26, 27], or Compressed Sensing [28]. However, most advanced reconstruction algorithms significantly increase the reconstruction time, making them less suitable for real-time tracking applications. Moreover, state-of-the-art 3D MR-image reconstruction methods are still struggling to achieve sufficiently high (> 5 Hz) temporal resolution, even with long offline reconstructions. Some previously proposed methods include [29–31], which achieved 1.4-2.0 Hz reconstructions. However, the reconstruction in Ong *et al.* [29], for example, took around 6 hours. Other methods achieved similar temporal resolutions with multi-plane 2D cine imaging [25, 32, 33]. The highest temporal resolution is typically achieved by surrogate signal models, which couple surrogate signals to volumetric image reconstruction [34], or motion-field reconstructions [35, 36]. A downside of these methods is that they use a very low-dimensional input, which could potentially not completely capture all motion in the data.

The conventional image-based strategy for motion estimation outlined above is evidently not very data-efficient. It repeatedly acquires k -space data of an internal anatomy that is only slightly changing over time. This leads to a large amount of redundancy in the acquired data, while the acquisition of the data required for image reconstruction is the most time-consuming step in the motion estimation pipeline. Hence, to improve the motion estimation latency, it could be advantageous to focus on what actually changes the internal anatomy of one reference image, namely the motion-fields. Moreover, motion-fields can naturally be modeled with few degrees of freedom, since a large part of the internal organ motion-fields can be assumed smooth in space and time. Spatially, organs move very rigidly, and neighboring organs move very similarly. Temporally, tissue smoothly transitions to a different location. This yields a smooth spatio-temporal motion-field, with minimal local variations. In general, this smoothness allows to model the motion-fields with few degrees of freedoms, or stated differently, represent the motion-fields in a low-dimensional space with few unknowns. Usually, few unknowns can be reconstructed from few samples. Under this assumption, few data would be required for motion estimation. Taken all considerations above into account, we therefore propose a new strategy for MR-based motion estimation: directly reconstruct motion-fields from k -space data and one reference image. This is the general idea behind the MR-MOTUS framework, which is an acronym for 'Model-based Reconstruction of MOTion-fields from Undersampled Signal.

A dynamic MR-signal model that accounts for subject motion

In order to perform motion-field reconstructions directly from k -space data, a signal model is required that relates these two quantities. We derive such a signal model by extending the standard Fourier MR-signal model to a dynamic setting. A high-level overview of the steps in this derivation is given below, we refer to Chapter 2 for more details.

Suppose we are imaging a d -dimensional dynamic object, with spin-density $\rho_t(\mathbf{r}) \in \mathbb{C}$ and transverse magnetization $m_t(\mathbf{r})$ at spatial coordinate $\mathbf{r} \in \mathbb{R}^d$ and time t . The acquired k -space signal $s_t(\mathbf{k})$ at time t and k -space coordinate $\mathbf{k} \in \mathbb{R}^d$ can then be formulated as the Fourier transform of $\mathbf{q}_t(\mathbf{r}) := \rho_t(\mathbf{r})m_t(\mathbf{r})$:

$$s_t(\mathbf{k}) = \int q_t(\mathbf{r}_t) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}_t} d\mathbf{r}_t. \quad (1.1)$$

1

To include motion in this signal model, the dynamic object q_t should be written in terms of motion-fields. This can be done by assuming that the dynamic q_t can be represented as a static reference image q_0 , i.e. the object at a reference time $t = 0$, warped by time-dependent motion-fields \mathbf{U}_t :

$$q_t(\mathbf{r}_t) \approx q_0(\mathbf{U}_t(\mathbf{r}_t)) (\det(\nabla\mathbf{U}_t)(\mathbf{r}_t)), \quad (1.2)$$

where the last term is the determinant of the Jacobian of \mathbf{U}_t (hereafter ‘Jacobian determinant’) which locally compensates for compression and expansion induced by the transformation \mathbf{U}_t . The motion-fields are denoted by \mathbf{U}_t , and are defined such that they transform the location \mathbf{r}_t at time t to a location $\mathbf{U}_t(\mathbf{r}_t) \in \mathbb{R}^d$ at time 0. It should be noted that this equation for q_t implies that both the transverse magnetization of an infinitesimally small area and its spin density do not change as it moves through the magnetic field due to motion. Effectively this means that the data should be acquired in a steady-state of the transverse magnetization, and that the magnetic field is very homogeneous. We discuss these and other assumptions underlying this signal model in more detail in Chapter 2.

To derive the signal model we proceed by substituting Eq. (1.2) in Eq. (1.1), and perform the change-of-variables $\mathbf{r}_t \mapsto \mathbf{T}_t(\mathbf{r})$, where \mathbf{T}_t is defined as the right inverse of \mathbf{U}_t . This yields the MR-MOTUS signal model:

$$s_t(\mathbf{k}) = F(\mathbf{T}_t, q_0)(\mathbf{k}) = \int q_0(\mathbf{r}) e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r})} d\mathbf{r}. \quad (1.3)$$

A visual overview of the strategy outlined above is provided in Figure 1.4.

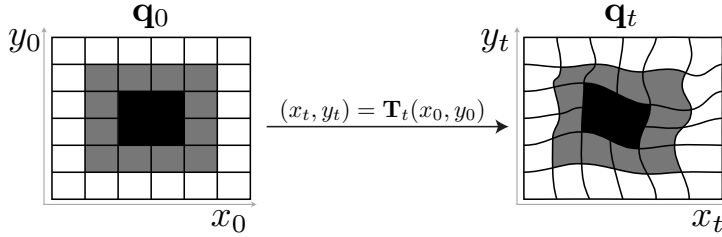


Fig. 1.4: The general idea behind MR-MOTUS: represent a dynamic object \mathbf{q}_t as changes of coordinates \mathbf{T}_t applied to a static reference object \mathbf{q}_0 .

Reconstruction strategy: solving the inverse problem

1

Now that a signal model is derived that relates motion-fields to k -space data and a reference image, a strategy should be determined to actually reconstruct the motion-fields from acquired k -space data. In what follows we assume that a reference image is available. The validity of this particular assumption and all other assumptions underlying the signal model are further discussed in Chapter 2. Mathematically, the model Eq. (1.3) defines a *forward model*, and effectively transforms a motion-field into an MR-signal. For a reconstruction, however, the opposite direction is desired. That is, to transform an acquired signal in motion-fields. Estimating motion-fields that correspond to an acquired signal is mathematically referred to as solving the (non-linear) *inverse problem* corresponding to the signal model F . For particular choices of F an explicit inverse function F^{-1} could be available, which allows to solve the inverse problem by simply applying F^{-1} to the observed data: $\mathbf{T}^* = F^{-1}(\mathbf{s})$. However, in case such a function is not available, as is the case for MR-MOTUS and many other inverse problems, more advanced strategies are required. One strategy is the variational method, which iteratively probes the output of the signal model by varying the input motion-fields. This iterative process continues until the difference between the signal model's output $F(\mathbf{T})$ and the actual observed data \mathbf{s} is at a minimum. However, when only a minimal amount of data is available, there could be many possible minimizers of this difference. That is, there is no unique minimizer and the set of viable solutions is large. As a result, the inverse problem is considered ill-posed in Hadamard's sense (1865-1963). To guide the minimization process to a 'good' solution, a regularization term can be added to the objective function that reduces the space of viable solutions by incorporating a priori information we may have about what we consider to be a good solution. For example, if we expect motion-fields to be smooth, we could penalize non-smooth motion-fields with a regularization based on the magnitudes of the spatial derivatives. In this thesis we consider several regularization terms. Chapter 2 introduces regularization based on the spatial curvature, Chapter 3 introduces a penalty based on the Jacobian determinant of the motion-fields [37], and Chapter 4 considers vectorial total variation [38].

The variational formulation of the MR-MOTUS inverse problem can be stated as

$$\min_{\mathbf{T}} \mathcal{D}(\mathbf{F}(\mathbf{T}), \mathbf{s}) + \lambda \mathcal{R}(\mathbf{T}), \quad (1.4)$$

where $\lambda \in \mathbb{R}^+$ is the regularization parameter that balances the optimization problem between a pure data-fit ($\lambda = 0$) and regularization. This minimization problem is solved in an iterative fashion, in which the solution is iteratively improved based on the gradient(s) of the objective function.

The difficulty of solving the optimization in Eq. (1.4) depends largely on the properties of F , and the amount of (informative) available data. The MR-MOTUS signal model results in a non-convex, non-linear, underdetermined inverse problem, which unfortunately is amongst the hardest inverse problems to solve. Nevertheless, some methods exist to do this. Unfortunately, the computation time of the iterative updates typically increases with the number of unknowns. Since many iterations (>100) may be required, the reconstruction time is typically long. Reducing this computation time was one of the main technical challenges in order to apply MR-MOTUS in a practical setting. This is addressed in Chapters 3 and 4 with a low-rank motion and a novel two-step reconstruction scheme to achieve real-time reconstructions.

Model order reduction

The regularization term in the variational formulation above implicitly reduces the space of possible solutions by favoring some solution over others. This effectively limits the degrees-of-freedom in the motion-fields, which in turn improves our chances to reconstruct motion-fields from reduced k -space data.

Another approach to reduce the space of viable solution is a model order reduction, i.e. explicitly reducing the number of parameters in the motion-fields. This was considered in all chapters through a cubic B-spline representation. Chapter 3 and 4 introduce an efficient low-rank motion model [29] which further reduces the model order. This compressed motion model represents the complete motion-field with a reduced number of parameters by leveraging a priori knowledge on the motion-fields. This thesis considers motion models that can be represented as follows:

$$\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{r}, t) = \mathbf{r} + \sum_i \theta_i \mathbf{B}_i(\mathbf{r}, t), \quad (1.5)$$

where \mathbf{B}_i denote spatio-temporal basis functions, $\boldsymbol{\theta}$ denotes the motion model coefficients to be estimated from the data.

Low-rank model for spatio-temporal respiratory motion-fields

A natural way to construct a 3D+t spatio-temporal motion model would be to compute a tensor product between a spatial 3D spline model and a temporal 1D spline model. This would effectively make copies of the spatial 3D spline model along the temporal dimension. However, if the 1D spline model has m_t parameters, this would thus result in a factor m_t increase in the total number of parameters. In other words, the number of parameters scales with the number of timepoints. As we will

consider reconstructions of high temporal resolution motion-fields, this is an inconvenient property in practice. Instead, a low-rank 3D+t motion model is introduced in Chapter 3. This model assumes that motion-fields can be modelled as partially separable functions [39], where the separation occurs between the spatial and the temporal dimension. This low-rank model allows to represent the 3D+t motion-fields with $R \cdot (m_s^3 + m_t)$ parameters rather than $m_s^3 \cdot m_t$, where $R \leq 5$ denotes the rank of the motion model. For practical parameter settings ($R \ll m_t$), this model results in much fewer parameters: $R \cdot (m_s^3 + m_t) \ll m_s^3 \cdot m_t$.

Intuitively, this low-rank model assumes that motion-fields can be approximated well by the multiplication of a spatial component - which is independent of time, i.e. static - and a temporal component - which is independent of spatial coordinates, i.e. global. The spatial component can be understood as modelling the directions and relative magnitudes of motion at every spatial coordinate. The temporal coordinate can be understood as modeling the global scaling along these directions and relative magnitudes in the spatial component. With just a single component, the direction and relative magnitudes of motion are fixed in time, i.e. tissue can only move in straight lines. To increase the complexity of the model multiple 'space x time' components may be added together. Increasing the number of components could for example allow curves in the motion paths, or model different types of body motion simultaneously, e.g. bulk motion and respiratory motion. The term 'low-rank' refers to the fact that in practice few of such components are required. Since the rank of the motion model is less than or equal to the number of components (the components are not necessarily orthogonal), few components imply a low-rank spatiotemporal motion-field.

1

1.4 Thesis overview

In Chapter 2, the MR-MOTUS method is introduced for 3D motion-field reconstructions. The approach to focus directly on motion-fields is motivated, and the MR-signal model Eq. (1.3) is derived. Particular attention is paid to the underlying assumptions. A reconstruction strategy is outlined and a proof-of-concept is demonstrated in silico and on retrospectively undersampled in vivo data, acquired with Cartesian trajectory. Three-dimensional rigid head motion and non-rigid respiratory motion are reconstructed. For the head motion an affine motion model is employed. For the respiratory motion the spatial smoothness is exploited with a cubic B-spline motion model and a regularization term that penalizes spatial curvature.

In Chapter 3, MR-MOTUS is extended to 3D+t spatio-temporal motion-field reconstructions from prospectively undersampled data. For the latter, the Cartesian acquisition of Chapter 2 was replaced with a non-Cartesian radial acquisition. The extension in this chapter allows to exploit the compressibility of motion-field in both space and time. A low-rank motion model is introduced to cope with the large amount of unknowns in the reconstructions of spatio-temporal motion-fields in the extended framework, and simultaneously exploit spatio-temporal compression. The low-rank model splits motion-fields in spatial and temporal components. Each component is represented with a B-spline model. Reconstructions with the low-rank

motion model are performed for 2D+t and 3D+t time-resolved respiratory motion, 3D+t respiratory-resolved respiratory motion, and 3D+t head-and-neck motion. A regularization term based on the Jacobian determinant of the motion-fields (see Section 1.3) is introduced to enforce anatomically plausible motion-fields [40].

In Chapter 4, the low-rank reconstructions of Chapter 3 are extended to real-time reconstructions on prospectively undersampled data, acquired with an MR-linac. This is done by splitting the reconstruction in two phases: 1) an offline phase that reconstructs the spatial motion-field components, and 2) a real-time online phase that only reconstructs the temporal components per dynamic using the result from the offline phase. The fundamental assumption that underlies this strategy is that the spatial components from the offline phase still allow to accurately represent motion in the real-time phase. In other words, similar motion is assumed in the offline and online phases. The real-time reconstructions are enabled by the small number of unknowns in the second phase (less than 5), and a minimal number of reconstruction iterations (see Section 1.3). The latter is achieved by a near-optimal initialization with the solution of the previous dynamic.

The assumption of similar motion in the offline and online phases in Chapter 4 enables real-time 3D motion reconstructions, but also introduces an uncertainty in the reconstructions. Several practical scenarios may reduce the validity of this assumption, e.g. during bulk motion, coughing, or a severe breathing pattern change. In the worst case scenario, employing a pre-trained motion model during such abnormal motion for which it was not built results in errors. In an adaptive MRgRT workflow, this would then eventually result in a difference between delivered dose and the radiation plan, and could thereby potentially harm the patient. In Chapter 5 we propose a framework for joint real-time motion and uncertainty estimation. The framework leverages the probabilistic machine learning methodology of Gaussian Processes (GP) and is independent of the MR-MOTUS approach described in Chapters 1-4. A similar two-phase strategy as Chapter 4 is employed. In an offline phase a low-rank motion model is built from respiratory-resolved 3D motion-fields obtained with image-registration, and a GP regression model is trained to infer low-dimensional motion-field representation coefficients from three readouts of k -space data. In the online phase the trained GP infers the most-likely representation coefficients and a corresponding measure of confidence from the k -space data. The framework is validated on MR-linac data, prospectively acquired during normal breathing and abnormal motions. It is demonstrated that normal breathing motion-fields can be inferred accurately, while abnormal motion can correctly be detected with the measure of confidence provided by the framework. During real-time aMRgRT the framework could be used to track the motion of the tumor and OAR. If the motion estimates are too uncertain, the radiation can be temporarily halted and continued whenever confidence is restored.

Chapter 2

MR-MOTUS: Model-based non-rigid motion estimation

Niek R.F. Huttinga
Cornelis A. T. van den Berg
Peter R. Luijten
Alessandro Sbrizzi

The following chapter is based on:

MR-MOTUS: Model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal k-space data. *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015004, 2020.

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.

2

Supporting Videos

The supporting videos corresponding to this chapter can be accessed through <https://doi.org/10.6084/m9.figshare.20480550.v2> or the following QR-code:



2.1 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 [10]. 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 [41]. 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 (\sim 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 [42] motion was estimated by co-registering corrupted images reconstructed using parallel imaging [26, 27] and compressed sensing [28]. 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 [32, 33, 43–46]. Another image-based method is PROMO [47], 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.* [48] 3D motion is estimated from low-resolution 3D images for the purpose of MRgRT, and it is reported that $5 \times 5 \times 5\text{mm}^3$ 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) [49, 50].

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 [35, 36] or the time evolution of noise covariances of an RF coil-array [51]. An application of these surrogate signal motion models is the GRICS framework [52, 53], 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 [52, 53].

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.* [54], Van der Kouwe *et al.* [55], Welch *et al.* [56], Pipe [57], and Stam *et al.* [58]. These methods are mainly based on the explicit relation between a linear transformation and k -space data due to properties of the Fourier transform [59] and are therefore limited to affine motion. In Prieto *et al.* [60] 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 [41]. 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 chapter 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 same undersampled data. Finally, we demonstrate the potential of

MR-MOTUS in practice by reconstructing motion-fields from prospectively under-sampled snapshot data consisting of 15 spokes, acquired in 60ms with a 2D golden angle acquisition during free-breathing.

2.2 Theory

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 = 2.30/1.15ms, 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 [61, 62]. One of the images and the obtained motion-field are shown in Figure 2.1.

Next, a cubic B-spline basis [63] 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[‡] (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 2.2. 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.

Signal model derivation

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

$$s_t(\mathbf{k}) = \int_{\Omega} q_t(\mathbf{r}) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}. \quad (2.1)$$

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 :

$$\mathbf{U}_t(\mathbf{r}) = \mathbf{r} + \boldsymbol{\delta}_t(\mathbf{r}), \quad (2.2)$$

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

$$q_t(\mathbf{r})d\mathbf{r} = q_0(\mathbf{U}_t(\mathbf{r})) |\det(\nabla\mathbf{U}_t)(\mathbf{r})|d\mathbf{r}, \quad (2.3)$$

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

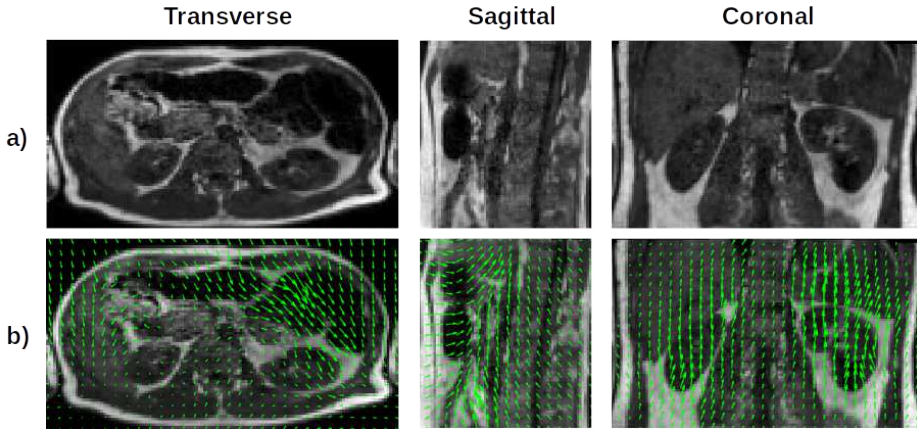


Fig. 2.1: Visualization of the data used in Section 2.2: a) Three slices of one of the images used for the registration, and b) three in-plane projections of the motion-field obtained with optical flow.

2

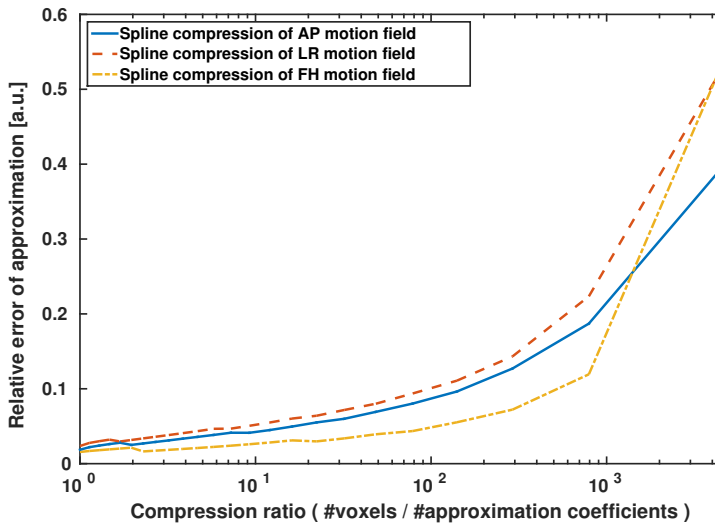


Fig. 2.2: Compression ratios for the three components of the motion-field of respiratory motion in a cubic B-spline basis, as described in Section 2.2. The curves show that the motion-fields are indeed very compressible: a maximum representation error of only 10% is made for all three components with 100 times as little approximation coefficients.

where $\nabla \mathbf{U}_t$ denotes the Jacobian of the motion-field \mathbf{U}_t . In the next two subsections 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 (2.3) into (2.1), followed by a change of coordinates:

$$s_t(\mathbf{k}) = \int_{\Omega} q_0(\mathbf{r}_0) e^{-i2\pi \mathbf{k} \cdot \mathbf{T}_t(\mathbf{r}_0)} d\mathbf{r}_0. \quad (2.4)$$

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

$$\mathbf{T}_t(\mathbf{r}) = \mathbf{r} + \boldsymbol{\eta}_t(\mathbf{r}), \quad \mathbf{U}_t \circ \mathbf{T}_t = \mathbf{T}_t \circ \mathbf{U}_t = \mathbf{Id}, \quad (2.5)$$

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 (2.4) with respect to motion-field.

To investigate under which assumptions (2.3) 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

$$q_t \equiv m_t \cdot \rho_t. \quad (2.6)$$

Next, we will derive temporal relations for m and ρ separately and combine them to obtain (2.3). Finally, we derive (2.4) by a substitution and a change of variables.

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 :

$$m_t(\mathbf{r}_t) = m_0(\mathbf{U}_t(\mathbf{r}_t)). \quad (2.7)$$

(Steady-state condition)

Note that for (2.7) 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 Tesla. Formally, (2.7) is exactly fulfilled under a few other technical assumptions, and for more mathematical details of the derivation we refer the reader to Supplementary Information Section 1.

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,

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

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

$$\int_{\mathbf{U}_t(X)} \rho_0(\mathbf{r}_0) \, d\mathbf{r}_0 = \int_X \rho_t(\mathbf{r}_t) \, d\mathbf{r}_t, \quad (2.8)$$

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 (2.5). We can then rewrite the left-hand side of (2.8) using the change of variables $\mathbf{r}_0 \mapsto \mathbf{U}_t(\mathbf{r}_t)$:

$$\int_{\mathbf{U}_t(X)} \rho_0(\mathbf{r}_0) \, d\mathbf{r}_0 = \int_X \rho_0(\mathbf{U}_t(\mathbf{r}_t)) |\det(\nabla \mathbf{U}_t)(\mathbf{r}_t)| \, d\mathbf{r}_t. \quad (2.9)$$

Combining (2.8) and (2.9) then yields

$$\int_X \rho_t(\mathbf{r}_t) \, d\mathbf{r}_t = \int_X \rho_0(\mathbf{U}_t(\mathbf{r}_t)) |\det(\nabla \mathbf{U}_t)(\mathbf{r}_t)| \, d\mathbf{r}_t, \quad (2.10)$$

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

$$\rho_t(\mathbf{r}_t) d\mathbf{r}_t = \rho_0(\mathbf{U}_t(\mathbf{r}_t)) |\det(\nabla \mathbf{U}_t)(\mathbf{r}_t)| d\mathbf{r}_t. \quad (2.11)$$

(Local spin conservation)

Derivation of the signal model

Combining (2.7) and (2.11) yields the previously described temporal relation (2.3) between the reference object, deforming object and motion-fields:

$$q_t(\mathbf{r}) d\mathbf{r} = q_0(\mathbf{U}_t(\mathbf{r})) |\det(\nabla \mathbf{U}_t)(\mathbf{r})| d\mathbf{r}. \quad (2.12)$$

Substituting (2.12) into the signal model (2.1) then yields

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

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

$$s_t(\mathbf{k}) = \int_{\mathbf{U}_t(\Omega)} q_0(\mathbf{r}_0) e^{-i2\pi \mathbf{k} \cdot \mathbf{T}_t(\mathbf{r}_0)} \, d\mathbf{r}_0. \quad (2.13)$$

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 (2.13) 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:

$$s_t(\mathbf{k}) = \int_{\Omega} q_0(\mathbf{r}_0) e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0)} d\mathbf{r}_0. \quad (2.14)$$

For a more formal derivation of the statements above we refer the reader to the Supplementary Information Section 1.

Inverse problem formulation

Note that (2.14) 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 (2.14). 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 (2.14) can then be rewritten in operator form as

$$\mathbf{s}_t = \mathbf{F}(\boldsymbol{\theta}_t | q_0), \quad (2.15)$$

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

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

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)[64, 65]. To reconstruct motion-fields the following minimization problem is solved:

$$\min_{\boldsymbol{\theta}_t} \|\mathbf{F}(\boldsymbol{\theta}_t) - \mathbf{s}_t\|_2^2 + \lambda \mathcal{R}(\mathbf{T}_t(\cdot | \boldsymbol{\theta}_t)), \quad (2.16)$$

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.

MR-MOTUS: Model-based reconstruction of motion-fields from undersampled signals

Regularization functional

A natural choice for \mathcal{R} in this setting, which was originally proposed in Fischer *et al.* [37], is to assume smooth motion-fields by penalizing the spatial curvature of the motion-fields:

$$\mathcal{R}(\mathbf{T}_t(\cdot | \boldsymbol{\theta}_t)) := \sum_{p \in \{x, y, z\}} \int_{\Omega} |\Delta T_t^p(\mathbf{r} | \boldsymbol{\theta}_t^p)|^2 d\mathbf{r}. \quad (2.17)$$

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 $\boldsymbol{\theta}_t = \{\boldsymbol{\theta}_t^x, \boldsymbol{\theta}_t^y, \boldsymbol{\theta}_t^z\}$:

$$\min_{\boldsymbol{\theta}_t} \|\mathbf{F}(\boldsymbol{\theta}_t) - \mathbf{s}_t\|_2^2 + \lambda \sum_{p \in \{x, y, z\}} \int_{\Omega} |\Delta T_t^p(\mathbf{r} | \boldsymbol{\theta}_t^p)|^2 \, d\mathbf{r}. \quad (2.18)$$

Motion models

In this work two motion models are considered: 3D affine transformations and free-form deformations (FFD) parameterized using cubic B-splines [63]. The 3D affine transformation is defined as

$$\mathbf{T}^{\text{aff}}(\mathbf{r} | \mathbf{A}, \mathbf{v}) = \mathbf{A}\mathbf{r} + \mathbf{v}, \quad (2.19)$$

where $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ is the affine matrix and $\mathbf{v} \in \mathbb{R}^{3 \times 1}$ is the shift vector. This results in $3 \times 3 + 3 = 12$ parameters. The free-form deformation is defined as

$$\mathbf{T}^{\text{FFD}}(\mathbf{r} | \mathbf{c}^x, \mathbf{c}^y, \mathbf{c}^z) = \mathbf{r} + \begin{pmatrix} \mathbf{b}^x(\mathbf{r})\mathbf{c}^x \\ \mathbf{b}^y(\mathbf{r})\mathbf{c}^y \\ \mathbf{b}^z(\mathbf{r})\mathbf{c}^z \end{pmatrix}, \quad (2.20)$$

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 $\mathbf{c}^p \in \mathbb{R}^{N_c^p \times 1}$ 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.

Optimization

Solving the optimization problem in (2.18) 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

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - [\mathbf{H}\mathbf{F}(\boldsymbol{\theta}^{(j)})]^{-1} \nabla \mathbf{F}(\boldsymbol{\theta}^{(j)}), \quad j \geq 0, \quad (2.21)$$

are performed. Here \mathbf{H} denotes the Hessian, ∇ denotes the gradient, and the superscript (j) denotes the iteration index. In this work the interior-point method was combined with an L-BFGS [66] 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.

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 (2.5)) which warp the reference image q_0 to the dynamic object through (2.12). Following [67], the inverse property of \mathbf{T} can be rewritten as a relation between the displacements:

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

We follow [67] again and perform fixed-point iterations to compute $\boldsymbol{\delta}$:

$$\boldsymbol{\delta}^{(0)}(\mathbf{r}) = 0, \quad (2.22)$$

$$\boldsymbol{\delta}^{(j)}(\mathbf{r}) = -\boldsymbol{\eta}(\mathbf{r} + \boldsymbol{\delta}^{(j-1)}(\mathbf{r})), \quad j \in \mathbb{N}. \quad (2.23)$$

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 (2.22)-(2.23) 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.

2.3 Methods

Signal model validation

A numerical phantom with known ground-truth forward and inverse motion-fields was used to validate the proposed signal model (2.14). The phantom consists of a main spherical compartment, filled with three smaller disjoint elliptical compartments. The forward (\mathbf{U}_t^{ph}) and inverse (\mathbf{T}_t^{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 $3.0 \times 3.0 \times 3.0 \text{ mm}^3$.

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^{ph} using (2.12). The signal from the deformed phantom was computed from (2.1) and compared with the signal obtained from the forward model (2.14) using the ground-truth motion-fields \mathbf{T}_1^{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 [68] (see right panel Figure 2.4 for a visualization). This trajectory will in practice be relatively insensitive to intra-acquisition motion [69] 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.

Analysis of spin flow assumption

To derive the final signal model (2.14) 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 2.3 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 (2.20) 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.

2

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.

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 2.3 and compared with the analytic ground-truth motion-fields \mathbf{T}_t^{ph} . The maximum displacement in this case was 15mm. 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 right panel Figure 2.4 for a visualization) was projected on the Cartesian grid with a nearest neighbor interpolation, resulting in a pseudo-cone trajectory. The snapshot data was simulated on the Cartesian points. The cubic B-spline motion model (2.20) 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

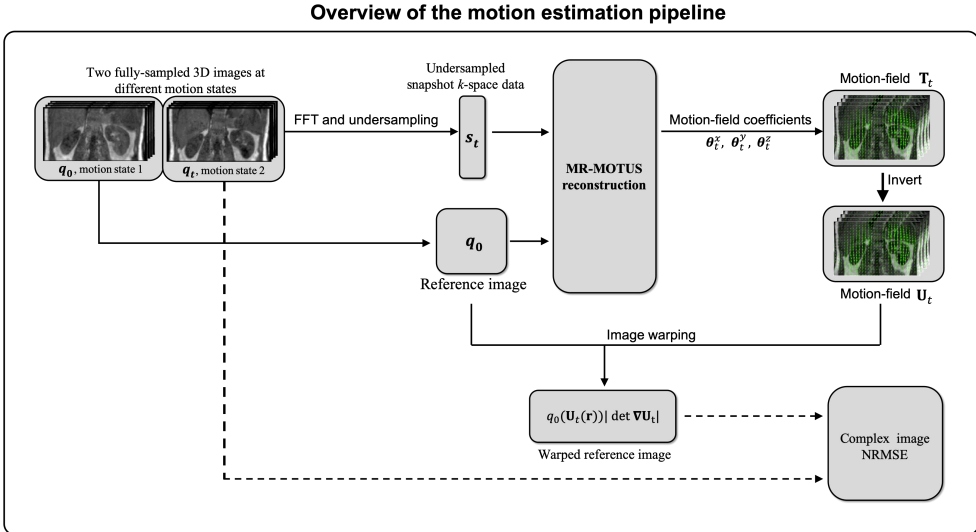


Fig. 2.3: An overview of the motion estimation pipeline described in Section 2.3. One of two fully-sampled images, each acquired at a different motion state (e.g. different position in the respiratory cycle), is transformed to k -space and retrospectively undersampled to simulate the snapshot k -space 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.

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.

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 under-sampled 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 (2.22)-(2.23) to obtain \mathbf{U}_t ;

Parameter	Retrospective <i>in vivo</i>		Prospective <i>in vivo</i>
	Head	Abdomen	Abdomen
FOV [m]	$0.25 \times 0.25 \times 0.13$	$0.28 \times 0.34 \times 0.34$	$0.48 \times 0.48 \times 0.010$
Acquisition size	$144 \times 144 \times 74$	$94 \times 128 \times 128$	$160 \times 160 \times 1$
Spatial resolution [mm]	$1.74 \times 1.74 \times 1.81$	$3.00 \times 2.70 \times 2.70$	$3.00 \times 3.00 \times 10.00$
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 SPGR	3D SPGR	2D Golden Angle
Scanner	Philips Ingenia 1.5T	Philips Ingenia 1.5T	Philips Ingenia 1.5T
Motion model	Affine	Cubic B-Splines	Cubic B-Splines

Table 2.1: Details of the *in vivo* experiments as described in Section 2.3

3. Warp the fully-sampled reference image \mathbf{q}_0 using (2.12) 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 2.3 for an overview of this workflow. The 3D affine motion model (2.19) was employed for the rigid head motion estimation, and the cubic B-spline motion model (2.20) 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 (2.22)-(2.23) 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 (SPGR) 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. 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 2.1.

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 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 under-sampled snapshot data and registered with optical flow image registration software [61, 62]. We refer to [61, 62] for more details on the optical flow method. To obtain the images, zero-filling reconstructions were performed on the 3D Cartesian down-sampled data, and compressed sensing reconstructions were performed for all other undersampling patterns (see Section 2.3) using the BART toolbox [70]. 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 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.

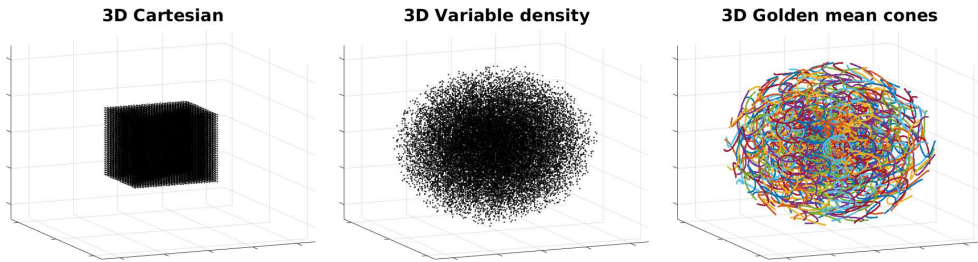


Fig. 2.4: Visualization of the undersampling trajectories with undersampling factor 63 for the three strategies discussed in Section 2.3 and reported in Table 2.4.

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 [68] (see Figure 2.4 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 2.3 for the motivation.

A 3D variable density undersampling scheme was included to facilitate an undersampling that is ideal for compressed sensing [71]. 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 [68] 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 [69]. 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 neighbor interpolation. Overlapping points after the projection were only used once.

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.

2

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 using a 2D golden angle radial acquisition with a total scan time of 15.84 seconds. Sequence parameters and other details can be found in Table 2.1. 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 seconds and ended with a free-breathing phase of 12.67 seconds. 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 seconds of free-breathing and these were used as snapshot data. The cubic B-spline model (2.20) with order 40 in every direction was employed, and motion-fields were reconstructed from the reference image and snapshot data by solving (2.18) and subsequently applying (2.22)-(2.23). The number of reconstruction iterations was fixed to 30 and the total reconstruction time was about 6 seconds. Similar to the other reconstructions presented in this work, we have employed the curvature regularization (2.17). The effect of this regularization was analyzed by comparing reconstructions with different

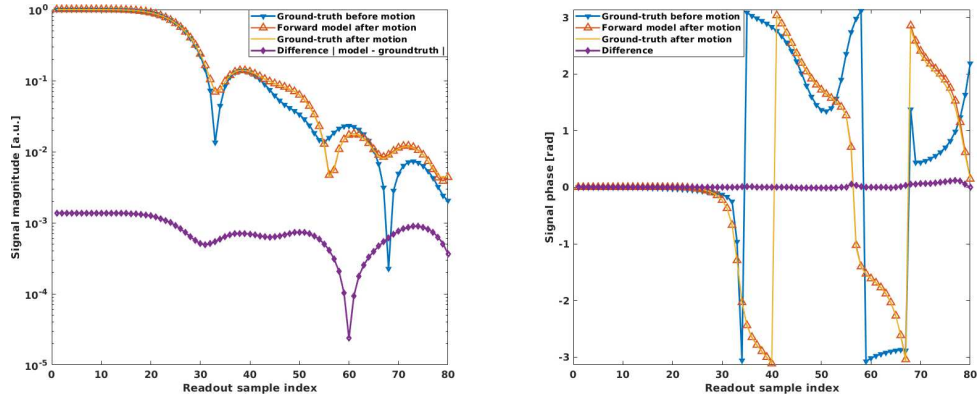


Fig. 2.5: Results of the signal model validation in Section 2.3 and 2.4. Plots of the magnitude (left) and phase (right) of the signal before deformation (blue), true signal after deformation (yellow), the signal calculated with the proposed signal model (2.14) (red), and the difference between the model’s 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.

2

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 124ms/frame. The image reconstruction was performed with the BART toolbox [70] 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.

2.4 Results

Signal model validation

The results of the validation are shown in Figure 2.5. 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.

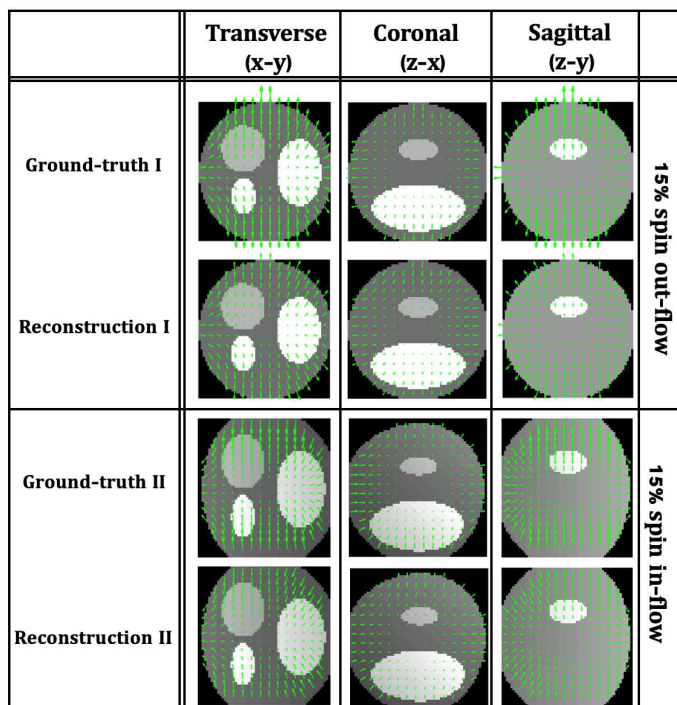


Fig. 2.6: Motion-field reconstructions in case the spin flow assumption is violated, as described in Section 2.3 and 2.4. Two cases are considered: 15% spin out-flow (top) and 15% spin in-flow (bottom).

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 2.6 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.

Undersampling factor	No noise				Gaussian noise, SNR \approx 80			
	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 2.2: Convergence analysis with analytic motion-fields as discussed in Section 2.3 and 2.4. The reported numbers are the RMSE between the reconstructed motion-fields and ground-truth motion-fields. The phantom resolution is 3mm³ and the maximum ground-truth displacement is 15 mm.

Motion estimation on retrospectively undersampled data

In silico motion estimation

The results on in silico motion reconstruction are presented in Table 2.2. 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 mm³. It can also be observed that the reconstructions are very robust 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 2.2 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.

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 minutes. In Table 2.3 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 inde-

Downsampling factor	3D Cartesian downsampling				
	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

Table 2.3: Results of the 3D rigid head motion estimation as discussed in Section 2.3 and 2.4. 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. In this table, ‘downsampling’ refers to the potential downsampling factors, if the data were to be acquired with an optimal trajectory. The downsampling factors 1,8,66,474 and 2551 correspond to 1534464,186624,3240 and 600 k -space points respectively.

pendent snapshot data, and are not simultaneously reconstructed. The results of this qualitative comparison are shown in Supporting Videos 2.1-2.3.[‡]

The quantitative results in Table 2.3 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 Supporting Videos 2.1-2.3 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 (Supporting Video 2.3) 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.

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 minutes. 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 Supporting

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

Undersampling factor	3D Cartesian				3D Variable density			3D Golden mean cones		
	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

Table 2.4: Results of the 3D respiratory motion estimation as discussed in Section 2.3 and 2.4. 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.

Videos 2.4 and 2.5.[‡] Moreover, the reconstructions for several undersampling factors and undersampling strategies are quantitatively compared with IR+OF in Table 2.4. Additionally, a visual comparison between MR-MOTUS and IR+OF reconstructions is provided in Figure 2.7 for 3D variable density undersampling with an undersampling factor of 63.

Results in Table 2.4 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 2.7 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.

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

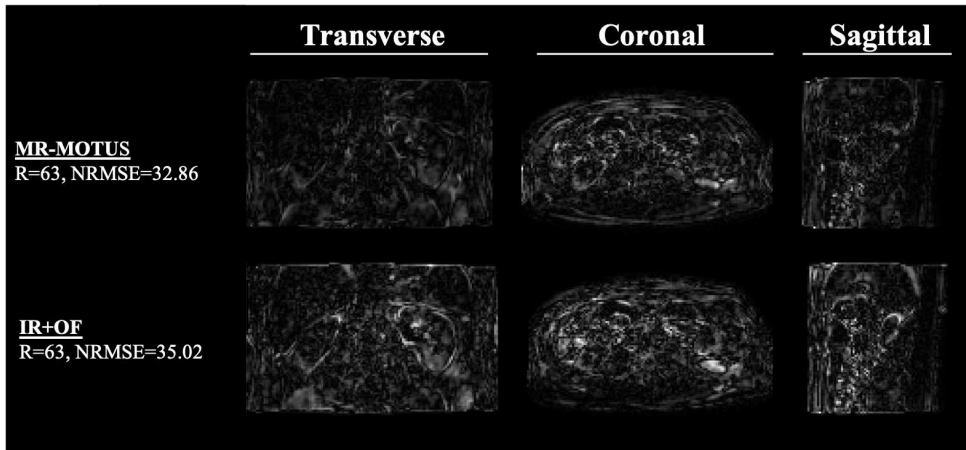


Fig. 2.7: Comparison between MR-MOTUS (top) and IR+OF (bottom) reconstruction quality for a specific case reported in Table 2.4; black is similar and white is different than the ground-truth. In this case 3D variable density with an undersampling factor of 63 was applied. The IR+OF clearly shows more quality degrading in comparison with MR-MOTUS.

2

Motion estimation on prospectively undersampled *in vivo* data

Figure 2.8 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 doesn't 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 2.9, 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.

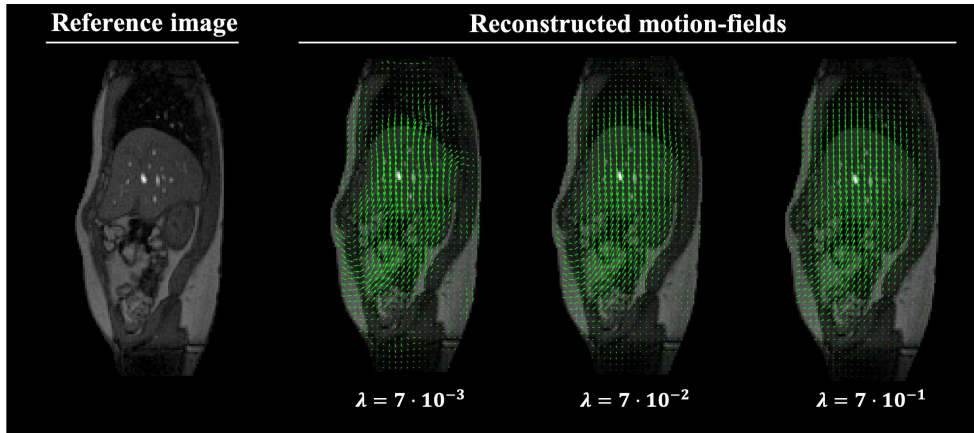


Fig. 2.8: Motion-fields reconstructed from prospectively undersampled data acquired during free breathing, as described in Section 2.3. Three different degrees of regularization are considered: $\lambda = 7 \cdot 10^{-3}$, $7 \cdot 10^{-2}$, $7 \cdot 10^{-1}$. See also Figure 2.9 for a qualitative analysis of the error in the reconstruction.

2

2.5 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 [52, 60]. We formally show 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 (2.13) to the excited FOV Ω . This assumption also appears in other works for similar reasons [52]. 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 2.7). 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

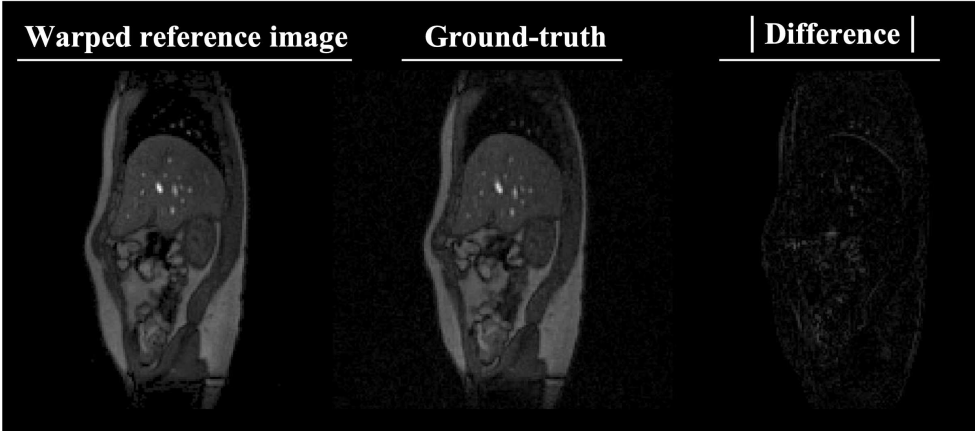


Fig. 2.9: A visual comparison between the reference image warped with the MR-MOTUS motion-field reconstructed with $\lambda = 7 \cdot 10^{-2}$, and the ground-truth image selected from a dynamic CS reconstruction, as described in Section 2.3.

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 1.5T.

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 60ms, 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 [69]. 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 $(128 \times 128 \times 94) / 63 \approx 24446$ 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 [68]. Hence, with a TR of 4ms and 1000 points per read-out, the acquisition time of the snapshot data can be approximated as $24446 / 1000 \cdot 4 = 98$ 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 Chapter is currently about 5 minutes, 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.* [64]. 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.* [64] 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 [36, 72].

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 is 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 2.8). 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 [41]. 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.* [73]. Applying MR-MOTUS in these settings will be considered for future research.

2.6 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.

2

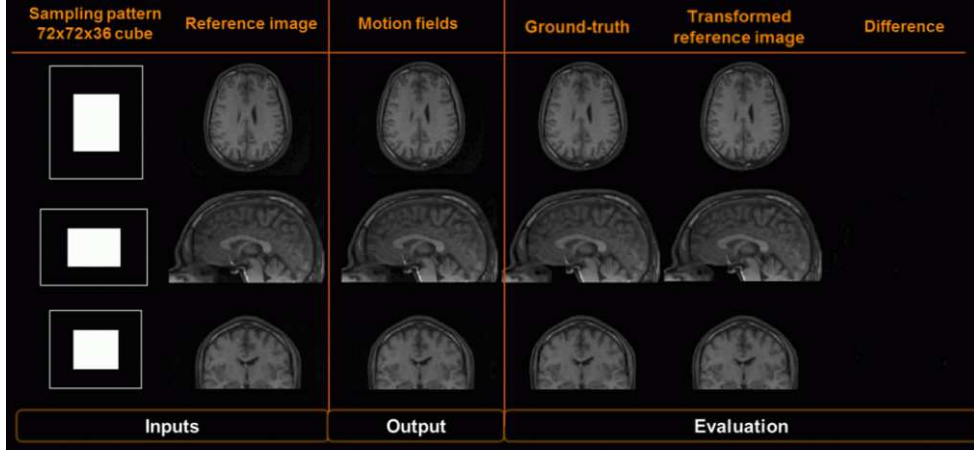
Acknowledgements

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 this chapter.

2.7 Supporting Information

Supporting Videos

The supporting videos are part of the main body of this chapter and can be found through the QR code below the abstract, or through the following link: <https://doi.org/10.6084/m9.figshare.20480550.v2>.



Supporting Figure 2.1: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480550.v2>.* 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 186624 k -space points per snapshot and potential effective total downsampling factor of about 8.2.

2

Detailed signal model derivation

Under the assumptions made in the main text (see Section 2.2) the steady-state equilibrium of all individual spins remains nearly constant during motion. Hence, the transverse magnetization at time t of the spins at location \mathbf{r}_t can be written as the transverse magnetization at time 0 of the same spin before deformation by \mathbf{U}_t . Defining the support of m_t as Σ_j , the observation above can mathematically be summarized as (see Supporting Figure 2.6)

$$m_t(\mathbf{r}_t) = m_0(\mathbf{U}_t(\mathbf{r}_t)), \quad \mathbf{r}_t \in \mathbf{T}_t(\Sigma_0) \cap \Sigma_t. \quad (\text{S2.1})$$

From the assumption on the conservation of spins it was derived in the main text that the following must hold for all $t = 0, 1, \dots$ and $\mathbf{r} \in \mathbb{R}^3$:

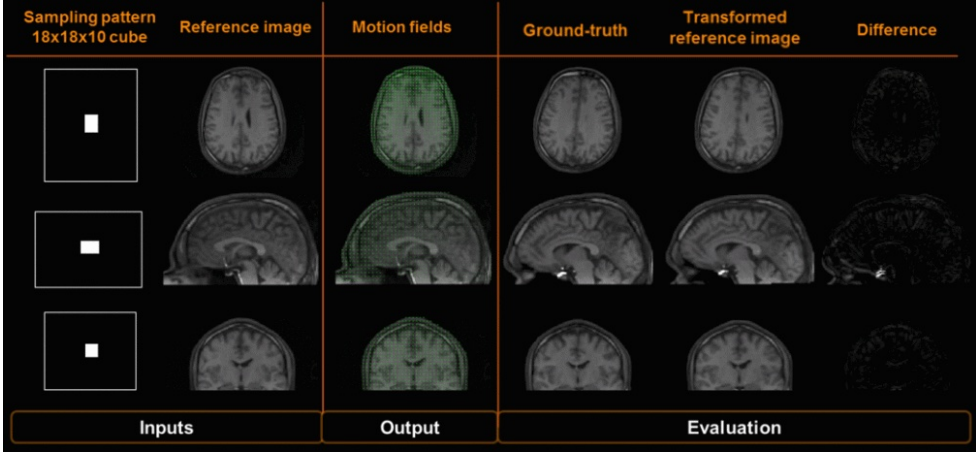
$$\rho_t(\mathbf{r}_t) d\mathbf{r}_t = \rho_0(\mathbf{U}_t(\mathbf{r}_t)) |\det(\nabla \mathbf{U}_t)(\mathbf{r}_t)| d\mathbf{r}_t. \quad (\text{S2.2})$$

Combining (S2.1) and (S2.2) into one equation, with $q_t \equiv \rho_t \cdot m_t$, yields,

$$q_t(\mathbf{r}_t) d\mathbf{r}_t = q_0(\mathbf{U}_t(\mathbf{r}_t)) |\det(\nabla \mathbf{U}_t)(\mathbf{r}_t)| d\mathbf{r}_t, \quad (\text{S2.3})$$

for $\mathbf{r}_t \in \mathbf{T}_t(\Sigma_0) \cap \Sigma_t$. The signal at time t is given as

$$s_t(\mathbf{k}) = \int_{\Sigma_t} q_t(\mathbf{r}_t) e^{-i2\pi \mathbf{k} \cdot \mathbf{r}_t} d\mathbf{r}_t, \quad (\text{S2.4})$$



Supporting Figure 2.2: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480550.v2>.* 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 potential effective total downsampling factor of about 474.

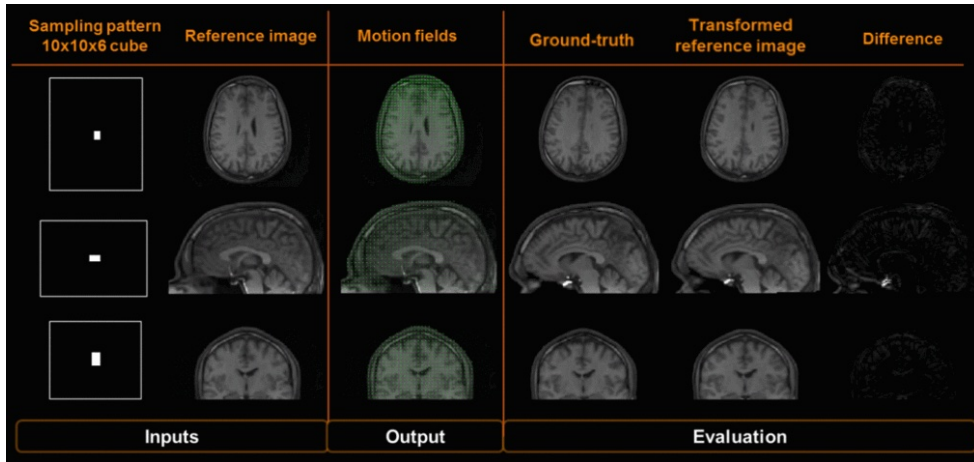
hence the substitution of (S2.3) into (S2.4) is only valid when $\Sigma_t \subseteq \mathbf{T}_t(\Sigma_0)$ holds. In other words, the substitution is valid when there is no in-flow of new spins between time 0 and time t . Under that assumption we get

$$s_t(\mathbf{k}) = \int_{\Sigma_t} q_0(\mathbf{U}_t(\mathbf{r}_t)) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}_t} |\det(\nabla\mathbf{U}_t)(\mathbf{r}_t)| d\mathbf{r}_t,$$

and after the change of variables $\mathbf{r}_t \mapsto \mathbf{T}_t(\mathbf{r}_0)$ we obtain the general signal model

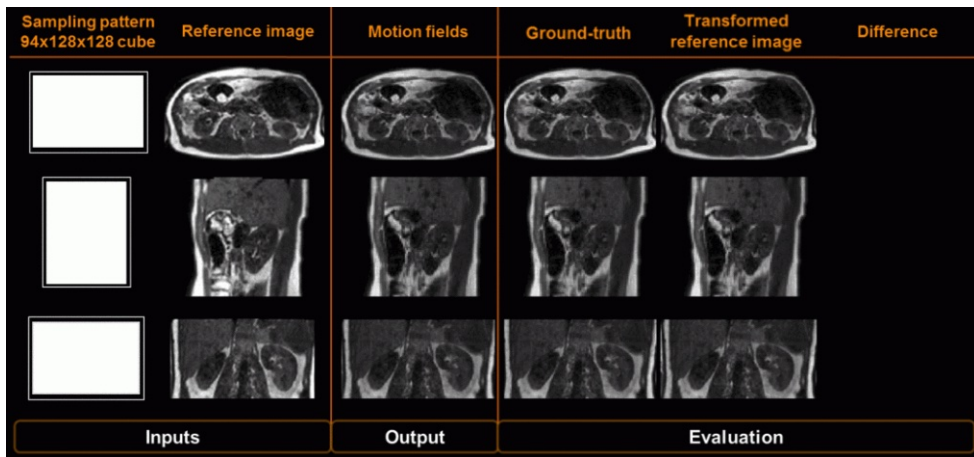
$$s_t(\mathbf{k}) = \int_{\mathbf{U}_t(\Sigma_t)} q_0(\mathbf{r}_0) e^{-i2\pi\mathbf{k}\cdot\mathbf{T}_t(\mathbf{r}_0)} d\mathbf{r}_0. \quad (\text{S2.5})$$

Note that the domain of integration in (2.13) depends on the (unknown) motion-field \mathbf{U}_t , which is inconvenient in practice. If $\Sigma_0 \subseteq \mathbf{U}_t(\Sigma_t)$, i.e. the support of q_0 is contained in $\mathbf{U}_t(\Sigma_t)$, then the integration domain can be changed to Σ_0 . However, since the substitution of (S2.3) into (S2.4) already assumes $\Sigma_t \subseteq \mathbf{T}_t(\Sigma_0)$, this implies that $\Sigma_0 = \mathbf{U}_t(\Sigma_t)$ must hold. In other words, the general model (S2.5) and the model (2.14) are equivalent when always the exact same spins are excited for the reference and all subsequent snapshots (see Supporting Figure 2.6 for an illustration), i.e. when no signal-contributing spins flow in or out of the FOV.

