

PAPER • **OPEN ACCESS**

# A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images

To cite this article: Jamie R McClelland *et al* 2017 *Phys. Med. Biol.* **62** 4273

View the [article online](#) for updates and enhancements.

## You may also like

- [Transthoracic echocardiography: from guidelines for humans to cardiac ultrasound of the heart in rats](#)  
Dorota Sztachman, Tymoteusz era, Katarzyna Czarzasta et al.
- [Crucial role of an aerophobic substrate in bubble-propelled nanomotor aggregation](#)  
Tao Wang, Meifen Zheng, Lina Wang et al.
- [Impact-induced acceleration by obstacles](#)  
N A Corbin, J A Hanna, W R Royston et al.

**JOIN US | ESTRO 2024**

**In-Booth Talks, Demos,  
& Lunch Symposium**

[Browse talk schedule >](#)



# A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images

Jamie R McClelland<sup>1</sup>, Marc Modat<sup>2</sup>, Simon Arridge<sup>1</sup>,  
Helen Grimes<sup>3</sup>, Derek D'Souza<sup>3</sup>, David Thomas<sup>4</sup>,  
Dylan O'Connell<sup>5</sup>, Daniel A Low<sup>5</sup>, Evangelia Kaza<sup>6</sup>,  
David J Collins<sup>6</sup>, Martin O Leach<sup>6</sup> and David J Hawkes<sup>1</sup>

<sup>1</sup> Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London, WC1E 6BT, United Kingdom

<sup>2</sup> Translational Imaging Group, Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, London, WC1E 6BT, United Kingdom

<sup>3</sup> Radiotherapy Physics Department, University College London Hospitals NHS FT, Euston Road, London, NW1 2PG, United Kingdom

<sup>4</sup> Department of Radiation Oncology, University of Colorado School of Medicine, 1665 Aurora Court, Suite 1032 MS F706—Aurora, CO 80045, United States of America

<sup>5</sup> Department of Radiation Oncology, University of California Los Angeles, 200 Medical Plaza Way, Suite B265, Los Angeles, CA 90095, United States of America

<sup>6</sup> CRUK Cancer Imaging Centre, Institute of Cancer Research and Royal Marsden Hospital, 123 Old Brompton Road, London, SW7 3RP, United Kingdom

E-mail: [j.mcclelland@ucl.ac.uk](mailto:j.mcclelland@ucl.ac.uk)

Received 22 August 2016, revised 10 January 2017

Accepted for publication 14 February 2017

Published 5 May 2017



## Abstract

Surrogate-driven respiratory motion models relate the motion of the internal anatomy to easily acquired respiratory surrogate signals, such as the motion of the skin surface. They are usually built by first using image registration to determine the motion from a number of dynamic images, and then fitting a correspondence model relating the motion to the surrogate signals. In this paper we present a generalized framework that unifies the image registration



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

and correspondence model fitting into a single optimization. This allows the use of ‘partial’ imaging data, such as individual slices, projections, or  $k$ -space data, where it would not be possible to determine the motion from an individual frame of data. Motion compensated image reconstruction can also be incorporated using an iterative approach, so that both the motion and a motion-free image can be estimated from the partial image data. The framework has been applied to real 4DCT, Cine CT, multi-slice CT, and multi-slice MR data, as well as simulated datasets from a computer phantom. This includes the use of a super-resolution reconstruction method for the multi-slice MR data. Good results were obtained for all datasets, including quantitative results for the 4DCT and phantom datasets where the ground truth motion was known or could be estimated.

**Keywords:** respiratory motion modelling, image registration, respiratory surrogate signals, motion compensated image reconstruction, super-resolution, CT, MR

 Supplementary material for this article is available [online](#)

(Some figures may appear in colour only in the online journal)