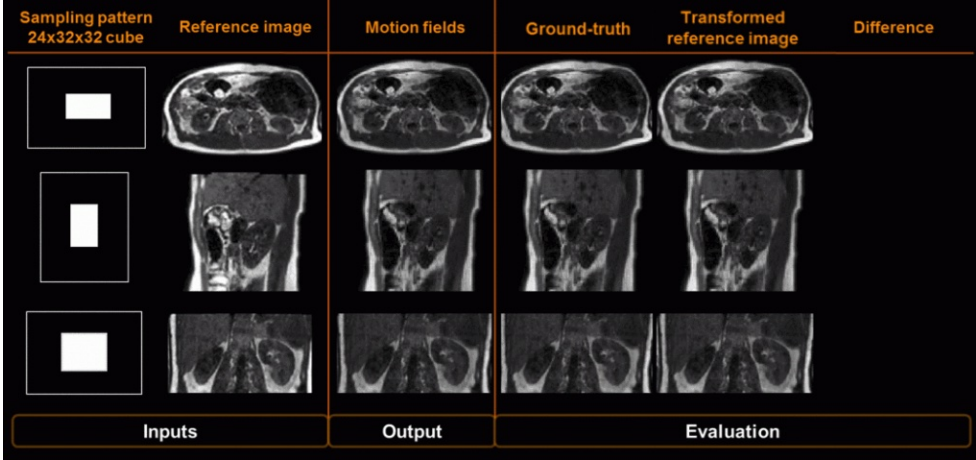


Supporting Figure 2.3: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480550.v2>. Affine motion reconstruction from retrospectively downsampled in vivo head data using MR-MOTUS, compared with the ground-truth image reconstructions. The 3D Cartesian technique was employed for the downsampling. A downsampling factor of 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 potential effective total downsampling factor of 2551.*

2



Supporting Figure 2.4: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480550.v2>. 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.*



Supporting Figure 2.5: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480550.v2>.* 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 directions, resulting in a potential effective total downsampling factor of 63.

2

Reconstruction implementation details

Motion models

The motion models are implemented in matrix form as

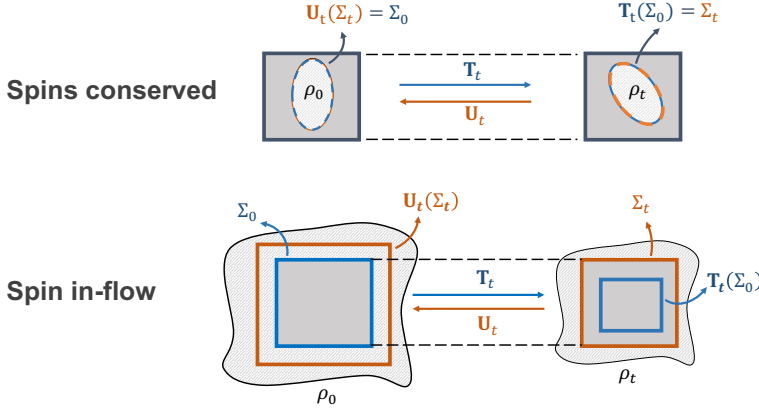
$$\mathbf{R}_t = \mathbf{R}_0 + \mathbf{B}\boldsymbol{\theta}_t = \begin{pmatrix} \mathbf{R}_0^x \\ \mathbf{R}_0^y \\ \mathbf{R}_0^z \end{pmatrix} + \begin{pmatrix} \mathbf{B}^x & & \\ & \mathbf{B}^y & \\ & & \mathbf{B}^z \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_t^x \\ \boldsymbol{\theta}_t^y \\ \boldsymbol{\theta}_t^z \end{pmatrix}.$$

Here \mathbf{R}_0 and \mathbf{R}_t denote the vectors with coordinates respectively at time 0 and after deformation by \mathbf{T}_t at time t , \mathbf{B} is the matrix determined by the motion model, and $\boldsymbol{\theta}_t$ is the vectors of unknowns. Both \mathbf{R}_0 and \mathbf{R}_t are obtained by first concatenating vertically per dimension and then stacking the dimensions vertically. Hence, for a total number of N_r voxels we have $\mathbf{R}_0, \mathbf{R}_t \in \mathbb{R}^{3N_r}$. The affine model (2.19) can be obtained by setting

$$\mathbf{B}^x = \mathbf{B}^y = \mathbf{B}^z = [\mathbf{R}_0^x \quad \mathbf{R}_0^y \quad \mathbf{R}_0^z \quad \mathbf{1}],$$

$$\begin{pmatrix} \boldsymbol{\theta}_t^x \\ \boldsymbol{\theta}_t^y \\ \boldsymbol{\theta}_t^z \end{pmatrix} = \text{vec} \left\{ \begin{pmatrix} \mathbf{A}_t^T - \mathbf{I} \\ \mathbf{v}_t^T \end{pmatrix} \right\},$$

where $\boldsymbol{\theta}_t^p \in \mathbb{R}^{4 \times 1}$, $\mathbf{1} \in \mathbb{R}^{N_r \times 1}$ denotes an all-one vector, and $\text{vec}(\cdot)$ the vectorization function. For the spline model (2.20) we get $\mathbf{B}^p \in \mathbb{R}^{N_r \times N_c^p}$ as B-spline basis matrices



Supporting Figure 2.6: Illustration of objects (light-gray) undergoing deformation using the notation introduced in Section 2.7: (top) a spin-conserving deformation and (bottom) a deformation with in-flowing spins. The dark-gray squares denote the excited regions at time 0 (left) and time t (right). $\mathbf{T}_t(\Sigma_0)$ denotes the new locations after deformation at time t , of spins in Σ_0 that were originally excited for the reference image. In this area the signal can be described with the model (2.13). $\mathbf{U}_t(\Sigma_t)$ are the source locations of the spins that are excited for the t -th snapshot image. If this area does not coincide with Σ_0 , then the signal from the area between Σ_0 and $\mathbf{U}_t(\Sigma_t)$ cannot be explained by the model. A similar observation can be made for deformations with spin out-flow, except that in that case $\mathbf{U}_t(\Sigma_t)$ will lie within Σ_0 .

with vertically concatenated $\mathbf{b}^p(\mathbf{r}) \in \mathbb{R}^{1 \times N_z^p}$ as entries. The $\boldsymbol{\theta}_t^p \in \mathbb{R}^{N_z^p \times 1}$ denote the basis coefficients.

Gradient of the objective function

To compute the gradient with respect to the objective function we first write the forward model in a way that is more convenient for differentiation:

$$\mathbf{F}(\boldsymbol{\theta}_t \mid \mathbf{q}_0) = \exp \left\{ -i2\pi \mathbf{K} [\mathbf{R}_t^x \quad \mathbf{R}_t^y \quad \mathbf{R}_t^z]^T \right\} \mathbf{q}_0,$$

where $\mathbf{K} \in \mathbb{R}^{M \times 3}$ is the matrix with all k -space coordinates of the applied trajectory, $\exp\{\cdot\}$ denotes element-wise application of the exponential function, and \mathbf{q}_0 denotes the vectorization of q_0 . Similarly, the curvature regularization function can be written as a simple L^2 -norm of a matrix-vector product by applying a finite difference scheme to discretize the derivatives, i.e.

$$\begin{pmatrix} \Delta \mathbf{T}_t^x \\ \Delta \mathbf{T}_t^y \\ \Delta \mathbf{T}_t^z \end{pmatrix} \approx \begin{pmatrix} \mathbf{L} & & \\ & \mathbf{L} & \\ & & \mathbf{L} \end{pmatrix} \begin{pmatrix} \mathbf{R}_0^x + \mathbf{B}^x \boldsymbol{\theta}_t^x \\ \mathbf{R}_0^y + \mathbf{B}^y \boldsymbol{\theta}_t^y \\ \mathbf{R}_0^z + \mathbf{B}^z \boldsymbol{\theta}_t^z \end{pmatrix} = \mathbf{L}(\mathbf{R}_0 + \mathbf{B}\boldsymbol{\theta}_t).$$

Here $\Delta \mathbf{T}^p$, $\mathbf{T}^p \in \mathbb{R}^{N_r \times 1}$ are vectorizations of the corresponding functions, and $\mathbf{L} \in \mathbb{R}^{N_r \times N_r}$ is defined as the matrix that applies the discretized Laplace operator to a

vector by left-multiplication. The discretization of the spatial derivatives was obtained using a central difference scheme with Neumann boundary conditions. Using this, we get

$$\mathcal{R}(\mathbf{T}_t) = \|\mathbf{L}(\mathbf{R}_0 + \mathbf{B}\boldsymbol{\theta}_t)\|_2^2.$$

The total objective function can now be written in discretized form as

$$E(\boldsymbol{\theta}_t) = \|\exp\left\{-i2\pi\mathbf{K}\tilde{\mathbf{R}}_t^T\right\}\mathbf{q}_0 - \mathbf{s}\|_2^2 + \lambda\|\mathbf{L}\mathbf{R}_t\|_2^2,$$

where we have defined $\tilde{\mathbf{R}}_t := [\mathbf{R}_t^x \ \mathbf{R}_t^y \ \mathbf{R}_t^z]$ as the new coordinates at time t dependent on $\boldsymbol{\theta}_t$. The Jacobian \mathbf{J}_F is defined as

$$\mathbf{J}_F := \left[\frac{\partial \mathbf{F}}{\partial \theta_t^x}, \frac{\partial \mathbf{F}}{\partial \theta_t^y}, \frac{\partial \mathbf{F}}{\partial \theta_t^z} \right] \in \mathbb{C}^{N_k \times N_c}.$$

Using matrix differentiation we can derive the following formulas for the derivatives

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial \theta_t^p} &= -i2\pi \text{diag}(\mathbf{k}^p) \cdot \dots \\ &\quad \exp\left\{-i2\pi\mathbf{K}\tilde{\mathbf{R}}_t^T\right\} \cdot \dots \\ &\quad \text{diag}(\mathbf{q}_0)\mathbf{B}^p, \end{aligned} \tag{S2.6}$$

$$\begin{aligned} \left(\frac{\partial \mathbf{F}}{\partial \theta_t^p}\right)^* &= i2\pi (\mathbf{B}^p)^T \text{diag}(\bar{\mathbf{q}}_0) \cdot \dots \\ &\quad \exp\left\{i2\pi\tilde{\mathbf{R}}_t\mathbf{K}^T\right\} \cdot \dots \\ &\quad \text{diag}(\mathbf{k}^p). \end{aligned} \tag{S2.7}$$

Note the complex conjugate $\bar{\mathbf{q}}_0$ and the change in signs of the imaginary variables in (S2.7). The gradient of the objective function is readily computed as

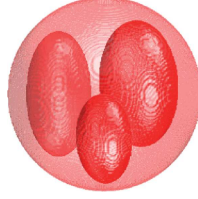
$$\frac{\partial E}{\partial \boldsymbol{\theta}_t} = \begin{pmatrix} \partial E / \partial \theta_t^x \\ \partial E / \partial \theta_t^y \\ \partial E / \partial \theta_t^z \end{pmatrix} = 2\text{Re} \{(\mathbf{J}_F)^*(\mathbf{F} - \mathbf{s}_t)\} + 2\lambda\mathbf{B}^T\mathbf{L}^T\mathbf{L}\mathbf{R}_t.$$

Here the superscript star denotes the Hermitian adjoint.

Jacobian-vector products

In the optimization algorithm frequent computations of the products $\mathbf{J}_F\mathbf{w}_1$ and $(\mathbf{J}_F)^*\mathbf{w}_2$ are required. By definition

$$\mathbf{J}_F\mathbf{w}_1 = \sum_{p \in \{x, y, z\}} \left(\frac{\partial \mathbf{F}}{\partial \theta_t^p}\right) \mathbf{w}_1^p, \quad (\mathbf{J}_F)^*\mathbf{w}_2 = \begin{pmatrix} \left(\frac{\partial \mathbf{F}}{\partial \theta_t^x}\right)^* \mathbf{w}_2 \\ \left(\frac{\partial \mathbf{F}}{\partial \theta_t^y}\right)^* \mathbf{w}_2 \\ \left(\frac{\partial \mathbf{F}}{\partial \theta_t^z}\right)^* \mathbf{w}_2 \end{pmatrix}.$$



Supporting Figure 2.7: 3D visualization of the spherical analytical phantom.

Here we have defined \mathbf{w}_1^p such that $\mathbf{w}_1 := \begin{pmatrix} \mathbf{w}_1^x \\ \mathbf{w}_1^y \\ \mathbf{w}_1^z \end{pmatrix}$. Using (S2.6)-(S2.7) we can compute

$$\mathbf{J}_F \mathbf{w}_1 = -i2\pi \sum_{p \in \{x, y, z\}} \left\{ \text{diag}(\mathbf{k}^p) \exp \left\{ -i2\pi \mathbf{K} \tilde{\mathbf{R}}_t^T \right\} \cdot \dots \right. \\ \left. \text{diag}(\mathbf{q}) \mathbf{B}^p \mathbf{w}_1^p \right\}$$

and

$$(\mathbf{J}_F)^* \mathbf{w}_2 = i2\pi \begin{pmatrix} (\mathbf{B}^x)^T \text{diag}(\bar{\mathbf{q}}) \exp \left\{ i2\pi \tilde{\mathbf{R}}_t \mathbf{K}^T \right\} \text{diag}(\mathbf{k}^x) \mathbf{w}_2 \\ (\mathbf{B}^y)^T \text{diag}(\bar{\mathbf{q}}) \exp \left\{ i2\pi \tilde{\mathbf{R}}_t \mathbf{K}^T \right\} \text{diag}(\mathbf{k}^y) \mathbf{w}_2 \\ (\mathbf{B}^z)^T \text{diag}(\bar{\mathbf{q}}) \exp \left\{ i2\pi \tilde{\mathbf{R}}_t \mathbf{K}^T \right\} \text{diag}(\mathbf{k}^z) \mathbf{w}_2 \end{pmatrix}.$$

Both products were implemented efficiently using the forward and backward type-3 NUFFT from Barnett *et al.* [64], respectively.

The analytical deformation phantom

Phantom

The phantom is defined by: $f(x, y, z) = \mathbf{1}_A + \mathbf{1}_B + \frac{1}{2}\mathbf{1}_C + \mathbf{1}_D$ where $\mathbf{1}_S$ denotes the indicator function corresponding to the set S . The sets A, B, C and D represent, respectively, the main spherical compartment of radius R and the three disjoint smaller ellipsoidal structures inside A (see Supporting Figure 2.7). They are defined as:

$$\begin{aligned} A &= \{(x, y, z) \mid \sqrt{x^2 + y^2 + z^2} \leq R\}; \\ B &= \{(x, y, z) \mid \sqrt{2(x + \frac{R}{3})^2 + (y + \frac{R}{3})^2 + \frac{1}{2}(z + \frac{R}{3})^2} \leq \frac{R}{3}\}; \\ C &= \{(x, y, z) \mid \sqrt{(x + \frac{R}{3})^2 + (y - \frac{R}{3})^2 + \frac{1}{4}z^2} \leq \frac{R}{3}\}; \\ D &= \{(x, y, z) \mid \sqrt{4(x - \frac{R}{2})^2 + 2y^2 + z^2} \leq \frac{2R}{3}\}. \end{aligned}$$

For a visualization of the phantom, see Supporting Figure 2.7

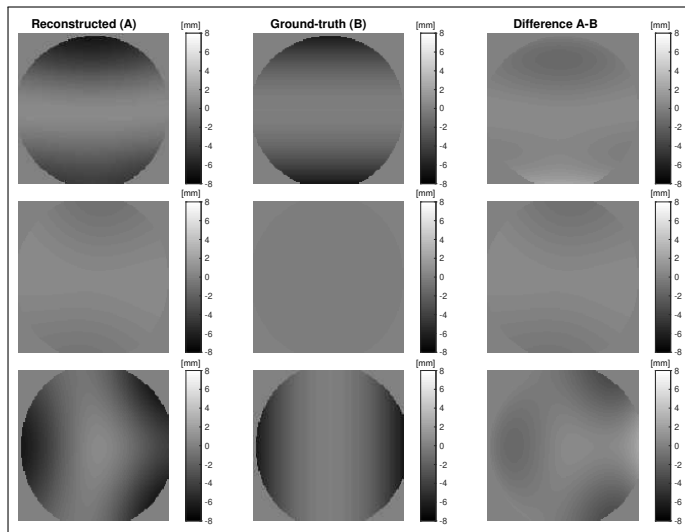
Invertible time-dependent transformation

A non-linear, time-dependent and invertible transformation, $\mathbf{U}^{\text{ph}} = (\mathbf{U}_x^{\text{ph}}, \mathbf{U}_y^{\text{ph}}, \mathbf{U}_z^{\text{ph}})$ is chosen such that (1) \mathbf{U}^{ph} is analytically invertible; (2) the determinant of the Jacobian is always and everywhere positive; (3) the rank of the matrix Φ , obtained by concatenating the motion-fields over time, is larger than 1. The following formulas define \mathbf{U}^{ph} and its inverse \mathbf{T}^{ph} such that $\mathbf{T}^{\text{ph}} \circ \mathbf{U}^{\text{ph}} = \mathbf{U}^{\text{ph}} \circ \mathbf{T}^{\text{ph}} = \mathbf{Id}$:

$$\begin{aligned} & \begin{cases} \mathbf{U}_x^{\text{ph}}(x, y, z) &= x - m\frac{\theta^2}{2}x^2; \\ \mathbf{U}_y^{\text{ph}}(x, y, z) &= y - m\theta y; \\ \mathbf{U}_z^{\text{ph}}(x, y, z) &= z + m\frac{\theta^2}{2}z^2; \end{cases} \\ \\ \text{If } \theta \neq 0 : & \begin{cases} \mathbf{T}_x^{\text{ph}}(x, y, z) &= \theta^{-2}m^{-1}(1 - S_x); \\ \mathbf{T}_y^{\text{ph}}(x, y, z) &= y(1 - m\theta)^{-1}; \\ \mathbf{T}_z^{\text{ph}}(x, y, z) &= \theta^{-2}m^{-1}(S_z - 1); \end{cases} \\ \\ \text{If } \theta = 0 : & \begin{cases} \mathbf{T}_x^{\text{ph}}(x, y, z) &= x; \\ \mathbf{T}_y^{\text{ph}}(x, y, z) &= y; \\ \mathbf{T}_z^{\text{ph}}(x, y, z) &= z. \end{cases} \end{aligned}$$

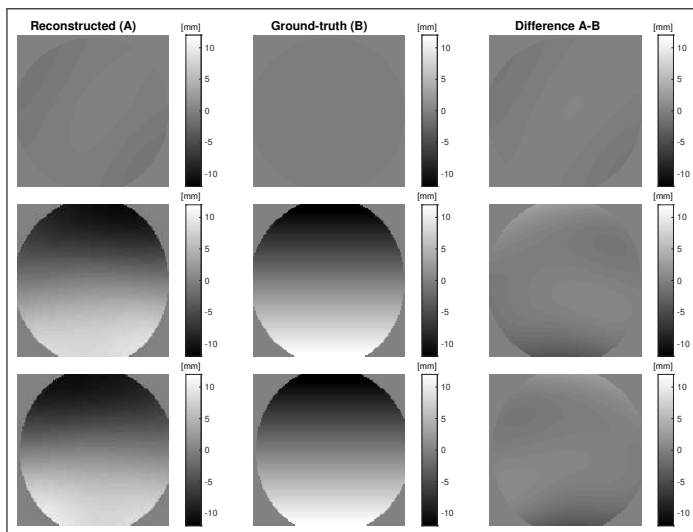
Here $S_x = \sqrt{1 - 2\theta^2 mx}$ and $S_z = \sqrt{1 + 2\theta^2 mz}$. The parameter m is fixed while θ is time varying. Note that for $\theta = 0$, \mathbf{U}^{ph} and \mathbf{T}^{ph} reduce to the identity transformation (no motion). The determinants of the corresponding Jacobian matrices are:

$$\begin{aligned} |\det(\nabla \mathbf{U}^{\text{ph}})| &= (1 - m\theta^2 x)(1 - m\theta)(1 + m\theta^2 z) \\ |\det(\nabla \mathbf{T}^{\text{ph}})| &= (S_x S_z (1 - m\theta))^{-1}. \end{aligned}$$

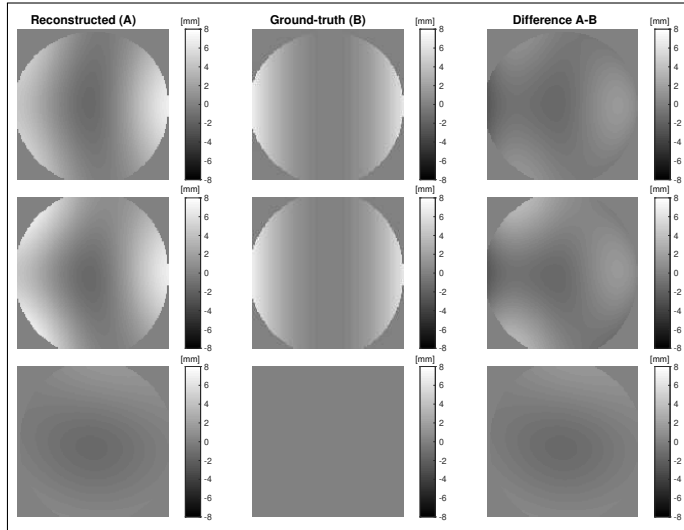


2

Supporting Figure 2.8: Motion-fields in x direction corresponding to the 15% spin in-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.

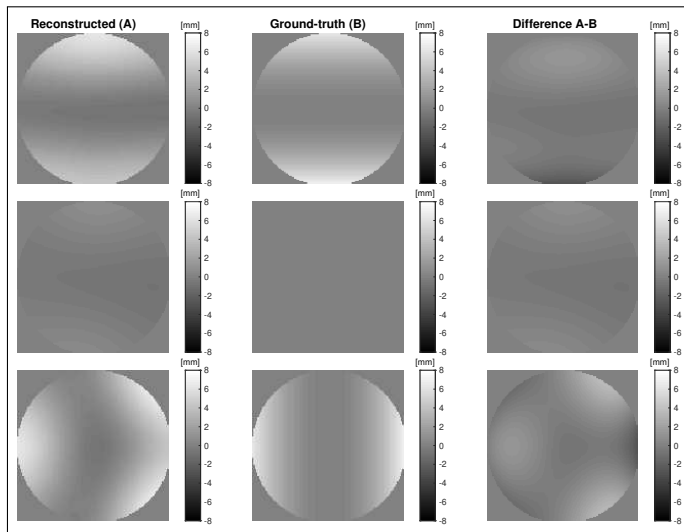


Supporting Figure 2.9: Motion-fields in y direction corresponding to the 15% spin in-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.

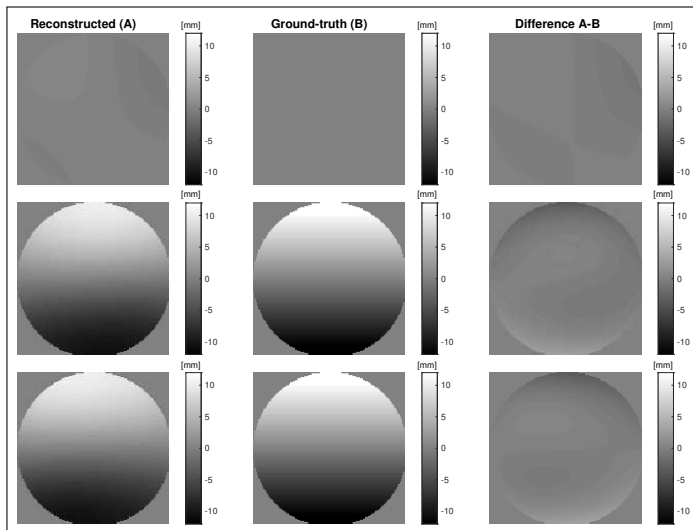


Supporting Figure 2.10: Motion-fields in z direction corresponding to the 15% spin in-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.

2

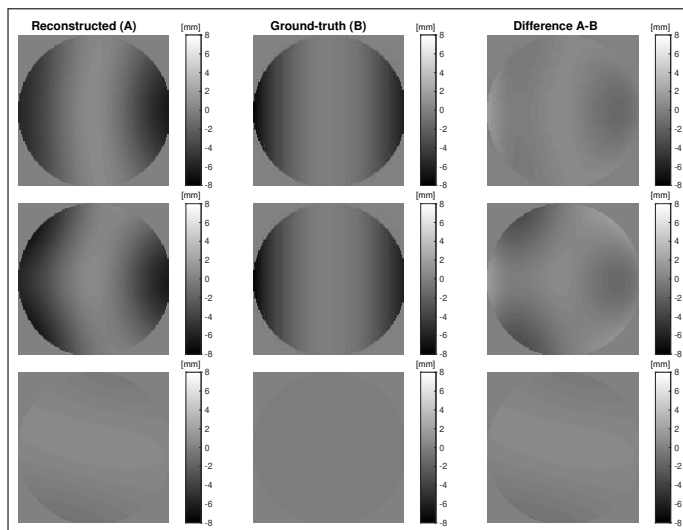


Supporting Figure 2.11: Motion-fields in x direction corresponding to the 15% spin out-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.



2

Supporting Figure 2.12: Motion-fields in y direction corresponding to the 15% spin out-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.



Supporting Figure 2.13: Motion-fields in z direction corresponding to the 15% spin out-flow experiment in Fig. 5 of this chapter. The columns correspond to (left) reconstructed motion-fields, (middle) ground-truth motion-fields, (right) difference between reconstruction and ground-truth.

Chapter 3

Non-rigid 3D motion estimation at high temporal resolution using low-rank MR-MOTUS

Niek R.F. Huttinga
Tom Bruijnen
Cornelis A. T. van den Berg
Alessandro Sbrizzi

The following chapter is based on:

Non-rigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS. *Magnetic Resonance in Medicine*, vol. 85, no. 4, p. 2309-2326, 2021.

Abstract

With the recent introduction of the MR-linac, an MR-scanner combined with a radiotherapy LINAC, MR-based motion estimation has become of increasing interest to (retrospectively) characterize tumor and organs-at-risk motion during radiotherapy. To this extent, we introduce low-rank MR-MOTUS, a framework to retrospectively reconstruct time-resolved non-rigid 3D+t motion-fields from a single low-resolution reference image and prospectively undersampled k-space data acquired during motion. Low-rank MR-MOTUS exploits spatio-temporal correlations in internal body motion with a low-rank motion model, and inverts a signal model that relates motion-fields directly to a reference image and k-space data. The low-rank model reduces the degrees-of-freedom, memory consumption and reconstruction times by assuming a factorization of space-time motion-fields in spatial and temporal components. Low-rank MR-MOTUS was employed to estimate motion in 2D/3D abdominothoracic scans and 3D head scans. Data were acquired using golden-ratio radial readouts. Reconstructed 2D and 3D respiratory motion-fields were respectively validated against time-resolved and respiratory-resolved image reconstructions, and the head motion against static image reconstructions from fully-sampled data acquired right before and right after the motion. Results show that 2D+t respiratory motion can be estimated retrospectively at 40.8 motion-fields-per-second, 3D+t respiratory motion at 7.6 motion-fields-per-second and 3D+t head-neck motion at 9.3 motion-fields-per-second. The validations show good consistency with image reconstructions. In conclusion, the proposed framework can estimate time-resolved non-rigid 3D motion-fields, which allows to characterize drifts and intra and inter-cycle patterns in breathing motion during radiotherapy, and could form the basis for real-time MR-guided radiotherapy.

Supporting Videos

The supporting videos corresponding to this chapter can be accessed through <https://doi.org/10.6084/m9.figshare.20480571.v1> or the following QR-code:



3.1 Introduction

Uncertainty in tumor and organs-at-risk locations due to unknown respiratory-induced organ motion diminishes the efficacy of radiotherapy in the abdomen and thorax in two ways. Firstly, tumors are irradiated with larger treatment margins, which results in increased radiation dose and toxicity to healthy tissue. Secondly, it prevents an accurate (retrospective) estimation of the actual dose accumulated in the targeted tumor and healthy surrounding tissue during the treatment.

Recently, the MR-linac was introduced as the combination of an MR-scanner and a linear accelerator (LINAC) in a single device [7–10], which has the potential to address both points above. Achieving this goal, however, poses the following technical challenge: real-time reconstructions at 5 Hz [22, 23] of internal body motion during the treatments. A fundamental step towards real-time reconstructions is the retrospective estimation of time-resolved motion-fields. Additionally, these retrospectively reconstructed motion-fields are valuable for the calculation of accumulated dose and can be taken into account for more accurate radiation planning of subsequent treatments. To this extent, we focus on the retrospective reconstruction of time-resolved 3D+t respiratory motion with a temporal resolution of 5 motion-fields-per-second. We envision that this framework could eventually be adapted to prospective real-time reconstructions [74].

In MR-guided radiotherapy, tumor and organs-at-risk motion is typically estimated from cine-MR-images followed by image registration. For time-resolved motion estimation, these cine-MR-images would thus require sufficient temporal resolution and spatial coverage to resolve the targeted motion. This is in general achievable in 2D, and also in 3D for slowly moving targets such as pelvic tumors [11]. However, in 3D it is more challenging for faster moving targets like lung tumors, that require at least 5 motion-fields-per-second [22, 23].

Several strategies have previously been proposed to extract tumor and organ-at-risk motion from MR-images, three of which will be reviewed below. With the first strategy, average respiratory motion is estimated from a respiratory-resolved 3D+t MRI. This approach retrospectively sorts image slices or k-space readouts in 3D acquisitions according to their respective respiratory phases, extracted using a respiratory motion surrogate (e.g. pneumatic belt, self-navigation signal or navigator). Examples include the works in Breuer *et al.* [75], Deng *et al.* [76], Han *et al.* [77], Cai *et al.* [78], and Feng *et al.* [79] (see Stemkens *et al.* [24] for a more complete overview). Although the retrospective sorting in these methods allows for efficient use of all acquired data, it makes strong assumptions on the periodicity of respiratory motion and characterizes only average 3D+t breathing motion. Although this is useful to reduce treatment margins, it may not be sufficient for accurate accumulation of the delivered dose.

A different strategy uses multi-slice/orthogonal 2D+t cine-MRI for 3D+t motion estimation [25, 32, 33, 35, 36, 43–45]. The reduction in the spatial dimension allows for higher temporal resolution, and is combined with a model that links the lower-dimensional image data to 3D motion-fields. This strategy assumes, however, that a good fit on lower-dimensional images implies a good fit in the full 3D domain. Although this is reasonable for small volumes, since slices cover a large fraction of

the volume in such a case, it may be less valid for larger volumes which may be required for dose accumulation.

The third strategy does not rely on sorting, but reconstructs images from highly undersampled k-space data. Even with parallel imaging [26, 27], this typically eventually results in lower SNR, lower spatial resolution, and/or undersampling artifacts. Nevertheless, it has been shown that motion-fields can be estimated from these images with sufficient accuracy [30, 48, 80, 81]. Additionally, iterative reconstructions based on compressed sensing [28] have been proposed to exploit the spatio-temporal sparsity of images. However, for the intended application the reported temporal resolution was too low [29–31], or the FOV was too small [82, 83].

Following a different approach, we have previously introduced MR-MOTUS [84] (Model-based Reconstruction of MOTion from Undersampled Signal), a new framework that allows to reconstruct non-rigid 3D motion-fields directly from k-space data. The key ingredient of MR-MOTUS is a signal model that explicitly relates dynamic k-space data to the combination of a static reference image and dynamic motion-fields. Assuming a reference image is available, and data is acquired in steady-state, motion-fields can be reconstructed directly from k-space data by solving the corresponding non-linear inverse problem. Since motion-fields are spatially correlated and therefore compressible, few data are required for the reconstructions.

The possibility to reconstruct motion from few k-space data makes MR-MOTUS a natural candidate for time-resolved 3D+t motion estimation, which is not directly restricted to the achievable temporal resolution in MR-images. Our work presented in Huttinga *et al.* [84], however, represents a proof-of-concept, and demonstrates MR-MOTUS in an experimental setting. Four points of improvement should be addressed for the extension of MR-MOTUS to time-resolved 3D+t motion estimation:

1. Only spatial correlation in motion-fields was exploited, and a single static motion-field was reconstructed for each single snapshot of k-space data. Additionally exploiting temporal correlation, and jointly reconstructing the 3D+t motion-field series at once, could improve the reconstruction quality and lower requirements of computing time and memory.
2. Only the body coil was used for data acquisition to obtain homogeneous coil sensitivity. Since typically multi-coil acquisitions are favored, this did not represent a practical setting.
3. The required reference image was obtained from a separate MR-scan during breath-hold. Ideally, no breath-holds are required and reconstructions can be performed on data acquired in free-breathing conditions.
4. 3D motion-fields were previously reconstructed from retrospectively undersampled Cartesian k-space data, while the motion estimation application requires prospectively undersampled acquisitions with an efficient non-Cartesian trajectory.

In this work we address the aforementioned points of improvement and extend the framework to experiments in a realistic setting, in which reference image and time-resolved 3D+t motion-fields can be reconstructed from linearly combined multi-coil,

free-breathing, prospectively undersampled non-Cartesian 3D k-space data. Time-resolved 3D+t motion-field reconstructions require the representation of 3D motion-fields over a large number of timepoints (>100), and thereby introduce a large number unknowns. We propose to use a spatio-temporal low-rank motion model to compress the representation of 3D+t motion-fields. Several works have previously proposed low-rank motion models for motion estimation [33, 85–89], and the analyses in Stemkens *et al.* [33], Zhang *et al.* [85], and Li *et al.* [86] suggest that a rank-2 motion model can accurately describe respiratory motion. Consequently, the low-rank motion model can reduce the number of unknowns by two orders of magnitude, thereby introducing a regularization in both space and time and significantly reducing memory consumption and reconstruction times for 3D+t reconstructions. We will refer to the extended framework as low-rank MR-MOTUS.

We demonstrate and validate low-rank MR-MOTUS in a total of 6 in vivo experiments on 2 healthy subjects and several moving anatomies. 2D/3D abdominothoracic respiratory motion is included in view of the MR-guided radiotherapy application, and 3D head-and-neck motion is included for additional validation and as a demonstration to handle different types of motion. The 2D respiratory motion reconstruction is validated against 2D time-resolved compressed sensing, the 3D respiratory motion reconstruction against respiratory-resolved 3D image reconstruction, and the 3D head-and-neck motion against 3D static images acquired right before and right after the motion.

3.2 Theory

Background MR-MOTUS

We assume a general d -dimensional setting, with targeted case $d = 3$, and we follow the convention that bold-faced characters denote vectorizations. We define $\mathbf{x}_0 \mapsto \mathbf{x}_t$ as the mappings from coordinates $\mathbf{x}_0 \in \mathbb{R}^d$ in a reference image to new locations $\mathbf{x}_t \in \mathbb{R}^d$ at time t . The mappings are characterized by the motion-fields \mathbf{d}_t through $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{d}_t(\mathbf{x}_0)$. This will be written in concatenated vector-form as

$$\mathbf{X}_t = \mathbf{X}_0 + \mathbf{D}_t, \quad (3.1)$$

where $\mathbf{X}_t, \mathbf{X}_0, \mathbf{D}_t \in \mathbb{R}^{Nd \times 1}$ denote the vertical concatenations over N spatial points in a d -dimensional setup. The MR-MOTUS forward model [84] explicitly relates the motion-fields \mathbf{D}_t and a static reference image $\mathbf{q}_0 \in \mathbb{C}^N$ to dynamic, single-channel (and possibly non-Cartesian) k-space measurements $\mathbf{s}_t \in \mathbb{C}^{N_k}$:

$$\mathbf{s}_t = \mathbf{F}(\mathbf{D}_t | \mathbf{q}_0) + \boldsymbol{\epsilon}_t. \quad (3.2)$$

Here $\boldsymbol{\epsilon}_t \in \mathbb{C}^{N_k}$ is the complex noise vector and $\mathbf{F} : \mathbb{R}^{Nd} \mapsto \mathbb{C}^{N_k}$ is the vectorization of the forward operator defined as

$$F(\mathbf{d}_t)[\mathbf{k}] = \int_{\Omega} q_0(\mathbf{x}_0) e^{-i2\pi \mathbf{k} \cdot [\mathbf{x}_0 + \mathbf{d}_t(\mathbf{x}_0)]} d\mathbf{x}_0, \quad (3.3)$$

where $\mathbf{k} \in \mathbb{R}^d$ denotes the k-space coordinate. By fitting the non-linear signal model in Eq. (3.3) to acquired k-space data, motion-fields can be reconstructed directly from k-space measurements.

Reconstruction problem formulation for space-time reconstructions

In this work we follow [90] and formulate the reconstruction problem for space-time motion-fields \mathbf{D} as follows:

$$\min_{\mathbf{D}} \sum_{t=1}^M \|\mathbf{F}(\mathbf{D}_t) - \mathbf{s}_t\|_2^2 + \lambda_R \mathcal{R}(\mathbf{D}). \quad (3.4)$$

Here $\mathcal{R}(\mathbf{D}) > 0$ is a regularization functional, with corresponding parameter $\lambda_R > 0$, which models a-priori assumptions in order to exploit correlations in both space and time.

Parameterization with a low-rank space-time motion model

A straightforward parameterization of \mathbf{D} considers one motion-field per dynamic, i.e. $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_M] \in \mathbb{R}^{Nd \times M}$. This is, however, impractical from a computational point-of-view, since the number of parameters scales with the number of dynamics: $|\mathbf{D}| = NMd \sim M$. For a typical scenario, $N \sim 10^6$, $M \sim 10^2$ and $d = 3$, in which case

$$|\mathbf{D}| \sim 10^8. \quad (3.5)$$

Hence, this parameterization results in high memory consumption and long reconstruction times.

We observe that internal body motion is typically very rigid, and movement occurs along a similar angle over time. We therefore hypothesize that a representation of motion-fields as the summation of motion directions per voxels with relative magnitude, multiplied by global scalings along these directions over time, can lead to an efficient representation with a reduced number of parameters. As an example, motion with fixed directions (and relative magnitude) per voxel can be represented with just one motion-field and one global 1D scaling along these directions over time. This representation can mathematically be captured with a low-rank motion model. The low-rank model simultaneously reduces the number of parameters for the reconstruction and introduces a natural regularization in both space and time by enforcing the following factorization in spatial and temporal contributions:

$$\mathbf{D} = \begin{pmatrix} | & & | \\ \mathbf{D}_1 & \dots & \mathbf{D}_M \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \Phi^1 & \dots & \Phi^R \\ | & & | \end{pmatrix} \begin{pmatrix} - & \Psi^1 & - \\ & \vdots & \\ - & \Psi^R & - \end{pmatrix} = \Phi \Psi^T. \quad (3.6)$$

Here R denotes the number of components of the model; $\Phi \in \mathbb{R}^{Nd \times R}$ denotes the matrix with spatial components, and $\Psi \in \mathbb{R}^{M \times R}$ denotes the matrix with temporal components. See also Figure 3.1 for a more intuitive explanation of the model. The model (3.6) will be referred to as the low-rank model, since $\text{rank}(\mathbf{D}) \leq R$. The upper

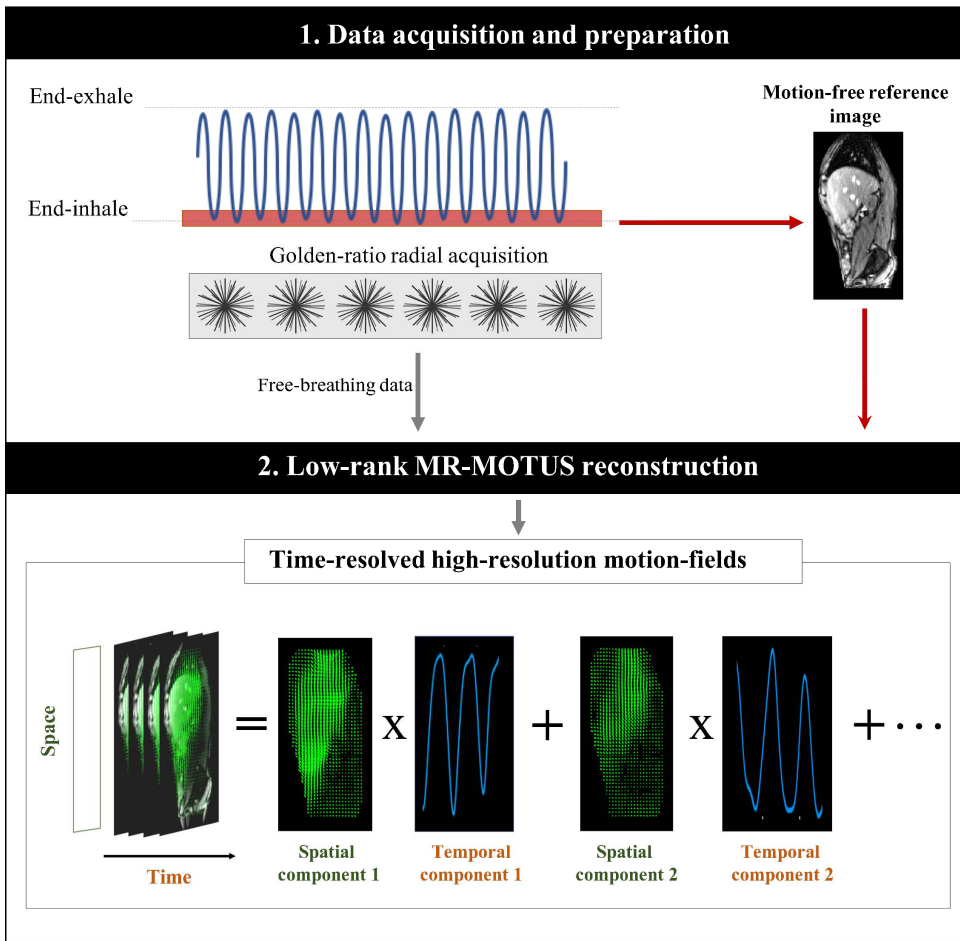


Fig. 3.1: Overview of the low-rank MR-MOTUS framework. First, data is acquired during free-breathing with a golden-ratio radial trajectory (2D: golden-angle radial [91], 3D: golden-mean radial kooshball [92]). Then, DC-based phase-binning is performed on end-inhale to reconstruct a motion-free reference image. Finally, the reference image and free-breathing data are fed into the low-rank MR-MOTUS reconstructions, resulting in time-resolved 3D motion-fields. The motion-fields are reconstructed with an explicit constraint on the maximum rank. That is, as a sum of component motion-fields with each a different temporal behavior. The number of such components is pre-determined.

limit is achieved for R linearly independent components. A similar explicit low-rank factorization was recently proposed in the context of image reconstructions in Ong *et al.* [29], with the same motivations as mentioned above.

The number of parameters in the low-rank model is $|\mathbf{D}| = |\Phi| + |\Psi| = (Nd + M)R$. Analyses in the works Stemkens *et al.* [33], Zhang *et al.* [85], and Li *et al.* [86] suggest that a motion model with rank 2 is sufficient to accurately model respiratory motion. For the typical scenario considered above ($N \sim 10^6, M \sim 10^2, d = 3$), this would then imply

$$|\mathbf{D}| \sim 10^6, \quad (3.7)$$

which is two orders of magnitude lower than Eq. (3.5). In Section 3.7 analyses are performed to confirm that the low-rank motion model with a small number of components R can indeed accurately represent motion-fields.

We follow a standard approach in non-rigid medical image registration [63] and represent both the spatial components Φ and the temporal components Ψ of the motion-fields in cubic B-spline bases. This results in representation coefficients α, β for respectively Φ and Ψ .

Regularization functional

The motion-field reconstruction problem in Eq. (3.4) is typically ill-posed, and requires incorporation of a-priori knowledge of the motion-fields. Since organs such as the liver, spleen and kidney consist of liquid filled tissue structures, they can be assumed incompressible and thus volume-preserving under motion [40]. The Jacobian determinant $\mathcal{J}(\mathbf{D}_t(\mathbf{r}))$ is the fraction of the volume at spatial coordinate \mathbf{r} after deformation by $\mathbf{D}_t(\mathbf{r})$, with respect to the reference volume before deformation. Hence, values between 0 and 1 indicate shrinkage, values around 1 indicate no compression nor expansion, and values above 1 indicate expansion. We enforce the incompressibility assumption by penalizing deviations of the Jacobian determinant from unity [93]:

$$\mathcal{R}(\mathbf{D}) := \sum_{t=1}^M \|\mathbf{W}(\mathcal{J}(\mathbf{D}_t) - \mathbf{1})\|_2^2. \quad (3.8)$$

Here $\mathcal{J}(\cdot)$ computes the determinant of the Jacobian, and \mathbf{W} is a diagonal matrix with weights per voxel. The weights are added to exclude regions where the regularization is less realistic, e.g. in the lungs. As weights we have taken the magnitude of the reference image, scaled to unit norm. For the implementation we follow [93], and compute spatial derivatives analytically using the spline parameterization of the motion-fields.

Acquisition details			
Parameter	2D resp. motion	3D resp. motion	3D head-and-neck motion
FOV [m]	$0.50 \times 0.50 \times 0.01$	$0.44 \times 0.44 \times 0.44$	$0.38 \times 0.38 \times 0.38$
Acquisition matrix size	$164 \times 164 \times 1$	$146 \times 146 \times 146$	$126 \times 126 \times 126$
Number of samples on readout	264	264	232
Spatial acq. resolution [mm]	$3.00 \times 3.00 \times 10.00$	$3.00 \times 3.00 \times 3.00$	$3.00 \times 3.00 \times 3.00$
Repetition time [ms]	4.90	4.40	5.40
Echo time [ms]	2.30	1.80	2.30
Flip angle [°]	20	20	20
Bandwidth [Hz]	298.72	541.48	284.73
Trajectory	2D golden-angle radial	3D golden-mean radial kooshball	3D golden-mean radial kooshball
Pulse sequence	2D SPGR	3D SPGR	3D SPGR
Coils (#Channels)	Anterior + Posterior (24)	Anterior + Posterior (24)	Anterior + Posterior (24)
Scanner	Philips Ingenia 1.5T	Philips Ingenia 1.5T	Philips Ingenia 1.5T
Reconstruction details			
Parameter	2D resp. motion	3D resp. motion	3D head-and-neck motion
Motion model components	$R = 3$	$R = 3$	$R = 6$
Reference image resolution [mm]	$6.70 \times 6.70 \times 10.00$	$6.70 \times 6.70 \times 6.70$	$9.05 \times 9.05 \times 9.05$
Regularization parameter	$\lambda_R = 1.5 \cdot 10^1$	$\lambda_R = 1.5 \cdot 10^1$	$\lambda_R = 1.4 \cdot 10^3$
Number of iterations	50	50	300
Splines per spatial dimension	18	16	3
Splines in time	1.28 / second	8.25 / second	5 / second
Temporal motion resolution	40.8 Hz: 5 spokes / dynamic	7.6 Hz: 30 spokes / dynamic	9.3 Hz: 20 spokes / dynamic
Reconstructed motion duration [s]	20	33	40
Reconstruction time	4 minutes	50 minutes	2 hours

Table 3.1: Details of the in vivo experiments as described in Supporting Information Section 3.3-Supporting Information Section 3.3: the top half lists acquisition details, and the bottom half lists reconstruction details for the time-resolved experiments. For the respiratory-resolved reconstruction in Supporting Information Section 3.3 the same parameters were used as listed in the ‘3D resp. motion’ column, but effectively resulted in a temporal motion resolution of about 5Hz, with 18062 spokes per dynamic, due to the sorting.

3

Final reconstruction problem formulation

Substituting the spline representation, low-rank model (3.6) and regularization (3.8) into the objective function (3.4) results in the following minimization problem to reconstruct space-time motion-fields:

$$\{\alpha^\dagger, \beta^\dagger\} = \underset{\Phi \Psi^T = [\mathbf{D}_1, \dots, \mathbf{D}_M]}{\operatorname{argmin}} \sum_{t=1}^M \|\mathbf{F}(\mathbf{D}_t) - \mathbf{s}_t\|_2^2 + \lambda_R \sum_{t=1}^M \|\mathbf{W}(\mathcal{J}(\mathbf{D}_t) - \mathbf{1})\|_2^2, \quad (3.9)$$

where $\lambda_R \in \mathbb{R}^+$ is the regularization parameter that balances the terms. Note that no temporal regularization is added, since the low-rank model already acts as a strong regularization in both space and time.

3.3 Methods

Experiments overview

The following data were acquired in three different experiments per volunteer for two volunteers:

1. 2D+t abdominothoracic data;

2. 3D+t abdominothoracic data;
3. 3D+t head-and-neck data.

The 2D+t abdominothoracic data allows for a validation against time-resolved image reconstruction at a high temporal resolution. The 3D+t abdominothoracic data is the targeted case for the application in MR-guided radiotherapy. The 3D+t head-and-neck data is included as a demonstration to handle different types of motion, and for additional validation. All reconstructions are analyzed by comparison with image reconstructions on the same data. As an additional sanity check, the Jacobian determinant of the transformation corresponding to the motion-fields is analyzed: $\mathbf{x}_0 \mapsto \mathbf{x}_0 + \mathbf{d}_t(\mathbf{x}_0)$. More details regarding the experiments are provided below, organized per subsection.

Data acquisition

All data were acquired on a 1.5T MRI scanner (Ingenia, Philips Healthcare, Best, the Netherlands) using a steady-state spoiled gradient echo sequence (SPGR) with anterior and posterior receive arrays. We employed golden-angle radial readouts for 2D [91], and golden-mean kooshball radial readouts for 3D [92]. The volunteers provided written informed consent prior to the scans, and all scans were approved by the institutional review board of the University Medical Center Utrecht and carried out in accordance with the relevant guidelines and regulations. See Table 3.1 for all relevant acquisition parameters.

Reconstruction details

We followed the approach outlined in Supporting Information Section 3.2, and reconstructed motion-fields from linearly combined multi-coil k-space data acquired during motion by solving the minimization problem (3.9) with L-BFGS [66], using the MATLAB implementation from Becker [94]. The low-rank MR-MOTUS workflow is schematically summarized in Figure 3.1. We refer to Table 3.1 for all parameter settings and to the Supporting Information of this work and the Supporting Information in Huttinga *et al.* [84] for more implementation details. Code that produces similar results as presented in this study is openly available at https://github.com/nrfhuttinga/LowRank_MRMOTUS.git.

Experiment 1: 2D+t in vivo respiratory motion reconstructions from abdominothoracic data

In the first experiment, a reference image and motion-fields were reconstructed from the same 2D+t data acquired during 20 seconds of free-breathing. The reference image was reconstructed from the end-inhale bin after phase-binning based on the self-navigation signal of $\mathbf{k} = 0$ values per readout (denoted as k_0 -values), see Section 3.7 for more details. The motion-fields were reconstructed at 40.8 Hz, i.e. 24.5 ms/frame, by assigning every 5 consecutive non-overlapping spokes to one dynamic. The low-rank model (3.6) was employed with $R = 3$, yielding motion-fields with rank

≤ 3 . Additional relevant reconstruction and acquisition parameters can be found in Table 3.1.

The motion-fields were analyzed by comparison with a time-resolved compressed sensing 2D+t reconstruction (CS2Dt) on the same free-breathing data, and by means of the Jacobian determinant. For the comparison with CS2Dt, the MR-MOTUS reference image was warped with the reconstructed motion-fields to obtain a dynamic image sequence as follows. First, the motion-fields are interpolated to the same spatial resolution as the image reconstruction using cubic interpolation. Second, the forward model (3.2) was evaluated on a Cartesian k-space grid using the reconstructed motion-fields \mathbf{D}_t . Finally, an inverse Fourier transform was performed to obtain one image per dynamic. The CS2Dt was reconstructed at a temporal resolution of 122.5 ms/frame by assigning every 25 consecutive non-overlapping spokes to one dynamic, and was performed with the BART toolbox [70] using spatial L_1 -wavelet and temporal total variation regularization. The temporal resolution of the CS2Dt was chosen as an integer multiple of the MR-MOTUS resolution to allow comparison at the coarser CS2Dt temporal resolution. The comparison was performed by means of the normalized root mean squared error (NRMSE). The NRMSE between vectors \mathbf{a}, \mathbf{b} was defined as $\text{NRMSE}(\mathbf{a}, \mathbf{b}) = \frac{\|\mathbf{a}-\mathbf{b}\|}{\|\mathbf{b}\|}$.

Experiment 2: 3D+t in vivo respiratory motion reconstructions from abdominothoracic data

3

In the second experiment we considered the targeted case for MR-guided radiotherapy: a reference image and motion-fields were reconstructed from 3D+t data acquired during 33 seconds of free-breathing. The targeted high temporal resolution does not allow for a straightforward validation by comparison with dynamic 3D image reconstruction. For validation purposes, we therefore compared MR-MOTUS with respiratory-resolved image reconstruction by performing both reconstructions on respiratory-sorted data.

Finally, we performed 3D+t time-resolved motion reconstruction to demonstrate the ability to reconstruct motion at high temporal resolution from time-resolved k-space data. The reference image for both reconstructions was reconstructed from the end-inhale bin after phase-binning based on the k_0 -value per readout (see Section 3.7), and the low-rank model (3.6) was employed with $R = 3$. See Table 3.1 for all reconstruction and acquisition parameters.

For the respiratory-resolved reconstructions phase-binning was performed in 20 equal-sized bins based on the k_0 -value per readout. The images were independently reconstructed for each bin using 28 iterations of CG-SENSE [95]. The motion-fields were reconstructed over all bins simultaneously with low-rank MR-MOTUS by solving Eq. (3.9) with 20 dynamics. The quality of the MR-MOTUS reconstruction was assessed by means of the Jacobian determinant and by comparison with the respiratory-resolved image reconstruction. For the latter, a reference image was warped with the reconstructed motion-fields to obtain a dynamic image sequence, as described in Section 3.3, and the two image sequences were compared in terms of NRMSE. The reference image that was warped using the MR-MOTUS motion-fields

was selected as the end-inhale phase of the respiratory-resolved image reconstruction (motion state #10) in order to reduce effects of image intensity, image quality, or contrast differences on the comparison of the two image sequences.

For the time-resolved 3D+t reconstructions, motion-fields were reconstructed at 7.6 Hz, i.e. 132 ms/frame, by assigning every 30 consecutive non-overlapping spokes to one dynamic. The reconstructions were analyzed by means of the Jacobian determinant and the average motion of the kidney was compared between the time-resolved and respiratory-resolved MR-MOTUS reconstructions. This motion was computed as the mean of the displacements over a manually segmented mask of the right kidney. For comparison between respiratory-resolved and time-resolved, the motion magnitudes of each respiratory bin in the respiratory-resolved reconstruction were assigned to the original, time-resolved, spoke indices that were sorted into that particular bin.

Experiment 3: 3D+t in vivo head-and-neck motion reconstructions

With the third experiment, 3D+t motion-fields were reconstructed from data acquired during head-and-neck motion. The subject was instructed to hold still in position 1 during the first 70 seconds of the acquisition, then move to position 2 and hold still for 70 seconds, then move freely for 40 seconds, and finally hold still afterwards in position 3 for 70 seconds. Data acquired in position 1 was used to reconstruct a reference image, data acquired during movement from position 2 to position 3 was used to reconstruct motion-fields, and position 2 and 3 were used as fully-sampled "checkpoints" to serve as validation; the beginning and end of the dynamic motion reconstruction should respectively coincide with positions 2 and 3. To verify this, the reference was warped with the reconstructed motion-fields as described in Supporting Information Section 3.3, and the first and last dynamic of the resulting image sequence were visually compared with the fully-sampled checkpoints. As a second analysis, the mean and standard deviation of the determinant of the Jacobian were computed for all dynamics, over all voxels within the body. The latter were determined by a threshold on the magnitude of the signal per voxel. The low-rank motion model was employed with $R = 6$ to accommodate the head-and-neck motion which includes rotations in multiple planes. The motion-fields were reconstructed at a temporal resolution of 9.3 Hz, i.e. 108 ms/frame, by assigning every 20 consecutive non-overlapping spokes to one dynamic. Additional reconstruction and acquisition parameters can be found in Table 3.1.

3.4 Results

Experiment 1: 2D in vivo respiratory motion reconstructions from abdominothoracic data

The time-resolved 2D respiratory motion was reconstructed with 40.8 motion-fields-per second. The Jacobian determinant and the comparison with CS2Dt is shown in Figure 3.2. The visual comparison with 2D+t compressed sensing image recon-

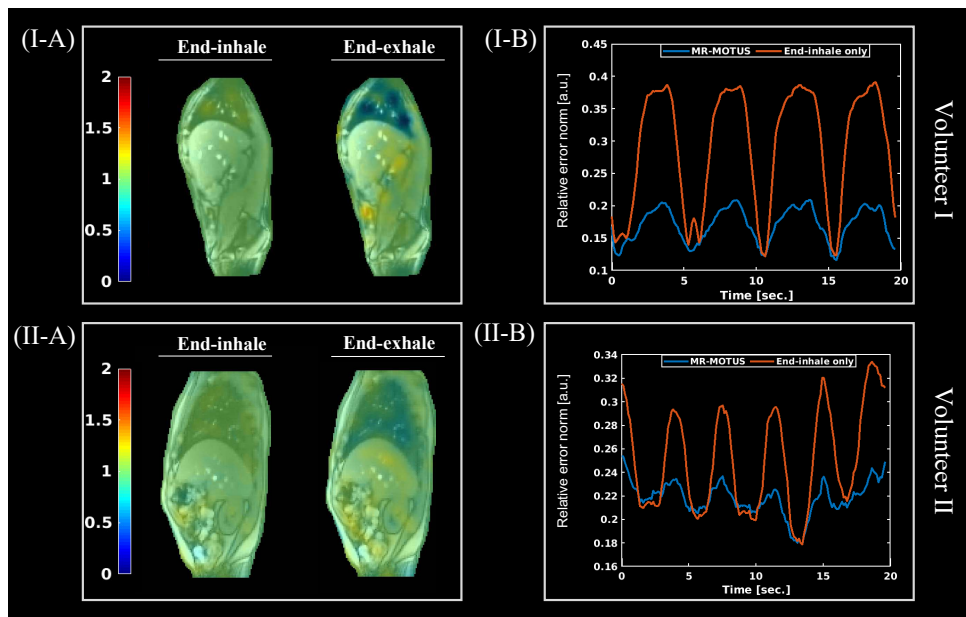


Fig. 3.2: A) Jacobian determinants of the reconstructed motion-fields in end-inhale (left) and end-exhale (right). The first end-exhale and second end-inhale positions were selected from all dynamics for this visualization. B) Normalized root mean squared error (NRMSE) between MR-MOTUS warped reference images and CS2Dt reconstruction over all dynamics (blue), and a baseline NRMSE between the fixed MR-MOTUS end-exhale warped reference image and CS2Dt. The top row (I) shows the results for volunteer 1, whereas the bottom row (II) shows the results for volunteer 2. The comparison is also visualized in Supporting Video 3.1, and the reconstructed motion-fields decomposed in the low-rank model components are visualized in Supporting Video 3.2 and Supporting Video 3.3.

struction corresponding to Figure 3.2B is shown in Supporting Video 3.1[‡]. It can be observed that good agreement is obtained for most phases of the respiratory cycle, with a small mismatch in end-exhale in the upper back near the spine-liver interface. The Jacobian determinants show small deviations from unity within the organs (green), and compression in the lungs (blue) except for the arteries. The qualitative results are supported by the quantitative results in Figure 3.2B, which show that the warped MR-MOTUS images considerably reduce the NRMSE.

The warped reference images corresponding to the reconstructed motion-field, overlaid with the motion-field are shown in Supporting Video 3.2 and Supporting Video 3.3. Moreover, these show the decomposition in the reconstructed low-rank components. For volunteer 1, the first two components show pseudo-periodic temporal behaviors, and the first is most prominent in magnitude. Both components show

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

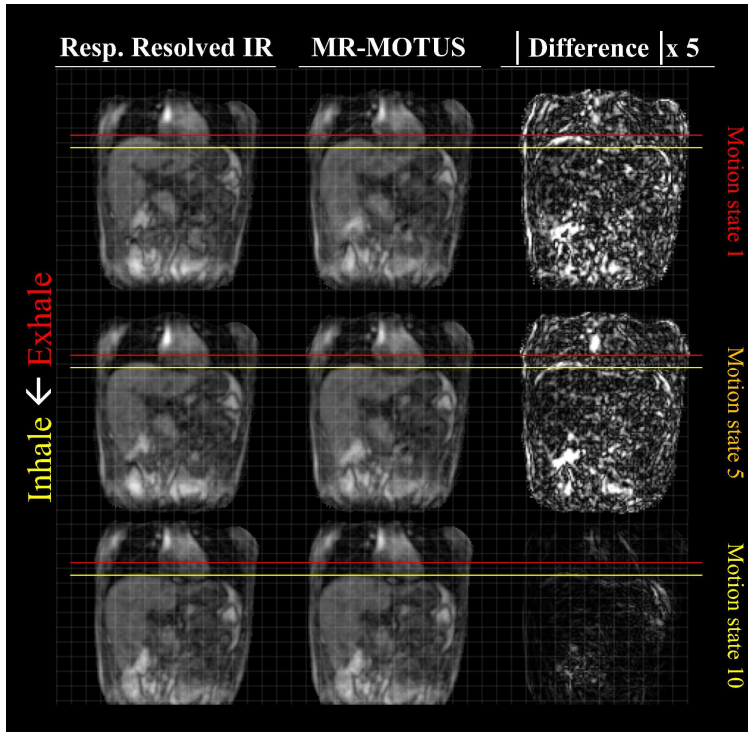


Fig. 3.3: Respiratory-resolved image reconstruction (Resp. Resolved IR, left), MR-MOTUS warped reference image (middle), and pixel-wise absolute difference between the two reconstructions (right), as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The red and yellow horizontal lines indicated respectively end-exhale and end-inhale positions. A video corresponding to this figure of volunteer 1 is provided in Supporting Video 3.4. A similar video for volunteer 2 is provided in Supporting Video 3.5.

realistic movement of organs such as the liver and kidney, but also small unrealistic motion in the spine near the liver in end-exhale. Interestingly, the third component shows a temporal behavior with a slight drift upwards, and the corresponding spatial motion-field indicates a global rotation. Similar movement can also be observed in the ground-truth CS2Dt reconstruction in Supporting Video 3.1. This movement could be caused by relaxation of the gluteus maximus muscle in the upper leg and buttocks. Similar motion patterns can be observed in Supporting Video 3.3 for volunteer 2, but the global rotation is less pronounced in the ground-truth CS2Dt reconstruction.

Experiment 2: 3D in vivo respiratory motion reconstructions from abdomen/thorax data

The comparison between MR-MOTUS and respiratory-resolved image reconstruction is shown in Figure 3.3, Figure 3.4, Supporting Video 3.4 and Supporting Video 3.5.

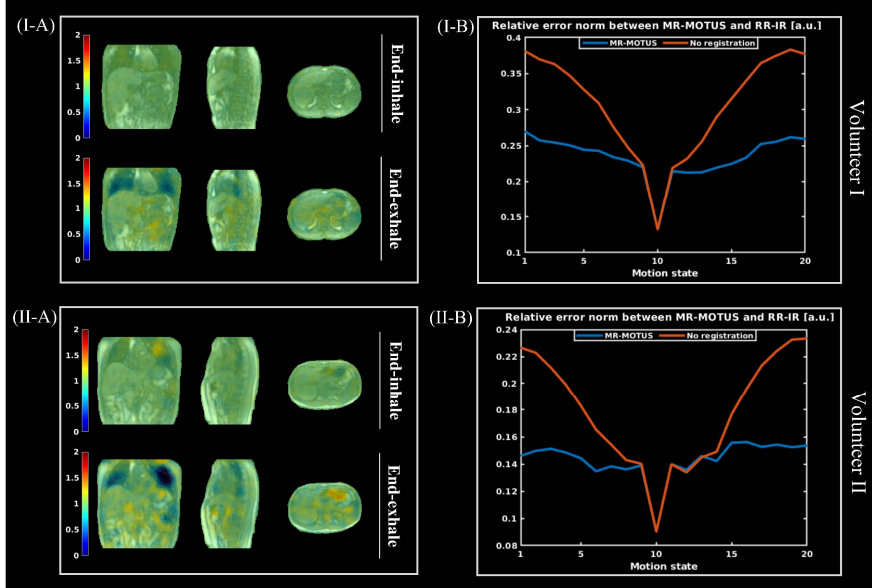


Fig. 3.4: A) Jacobian determinants of the reconstructed respiratory-resolved motion-fields in end-inhale (top) and end-exhale (bottom). The first end-exhale and second end-inhale positions were selected from all dynamics for this visualization. B) Normalized root mean squared error (NRMSE) with respiratory-resolved image reconstruction (RR-IR) for every motion state. The blue graph indicates the NRMSE between MR-MOTUS and respiratory-resolved image reconstruction. The orange graph indicates a baseline comparison between the (fixed) end-inhale image of the MR-MOTUS reconstruction and the (dynamic) respiratory-resolved image reconstruction. The sharp peak is caused by taking the 10th dynamic as the reference image for this comparison. The top row shows the results for volunteer 1, and the bottom row shows the results for volunteer 2. Videos corresponding to the comparisons in (B) are provided in Supporting Video 3.4 and Supporting Video 3.5.

It can be observed that good visual agreement is obtained between the two reconstructions for both volunteers. This is especially visible from the position of the top of the liver dome. The Jacobian determinants of the reconstructed motion-fields are shown in Figure 3.4A. The lungs show compression (blue), except for the arteries, and small deviations from unity can be observed in the rest of the body. Deviation from unity can be observed at the spine-liver interface, where a large volumetric compression is reconstructed. We expect this is related to the attachment of liver tissue to the spine during exhalation. The quantitative comparison in Figure 3.4B shows best agreement at motion state 10 (inhale) and worst agreement in motion state 19 (exhale). The sharp peak at motion state 10 can be explained by the fact that we took motion state 10 as the reference image to compute the warped reference images for MR-MOTUS. The warped reference images reconstructed from the respiratory-sorted data, overlaid with the motion-field, are visualized for both vol-

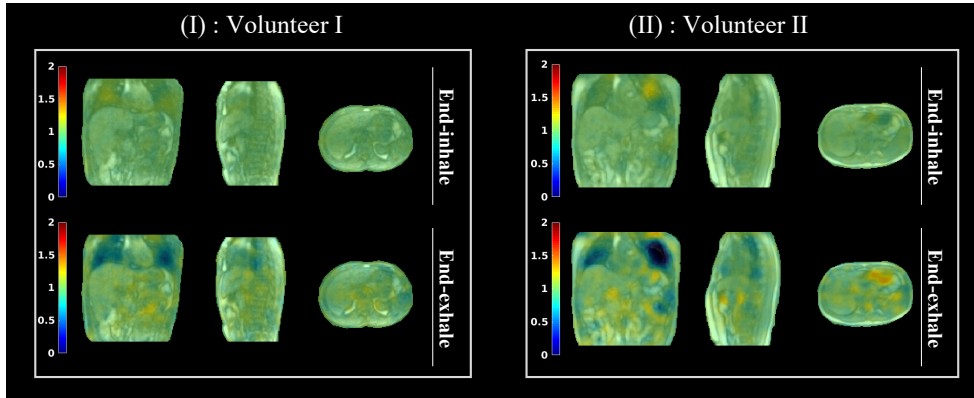


Fig. 3.5: Jacobian determinants of the reconstructed time-resolved motion-fields in end-inhale (top) and end-exhale (bottom). The left figure shows the results for volunteer 1 and the right figure the results for volunteer 2. Videos corresponding to the reconstructions in this figure are provided in Supporting Video 3.8 and Supporting Video 3.9.

3

unteers in Supporting Video 3.6 and Supporting Video 3.7. Moreover, these show the decomposition in the reconstructed low-rank components. For both volunteers the first component shows a pseudo-periodic behavior in time and is most prominent in magnitude; the other components make only minor contributions. These large contributions of pseudo-periodic components could be due to the periodicity assumption underlying the respiratory-sorting. Small unrealistic motion can be observed for volunteer 1 at the spine-liver interface and at the back of the spine, similar to the 2D reconstructions. Additionally, a small rotating motion can be observed in the motion-field for volunteer 1 at the interface with the rib cage in the coronal slice on the bottom right. We expect the latter is caused by a combination of the volume-preserving regularization and the inability of the motion model to resolve the sliding motion that is present in this area.

The time-resolved 3D respiratory motion was reconstructed with 7.6 motion-fields-per-second. The warped reference images reconstructed from the time-resolved data, overlaid with the motion-field, are visualized for both volunteers in Supporting Video 3.8 and Supporting Video 3.9. Similar motion is obtained as with the respiratory-sorted data, but the reconstructed motion components are now similar in magnitude. All components show pseudo-periodic temporal behavior, and the first component of volunteer 1 indicates a small drift. Similar to the respiratory-resolved reconstructions, small unrealistic motion at the spine-liver interface and anterior side of the spine can be observed for volunteer 1. Additionally, the same small rotation can be observed near the rib cage in the bottom right of the coronal slice. The Jacobian determinants of the reconstructed motion-fields are shown in Figure 3.5. Similar patterns can be observed in end-exhale as for the respiratory-resolved motion reconstructions. Interestingly, the end-inhale image for volunteer 1 shows a small expansion in the lungs, possibly indicating that a deeper inhale than the reference

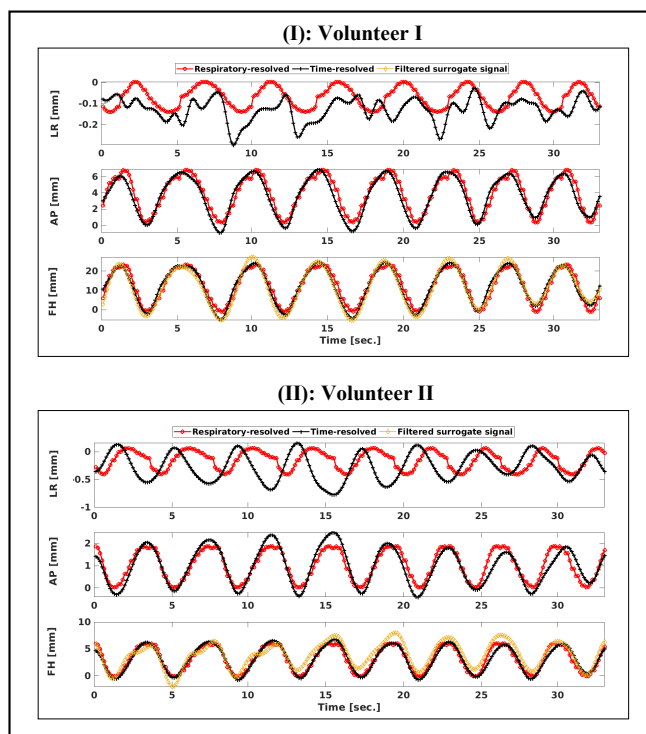


Fig. 3.6: This figure shows the average motion of the right kidney over time, for both the respiratory-resolved and the time-resolved MR-MOTUS reconstructions mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The respiratory-resolved MR-MOTUS reconstruction was projected back on the time axis, as described in Supporting Information Section 3.3 and Supporting Information Section 3.4. The average motion magnitudes were computed over a manually segmented mask of the right kidney. Videos of reconstructions corresponding to these figures are provided in Supporting Video 3.8 and Supporting Video 3.9.

image was reconstructed while the reference image was obtained using respiratory-sorting on end-inhale. Finally, the comparison between the average kidney motion in the time-resolved and respiratory-resolved MR-MOTUS reconstructions is visualized in Figure 3.6. The phase of the reconstructions are most similar in feet-head (FH) and anterior-posterior (AP), while in left-right (LR) different patterns can be observed. However, it should be noted that the motion in FH and AP is two orders of magnitude higher than in LR. The motion magnitude is similar for both reconstructions, but the respiratory-resolved reconstruction shows a constant amplitude over time since it only reconstructs an average breathing cycle. The time-resolved reconstruction shows changing motion amplitudes over time. The phase difference between the two reconstructions may be explained by imperfect respiratory-sorting.

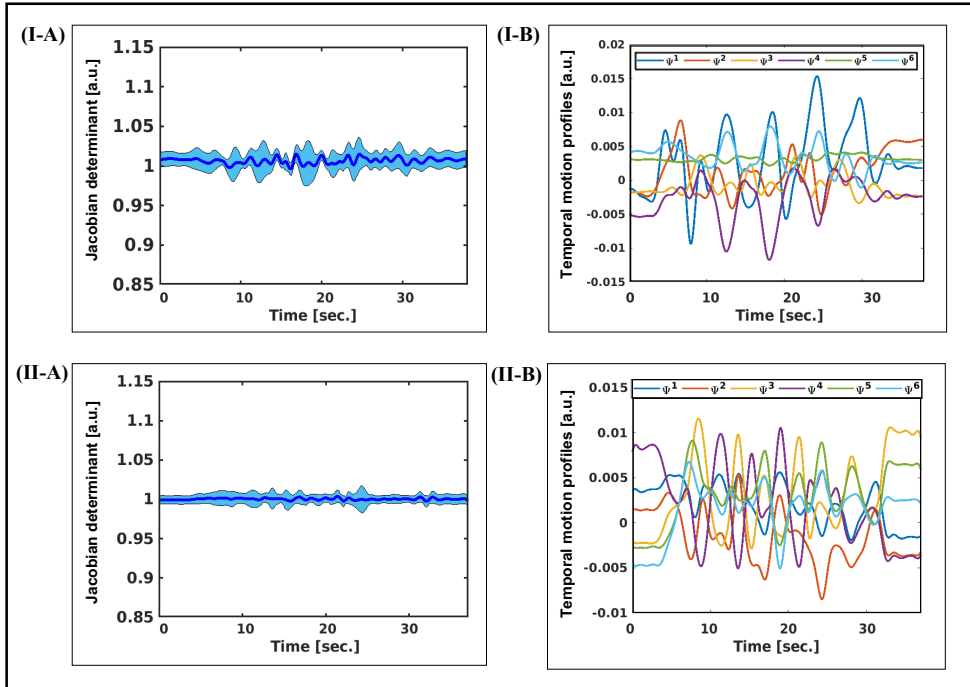


Fig. 3.7: This figure corresponds to the head-and-neck reconstructions in Supporting Information Section 3.3 and Supporting Information Section 3.4. A) The mean (solid line) and standard deviation (shaded area) of the Jacobian determinants of the reconstructed motion-fields over time. B) The reconstructed temporal profiles Ψ^i , scaled by the norm of the corresponding Φ^i to be able to compare their magnitudes. The top row and bottom rows respectively show the results for volunteer 1 and 2. Videos corresponding to the reconstructions in these figures are provided in Supporting Video 3.10 and Supporting Video 3.11.

Experiment 3: 3D in vivo head-and-neck motion reconstructions

The time-resolved 3D head-and-neck motion was reconstructed with 9.3 motion-fields-per-second. The MR-MOTUS warped reference images from 3D data acquired during head-and-neck motion are visualized for both volunteers in Supporting Video 3.10[‡] and Supporting Video 3.11. Clearly, rigid motion-fields are reconstructed within the skull, and non-rigid motion-fields at the neck. Figure 3.7 shows the Jacobian determinants of the reconstructed motion-field over time (A), and the reconstructed temporal components (B) for both volunteers. The Jacobian determinant is close to 1 over the whole reconstructed time, with slightly more deviations for volunteer 1. These can be attributed to larger and more irregular motion than volunteer 2. The temporal components are relatively flat at the start and the end, corresponding

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

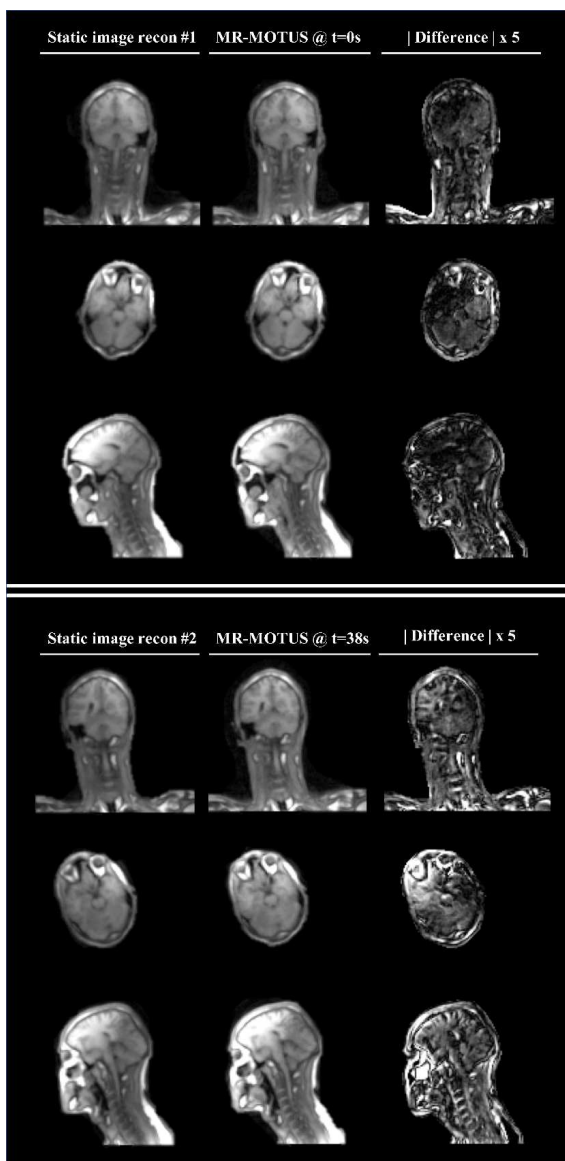


Fig. 3.8: This figure shows the checkpoint validation for the head-and-neck reconstructions of volunteer 2, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The left columns shows the fully-sampled checkpoint image, the middle column shows the MR-MOTUS warped reference images and the right column shows the absolute pixel-wise difference. The top part corresponds to the comparison with the checkpoint acquired right before the start of the motion, and the bottom part corresponds to the checkpoint acquired right after the start of the motion. A video corresponding to this figure is provided in Supporting Video 3.11. A similar video for volunteer 1 is provided in Supporting Video 3.10.

3

to the static begin and end positions. The more extreme motion of volunteer 1 can also be observed from the larger magnitudes of the temporal components and from Supporting Video 3.10. Figure 3.8 shows the checkpoint validation for volunteer 2. It can be observed that good agreement is obtained between the fully-sampled checkpoint images and the MR-MOTUS reconstructions.

3.5 Discussion

We have previously introduced MR-MOTUS [84], a framework to estimate motion directly from minimal k-space data and a reference image by exploiting spatial correlation in internal body motion. In this work, we introduce low-rank MR-MOTUS: an extension of MR-MOTUS from 3D to 3D+t reconstructions in a realistic experimental setting, where both reference image and motion-fields are reconstructed from data acquired during free-breathing. Low-rank MR-MOTUS employs a low-rank motion model that constrains the degrees of freedom in space and time, thereby reducing memory consumption and functioning as a regularization in both space and time. It was demonstrated that the proposed method can reconstruct high quality 3D motion-fields with a temporal resolution of more than 7.6 motion-fields-per-second, while showing consistency with static, respiratory-resolved and time-resolved image reconstructions. Prospectively undersampled data were acquired with a non-Cartesian trajectory and multi-channel receivers, thereby bridging the gap towards clinical application.

The ability of the proposed framework to estimate time-resolved rather than respiratory-resolved motion is promising as it allows to characterize drifts and intra and inter-cycle breathing patterns. This is in contrast with respiratory-resolved methods that require sorting to obtain suitable images [33, 34, 75–79, 96, 97]. The sorting effectively results in (a motion model for) average breathing motion, which may have trouble capturing drifts and inter-cycle variations. Some works have been proposed to reconstruct time-resolved MR-images without the need of retrospective sorting. However, the reported temporal resolution was too low [29–31], or the FOV was too small [82, 83]. The time-resolved motion estimation of low-rank MR-MOTUS in combination with an MR-linac can be particularly beneficial for MR-guided radiotherapy; the (retrospective) reconstruction of 3D+t time-resolved tumor and organs-at-risk motion during treatment can be used for accurate dose accumulation [88], allowing for an accurate assessment of the treatments.

The resulting motion model explicitly separates a high-dimensional static spatial component from a low-dimensional dynamic temporal component. The low-dimensionality of the dynamic behavior could be exploited to reduce the number of parameters and reconstruction times of future real-time reconstructions, analogously to recently proposed approaches in Stemkens *et al.* [33], McClelland *et al.* [36], Sbrizzi *et al.* [72], and Huttinga *et al.* [74]. Our method could thereby form the basis for future work on real-time MR-based motion estimation, where reconstructions are performed on-the-fly to track tumor and organs-at-risk motion.

Low-rank models in the context of motion estimation have been investigated before in several works, most of which retrospectively perform compression to a low-rank model using principal component analysis [33, 85, 87, 88]. Others decouple the motion-fields into spatial components and temporal components based on surrogate signals [35, 36, 89]. The approach in this work is different in the sense that it explicitly and a-priori enforces a structure that yields low-rank motion-fields, and does not assume dependence on surrogate signals for the motion model. Similar approaches have been studied in the context of image reconstruction [29, 39, 82, 98–101].

This work includes some limitations and assumptions that should be addressed. Both the respiratory-resolved and time-resolved 3D respiratory motion reconstructions in Supporting Information Section 3.3 and Supporting Information Section 3.4 look realistic in general. Yet, small unrealistic motion is reconstructed near discontinuities in the true motion-fields that are present near sliding or attaching/detaching organ surfaces. This can be observed in for example Supporting Video 3.2[‡], at the spine/liver interface in end-exhale. This could possibly be resolved with region-specific [102] or non-parametric motion models [103], but is beyond the scope of this work.

The respiratory motion reconstruction parameters in this work were fixed and were based on the results of a grid search for a single volunteer. We have empirically observed that the parameters from the grid search for a single volunteer are generalizable and yield acceptable results for all respiratory motion reconstructions in this work. Hence, in a realistic setting where grid searches may not be feasible due to time constraints, similar reconstruction parameter settings could be employed. The number of components R may be more subject-dependent, and can be determined with analyses similar to the ones described in Section 3.7.

This work is based on the assumption that a motion-free reference image is available, and that it can be warped with unknown motion-fields into dynamic image series of a moving anatomy. For respiratory motion, these reference images could be extracted by means of respiratory-binning, but this may be less straightforward for other body motion such as bladder filling, bowel movements and prostate movement. However, these types of motion take place on a slower temporal scale, so more time would be available to gather information and update the reference image. Additionally, the modeling assumption regarding the warping of a single reference image into the dynamic image series may be partly violated due to unwanted contrast effects such as susceptibility-induced B_0 variations. Yet, these effects can be assumed small at the targeted field strength of 1.5T. Nevertheless, future work will consider incorporating such additional contrast effects in the signal model, and performing joint reconstructions of the motion-fields and the reference image.

Resulting from the grid search for optimal reconstruction parameters, we have used relatively low-resolution reference images. This delivered the best performance, and allowed a factor 3 reduction of the reconstruction time. These low-resolution reference images could allow for high-resolution motion-fields by means of interpolation, but do, however, provide less precise information regarding organ boundaries.

Contrary to standard coil compression techniques, the aim of the compression of multi-channel data to a single channel in this work is homogeneous coil sensitivity. Consequently, this compression is suboptimal in terms of SNR [104, 105]. Supporting Figure 3.3 analyzes the loss between between a Roemer coil combination [104] and the proposed coil compression on the 2D data. This shows an SNR loss factor between 1.5 and 2.5 in most of the body, which increases towards the boundary of the body. Good results were obtained with the coil compression introduced in this work, but

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

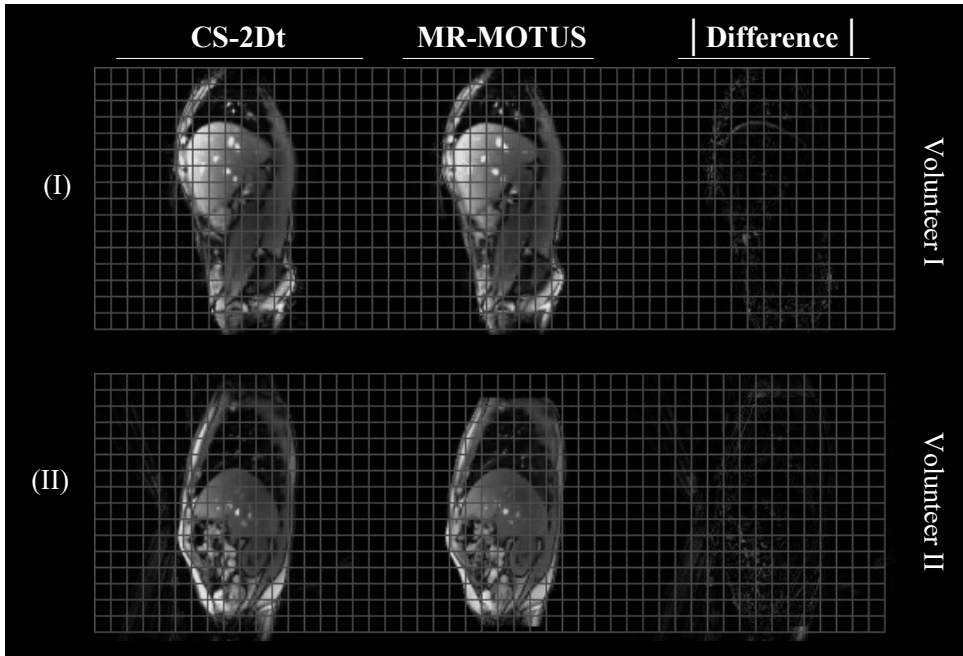
more advanced techniques could possibly be used to improve the SNR after the compression. We refer to Section 3.7 for a more extensive discussion.

The last point of improvement is the validation of time-resolved 3D+t motion-fields. In general this is not straightforward, and we considered three viable options for this: (1) *in silico* with a digital phantom, (2) with an MR-compatible motion phantom, and (3) *in vivo* with respiratory-resolved image reconstruction. We have opted for the third option, since this was considered the closest to a practical use-case. The *in silico* validation does not consider real acquisition-related data corruption (e.g. eddy currents, flow effects), and can, in case of e.g. the XCAT phantom [106], yield unstable motion-fields [107]. MR-compatible motion phantoms, although useful for proof-of-principle validations, represent simplified *in vivo* anatomies.

The intended application of MR-MOTUS is MR-guided radiotherapy, possibly in real-time. However, the current reconstruction times in MATLAB on a desktop workstation are around 4 minutes for 2D+t with 40 motion-fields/second, around 6 minutes for the respiratory-resolved 3D reconstruction, and around 50 minutes for 3D+t time-resolved respiratory motion with 7.6 motion-fields/second. Hence, the current implementation of the method is not directly applicable for real-time processing, but reconstruction times may be reduced with a different programming language, improved hardware, GPU-accelerations or deep learning.

3.6 Conclusion

We have introduced low-rank MR-MOTUS, an extension of MR-MOTUS, that allows to retrospectively reconstruct time-resolved 3D+t motion-fields from prospectively undersampled k-space data and one reference image. Reconstructions were performed for 2D/3D respiratory motion and 3D head-and-neck motion. A temporal resolution of more than 7.8 motion-fields-per-second was obtained, and the motion-fields were consistent with image reconstructions. For MR-guided radiotherapy, the time-resolved 3D motion-fields could be used to reconstruct the respiratory-motion-compensated accumulated dose during the treatment. Furthermore, the explicit decomposition of motion-fields in static and dynamic components could form the basis for future work towards real-time MR-guided radiotherapy.



Supporting Video 3.1: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>. 2D-t compressed sensing reconstruction (left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The top row shows reconstructions for volunteer 1, and the bottom row for volunteer 2.*

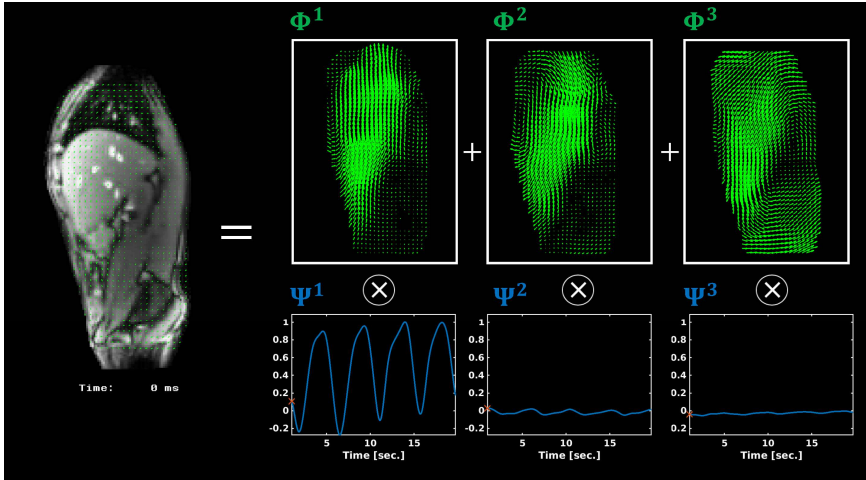
Data availability statement

Code that produces similar results as presented in this study is openly available at https://github.com/nrfhuttinga/LowRank_MRMOTUS.git.

3.7 Supporting Information

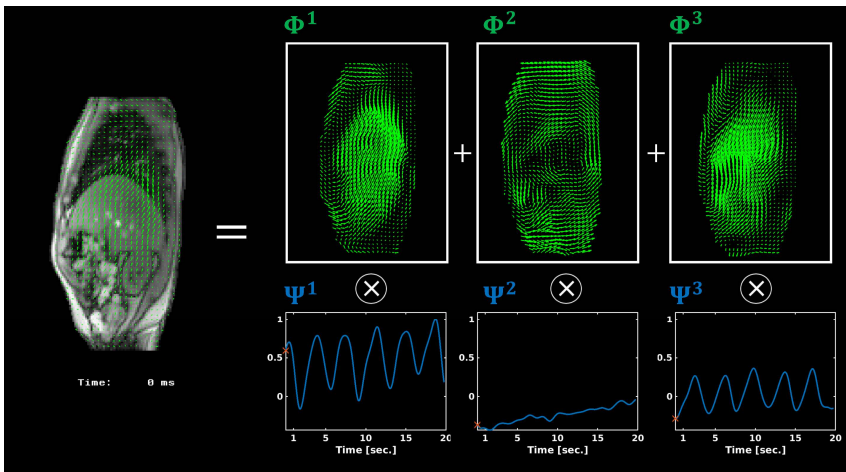
Supporting Figures

The supporting videos are part of the main body of this chapter and can be found through the QR code below the abstract or through the following link: <https://doi.org/10.6084/m9.figshare.20480571.v1>.

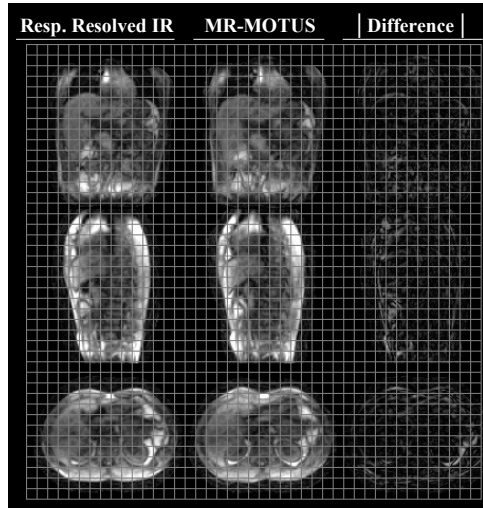


Supporting Video 3.2: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>. MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from 2D time-resolved data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.*

3

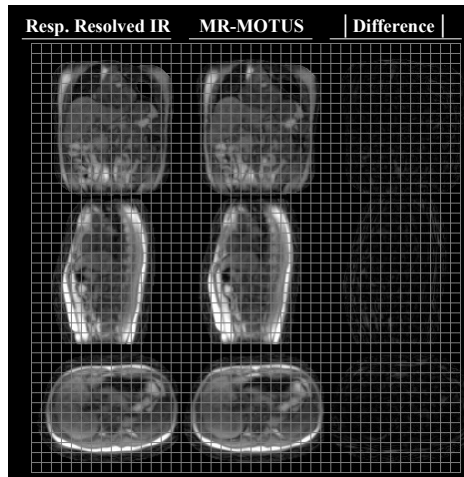


Supporting Video 3.3: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>. MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from 2D time-resolved data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.*

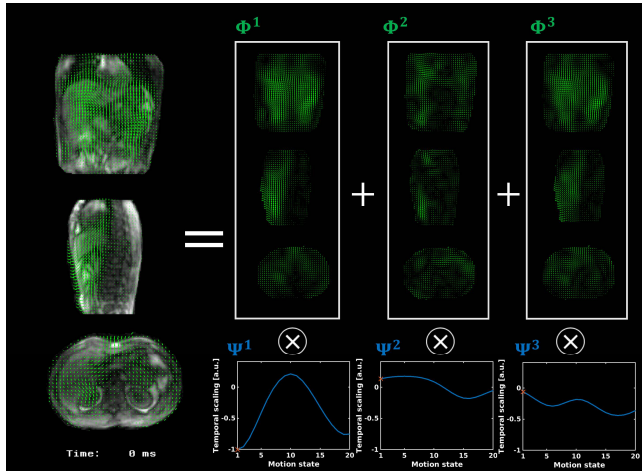


Supporting Video 3.4: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The visualization shows data from volunteer 1.

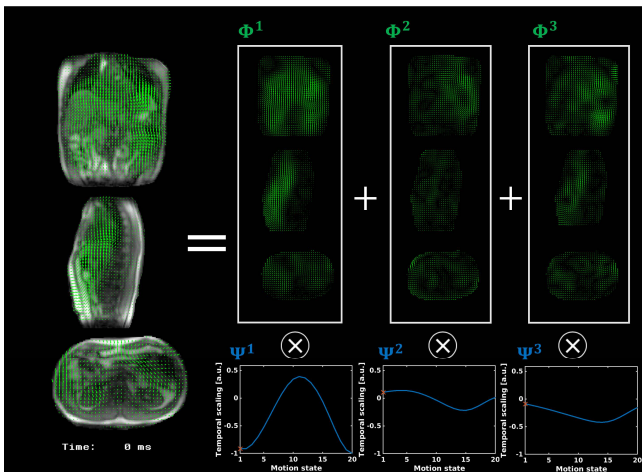
3



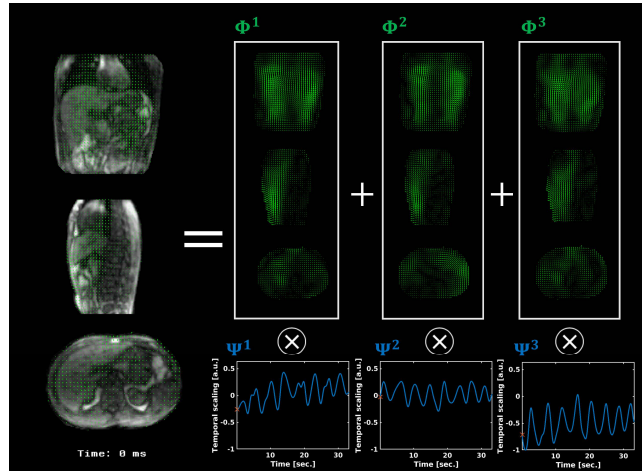
Supporting Video 3.5: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The visualization shows data from volunteer 2.



Supporting Video 3.6: *This is an animated figure and should be viewed on-line at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.

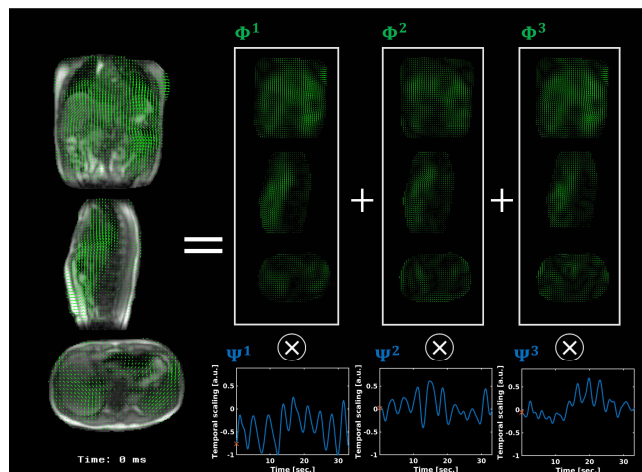


Supporting Video 3.7: *This is an animated figure and should be viewed on-line at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.

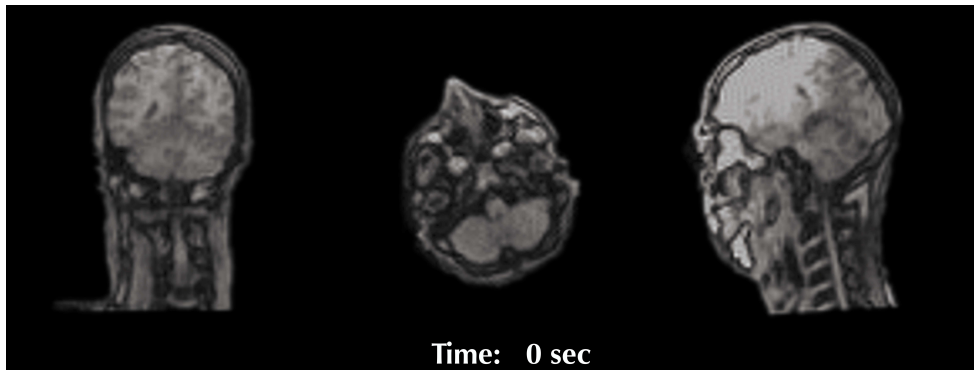


Supporting Video 3.8: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>. MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.*

3

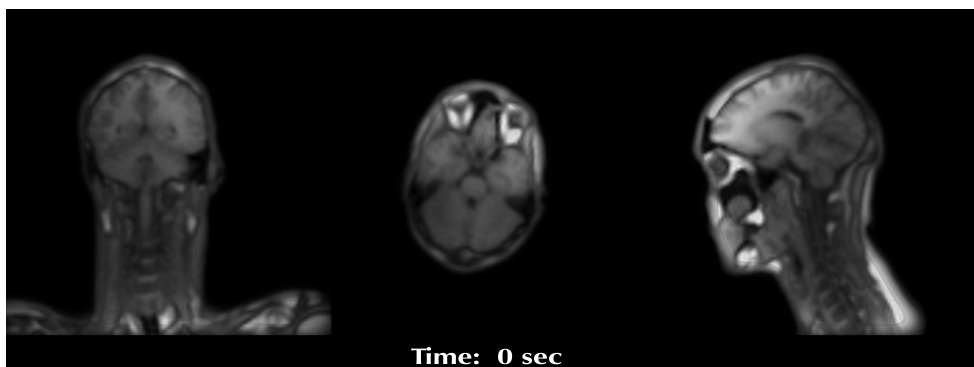


Supporting Video 3.9: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>. MR-MOTUS warped reference images overlaid with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4. The image shows a decomposition in the reconstructed components Φ^i (spatial) and Ψ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.*

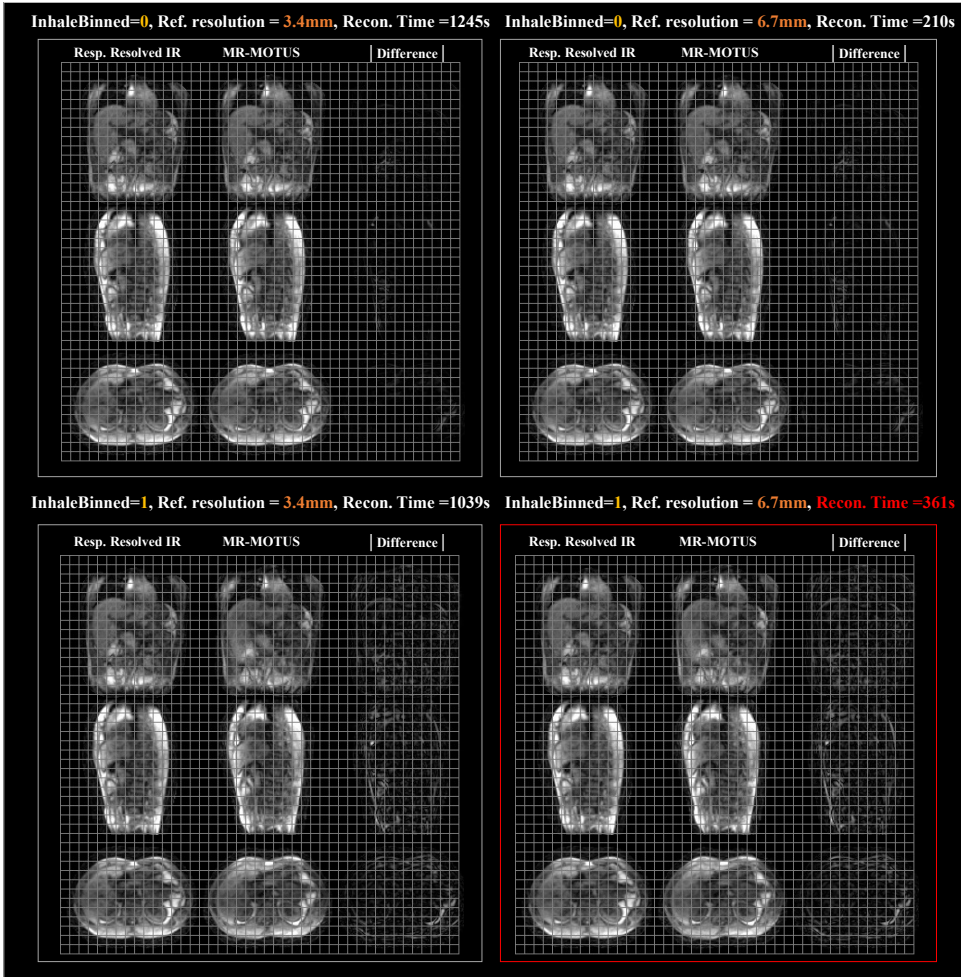


Supporting Video 3.10: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* MR-MOTUS warped reference images resulting from the 3D head-and-neck motion reconstructions for volunteer 1, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4.

3

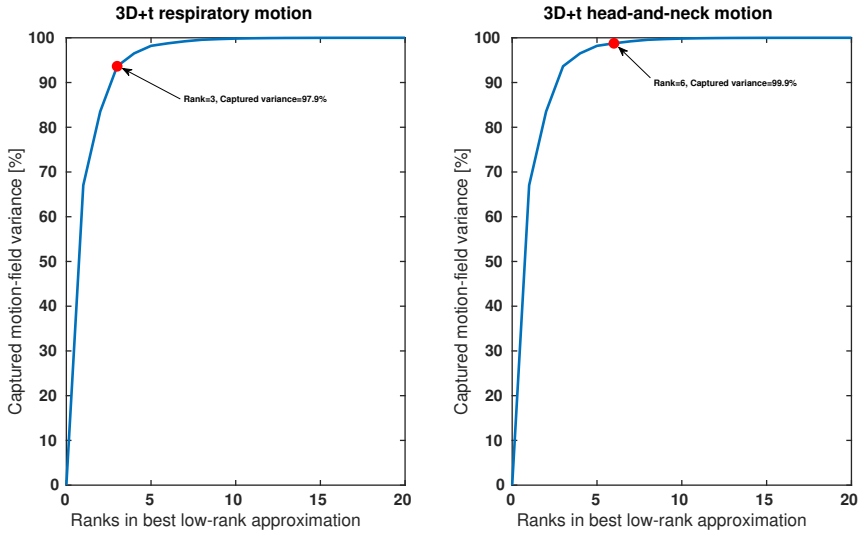


Supporting Video 3.11: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* MR-MOTUS warped reference images resulting from the 3D head-and-neck motion reconstructions for volunteer 2, as mentioned in Supporting Information Section 3.3 and Supporting Information Section 3.4.



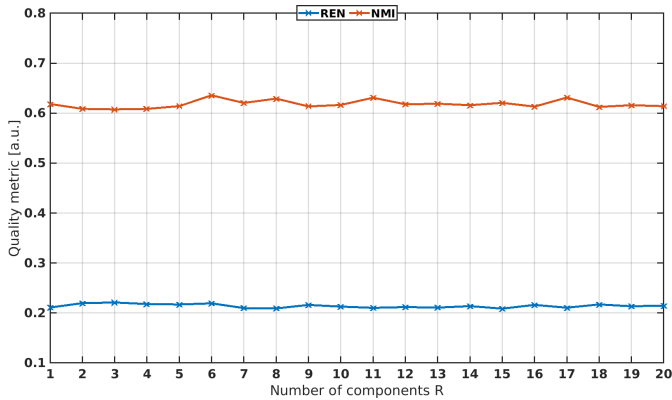
3

Supporting Video 3.12: *This is an animated figure and should be viewed online at <https://doi.org/10.6084/m9.figshare.20480571.v1>.* Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Supporting Information Section 3.7 and Supporting Information Section 3.3. The four blocks show reconstructions with different reconstruction parameter settings. ‘InhaleBinned’ denotes whether the reference image is binned in inhale (1) or exhale (0). ‘Ref. resolution’ denotes the resolution of the reference image in millimeters. All motion-fields were reconstructed without regularization and with 9 cubic spline functions in every direction.

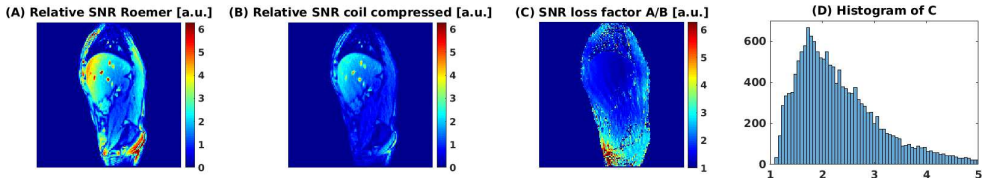


Supporting Figure 3.1: Results of the singular value analyses in Section 3.7 for 3D+t respiratory motion (left), and 3D+t head-and-neck motion (right). This figure clearly indicates that 3D+t motion-fields possess the low-rank property; models with $R = 3$ and $R = 6$ can respectively capture 97.9% and 99.9% of the variance of 3D+t respiratory motion and 3D+t head-and-neck motion, allowing for a significant reduction in the number of unknowns.

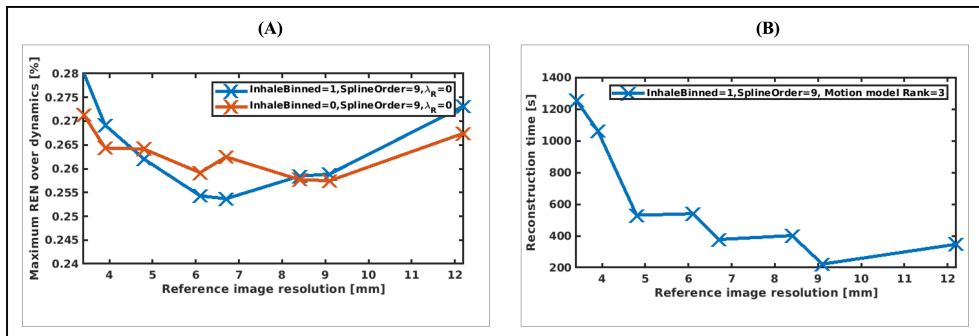
3



Supporting Figure 3.2: This figure shows the effect of a different number of components $R = 1, \dots, 20$ on the respiratory motion reconstructions, as discussed in Section 3.7. The metrics were evaluated on respiratory-resolved MR-MOTUS and image reconstructions, and the means over all 20 reconstructed respiratory phases are visualized in this figure. Only a minimal change can be observed in both metrics, showing that the effect of R on the results is minimal for respiratory motion. NMI = Normalized Mutual Information, NRMSE = Normalized Root Mean Squared Error.



Supporting Figure 3.3: (A) Roemer reconstruction [104]. (B) The coil compression with $\lambda_{CC} = 5 \cdot 10^5$, as discussed in Section 3.7. (C) SNR loss factor between (A) and (B). (D) The histogram of the SNR loss factor in (C). The SNR loss factor is between 1.5 and 2.5 in most of the body and increases towards the boundary of the body.



Supporting Figure 3.4: Results of the parameter search as mentioned in Section 3.7 and Supporting Information Section 3.3. (A) The effect of the reference image resolution and reference image respiratory binning-phase on the reconstruction quality. (B) The effect of the reference image binning on the reconstruction time. In the figure, ‘InhaleBinned’ refers to the binning phase for the reference image (InhaleBinned=1 for inhale, InhaleBinned=0 for exhale), ‘Resolution’ denotes the spatial resolution of the reference image, and ‘SplineOrder’ denotes the number of spline basis functions defined per spatial dimension.

Analysis of the number of components R in the low-rank motion model

To test the hypothesis that the low-rank motion model can accurately represent motion-fields with a small number of components R , a singular value analysis was performed on 3D+t abdominothoracic and head motion-fields. Motion-fields with a large degree of freedom in the reconstructions were reconstructed from Eq. (3.9) by setting the number of components to 20, and the regularization to zero. Then, singular values σ_j were computed by means of an SVD on the Casorati matrix of the resulting spatio-temporal motion-fields (spatio-temporal motion-fields ordered in a matrix with rows as space, and columns as time). Finally, the variance of the motion-fields captured by models with a reduced number of components R was analyzed by computing $\sum_{j=1}^R \sigma_j^2 / \sum_{j=1}^{20} \sigma_j^2$. See Supporting Figure 3.1 for the results of the analyses.

For both 3D+t abdominothoracic and head motion-fields, the resulting variance plots show a steep increase. This indicates that a low-rank assumption on the motion-fields is reasonable. The figure also shows that models with $R = 3$ and $R = 6$ for respectively 3D+t respiratory and head motion capture a significant amount of the variance (97.8% and 99.9% respectively) in spatio-temporal motion-fields that were reconstructed with a large degree of freedom.

In addition to the singular value analysis, the effects of the model order R on the results were quantitatively analyzed for 3D+t respiratory-resolved motion-fields. For this analysis the normalized root mean squared error (NRMSE), and normalized mutual information (NMI) were computed between MR-MOTUS reconstructions with $R = 1, \dots, 20$ and image reconstruction on respiratory-resolved data. For the MR-MOTUS reconstructions, the reconstruction parameters listed in Table 3.1 were used. The means over all 20 reconstructed respiratory phases are shown in Supporting Figure 3.2. Only minor changes in the metrics can be observed between the different R -values. Based on these analyses, combined with visual inspections of the results, we have set $R = 3$ for respiratory motion, and $R = 6$ for head-and-neck motion.

Extension of MR-MOTUS to multi-coil acquisitions

In Huttinga *et al.* [84], the model in Eq. (3.3) was derived by substituting

$$q_t(\mathbf{x}) = q_0(\mathbf{U}_t(\mathbf{x})) \det \nabla(\mathbf{U}_t(\mathbf{x})) \quad (\text{S3.1})$$

in the conventional MRI signal equation

$$s_t(\mathbf{k}) = \int_{\Omega} q_t(\mathbf{x}) e^{-i2\pi\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}, \quad (\text{S3.2})$$

followed by a change-of-variables. In a multi-coil acquisition, k-space data $s_t^j(\mathbf{k})$ from coil j with coil sensitivity $S^j(\mathbf{x}) \in \mathbb{C}$ can be modelled as

$$s_t^j(\mathbf{k}) = \int_{\Omega} S^j(\mathbf{x}) q_t(\mathbf{x}) e^{-i2\pi\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}. \quad (\text{S3.3})$$

Substitution of Eq. (S3.1) in Eq. (S3.3), followed by a change-of-coordinates $\mathbf{x} \mapsto \mathbf{U}^{-1}(\mathbf{x})$ results in a signal model similar to the one used in this work. The change-of-coordinates will make the intensity variations due to the coil sensitivities dependent on the unknown motion-fields:

$$s_t^j(\mathbf{k}) = \int_{\Omega} S^j(\mathbf{U}_t^{-1}(\mathbf{x}))q_0(\mathbf{x})e^{-i2\pi\mathbf{k}\cdot\mathbf{U}_t^{-1}(\mathbf{x})}d\mathbf{x}. \quad (\text{S3.4})$$

Unfortunately, there is a practical problem with the model Eq. (S3.3): the reconstruction times are increased by a factor N_c for N_c coils, since N_c times more NUFFT are required. Additionally, it will require a warping $S^j(\mathbf{U}_t^{-1}(\mathbf{x}))$ at every dynamic, which further increases the reconstruction times. Altogether, this makes the reconstruction less practical.

In theory, the model could be combined with any available k-space-based coil compression technique that gives access to the coil sensitivities that modulate the compressed k-space data. In practice, this requires knowledge of the coil sensitivities of the compressed data, which are possibly different at every dynamic. We have therefore opted for a linear coil compression with coefficients $c^j \in \mathbb{C}$ that produces a trade-off between homogeneous coil sensitivities and SNR-loss. The resulting approximately homogeneous coil sensitivity allows to neglect the coil sensitivities in the signal model, since $\sum_j c^j S^j(\mathbf{U}_t^{-1}(\mathbf{x})) \approx 1$ for all \mathbf{x} . Combining both requirements, the compression coefficients can be obtained by solving the following minimization problem

$$\min_{\mathbf{c}} \|\mathbf{S}\mathbf{c} - \mathbf{1}\|_2^2 + \lambda_{cc}\|\mathbf{U}\mathbf{c}\|_2^2. \quad (\text{S3.5})$$

Here $\mathbf{S} \in \mathbb{C}^{N \times N_c}$ denotes the matrix of coil sensitivities, $\mathbf{1} \in \mathbb{C}^{N \times 1}$ denotes a vector with all entries equal to one, and \mathbf{U} denotes Cholesky factor of the noise covariance matrix $\mathbf{\Sigma}$ such that $\mathbf{\Sigma} = \mathbf{U}^H \mathbf{U}$. The minimization problem in Eq. (S3.5) has a unique solution given by

$$\mathbf{c} = [\mathbf{S}^H \mathbf{S} + \lambda_{cc} \mathbf{\Sigma}]^{-1} \mathbf{S}^H \mathbf{1}, \quad (\text{S3.6})$$

where the superscript H denotes the conjugate transpose.

Applying the compression coefficients in Eq. (S3.6) to both sides of Eq. (S3.3) yields:

$$\tilde{\mathbf{s}}_t = \mathbf{F}(\mathbf{D}_t | \tilde{\mathbf{q}}_0). \quad (\text{S3.7})$$

Here $\tilde{\mathbf{s}}_t := \sum_j c^j \mathbf{s}_t^j$ and $\tilde{\mathbf{q}}_0 := \sum_j c^j S^j(\mathbf{x})\mathbf{q}_0(\mathbf{x}) \approx \mathbf{q}_0(\mathbf{x})$ denote the coil-compressed k-space data and reference image, respectively, and can be pre-computed before the motion-field reconstruction. Note that this includes the same operator as in Eq. (4.2), such that the same reconstruction strategy as for single-channel data can now be employed to reconstruct motion-fields from multi-coil acquisitions. This thus allows for similar reconstruction times with multi-coil data, but comes at the cost of SNR loss. For all experiments considered in this work we have set the compression parameter $\lambda_{CC} = 5 \cdot 10^5$; this yielded an empirically good trade-off between SNR loss and homogeneous sensitivities.

As mentioned above, the proposed compression is not optimal in terms of SNR [104, 105]. To this extent, Supporting Figure 3.3 analyzes the loss in SNR between a Roemer coil combination [104] and the proposed coil compression on the

2D respiratory data. This shows an SNR loss factor between 1.5 and 2.5 in most of the body, which increases towards the boundary of the body. k-Space-based parallel imaging methods such as GRAPPA could possibly be used to improve the SNR after the compression.

Reference image reconstruction from free-breathing, multi-coil acquisitions

MR-MOTUS requires a motion-free reference image with slowly varying coil sensitivity. To be consistent with the forward model in Eq. (S3.7), the reference image $\tilde{\mathbf{q}}_0$ should be reconstructed from the data $\tilde{\mathbf{s}}_t$ after coil compression. This data was, however, acquired during free-breathing which makes the reconstruction of a motion-free image not trivial. The extreme positions in the respiratory motion (i.e. end-inhale, end-exhale) can be assumed relatively constant over different breathing cycles. To extract motion-free k-space data from the free-breathing data, we retrospectively selected only the readouts acquired at end-inhale; this yielded consistently better results than end-exhale. To enable this selection, we performed phase-binning on every readout based on the magnitude of the signal at $\mathbf{k} = \mathbf{0}$ (denoted as k_0 -values). A low-pass filter was applied to the time-series of k_0 -values prior to the binning, based on the expected breathing frequency. Next, phase-binning was performed with 10 phases per breathing cycle. Finally, the reference image was reconstructed from the end-inhale bin, using a density-compensated preconditioned conjugate-gradient reconstruction.

Motion reconstruction details

We followed the approach outlined in Supporting Information Section 3.2, and reconstructed motion-fields from multi-coil k-space data acquired during motion by solving the minimization problem (3.9). The low-rank MR-MOTUS workflow is schematically summarized in Figure 3.1. The reconstruction problem (3.9) was solved with L-BFGS [66], using the MATLAB implementation from Becker [94]. The L-BFGS memory parameter was set to 20 and sampling density compensation [95, 108] was applied to improve the conditioning of the reconstruction. The elements in both of the matrices Φ and Ψ were initialized as uniformly random draws from $[-0.5, 0.5]$, after which the matrices were scaled to unit Frobenius norm. This initialization yields equally sensitive updates in Φ and Ψ during the starting phase of the reconstruction, and lead to consistently good results both in simulations and experimental settings despite the non-convex nature of the reconstruction problem. The multi-coil data was compressed to a single channel prior to all reconstructions, see Section 3.7 for more details. The regularization parameter λ_R was chosen according to $\lambda_R \sim 1/M$ and the data \mathbf{s}_t were scaled by the norm of the density-compensated k-space data in order to obtain consistent values between experiments. The reconstruction parameters were determined with a heuristic parameter search (see Section 3.7). The forward and adjoint operators were implemented using the type-3 NUFFT from Barnett *et al.* [64]. We refer to Table 3.1 for all parameter settings and to Section 3.7 and the Supporting Information in Huttinga *et al.* [84] for more implementation details.

Code that produces similar results as presented in this study is openly available at https://github.com/nrfhuttinga/LowRank_MRMOTUS.git.

The forward model in (3.3), and corresponding derivatives (see Supporting Information in Huttinga *et al.* [84]) required for the motion-field reconstruction problem (3.9) are computed using the NUFFT-operator from Barnett *et al.* [64]. We observe that the evaluation of the data-fidelity term $E[\cdot]$ in the objective function (3.4) includes a sum over M independent terms. We found that a so-called ‘embarrassingly parallel’ evaluation of the data-fidelity term can result in a reduced computation time. This requires a single-thread compilation of the NUFFT, and a parallelization over the M independent terms in the summation. The parallelization was realized in MATLAB with a ‘parfor-loop’ over the dynamics. For 3D this resulted in a significant improvement in computation times, while for 2D the speed-up due to parallelization was smaller than the thread initialization overhead. Hence, for 2D we consider the non-parallelized evaluation and for 3D we consider the parallelized evaluation.

Heuristic search for reconstruction parameter settings

We have heuristically determined the reconstruction parameter settings based on the quality of the respiratory-resolved MR-MOTUS reconstructions described in Supporting Information Section 3.3. The number of dynamics was set equal to the number of bins, i.e. 20, and as a quality metric we have taken the maximum NRMSE between the reconstructions over all 20 dynamics. This allowed to select the parameters with the best worst-case performance over all respiratory phases. The reconstruction parameters included the resolution of the reference image, the binning-phase of the reference image, the spline order for the motion-field basis and the number of motion model components R . The optimal reconstruction parameters were determined as follows. First, the spline order was fixed to 9 (i.e. 9 spline functions over the whole FOV, in every direction), in order to have at least the 4 points required to correctly define a single third order B-spline on every resolution level. A grid search was then performed for the reference image resolution, binning-phase and the number of motion model components, with as metric the maximum NRMSE over all dynamics. Next, the optimal parameters resulting from the grid search were selected, the spline order was increased to the maximum (i.e. 1 spline function per 4 grid-points), and a second grid search was performed for only the regularization parameter λ_R . To vary the spatial resolution of the reference image we have retrospectively varied the k_{\max} of the data, resulting in varying matrix sizes and, consequently, reconstruction times. To vary the respiratory-phase, the binning as described in Supporting Information Section 3.7 is performed to select only the required phase.

Supporting Figure 3.4 visualizes the quantitative results of the parameter search. Supporting Figure 3.4A shows the NRMSE of the worst-case dynamic, minimized over the number of motion model components ($R = 1 - 6$), versus the reference image resolution. No regularization was applied yet. The best worst-case performance was obtained for a reference image resolution binned in end-inhale with a spatial resolution of 6.7 mm isotropic, and motion model with $R = 3$. This reference resolution was subsequently fixed, the motion resolution was set to the maximum of 1 spline function per 4 grid-points (i.e. spline order 16), and the optimal value for

the regularization parameter was determined with a grid-search. This resulted in an optimum value of $\lambda_R = 15$. Supporting Video 3.12 qualitatively compares the MR-MOTUS reconstructions with respiratory-resolved image reconstructions for different reference image binning-phases and reference image resolution. Although the visual differences between the reconstructions are minimal, it can be observed from Supporting Figure 3.4B that the reconstruction times are about a factor 3 lower for a reference image resolution of 6.7 mm, than for a resolution of 3.4 mm.

Chapter **4**

Real-time non-rigid 3D respiratory motion estimation using MR-MOTUS

Niek R.F. Huttinga
Tom Bruijnen
Cornelis A. T. van den Berg
Alessandro Sbrizzi

The following chapter is based on:

Real-time non-rigid 3D respiratory motion estimation for MR-guided radiotherapy using MR-MOTUS. IEEE Transactions on Medical Imaging, vol. 41, no. 2, p. 332-346, 2022.

Abstract

The MR-linac is a combination of an MR-scanner and radiotherapy linear accelerator (linac) which holds the promise to increase the precision of radiotherapy treatments with MR-guided radiotherapy by monitoring motion during radiotherapy with MRI, and adjusting the radiotherapy plan accordingly. Optimal MR-guidance for respiratory motion during radiotherapy requires MR-based 3D motion estimation with a latency of 200-500 ms. Currently this is still challenging since typical methods rely on MR-images, and are therefore limited by the 3D MR-imaging latency. In this work, we present a method to perform non-rigid 3D respiratory motion estimation with 170 ms latency, including both acquisition and reconstruction. The proposed method called real-time low-rank MR-MOTUS reconstructs motion-fields directly from k -space data, and leverages an explicit low-rank decomposition of motion-fields to split the large scale 3D+t motion-field reconstruction problem posed in our previous work into two parts: (I) a medium-scale offline preparation phase and (II) a small-scale online inference phase which exploits the results of the offline phase for real-time computations. The method was validated on free-breathing data of five volunteers, acquired with a 1.5T Elekta Unity MR-linac. Results show that the reconstructed 3D motion-field are anatomically plausible, highly correlated with a self-navigation motion surrogate ($R = 0.975 \pm 0.0110$), and can be reconstructed with a total latency of 170 ms that is sufficient for real-time MR-guided abdominal radiotherapy.

Supporting Videos

The supporting videos corresponding to this chapter can be accessed through <https://doi.org/10.6084/m9.figshare.20480667.v2> or the following QR-code:



4.1 Introduction

The physiological movement of organs during radiotherapy is a source of uncertainty, and generally reduces the precision of the treatments. Such motion can typically be related to respiration, digestion and cardiac contractions. Recently, the MR-linac was introduced as a combination of an MR-scanner and a radiotherapy linear accelerator (Linac) [7–10]. The MR-linac holds the promise to increase the precision of radiotherapy treatments through MR-guided radiotherapy (MRgRT) by monitoring physiological motion with the MRI, and performing corresponding radiotherapy plan adaptations on the LINAC. MRgRT consists of a two-fold implementation: inter-fraction and intra-fraction MRgRT. In inter-fraction MRgRT, the radiation plan should be adjusted between treatments based on the daily changes in anatomy detected with pre-treatment MRI [109]. Inter-fraction MRgRT is feasible with the currently available techniques, and is already being applied in the clinic with success [41, 110]. In intra-fraction real-time adaptive MRgRT, MR-based motion monitoring and radiation plan adaptations should be performed during the treatments in a real-time loop, for which the required latency is determined by the speed of the targeted motion. A maximum total latency (i.e. acquisition and tracking) of 200-500 ms would be required for 3D motion estimation to achieve the full potential of the MR-linac and optimally compensate for respiratory motion with MRgRT [22, 23]. However, due to the relatively slow imaging speed of an MR-scanner, achieving this latency for 3D motion estimation is still a technical challenge.

Several respiratory motion estimation methods have been proposed in the context of MRgRT over the last years, most of which estimate motion from images. Unfortunately, these methods are directly limited by the latency of MR-imaging, including both acquisition and reconstruction time. To circumvent this, a typical approach is to estimate 3D motion from lower-dimensional cine-MR-images. For example, orthogonal 2D cine-images (i.e. 2.5D) can be acquired in order to estimate 3D motion at high temporal resolution [25, 32, 111]. In Stemkens *et al.* [33], a pre-trained 3D motion model was fit to incoming 2D cine-images to obtain fast motion estimates. Additionally, surrogate-driven motion models have been proposed that relate cine-MRI-derived surrogate signals to motion-fields [36, 112]. Another recent work by Feng *et al.* [34] rapidly generates 3D MR-images by determining the best match between the current 1D motion state and 1D motion states in a 3D+t respiratory-resolved image reconstruction. For a more detailed review of related methods for MRgRT we refer to Stemkens *et al.* [24] and Paganelli *et al.* [25], and for an in silico comparison of several related methods we refer to Paganelli *et al.* [113].

In this work, we focus on the aforementioned technical challenge in MRgRT and propose real-time low-rank MR-MOTUS [84, 114] for real-time estimation of non-rigid 3D respiratory motion-fields directly from prospectively undersampled 3D k -space data, with a total latency (acquisition and reconstruction) below 200 ms. The MR-MOTUS signal model relates motion-fields and a reference image to k -space data, allowing to reconstruct motion-fields directly from k -space data, given a reference image [84]. In Huttinga *et al.* [114], MR-MOTUS was extended to reconstruct 3D+t motion-fields in the order of minutes, using an explicit low-rank factorization in static spatial motion components and dynamic temporal motion components. Here, the

same low-rank factorization is assumed, and the framework is extended to perform real-time computations by observing that a low-rank factorization allows to split the reconstruction in two phases: (I) an offline preparation phase that separates static spatial motion components from dynamic temporal motion components, and (II) an online phase in which the pre-trained static motion components are fixed and dynamic motion components are estimated with real-time acquisitions and real-time model-based reconstructions. Real-time low-rank MR-MOTUS is applied *in silico* to a digital anatomical phantom, and *in vivo* to 5 volunteers whose data were acquired with an MR-linac. The reconstructed motion-fields are validated *in silico* in terms of end-point-errors with ground-truth motion-fields, and *in vivo* in terms of anatomical plausibility, accuracy and correspondence with a conventional respiratory-resolved compressed sensing reconstruction.

4.2 Theory

Background MR-MOTUS

Forward signal model

We assume a general d -dimensional setting, with targeted case $d = 3$, and we follow the convention that bold-faced characters denote vectorizations. We define $\mathbf{x}_0 \mapsto \mathbf{x}_t$ as the mappings from coordinates $\mathbf{x}_0 \in \mathbb{R}^d$ in a reference image to new locations $\mathbf{x}_t \in \mathbb{R}^d$ at time t . The mappings are characterized by the motion-fields \mathbf{d}_t through $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{d}_t(\mathbf{x}_0)$. This will be written in concatenated vector-form as

$$\mathbf{X}_t = \mathbf{X}_0 + \mathbf{D}_t, \quad (4.1)$$

where $\mathbf{X}_t, \mathbf{X}_0, \mathbf{D}_t \in \mathbb{R}^{Nd \times 1}$ denote the vertical concatenations over N spatial points in a d -dimensional setup. The MR-MOTUS forward model [84] explicitly relates the motion-fields \mathbf{D}_t and a static reference image $\mathbf{q}_0 \in \mathbb{C}^N$ to dynamic, single-channel (and possibly non-Cartesian) k -space measurements $\mathbf{s}_t \in \mathbb{C}^{N_k}$:

$$\mathbf{s}_t = \mathbf{F}(\mathbf{D}_t | \mathbf{q}_0) + \boldsymbol{\epsilon}_t. \quad (4.2)$$

Here $\boldsymbol{\epsilon}_t \in \mathbb{C}^{N_k}$ is the complex noise vector and $\mathbf{F} : \mathbb{R}^{Nd} \mapsto \mathbb{C}^{N_k}$ is the discretization of the forward operator defined as

$$F(\mathbf{d}_t)[\mathbf{k}] = \int_{\Omega} q_0(\mathbf{x}_0) e^{-i2\pi \mathbf{k} \cdot [\mathbf{x}_0 + \mathbf{d}_t(\mathbf{x}_0)]} d\mathbf{x}_0, \quad (4.3)$$

where $\mathbf{k} \in \mathbb{R}^d$ denotes the k -space coordinate. Motion-fields can be reconstructed directly from k -space measurements by exploiting the availability of a fixed reference image \mathbf{q}_0 , and subsequently fitting the nonlinear signal model Eq. (4.3) to acquired k -space data. We refer the reader to our previous works [84, 114] for an extensive discussion on the assumptions underlying the signal model Eq. (4.3).

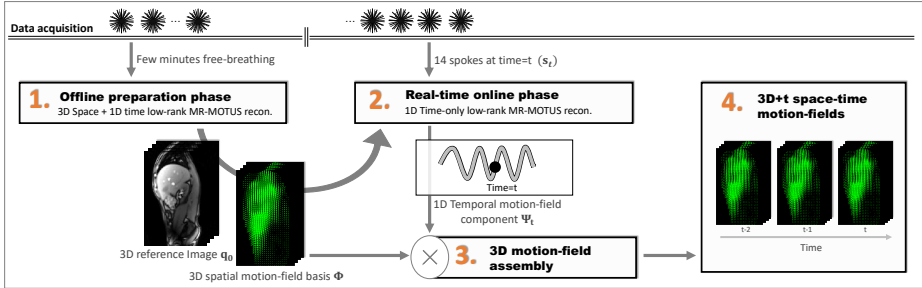


Fig. 4.1: An overview of the real-time MR-MOTUS reconstructions, as described in Section 4.2. (1) A reference image and a spatial motion-field basis are reconstructed in an offline preparation phase from 10 minutes of data acquired in free-breathing. (2) The result of the offline phase is exploited, and only the low-dimensional dynamic representation coefficients in the spatial motion-field basis are reconstructed from just 14 spokes with a latency of just 170 ms (data acquisition + reconstruction). (3) A 3D motion-field per time instance is assembled using the spatial motion-field basis (offline) and representation coefficients (online). (4) Finally, this results in the real-time reconstruction of 3D+t motion-fields.

Retrospective MR-MOTUS reconstructions of space-time motion-fields

Reconstructing motion-fields over a longer period of time allows to exploit correlations in both space and time. However, this requires a large number of parameters, and therefore significantly increases memory consumption and reconstruction times. As shown in Huttinga *et al.* [114], a low-rank motion model can be employed that reduces memory consumption and adds a regularization in both space and time simultaneously. This model splits space-time motion-fields $\mathbf{D} \in \mathbb{R}^{Nd \times M}$ in a spatial component and temporal component as follows

$$\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_M] = \Phi \Psi^T. \quad (4.4)$$

The first component Φ , is the spatial component that models directions and magnitude of motion per voxel. The second, Ψ , is the temporal component that models the global scalings along these directions. Both components consist of R rank-1 sub-components, i.e. $\Phi \in \mathbb{R}^{Nd \times R}$ and $\Psi = [\Psi_1^T, \dots, \Psi_M^T]^T \in \mathbb{R}^{M \times R}$, where M denotes the number of dynamics. Since internal body motion (e.g. respiratory motion) typically occurs along similar directions over time (e.g. feet-head), this motion model allows for significantly compressed representation with $R \ll M$. Since $\text{rank}(\Phi \Psi_t^T) \leq R \ll M$, we refer to Eq. (4.4) as a low-rank motion model. The motion model components can be obtained by solving the following reconstruction problem[114]:

$$\underset{\Phi, \Psi}{\text{argmin}} \sum_{t=1}^M \|\mathbf{F}(\Phi \Psi_t^T) - \mathbf{s}_t\|_2^2 + \lambda \mathcal{R}(\Phi \Psi_t^T), \quad (4.5)$$

and can subsequently be assembled to a space-time motion-field through Eq. (4.4). In Eq. (4.5), the first term models the data fidelity, \mathcal{R} is a regularization term that incorporates a-priori assumptions, and $\lambda \in \mathbb{R}^+$ is the regularization parameter that balances both terms. The reconstruction time for Eq. (4.5) scales with the number of dynamics M , which is typically large to capture large-scale dynamics. Alternatively, motion model components can be reconstructed on respiratory-resolved (rr) data,

$$\operatorname{argmin}_{\Phi, \Psi} \sum_{t=1}^{M^{\text{rr}}} \|\mathbf{F}(\Phi \Psi_t^T) - \mathbf{s}_t^{\text{rr}}\|_2^2 + \lambda \mathcal{R}(\Phi \Psi_t^T), \quad (4.6)$$

where \mathbf{s}_t^{rr} denotes k -space data that is sorted into M^{rr} respiratory phases. Since typically $M^{\text{rr}} \ll M$, and the number of readouts per dynamic in $|\mathbf{s}_t^{\text{rr}}|$ is larger than in $|\mathbf{s}_t|$, this results in a reconstruction problem with better conditioning and a reduced reconstruction time that is beneficial in practice. Due to the practical benefits we will consider Eq. (4.6) in this work.

Extension to real-time reconstructions: framework overview

In Huttinga *et al.* [114] it was shown that high temporal resolution space-time motion-fields can be reconstructed by solving Eq. (4.5) or Eq. (4.6). However, despite the low-rank factorization and respiratory sorting, Eq. (4.6) is still a medium-scale reconstruction problem, resulting in reconstruction times in the order of minutes. Retrospectively, the reconstructed space-time motion-fields are valuable for e.g. the assessment of dose accumulated during treatments, but the long reconstruction times prevent the direct application of the framework to real-time MRgRT.

We observe that reconstructions with the low-rank model yield a convenient representation that allows for significant reduction in computation times. All motion-fields $\mathbf{D}_1, \dots, \mathbf{D}_M$ are represented as a linear combination of the R columns of Φ , with the R representation coefficients given as the columns of Ψ (Eq. (4.4)). It has empirically been shown that realistic respiratory motion can be represented with $R = 1 \dots 3$, i.e. with few basis functions and few representation coefficients [33, 114].

These observations suggest a strategy to reduce the computational burden of Eq. (4.6) by splitting the reconstruction in two phases: I) a medium-scale offline preparation phase that reconstructs Φ from data acquired during representative motion, and II) a small-scale inference phase that exploits the availability of Φ and only reconstructs the few representation coefficients per dynamic in real-time:

$$\{\Phi^{\text{rr}}, \Psi^{\text{rr}}\} = \operatorname{argmin}_{\Phi, \Psi} \sum_{t=1}^{M^{\text{rr}}} \|\mathbf{F}(\Phi \Psi_t^T) - \mathbf{s}_t^{\text{rr}}\|_2^2 + \lambda \text{TV}(\Phi \Psi_t^T), \quad (4.7)$$

$$\{\Psi_t\} = \operatorname{argmin}_{\Psi_t} \|\mathbf{F}(\Phi^{\text{rr}} \Psi_t^T) - \mathbf{s}_t\|_2^2 + \mu \|\Psi_t - \Psi_{t-1}\|_2^2. \quad (4.8)$$

Here $\mu > 0$ is a regularization parameter that stabilizes the real-time reconstructions by penalizing large deviations from the solution at the previous dynamic. Moreover, "TV" is defined as the vectorial total variation, computed as the L^2 -norm over the

total variation per motion-field direction [38]:

$$\text{TV}(\mathbf{D}_t) := \sqrt{\sum_{p=1\dots d} \left(\sum_i \|\nabla \mathbf{D}_t^p\|_2 \right)^2},$$

where $[\nabla \cdot]_i$ denotes the gradient at the i -th spatial coordinate, and the superscript p denotes the motion-field direction. The first phase of the reconstruction, Eq. (4.7), consists of a medium-scale reconstruction problem which can be solved offline in the order of minutes on a desktop PC, for respiratory-sorted data \mathbf{s}^{rf} in a training set. The second phase, Eq. (4.8), consists of an extremely small-scale reconstruction problem with typically only 1-3 unknowns per dynamic, which can be solved online in the order of few milliseconds on a desktop PC, for dynamics not present in the training set. In practice, the first phase can be performed offline during the radiotherapy treatment preparation, and the second phase online with minimal latency during irradiation. Figure 4.1 schematically illustrates the main steps of the workflow.

4.3 Methods

To evaluate the (source of) local motion-field errors in the proposed method, several validation experiments were performed with a digital anatomical XCAT phantom [106] with realistic respiratory motion. Real-time MR-MOTUS reconstructions were performed for 5 volunteers whose data were acquired on an MR-linac, during free-breathing, and with a multi-channel radiolucent receive array. The anatomical plausibility of the reconstructed in vivo motion-fields was evaluated by means of the Jacobian determinant, allowing to detect possibly unrealistic compression or expansion induced by the reconstructed motion-fields. Additionally, the global accuracy of the in vivo motion-fields was assessed by means of the Pearson correlation and Bland-Altman difference plots between the reconstructed motion in feet-head direction, and a 1D respiratory motion surrogate. This allowed for a validation of the 3D MR-MOTUS motion-fields at high temporal resolution (6.7 Hz). Finally, the reconstructions in the offline phase were qualitatively compared to respiratory-resolved compressed sensing reconstructions in 3D. More details on the in silico and in vivo experiments will be discussed in, respectively, Section 4.3 and Section 4.3. We first describe the complete reconstruction pipeline in Section 4.3, which is also visualized in detail in Figure 4.2. All computations in this work were performed in Matlab 2019a (The MathWorks Inc., Naticks, Massachusetts). Representative code to perform similar reconstructions will be made available at https://github.com/nrfhuttinga/Realtime_MRMOTUS.

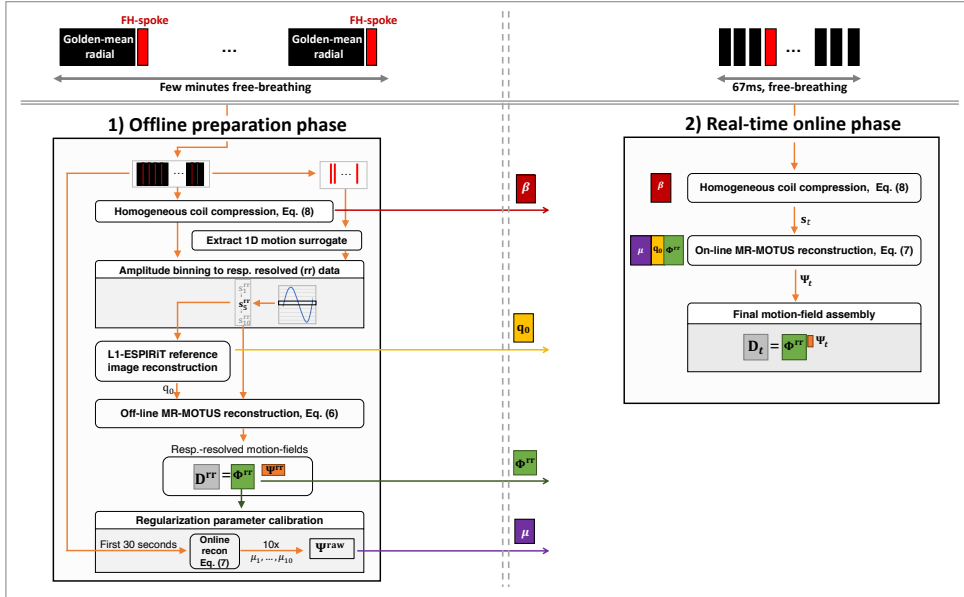


Fig. 4.2: A detailed overview of the reconstruction pipeline. Several components are reconstructed in the offline phase and exploited in the real-time phase to reduce the computation time to 200 ms per dynamic: coil compression coefficients β , reference image q and spatial motion-field basis Φ^{rr} . All steps in the figure are elucidated in Section 4.3.

4

Real-time reconstruction pipeline

Data acquisition/simulation

In practice all data were acquired or simulated during free-breathing. In vivo, a multi-element receive array was used and a steady-state spoiled gradient echo sequence (SPGR) was employed, $TR = 4.8$ ms, $TE = 1.8$ ms, $FA = 20^\circ$, $FOV = 30$ cm \times 30 cm \times 30 cm, and $BW = 540$ Hz. A 3D golden-mean (GM) radial kooshball trajectory [92] was employed, interleaved every 31 spokes with a self-navigation spoke oriented along the feet-head (FH) direction. The 3D GM kooshball trajectory efficiently acquires 3D k -space data with relatively uniform angular distribution at all temporal resolutions [92], while the self-navigation spoke yields a motion surrogate every $31 \cdot TR = 148.8$ ms, allowing for retrospective validation of the reconstructed motion-fields in FH direction at 6.7 Hz temporal resolution [79, 115].

Offline preparation phase

The offline preparation phase consists of several steps, outlined below.

Homogeneous coil compression. The MR-MOTUS signal model requires single-channel k -space data with approximately homogeneous coil sensitivity [114]. To achieve this, we followed [114] and linearly compressed all data to a single virtual channel prior to MR-MOTUS reconstructions. Compression coefficients $\beta \in \mathbb{C}^{N_c}$ were obtained by solving:

$$\min_{\beta} \|\mathbf{S}\beta - \mathbf{1}\|_2^2 \Rightarrow \beta = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{1}. \quad (4.9)$$

Here N_c denotes the number of channels, $\mathbf{S} \in \mathbb{C}^{N \times N_c}$ the coil sensitivities, and $\mathbf{1} \in \mathbb{R}^N$ an all-one vector.

Surrogate signal extraction and binning. A surrogate was extracted from the self-navigation spoke along FH direction for the purpose of respiratory binning and validation at high temporal resolution. For this we follow the principal component analysis (PCA) approach of Feng *et al.* [79], and extract the surrogate as the principal component with the highest spectral density in the respiratory motion frequency range 0.1 - 0.5 Hz. A cluster of coil elements was determined for which the extracted respiratory motion surrogate signals show high correlation by following the method in Zhang *et al.* [116]. The final surrogate was extracted from k -space data averaged over this cluster. Finally, a low-pass filter was applied to remove remaining high-frequency oscillations.

Subsequently, all acquired data were sorted based on the amplitude of the extracted motion surrogate. This required a surrogate signal value per readout, so a nearest-neighbor interpolation was performed to interpolate the motion surrogate from the temporal resolution of the self-navigation spokes to the temporal resolution of a single spoke. A total of 10 respiratory bins were selected, and we denote the resulting respiratory-resolved data as $\mathbf{s}_1^{\text{rr}}, \dots, \mathbf{s}_{10}^{\text{rr}}$.

Reference image reconstruction. The bin in the respiratory phase halfway between exhale and inhale, i.e. mid-ventilation, containing around 12000 spokes was selected for reference image reconstruction. Two mid-ventilation reference images were reconstructed with L^1 -wavelet ESPiRiT using the BART MRI reconstruction toolbox [70] (regularization parameter: 5e-3, iterations: 550). The first is a low-resolution reference image (6.7 mm isotropic) for subsequent MR-MOTUS motion-field reconstructions, and the second a higher resolution reference image (3 mm isotropic) for visualization purposes.

Offline MR-MOTUS reconstructions. The sorted data was used to perform a respiratory-resolved MR-MOTUS reconstruction with Eq. (4.7), following [114] and using the code that was made available online at https://github.com/nrfhuttinga/LowRank_MRMOTUS. The number of respiratory phases was set to 10, the regularization parameter λ is empirically tuned once and then fixed for all volunteers. In accordance with previous work [33], the number of ranks was set to 1 ($R = 1$). The motion-fields were parameterized with cubic B-spline bases in space (24 splines in AP, LR, and 16 splines in FH) and time (5 splines). All spline coefficients were randomly initialized in $[-0.5, 0.5]$, and scaled such that the sets of spatial and temporal coefficients both had unit norm. The reconstruction was performed with 60 iterations of L-BFGS [66] (using the Matlab wrapper [94]), with reconstruction times in the order of minutes. This reconstruction resulted in the spatial component Φ^{rr} required for the online reconstructions.

Online inference phase

In the online inference phase, k -space data was grouped into dynamics, with 14 spokes per dynamic, and Eq. (4.8) was solved per dynamic. The reconstruction for time t was initialized with the reconstruction at time $t - 1$, and the reconstruction problem was solved with a single iteration of a GPU-accelerated Gauss-Newton scheme. Finally, a 3D motion-field \mathbf{D}_t was assembled per dynamic using the offline reconstructed spatial components Φ^{rr} and the online reconstructed temporal component: $\mathbf{D}_t = \Phi^{\text{rr}} \Psi_t^T$.

Several aspects of the reconstruction were considered to speed up computations. The forward model was implemented with an explicit matrix-vector multiplication, rather than with the type-3 non-uniform FFT (NUFFT) that was used for Eq. (4.7) due to the relatively large overhead of the NUFFT for few k -space samples. The whole online phase was performed on a GPU (Nvidia Quadro K620 2GB) using MATLAB's `gpuArrays`, which resulted in a factor 8 reduction in computation time in comparison with the CPU implementation. The pseudo-code of the reconstruction algorithm and more details on the speed-up steps are provided in the Supporting Information at the end of this Chapter.

The total processing time of all steps above depends on the number of spokes per dynamic. Ideally, a large number of spokes should be selected per dynamic to improve the conditioning of the reconstruction problem, but this increases both the acquisition and the reconstruction time. Therefore, a trade-off has to be made to satisfy the latency requirement of 200 ms for real-time MRgRT. The dependency of the total processing time on the number of spokes per dynamic was analyzed by calculating the mean and standard deviations over 2000 online reconstructions (see Section 4.4, Supporting Information Figure S1). Based on this analysis, only 14 spokes and 8 samples per spoke were selected per dynamic, which resulted in a mean real-time reconstruction time of 103 ms per dynamic. With an acquisition time of $14 \cdot \text{TR} = 67\text{ms}$, this resulted in a total latency of 170 ms, which is well below the latency requirement for real-time MRgRT [22, 23].

In-silico validation: error analysis of real-time 3D MR-MOTUS reconstructions

In-silico validations were performed to evaluate the local errors of the proposed two-step motion-field reconstruction approach, i.e. inferring time-resolved motion-fields using a motion model built on respiratory-resolved data. We analyzed the contributions of both the offline phase and the online phase to this error. Data was simulated using the XCAT digital phantom for respiratory motion [106], to which MR-contrast was manually added, see Figure 4.3 for the resulting XCAT phantom.

Realistic motion-fields

The following aspects were considered to obtain realistic motion-fields. To ensure motion-fields that are invertible and consistent with the deformed XCAT images, the original XCAT motion-fields were post-processed with the recently published framework by Eiben *et al.* [107]. To simulate lower velocity in exhale than inhale, \cos^4 waveforms were used as input to the XCAT framework. Hysteresis was simulated by a phase-delay between chest and abdominal input waveforms. To simulate pseudo-periodic motion, end-exhale and end-inhale position deviations were randomly generated within a range of respectively 1% and 2% of the waveform amplitude. No cardiac motion was applied. See Figure 4.4 for the input waveforms.

Translational error of the two-step reconstruction approach

To evaluate the translational performance of the framework in case of normal breathing in the offline phase and different breathing patterns in the online phase, the offline phase was performed on respiratory-resolved data simulated during normal breathing, and the online real-time phase on data simulated during four breathing patterns: normal breathing, chest-only, abdominal-only and amplitude drifts (see Figure 4.4). Errors in this experiment specifically due to the offline phase were assessed by comparing real-time reconstructions with an offline reconstructed spatial basis Φ , and with a ground-truth Φ , both obtained during normal breathing. The columns of the ground-truth Φ were obtained as the left-singular vectors of the ground-truth motion-fields during normal-breathing.

Data simulation

To simulate data on dynamic XCAT images we proceeded as follows. An end-exhale XCAT image was taken as the reference image, and smooth magnitude and phase variations were added to obtain a complex image with some intra-organ. This reference image was deformed with the post-processed motion-field by cubic interpolation. Each breathing phase consisted of 100 dynamics and five breathing cycles, each with a period of 5 seconds. Data was generated from the deformed images with a type-2 NUFFT [64] evaluated on the same trajectory as for the in vivo experiments, i.e. a golden-mean 3D radial trajectory interleaved with a self-navigation FH spoke every 31 spokes. Complex noise was added to achieve an SNR of approximately 50, and 400



Fig. 4.3: XCAT reference volume for the simulations described in Section 4.3 (left to right: coronal, sagittal, axial) with manually added MR-contrast and a spherical lesion in the liver.

spokes were simulated per dynamic. With the current in vivo acquisition parameters (see Section 4.3) this equals a free-breathing acquisition time of around 3 minutes per breathing phase.

Data processing and reconstruction

The data processing for the offline phase was kept similar to that for in vivo reconstructions as described in Section 4.3, including the reconstruction of the reference image from binned data. Some exceptions are that the coil compression was not required since single-channel data was generated. Moreover, the reference image was reconstructed in end-exhale, and the binning and data-sorting was performed for inhale and exhale separately to increase the sensitivity for the different breathing types. For the online reconstructions, 14 spokes were simulated per dynamic, similar to the in vivo experiments, and online reconstructions were performed as described in Section 4.3. A similar spline basis as for the in vivo experiments was used, and no regularizations were employed ($\mu, \lambda = 0$).

Performance evaluation

Performance in all experiments above was analyzed in terms of end-point-errors (EPEs) between reconstructed and ground-truth motion-fields, evaluating both the complete volumetric spatial distribution and the mean over an ROI defined as a spherical tumor insert in the liver. The maximum displacement of this spherical tumor insert in the normal breathing scenario in Figure 4.4 was 14.8 mm (7 mm in AP, 13 mm in FH, 0 mm in LR), for chest-only it was 7 mm (7 mm in AP, 0 mm in FH, 0 mm in LR), and for abdominal-only 13 mm (0 mm in AP, 13 mm in FH, 0 mm in LR). For the amplitude drift scenario all displacements of the normal breathing scenario were scaled with a factor that linearly increased from 1 to 1.5, resulting in a maximum displacement of 22.1 mm (10.5 mm in AP, 19.5 mm in FH, 0 mm in LR).

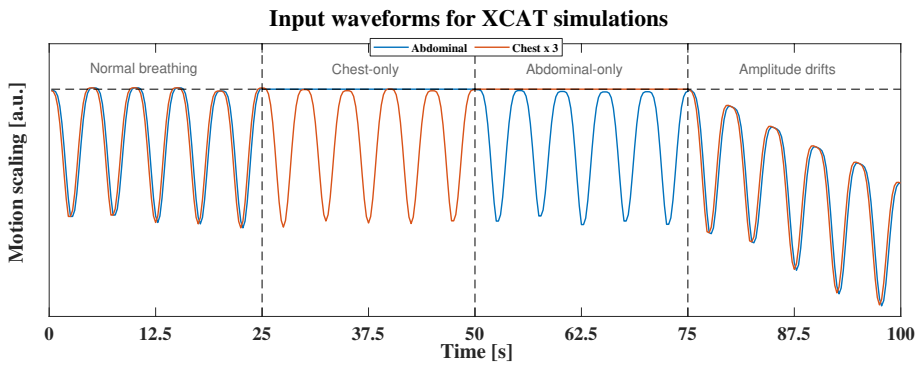


Fig. 4.4: Input abdominal (blue) and chest (orange) waveforms for the XCAT simulation study described in Section 4.3. Here, ‘Chest x 3’ denotes that the amplitude of the actual chest waveform is three times lower than that of the abdominal waveform, resulting in relatively smaller chest motion. Note the varying end-inhale and end-exhale positions, the phase delay causing hysteresis, and the different breathing patterns.

Real-time in vivo reconstructions

For the real-time in vivo reconstruction we follow the pipeline outlined in Section 4.3. Prior to these reconstructions data is acquired and processed according to the steps below.

4

Data acquisition

All data were acquired on a 1.5T MR-linac (Elekta Unity, Elekta AB, Stockholm, Sweden) from 5 healthy volunteers ($BMI \in [19.5; 27.2]$, $Age \in [22; 44]$) during 11:40 minutes of free-breathing, no breathing coaching was performed. The first 10 minutes of data were used for the offline phase (Section 4.3), the last 1:40 minutes were used for the real-time phase (Section 4.3). All experiments were approved by the institutional review board, carried out in accordance with the relevant guidelines and regulations, and written informed consent was obtained from all volunteers prior to the experiments.

Online reconstruction

A total of 100 seconds of dynamics were reconstructed per volunteer, of which the first 30 seconds were used to empirically tune the regularization parameter μ (Eq. (4.8)) per volunteer so as to remove high-frequency oscillations, without over-smoothing.

In vivo anatomical plausibility test with Jacobian determinants

Organs such as the liver, spleen and kidney consist of liquid-filled tissue structures, and can therefore be assumed incompressible and thus volume-preserving during deformation [40]. Volume change due to deformation-fields can be quantified by the determinant of the local deformation-field's Jacobian matrix, which will be referred to as the Jacobian determinant. Hence, for anatomically plausible motion-fields, the Jacobian determinant should be close to unity within organs such as the liver.

The anatomical plausibility of the reconstructed motion-fields was therefore evaluated by means of the Jacobian determinant for the two most extreme positions with respect to the mid-ventilation reference image: end-exhale and end-inhale. The spatial derivatives required for the computation of the Jacobian determinants were obtained through finite differences.

In vivo global accuracy test at high temporal resolution

Alongside the anatomical plausibility, the accuracy of motion-fields is also of great importance for applications such as real-time adaptive MRgRT. However, it is evidently not straightforward to validate 3D motion-fields at high temporal resolution using MR-images, due to the relatively slow imaging speed of MRI. We propose an alternative strategy for to check the global accuracy of the motion-fields. The magnitude of respiratory motion is typically dominant in FH direction [117, 118], and its temporal profile can be extracted at high temporal resolution with PCA on self-navigation spokes along the FH direction[79, 115]. The reconstructed motion-fields should resemble the same temporal profile, since the FH motion that can be extracted from self-navigation spokes along FH should also be present in the reconstructed motion-fields. For these reasons, we validated the motion-fields by correlation analyses between the temporal component of the reconstructed motion-fields Ψ , and a 1D motion surrogate extracted from the self-navigation spokes along FH direction. The surrogate was extracted as described in Section 4.3. In addition to the surrogate, a 1D FFT on the FH self-navigation spokes also yielded high temporal resolution projections of the moving anatomy on the FH-axis.

Qualitative analysis was performed by visual comparison between these projection, the 1D FH motion surrogate extracted with PCA, and the reconstructed temporal MR-MOTUS profile Ψ . Since both signals represent a global scaling, they cannot be compared directly. Therefore, the signals were normalized by subtracting the mean and dividing by their respective standard deviations. Additionally, Bland-Altman plots were generated to assess the similarity. Finally, quantitative analysis was performed by computing the Pearson correlation coefficient between the MR-MOTUS temporal profile and the 1D motion surrogate for all volunteers.

Comparison with respiratory-resolved compressed sensing

The offline-reconstructed respiratory-resolved MR-MOTUS reconstructions were compared to offline respiratory-resolved compressed sensing reconstructions as follows. The reconstructed motion-fields were used to warp the reference image, resulting in respiratory-resolved volumetric images. A compressed sensing reconstruction was performed with BART [70] on the same single-channel respiratory-binned data as was used for the offline MR-MOTUS reconstructions, also resulting in respiratory-resolved volumetric images. The reconstruction was performed over all respiratory bins simultaneously, using 550 iterations, L1-wavelet spatial, and total-variation temporal regularization. Finally, to evaluate the differences between the two reconstructions, the two sets of volumetric images were visually compared side-by-side and in terms of absolute differences.

4.4 Results

In-silico validation: error analysis of real-time 3D MR-MOTUS reconstructions

The end-point-errors between the real-time MR-MOTUS XCAT reconstructions described in Section 4.3 are evaluated over an ROI (spherical lesion in the liver) in Figure 4.5, and over the whole volume in Figure 4.6.

Several conclusions can be drawn from Figure 4.5. The translational performance can be assessed by comparing the same colors over the different scenarios. This shows that an extreme change of breathing patterns, i.e. from normal breathing to either chest-only or abdominal-only, increases the errors in most cases. An exception is the rank-2 model with offline-reconstructed Φ , which shows an improved performance for the extreme scenarios. Comparing the same tints of the different colors per breathing pattern shows an improved performance of the rank-2 models over the rank-1 models in all cases, both with offline reconstructed and ground-truth Φ .

The performance of the complete reconstruction pipeline for the $R = 1$ and $R = 2$ models are visualized in, respectively, light orange and light green in Figure 4.5. The model in light orange was considered for the in vivo experiments described in Section 4.3, and shows acceptable performance for normal breathing and amplitude drifts, but is outperformed by the rank-2 models in the other two extreme scenarios. This indicates that more degrees of freedom would be favorable to model extreme changes in breathing pattern. Interestingly, the results on the rank-2 models show that a model trained on normal breathing can cleanly separate chest and abdominal motion-field components in the online phase.

The contribution of the offline reconstruction phase can be assessed by comparing reconstructions with the offline-reconstructed and ground-truth Φ per breathing pattern. For rank-1 reconstruction this difference is minimal, but for rank-2 reconstructions the steep drop from the errors in the reconstructions with an offline reconstructed Φ to the reconstructions with ground-truth Φ shows that a large portion of the remaining errors can be attributed to the offline phase. An error below 0.75 mm is obtained in all scenarios with $R = 2$ real-time MR-MOTUS reconstructions.

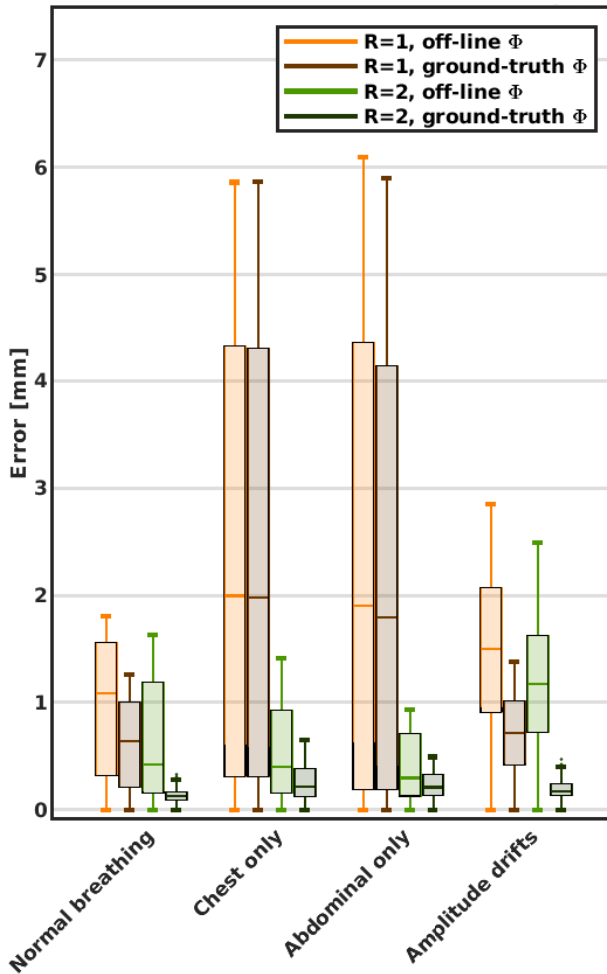


Fig. 4.5: Mean end-point-errors (EPEs) over an ROI (tumor insert) for real-time MR-MOTUS reconstructions on simulated XCAT data (see Section 4.3, Section 4.4).

4

tions and a ground-truth Φ (dark green). The dark green bars also show that the contribution of the rank-2 online reconstruction to the overall error is very minimal. In conclusion, the in-silico results indicate that higher-rank models with improved offline reconstruction quality could improve the overall quality of the proposed real-time reconstruction pipeline.

The spatial distribution of the EPEs in Figure 4.6 shows acceptable errors within the ROI, the lungs and the liver. Higher errors are visible at the organ interfaces, which could be attributed to the disability of the smooth spline motion model to represent inter-organ discontinuities.

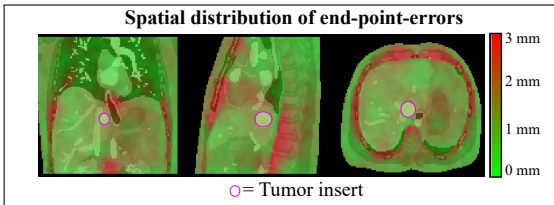


Fig. 4.6: Spatial distribution of EPEs of real-time MR-MOTUS reconstruction with $R = 2$ and offline reconstructed Φ (see Section 4.3, Section 4.4). The dynamic with the largest errors during normal breathing was selected for visualization.

In vivo real-time 3D MR-MOTUS reconstructions

Offline preparation phase

In the offline preparation phase, Eq. (4.7) was solved with respiratory-sorted data. Figure 4.7 shows a snapshot of the reconstructed motion-fields for all volunteers in a coronal plane, and for volunteer 4 in three mutually orthogonal planes. These results should be viewed in the corresponding videos of the respiratory-resolved reconstructions: Supporting Video 4.1[‡] and Supporting Video 4.2. Please see the Section 4.6 for an overview of all Videos. Little to no movement was reconstructed in organs not subject to respiratory motion such as the spine. Clear differences between the motion-fields for different volunteers can also be appreciated, most notably the large differences in the breathing motion amplitudes between e.g. volunteer 3 (large amplitude) and volunteer 5 (small amplitude).

Online inference phase

The real-time reconstructed motion-fields are visualized for volunteer 3 and volunteer 5, in respectively Supporting Video 4.3 and Supporting Video 4.4. Overall realistically looking motion-fields are reconstructed. It can be observed that the motion-fields are mostly smooth in time, except for small high frequency fluctuations which are mostly visible in end-exhale for both volunteers. These fluctuations may be caused by hardware imperfections [119] or by sensitivity to cardiac motion, which manifest themselves as high frequency fluctuations on top of respiratory motion.

Supporting Figure 4.1 shows the dependency of the total processing time on the number of spokes per dynamic, and the latency target of 200 ms per dynamic for real-time MRgRT. Taking into account the fluctuations in the processing times, 14 spokes were selected to stay well below the latency target, while maximizing the number of spokes per dynamic. With 67 ms of acquisition time for 14 spokes, and 103 ms of reconstruction time per dynamic, this resulted in a 3D motion-field every 170 ms.

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

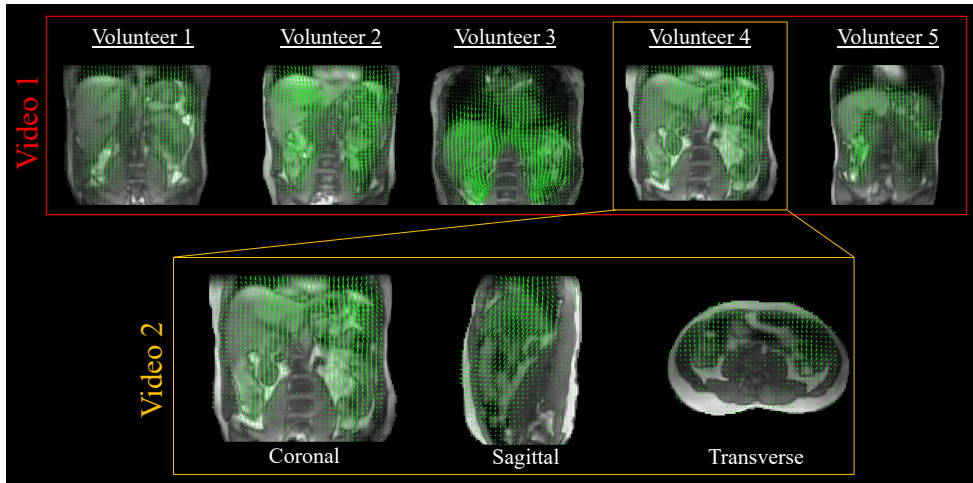


Fig. 4.7: Snapshot of the reconstructed respiratory-resolved motion-fields, as described in Section 4.3 and Section 4.4. These results should be viewed in the videos in the supporting files.

In vivo anatomical plausibility test with Jacobian determinants

4

Figure 4.8 shows the validation of the anatomical plausibility of the reconstructed motion-fields with its Jacobian determinant. For both volunteers it can be observed that most organs such as the liver preserve volume during deformation, which is in accordance with literature [40]. Bright red spots in the lungs indicate compression during inhalation, and bright blue spots indicate expansion during exhalation. Expansion and compression values are twice as high for volunteer 3 as for volunteer 5, indicating a relatively large breathing amplitude. Large values are also present at the interface between the top of the liver and the spine, where sliding motion occurs.

In vivo global accuracy test at high temporal resolution

The self-navigation spokes along feet-head, that are interleaved with the golden-mean 3D radial koshball acquisition, yield 1D respiratory motion information at high temporal resolution as projections of the whole excited FOV onto the feet-head axis. Figure 4.9A and Figure 4.9B show the validation of the reconstructed motion-fields with the projected profiles of the self-navigation spokes at 6.7 Hz and the surrogate signal extracted from these projections. Similar dynamic behavior of the projections, surrogate signal, and MR-MOTUS reconstructions can visually be observed for both volunteers, albeit that small high frequency oscillations remain present in the real-time MR-MOTUS reconstructions. Furthermore, Figure 4.9B shows that the irregular breathing pattern of volunteer 5 is also reconstructed with MR-MOTUS. The dynamic behavior of the real-time reconstructions is quantitatively analyzed in Figure 4.9C and Supporting Figure 4.2 by means of the Pearson correlation between the dynamic component Ψ of the real-time MR-MOTUS reconstructions and the 1D

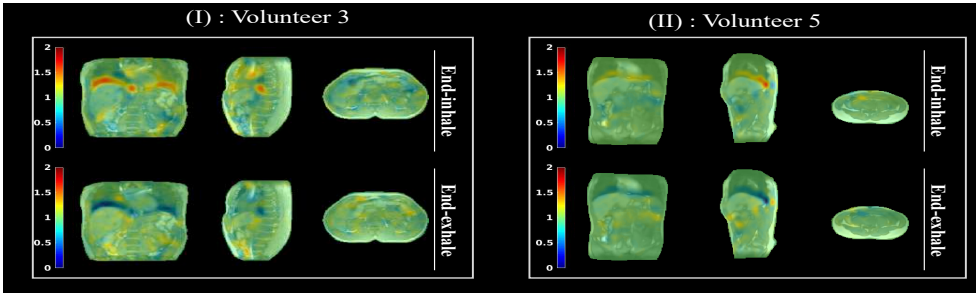


Fig. 4.8: Validation of the motion-fields by means of their Jacobian determinants, as mentioned in Section 4.3 and Section 4.4. The value of the Jacobian determinant can be interpreted as the volume fraction after deformation with respect to a reference image. In this case the reference image is reconstructed in mid-position. Compression and expansion can be observed in the lungs, while organs such as the liver mostly preserve volume.

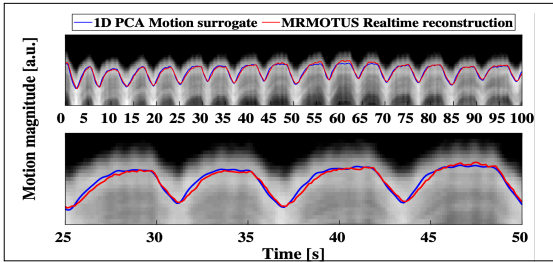
PCA motion surrogates. A linear correlation of 0.975 ± 0.0110 was found across all volunteers, further substantiating highly similar dynamic behavior between the 1D PCA motion surrogate and the real-time MR-MOTUS reconstructions. Additionally, Bland-Altman analyses are shown in Supporting Figure 4.3.

Comparison with respiratory-resolved compressed sensing

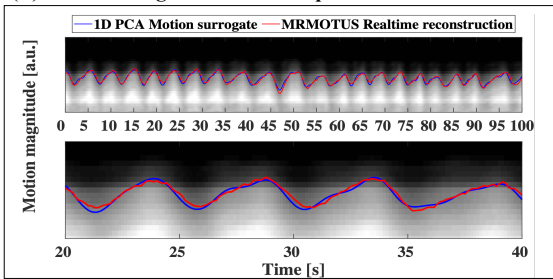
The comparison of the respiratory-resolved offline MR-MOTUS and compressed sensing (CS) reconstructions is shown in Supporting Video 4.5[‡] and Supporting Video 4.6 for respectively volunteer 1 and volunteer 4. A snapshot of Video 5 in end-inhale is shown in Figure 4.10, this respiratory phase showed most differences between the two reconstruction. The colored horizontal lines show minimal differences between locations of anatomical landmarks. From the videos it can be observed that the dynamics in the two reconstructions are very similar. From the last column it can be observed that only minimal differences remain present between the top of the liver. It can also be observed that some motion is reconstructed in the top of the spine, which is not visible in the CS reconstructions. The rest of the spine, however, remains static. Finally, the CS reconstructions contain pulsations in the aorta, which by construction cannot be visualized in the MR-MOTUS reconstruction. In general it should be noted that contrast variations between the two reconstructions will also contribute to the difference images in the last column.

[‡]The supporting videos can be found by following the link or QR-code on the title page. A description of the videos is available at the end of this chapter.

(A) Validation against feet-head spokes – Volunteer 3



(B) Validation against feet-head spokes – Volunteer 5



(C) Correlations with 1D PCA motion surrogate

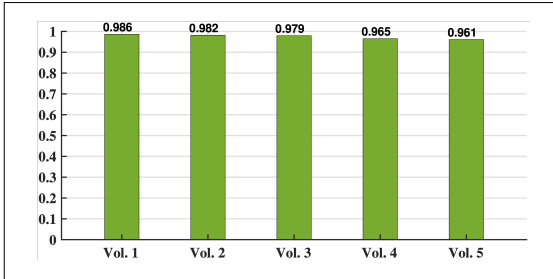


Fig. 4.9: Results of the qualitative validation (A-B) and quantitative validation (C) of the reconstructed motion-fields at 6.7 Hz, as described in Section 4.3 and Section 4.4. (A-B) show clear visual similarity between the projection profiles, motion surrogate and MR-MOTUS reconstructions, and (C) substantiates this with high Pearson correlations of 0.975 ± 0.0110 . Scatter plots corresponding to the Pearson correlations can be found in Supporting Figure 4.2, and additional Bland-Altman analyses in Supporting Figure 4.3.

4.5 Discussion

In light of MR-guided radiotherapy, we have presented a method to perform real-time deformable 3D respiratory motion estimation with 170 ms latency including both data acquisition and reconstruction. The proposed method relies on splitting the large-scale motion reconstruction problem formulated in Huttinga *et al.* [114] into two parts: (I) a medium-scale offline preparation phase and (II) a small-scale online inference phase which exploits the results of the offline phase for real-time computations. The method was validated on free-breathing data of 5 volunteers, acquired with an Elekta Unity MR-linac.

The results show that motion-fields reconstructed from data acquired with an MR-linac are anatomically plausible [40], have high correlation with a 1D motion surrogate [79, 115], and can be reconstructed with a latency of 170 ms that is sufficient for real-time adaptive MR-guided radiotherapy [22, 23]. Hence, the proposed

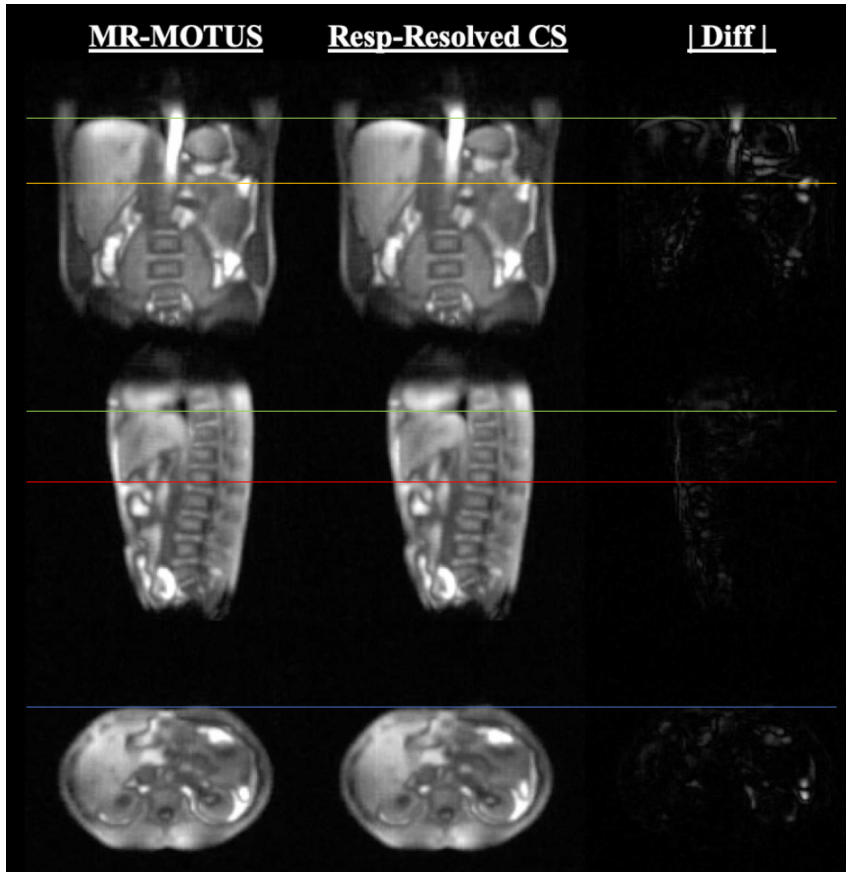


Fig. 4.10: Snapshot of Supporting Video 4.5, showing a comparison between respiratory-resolved MR-MOTUS and compressed sensing reconstructions for volunteer 1. The end-inhale phase was visualized, which showed the largest differences between the two reconstructions. The colored horizontal lines compare the vertical positions of anatomical landmarks: top of the liver dome (green), liver vessel (orange), bottom of the liver (red) and the anterior side of the lower abdomen (blue).

framework could be of significant value in a future clinical workflow for adaptive real-time MR-guided radiotherapy. Moreover, the motion-fields reconstructed in real-time during radiotherapy treatments could also be used for retrospective radiation dose accumulation calculations[120]. Furthermore, the current framework relies on a golden-mean 3D radial trajectory acquisition such that data required for the offline MR-MOTUS preparations, and the pre-treatment 3D+t respiratory-resolved MRI required for the radiotherapy workflow [109], could be acquired simultaneously.

The low-rank representation of respiratory motion [114] is an important component of this work, which has shown to represent realistic respiratory motion-fields with few parameters. The apparent strength of this representation is in agreement

with results reported in several other works in the context of motion estimation [25, 31, 33, 85–88, 121]. The strategy to split the motion-fields into spatial low-rank motion-field components and temporal motion-field components for fast inference was proposed before, e.g. for real-time CT-based motion estimation [122, 123], and for MR-based motion estimation [33, 121]. However, a notable difference is that in this work and [114], the low-rank structure of motion-fields is enforced a-priori; the low-rank components are obtained with a model-based reconstruction that ensures consistency with 3D k -space data of a training phase. In the other works [33, 121–123], the low-rank components are retrospectively obtained with PCA *after* a model-based reconstruction, and thus do not necessarily ensure consistency with data of the training phase. Similar approaches that reconstruct all low-rank components directly from the data were also proposed for dynamic MR-image reconstruction [29, 39], but were not extended to real-time reconstructions.

There are two large differences between the proposed method and other methods in the context of real-time motion estimation: the dimensionality of the input data (1D, 2D, 2.5D or 3D) and the processing domain (image domain or k -space domain). The former can be reduced to increase the temporal resolution of 3D reconstructions. For example, several works use multiple orthogonal 2D-cine planes, i.e. 2.5D+t, to reconstruct volumes (motion-fields or images), achieving a frame rate of about 2 Hz excluding reconstruction [25, 32, 33, 43, 121]. Further reducing the dimensionality, [44] achieved 5 Hz with 2D input data, and [34] generated 3D volumes at 3.3 Hz (acquisition + reconstruction) with 1D input data. Slightly different type of methods are based on surrogate signal models [35, 36, 51, 89, 112] that - similarly to this work - also use a bi-linear motion model, but directly incorporate 1D surrogate signals in this model to infer 3D motion-fields from 1D input data. These methods can thus achieve high temporal resolution, but rely heavily on the quality of the motion surrogate. In contrast with the other works, MR-MOTUS uses 3D input data and thereby has the ability to take higher dimensional motion information into account, but possibly at a lower spatial resolution. The other large difference is the processing domain; MR-MOTUS fits motion-fields directly in the k -space domain, whereas most other methods fit in the image domain. Fitting directly in k -space has the advantage of being more flexible in terms of temporal resolution and dimensionality of input data, but fitting on single-channel k -space data comes at the cost of a reduction in SNR and an increased sensitivity to hardware imperfections that cause temporal signal fluctuations such as eddy-currents. We have empirically observed that both can be controlled well with the proposed regularization techniques. However, better control of e.g. eddy-currents on an MR-linac can further improve the data quality [124], and may thereby improve the results in this work.

There are several other points that should be discussed. Firstly, the offline reconstruction times are currently in the order of minutes, of which most time is taken up by type-3 non-uniform FFTs (NUFFTs) [84]. Reconstruction times could thus be improved significantly by faster NUFFT computations.

The data acquisition in the offline phase was not optimized; the 10 minutes used for the in vivo results presented in this work was chosen on forehand to ensure sufficient data to reconstruct a reference image. This should be considered as a very conservative scenario to demonstrate the potential of the method. Results on the dig-

ital motion phantom and preliminary in vivo results indicate that around 3 minutes of free-breathing data would be sufficient to perform the same reconstructions. In practice, the data acquisition may be performed simultaneously with the pre-treatment MRI that is required for radiotherapy [109], or in the idle time during the treatment plan optimization, which takes 3-4 minutes in our current MR-linac workflow. In the latter scenario no additional time would be added to the treatment. Finally, radiolucent coils with more SNR that are currently being developed [125] could further improve our acquisition protocol. The optimization of the data acquisition protocol will be considered in a future work.

There is also still room for improvement in the data acquisition in the online phase. For example, 14 spokes per dynamic and only 8 samples per spoke were sufficient to perform the real-time reconstructions, but more samples were acquired per spoke. Although the radial readouts considered in this work yielded good performance, the proposed framework is in theory not limited to a specific type of readout but in practice a readout is required that contains sufficient motion information. Thus, different, more efficient trajectories could further reduce the latency of the online phase. However, designing such a trajectory is not trivial; a trajectory that only traverses the 8 samples required for the online reconstructions across multiple spokes would be affected by different system imperfections (zeroth and first order eddy currents in particular) compared to the golden mean spokes. This could induce a discrepancy between the offline and online data that could potentially affect the real-time motion estimation.

In this work a spatial motion-field basis was reconstructed in the offline phase, and was used in the online phase for fast inference. The spatial basis was reconstructed from respiratory-sorted data acquired over several minutes, and should therefore be able to represent a wide range of respiratory motion. The proposed framework, including both offline and online reconstruction phases, is in theory compatible with any pre-specified rank in the motion model that could be required to model a wide range of motion. The in silico experiments show that a higher-rank model built for normal breathing can be used for real-time reconstructions of different breathing patterns, and in vivo results in Figure 4.9B also show good reconstructions for the heavily varying breathing pattern of volunteer 5. The current work used a rank-1 motion model for in vivo reconstructions, since we empirically observed that larger motion models did not significantly improve in vivo reconstruction quality, but did increase real-time computation times. However, to accurately model in vivo respiratory motion on larger time scales, it may be required to either update the spatial basis during the real-time reconstructions or use higher-rank motion models as was done for the in silico experiments. The in silico results also indicate that improving the quality of the offline reconstructions of the less-dominant motion modes can most significantly improve the overall quality of the proposed real-time reconstruction pipeline. Future work will focus on improving reconstructions of higher-rank in vivo motion models.

The regularization added to the real-time reconstructions was shown to reduce high-frequency oscillations, but small oscillations still remain present. We expect these are caused by either high-frequency physiological motion such as cardiac motion, or by hardware-related system imperfections such as eddy-currents that are

known to affect the data quality of the Elekta Unity MR-linac [124]. Furthermore, the real-time reconstructions were shown to have only 170 ms latency, but in a practical application even this delay will have to be compensated. Possible directions could be Gaussian Processes [126] or other Bayesian filters [127], to simultaneously perform adaptive filtering and short-term predictions.

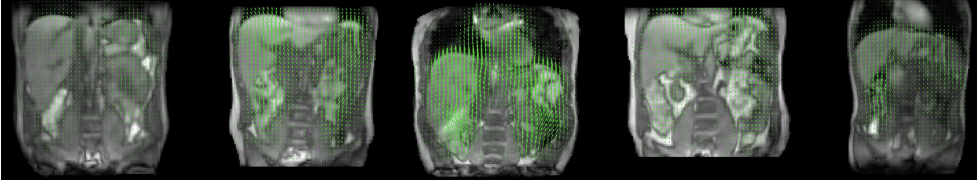
In this work we have chosen to perform an extensive *in silico* validation of the local reconstruction errors to get an impression of the potential weaknesses of the proposed framework. The *in vivo* validation of 3D+t motion-fields at high temporal resolution, however, remains challenging, and we have chosen to evaluate the correlation with a 1D feet-head motion surrogate that can be extracted at high temporal resolution with PCA. This surrogate has been shown to be highly correlated with physical translation in feet-head direction [115]. Results show high correlation between the global scaling of MR-MOTUS motion-fields and the PCA-based motion surrogate, indicating that the global dynamic behavior of the *in vivo* MR-MOTUS motion-fields is very similar to a widely used motion surrogate. However, the PCA-based surrogate is one-dimensional, and therefore does not allow for a local 3D validation of the accuracy of the motion-fields *in vivo*. To completely assess the *in vivo* performance of the proposed method, more validations are required with e.g. with a higher dimensional image navigator, fiducial marker tracking, or external respiratory motion sensors [128].

A different possible direction for future work is the extension of the proposed framework to other types of motion. The flexibility of low-rank MR-MOTUS was demonstrated in Huttinga *et al.* [114] by reconstructing both respiratory motion and head motion with the same low-rank motion decomposition model, but with a different number of components. The present work extends low-rank MR-MOTUS to real-time reconstructions without making direct assumptions on the type of motion, so the proposed real-time framework may be extended as was done in Huttinga *et al.* [114].

4

4.6 Conclusion

We have demonstrated real-time low-rank MR-MOTUS, a framework that can reconstruct 3D nonrigid motion-fields in real-time with a total latency of 170 ms. The proposed method was validated *in silico* and evaluated on a hybrid 1.5T MR-linac, and could reconstruct motion-fields that are anatomically plausible and are highly correlated with respiratory surrogate signals. We anticipate that low-rank MR-MOTUS could provide a novel practical solution for real-time MR-guided abdominal radiotherapy in the future.



Supporting Video 4.1: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>.* Offline reconstructed respiratory-resolved 3D+t motion-fields in a coronal plane for all volunteers. The moving images are obtained by warping the reference image with the motion-field. The visualized motion-fields (green arrows) show the displacement magnitude and direction projected on the selected planes.

4.7 Supporting Information

Supporting Figures

The supporting videos are part of the main body of this chapter and can be found through the QR code below the abstract, or through the following link: <https://doi.org/10.6084/m9.figshare.20480667.v2>.

Real-time implementation details

Preliminaries

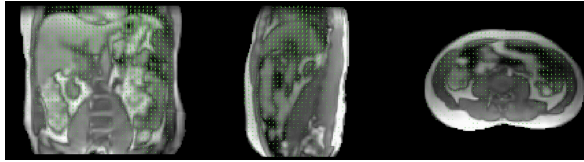
Assume a rank R motion model, i.e.

$$\mathbf{X}(\mathbf{r}, t) = \mathbf{X}(\mathbf{r}, 0) + \mathbf{D}(\mathbf{r}, t) = \mathbf{X}(\mathbf{r}, 0) + \Phi(\mathbf{r})\Psi^T(t),$$

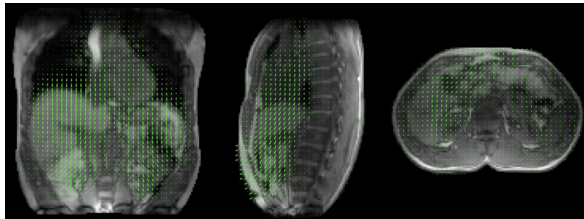
for $\mathbf{X}(\mathbf{r}, t), \mathbf{X}(\mathbf{r}, 0), \mathbf{D}(\mathbf{r}, t) \in \mathbb{R}^{d \times 1}$, $\Phi(\mathbf{r}) = [\Phi_1(\mathbf{r}), \dots, \Phi_R(\mathbf{r})] \in \mathbb{R}^{d \times R}$, and $\Psi(t) \in \mathbb{R}^{1 \times R}$. For the online inference phase, we will only deal with data and unknowns at a specific point in time, and for ease of notation we will therefore drop the continuous time-dependency and write $\Psi(t)$ as $\Psi_t := [\Psi_{t1}, \dots, \Psi_{tR}] \in \mathbb{R}^{1 \times R}$ in what follows. Substitution of the motion model in the signal model (4.3) yields

$$F(\Phi, \Psi_t)[\mathbf{k}] = \int_{\Omega} q(\mathbf{r}) e^{-i2\pi \mathbf{k}^T (\mathbf{r} + \Phi(\mathbf{r})\Psi_t^T)} d\mathbf{r}, \quad \mathbf{k} \in \mathcal{K}_t,$$

where \mathcal{K}_t denotes the set of k -space coordinates assigned to the acquisition at time t . We have split the arguments in F for ease of notation in the derivation of the derivatives in the next subsection.

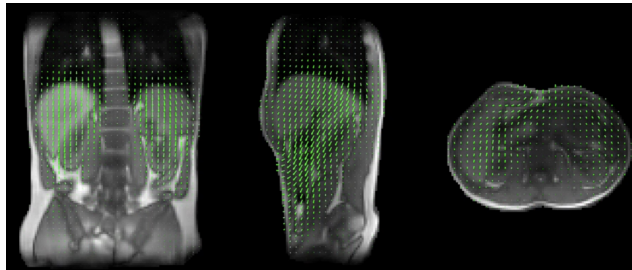


Supporting Video 4.2: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>.* Offline reconstructed respiratory-resolved 3D+t motion-fields in a coronal plane for in all planes for volunteer 5. The moving images are obtained by warping the reference image with the motion-field. The visualized motion-fields (green arrows) show the displacement magnitude and direction projected on the selected planes.

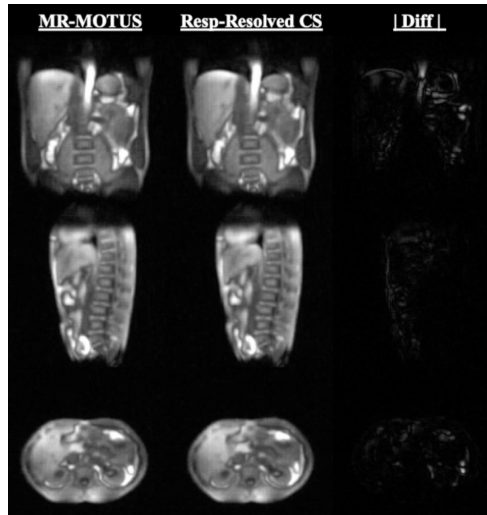


4

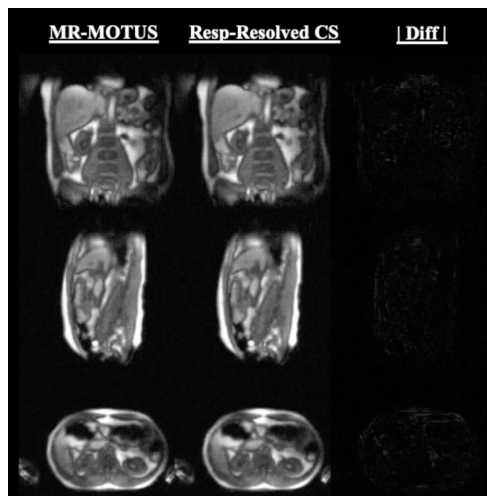
Supporting Video 4.3: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>.* Online time-resolved 3D+t motion-fields for volunteer 3, reconstructed at 16.6 Hz in 170 ms per dynamic, and visualized at 8.1 Hz. The moving images are obtained by warping the reference image with the motion-field. The visualized motion-fields (green arrows) show the displacement magnitude and direction projected on the selected planes.



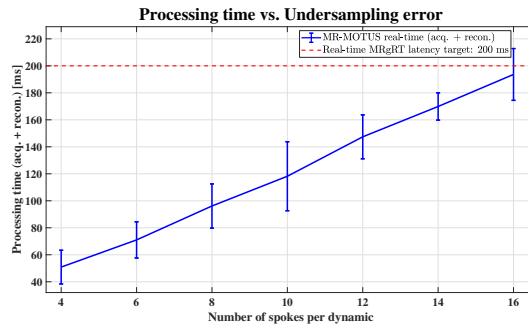
Supporting Video 4.4: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>.* Online time-resolved 3D+t motion-fields for volunteer 4, reconstructed at 16.6 Hz in 170 ms per dynamic, and visualized at 8.1 Hz. The moving images are obtained by warping the reference image with the motion-field. The visualized motion-fields (green arrows) show the displacement magnitude and direction projected on the selected planes.



Supporting Video 4.5: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>. Comparisons between offline-reconstructed respiratory-resolved MR-MOTUS and compressed sensing reconstructions, both resulting in respiratory-resolved volumetric images. The videos show results for volunteer 1. To obtain the MR-MOTUS images, the reference image was warped with the offline-reconstructed motion-fields.*

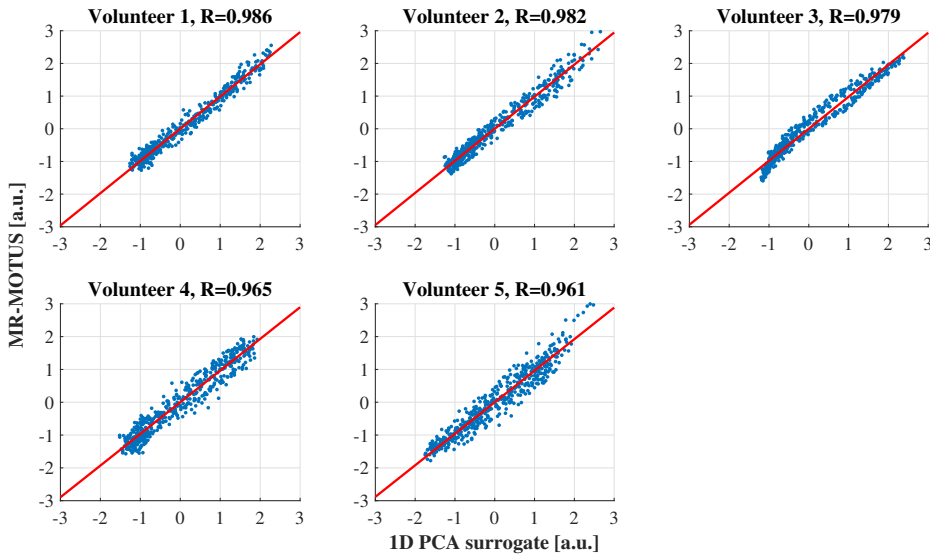


Supporting Video 4.6: *This is an animated figure that should be viewed online at <https://doi.org/10.6084/m9.figshare.20480667.v2>. Comparisons between offline-reconstructed respiratory-resolved MR-MOTUS and compressed sensing reconstructions, both resulting in respiratory-resolved volumetric images. The videos show results for volunteer 4. To obtain the MR-MOTUS images, the reference image was warped with the offline-reconstructed motion-fields.*

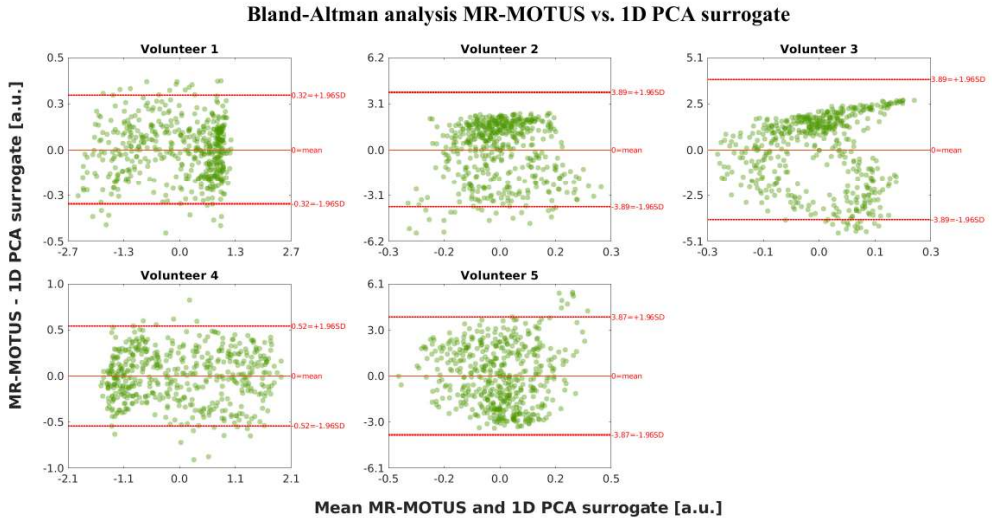


Supporting Figure 4.1: The dependency of the processing time (acquisition + reconstruction) on the number of spokes per dynamic. The solid blue line denotes the mean over 2000 real-time reconstructions for volunteer 4, and the vertical bars denote the standard deviations. We selected 14 spokes per dynamic, taking into account the fluctuations in the processing time, while maximizing the data per dynamic, and remaining below the real-time MR-guided radiotherapy target of 200 ms (dashed red line).

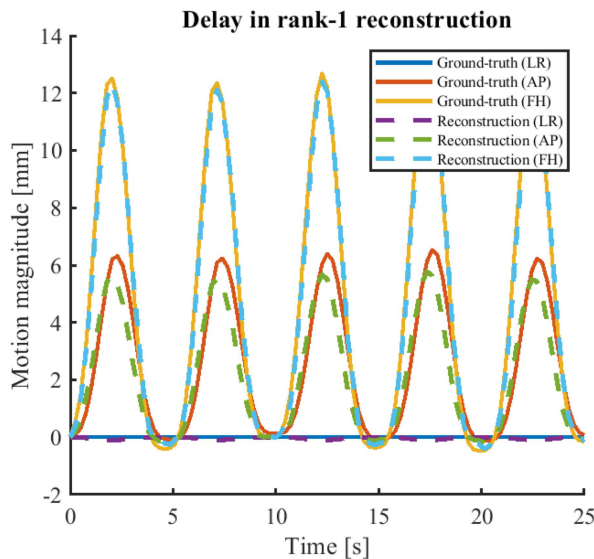
4



Supporting Figure 4.2: This figure visualizes for all volunteers scatter plots of the real-time reconstructed MR-MOTUS dynamic motion component Φ (y-axis), versus the 1D PCA motion surrogate (x-axis). Both quantities are normalized by subtracting the mean and dividing by the standard deviation. The red line shows the least-squares line through the samples. The slope is denoted by R , which directly corresponds to the Pearson correlation coefficient.



Supporting Figure 4.3: This figure visualizes Bland-Altman difference plots between the in vivo real-time reconstructions and 1D PCA surrogate for all volunteers. Both the MR-MOTUS and 1D PCA surrogate were centered and scaled by their respective standard deviations in order to make this comparison possible.



Supporting Figure 4.4: Visualization of the delay in AP-direction caused by a rank-1 model applied to an XCAT model with hysteresis; this model will mostly capture the most dominant motion mode in FH direction (Supporting Figure S5), and thereby makes an error in the less dominant motion mode. Due to the phase delay between AP and FH components during hysteresis, this error manifests as a phase delay.

Jacobian and Hessian of the forward model

In the online inference phase we are interested in real-time reconstruction of Ψ_t , given Φ and data, which requires extremely fast evaluations of both forward model and derivatives with respect to the unknowns Ψ_t . In this work we employ a Gauss-Newton algorithm, which requires a Jacobian matrix \mathbf{J} with first order derivatives and approximates the Hessian matrix as $\mathbf{J}^H \mathbf{J}$. Alternatively, a Newton method can be followed by using the true Hessian matrix, but we have experienced minor improvement, and increased reconstruction times with the Newton method as opposed to the Gauss-Newton method. For clarity we derive both the Jacobian and true Hessian matrix. We first derive the required derivatives:

$$\left(\frac{\partial^{a+b} F}{\partial [\Psi_{tm}]^a \partial [\Psi_{tn}]^b} \right) [\mathbf{k}] = \int_{\Omega} q(\mathbf{r}) e^{-i2\pi \mathbf{k}^T (\mathbf{r} + \Phi(\mathbf{r}) \Psi_t^T)} \cdot [-i2\pi \mathbf{k}^T \Phi_m(\mathbf{r})]^a [-i2\pi \mathbf{k}^T \Phi_n(\mathbf{r})]^b \, d\mathbf{r},$$

where $\mathbf{k}_l \in \mathbb{C}^{d \times 1}$ is the l -th d -dimensional k -space coordinate in the sequence \mathcal{K}_t of k -space coordinates in the dynamic at time t . Similarly as in the main part of this Chapter, we denote the total number of k -space coordinates at dynamic t as $N_k := |\mathcal{K}_t|$. Hence, the Jacobian $\mathbf{J} \in \mathbb{C}^{N_k \times R}$ and the Hessian matrix $\underline{\mathbf{H}} \in \mathbb{C}^{N_k \times R \times R}$ of $\mathbf{F}(\Phi, \cdot) : \mathbb{R}^R \mapsto \mathbb{C}^{N_k}$, evaluated at Ψ_t , are respectively computed as

$$[\mathbf{J}]_{l,m} := \int_{\Omega} q(\mathbf{r}) e^{-i2\pi (\mathbf{k}_l)^T (\mathbf{r} + \Phi(\mathbf{r}) \Psi_t^T)} \cdot [-i2\pi (\mathbf{k}_l)^T \Phi_m(\mathbf{r})],$$

$$[\underline{\mathbf{H}}]_{l,m,n} := \int_{\Omega} q(\mathbf{r}) e^{-i2\pi (\mathbf{k}_l)^T (\mathbf{r} + \Phi(\mathbf{r}) \Psi_t^T)} \cdot [-i2\pi (\mathbf{k}_l)^T \Phi_m(\mathbf{r})] [-i2\pi (\mathbf{k}_l)^T \Phi_n(\mathbf{r})] \, d\mathbf{r}.$$

Note that the Hessian of the vector-valued function $\mathbf{F}(\Phi, \cdot)$ is a 3-tensor. In what follows the notations of all tensors will be denoted by capital, underlined bold letters and in general we follow the tensor notations in Cichocki *et al.* [129].

Jacobian and Hessian of the objective function

We now assume subscripts denote tensor indices, and a colon selects all elements along a dimension, \odot denotes element-wise multiplication, and $\overline{\times}_n$ n -mode tensor-vector products (i.e. dot-products along the n -th tensor dimension). Furthermore, we define $\mathbf{K} \in \mathbb{R}^{N_k \times d \times N_t}$ as the tensor with N_k d -dimensional k -space coordinates at times $t \in [1, \dots, N_t]$, and $\hat{\Phi} \in \mathbb{R}^{N \times d \times R}$ as the discretization of $\Phi(\mathbf{r})$. Finally, we define the following matrices that are required for efficient computations in the real-time reconstructions:

$$\begin{aligned} [\mathbf{L}]_{:,:,m} &:= -i2\pi ([\mathbf{K}]_{:,:,t}) [\hat{\Phi}]_{:,:,m}^T, \\ [\mathbf{G}]_{:,j} &:= -i2\pi ([\mathbf{K}]_{:,:,t}) \mathbf{r}_j, \\ \mathbf{A}_t &:= \exp(\mathbf{G} + \underline{\mathbf{L}} \overline{\times}_3 \Psi_t), \end{aligned}$$

with $\underline{\mathbf{L}} \in \mathbb{C}^{N_k \times N \times R}$, $\mathbf{G} \in \mathbb{R}^{N_k \times N}$, $\mathbf{A}_t \in \mathbb{C}^{N_k \times N}$, and denote the vectorization of $q(\mathbf{r})$ as $\mathbf{q}_0 \in \mathbb{C}^{N \times 1}$, with N spatial points and d spatial dimensions. Then, the

evaluations of the forward model, Jacobian and Hessian at Ψ_t can respectively be written in tensor-vector form as

$$\mathbf{F}(\cdot, \Psi_t) = \mathbf{A}_t \mathbf{q}_0 \in \mathbb{C}^{N_k \times 1} \quad (4.1)$$

$$[\mathbf{J}]_{:,m} = (\mathbf{A}_t \odot [\mathbf{L}]_{:,m}) \mathbf{q}_0 \in \mathbb{C}^{N_k \times 1} \quad (4.2)$$

$$[\mathbf{H}]_{:,m,n} = (\mathbf{A}_t \odot [\mathbf{L}]_{:,m} \odot [\mathbf{L}]_{:,n}) \mathbf{q}_0 \in \mathbb{C}^{N_k \times 1 \times 1}. \quad (4.3)$$

The objective function for dynamic t can now be formulated as

$$E(\Psi_t) := \|\mathbf{e}_t\|_2^2, \quad \mathbf{e}_t := \mathbf{A}_t \mathbf{q}_0 - \mathbf{s}_t \in \mathbb{C}^{N_k \times 1}.$$

The gradient ∇_E , true Hessian \mathbf{H}_E , and Gauss-Newton Hessian approximation $\hat{\mathbf{H}}_E$ of E can be derived using (4.1)-(4.3) as

$$\nabla_E = 2\Re(\mathbf{J}^H \mathbf{r}_t) \in \mathbb{R}^{R \times 1} \quad (4.4)$$

$$\mathbf{H}_E = 2\Re(\mathbf{J}^H \mathbf{J} + \overline{\mathbf{H}} \overline{\mathbf{x}}_1 \mathbf{e}_t) \in \mathbb{R}^{R \times R} \quad (4.5)$$

$$\hat{\mathbf{H}}_E = 2\Re(\mathbf{J}^H \mathbf{J}) \in \mathbb{R}^{R \times R}, \quad (4.6)$$

where the superscript H denotes the conjugate transpose. As argued in the main part of this Chapter, R will typically be small, so these Jacobian and Hessian are very small, allowing for fast (sub-millisecond) processing.

Algorithm 1 Gauss-Newton algorithm for real-time MR-MOTUS

```

% Initialize solution variables
 $\Psi_{t=M_{\text{train}}} := \mathbf{0}$ 
 $\hat{\Phi} := \text{reshape}(\Phi, N, d, R)$ 
%  $\Psi_{t=M_{\text{train}}} \in \mathbb{R}^{R \times 1}$ 
%  $\hat{\Phi} \in \mathbb{R}^{N \times d \times R}$ 

for  $t > M_{\text{train}}$  do

    % Pre-compute  $t$ -dependent tensors:
     $[\underline{\mathbf{L}}]_{::,m} := -i2\pi ([\underline{\mathbf{K}}]_{::,t}) [\hat{\Phi}]_{::,m}^T$ 
     $[\underline{\mathbf{G}}]_{:,j} := -i2\pi ([\underline{\mathbf{K}}]_{::,t}) \mathbf{r}_j$ 
    %  $\underline{\mathbf{L}} \in \mathbb{C}^{N_k \times N \times R}$ 
    %  $\underline{\mathbf{G}} \in \mathbb{C}^{N_k \times N}$ 

    % Initialize Gauss-Newton iterations with solution at previous time-index:
     $\mathbf{x}_t^{(1)} = \Psi_{t-1}$ 
    %  $\mathbf{x}_t^{(1)} \in \mathbb{R}^{R \times 1}$ 

    for Gauss-Newton iterations  $k = 1 \dots N_{GN}$  do

        % Compute the forward model matrix and the residuals at the current iterate

         $\mathbf{A}_t := \exp(\underline{\mathbf{G}} + \underline{\mathbf{L}} \bar{\times}_3 \mathbf{x}_t^{(k)})$ 
         $\mathbf{e}_t := \mathbf{A}_t \mathbf{q}_0 - \mathbf{s}_t$ 
        %  $\mathbf{A}_t \in \mathbb{C}^{N_k \times N}$ 
        %  $\mathbf{e}_t \in \mathbb{C}^{N_k \times 1}$ 

        % Compute Jacobian of the forward model, and the gradient and approxi-
        % mated Hessian matrix of objective function w.r.t. solution variables
         $[\underline{\mathbf{J}}]_{:,m} := (\mathbf{A}_t \odot \underline{\mathbf{L}}_{::,m}) \mathbf{q}_0$ 
         $\nabla_E := 2\Re(\underline{\mathbf{J}}^H \mathbf{e}_t)$ 
         $\hat{\mathbf{H}}_E := 2\Re(\underline{\mathbf{J}}^H \underline{\mathbf{J}})$ 
        %  $\underline{\mathbf{J}} \in \mathbb{C}^{N_k \times R}$ 
        %  $\nabla_E \in \mathbb{R}^{R \times 1}$ 
        %  $\hat{\mathbf{H}}_E \in \mathbb{R}^{R \times R}$ 

        % Compute Gauss-Newton step-direction
         $\delta_{\mathbf{x}}^{(k)} = \text{argmin}_{\delta} \|\hat{\mathbf{H}}_E \delta + \nabla_E\|_2^2$ 
        %  $\delta_{\mathbf{x}}^{(k)} \in \mathbb{R}^{R \times 1}$ 

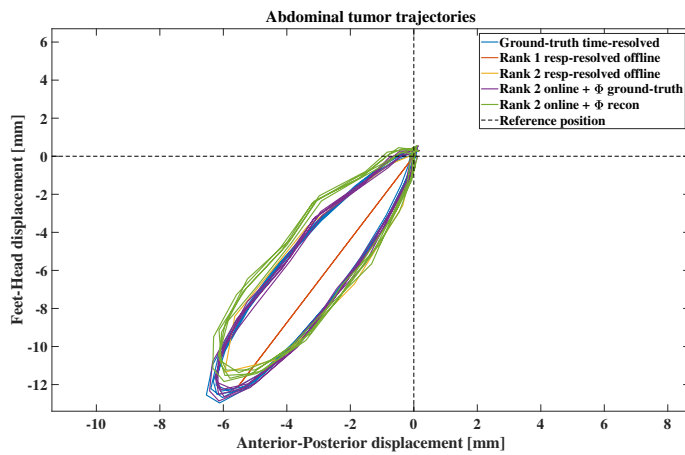
        % Update solution variable
         $\mathbf{x}_t^{(k+1)} = \mathbf{x}_t^{(k)} + \delta_{\mathbf{x}}^{(k)}$ 
        %  $\mathbf{x}_t^{(k+1)} \in \mathbb{R}^{R \times 1}$ 

    end for

    % Assemble 3D motion-field at time  $t$ 
     $\Psi_t = \mathbf{x}_t^{(N)}$ 
     $\mathbf{D}_t = \Phi \Psi_t^T$ 
    %  $\Psi_t \in \mathbb{R}^{R \times 1}$ 
    %  $\mathbf{D}_t \in \mathbb{R}^{N \times d \times R}$ 

end for

```



Supporting Figure 4.5: Analysis of the reconstructed tumor trajectories for different scenarios. This shows the expected elliptical trajectory. Moreover, the rank-1 seems to estimate the major axis of the ellipse, corresponding to the most dominant motion mode, whereas the rank-2 model is able to estimate the complete ellipse.

Chapter 5

Gaussian Processes for real-time 3D motion & uncertainty estimation

Niek R.F. Huttinga
Tom Bruijnen
Cornelis A. T. van den Berg
Alessandro Sbrizzi

The following chapter is based on:
Gaussian Processes for real-time 3D motion and uncertainty estimation. Medical Image Analysis (in submission).

Abstract

Respiratory motion during radiotherapy causes uncertainty in the tumor's location, which is typically addressed by an increased radiation area and a decreased dose. As a result, the treatments' efficacy is reduced. The recently proposed hybrid MR-linac scanner holds the promise to efficiently deal with such respiratory motion through real-time adaptive MR-guided radiotherapy (MRgRT). For MRgRT, motion-fields should be estimated from MR-data and the radiotherapy plan should be adapted in real-time according to the estimated motion-fields. All of this should be performed with a total latency of maximally 200 ms, including data acquisition and reconstruction. A measure of confidence in such estimated motion-fields is highly desirable, for instance to ensure the patient's safety in case of unexpected and undesirable motion. In this work, we propose a framework based on Gaussian Processes to infer 3D motion-fields and uncertainty maps in real-time from only three readouts of MR-data. We demonstrated an inference frame rate up to 69 Hz including data acquisition and reconstruction, thereby exploiting the limited amount of required MR-data. Additionally, we designed a rejection criterion based on the motion-field uncertainty maps to demonstrate the framework's potential for quality assurance. The framework was validated *in silico* and *in vivo* on healthy volunteer data ($n = 5$) acquired using an MR-linac, thereby taking into account different breathing patterns and controlled bulk motion. Results indicate end-point-errors with a 75th percentile below 1 mm *in silico*, and a correct detection of erroneous motion estimates with the rejection criterion. Altogether, the results show the potential of the framework for application in real-time MR-guided radiotherapy with an MR-linac.

Supporting Videos

The supporting videos corresponding to this chapter can be accessed through <https://doi.org/10.6084/m9.figshare.20480697.v1> or the following QR-code:



5.1 Introduction

Motion during abdominal radiotherapy decreases the efficacy of treatments due to an uncertain tumor location. This uncertainty can be reduced in several ways, one of which is to estimate the tumor's motion from MR-data acquired during radiation with an MR-linac [7–10], a hybrid device which combines an MR-scanner and a radiotherapy LINAC. The ultimate goal with the MR-linac is real-time adaptive MR-guided radiotherapy (MRgRT). This requires a continuous loop comprising tumor motion estimation, followed by corresponding radiation beam adjustments. MR-guided radiotherapy introduces several major technical challenges, one of which is to reconstruct accurate 3D motion-fields in real-time, from a stream of MR-data. The required speed for this reconstruction is determined by the expected velocities; for slowly-moving tumors such as prostate tumors 1 Hz could be sufficient, but tumors subject to respiratory motion requires at least 5 Hz [22, 23].

Here, we focus on the latter category, which requires MR-data acquisition and motion reconstruction with a frame rate of at least 5 Hz. Motion-fields can be estimated from MR-images by means of image registration (image-based), or directly from k -space data (k -space-based) [84]. Reconstructing 3D motion-fields of the abdomen and thorax at this rate is currently still challenging. To increase the achievable frame rate of 3D motion reconstructions, prior assumptions of images and/or motion-fields need to be included in the reconstruction. A frequently used strategy is to exploit an a priori built model with a two-step approach: 1) a calibration or training phase to build a patient-specific motion or image model prior to radiation; 2) a real-time phase during treatment, in which 3D information is reconstructed from a minimal amount of rapidly acquired MR-data by exploiting the model from the training phase. Examples of image-based methods proposed for radiotherapy include MR-SIGMA by Feng *et al.* [34], and approaches based on interleaved orthogonal cine images [25, 32, 33, 111]. An example of a k -space-based method for radiotherapy is the authors' real-time low-rank MR-MOTUS [84, 114, 130]. MR-SIGMA estimated 3D MR-images at 3.3 Hz, Stemkens *et al.* [33] achieved 3D motion-field reconstructions at about 2 Hz using image registration and cine-MRI, and MR-MOTUS achieved 3D motion-field reconstructions directly from k -space data at 6.7 Hz. Although considerably different in the modelling aspect, all the methods mentioned above employed a two-step approach.

It should be noted, however, that this two-step approach relies on the assumption that the motion in the training and real-time phases is similar. Although this is likely true in most cases, several practical scenarios such as bulk motion, or a change in breathing pattern could reduce the validity of this assumption. This could therefore result in erroneous motion estimates, which - if left undetected - could eventually lead to harmful radiation to the patient. To warrant the patient's safety in such scenarios, methods for real-time MRgRT should therefore ideally not only estimate motion in real-time, but should also provide some measure of reliability. In a practical setting, this could be used for real-time quality assurance during radiotherapy, e.g. to halt and resume the radiation treatment according to the degree of confidence.

In this work, we present a probabilistic framework to simultaneously quantify 3D motion and provide a measure of reliability in real-time. It thereby addresses

two critical needs for real-time MRgRT. The framework is based on the previously discussed two-step reconstruction approach. Firstly, in the training phase, a model for respiratory-resolved motion-fields is built that allows to represent 3D motion-fields with few coefficients. Secondly, in the inference phase, these representation coefficients are estimated from just three mutually orthogonal readouts of MR-data in real-time, thereby exploiting the motion model built in the training phase.

The idea to extract motion information directly from few readouts was motivated by the authors' promising results with (real-time) MR-MOTUS method [130], which exploits the MR-MOTUS signal model that explicitly relates motion-fields to k -space. Moreover, it employs a two-step reconstruction as described above. During both the training and inference phase, a non-linear inversion of the signal model is performed to estimate motion-fields from (few) k -space data acquired with a 3D golden-mean radial trajectory (GM3DR) [92]. As an alternative, this work proposes to reconstruct motion-field coefficients from three mutually orthogonal readouts - which will hereafter be referred to as self-navigation (SN) spokes - without assuming any signal model. This idea was motivated as follows. A single readout of k -space data that crosses the k -space center effectively contains a projection of the excited FOV in the readout direction. This can be seen by transforming the readout to image space. It therefore mostly contains information of motion in the direction of the readout. Consequently, a set of three mutually orthogonal readouts contains information of motion in all directions. Based on the two observations above, we hypothesize that the low-dimensional motion-field representation coefficients can directly be inferred from the three mutually orthogonal readouts. That is, we assume the motion-field representation coefficients are a function of the data on the three readouts.

Here, we propose to learn this underlying function via a probabilistic machine learning regression technique called Gaussian Processes (GP). A GP requires a calibration phase to tune its internal parameters based on a small training set, which takes ≈ 0.5 seconds in this work. The trained GP can then be used for real-time inference of the posterior distribution of the 3D motion-field representation coefficients, given the three self-navigation spokes. This step exploits the availability of a closed-form analytical expression for the posterior of a GP, which allows for sub-millisecond computations (≈ 0.1 milliseconds per dynamic in this work). Combining the resulting motion-field representation coefficients with the motion model, this eventually allowed for 3D motion-field reconstruction at frame rates as high as 69 Hz. Moreover, the posterior distribution as inferred by the GP captures not only the most likely motion-field estimate corresponding to the input data, but also the corresponding estimation uncertainty. The latter provides a measure of both the model-related and measurement-related uncertainties (respectively the epistemic and aleatoric uncertainties). As a consequence, a measure of motion model reliability is provided. We therefore hypothesize that the GP posterior uncertainty can be used for real-time quality assurance, i.e. to detect potentially erroneous motion estimates of the proposed framework. This possibility of real-time quality assurance would be a step forward with respect to other frequently used two-step approaches, such as real-time MR-MOTUS [130], since it is not trivial to give a degree of confidence in the estimates provided by such methods, even though they make the potentially problematic assumption of the transferability of a motion model from the training to

the inference phase.

This work is an extension of the preliminary work presented as conference abstracts in Sbrizzi *et al.* [72] and Huttinga *et al.* [131]. More specifically, in Sbrizzi *et al.* [72] we showed the feasibility of estimating motion-field representation coefficients with GPs from k -space data simulated from 2D DICOMs and a 3D digital phantom, and extracted a motion model using PCA and image registration. In Huttinga *et al.* [131] we extended this to multi-channel 3D k -space, prospectively acquired on an MR-linac, and built a rank-one subspace-constrained motion model using MR-MOTUS [114]. This work combines and extends both previous works by constructing a more expressive motion model using image registration on images reconstructed from prospectively acquired 3D MR-linac data.

Altogether, the framework could perform simultaneous real-time 3D motion-field estimation with real-time quality assurance. We assessed the accuracy of the presented framework based on several in silico and in vivo tests, and tested our hypothesis regarding the value of the GP posterior uncertainty for real-time quality assurance. For the latter, we designed a rejection criterion based on the uncertainty that flags dynamics with potential erroneous motion estimates in real-time. Both the motion estimation and rejection criterion were evaluated on data simulated using a digital XCAT phantom and in vivo data of five healthy volunteers acquired on an MR-linac, thereby considering different types of breathing and bulk motion.

5.2 Theory

A general introduction to Gaussian Processes

A Gaussian Process (GP) [126] models a Gaussian probability distribution over functions, and can therefore be considered as an extension of the multi-variate Gaussian distribution, which models a Gaussian probability distribution over vectors. Gaussian Processes are frequently applied to regression problems by assuming the following model for noisy measurements $\mathbf{y}_t \in \mathcal{Y} \subset \mathbb{R}^{N_y}$ at samples $\mathbf{x}_t \in \mathcal{X} \subset \mathbb{C}^{N_x}$:

$$\mathbf{y}_t = \mathbf{y}(\mathbf{x}_t) + \boldsymbol{\epsilon}, \quad (5.1)$$

where $\boldsymbol{\epsilon} \in \mathcal{X}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, and $\mathbf{y} : \mathcal{X} \rightarrow \mathcal{Y}$ is the underlying process. In case a function is drawn from a GP, and evaluated at a finite collection of $N_{\mathcal{T}}$ samples, $\mathbf{x}_{\mathcal{T}} := [\mathbf{x}_1, \dots, \mathbf{x}_{N_{\mathcal{T}}}] \in \mathbb{R}^{N_x \cdot N_{\mathcal{T}}}$, the vertically concatenated corresponding function values $\mathbf{y}_{\mathcal{T}} := [\mathbf{y}_1; \dots; \mathbf{y}_{N_{\mathcal{T}}}] \in \mathbb{R}^{N_y \cdot N_{\mathcal{T}}}$ follow a multi-variate Gaussian distribution. A GP is completely characterized by a mean *function* $m(\mathbf{x})$, and kernel covariance *function* $k(\mathbf{x}_i, \mathbf{x}_j)$, which in turn specify the mean vector and covariance matrix of the corresponding multi-variate Gaussian distribution:

$$\mathbf{y}_{\mathcal{T}} | \mathbf{x}_{\mathcal{T}} \sim \mathcal{N}(\mathbf{m}_{\mathcal{T}}, \mathbf{K}_{\mathcal{T}, \mathcal{T}}). \quad (5.2)$$

Here, \mathcal{T} is defined as the training set $\mathcal{T} := \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_{\mathcal{T}}}$, and $\mathbf{m}_{\mathcal{T}}$ and $\mathbf{K}_{\mathcal{T}, \mathcal{T}}$ denote the mean vector and covariance matrix, which are computed by evaluating respectively the mean and kernel function for all training samples in \mathcal{T} .

The kernel function characterizes the properties of the underlying process (in this case y), such as smoothness or periodicity. Without loss of generality, a zero-mean GP is typically assumed with $m \equiv 0$, in which case the GP is completely characterized by the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$. The kernel function is typically a function of the distance between its two inputs, and is parameterized by hyperparameters $\boldsymbol{\theta}$ that determine its form: $k(\mathbf{x}_i, \mathbf{x}_j) := f(\|\mathbf{x}_i - \mathbf{x}_j\|_2 | \boldsymbol{\theta})$.

Evidently, correct tuning of $\boldsymbol{\theta}$ directly influences the GP's regression performance. Therefore, two steps are followed to perform GP-based regression. In the first step, the GP hyperparameters $\boldsymbol{\theta}$ are estimated by maximum likelihood estimation (MLE) on the joint Gaussian likelihood (5.2). The result is a fully-determined GP kernel function, specifying the properties of the functions that best fit the training measurements. In the second step, a posterior distribution over the function values $y_{\mathcal{Q}}$ at samples \mathcal{Q} is computed by conditioning on the training data \mathcal{T} . This yields the Gaussian posterior distribution [126, Eq. [4.6]]:

$$\mathbf{y}_{\mathcal{Q}} | \mathcal{T} \sim \mathcal{N}(\mathbf{m}_{\mathcal{Q}}, \boldsymbol{\Sigma}_{\mathcal{Q}}), \quad (5.3)$$

with

$$\mathbf{m}_{\mathcal{Q}} = \mathbf{K}_{\mathcal{T}, \mathcal{Q}}^T (\mathbf{K}_{\mathcal{T}, \mathcal{T}} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}_{\mathcal{T}}, \quad (5.4)$$

$$\boldsymbol{\Sigma}_{\mathcal{Q}} = \mathbf{K}_{\mathcal{Q}, \mathcal{Q}} - \mathbf{K}_{\mathcal{T}, \mathcal{Q}}^T (\mathbf{K}_{\mathcal{T}, \mathcal{T}} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{K}_{\mathcal{T}, \mathcal{Q}}. \quad (5.5)$$

In particular, the diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_{\mathcal{Q}}$ define the GP prediction uncertainty.

One of the main strengths of the GP framework is the availability of the closed-form analytical expressions in Eq. (5.4)-(5.5) that completely characterize the posterior distribution in Eq. (5.3). The computations involve the matrices $\mathbf{K}_{\mathcal{T}, \mathcal{Q}} \in \mathbb{R}^{N_{\mathcal{Y}} N_{\mathcal{T}} \times N_{\mathcal{Y}}}$, $\mathbf{K}_{\mathcal{T}, \mathcal{T}} \in \mathbb{R}^{N_{\mathcal{Y}} N_{\mathcal{T}} \times N_{\mathcal{Y}} N_{\mathcal{T}}}$ and $\mathbf{K}_{\mathcal{Q}, \mathcal{Q}} \in \mathbb{R}^{N_{\mathcal{Y}} \times N_{\mathcal{Y}}}$, where $N_{\mathcal{T}}$ denotes the number of training samples and $N_{\mathcal{Y}}$ the number of outputs.

5

5.3 Methods

Proposed framework: integrating Gaussian Processes in motion modeling

The technical challenge for this application is to perform a real-time reconstruction of a complete 3D motion-field and corresponding reconstruction uncertainties from few readouts that can be rapidly acquired. This work considers Gaussian Processes for this purpose for two main reasons. Firstly, for the targeted application the computations in Eq. (5.4)-(5.5) can be performed in the order of milliseconds. Secondly, the probabilistic nature of GPs gives access to reconstruction uncertainties, which will prove valuable for quality assurance.

We identify 3 practical challenges when trying to apply GPs in our context. Firstly, GPs suffer from the curse of dimensionality, and are therefore challenging to apply to high-dimensional inputs (e.g. > 300 total samples on the three mutually

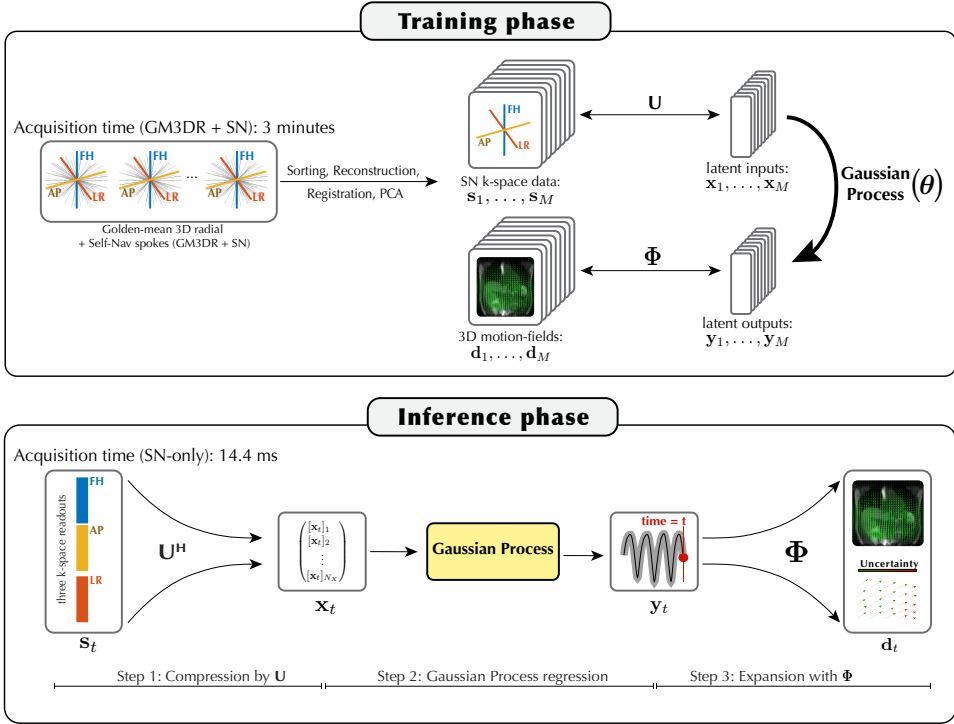


Fig. 5.1: Framework overview. In the training phase, GM3DR+SN data is acquired during 3 minutes of free-breathing. The motion model and spoke compression basis are built with PCA on respectively respiratory-sorted GM3DR+SN and SN spokes. The motion model requires motion-fields, which are obtained by E_Volution [132] on respiratory-sorted images. Finally, GP hyperparameters θ are obtained by maximum likelihood estimation on the training set $\mathcal{T} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^M$, consisting of the compressed spoke representations \mathbf{x}_i and motion-fields representation coefficients \mathbf{y}_i . In the inference phase three steps are followed: 1) SN spokes are compressed to $\mathbf{x}_t = \mathbf{U}^H \mathbf{s}_t$; 2) the posterior distribution $P(\mathbf{y}_t | \mathcal{T})$ is computed with the GP from \mathbf{x}_t ; 3) a 3D motion-field \mathbf{d}_t and a corresponding uncertainty map are inferred by expanding the GP posterior distribution with the motion model Φ .

orthogonal readouts); distances in high-dimensional input-space become uninformative [133], while these distances - as evaluated by the kernel function - form the basis of the GP theory.

Secondly, GPs are mostly applied to regression problems with scalar functions, since modeling correlations in outputs is a challenging task and requires a non-trivial extension of the framework [134]. To overcome these first two challenges, we propose a linear compression of both the input space and the output space, and learn a GP that maps from the latent input space $\mathcal{X} \subset \mathbb{R}^{N_x}$ to the latent output space $\mathcal{Y} \subset \mathbb{R}^{N_y}$. The schematic pipeline from few readouts $\mathbf{s}_t \in \mathcal{S} \subset \mathbb{C}^{N_s}$, to motion-fields

$\mathbf{d}_t \in \mathcal{D} \subset \mathbb{R}^{3N}$, then becomes

$$\mathcal{S} \xrightarrow{\mathbf{U}^H} \mathcal{X} \xrightarrow{\text{GP}} \mathcal{Y} \xrightarrow{\Phi} \mathcal{D}, \quad (5.6)$$

where \mathbf{U} and Φ denote orthogonal bases of the latent input and output space, and \mathbf{U}^H the Hermitian transpose of \mathbf{U} . This pipeline is also visually outlined in Figure 5.1. Assuming $N_y \geq 1$ uncorrelated dimensions, each element $[\mathbf{y}_t]_j$ ($j = 1, \dots, N_y$) of \mathbf{y}_t can be modelled with a separate GP, each of which takes \mathbf{x}_t as input.

Now that the general pipeline is outlined, the third practical challenge can be identified. Training of the GPs in the pipeline (5.6) requires training sets $\mathcal{T}_j = \{(\mathbf{x}_t, [\mathbf{y}_t]_j)\}_{t=1, \dots, N_T}$, for $j = 1, \dots, N_y$. On the one hand, the GPs are required to constantly evaluate distances between \mathbf{x}_t 's. To do this consistently, all \mathbf{x}_t should thus measure the same k -space locations but at a different time instance, which in turn requires the same k -space locations for all t . On the other hand, the targets \mathbf{y}_t will be derived from 3D images, and thus require k -space samples from the entire range of k -space coordinates. To meet both requirements, we employed an interleaved acquisition of mutually orthogonal self-navigation k -space spokes (SN spokes) and golden-mean 3D radial (GM3DR) spokes [92]. The SN spokes are acquired along the three mutually orthogonal axes: feet-head (FH), anterior-posterior (AP), and left-right (LR). Before further processing the k -space spokes are transformed to image space through an inverse FFT along the readout direction. Next, the vertically concatenated SN spokes at dynamic t , denoted as $\mathbf{s}_t \in \mathcal{S}$ are processed to GP inputs $\mathbf{x}_t \in \mathcal{X}$, and GM3DR spokes at dynamic t to GP outputs $\mathbf{y}_t \in \mathcal{Y}$. For the SN spokes, the processing includes the computation of a compressed representation with respect to an orthonormal basis \mathbf{U} . For GM3DR spokes, the processing comprises a respiratory-resolved image reconstruction, followed by an image registration, the construction of an orthogonal motion model Φ via PCA, and finally the computation of a representation with respect to Φ . More details on these processing steps will be given in the subsequent sections. The construction of the GP training sets \mathcal{T}_j from the acquired SN and GM3DR spokes is also visualized in Figure 5.1.

The inference pipeline Eq. (5.6) requires three main steps: 1) a compression of readout spokes to the latent input space using \mathbf{U} , 2) regression between the latent input and output space by a GP with hyperparameters θ , 3) an expansion of the latent output space to 3D motion-fields using a motion model Φ . These steps are also visualized at the bottom of Figure 5.1. All three sets of parameters (\mathbf{U} , θ , Φ) mentioned above are obtained from respiratory-resolved training data, which was in turn acquired by sorting all GM3DR spokes with a 1D respiratory motion surrogate extracted from the closest feet-head spoke [79, 114]. Details including how the required parameters \mathbf{U} , θ and Φ in the pipeline Eq. (5.6) are obtained will be discussed next, organized in sections. The first step in the pipeline, i.e. a compression of the acquired space, will be discussed in Section 5.3. The second step, i.e. the GP regression, will be discussed in Section 5.3. The third step, i.e. the expansion from GP outputs to motion-fields, will be discussed in Section 5.3. Finally, we also investigate the potential of the framework outlined above for quality assurance. To this extent, we designed and evaluated a rejection criterion based on the GP's reconstruction uncertainty. The rejection criterion is evaluated in parallel with the inference pipeline,

rejecting unreliable motion estimates due to undesirable motion. Details on this are discussed in Section 5.3.

Step 1: Input space compression with \mathbf{U}

We empirically observed a sensitivity of the reconstruction to cardiac motion, which resulted in high-frequency oscillations on top of the respiratory motion. Since this is undesirable, the k -space readouts were first transformed to image space with an inverse FFT along the readout direction, and subsequently samples on the readouts (radial spokes) were thresholded based on their temporal frequency spectrum; any sample with a large contribution from the cardiac frequency range was removed. More specifically, individual samples on any spoke were removed for which the energy content above 0.5-0.8 Hz (depending on which data) exceeded 25% of the total energy. The thresholding was performed at both training and test time; the indices were determined from the training data, and were used to mask the samples in both training and test phases. This thresholding typically resulted in an automatic selection of the liver, and excluded the heart and large arteries. After thresholding, respiratory-sorting was performed on the spokes, and PCA was performed along the respiratory dimension on the respiratory-sorted matrix of spokes $\mathbf{S} := [\mathbf{s}_1, \dots, \mathbf{s}_{20}]$, such that $\mathbf{S} \approx \mathbf{U}\mathbf{V}^H$. For the sorting, a hybrid strategy was employed; inhale and exhale were first separated with a peak detection, and each was separately sorted based on the amplitudes of the surrogate signal. This will be similar to a clustering approach on the surrogate signal and its first temporal derivative. It should be noted, however, that even for simple amplitude binning, a rank higher than one may be required to accurately describe the motion in the resulting respiratory-resolved image sequence. Here, and in the rest of this work, the number of respiratory bins was set to 20. The \mathbf{s}_i ($i = 1, \dots, 20$) were obtained by averaging over all SN spokes that ended up in the i -th bin. The number of columns in \mathbf{U} and \mathbf{V} for the approximation $\mathbf{S} \approx \mathbf{U}\mathbf{V}^H$ was set to 4. This restricts the compression of the inputs to no less than 4 samples. This was empirically observed to improve the flexibility of the framework to correctly perform the regression. Finally, data compression at time t was performed by $\mathbf{x}_t = \mathbf{U}^H \mathbf{s}_t$.

Step 2: GP regression between latent input and output space

For the GPs we chose to use a Matérn kernel with $\nu = 3/2$ and automatic relevance determination (ARD). The Matérn class of kernel functions is an extension of the class of radial basis functions, and restricts the functions that can be drawn from the GP to be $\lfloor \nu \rfloor$ times differentiable. With $\nu = 3/2$, this thus results in a GP that models once-differentiable functions. In general, kernel functions return the correlation between its two inputs. In case of an ARD kernel, the relevance of each input dimension in this correlation computation is specified by the length-scales l_1, \dots, l_{N_x} . For the Matérn-3/2 kernel, we have

$$k(\mathbf{x}_i, \mathbf{x}_j | \sigma_n, l_1, \dots, l_{N_x}) = \sigma_n^2 \left(1 + \sqrt{3} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{L}} \right) \exp \left(-\sqrt{3} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{L}} \right), \quad (5.7)$$

Here, $\boldsymbol{\theta} := \{\sigma_n, l_1, \dots, l_{N_{\mathcal{X}}}\}$ denote the kernel hyperparameters, and we defined $\mathbf{L} := \text{diag}(1/l_1^2, \dots, 1/l_{N_{\mathcal{X}}}^2)$ and the norm of a vector \mathbf{a} with respect to a positive definite matrix \mathbf{B} as $\|\mathbf{a}\|_{\mathbf{B}} := \sqrt{\mathbf{a}^T \mathbf{B} \mathbf{a}}$. Theoretically, the standard deviation σ_n of the target noise $\boldsymbol{\epsilon}$ in Eq. (5.1) could also be optimized, but here we manually fixed σ_n such that the 95% confidence interval approximately contained the neighboring bin's training targets (see first column in Figure 5.3). With inputs $\mathbf{x}_t \in \mathbb{R}^{N_{\mathcal{X}}}$, this resulted in a total of $N_{\mathcal{X}}$ optimizable GP hyperparameters. Finally, $N_{\mathcal{Y}}$ pair-wise training sets were constructed as $\mathcal{T}_j := \{(\mathbf{U}^H \mathbf{s}_t, [\mathbf{y}_t]_j)\}_{t=1, \dots, 20}$, and $N_{\mathcal{Y}}$ sets of hyperparameters - one for each GP - were obtained through MLE on Eq. (5.2). The total time of the training was about 0.5 seconds. For the GP hyperparameter optimization we made use of the Gaussian Processes for Machine Learning (GPML) Matlab toolbox [135]. Once the training is performed, the resulting GP can infer a posterior distribution $P(\mathbf{y}_t | \mathbf{x}_t)$ of the motion-field representation coefficients \mathbf{y}_t , given the compressed representation of the spokes $\mathbf{x}_t = \mathbf{U}^H \mathbf{s}_t$. This inference process took about 0.1 milliseconds per dynamic.

Step 3: Expansion from latent output space to motion-fields

To expand the posterior distribution on \mathbf{y}_t to motion-fields, a motion model is required. To derive the motion model $\boldsymbol{\Phi}$, first respiratory-resolved MR-images $\mathbf{I} := [\mathbf{I}_1, \dots, \mathbf{I}_{20}] \in \mathbb{C}^{N \times 20}$, were reconstructed with an L1-ESPIRiT reconstruction without temporal regularization with the BART toolbox [70], where N denotes the number of voxels. Subsequently, respiratory-resolved motion-fields were obtained by image registration with EVOlution [132] between the end-exhale dynamic and all other dynamics. The result was stored in a matrix $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_{20}] \in \mathbb{R}^{3N \times 20}$. Similarly as in the two-phase motion-field reconstruction approaches, PCA was performed along the respiratory dimension of \mathbf{D} to obtain a linear motion model $\boldsymbol{\Phi}$. Typically, most information of \mathbf{D} is concentrated in its first principal components [33, 85, 86, 114], allowing to truncate the number of columns of $\boldsymbol{\Phi}$ from 20 to $N_{\mathcal{Y}} \ll 20$, without sacrificing too much representational power. Thus, $\boldsymbol{\Phi} \in \mathbb{R}^{3N \times N_{\mathcal{Y}}}$, and \mathbf{D} is approximately low-rank, with rank $N_{\mathcal{Y}}$. In all experiments in this work, we normalized all columns in $\boldsymbol{\Phi}$ to unit norm, and determined the rank via the L-curve of the explained variance of \mathbf{D} as a function of the number of components. This resulted in $1 \leq N_{\mathcal{Y}} \leq 3$ for all performed tests, and a motion model which approximates high-dimensional motion fields $\mathbf{d}_t \in \mathbb{R}^{3N \times 1}$ with low-dimensional representation coefficients $\mathbf{y}_t \in \mathbb{R}^{N_{\mathcal{Y}}}$: $\mathbf{d}_t \approx \boldsymbol{\Phi} \mathbf{y}_t$.

A posterior distribution of the motion-field at time t can be derived by applying the motion model to the posterior distribution of the representation coefficients - $P(\mathbf{y}_t | \mathbf{s}_t) := \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ - as output by the GP. Using basic properties of the multivariate Gaussian distribution, we can derive

$$\mathbf{d}_t := \boldsymbol{\Phi} \mathbf{y}_t \sim \mathcal{N}(\boldsymbol{\Phi} \boldsymbol{\mu}_t, \boldsymbol{\Phi} \boldsymbol{\Sigma}_t \boldsymbol{\Phi}^T). \quad (5.8)$$

The spatio-temporal uncertainties \mathbf{P}_t in the motion-fields can be obtained as the diagonal of the covariance matrix in Eq. (5.8), i.e. $\mathbf{P}_t := \text{diag}(\boldsymbol{\Phi} \boldsymbol{\Sigma}_t \boldsymbol{\Phi}^T)$. Note that this uncertainty map should be treated with caution. For the derivation above it is

assumed that Φ is valid, and that there is no uncertainty associated with it. Hence, the spatio-temporal uncertainties in the motion-fields \mathbf{d}_t can only be reasonably interpreted when this assumption is true. Unlike the uncertainty on \mathbf{d}_t , the uncertainty in y_t has a useful meaning also when Φ is no longer valid. This forms the basis of the rejection criterion, as will be discussed next.

Online quality assurance via posterior uncertainty

We propose a rejection criterion based on the posterior uncertainty, which flags dynamics with uncertain predictions as follows. A new measurement can be considered unreliable when the corresponding prediction uncertainty is substantially higher than those evaluated on the training set. To this extent, we defined the rejection threshold τ such that

$$P(\sigma_t < \tau | t \in \mathcal{T}) = \alpha, \quad (5.9)$$

where σ_t is a measure of uncertainty. Given this definition of τ , any prediction with an uncertainty exceeding τ will be rejected. Effectively, this ensures a false positive rejection rate - i.e. the chance of incorrect rejections - of approximately $1-\alpha$. Note that τ in Eq. (5.9) is equivalent to the α -th percentile over all σ_t evaluated on the training set data. In all experiments in this work we set $\alpha = 0.95$. Each of the N_y GPs outputs a measure of uncertainty (i.e. the variance), which are combined to a single measure of uncertainty in order to evaluate the rejection criterion. To this end, we took a weighted linear combination, where the weights were determined by the singular values in the motion model. This leads to $\sigma_t = \sum_{j=1}^{N_y} s_j^2 [\Sigma_t]_{j,j}$, where s_j is the singular value corresponding to the j -th singular vector of the motion model Φ .

5.4 Experiments

In vivo data acquisition

For the in vivo experiments described later, data were acquired on a 1.5T Elekta Unity MR-linac (Elekta AB, Stockholm, Sweden) from healthy volunteers during free-breathing. All experiments were approved by the institutional review board, carried out in accordance with the relevant guidelines and regulations, and written informed consent was obtained from all volunteers prior to the experiments. In all experiments, we used the standard clinical 8-element radiolucent array with anterior and posterior coils and a steady-state spoiled gradient echo sequence (SPGR) with TR = 4.8 ms, TE = 1.8 ms, FA = 20°, FOV = 30 cm x 30 cm x 30 cm, BW = 540 Hz, resolution = 3 mm x 3 mm x 3 mm. We considered two modes for the data acquisition. In the first mode, data were acquired with a golden-mean 3D radial (GM3DR) kooshball trajectory [92], interleaved every 31 spokes with three self-navigation spokes (SN-spokes); along feet-head (FH), anterior-posterior (AP), and left-right (LR). This first mode is referred to as the GM3D+SN mode. Since the proposed framework can infer a 3D motion-field from only a single set of SN-spokes, this interleaved acquisition allows for inference at $1000/(31 \times 4.8) \approx 6.7$ Hz. The

second mode served to test the feasibility of high-speed inference. In this mode only the SN-spokes were acquired, which allows for inference at $1000/(3 \times 4.8) \approx 69$ Hz. The second mode is referred to as the SN-only mode.

In silico XCAT experiment

Data simulation

Data were generated using the digital XCAT phantom [106] with manually added MR-contrast. XCAT requires two inputs, an abdominal and chest waveform, and returns dynamic volumetric images with respiratory motion (cardiac motion was disabled) and corresponding ground-truth motion-fields. To simulate realistic motion, \cos^4 waveforms were used, hysteresis was simulated through a phase-delay between the two input waveforms, and the variations in the extreme respiratory positions were randomly generated as 1% and 2% of the waveform amplitudes for respectively end-exhale and end-inhale. All simulated motion-fields were post-processed with cid-X [107] to ensure realistic and invertible motion-fields. A complex reference volume in end-exhale was generated by adding a smoothly varying phase, and the resulting motion-fields were applied to this reference volume by cubic interpolation, resulting in a series of dynamics. Multi-channel images were simulated by multiplying each dynamic with 8 (static) coil sensitivity maps simulated with BART [70]. Finally, 800 readouts of k -space data were simulated for each dynamic with a NUFFT [64] evaluated on the GM3DR+SN trajectory, which is equivalent to the trajectory used for in vivo acquisitions. Complex noise was added to the resulting k -space data to achieve an SNR of 60. In total four types of breathing modes were simulated for our experiments: normal, chest-only, abdominal-only, and amplitude drifts. Chest-only and abdominal-only breathing were simulated by setting respectively the abdominal and chest waveforms to zero, amplitude drifts were simulated by adding a linearly increasing shift to the normal breathing waveforms. For each breathing mode, approximately five breathing cycles were generated, resulting in a total of 100 dynamics. The total data per breathing mode corresponds to an acquisition of approximately 6 minutes. As described later in Section 5.4, half of the normal breathing phase data were used for training, and the other half for testing. Hence, the in silico training was performed on data that could in practice be acquired in 3 minutes. This is similar to the in vivo experiments, as will be discussed later in Section 5.4 and Section 5.4.

Performance evaluation: end-point-error analysis and model applicability tests

To validate our framework, we performed an in silico end-point-error (EPE) analysis. These EPEs were computed per voxel, as the magnitude of the difference between the reconstructed $\hat{\mathbf{d}}_t$ and ground-truth motion-fields \mathbf{d}_t :

$$\text{EPE}([\hat{\mathbf{d}}_t]_i, [\mathbf{d}_t]_i) := \left\| [\hat{\mathbf{d}}_t]_i - [\mathbf{d}_t]_i \right\|_2, \quad (5.10)$$

where $[\cdot]_i$ denotes the i -th voxel. We considered the scenario where we train only on data simulated during normal breathing, and performed inference on all types of

breathing: normal, chest-only and abdominal-only breathing, and amplitude drifts. The normal breathing data were separated in two consecutive sets of equal size, a train and a test set. We analyzed the results in two ways. Firstly, we compared boxplots of the EPEs of a model trained for normal breathing, and applied to all breathing patterns. We hereby separated the EPEs statistics into dynamics rejected by our framework and the non-rejected dynamics, which allowed to see the effect of the proposed rejection criterion on the EPEs. Secondly, we analyzed the correlation between the GP posterior uncertainty and EPEs.

Synthetic time-resolved in vivo data

Data simulation

To assess the performance of our framework in a more realistic setting which still allows for a comparison with ground-truth motion-fields, experiments were performed on synthetic time-resolved in vivo data. To generate this data, realistic 3D motion-fields were extracted from 2D cine-MRI, which were subsequently used to warp a 3D image and finally generate k -space data.

More specifically, a 3D motion model was built as described in Section 5.3. A free-breathing 2D sagittal cine of approximately 1 minute was acquired, and all images were registered to end-exhale with image registration [132]. From the resulting motion-fields two displacement curves were extracted: one of the diaphragm (in feet-head) and one of the chest wall (in anterior-posterior). Subsequently, the diaphragm and chest wall were identified on 3D images, and motion-field representation coefficients were estimated such that the 3D motion-fields coincide with the 2D displacement curves at both anatomical landmarks.

In more detail, the motion-field representation coefficients $\mathbf{c}_t \in \mathbb{R}^{N_y \times 1}$ were estimated by solving the linear system

$$\mathbf{A}\mathbf{c}_t = \mathbf{b},$$

for every dynamic t in the 2D cine, with

$$\mathbf{A} := \begin{pmatrix} \Phi_{3D}^{AP}(\mathbf{r}_{c,3D}) \\ \Phi_{3D}^{FH}(\mathbf{r}_{d,3D}) \end{pmatrix} \in \mathbb{R}^{2 \times N_y}, \quad \mathbf{b} := \begin{pmatrix} \mathbf{d}_{t,2D}^{AP}(\mathbf{r}_{c,2D}) \\ \mathbf{d}_{t,2D}^{FH}(\mathbf{r}_{d,2D}) \end{pmatrix} \in \mathbb{R}^{2 \times 1}.$$

Here superscripts AP and FH denote respectively the extracted AP and FH components of the vector-fields, \mathbf{d}_t denotes the motion-fields at dynamic t , and $\mathbf{r}_{d,3D}$, $\mathbf{r}_{d,2D}$ and $\mathbf{r}_{c,3D}$, $\mathbf{r}_{c,2D}$ denote respectively the corresponding spatial coordinates of the diaphragm (3D, 2D) and chest wall (3D, 2D). The linear system was solved for each dynamic t as

$$\mathbf{c}_t = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b},$$

for a small value of $\lambda \in \mathbb{R}^+$, which prevents noise amplifications.

Solving the linear system above resulted in realistic looking 3D motion-fields that follow the dynamics of the diaphragm and chest wall in the corresponding 2D cine. Next, the end-exhale image from the 3D respiratory-resolved image sequence

was warped with the resulting motion-fields, resulting in a 3D time-resolved free-breathing image sequence. Following this procedure, the generated 3D motion-fields could now serve as the ground-truth motion-fields for further validations.

Finally, multi-channel k -space data was generated using the same GM3DR+SN trajectory as for in vivo acquisitions (Section 5.4) and coil sensitivities estimated from the acquired 3D in vivo data. The process above was performed for a total of two volunteers, resulting in two 3D time-resolved datasets, each about 1 minute in length with a frame rate of 4 Hz, and accompanied with ground-truth motion-fields for validation.

Performance evaluation

Several aspects of our pipeline are evaluated on the synthetic time-resolved data. Firstly, a comparison was made between Gaussian Process regression and ordinary linear regression to evaluate whether the added benefit of uncertainty estimation with Gaussian Processes comes at the cost of performance.

Secondly, to assess the effect of the image registration on the final performance, the difference in performance was evaluated between two motion models. The first motion model was extracted from the generated synthetic 3D+t time-resolved data (see above). For this we follow our conventional pipeline with image registration on respiratory-sorted images. To obtain the binned images, binning was performed in image-space by averaging the time-resolved complex images per bin. We refer to the first motion model as the ‘Approximate motion model’. For the second motion model, PCA was performed on the ground-truth motion-fields. To make a fair comparison, the ground-truth time-resolved motion-fields were binned to respiratory-resolved motion-fields in the same way as was done in the conventional pipeline prior to extracting the motion model with PCA. We refer to the second model as the ‘True model’. It should be noted that the difference between the ‘Approximate motion model’ and the ‘True model’ is given by the image registration; the former performs image registration on the binned images, whereas the latter extracts the ideal respiratory-resolved motion-fields directly from the ground-truth motion-fields. The performance difference between these two models shows the effect of the image registration on the final error.

To prevent a bias in the evaluation due to the time-varying breathing pattern, 6-fold cross validation was performed for all analyses. More specifically, the time-resolved data was partitioned into 6 contiguous but disjoint sets. Next, the training and inference were performed 6 times, once for each possible combination of train-test sets according to a 5:1 ratio. The mean and standard deviations were computed over all experiments and reported in Table 5.1.

In silico and in vivo robustness test

To assess the robustness of the proposed framework we trained the model on three minutes of free-breathing and performed inference over all remaining GM3DR spokes. Data were simulated with the digital XCAT phantom, and were acquired from 5 volunteers on the MR-linac, as described in Section 5.4. The volunteers were not instructed regarding their breathing. The preparation phase was performed on three minutes of data, as described in Section 5.3, Section 5.3 and Section 5.3. Inference was performed on the remainder of the data according to the pipeline in Eq. (5.6), and finally the rejection criterion was evaluated as described in Section 5.3.

The observable motion in FH spokes is typically used as a surrogate for respiratory motion [79]. The first principal component of the motion model most significantly contributes to the final reconstructed motion-field. Hence, the motion in the FH spokes should be similar to the motion in the first principal component of the motion model, i.e. the output of the first GP. Therefore, we qualitatively compared the output of the first GP (posterior mean) with the projection images obtained from these FH k -space spokes. These projections were obtained by performing an FFT along the readout direction of the spokes. The comparison was performed on data acquired about two minutes after the training data.

In vivo feasibility tests of high-speed inference at 69 Hz

In order to perform the input data dimension reduction, train the GP, and build a motion model, the presented framework requires both golden-mean 3D radial kooshball and SN spokes in the training phase. In the inference phase, however, the golden-mean radial spokes are not required, and the use of only the feet-head spokes would lead to a much higher inference frequency. With this experiment we investigate the feasibility of such high-speed inference.

For this purpose, data were acquired in two phases. In the first phase, GM3DR+SN data were acquired in order to train the model. In the second phase, only the SN spokes were acquired at 69 Hz. Data were continuously acquired from four healthy volunteers (volunteers 2-5) over 9 minutes, with the same acquisition parameters as described in Section 5.4. The first 7 minutes of data were acquired with GM3DR+SN, and the last 2 minutes with SN-only. To mimic a realistic setting, only the first three minutes of the GM3DR+SN were used to perform all preparation steps in the training phase. The rejection criterion was calibrated on the last 30 seconds of the 3 minutes of GM3DR+SN training data. Finally, inference was performed on all available SN-only data, which covered around 2 minutes for most volunteers. This inference was performed on very high temporal resolution. The motion model was built to represent average breathing motion, and it is therefore likely that the data acquired at this high temporal resolution will represent motion states that slightly differ from the average breathing motion states. This could decrease the correlation between the training data and the inference data, which could in turn result in a high uncertainty. To make the rejection criterion less sensitive to this effect, we evaluated it at 5 Hz rather than at 69 Hz by only flagging dynamics for which the current dynamic and all 13 preceding dynamics (acquired in around 200 ms) exceeded the

rejection threshold. Note that this temporal resolution for the rejection criterion of 5 Hz is still sufficiently fast for MRgRT [22, 23]. Alternatively, the SN input spokes could be averaged in time, allowing for a trade-off between SNR of the input data and the temporal resolution of the outputs. Preliminary results in Supporting Information Figure S1 indicate that this is indeed feasible.

We also evaluated the effectiveness of the rejection criterion based on the GP posterior uncertainty by instructing volunteers 3 and 5 in the final minutes of the SN-only phase. Volunteer 3 was instructed to perform in sequence: 1) normal breathing, 2) a switch to chest-only motion, 3) bulk motion, and 4) normal breathing. Volunteer 5 was instructed to perform 1) irregular breathing, and 2) a bulk motion of the abdomen/thorax. During these undesirable movements, we compared the rejected dynamics with another independent method based on the center-of-mass (COM) per coil. The 3D COM coordinates were computed from the 3D radial spokes [136], and the comparison was performed with the left-right and feet-head COM coordinates from a single channel of k -space data. The general idea is that changes in COM coordinates indicate large body motions, and can therefore be used for bulk motion detection [137]. This is especially pronounced for COM coordinates in AP and LR, since these are relatively less affected by respiratory motion.

Finally, we performed an additional analysis to gain further insight in the uncertainty estimated by the GP, and thereby the dynamics rejected by the proposed method. We analyzed the temporal behavior of the total uncertainty σ_t (see Section 5.3) over the course of the whole acquisition, which was around 10 minutes for some volunteers. A gradual build up of this uncertainty would indicate a gradual decrease in the correlation between real-time measurements and the training data, and could therefore potentially give insights in organ drifts that could affect the inference during the SN-only phase.

5.5 Results

In silico XCAT experiment

The left part of Figure 5.2 shows the results of the correlation analysis between GP posterior uncertainty and EPEs, indicating a positive correlation. Additionally, a statistical correlation analysis was performed between all breathing patterns other than the normal breathing pattern that was used for training. For respectively abdominal-only, chest-only and amplitude drifts, this resulted in correlation coefficients $\rho = 0.99 \pm 0.00$, $\rho = 0.99 \pm 0.00$ and $\rho = 0.90 \pm 0.02$, which all indicate a strong positive correlation between the GP posterior uncertainty and EPEs.

The right part of Figure 5.2 shows notched boxplots of EPEs on dynamics rejected (red) and not rejected (green) by the rejection criterion. The results for normal breathing data, which was not exactly the same as the training data, show an interquartile range (IQR) of 0.36 mm - 0.78 mm. Here, the IQR is defined as the data between the 25th and 75th percentiles. For abdominal, chest and drifts, the IQR was respectively 0.36 mm - 0.68 mm, 0.29 mm - 0.61 mm and 0.43 mm - 0.88 mm. Overall, the maximum EPE of rejected dynamics was 1.31 mm. The notches

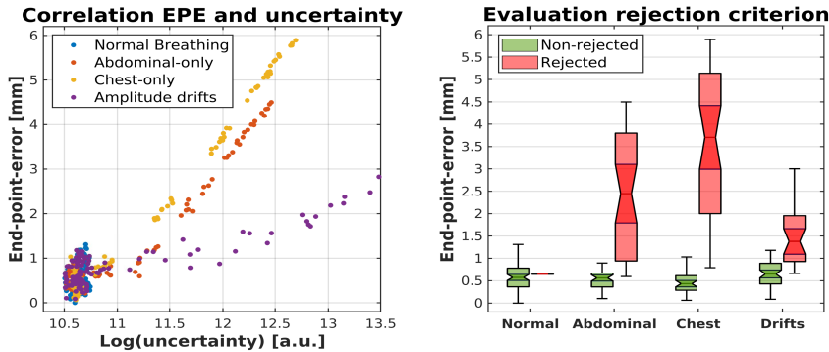


Fig. 5.2: Left: correlation between end-point-error (EPE) and GP posterior uncertainty for all simulated breathing patterns. Note the log-scale on the x -axis. Additional analyses indicated for abdominal-only, chest-only and amplitude drifts, a correlation coefficient of respectively $\rho = 0.99 \pm 0.00$, $\rho = 0.99 \pm 0.00$ and $\rho = 0.90 \pm 0.02$ 0.90 ± 0.01 . Right: evaluation of the rejection criterion for all simulated breathing patterns; EPEs are computed on rejected and non-rejected dynamics. The notches indicate the 95% confidence interval around the medians, and show strong statistical evidence that the EPEs of rejected dynamics are higher than non-rejected dynamics. This shows the effectiveness to reject erroneous motion estimates.

indicate the 95% confidence interval around the medians, and show strong statistical evidence that the EPEs of rejected dynamics are higher than non-rejected dynamics.

Both results in Figure 5.2 empirically confirm our hypothesis that the GP posterior uncertainty can be used for real-time quality assurance, since the uncertainty is highly correlated with the error in the motion estimates. Especially the right part of Figure 5.2 highlights that the rejection criterion based on the GP posterior uncertainty rejects estimates with high EPEs and preserves low EPEs during undesirable breathing patterns.

Synthetic time-resolved in-vivo data

Table 5.1 shows the results of the synthetic time-resolved reconstructions. Firstly, this shows that Gaussian Processes and Linear regression perform similarly, but the Gaussian Process also provides a measure of uncertainty for its fit. Both achieve high accuracy with errors below 1 mm on average. Secondly, both methods perform on average better with the ‘True model’ (i.e. derived from the ground-truth respiratory-resolved motion-fields), as opposed to the ‘Approximate model’ (i.e. derived through image registration). Further analysis in the second-to-last column shows that the registration errors may be the largest source of errors.

	Linear		Gaussian Processes		Other metrics		
	Approx. model	True model	Approx. model	True model	Registration error	Amplitude	
Vol. 1	diaphragm	0.78 +/- 0.52	0.35 +/- 0.22	0.78 +/- 0.52	0.30 +/- 0.19	0.68 +/- 0.54	15.00
	left kidney	0.58 +/- 0.29	0.45 +/- 0.25	0.59 +/- 0.30	0.47 +/- 0.29	0.35 +/- 0.22	7.51
	right kidney	0.89 +/- 0.59	0.26 +/- 0.19	0.90 +/- 0.59	0.20 +/- 0.13	0.86 +/- 0.57	12.79
Vol. 2	diaphragm	0.88 +/- 0.37	0.76 +/- 0.38	0.88 +/- 0.34	0.47 +/- 0.26	0.47 +/- 0.19	11.25
	left kidney	0.51 +/- 0.34	0.47 +/- 0.35	0.55 +/- 0.34	0.03 +/- 0.02	0.25 +/- 0.03	9.70
	right kidney	0.88 +/- 0.37	0.45 +/- 0.34	0.88 +/- 0.35	0.12 +/- 0.07	0.71 +/- 0.25	9.35

Table 5.1: Synthetic time-resolved reconstruction results [mm]. A comparison in end-point-errors between linear and Gaussian Process regression, using an approximate model derived with image registration (Section 5.3) and the True model derived from ground-truth respiratory-resolved motion-fields (Section 5.4).

In silico and in vivo robustness test

Figure 5.3 shows the results of the training phase (first column), an overview of the inference phase (second column), and a zoom in a region with increased estimation uncertainty. The number of ranks in the motion model per volunteer - and thus number of GPs trained per volunteer - was determined through the L-curve approach on the explained variances, and resulted for respectively the digital phantom and volunteers 1-5 in the ranks $R = 2, 3, 3, 3, 1, 3, 2$. The figure shows the output of the first GP, since this is generally the most interpretable component; an example of the other GP's outputs can be found in Supporting Information Figure S2. In the first column the confidence interval should be observed; this was manually set in the training phase to account for potential errors in the training targets. The results in the overview in column 2 show realistic motion traces, and differences in breathing pattern can be observed between the different volunteers (different rows). In the zoom in column 3 it can be observed that the uncertainty increases in several scenarios: during a switch from normal to abdominal-only breathing (for the digital phantom, $t \approx 25$ s), during deep inhales (volunteers 4 and 5) and during deep exhales (volunteers 1 and 2). The increased uncertainty is especially pronounced for dynamics with breathing amplitudes outside the range of the training data, i.e. larger than average exhales or inhales. This observation is explainable, since the GP posterior uncertainty increases with the distance to the training data, as evaluated by the kernel function. Animated Figure 1 in the supporting files shows the posterior mean and spatial estimation uncertainty maps for volunteer 1, as derived in (5.8). The animated figure shows the inference over the first 35 seconds in the second column in Figure 5.3, and visualizes every 4th dynamic with a total 60 frames at 4 Hz. Similar to the results in Figure 5.3, the respiratory traces appear smooth overall, however small high frequency oscillations can be observed which could be related to cardiac motion or measurement imperfections (eddy currents). We refer the reader to 4.6 for an overview of all animated figures.

The comparison with the projection images obtained from FH k -space spokes is shown in Figure 5.4. Moreover, the figure shows the results of the application of the rejection criterion, as discussed in Section 5.3. Overall the posterior mean of the first GP coincides with the projection images. The framework mostly rejected end-exhale dynamics, which were likely outside the range of the training data. Interestingly,

for volunteer 2 the frequency of rejections is relatively large at the beginning, and decreases after about 400 seconds in the acquisition. For volunteers 4 and 5, it can be observed that the framework also rejected deep inhales, which were substantially different from the average breathing in the training data. The change in the frequency of rejections could be an indication of a breathing pattern change. The large number of rejections in exhale could be due to drifts of internal organs, resulting in a different end-exhale position. These hypotheses are further analyzed in Figure 5.5.

In vivo experiments

In vivo feasibility tests of high-speed inference at 69 Hz

In this experiment, inference was performed at 69 Hz, using only three mutually orthogonal readouts as input. The results are shown in Figure 5.6, which compares the GP inference with projection images (background). The Animated Figures 2-5 in the supporting files show the posterior mean and spatial estimation uncertainty maps for volunteers 2-5, as derived in (5.8). These animated figures show the inference over the first 35 seconds of the data in Figure 5.6. The animation shows every 40th dynamic with a total 60 frames, visualized at 4 Hz. An exception to this is volunteer 4, for whom every 80th dynamic is visualized at 4 Hz because of a low breathing frequency. We refer the reader to 4.6 for an overview of all animated figures.

Overall the inference during normal breathing returns plausible motion traces. Volunteer 2 shows a regular breathing pattern, which changes to larger breathing amplitudes halfway the scan, and eventually shallow breathing at the very end of the scan. Volunteers 3 and 4 show a regular breathing pattern, and volunteer 5 shows a very irregular breathing pattern. The irregularities in the breathing are mostly accompanied by larger uncertainty, which is most pronounced for motion states that fall outside the range of motion in the training phase. The spatial uncertainty map for volunteer 3 shows a larger uncertainty than the other volunteers, although the visualized uncertainty range of 0 mm - 1 mm should be noted here. The relatively larger uncertainty could be related to the relatively large breathing amplitudes of volunteer 3.

For volunteer 2, hardly any dynamics were rejected until about 640 seconds into the acquisition, after which the breathing pattern seems to change to very shallow breathing, resulting in many rejections. For volunteers 3 and 5, hardly any dynamics were rejected before the instructed motion changes, which started at respectively 535 seconds and 718 seconds in the acquisition. For volunteer 4, almost all end-exhale dynamics were rejected. Since this occurs after more than 10 minutes in the acquisition, this could be caused by organ drifts that changes the internal positions of the organs in such a way that it leads to rejections. This hypothesis is further analyzed in Figure 5.5. The results show that for most volunteers the posterior uncertainty gradually increases over the course of the acquisition, which indicates a gradual decrease of the correlation between the acquired data and the data in the training set. This could be explained by physiological drifts, which are also visually observable in Figure 5.3 and the projection images over the course of the whole scan (not shown).

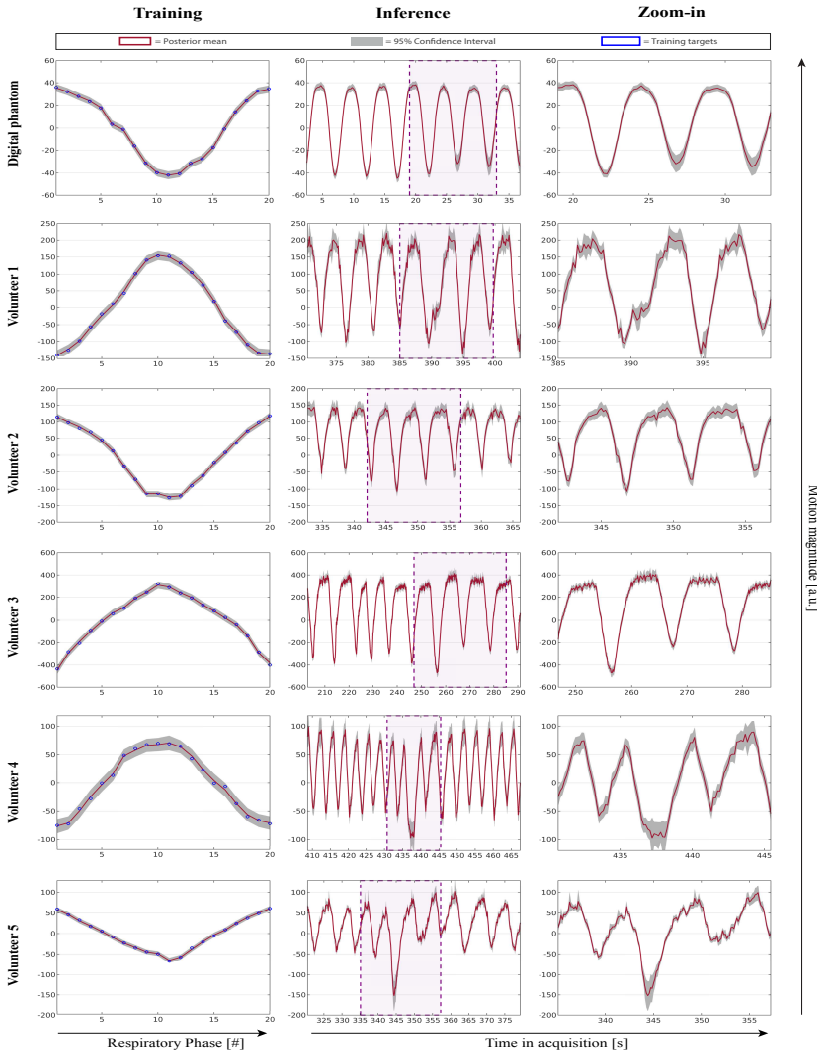


Fig. 5.3: Qualitative results of GP motion and uncertainty estimations in a digital phantom and five volunteers. The blue dots represent the training targets, the red line the posterior mean as estimated by the first GP - which corresponds to the first principal component of the motion model - from three readouts, and the gray shaded areas are the 95% confidence intervals obtained from the estimation uncertainty as output by the GP. The first and second columns show an overview of respectively the training and the inference phase, the third column shows a zoom-in on a region with a slightly different motion pattern or increased estimation uncertainty. The increased uncertainty mostly occurs at the highest motion amplitude levels. Note that the units on the vertical axes are arbitrary; the scaling depends on the magnitudes of the principal components in the motion model. However, in all cases the higher values indicate exhales, while the lower values indicate inhales.

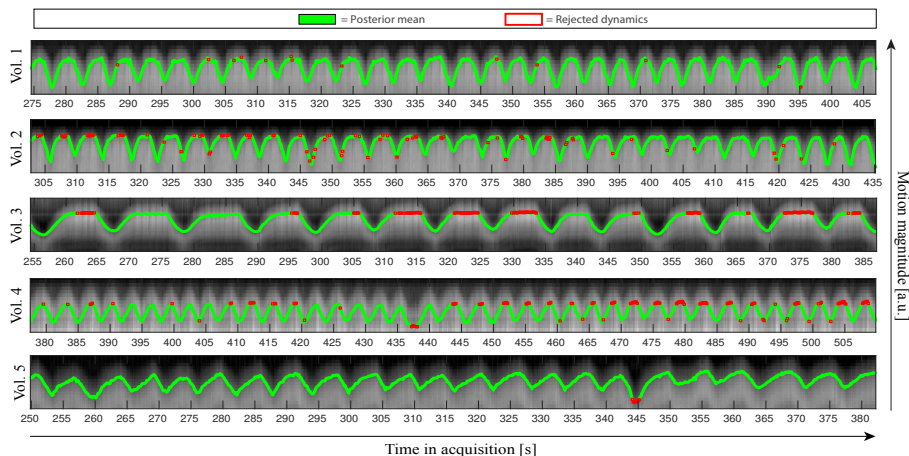


Fig. 5.4: Comparison between GP predictions and projection images. The image shows the comparison for all five volunteers between the posterior mean as estimated by the first GP corresponding to the first principal component of the motion model (green), projections on the FH-axis as obtained by an FFT over the readout of FH spokes (background), and rejected dynamics (red marks). The inference was performed on data acquired about two minutes after the training data were acquired.

To evaluate the rejection criterion, volunteers 3 and 5 were asked to perform a specific motion that forced a deviation from the data in the training phase and would render the motion model inapplicable. The results are shown in Figure 5.7. The top figure shows a comparison with the projection images obtained with an FFT along the readout of the FH k -space spokes, as well as the evaluation of the rejection criterion, as discussed in Section 5.3. The bottom part of the figure shows the center-of-mass (COM) coordinates in left-right (orange) and feet-head (purple), extracted from the 4th channel of k -space data. The COM coordinates serve as an independent visualization of the volunteer’s behavior during this experiment. The black arrows indicate the changes in motion, which are accompanied by a change in the pattern of the COM. Volunteer 3 (left), changed breathing pattern at arrow #1, performed bulk motion at arrow #2, and returned to a normal breathing pattern at arrow #3. Volunteer 5 shows irregular breathing with a slight drift towards inhale up to arrow #4, and bulk motion after arrow #4. The framework successfully detected abnormal motion, and rejected many dynamics. It should be noted that Volunteer 3 appears to return to a different position after the bulk motion, since the COM in left-right is slightly higher in the final part of the figure. Although this is not reflected by the framework’s rejections, in practice the large number of consecutive rejections preceding this event should already be indicative that the model’s predictions are too uncertain, and that the motion model should thus be updated. For volunteer 5, almost all dynamics were rejected during bulk motion. Only two sets of consecutive dynamics were not rejected during bulk motion, during which the COM coordinates indicate a similar position as in the preceding normal breathing phase.

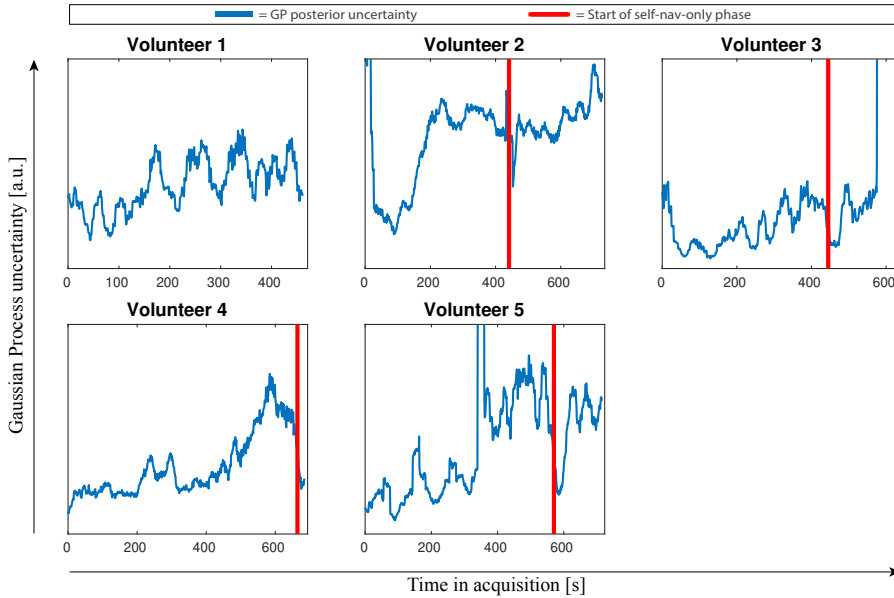


Fig. 5.5: Analysis of the GP posterior uncertainties over the course of the whole acquisition. For volunteers 2-5 the acquisition switched to an SN-only acquisition at the end of the acquisition, indicated by the solid red line. A moving average filter with a window of 10 seconds is applied to the uncertainties to highlight the patterns over the whole acquisition rather than within the breathing cycles.

5.6 Discussion

Summary of innovations

In this work we presented a probabilistic framework which addresses two major technical hurdles towards real-time adaptive MRgRT simultaneously: real-time 3D motion-field estimation and uncertainty quantification. The framework was built on the idea that low-dimensional motion information can be extracted from few read-outs of k -space data [130]. This idea was exploited via a two-step reconstruction approach, in which first a motion model was built, and subsequently the model's coefficients were inferred from the data.

For the inference, a probabilistic machine learning regression technique based on Gaussian Processes was used. Due to its probabilistic nature, this technique not only estimated the most likely motion model coefficients, but also provided a measure of estimation uncertainty. The inferred model coefficients combined with the motion model yielded the motion-fields required for MRgRT. The inferred uncertainty was hypothesized to be useful for real-time quality assurance during radiotherapy. This hypothesis was empirically confirmed in simulations, in which it was shown to enable the detection of erroneous motion estimates, which - if left undetected - could in practice result in harmful radiation to organs-at-risk.

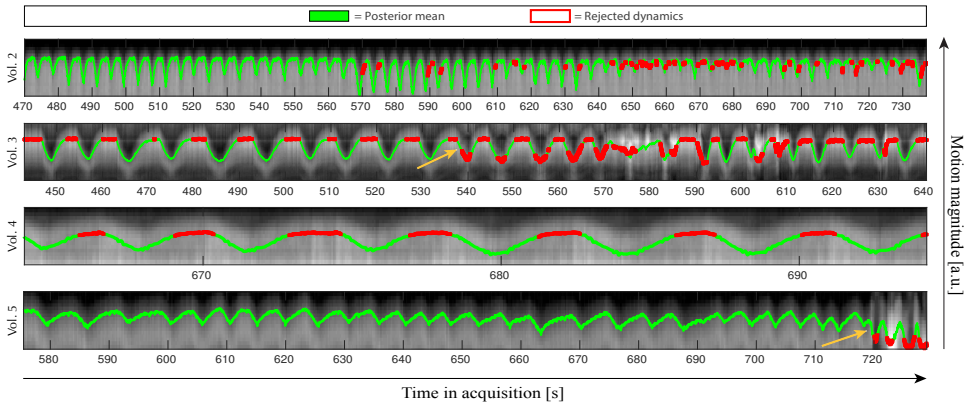


Fig. 5.6: High-speed in-vivo inference and quality assurance. The posterior mean of the first GP (green), and rejection criterion evaluations (red), are compared with projections on the FH-axis for all four volunteers with SN-only data available (2-5). Volunteers 3 and 5 were instructed to perform motion which would render the motion model invalid. Volunteer 3 performed in sequence: 1) normal breathing; 2) a switch to chest-only breathing; 3) bulk motion; 4) normal breathing, and the first abnormal event started around 535 seconds in the acquisition, as indicated by the yellow arrow. Volunteer 5 performed in sequence: 1) normal breathing; 2) irregular breathing; 3) bulk motion, and the first abnormal event started around 718 seconds in the acquisition, as indicated by the yellow arrow. The GP posterior means are scaled to visually overlap the pattern in the projection images.

Rejection criterion vs. center-of-mass (COM)

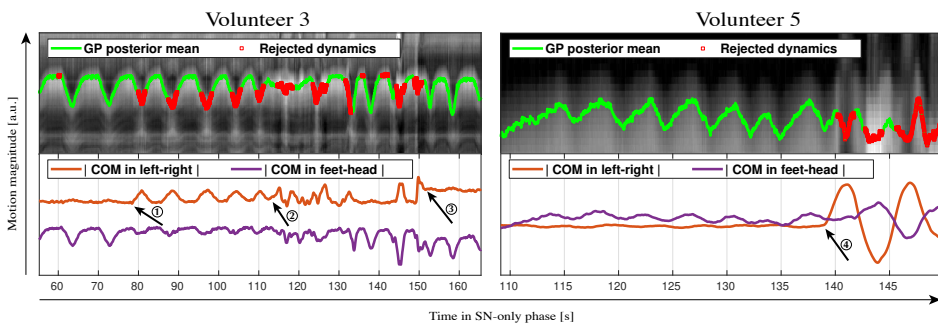


Fig. 5.7: This figure shows the result of the high-speed inference for volunteers 3 (left) and 5 (right), as discussed in Section 5.4 and Section 5.5. The bottom graphs show a comparison with center-of-mass (COM) coordinates in left-right (orange) and feet-head (purple). The black arrows, labeled 1-4, indicate four events where the motion was changed, as can also be observed from the changing patterns in the COM-coordinates; 1) volunteer 3 switches breathing pattern, 2) volunteer 3 performs bulk motion, 3) volunteer 3 returns to normal breathing, 4) volunteer 5 performs bulk motion.

As opposed to other machine learning techniques such as neural networks, the training of the GPs could be done in about half a second, and required only a minimal amount of training samples (20 in this work). Moreover, the inference time was very short (≈ 0.1 milliseconds per dynamic) due to the availability of a rapidly computable closed-form analytical expression for the GP's posterior distribution. Altogether, this makes the GP a natural fit in a real-time MRgRT workflow where little latency is essential.

The complete framework was validated *in silico* with simulations using a digital phantom, and *in vivo* with MR-linac data of five volunteers, thereby taking into account different breathing patterns and bulk motions.

Potential impact

The presented framework can have several applications in MR-guided radiotherapy. Firstly, the framework allows to infer 3D motion-fields at 69 Hz. The inference at this speed is more than sufficient to resolve abdominothoracic motion during MR-guided radiotherapy [22, 23]. Although 5 Hz should be sufficient for this application, the feasibility of inference at 69 Hz could open up possibilities of applying radiotherapy to tumors subject to cardiac motion, such as central lung tumors. Another application could be cardiac radio-ablation [138]. This is an emerging non-invasive treatment technique of cardiac arrhythmias with highly focused radiotherapy, which could benefit from high-speed tracking of myocardial landmarks. However, it should be noted that the extension to cardiac is not trivial, as it will require to image and model cardiac motion at high speeds in addition to respiratory motion.

Secondly, the proposed framework not only estimates motion-fields, but also provides a measure of estimation confidence. We have demonstrated an example of how this measure of confidence can be used for real-time quality assurance by designing a rejection criterion based on this estimation confidence. In practice, this could be useful to detect unexpected motions during radiotherapy, such as a change of breathing pattern or bulk motions, which could result in erroneous motion estimates due to an unsuitable motion model built for normal breathing. Without any measure of quality assurance, such estimates could result in harmful radiation to organs-at-risk. With the real-time quality assurance proposed in this work, the treatments could be (temporarily) halted to assure the patient's safety during such potentially erroneous motion estimates.

Thirdly, in the current proof-of-concept work we have designed the rejection criterion around a global measure of uncertainty, which assumes every voxel receives an equal weight in the uncertainty measure (see Section 5.3). However, from Eq. (5.8) it can be observed that this global measure of uncertainty could be transformed into a local measure of uncertainty by including the actual motion model Φ . Such a local measure would allow to derive multiple local rejection criteria that could allow for more accurate local quality control of the motion-fields.

Finally, the availability of time-resolved 3D motion-fields over the course of a radiotherapy treatment could be used for retrospective dose accumulation calculations [120]. Such calculations provide insights in the actual dose deposited to the target

tumor and surrounding organs-at-risk during a treatment, which can be taken into account to improve subsequent treatment planning.

Related work

Real-time 3D inference as proposed in this work was shown before. For example, our own method MR-MOTUS [130] achieved 3D motion-field reconstruction at 6.7 Hz, Li *et al.* [122] reconstructed 3D CT volumes and motion-fields at about 4 Hz, MR-SIGMA [34] estimated 3D MRI volumes at 3.3 Hz, and cine-based MRI methods such as proposed in Stemkens *et al.* [33] achieved 3D motion-field reconstruction at about 2 Hz. Slightly different type of methods are based on surrogate signal models [35, 36, 51, 89, 112], and also inferred 3D motion-fields at high temporal resolution from (several) 1D surrogate signals. Recently, also several deep learning (DL) based methods were proposed for real-time 3D inference. For example, Romaguera *et al.* [139] inferred 3D motion-fields from a pre-treatment volume and real-time 2D images, and also predicted the motion for future timepoints. Terpstra *et al.* [140] proposed TEMPEST, a network that estimates 3D motion-fields directly from highly undersampled, aliased, 3D images with a frame rate up to 5 Hz.

Frequently, non-DL-based real-time inference methods exploit a low-rank motion model similar to the one employed in this work, which is typically obtained by a retrospective compression of motion-fields using principal component analysis [31, 33, 85, 87, 88], or by decoupling the motion-fields into spatial components and temporal components based on surrogate signals [35, 36, 89]. In Huttinga *et al.* [114] and Huttinga *et al.* [130], we also employed a low-rank model, but estimated its components directly from k -space data by solving a large-scale non-linear optimization problem [84].

Most of the methods above only estimate motion-fields, without any measure of confidence or motion model applicability. An exception is the work by King *et al.* [31], where a low-rank 3D motion model was fit to incoming 2D navigator MR-images and the motion model's applicability was constantly evaluated as the image similarity after registration. A notable difference is that King *et al.* [31] used 2D cine navigator images, whereas we used three 1D spokes. The network proposed by Romaguera *et al.* [139] also outputs estimation uncertainties, but these were not used for quality assurance. The proposed rejection criterion is therefore most similar to the model applicability test in King *et al.* [31].

The employed dimension reduction techniques for both the input and output space were proposed to make GPs fit in a high-dimensional pipeline. Similar dimension reduction techniques on the output space were previously proposed in PCA-GPs [141], with the aim to overcome the same challenges of combining GPs with high-dimensional output data. In this work, GPs were used to simultaneously perform regression over multiple scalar functions with multi-dimensional inputs, thereby assuming no correlations between the individual GP outputs. Alternatively, correlation in the outputs could be modeled with multi-task GPs [134]. Although this could improve the performance in theory, our preliminary results in Sbrizzi *et al.* [72] with multi-task GPs showed little to no improvement.

Points of improvement and future work

Several aspects of this work could be improved for a clinical application. The proposed GP framework provides all the utilities required for real-time adaptive MRgRT. Further improvements on the method are modular and not strictly necessary. We expect the largest gains in performance to reside in improvements to the motion model, for example by better image reconstructions, or better image registration methods. Moreover, currently an average breathing motion model is used to estimate time-resolved breathing. The accuracy could possibly be improved, and the number of rejected dynamics could be reduced, if a more expressive motion model were to be used. For example, a model built on 3D+t time-resolved cine MR-images, as proposed in King *et al.* [31], or built from unsorted k-space data with low-rank MR-MOTUS [114]. Additional improvements could be obtained by using an adaptive (possibly respiratory-resolved) motion model, which is updated in parallel with the real-time estimates according to the measure of uncertainty, e.g. similar to Baumgartner *et al.* [142].

We have validated our framework *in silico* with end-point-errors, investigating four different breathing patterns and synthetic time-resolved 3D data. Moreover, a comparison was made with *in vivo* FH projection images. For a practical application of the proposed method, but also for other similar methods, more extensive *in vivo* validations are required. Unfortunately this is a very challenging task, since no ground-truth deformation-fields will be available. Possible validation strategies could be comparisons with manually tracked targets on 2D cine MR-images, that are interleaved with the acquisition required for the real-time 3D motion estimation. Another option would be the use of realistic *ex vivo* phantoms with controlled movement that can be measured independently, or a comparison with quantitative displacement estimates obtained with external sensors. More extensive validations along these lines will be the subject of future work.

For several volunteers a sensitivity of the framework was observed to high-frequency oscillations in the data. These oscillations could be physiological, e.g. due to local cardiac motion or related blood flows. In the current framework, these high-frequency oscillations could not be coupled to local motion since this was not incorporated in the motion model. For this reason an attempt was made to remove the high-frequency oscillations by thresholding the input data based on the relative energy content above the realistic breathing frequency range (see Section 5.3). More complex motion models could possibly allow to identify the source of these oscillations. Alternatively, since inference could be performed with a speed of 69 Hz - which is about 10 times the required speed for MRgRT - the oscillations could simply be filtered out by performing a temporal averaging on the GP inputs. The averaging would allow for a trade-off between temporal resolution and smoothed outputs. Preliminary results in Supporting Information Figure S1 indicate that this is indeed feasible.

In this work, only correlations in the data were considered, but temporal correlations were not yet taken into account. Such an extension would require to model temporal correlations, but could prove valuable for near-future predictions. A possible downside of including temporal correlations is that it could impose strong restrictions on the temporal behavior of the motion (e.g. periodicity), which is why we

have not considered it in this work.

In general, it should be noted that all improvements discussed above are modular, and would require little to no changes in the general pipeline of the proposed framework.

5.7 Conclusion

We have presented a probabilistic framework for simultaneous real-time 3D motion and uncertainty estimation. The complete framework, including the rejection criterion, allowed to preserve low EPEs (75th percentiles ≤ 0.88 mm) during four different breathing patterns in simulations. Without the proposed rejection criterion these breathing patterns would have resulted in EPEs up to almost 6 mm, which - if left undetected - could lead to harmful radiation to organs-at-risk. The framework estimated in vivo motion that corresponds well with FH projections. Moreover, it flagged dynamics during which bulk motion and changes of breathing patterns were performed. This flagging strategy could be used to ensure the safety of the patient by (temporarily) halting the radiation, and continuing whenever confidence in the estimated motion is restored. Altogether, the framework tackles two major technical challenges for real-time adaptive MR-guided radiotherapy, real-time 3D MR-based motion estimation and uncertainty quantification, and it thereby paves the way to reach the ultimate potential of the MR-linac.

Acknowledgements

This work was supported in part by the Dutch Research Council (NWO) under Grant 15115.

5.8 Supporting Figures and Videos

The supporting videos are part of the main body of this chapter and can be found through the QR code below the abstract, or through the following link: <https://doi.org/10.6084/m9.figshare.20480697.v1>. A snapshot of one of the videos is shown below.

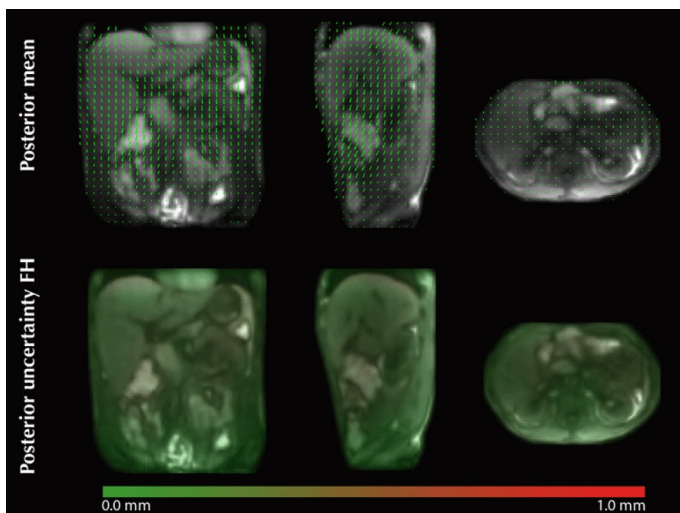
Caption of Supporting Video 1

This video shows the posterior mean and spatial estimation uncertainty maps for volunteer 1, as derived in Eq. (5.8). The video shows the inference over the first 35 seconds in the second column in Figure 5.3, and visualizes every 4th dynamic with a total of 60 frames at 4 Hz.

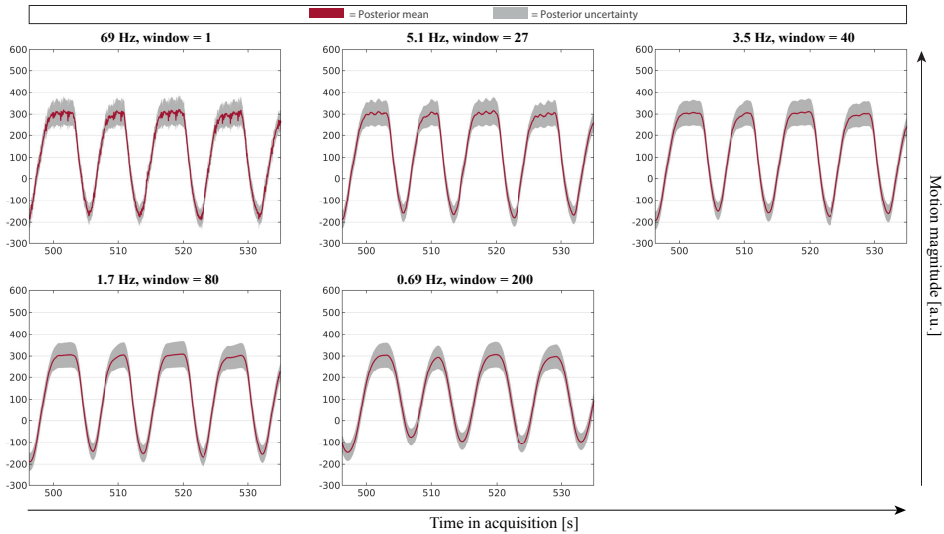
Captions of Supporting Videos 2-5

These videos show the posterior mean and spatial estimation uncertainty maps for volunteers 2-5, as derived in Eq. (5.8). These videos show the inference over the first 35 seconds of the data in Figure 5.6. The video shows every 40th dynamic with a total of 60 frames, visualized at 4 Hz. An exception to this is volunteer 4, for whom every 80th dynamic is visualized at 4 Hz because of a low breathing frequency.

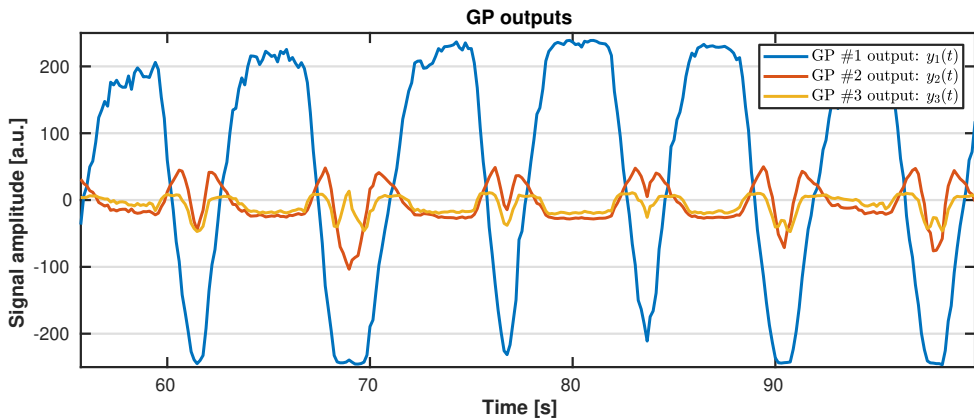
5



Supporting Video 5.1: Snapshot of one of the Supporting Videos



Supporting Figure 5.1: The effect of temporally averaging the input SN-spokes. Several reconstructions of the first GP for volunteer 3 are shown for which a symmetric moving mean filter is applied with a varying width: 1, 27, 40, 60, 80, and 200 dynamics. This averaging increases the temporal latency, as indicated in the titles, since the data required by the second half of the window lies in the future. However, evidently, the filtering also results in smoother reconstructions. Hence, temporal filtering allows to make a trade-off between the temporal resolution of the reconstructions and the SNR. Temporal averaging with a width of 27 dynamics would result in sufficient temporal resolution for real-time adaptive MRgRT [22, 23].



Supporting Figure 5.2: All GP outputs for volunteer 3.

Chapter 6

General discussion

6.1 Summary

The ultimate potential of the MR-linac is real-time adaptive MR-guided radiotherapy (aMRgRT), i.e. adapt the radiation plan in real-time according to real-time 3D motion estimates. One of the major technical roadblocks towards achieving this goal is the real-time 3D motion estimation. This thesis presented two new approaches in this context. The main method is called MR-MOTUS and is the subject of Chapters 2-4. MR-MOTUS was motivated by the observation that internal body motion exhibits a high level of spatio-temporal correlation. That is, motion-fields are compressible, or low-dimensional. A different method, but motivated by the same observation, is introduced in Chapter 5. Instead of reconstructing motion-fields from k -space data through the MR-MOTUS signal model, a machine learning method based on Gaussian Processes (GP) was trained to perform the reconstructions. Since GPs allow real-time inference and also provide a measure of estimation uncertainty, this method could open up new possibilities for real-time quality assurance. We describe the content of the chapters in more detail below.

Chapter 2 demonstrates the proof of concept. First, the MR-MOTUS signal model was derived that explicitly relates a motion-field and a reference image to k -space data. The signal model was validated, and a non-linear inverse problem with respect to the signal model was formulated to reconstruct motion-fields from k -space data. A reconstruction strategy with a quasi-Newton scheme was outlined, and the performance of the reconstruction algorithm was analyzed in silico by comparison with ground-truth motion-fields. The proof-of-concept was demonstrated by reconstructing in vivo 3D rigid head motion and in vivo 3D non-rigid respiratory motion from retrospectively highly undersampled k -space data. Moreover, the practical feasibility of the approach was demonstrated with a 2D non-rigid respiratory motion-field reconstruction on prospectively undersampled data, acquired with a 2D golden angle trajectory. Three retrospective undersampling strategies were investigated: low-resolution 3D Cartesian, 3D variable density, and 3D golden mean cones [68]. These MR-MOTUS reconstructions were compared with state-of-the-art image registration on image reconstructions with the same amount of undersampled data. Results show that MR-MOTUS outperformed the image registration, most notably for higher undersampling factors. Moreover, for high undersampling factors the 3D variable density and 3D cone undersampling outperformed the Cartesian undersampling.

The proof-of-concept presented in Chapter 2 has several points of improvements. For example, the framework required a breath-hold to reconstruct the reference image, the signal model assumed a single-coil setting, and 3D reconstructions on prospectively undersampled data were not yet demonstrated. Moreover, the framework could not yet exploit temporal correlation in motion-fields, because reconstruction were performed per dynamic rather than simultaneously over all dynamics. Chapter 3 addresses these points of improvements in the proof-of-concept, and tightens the gap to clinical application. In particular, highly undersampled 3D k -space data was prospectively acquired using a 3D golden-mean radial trajectory (GM3DR), the reference image was reconstructed from free-breathing using respiratory binning, and a novel coil compression technique was introduced that allowed to combine multi-

channel acquisitions with the single-channel signal model from Chapter 2. Finally, the framework was extended to 3D+t spatio-temporal motion-field reconstructions by using a low-rank motion model, which naturally models motion-fields in two components: a spatial component, and a temporal component. The low-rank model reduced the number of unknowns for space-time motion-fields by two orders of magnitude. This enabled 3D+t motion-field reconstruction with high temporal resolution on a desktop PC. The separation between spatial and temporal components also enabled separate regularization strategies in space and time. To demonstrate the framework's flexibility, reconstructions were performed on 2D+t and 3D+t abdominothoracic data, and 3D+t head-and-neck motion. Framework validations were performed with Jacobian determinants and by comparisons of the warped reference image with image reconstructions.

In Chapter 4, a two-step reconstruction approach was proposed for low-rank MR-MOTUS which enabled real-time reconstruction at 6.7 Hz. Moreover, several processing steps of the framework were improved that enabled MR-MOTUS reconstructions on prospective data acquired with the Elekta Unity MR-linac. The reconstructions exploited the separation of spatial and temporal components by the low-rank motion model introduced in Chapter 3. In the first phase of the two-step approach, the spatial component that was assumed to be fixed in time over several minutes is obtained with an offline reconstruction. In the second phase, the temporal component that encodes the dynamics in the motion-field is reconstructed per dynamic in an online reconstruction. The main rationale behind this approach is that the temporal component has few degrees of freedom (<10), and can therefore be reconstructed in real-time from few k -space data once the spatial component is available. Real-time MR-MOTUS was validated *in silico* on the digital XCAT phantom, and *in vivo* on free-breathing data of five healthy volunteers, acquired on the Elekta Unity MR-linac. The reconstruction were validated with the Jacobian determinants, by comparison with image reconstruction, and by comparison with respiratory motion surrogate signal. Results indicate the feasibility real-time 3D motion-field reconstructions on an MR-linac with a total latency of 170 ms (67 ms data acquisition, 103 ms reconstruction time).

A two-step approach as proposed in Chapter 4 is a natural way to enable real-time reconstruction with a limited amount of data. For similar reasons this was also previously proposed in the literature. A downside of this approach is that it assumes similar motion in the training and inference phase. Although this is true in most cases, this assumption could be violated in practice due to abnormal motion such as coughing or bulk motion. Evidently, a measure of confidence in real-time motion estimates would be beneficial for quality assurance during real-time adaptive MRgRT. In Chapter 5, we therefore proposed a probabilistic framework for joint real-time motion and uncertainty estimation, based on the machine learning theory of Gaussian Processes (GPs). In a preparation phase, a low-rank motion model was reconstructed offline, and a GP was trained to infer a probability distribution on motion-field representation coefficients from k -space data. In the inference phase, the trained GP was employed to infer the most-likely motion-field representation coefficients from just 3 readouts of k -space data. This minimal number of k -space data combined with the efficient inference of GPs resulted in an inference frame-

rate of 69 3D motion-fields-per-second. To additionally demonstrate the quality assurance capabilities, a rejection criterion based on the estimation uncertainty was designed to flag dynamics with potentially erroneous motion estimates. The proposed strategy preserved low end-point-errors (75th percentiles ≤ 0.88 mm) during four different breathing patterns in simulations. Without the proposed rejection criterion these breathing patterns would have resulted in end-point-errors up to almost 6 mm. Besides simulations, the motion reconstruction robustness of the framework was assessed on MR-linac data of five healthy volunteers. Two volunteers were asked to perform abnormal motion, which was detected by the framework. The detection of potentially erroneous motion estimates, as presented in this chapter, could play a crucial role to ensure patient safety during real-time adaptive MRgRT.

6.2 Discussion and future perspectives

Signal model imperfections

The signal model derived in Chapter 2 assumes the availability of a fixed reference image. This image is used in combination with the motion-fields to model the dynamic object in the scanner.

Since the image is assumed to be fixed, the model does not take local image variations into account. These could arise due to inhomogeneous B_0 or B_1 fields, or mass flow across the FOV boundaries. Chapter 2 stated that static B_0 and B_1^+ inhomogeneities are minimal for the targeted field strength of 1.5T, and the mass flow effects were shown to have minimal effect in simulations in Chapter 2. However, the lack of B_1^- terms in the model is more troublesome, since this made the extension to multi-channel receive array acquisitions challenging.

In Chapter 3, a novel coil combination was proposed, which allowed to compress multi-channel data to a single virtual coil with homogeneous sensitivity profile. This strategy worked well to make the framework compatible with multi-channel acquisitions, but this coil compression did reduce SNR. Moreover, such a coil compression does not actually exploit the possibilities of parallel imaging. Hence, extending the model to a true multi-channel model could significantly lower the amount of required data, and thereby improve the temporal resolution even more. Unfortunately this requires substantial adaptations. The reason for this is that for a fixed coil, the internal body motion is relative to the coil. As a result, the intensity attenuation induced by the coil sensitivities will dynamically vary, and the reference image will become a function of the motion-fields. For a more detailed discussion on this topic we refer the reader to the Supporting Information of Chapter 3.

Reference image

The acquisition of a reference image is not always straightforward. In an MR-guided radiation therapy workflow, a pre-treatment MRI will always be available to serve as a reference image. This is in contrast with other applications such as cardiac MR (CMR), where the final goal is usually to obtain such a fixed artefact-free image.

Chapter 2 reconstructed the reference image from breath holds, which is inconvenient in practice. Chapter 3, 4 and 5 improved upon this, and reconstructed a reference image from free-breathing data by respiratory binning. A downside is that respiratory binning reduces the acquisition efficiency with about a factor 10. As a result, about 10 times longer acquisitions are required for free-breathing reference image reconstructions, as opposed to breath-hold acquisitions. In practice, three minutes of free-breathing data were required, which is still sufficiently short to fit in the pre-treatment phase.

In future work a joint reconstruction approach should be considered, where motion-fields and a reference image are reconstructed simultaneously with an alternating scheme (see e.g. Burger *et al.* [143]). Preliminary simulation results (not shown) already indicated the feasibility of such an approach in case of severe (nonrigid) affine motion. Another promising direction would be to combine the motion-corrected image reconstructions in Bruijnen *et al.* [144] with MR-MOTUS.

Data acquisition

MR-signal acquisition is typically subject to several unwanted effects that corrupt the acquired signal. As a result, the signal model does not accurately model the acquired signal, and artifacts arise in the reconstructed images. At 1.5T two of the most prominent sources of signal corruptions are local field inhomogeneity and hardware imperfections. In image-space, the signal corruptions resulting from these sources mostly corrupt the image phase rather than the image magnitude. Fortunately the phase is of low diagnostic value in most applications. As a result, the signal corruptions can largely be removed by simply taking the absolute value of the complex MR-image, without losing valuable information.

The proposed MR-MOTUS framework allows to reconstruct motion-fields directly from k -space data. Operating in k -space makes the framework flexible in many ways, but it also makes dealing with model imperfections more difficult. These imperfections are usually due to corrupted image phase caused by processes that are not modeled, e.g. local field inhomogeneities or hardware imperfections. Besides causing a model mismatch, the image phase also contains information that is in fact useful for motion estimation. Applying MR-MOTUS in practice therefore requires a good understanding of the model's shortcomings, and strengths, such that a reasonable balance can be achieved between the two. As opposed to image space, the signal corruptions cannot simply be removed k -space. Instead, the data acquisition protocol should be designed to reduce these effects as much as possible. In practice, this meant using a high receiver bandwidth, and a short TE, and designing a coil combination that achieves an approximately homogeneous receive sensitivity within body in the FOV. The increase in receiver bandwidth serves two purposes: to reduce chemical

shift artifacts, and to reduce susceptibility artifacts, both of which are present for abdominothoracic imaging. The chemical shift artifacts cause certain molecules such as fat, to precess at a different frequency than proton. This causes a model mismatch, since fat will be shifted with respect to water in the reference image. As a result, the motion of voxels that contain both fat and water is incorrectly modeled by Eq. (1.2). Even with a reference image corrected for this, the forward model would still have to be adapted to account for the different precession frequency of fat. In other words, it would require to model the reference image in terms of a fat component and a water component. Increasing the bandwidth reduces both the water-fat-shift and susceptibility artifacts in the reference image, and a shorter TR minimizes the model mismatch since it minimizes the time for dephasing due to the off-resonance.

Another important consideration for the acquisition protocol is to capture sufficient information for reconstruction of the targeted motion-fields. The encoding capabilities are mostly governed by the order in which k -space data is acquired: the k -space trajectory. Evidently, the trajectories should cover as much of k -space as fast as possible to achieve high spatio-temporal resolution. Spatio-temporal motion-field reconstruction were in this thesis performed by splitting the acquired data in dynamics, each consisting of a finite number of sequentially acquired readouts, and subsequently performing the reconstruction in space and time simultaneously. To achieve a similar reconstruction quality for each dynamic, it is desirable to have sufficient and similar SNR for each dynamic. Based on variable density trajectories in compressed sensing MRI [28], this can be achieved with trajectories that for each dynamic have a high density in the center, and lower density on the periphery of k -space. To reduce the acquisition of redundant data, the cumulative sampling density should also linearly increase in time. This is also favorable for the spatial motion resolution, given the repetitive nature of most types of internal body motion.

Although Cartesian trajectories are still the standard for most clinical applications, they are not optimal for motion reconstruction due to the lack of the properties described above. We therefore considered non-Cartesian trajectories. One of the first non-Cartesian trajectories we considered was the 3D cones trajectory [145, 146], which was proposed for rapid and ultra-short TE (UTE) MRI. This trajectory acquires 3D k -space with very high time-efficiency, and was therefore suitable for initial tests with head motion estimation that did not necessarily require high spatial resolution. We refer the reader to our 2019 ISMRM abstract [90] for more details. Well-known trajectories that satisfy all properties above are the 2D golden-angle radial trajectory (2DGA) [91], and its 3D extension: the 3D golden-mean radial trajectory (GM3DR) [92]. Both of these golden-angle-based radial trajectories ensure approximately homogeneous sampling density in the angular directions. One downside regarding motion encoding, however, is the linear readouts that form the basis of these trajectories, since a linear readout only encodes motion in the readout direction. Although this direction constantly rotates with the golden angle over time, it would be more efficient to probe more than one direction within a readout as well. An extension of the GM3DR trajectory that does this is the 3D golden-mean cone (GM3DC) trajectory [68]. In essence this trajectory is similar to GM3DR, but the linear radial readouts are substituted with cones, or hybrid-radial readouts. The GM3DC combines the improved encoding of cones as opposed to purely radial, with

the golden-mean rotations of GM3DR to satisfy all properties specified above.

We considered the GM3DC trajectory for our preliminary work on real-time motion estimation in our 2020 ISMRM abstract [74]. In general a tremendous effort was put into making real-time reconstructions feasible with this trajectory. This included implementation on the pulse programming software on the scanner, processing with gradient impulse response functions (GIRF) to compensate for hardware imperfections [124, 147], and designing preconditioners for the reconstruction. Real-time reconstruction were possible for several volunteers, but unfortunately this trajectory turned out to be too unstable for our 1.5T Philips Igenia system. We expect this was mainly related to large eddy-currents due to fast gradient switching, and local inhomogeneity effects that are more pronounced with the long cone readouts. Eventually the framework would have to be applied to data acquired with an Elekta Unity MR-linac, which we expected to be even more susceptible to such effects due to its split gradient coil [7, 10]. We therefore decided to discontinue the experiments with the GM3DC trajectory for respiratory motion and fell back to the GM3DR trajectory. All data in Chapters 3-5 is acquired with the GM3DR trajectory.

Validation of reconstructed motion-fields

Validation is important for new technical methods, especially for medical applications. In this thesis we considered several validation strategies: 1) in silico with the digital XCAT phantom [106], 2) an MR-compatible motion phantom (QUASAR MRI 4D Motion Phantom, Modus QA, Ontario, Canada), 3) in vivo with image reconstructions, 4) in vivo with image reconstruction followed by Optical Flow [62], 5) in vivo with feet-head projection profiles.

Each strategy has advantages and disadvantages. The in silico validation does not consider real acquisition-related data corruption (e.g. eddy currents, flow effects), and can, in case of e.g. the XCAT phantom [106], yield unstable motion-fields [107]. MR-compatible motion phantoms, although useful for proof-of-principle validations, have limitations regarding the representation of realistic in vivo anatomies. Nevertheless, in silico and motion phantom validations allow for a precise assessment of the errors since ground-truth motion-fields are typically available. MR-image-based validations are limited in the sense that image reconstructions are required, which may not be feasible in 3D at frame-rates considered in this thesis. And, in general, the value of comparisons with other methods is questionable if these have not been thoroughly validated themselves. For example, the recently proposed novel framework of Extreme MRI [29] could enable 3D image reconstruction at a sufficiently high frame-rate, but the combination of this framework with image registration has only recently been investigated [148], and requires more thorough validation.

As the field of radiation therapy moves more and more towards real-time adaptive treatments, MR-based motion estimation methods and the validations thereof will become increasingly more important. The current MR-imaging speed may be insufficient to validate 3D motion-fields at the frame-rate of 5 Hz that is required to resolve respiratory motion [22]. Hence, external motion measurements will have to be incorporated in the validation and quality assurance. For example, optical tracking could be done with a stereo camera system [128, 149], in-bore camera system [150–

152], or time-of-flight camera system [153]. Other options include MR-compatible inertial sensors [154], fiber optic sensors [155, 156] or ultra-wideband electromagnetic sensors [157]. However, instead of just validating one motion estimation method with another, it would actually also be desirable for adaptive radiotherapy to use multiple motion estimation methods simultaneously. Each motion estimation strategy inevitably comes with its own uncertainties, and temporal resolution, but their agreement would be a good indication of the actual ground-truth motion. In fact, such strategies are already the de facto standard in the context of self-driving cars and robot navigation, where many simultaneous sensor measurements are fused to determine the 'ground-truth' environment [158, 159].

Uncertainty estimation

The topic of uncertainty estimation will also become increasingly more important for adaptive radiotherapy, especially because the motion estimates will be used to adapt a radiation plan; if these estimates contain errors, are we then actually better off by adjusting the plan? Unfortunately, uncertainty estimation is not a straightforward task. Chapter 5 briefly touched upon this subject with a probabilistic framework that estimates motion-fields, accompanied by a measure of its estimation uncertainty. In this framework, the estimation uncertainty is based on the distance between newly acquired data and in-distribution training data, and is by definition a measure of both motion model uncertainty (epistemic) and data-related uncertainty (aleatoric). The estimation uncertainty was shown to be strongly correlated with motion-field estimation errors in simulations. Such a measure could thus be used for real-time quality assurance: the radiation may temporarily be halted in case of high uncertainty, and may be continued when the confidence is restored. Chapter 5 also provided a way to generate a complete spatial uncertainty map, i.e. the motion-field's uncertainty as a function of spatial coordinates. However, the interpretation of this uncertainty is only reasonable during normal breathing. Nevertheless, in general similar uncertainty maps would most desirable for adaptive radiation therapy. As briefly explained in Section 1.2, a radiation plan takes geometrical uncertainties in the tumor's location into account via the PTV, and adaptive radiation therapy aims to reduce this PTV with beam-on motion estimation. Evidently, given a spatial uncertainty map corresponding to the total uncertainty in the tumor's position - including geometric uncertainties, and motion-field estimation uncertainties - the PTV can be reduced as adequately as possible. Hence, such a concept of total uncertainty could allow for the most efficient treatments.

In general it should be noted that any model will introduce uncertainty, however realistic the model may be. The definition of uncertainty is usually closely related to the actual ground-truth - given an uncertainty estimate, a confidence interval can be derived that will contain the ground-truth with high probability [160]. Without access to a ground-truth to refine your models - as is usually the case with motion-fields - a simple and general way to reduce any uncertainty is to make more observations or more estimates with different methods. As also mentioned above, this concept is already exploited in (semi) self-driving cars, where it is paramount that the locations of humans or barricades are estimated with high accuracy. In self-driving

cars this is achieved by using multiple independent measurements of these locations simultaneously, through the use of different sensors. Whenever many independent sensors agree on the location of an object, the uncertainty in this location is low. The same concept could be applied in radiotherapy, for example by using multiple motion estimation methods and/or positioning systems simultaneously. In this thesis we have developed two independent MR-based motion estimation methods (MR-MOTUS, and the probabilistic framework based on Gaussian Processes) that may be used simultaneously to improve motion estimation quality. Below we briefly describe how this could be done.

Application perspectives

Adaptive MR-guided radiation therapy

All methods presented in this thesis were designed with the application to real-time adaptive MRgRT in mind. We envision the following treatment workflow in which the presented techniques are applied.

In the pre-treatment phase, data will be acquired in about three minutes of free-breathing. This data will serve to reconstruct a pre-treatment MRI for day-to-day treatment plan adaptations and a reference image for MR-MOTUS, and to build a low-rank motion model. Optimally, all of this is performed with data acquired in just a single scan. However, MR-MOTUS relies on an acquisition with a non-Cartesian trajectory, which will likely not be able to provide the T_2 contrast which is desired for delineations on the pre-treatment MRI [24]. A possible extension that would enable the reconstruction of all three components described above from a single scan, would be to acquire all data with (rebound) Cartesian spiral readouts (rCASPR). This type of acquisition can be thought of as a hybrid between Cartesian and non-Cartesian acquisitions with variable sampling density in-plane [144], which favors MR-MOTUS reconstructions. Moreover, it enables free-breathing TSE acquisitions that allow for the acquisition of T_2 -weighted MR-images. Another benefit of this extension would be the possibility of improving the pre-treatment MRI with a motion-corrected image reconstruction [144, 161] using the MR-MOTUS motion-fields. Such a motion-corrected reconstruction increases the data-efficiency of the pre-treatment phase as much as possible by using all free-breathing data to reconstruct a single image. This could possibly result in pre-treatment MRIs with higher SNR and/or spatial resolution while requiring the same acquisition time. Alternatively to the rCASPR acquisitions, the framework could be extended to acquisitions with a 3D stack-of-star trajectory. This trajectory is closer to conventional Cartesian readouts, naturally provides a way for quality assurance through the inherently available 2D+t projection images, and can also be used to achieve the desired T_2 -weighted contrast when combined with TSE pulse sequences [162, 163]. Moreover, it has already successfully been applied for respiratory-resolved 3D+t image reconstructions [79, 164], which also increases the chances of success for 3D+t motion reconstructions. Both of these extensions to acquisitions with TSE pulse sequences however pose a technical challenge that is easily overlooked: modeling the T_2 signal decay of the echo trains. Currently this decay is not account for in the model, which assumes a steady state

of the transverse magnetization, and would thus require an extension along the lines of [165].

In the beam-on phase a combination of real-time MR-MOTUS (Chapter 4) and the GP framework (Chapter 5) will be performed. Data will be acquired with a GM3DR trajectory, interleaved at around 5 Hz with self-navigation spokes. Reconstructions could be performed that simultaneously minimize an MR-MOTUS data-fidelity term, and a regularization term based on the learned GP. The weight of the regularization can be taken inversely proportional to the GP's estimation uncertainty. This approach can be understood from a Bayesian perspective, where the likelihood is modeled with a Gaussian distribution, and the learned Gaussian Process prior is simultaneously placed on the temporal components of the motion-fields. The reconstruction described above is then equivalent with maximum likelihood estimation (MLE) of the posterior distribution, which is given as the product of the likelihood and the prior due to Bayes' rule. Any other independent measure of the motion may also be incorporated in a similar way. Eventually, the resulting motion-fields will be used to adapt the radiation plan in real-time, thereby dynamically changing the PTV according to the spatial uncertainties in the estimated motion-fields. Real-time quality assurance will also play an important role in this phase. Besides the uncertainty estimation as provided by the GP framework, visual feedback will also be important. This allows clinician to remain in control during the treatments, and manually terminate the treatment whenever required. Such visual feedback could be provided with real-time low-resolution images [166] that can be reconstructed with both the GM3DR acquisition and the rCASPR acquisition.

Post-treatment, time-resolved MR-MOTUS motion-fields will retrospectively be reconstructed over the course of the beam-on phase. Using these time-resolved motion-fields, and the real-time adapted radiation plan, the dose accumulated during the treatment can be computed [88, 120].

With the next fraction in mind, the time-resolved motion-fields could be used to further improve a static 3D MRI of the internal anatomy with motion-corrected image reconstruction [161]. This high-resolution MRI can serve as an initialization of the pre-treatment MRI reconstruction in the next fraction. Subsequently, the motion model may also be transferred to the next fraction by registering the pre-treatment MRI of both fractions. Furthermore, the radiation plan for the next fraction can be further optimized by taking into account the deposited dose, and all observed motion during the treatment.

Target tracking applications

Target tracking is another promising application for the GP framework presented in Chapter 5. The framework presented in Chapter 5 trained a GP to infer motion-field representation coefficients with respect to an a priori built motion model from few readouts of k -space data. The motion model was in that scenario required to upscale the low-dimensional output of the GP to a full high-dimensional motion-field. In case of target tracking, such a motion model would not be required, since the targeted output coordinates would already be low-dimensional. We have explored the feasibility of this approach in our most recent 2022 ISMRM abstract [167], wherein the

framework was applied to myocardial landmark tracking at more than 50 Hz. This type of high-speed tracking is paramount for stereotactic arrhythmia radio ablation [138] (STAR) with an MR-linac. STAR is an upcoming non-invasive treatment technique for ventricular tachycardia (VT) patients, where the idea is to apply highly focused radiation therapy to a fast-moving myocardial target that causes the arrhythmia. The radiation therapy scars the myocardial tissue, and should thereby reduce the chance of future arrhythmic events. The results in Huttinga *et al.* [167] were promising, especially because the presented method could handle severe signal artifacts caused by electronic cardiac implants that are typically present in VT patients.

Besides myocardial tracking, the framework in Huttinga *et al.* [167] could also be applied to catheter tracking during MR-guided interventions. Previously published methods in this context have proposed to track a fiducial marker attached to the catheter [168]. This marker shows up as a clear bright dot in the MR-image, and would therefore be an excellent target to track at high speed with our framework presented in Huttinga *et al.* [167].

Beyond radiotherapy: motion correction in PET/MR

Besides MR-guided radiation therapy, another application of the presented methods could be motion correction in PET/MRI scanners. Currently, a PET/MR scanner is being installed at the radiotherapy department of the UMC Utrecht. One of the most promising applications of such a PET/MR for radiotherapy is the detection of small tumors and lymph nodes spread all over the body in oligometastases [169]. When integrated in an MRgRT pipeline, the localized lesions could subsequently be radiated simultaneously on the MR-linac in a single fraction [170].

The simultaneous acquisition of PET and MR-data would furthermore allow to accurately localize small lesions subject to respiratory motion. The general idea is to perform a motion-corrected PET reconstructions, using MR-based motion-fields. The PET/MR currently being installed in our department has the same split gradient system as the MR-linac [7, 10], and the MR-system is therefore expected to perform similarly as the MR-linac. For respiratory motion correction, these MR-based motion-fields could therefore be estimated with low-rank MR-MOTUS. Exactly the same acquisition and reconstructions could be used as described for the MR-linac in Chapters 3 and 4.

Another promising application of the PET/MR is pediatric imaging. Pediatric cancer patients typically have to undergo several imaging sessions. Many sessions currently include PET/CT imaging, in which the CT deposits ionizing radiation that is particularly harmful for the developing bodies of the children. Moreover, this PET/CT sequentially acquires PET and CT images, and therefore does not allow to correct PET data using motion estimated from CT data. Therefore, a severe limitation of the PET/CT is the requirement that children need to lie still in the scanner. This is typically enforced by the administration of anesthesia, which both causes anxiety and could be harmful for the developing brains of the children. A PET/MR would allow for MR-scans, rather than CT scans, and thereby reduce the risk to the children. Moreover, the PET/MR will allow to track and correct for the

children's motion, with the simultaneously acquired MR-based motion-fields. This alleviates the restriction for the children to lie still in the scanner, allowing the use of light anesthetics instead of a full narcosis. The motion-fields required to do this could be acquired with MR-MOTUS, for example by using an affine motion model with temporally smoothly varying parameters, as was done in Huttinga *et al.* [90]. The result would be result in an affine transformation per dynamic, which allows for a simple retrospective motion correction by applying the inverse affine transformation to the simultaneously acquired PET and MR data.

Chapter 7

Appendices

Nederlandse samenvatting

Een MR-linac is de combinatie van een MRI met een radiotherapy behandelings-apparaat, ofwel LINAC. Het ultieme doel van de MR-linac is realtime adaptieve MR-gestuurde radiotherapie (AMRgRT), zodat ook bewegende tumoren nauwkeurig en effectief behandeld kunnen worden. Dat wil zeggen: het in realtime aanpassen van het bestralingsplan volgens realtime 3D-bewegingsschattingen. Een van de moeilijkste en belangrijkste technische obstakels om dit doel te bereiken, is de realtime 3D-bewegingsschatting. Dit proefschrift presenteert twee nieuwe grensverleggende benaderingen die dit mogelijk maken door slim gebruik te maken van MR-data. De belangrijkste methode heet MR-MOTUS en is het onderwerp van de Hoofdstukken 2-4. MR-MOTUS werd gemotiveerd door de observatie dat de interne beweging van organen een hoge mate van ruimte-tijdcorrelatie vertonen. Hierdoor zijn bewegingsvelden goed comprimeerbaar, of laagdimensionaal, en kunnen beschreven worden met relatief weinig parameters. Dat de bewegingsvelden waarin wij geïnteresseerd zijn beschreven kunnen worden met relatief weinig parameters, biedt mogelijkheden voor snelle reconstructies, omdat weinig parameters wellicht kunnen worden geschat op basis van een klein aantal snelle metingen. Een andere methode, maar gemotiveerd door dezelfde observatie, wordt geïntroduceerd in Hoofdstuk 5. In plaats van bewegingsvelden te reconstrueren via het MR-MOTUS-signaalmodel uit k -ruimte MRI metingen, werd een machine learning methode - gebaseerd op Gaussiaanse Processen (GP) - gebruikt om de reconstructies uit te voeren. Aangezien GPs realtime bewegingsschattingen mogelijk maken en ook een maatstaf voor schattingsonzekerheid bieden, kunnen we met deze methode ook een idee krijgen hoe betrouwbaar de geschatte beweging daadwerkelijk is. Dit is van cruciaal belang voor bijvoorbeeld realtime kwaliteitsborging tijdens radiotherapie behandelingen. Hieronder lichten we de inhoud van de hoofdstukken nader toe.

Hoofdstuk 2 demonstreert de proof-of-concept. Eerst werd het MR-MOTUS-signaalmodel afgeleid dat een bewegingsveld en een referentiebeeld expliciet relateert aan k -ruimte signalen. Het signaalmodel werd gevalideerd en een niet-lineair invers probleem met betrekking tot het signaalmodel werd geformuleerd om bewegingsvelden te reconstrueren uit k -ruimte signalen. Een reconstructiestrategie met een quasi-Newton-schema werd geschetst en de prestaties van het reconstructie-algoritme werden in silico geanalyseerd in vergelijking met de daadwerkelijke 'ground-truth' bewegingsvelden. De proof-of-concept werd gedemonstreerd door in vivo 3D rigide hoofdbeweging en in vivo 3D niet-rigide ademhalingsbeweging te reconstrueren uit retrospectief sterk onderbemonsterde (hierna: undersampled) k -ruimte signalen. Boven-

dien werd de praktische haalbaarheid van de aanpak aangetoond met een reconstructie van niet-rigide 2D bewegingsveldreconstructie van ademhaling vanuit prospectief undersampled data, verkregen met een 2D golden-angle acquisitie. Drie retrospectieve undersamplestrategieën werden onderzocht: lage resolutie 3D cartesiaans, 3D variabele dichtheid en 3D golden-mean cones [68]. Deze MR-MOTUS-reconstructies werden vergeleken met state-of-the-art beeldregistratie op beeldreconstructies verkregen uit dezelfde hoeveelheid undersampled data. De resultaten laten zien dat MR-MOTUS beter presteerde dan de beeldregistratie, met name bij hogere undersample factoren. Bovendien presteerden de 3D-variabele dichtheid en 3D golden-mean cones voor hoge undersamplefactoren beter dan de cartesiaanse undersampling.

De proof-of-concept methode gepresenteerd in Hoofdstuk 2 heeft een aantal verbeterpunten. Het vereiste bijvoorbeeld een adempauze om het referentiebeeld te reconstrueren, het signaalmodel ging uit van een single-coil-instelling en de haalbaarheid van 3D-reconstructies op prospectief undersampled data was nog niet aangetoond. Bovendien kon de methode nog geen gebruik maken van temporele correlatie in bewegingsvelden, omdat de reconstructie per tijdstip werd uitgevoerd in plaats van gelijktijdig over alle tijdstippen. Hoofdstuk 3 behandelt deze verbeterpunten in de proof-of-concept methode uit Hoofdstuk 2, en verkleint de kloof naar klinische toepassing. In het bijzonder werden zeer undersampled 3D k -ruimte signalen prospectief geacquireerd met behulp van een 3D golden-mean radial (3DGMR) schema, het referentiebeeld werd gereconstrueerd vanuit vrije ademhaling met behulp van respiratoire binning, en een nieuwe spoelcompressietechniek werd geïntroduceerd die het mogelijk maakte om meerkanaals acquisities te combineren met het enkelkanaals signaalmodel uit Hoofdstuk 2. Ten slotte werd de methode uitgebreid tot 3D+t ruimte-tijd bewegingsveldreconstructies door gebruik te maken van een bewegingsmodel van lage rang, dat per constructie bewegingsvelden modelleert in twee componenten: een ruimtelijke component en een temporele component. De ruimtelijke component modelleert de bewegingsvectoren met de meeste invloed. Per voxel coderen zij de hoek waarlangs het weefsel beweegt, en een relatieve bewegingsgrootte. De temporele component modelleert de globale schaling van de bewegingsvelden langs de vectoren in de ruimtelijke component. Het lage-rang model verminderde het aantal onbekenden voor ruimte-tijd bewegingsvelden met twee ordes van grootte. Dit maakte 3D+t motion-field-reconstructie mogelijk met een hoge temporele resolutie op een desktop-PC. De scheiding tussen ruimtelijke en temporele componenten maakte ook afzonderlijke regularisatiestrategieën in ruimte en tijd mogelijk. Om de flexibiliteit van de methode te demonstreren, werden reconstructies uitgevoerd op 2D+t en 3D+t abdominothoracale data en 3D+t hoofd-halsbewegingen. Validaties werden uitgevoerd met de determinant van de Jacobiaan en door middel van vergelijking van het vervormde referentiebeeld met beeldreconstructies.

In Hoofdstuk 4 werd een tweestaps-reconstructie voorgesteld voor low-rank MR-MOTUS die de realtime reconstructie op 6,7 Hz mogelijk maakte. Bovendien zijn verschillende verwerkingsstappen van de methode verbeterd, waardoor MR-MOTUS-reconstructies mogelijk zijn op prospectieve gegevens die zijn verkregen met de Elekta Unity MR-linac. De reconstructies maakten gebruik van de scheiding van ruimtelijke en temporele componenten door het low-rank bewegingsmodel geïntroduceerd in Hoofdstuk 3. In de eerste fase van de tweestaps-reconstructie wordt de ruimtelijke

component, waarvan werd aangenomen dat deze statisch in de tijd is, verkregen met een offline reconstructie van enkele minuten. In de tweede fase wordt de temporele component die de dynamiek in het bewegingsveld codeert per tijdstip gereconstrueerd in een online reconstructie. De belangrijkste gedachte achter deze benadering is dat de temporele component weinig vrijheidsgraden heeft (< 10), en daarom in realtime kan worden gereconstrueerd uit enkele k -ruimte signalen zodra de ruimtelijke component beschikbaar is. Realtime MR-MOTUS werd in silico gevalideerd op het digitale XCAT-fantoom en in vivo op data van vijf gezonde vrij-ademhalende vrijwilligers geacquireerd op de Elekta Unity MR-linac. De reconstructie werd gevalideerd met de determinanten van de Jacobiaan, door vergelijking met beeldreconstructie en door vergelijking met een surrogaatsignaal voor ademhaling. De resultaten geven de haalbaarheid aan van realtime 3D bewegingsveldreconstructies op een MR-linac met een totale wachttijd van 170 ms (67 ms data acquisitie, 103 ms reconstructietijd).

Een benadering in twee stappen, zoals voorgesteld in Hoofdstuk 4, is een natuurlijke manier om realtime reconstructie mogelijk te maken met een beperkte hoeveelheid MR-data. Om soortgelijke redenen werd dit ook eerder in de literatuur voorgesteld. Echter, een nadeel van deze benadering is dat deze aanneemt dat de beweging in de trainings- en inferentiefase vergelijkbaar is. Hoewel dit in de meeste gevallen een goede aanname is, kan dit in de praktijk minder realistisch zijn door abnormale bewegingen zoals hoesten of verplaatsingen van het volledige lichaam. Dit voorbeeld laat zien dat een mate van vertrouwen in realtime bewegingsschattingen gunstig zou zijn voor kwaliteitsborging tijdens realtime adaptieve MRgRT. In Hoofdstuk 5 hebben we daarom een probabilistische method voorgesteld voor de gelijktijdige realtime schatting van beweging en onzekerheid, gebaseerd op de machine learning-theorie van Gaussiaanse Processen (GP). In een voorbereidingsfase werd offline een low-rank bewegingsmodel gereconstrueerd en werd een GP getraind om een kansverdeling voor bewegingsveldrepresentatiecoëfficiënten af te leiden uit k -ruimte data. In de inferentiefase werd de getrainde GP gebruikt om de meest waarschijnlijke coëfficiënten af te leiden uit slechts 3 uitlezingen van k -ruimte data. Deze minimale hoeveelheid data in combinatie met de snelle reconstructie van de GP resulteerde in een reconstructiesnelheid van 69 3D-bewegingsvelden per seconde. Om bovendien de mogelijkheden voor kwaliteitsborging aan te tonen, werd een afwijzingscriterium op basis van de schattingsonzekerheid ontworpen om tijdstippen met mogelijk foutieve bewegingsschattingen te markeren. De voorgestelde strategie resulteerde in lage eindpuntfouten (75e percentielen $\leq 0,88$ mm) tijdens vier verschillende ademhalingspatronen in simulaties. Zonder het voorgestelde afwijzingscriterium zouden deze ademhalingspatronen hebben geleid tot eindpuntfouten van bijna 6 mm. Naast simulaties werd de robuustheid van deze bewegingsreconstructie door middel van MR-linac-acquisities van vijf gezonde vrijwilligers. Twee vrijwilligers werden gevraagd om abnormale bewegingen uit te voeren, die succesvol werden gedetecteerd door de methode. De detectie van mogelijk foutieve bewegingsschattingen, zoals gepresenteerd in dit hoofdstuk, zou een cruciale rol kunnen spelen om de patiëntveiligheid te waarborgen tijdens realtime adaptieve MRgRT.

References

- [1] H. Sung *et al.*, “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”, *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] M. K. Thompson *et al.*, “Practice-changing radiation therapy trials for the treatment of cancer: where are we 150 years after the birth of Marie Curie?”, *British journal of cancer*, vol. 119, no. 4, pp. 389–407, 2018.
- [3] R. Otazo *et al.*, “MRI-guided Radiation Therapy: An Emerging Paradigm in Adaptive Radiation Oncology”, *Radiology*, vol. 298, no. 2, pp. 248–260, 2021.
- [4] C. Ménard and U. A. van der Heide, “Introduction: Magnetic resonance imaging comes of age in radiation oncology”, in *Seminars in radiation oncology*, vol. 24, 2014, pp. 149–150.
- [5] H. Chandarana, H. Wang, R. Tijssen, and I. J. Das, “Emerging role of MRI in radiation therapy”, *Journal of Magnetic Resonance Imaging*, vol. 48, no. 6, pp. 1468–1478, 2018.
- [6] M. A. Schmidt and G. S. Payne, “Radiotherapy planning using MRI”, *Physics in Medicine & Biology*, vol. 60, no. 22, R323, 2015.
- [7] J. J. W. Lagendijk *et al.*, “MRI/linac integration”, *Radiotherapy and Oncology*, vol. 86, no. 1, pp. 25–29, Jan. 2008.
- [8] S. Mutic and J. F. Dempsey, “The ViewRay System: Magnetic Resonance-Guided and Controlled Radiotherapy”, *Seminars in Radiation Oncology*, vol. 24, no. 3, pp. 196–199, 2014.
- [9] P. J. Keall, M. Barton, and S. Crozier, “The Australian Magnetic Resonance Imaging-Linac Program”, *Seminars in Radiation Oncology*, vol. 24, no. 3, pp. 203–206, 2014.
- [10] B. W. Raaymakers *et al.*, “Integrating a 1.5 T MRI scanner with a 6 MV accelerator: Proof of concept”, *Physics in Medicine & Biology*, vol. 54, no. 12, N229–N237, May 2009.
- [11] D. M. de Muinck Keizer *et al.*, “Soft-tissue prostate intrafraction motion tracking in 3D cine-MR for MR-guided radiotherapy”, *Physics in Medicine & Biology*, vol. 64, no. 23, p. 235 008, Dec. 2019.
- [12] H. Holthusen, “Erfahrungen über die Verträglichkeitsgrenze für Röntgenstrahlen und deren Nutzenwendung zur Verhütung von Schäden”, *Strahlentherapie*, vol. 57, no. 254, 51a, 1936.
- [13] P. Mayles, A. Nahum, and J.-C. Rosenwald, *Handbook of radiotherapy physics: theory and practice*. CRC Press, 2007.
- [14] J. F. Ward, “DNA damage produced by ionizing radiation in mammalian cells: Identities, mechanisms of formation, and reparability”, *Progress in nucleic acid research and molecular biology*, vol. 35, pp. 95–125, 1988.
- [15] S. P. Jackson and J. Bartek, “The DNA-damage response in human biology and disease”, *Nature*, vol. 461, no. 7267, pp. 1071–1078, 2009.
- [16] G. Palma, S. Monti, M. Conson, R. Pacelli, and L. Cella, “Normal tissue complication probability (NTCP) models for modern radiation therapy”, in *Seminars in oncology*, Elsevier, vol. 46, 2019, pp. 210–218.

- [17] M. Van Herk, “Errors and margins in radiotherapy”, in *Seminars in radiation oncology*, Elsevier, vol. 14, 2004, pp. 52–64.
- [18] M. Carol *et al.*, “Initial clinical experience with the Peacock intensity modulation of a 3-D conformal radiation therapy system”, *Stereotactic and functional neurosurgery*, vol. 66, no. 1-3, pp. 30–34, 1996.
- [19] D. Verellen, M. De Ridder, N. Linthout, K. Tournel, G. Soete, and G. Storme, “Innovations in image-guided radiotherapy”, *Nature Reviews Cancer*, vol. 7, no. 12, pp. 949–960, 2007.
- [20] C. Kontaxis *et al.*, “Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac”, *Physics in Medicine & Biology*, vol. 62, no. 18, pp. 7233–7248, Aug. 2017.
- [21] P. Keall *et al.*, “The clinical implementation of respiratory-gated intensity-modulated radiotherapy”, *Medical Dosimetry*, vol. 31, no. 2, pp. 152–162, 2006.
- [22] P. J. Keall *et al.*, “The management of respiratory motion in radiation oncology report of AAPM Task Group 76 a”, *Medical physics*, vol. 33, no. 10, pp. 3874–3900, 2006.
- [23] M. J. Murphy, M. Isaakson, and J. Jalden, “Adaptive filtering to predict lung tumor motion during free breathing”, *CARS 2002 Computer Assisted Radiology and Surgery*, pp. 539–544, 2002.
- [24] B. Stemkens, E. S. Paulson, and R. H. Tijssen, “Nuts and bolts of 4D-MRI for radiotherapy”, *Physics in Medicine & Biology*, vol. 63, no. 21, 21TR01, 2018.
- [25] C. Paganelli *et al.*, “MRI-guidance for motion management in external beam radiotherapy: current status and future challenges”, *Physics in Medicine & Biology*, vol. 63, no. 22, 22TR03, 2018.
- [26] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: Sensitivity encoding for fast MRI”, *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [27] M. A. Griswold *et al.*, “Generalized autocalibrating partially parallel acquisitions (GRAPPA)”, *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [28] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI”, *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [29] F. Ong *et al.*, “Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions”, *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 1763–1780, 2020.
- [30] J. Yuan, O. L. Wong, Y. Zhou, K. Y. Chueng, and S. K. Yu, “A fast volumetric 4D-MRI with sub-second frame rate for abdominal motion monitoring and characterization in MRI-guided radiotherapy”, *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 7, p. 1303, 2019.
- [31] A. P. King, C. Buerger, C. Tsoumpas, P. K. Marsden, and T. Schaeffter, “Thoracic respiratory motion estimation from MRI using a statistical model and a 2-D image navigator”, *Medical Image Analysis*, vol. 16, no. 1, pp. 252–264, Jan. 2012.
- [32] T. Bjerre *et al.*, “Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes”, *Physics in Medicine & Biology*, vol. 58, no. 14, pp. 4943–4950, 2013.
- [33] B. Stemkens, R. H. N. Tijssen, B. D. de Senneville, J. J. W. Lagendijk, and C. A. T. van den Berg, “Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy”, *Physics in Medicine & Biology*, vol. 61, no. 14, p. 5335, 2016.
- [34] L. Feng, N. Tyagi, and R. Otazo, “MRSIGMA: Magnetic Resonance SIGNature MAtching for real-time volumetric imaging”, *Magnetic Resonance in Medicine*, vol. 84, no. 3, pp. 1280–1292, 2020.
- [35] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King, “Respiratory motion models: a review”, *Medical Image Analysis*, vol. 17, no. 1, pp. 19–42, 2013.

- [36] J. R. McClelland *et al.*, “A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images”, *Physics in Medicine & Biology*, vol. 62, no. 11, pp. 4273–4292, 2017.
- [37] B. Fischer and J. Modersitzki, “Curvature based image registration”, *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 81–85, 2003.
- [38] P. Blomgren and T. F. Chan, “Color TV: Total variation methods for restoration of vector-valued images”, *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, 1998.
- [39] Z.-P. Liang, “Spatiotemporal imaging with partially separable functions”, in *2007 4th IEEE international symposium on biomedical imaging: from nano to macro*, IEEE, 2007, pp. 988–991.
- [40] C. Zachiu, B. D. de Senneville, C. T. Moonen, B. W. Raaymakers, and M. Ries, “Anatomically plausible models and quality assurance criteria for online mono-and multi-modal medical image registration”, *Physics in Medicine & Biology*, vol. 63, no. 15, p. 155 016, 2018.
- [41] B. W. Raaymakers *et al.*, “First patients treated with a 1.5 T MRI-Linac: Clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment”, *Physics in Medicine & Biology*, vol. 62, no. 23, pp. L41–L50, 2017.
- [42] G. R. Lee, “Robust rigid-body motion estimation from extremely short subsets of 3D Cartesian scans”, in *26th annual meeting ISMRM, Paris, France*, 2018.
- [43] E. Tryggestad *et al.*, “4D tumor centroid tracking using orthogonal 2D dynamic MRI: Implications for radiotherapy planning”, *Medical physics*, vol. 40, no. 9, p. 091 712, 2013.
- [44] L. Brix, S. Ringgaard, T. S. Sørensen, and P. R. Poulsen, “Three-dimensional liver motion tracking using real-time two-dimensional MRI”, *Medical physics*, vol. 41, no. 4, p. 42 302, 2014.
- [45] M. Seregni *et al.*, “Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI”, *Physics in Medicine & Biology*, vol. 61, no. 2, p. 872, 2016.
- [46] L. I. Cervino, J. Du, and S. B. Jiang, “MRI-guided tumor tracking in lung cancer radiotherapy”, *Physics in Medicine & Biology*, vol. 56, no. 13, p. 3773, 2011.
- [47] N. White *et al.*, “PROMO: Real-time prospective motion correction in MRI using image-based tracking”, *Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 91–105, 2010.
- [48] M. Glitzner, B. D. de Senneville, J. J. W. Lagendijk, B. W. Raaymakers, and S. Crijs, “On-line 3D motion estimation using low resolution MRI”, *Physics in Medicine & Biology*, vol. 60, no. 16, N301, 2015.
- [49] M. Usman *et al.*, “Motion corrected compressed sensing for free-breathing dynamic cardiac MRI”, *Magnetic Resonance in Medicine*, vol. 70, no. 2, pp. 504–516, 2013.
- [50] R. R. Ingle *et al.*, “Nonrigid autofocus motion correction for coronary MR angiography with a 3D cones trajectory”, *Magnetic Resonance in Medicine*, vol. 72, no. 2, pp. 347–361, 2014.
- [51] A. Andreychenko *et al.*, “Thermal noise variance of a receive radiofrequency coil as a respiratory motion sensor”, *Magnetic Resonance in Medicine*, vol. 77, no. 1, pp. 221–228, 2017.
- [52] F. Odille, N. Cindea, D. Mandry, C. Pasquier, P.-A. Vuissoz, and J. Felblinger, “Generalized MRI reconstruction including elastic physiological motion and coil sensitivity encoding”, *Magnetic Resonance in Medicine*, vol. 59, no. 6, pp. 1401–1411, 2008.
- [53] F. Odille, P.-A. Vuissoz, P.-Y. Marie, and J. Felblinger, “Generalized reconstruction by inversion of coupled systems (GRICS) applied to free-breathing MRI”, *Magnetic Resonance in Medicine*, vol. 60, no. 1, pp. 146–157, 2008.
- [54] Z. W. Fu *et al.*, “Orbital navigator echoes for motion measurements in magnetic resonance imaging”, *Magnetic Resonance in Medicine*, vol. 34, no. 5, pp. 746–753, 1995.
- [55] A. J. Van der Kouwe, T. Benner, and A. M. Dale, “Real-time rigid body motion correction and shimming using cloverleaf navigators”, *Magnetic Resonance in Medicine*, vol. 56, no. 5, pp. 1019–1032, 2006.

- [56] E. B. Welch, A. Manduca, R. C. Grimm, H. A. Ward, and C. R. Jack Jr, “Spherical navigator echoes for full 3D rigid body motion measurement in MRI”, *Magnetic Resonance in Medicine*, vol. 47, no. 1, pp. 32–41, 2002.
- [57] J. G. Pipe, “Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging”, *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 963–969, 1999.
- [58] M. K. Stam *et al.*, “Navigators for motion detection during real-time MRI-guided radiotherapy”, *Physics in Medicine & Biology*, vol. 57, no. 21, p. 6797, 2012.
- [59] S. Wisetphanichkij and K. Dejhan, “Fast Fourier transform technique and affine transform estimation-based high precision image registration method”, *GESTS Int’l Trans. Computer Science and Engr*, vol. 20, no. 1, pp. 179–191, 2005.
- [60] C. Prieto, P. G. Batchelor, D. Hill, J. V. Hajnal, M. Guarini, and P. Irarrazaval, “Reconstruction of undersampled dynamic images by modeling the motion of object elements”, *Magnetic Resonance in Medicine*, vol. 57, no. 5, pp. 939–949, 2007.
- [61] C. Zachiu, B. D. de Senneville, C. Moonen, and M. Ries, “A framework for the correction of slow physiological drifts during MR-guided HIFU therapies: Proof of concept”, *Medical physics*, vol. 42, no. 7, pp. 4137–4148, 2015.
- [62] C. Zachiu, N. Papadakis, M. Ries, C. Moonen, and B. D. de Senneville, “An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs”, *Physics in Medicine & Biology*, vol. 60, no. 23, p. 9003, 2015.
- [63] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, “Non-rigid registration using free-form deformations: application to breast MR images”, *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [64] A. H. Barnett, J. F. Magland, and L. af Klinteberg, “A parallel non-uniform fast Fourier transform library based on an "exponential of semicircle" kernel”, *SIAM Journal on Scientific Computing*, vol. 41, no. 5, pp. C479–C504, 2019.
- [65] J.-Y. Lee and L. Greengard, “The type 3 nonuniform FFT and its applications”, *Journal of Computational Physics*, vol. 206, no. 1, pp. 1–5, 2005.
- [66] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization”, *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.
- [67] M. Chen, W. Lu, Q. Chen, K. J. Ruchala, and G. H. Olivera, “A simple fixed-point approach to invert a deformation field”, *Medical physics*, vol. 35, no. 1, pp. 81–88, 2008.
- [68] K. M. Johnson, “Hybrid radial-cones trajectory for accelerated MRI”, *Magnetic Resonance in Medicine*, vol. 77, no. 3, pp. 1068–1081, 2017.
- [69] K. T. Block *et al.*, “Towards routine clinical use of radial stack-of-stars 3D gradient-echo sequences for reducing motion sensitivity”, *Journal of the Korean Society of Magnetic Resonance in Medicine*, vol. 18, no. 2, pp. 87–106, 2014.
- [70] M. Uecker *et al.*, “Berkeley advanced reconstruction toolbox”, in *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 23, 2015, p. 2486.
- [71] N. Chauffert, P. Ciuciu, and P. Weiss, “Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies”, in *2013 IEEE 10th International Symposium on Biomedical Imaging*, IEEE, 2013, pp. 298–301.
- [72] A. Sbrizzi, **N. R. F. Huttinga**, and C. A. T. van den Berg, “Acquisition, reconstruction and uncertainty quantification of 3D non-rigid motion fields directly from k-space data at 100 Hz frame rate”, in *Proceedings 27th annual meeting ISMRM, Montréal, Canada*, 2019.
- [73] F. Lam, J. P. Haldar, and Z.-P. Liang, “Motion compensation for reference-constrained image reconstruction from limited data”, in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2011, pp. 73–76.
- [74] **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS”, in *Proceedings 28th annual meeting ISMRM, Sydney, Australia*, 2020, p. 0598.

- [75] K. Breuer *et al.*, “Stable and efficient retrospective 4D-MRI using non-uniformly distributed quasi-random numbers”, *Physics in Medicine & Biology*, vol. 63, no. 7, p. 075 002, 2018.
- [76] Z. Deng *et al.*, “Four-dimensional MRI using three-dimensional radial sampling with respiratory self-gating to characterize temporal phase-resolved respiratory motion in the abdomen”, *Magnetic Resonance in Medicine*, vol. 75, no. 4, pp. 1574–1585, 2016.
- [77] F. Han, Z. Zhou, M. Cao, Y. Yang, K. Sheng, and P. Hu, “Respiratory motion-resolved, self-gated 4D-MRI using rotating cartesian k-space (ROCK)”, *Medical physics*, vol. 44, no. 4, pp. 1359–1368, 2017.
- [78] J. Cai, Z. Chang, Z. Wang, W. Paul Segars, and F.-F. Yin, “Four-dimensional magnetic resonance imaging (4D-MRI) using image-based respiratory surrogate: A feasibility study”, *Medical physics*, vol. 38, no. 12, pp. 6384–6394, 2011.
- [79] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, “XD-GRASP: golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing”, *Magnetic Resonance in Medicine*, vol. 75, no. 2, pp. 775–788, 2016.
- [80] B. Stemkens *et al.*, “Optical flow analysis on undersampled radial acquisitions for real-time tracking of the pancreas in MR guided radiotherapy”, in *Proc Int Soc Magn Reson Med*, vol. 21, 2013, p. 4325.
- [81] S. Roujol, M. Ries, C. Moonen, and B. D. de Senneville, “Automatic nonrigid calibration of image registration for real time MR-guided HIFU ablations of mobile organs”, *IEEE Transactions on Medical Imaging*, vol. 30, no. 10, pp. 1737–1745, 2011.
- [82] M. Fu *et al.*, “High-frame-rate full-vocal-tract 3D dynamic speech imaging”, *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1619–1629, 2017.
- [83] M. Burdumy *et al.*, “One-second MRI of a three-dimensional vocal tract to measure dynamic articulator modifications”, *Journal of magnetic resonance imaging*, vol. 46, no. 1, pp. 94–101, 2017.
- [84] **N. R. F. Huttinga**, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, “MR-MOTUS: Model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal k-space data”, *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015 004, 2020.
- [85] Q. Zhang *et al.*, “A patient-specific respiratory model of anatomical motion for radiation treatment planning”, *Medical physics*, vol. 34, no. 12, pp. 4772–4781, 2007.
- [86] R. Li *et al.*, “On a PCA-based lung motion model”, *Physics in Medicine & Biology*, vol. 56, no. 18, pp. 6009–6030, 2011.
- [87] P. Mishra *et al.*, “An initial study on the estimation of time-varying volumetric treatment images and 3D tumor localization from single MV cine EPID images”, *Medical physics*, vol. 41, no. 8.1, p. 081 713, 2014.
- [88] W. Cai *et al.*, “3D delivered dose assessment using a 4DCT-based motion model”, *Medical physics*, vol. 42, no. 6.1, pp. 2897–2907, 2015.
- [89] D. A. Low *et al.*, “Novel breathing motion model for radiotherapy”, *International Journal of Radiation Oncology, Biology, Physics*, vol. 63, no. 3, pp. 921–929, 2005.
- [90] **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and P. R. Luijten, “Prospective 3D+t non-rigid motion estimation at high frame-rate from highly undersampled k-space data: Validation and preliminary in-vivo results”, in *Proceedings 27th annual meeting ISMRM, Montréal, Canada*, 2019, p. 1180.
- [91] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, “An optimal radial profile order based on the Golden Ratio for time-resolved MRI”, *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 68–76, 2006.
- [92] R. W. Chan, E. A. Ramsay, C. H. Cunningham, and D. B. Plewes, “Temporal stability of adaptive 3D radial MRI using multidimensional golden means”, *Magnetic Resonance in Medicine*, vol. 61, no. 2, pp. 354–363, 2009.

- [93] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs, “Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint”, *IEEE Transactions on Medical Imaging*, vol. 22, no. 6, pp. 730–741, 2003.
- [94] S. Becker, *L-BFGS-B, converted from Fortran to C, with Matlab wrapper*, <https://github.com/stephenbecker/L-BFGS-B-C>, 2019.
- [95] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, “Advances in sensitivity encoding with arbitrary k-space trajectories”, *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, 2001.
- [96] C. M. Rank *et al.*, “4D respiratory motion-compensated image reconstruction of free-breathing radial MR data with very high undersampling”, *Magnetic Resonance in Medicine*, vol. 77, no. 3, pp. 1170–1183, 2017.
- [97] W. Jiang *et al.*, “Motion robust high resolution 3D free-breathing pulmonary MRI using dynamic 3D image self-navigator”, *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 2954–2967, 2018.
- [98] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, “Accelerated dynamic MRI exploiting sparsity and low-rank structure: Kt SLR”, *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.
- [99] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z. P. Liang, “Image reconstruction from highly under sampled (k, t)-space data with joint partial separability and sparsity constraints”, *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1809–1820, 2012.
- [100] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, “Low rank matrix recovery for real-time cardiac MRI”, in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, 2010, pp. 996–999.
- [101] J. P. Haldar and Z. P. Liang, “Spatiotemporal imaging with partially separable functions: A matrix recovery approach”, in *2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2010 - Proceedings*, 2010, pp. 716–719.
- [102] V. Delmon, S. Rit, R. Pinho, and D. Sarrut, “Registration of sliding objects using direction dependent B-splines decomposition.”, *Physics in medicine and biology*, vol. 58, no. 5, pp. 1303–1314, 2013.
- [103] Y. Fu, S. Liu, H. H. Li, H. Li, and D. Yang, “An adaptive motion regularization technique to support sliding motion in deformable image registration:” *Medical Physics*, vol. 45, no. 2, pp. 735–747, 2018.
- [104] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, “The NMR phased array”, *Magnetic Resonance in Medicine*, vol. 16, no. 2, pp. 192–225, Nov. 1990.
- [105] P. Kellman and E. R. McVeigh, “Image reconstruction in SNR units: A general method for SNR measurement”, *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1439–1447, Dec. 2005.
- [106] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, “4D XCAT phantom for multimodality imaging research”, *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, Aug. 2010.
- [107] B. Eiben, J. Bertholet, M. J. Menten, S. Nill, U. Oelfke, and J. R. McClelland, “Consistent and invertible deformation vector fields for a breathing anthropomorphic phantom: a post-processing framework for the XCAT phantom”, *Physics in Medicine & Biology*, vol. 65, no. 16, p. 165 005, 2020.
- [108] J. G. Pipe and P. Menon, “Sampling density compensation in MRI: Rationale and an iterative numerical solution”, *Magnetic Resonance in Medicine*, vol. 41, no. 1, pp. 179–186, 1999.
- [109] D. Winkel *et al.*, “Adaptive radiotherapy: The Elekta Unity MR-linac concept”, *Clinical and Translational Radiation Oncology*, vol. 18, pp. 54–59, 2019.

- [110] A. M. Werensteijn-Honingh *et al.*, “Feasibility of stereotactic radiotherapy using a 1.5 T MR-linac: Multi-fraction treatment of pelvic lymph node oligometastases”, *Radiotherapy and Oncology*, vol. 134, pp. 50–54, 2019.
- [111] N. J. Mickevicius and E. S. Paulson, “Simultaneous orthogonal plane imaging”, *Magnetic Resonance in Medicine*, vol. 78, no. 5, pp. 1700–1710, 2017.
- [112] E. H. Tran *et al.*, “Evaluation of MRI-derived surrogate signals to model respiratory motion”, *Biomedical Physics and Engineering Express*, vol. 6, no. 4, p. 45 015, 2020.
- [113] C. Paganelli *et al.*, “Time-resolved volumetric MRI in MRI-guided radiotherapy: an in silico comparative analysis”, *Physics in Medicine & Biology*, vol. 64, no. 18, p. 185 013, 2019.
- [114] **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS”, *Magnetic Resonance in Medicine*, vol. 85, no. 4, pp. 2309–2326, 2020.
- [115] J. Pang *et al.*, “ECG and navigator-free four-dimensional whole-heart coronary MRA for simultaneous visualization of cardiac anatomy and function”, *Magnetic Resonance in Medicine*, vol. 72, no. 5, pp. 1208–1217, 2014.
- [116] T. Zhang, J. Y. Cheng, Y. Chen, D. G. Nishimura, J. M. Pauly, and S. S. Vasanawala, “Robust self-navigated body MRI using dense coil arrays”, *Magnetic Resonance in Medicine*, vol. 76, no. 1, pp. 197–205, 2016.
- [117] H. D. Heerkens *et al.*, “MRI-based tumor motion characterization and gating schemes for radiation therapy of pancreatic cancer”, *Radiotherapy and Oncology*, vol. 111, no. 2, pp. 252–257, 2014.
- [118] M. Von Siebenthal, G. Székely, U. Gamper, P. Boesiger, A. Lomax, and P. Cattin, “4D MR imaging of respiratory organ motion and its variability”, *Physics in Medicine & Biology*, vol. 52, no. 6, p. 1547, 2007.
- [119] T. Bruijnen, B. Stemkens, C. A. T. van den Berg, and R. H. N. Tijssen, “Prospective GIRF-based RF phase cycling to reduce eddy current-induced steady-state disruption in bSSFP imaging”, *Magnetic Resonance in Medicine*, vol. 84, no. 1, pp. 115–127, 2020.
- [120] C. Kontaxis *et al.*, “Delivered dose quantification in prostate radiotherapy using online 3D cine imaging and treatment log files on a combined 1.5T magnetic resonance imaging and linear accelerator system”, *Physics and Imaging in Radiation Oncology*, vol. 15, no. May, pp. 23–29, 2020.
- [121] P. T. Borman, C. Bos, B. Stemkens, C. T. Moonen, B. W. Raaymakers, and R. H. Tijssen, “Assessment of 3D motion modeling performance for dose accumulation mapping on the MR-linac by simultaneous multislice MRI”, *Physics in Medicine & Biology*, vol. 64, no. 9, 2019.
- [122] R. Li *et al.*, “Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy”, *Medical Physics*, vol. 37, no. 6, pp. 2822–2826, 2010.
- [123] R. Li *et al.*, “3D tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy”, *Medical Physics*, vol. 38, no. 5, pp. 2783–2794, 2011.
- [124] T. Bruijnen, B. Stemkens, J. J. W. Lagendijk, C. A. T. van den Berg, and R. H. N. Tijssen, “Gradient system characterization of a 1.5 T MRI-Linac with application to UTE imaging”, in *Proceedings 26th annual meeting ISMRM, Paris, France*, 2018, p. 235.
- [125] S. E. Zijlema *et al.*, “Design and feasibility of a flexible, on-body, high impedance coil receive array for a 1.5 T MR-linac”, *Physics in Medicine & Biology*, vol. 64, no. 18, 2019.
- [126] C. E. Rasmussen, “Gaussian processes in machine learning”, in *Summer school on machine learning*, Springer, 2003, pp. 63–71.
- [127] Z. Chen *et al.*, “Bayesian filtering: From Kalman filters to particle filters, and beyond”, *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

- [128] M. Hoogeman, J. B. Prévost, J. Nuyttens, J. Pöll, P. Levendag, and B. Heijmen, “Clinical Accuracy of the Respiratory Tumor Tracking System of the CyberKnife: Assessment by Analysis of Log Files”, *International Journal of Radiation Oncology, Biology, Physics*, vol. 74, no. 1, pp. 297–303, 2009.
- [129] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [130] **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time non-rigid 3D Respiratory motion estimation for MR-guided radiotherapy using MR-MOTUS”, *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 332–346, 2022.
- [131] **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Joint 3D motion-field and uncertainty estimation at 67Hz on an MR-LINAC”, in *Proceedings 29th annual meeting ISMRM, Virtual*, 2021, p. 0625.
- [132] B. D. de Senneville, C. Zachiu, M. Ries, and C. Moonen, “EVolution: an edge-based variational method for non-rigid multi-modal image registration”, *Physics in Medicine & Biology*, vol. 61, no. 20, p. 7377, 2016.
- [133] Y. Bengio, O. Delalleau, and N. Le Roux, “The curse of highly variable functions for local kernel machines”, *Advances in neural information processing systems*, vol. 18, p. 107, 2006.
- [134] E. V. Bonilla, K. Chai, and C. Williams, “Multi-task Gaussian Process Prediction”, *Advances in Neural Information Processing Systems*, vol. 20, pp. 153–160, 2007.
- [135] C. E. Rasmussen and H. Nickisch, “Gaussian processes for machine learning (GPML) toolbox”, *The Journal of Machine Learning Research*, vol. 11, pp. 3011–3015, 2010.
- [136] A. G. Anderson III, J. Velikina, W. Block, O. Wieben, and A. Samsonov, “Adaptive retrospective correction of motion artifacts in cranial MRI with multicoil three-dimensional radial acquisitions”, *Magnetic Resonance in Medicine*, vol. 69, no. 4, pp. 1094–1103, 2013.
- [137] E. M. Schrauben, J. M. Lim, D. S. Goolaub, D. Marini, M. Seed, and C. K. Macgowan, “Motion robust respiratory-resolved 3D radial flow MRI and its application in neonatal congenital heart disease”, *Magnetic Resonance in Medicine*, vol. 83, no. 2, pp. 535–548, 2020.
- [138] P. S. Cuculich *et al.*, “Noninvasive cardiac radiation for ablation of ventricular tachycardia”, *New England Journal of Medicine*, vol. 377, no. 24, pp. 2325–2336, 2017.
- [139] L. V. Romaguera, T. Mezheritsky, R. Mansour, J.-F. Carrier, and S. Kadoury, “Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy”, *Medical Image Analysis*, vol. 74, p. 102250, 2021.
- [140] M. L. Terpstra, M. Maspero, T. Bruijnen, J. J. Verhoeff, J. J. Lagendijk, and C. A. T. van den Berg, “Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks”, *Medical Physics*, 2021.
- [141] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, “Computer model calibration using high-dimensional output”, *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 570–583, 2008.
- [142] C. F. Baumgartner, C. Kolbitsch, J. R. McClelland, D. Rueckert, and A. P. King, “Autoadaptive motion modelling for MR-based respiratory motion estimation”, *Medical Image Analysis*, vol. 35, pp. 83–100, 2017.
- [143] M. Burger, H. Dirks, and C.-B. Schönlieb, “A Variational Model for Joint Motion Estimation and Image Reconstruction”, *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 94–128, 2016. arXiv: [arXiv:1607.03255v1](https://arxiv.org/abs/1607.03255v1).
- [144] T. Bruijnen *et al.*, “Free-breathing 3D T2-weighted TSE using Cartesian acquisition with rewinded spiral profile ordering (rCASPR) for abdominal radiotherapy”, in *Proceedings 28th annual meeting ISMRM, Sydney, Australia*, 2020.
- [145] P. T. Gurney, B. A. Hargreaves, and D. G. Nishimura, “Design and analysis of a practical 3D cones trajectory”, *Magnetic Resonance in Medicine*, vol. 55, no. 3, pp. 575–582, 2006.

- [146] P. Irarrazabal and D. G. Nishimura, “Fast three dimensional magnetic resonance imaging”, *Magnetic Resonance in Medicine*, vol. 33, no. 5, pp. 656–662, 1995.
- [147] S. J. Vannesjo *et al.*, “Gradient system characterization by impulse response measurements with a dynamic field camera”, *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 583–593, 2013.
- [148] Z. Miller, L. Torres, S. Fain, and K. Johnson, “Motion Compensated Extreme MRI: Multi-Scale Low Rank Reconstructions for Highly Accelerated 3D Dynamic Acquisitions (MoCo-MSLR)”, *arXiv preprint arXiv:2205.00131*, 2022.
- [149] M. Zaitsev, C. Dold, G. Sakas, J. Hennig, and O. Speck, “Magnetic resonance imaging of freely moving objects: Prospective real-time motion correction using an external optical motion tracking system”, *Neuroimage*, vol. 31, no. 3, pp. 1038–1050, 2006.
- [150] L. Qin, P. Gelderen, J. Zwart, F. Jin, Y. Tao, and J. Duyn, “Head movement correction for MRI with a single camera”, in *Proceedings 16th annual meeting ISMRM, Toronto, Canada*, 2008, p. 1467.
- [151] J. Maclaren *et al.*, “Measurement and correction of microscopic head motion during magnetic resonance imaging of the brain”, *PLoS one*, vol. 7, no. 11, e48088, 2012.
- [152] J. Schulz *et al.*, “An embedded optical tracking system for motion-corrected magnetic resonance imaging at 7T”, *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 25, no. 6, pp. 443–453, 2012.
- [153] G. Kudielka, A. Menini, P.-A. Vuissoz, J. Felblinger, and F. Wiesinger, “Technical Feasibility and Potential Applications of an Optical Time-Of-Flight Camera Mounted Inside the MR Scanner”, in *Proceedings 23rd annual meeting ISMRM, Toronto, Canada*, 2015, p. 0588.
- [154] B. Chen *et al.*, “Design and Validation of a Novel MR-Compatible Sensor for Respiratory Motion Modeling and Correction”, *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 123–133, 2017.
- [155] A. F. Silva, J. P. Carmo, P. M. Mendes, and J. H. Correia, “Simultaneous cardiac and respiratory frequency measurement based on a single fiber Bragg grating sensor”, *Measurement Science and Technology*, vol. 22, no. 7, p. 075 801, 2011.
- [156] M. Megens *et al.*, “Shape accuracy of fiber optic sensing for medical devices characterized in bench experiments”, *Medical Physics*, 2021.
- [157] F. Thiel, O. Kosch, and F. Seifert, “Ultra-wideband sensors for improved magnetic resonance imaging, cardiovascular monitoring and tumour diagnostics”, *Sensors*, vol. 10, no. 12, pp. 10 778–10 802, 2010.
- [158] F. Gustafsson, *Statistical sensor fusion*. Studentlitteratur, 2010.
- [159] M. Kam, X. Zhu, and P. Kalata, “Sensor fusion for mobile robot navigation”, *Proceedings of the IEEE*, vol. 85, no. 1, pp. 108–119, 1997.
- [160] A. N. Angelopoulos and S. Bates, “A gentle introduction to conformal prediction and distribution-free uncertainty quantification”, *arXiv:2107.07511*, 2021.
- [161] P. G. Batchelor, D. Atkinson, P. Irarrazabal, D. L. G. Hill, J. Hajnal, and D. Larkman, “Matrix description of general motion correction applied to multishot images”, *Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1273–1280, 2005.
- [162] M. B. Keerthivasan *et al.*, “An efficient 3D stack-of-stars turbo spin echo pulse sequence for simultaneous T2-weighted imaging and T2 mapping”, *Magnetic Resonance in Medicine*, vol. 82, no. 1, pp. 326–341, 2019.
- [163] M. B. Keerthivasan *et al.*, “Abdominal T2-Weighted Imaging and T2 Mapping Using a Variable Flip Angle Radial Turbo Spin-Echo Technique”, *Journal of Magnetic Resonance Imaging*, 2021.
- [164] L. Feng *et al.*, “Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI”, *Magnetic Resonance in Medicine*, vol. 72, no. 3, pp. 707–717, 2014.

- [165] K. T. Block, M. Uecker, and J. Frahm, “Model-based iterative reconstruction for radial fast spin-echo MRI”, *IEEE Transactions on Medical Imaging*, vol. 28, no. 11, pp. 1759–1769, 2009.
- [166] T. Bruijnen, P. T. S. Borman, T. Schakel, and B. Raaymakers, “Parallel imaging stream for multi-purpose real-time adaptive MRI-guided prostate radiotherapy”, *Physics in Medicine & Biology* (in submission), 2022.
- [167] **N. R. F. Huttinga**, O. Akdag, M. Fast, J. Verhoeff, A. Sbrizzi, and S. Mandija, “Real-time myocardial landmark-tracking for MRI-guided cardiac radio-ablation using Gaussian Processes”, in *Proceedings 30th annual meeting ISMRM, London, United Kingdom*, 2022.
- [168] T. Kuehne, R. Fahrig, and K. Butts, “Pair of resonant fiducial markers for localization of endovascular catheters at all catheter orientations”, *Journal of Magnetic Resonance Imaging*, vol. 17, no. 5, pp. 620–624, 2003.
- [169] C. Würslin *et al.*, “Respiratory motion correction in oncologic PET using T1-weighted MR imaging on a simultaneous whole-body PET/MR system”, *Journal of nuclear medicine*, vol. 54, no. 3, pp. 464–471, 2013.
- [170] D. A. Palma, G. S. Bauman, and G. B. Rodrigues, “Beyond oligometastases”, *International Journal of Radiation Oncology, Biology, Physics*, vol. 107, no. 2, pp. 253–256, 2020.

Publications & Honors

Journal publications

- 2023 **N. R. F. Huttinga** *et al.*, “Real-time myocardial landmark tracking for MRI-guided cardiac radio-ablation using Gaussian Processes”, *Physics in Medicine & Biology (in submission)*, 2023.
- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Gaussian Processes for real-time 3D motion and uncertainty estimation during MR-guided radiotherapy”, *Medical Image Analysis (in submission)*, *arXiv preprint arXiv:2204.09873*, 2023.
- G. Rizzuti, T. Schakel, **N. R. F. Huttinga**, J. W. Dankbaar, T. van Leeuwen, and A. Sbrizzi, “Towards retrospective motion correction and reconstruction for clinical 3D brain MRI protocols with a reference contrast”, *NMR in Biomedicine (in submission)*, *arXiv preprint arXiv:2301.01106*, 2023.
- 2022 **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time non-rigid 3D Respiratory motion estimation for MR-guided radiotherapy using MR-MOTUS”, *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 332–346, 2022.
- 2021 M. H. C. Van Riel, **N. R. F. Huttinga**, and A. Sbrizzi, “Spectro-Dynamic MRI: Characterizing Mechanical Systems on a Millisecond Scale”, *IEEE Access*, vol. 10, pp. 271–285, 2021.
- 2020 **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS”, *Magnetic Resonance in Medicine*, vol. 85, no. 4, pp. 2309–2326, 2020.
- N. R. F. Huttinga**, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, “MR-MOTUS: Model-based non-rigid motion estimation for MR-guided radiotherapy using a reference image and minimal k-space data”, *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015 004, 2020.
- 2019 S. Mandija, E. F. Meliadó, **N. R. F. Huttinga**, P. R. Luijten, and C. A. T. van den Berg, “Opening a new window on MR-based electrical properties tomography with deep learning”, *Nature Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

Conference publications

- 2023 **N. R. F. Huttinga**, S. Anand, C. A. T. van den Berg, A. Sbrizzi, and M. Lustig, “Three-dimensional rigid head motion correction using the Beat Pilot Tone and Gaussian Processes”, in *Proceedings 31st annual meeting ISMRM, Toronto, Canada*, 2023.
- T. E. Olausson, C. Beijst, A. Sbrizzi, C. A. T. van den Berg, and **N. R. F. Huttinga**, “Time-resolved cardiac imaging and motion analysis using a multi-scale dynamics decomposition”, in *Proceedings 31st annual meeting ISMRM, Toronto, Canada*, 2023.
- G. Rizzuti, T. Schakel, **N. R. F. Huttinga**, T. van Leeuwen, and A. Sbrizzi, “Retrospective motion correction and reconstruction for clinical 3D brain MRI protocols with areference contrast”, in *Proceedings 31st annual meeting ISMRM, Toronto, Canada*, 2023.
- 2022 **N. R. F. Huttinga**, O. Akdag, M. Fast, J. Verhoeff, A. Sbrizzi, and S. Mandija, “Real-time myocardial landmark-tracking for MRI-guided cardiac radio-ablation using Gaussian Processes”, in *Proceedings 30th annual meeting ISMRM, London, United Kingdom*, 2022.
- N. R. F. Huttinga**, T. Bruijnen, S. Anand, S. Mandija, M. Lustig, and A. Sbrizzi, “Gaussian Processes for real-time motion estimation, target tracking & quality assurance”, in *ISMRM Motion detection & Correction Workshop, Oxford, United Kingdom*, 2022.
- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time quality assurance for volumetric motion estimation during MR-guided radiotherapy”, in *Proceedings 30th annual meeting ISMRM, London, United Kingdom*, 2022, p. 2127.
- G. Rizzuti, **N. R. F. Huttinga**, A. Sbrizzi, and T. van Leeuwen, “Retrospective motion correction with structural priors for clinical MRI protocols”, in *Proceedings 30th annual meeting ISMRM, London, United Kingdom*, 2022, p. 1957.
- M. H. C. Van Riel, **N. R. F. Huttinga**, T. Bruijnen, and A. Sbrizzi, “Protocol Optimization of Spectro-Dynamic MRI”, in *Proceedings 30th annual meeting ISMRM, London, United Kingdom*, 2022, p. 1706.
- 2021 **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Robust real-time 3D motion estimation with MR-MOTUS on an MR-LINAC: a multi-subject validation study”, in *Proceedings 29th annual meeting ISMRM, Virtual*, 2021.
- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Robust real-time 3D motion estimation with MR-MOTUS on an MR-LINAC: a multi-subject validation study”, in *Proceedings ISMRM Benelux*, 2021.
- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Joint 3D motion-field and uncertainty estimation at 67Hz on an MR-LINAC”, in *Proceedings 29th annual meeting ISMRM, Virtual*, 2021, p. 0625.
- 2020 **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS”, in *Proceedings 28th annual meeting ISMRM, Sydney, Australia*, 2020, p. 0598.
- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and A. Sbrizzi, “Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS”, in *Proceedings ISMRM Benelux*, 2020.
- 2019 **N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and P. R. Luijten, “Prospective 3D+t non-rigid motion estimation at high frame-rate from highly undersampled k-space data: Validation and preliminary in-vivo results”, in *Proceedings 27th annual meeting ISMRM, Montréal, Canada*, 2019, p. 1180.

- N. R. F. Huttinga**, T. Bruijnen, C. A. T. van den Berg, and P. R. Luijten, “Prospective 3D+t non-rigid motion estimation at high frame-rate from highly undersampled k-space data: Validation and preliminary in-vivo results”, in *Proceedings ISMRM Benelux*, 2019.
- S. Mandija, E. F. Meliadó, **N. R. F. Huttinga**, P. R. Luijten, and C. A. T. van den Berg, “Deep Learning based MR Electrical Properties Tomography”, in *Proceedings 27th annual meeting ISMRM, Montréal, Canada*, 2019, p. 0326.
- A. Sbrizzi, **N. R. F. Huttinga**, and C. A. T. van den Berg, “Acquisition, reconstruction and uncertainty quantification of 3D non-rigid motion fields directly from k-space data at 100 Hz frame rate”, in *Proceedings 27th annual meeting ISMRM, Montréal, Canada*, 2019.
- 2018 **N. R. F. Huttinga**, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, “A dynamic MR-signal model to capture 3D motion-fields at ultra-high frame-rate”, in *Proceedings 26th annual meeting ISMRM, Paris, France*, 2018, p. 4091.
- N. R. F. Huttinga**, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, “A dynamic MR-signal model to capture 3D motion-fields at ultra-high frame-rate”, in *Proceedings ISMRM Benelux*, 2018.
- N. R. F. Huttinga**, C. A. T. van den Berg, P. R. Luijten, and A. Sbrizzi, “A dynamic MR-signal model to capture 3D motion-fields at ultra-high frame-rate”, in *Proceedings ImagO conference*, 2018.

Honors

- 2022 **Invited talk: ISMRM seminar Motion Detection & Correction.**
Virtual, first author. “Towards real-time 3D motion estimation with MRI: from MR-images to MR-MOTUS.”
- 2022 **Selected MRM highlight: special issue on Motion in MRI.**
MRM, first author. “Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k-space data using low-rank MR-MOTUS.”
- 2022 **Invited talk: Lung MRI seminar.**
UC Berkeley, virtual, first author. “Towards real-time 3D motion estimation with MRI: from MR-images to MR-MOTUS.”
- 2022 **Best oral presentation award.**
ISMRM 2022, MR in RT study group, first author. “Real-time myocardial landmark-tracking for MRI-guided cardiac radio-ablation using Gaussian Processes.”
- 2022 **Top 5% abstracts ISMRM: Summa Cum Laude merit award.**
ISMRM 2022, Power Pitch, first author. “Real-time myocardial landmark-tracking for MRI-guided cardiac radio-ablation using Gaussian Processes.”
- 2022 **Top 15% abstracts ISMRM: Magna Cum Laude merit award.**
ISMRM 2022, Poster, first author. “Real-time quality assurance for volumetric motion estimation during MR-guided radiotherapy.”
- 2022 **Personal Van Leersum travel grant.**
5-month research at UC Berkeley with Miki Lustig; 7000 Euro’s sponsored by Royal Dutch Academy of Sciences (KNAW). “Three-dimensional rigid head motion correction using the Beat Pilot Tone and Gaussian Processes.”
- 2021 **Third best oral presentation award.**
ISMRM 2021, Motion Correction study group, first author. “Joint 3D motion-field and uncertainty estimation at 67Hz on an MR-Linac.”

- 2021 **Top 15% abstracts ISMRM: Magna Cum Laude merit award.**
ISMRM 2021, Oral, first author. *“Robust real-time 3D motion estimation with MR-MOTUS on an MR-Linac: a multi-subject validation study.”*
- 2020 **Top 5% abstracts ISMRM: Summa Cum Laude merit award.**
ISMRM 2021, Oral, first author. *“Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS.”*
- 2020 **Selected ISMRM highlight by Berkin Bilgic.**
ISMRM 2020, first author. *“Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS.”*
- 2020 **Best power poster award.**
ISMRM Benelux 2020, power poster, first author. *“Real-time 3D respiratory motion estimation for MR-guided radiotherapy using low-rank MR-MOTUS.”*
- 2019 **Best oral presentation award.**
ISMRM 2019, MR in RT study group, first author. *“Prospective 3D+t non-rigid motion estimation at high frame-rate from highly undersampled k-space data.”*
- 2019 **Top 15% abstracts ISMRM: Magna Cum Laude merit award.**
ISMRM 2019, oral, first author. *“Prospective 3D+t non-rigid motion estimation at high frame-rate from highly undersampled k-space data.”*
- 2019 **Top 15% abstracts ISMRM: Magna Cum Laude merit award.**
ISMRM 2019, oral, third author. *“Deep Learning based MR Electrical Properties Tomography.”*

Acknowledgements

How it all started...

When I first arrived at the UMC I was welcomed by one Italian and one semi-Italian: Alessandro and Nico. The first few minutes were filled with Alessandro's jokes and Nico's laughter; I had no idea what to expect. Later, Alessandro enthusiastically convinced me of his ideas for a new '3D MRI Action Camera' for motion estimation with MRI in radiotherapy. Through many mathematical formulas on the whiteboard this approach became clear, and from that moment on it was clear what my goal would be for the coming years: real-time motion estimation for MR-guided radiotherapy with the MR-linac. To reflect on this journey, this has also become the title of my thesis. Needless to say, this thesis would not have been what it currently is without the help of several key persons, that I would therefore like to thank.

Mentors

Alessandro, your enthusiasm, positivity, and problem-solving skills have continued to inspire me ever since day one. I was always looking forward to our weekly meetings. And even besides the meetings, there was always time for a chat, a stupid joke, or a problem to solve. In times of doubt, these meetings really motivated me to continue. In particular, when everything already worked in theory, but nothing worked in practice on the scanner, your positivity and down-to-earth attitude helped a lot. Fortunately, things eventually also worked out in practice, so the first paper could be published. One problem: the name 3D MRI Action Camera. Some 20 acronyms and contradictory opinions between you, me, and Nico later, MR-MOTUS was born. I also like how you could be critical at times; no boring low-hanging fruits, let's do something new. Eventually we managed to do 'something new', and demonstrated real-time motion estimation with MR-MOTUS. Without any doubt, everything presented in this thesis would not have been possible without you. I could not have imagined a better daily supervisor. Thank you!

Nico, if I have learned one thing from you the past years, it is to never stop dreaming. You always come up with the craziest ideas, usually late in the afternoon just before you go home. I remember countless occasions of you storming into the office with the question: 'Niek, can we do this??' Mostly the answer was yes, so your next grant could be written. Lately our discussion subjects frequently evolved

into kernels and Gaussian Processes, which still seem to haunt your dreams till this day. Regarding MR-MOTUS, you were always the person to put the mathematical theories into a practical, clinical perspective. This has helped me a lot in learning to comprehensibly explain theoretical concepts. This has resulted in several presentation awards to which you have contributed very significantly. At the conferences, you were also always present at the parties till late in the night. I remember commenting on your dance moves once, and your reply was: ‘Ha, can’t fit that with your B-splines right?!’ Thank you!

PhD committee

PhD committee members, and specifically the reading committee Martin Fast, Miki Lustig, Jan Legendijk, Bas Raaymakers and Rob Bisseling: thank you for your time and effort going through this immense manuscript. In particular, Jan and Bas, thank you for your inspirational vision regarding the MR-linac and its ultimate future, and your positivity regarding MR-MOTUS; this motivated a lot!

Colleagues & friends

Tom, it is safe to say that without your persistence, effort, and knowledge, it would not have been possible to achieve the results presented in this thesis: you have enabled the translation of MR-MOTUS into practice. For me this was the most difficult aspect of the research, and a hard period during my PhD. We have spent countless nights over at least a year trying to debug radial, cone and spiral trajectories, pondering what was wrong with the scanner. Unfortunately, the best theoretical solutions (cones) did not work at all in practice, and we were forced to abandon that idea in view of time. Nevertheless, we managed to get something else to work in the end. I look back on a very nice collaboration: thank you!

Stefano, you have been my fellow office mate throughout my entire PhD. I know you as a very persistent and patient person, who drinks too much chocolate milk and walks around the hospital on fluffy white slippers, and in hospital trousers when it’s been raining. You have provided me with good advice regarding almost anything related to research and life. I’m happy to have also collaborated with you on several occasions. Especially after the pandemic, when my motivation was low, our late-night abstract and team effort really helped me to get my motivation back up again. Thank you!

To all my other office mates throughout the years, Anna, Hanna, Mike, Flavio, Soraya, Jordi and Maarten: thank you for all the conversations, discussions, humor and collaborations. Flavio, you were always the most silent person with the best ideas. I look back on good English lessons, a good collaboration and good laughs. Hanna, good luck on finishing your thesis; I’m sure you’ll eventually solve all the problems that you currently ‘don’t know’ how to solve, just like all the other solved problems you ‘didn’t know’ how to solve. Maarten, thank you for your humor in the office, great ideas, listening ears, and the great trip to London together.

To my other, not yet mentioned, collaborators, Martin, Osman, Gabrio, Peter Luijten, and Max: thank you for our great collaborations, and thank you for taking my always critical eye with a grain of salt sometimes. Luis and Zach, thank you for the collaborations across the ocean. Continue the good work!

Arjan, Janot, Oscar, Cyril, Tim, Peter, Maarten, Robin, Stefan, Tom, Mike, thank you for the occasional fussball matches; they were a great distraction!

PET/MR team, Casper, Rodrigo, Rein and Thomas, thank you for your hard work on the PET/MR project, I'm sure this team can make it a success. Thomas, I'm happy you're continuing with MR-MOTUS, and I'm confident you'll make 3D time-resolved cardiac motion estimation work one day! Continue the good work!

To all my other colleagues over the past years: there's too many to name you all, but thank you for the coffees, great atmosphere at the conferences, Basket drinks, sailing trips, graduation parties, indoor soccer matches, and pub quizzes. Tim, thank you for the laughs, coffees, drinks and parties, and advice regarding anything MRI-related. Oscar, thank you for your patient explanations of and discussions on reconstruction algorithms, for sharing my opinion that the UMC coffee is the worst, the cappuccino moments, hooking me up with a Gaggia Classic and coffee grinder (they still work!), and teaching me how to make latte art. Bart, thank you for the drinks, great company in Paris when the plane got cancelled, conference parties, and interesting discussions on research. Lisa, Anneloes, Erwin and Janot, thank you for the drinks and parties in the early PhD days. Arjan, thank you for being a reliable factor in any party; you were always there.

MikGroupers & others at UC Berkeley, Miki, Suma, Ekin, Tanya, Yuhan, Victor, Julian, Engin, Paul, Gopal, Alan, Efrat, Katie, Shreya, Alfredo, Jingjia and Ke: thank you for the great times in Berkeley, you've helped to make this a wonderful experience! Miki, thank you for giving me this opportunity, I've learned a lot from you and your group; it's been a great experience. Additionally, thanks for joining my PhD committee and the time and effort to go through the thesis. Suma, thanks for the great collaboration, practical help with any problems I had to get set up, your help and patience with all the scan session we did, and inviting me to all kinds of social activities. Efrat, thank you for the great discussions, great company, and helping me out with all my questions in the beginning and right before I arrived in Berkeley. Gopal, Engin and Paul, thanks for the occasional drinks, good discussions, city trips and mountain biking!

Vrienden

Joanan, Normen, Niels, Joris, en vrouwelijke aanhang. Bedankt voor de afleiding tijdens de vele (Kungar) avonden in Utrecht, een luisterend oor, maar ook zeker de kritische blik. Ook al zagen we elkaar in Enschede al veel, ik denk dat we door deze avonden in Utrecht nog een stuk beter hebben leren kennen. Joanan, in het bijzonder bedankt voor het maken van de cover van dit proefschrift. Ik heb overwogen het zelf te proberen, maar dan was het nooit zo mooi geworden als nu. Daarnaast waardeer ik de onvoorwaardelijke band die we de laatste jaren hebben opgebouwd heel erg, en ik heb er geen twijfel over dat dat de rest van ons leven zo blijft. Heel erg bedankt!

Normen, door jou ben ik een PhD gaan doen in het UMC Utrecht; jij sloeg deze baan af, kreeg een andere baan waar ik je heb aangedragen, en ik kreeg even later de baan bij het UMC. Bedankt voor de vele avonden slap geouwehoer, samenwonen, ontelbaar vaak eten van wraps, festivals, en vakanties. Ik kan nog wel even door gaan maar eigenlijk was je bijna bij alles in mijn leven betrokken, en het kwam dan ook als een klap dat dat plotseling anders was. Inmiddels ben je voorzichtig aan de beterende hand en ik kijk er naar uit alle activiteiten met je weer op te pakken. Niels, bedankt voor je betrouwbaarheid, je attentheid, het altijd beschikbaar zijn als er écht problemen zijn, de vakantie in Colombia, en voor de menige festivals. Joris, bedankt voor je positiviteit, luisterende oor, slap geouwehoer en grappen, en het (terecht) corrigeren van mijn gedrag door middel van het aanreiken van menig beerbong.

Camiel, Siemen, Wout en Leendert: bedankt voor de vele feestjes en festivals! Camiel, bedankt voor de mooie tijd in Colombia, en Siemen, bedankt voor je betrokkenheid en attentheid, en dat je me continu blijft verbazen.

Thoridor, en in het bijzonder Rob, Ruben, Sjoerd, Olaf, Kevin, Rogier, Guus, Ruben, Vincent, Cor, Harm en Erwin: bedankt voor de mooie studententijd, weekendjes weg, etentjes, feestjes en festivals. Olaf, bedankt voor de mooie tijd aan de Livingstonelaan. Kevin, bedankt voor alle interessante discussies op het gebied van onderzoek, startups en AI. Erwin, bedankt voor discussies omtrent onder andere de toekomst. BT!

Vrienden uit Keijenborg e.o.: bedankt voor de nodige biertjes, weekendjes weg en slap geouwehoer!

Sterrenteam Zwaluwen 3. Bedankt voor de nodige afleiding in de weekenden, trainingen op donderdag (drie keer raken derde keer spelen), teamweekenden, TD's, gezeik tegen de scheidsrechters (terecht), eigen doelpunten en biljetjes! Volgend seizoen worden we écht kampioen!

En last but not least: mijn paranimfen Peter en Sander. Dan is de driehoek eindelijk rond en zijn we alle drie gepromoveerd (bijna dan). Ik wil jullie bedanken voor alle extra moeite en tijd die jullie er in hebben gestoken om hier een onvergetelijke dag van te maken. Daarnaast wil ik jullie bedanken voor de mooie avonden met spelletjes en discussies, fietstochten, feestjes, zaalvoetbal, conferenties, en tafelvoetbal en Footy potjes. Ik kijk ik uit naar de vele whisky avonden (en fietstochten) die ongetwijfeld zullen volgen! Heel erg bedankt!

Familie

Papa en mama: ik wil jullie bedanken voor jullie onvoorwaardelijke steun gedurende mijn leven en in het bijzonder de PhD. Het ging met ups en downs maar jullie stonden altijd voor mij klaar en dat waardeer ik heel erg. Papa, er is veel gebeurd de afgelopen jaren, maar ik ben enorm trots en blij dat we deze mijlpaal samen mogen meemaken. Mama, je bent de sterkste persoon die ik ken en ik neem in veel dingen nog altijd een voorbeeld aan je. Bedankt voor alles!

Stef, Nienke, Vincent, Marieke, Cato en Steijn: bedankt voor jullie steun en afleiding in het proces. Stef, bedankt voor al je hulp aan papa en mama in Keijenborg, en het nodige goede advies aan mij. Als kleine broer geef jij mij uiteindelijk vaker

advies dan andersom. Vincent en Marieke, het is fijn jullie de afgelopen 10 jaren beter te kunnen hebben leren kennen: bedankt voor de borrels, vakantie in California en Frankrijk, gezelligheid en afleiding.

Rene, Lida, Kelly en Zayn: bedankt voor de gezelligheid, etentjes en verjaardagen.

En dan de belangrijkste, Tara. Na een aantal persoonlijke tegenslagen en een globale pandemie, was er ineens toch een lichtpuntje: tussen de lockdowns door leerde ik jou kennen. Het is moeilijk onder woorden te brengen hoe onmisbaar en belangrijk jij voor mij bent geworden in een relatief korte tijd. We leerde elkaar richting het einde van mijn PhD kennen, maar dat was achteraf verreweg het zwaarste gedeelte. Ik heb erg veel aan je gehad in deze tijd, en je bent ook een van de weinige die weet hoe lastig het soms was. Jouw positiviteit, humor, gekheid, zorgzaamheid en onvoorwaardelijke steun hebben mij zeker door deze tijd heen gesleept, en wisten me vaak op andere gedachten te brengen als dat nodig was. Ik ben blij dat ik me tijdens de lockdowns toch maar eens aan datingapps gewaagd heb, en ik ben enorm dankbaar dat ik jou in mijn leven heb. We hebben al veel uitdagingen overwonnen samen en ik kijk heel erg uit naar onze toekomst samen. Dankjewel voor alles!!

Curriculum vitae

I was born on the 18th of Februari, 1993, in Arnhem, the Netherlands. I completed high school at the Sint Ludger College in Doetinchem, with a specialization in Nature, Technology and Computer Science.

After high school, I started the Bachelor Applied Mathematics at the University of Twente in 2011. I graduated in 2014 (cum-laude) and continued with the Master Applied Mathematics at the University of Twente, with a specialization in Mathematical Physics.

As part of my Master, I completed an internship in 2016 at the University of Waterloo, Canada, supervised by Sander Rhebergen, in which I implemented a new Finite Element Galerkin method into an existing C++ numerical PDE solver. Later that year I started my graduation project under supervision of Christoph Brune, in which I researched the relations between variational methods for inverse problems and machine and deep learning methods that were emerging at that time, and implemented a deep learning circulating tumor cell classification network. I graduated in 2017 (GPA 8/10).

In June 2017, I started my PhD in Applied Mathematics at the University Medical Center in Utrecht, under supervision of Alessandro Sbrizzi and Nico van den Berg. You have the result of this thesis in front you at this very moment.

During the final year of my PhD I was fortunate to receive a personal travel grant from the Royal Dutch Academy of Sciences (KNAW) for a 5-month research project at the University of California, Berkeley. During this project I was supervised by Miki Lustig and investigated the feasibility of three-dimensional rigid head motion correction using the Beat Pilot Tone and Gaussian Processes.

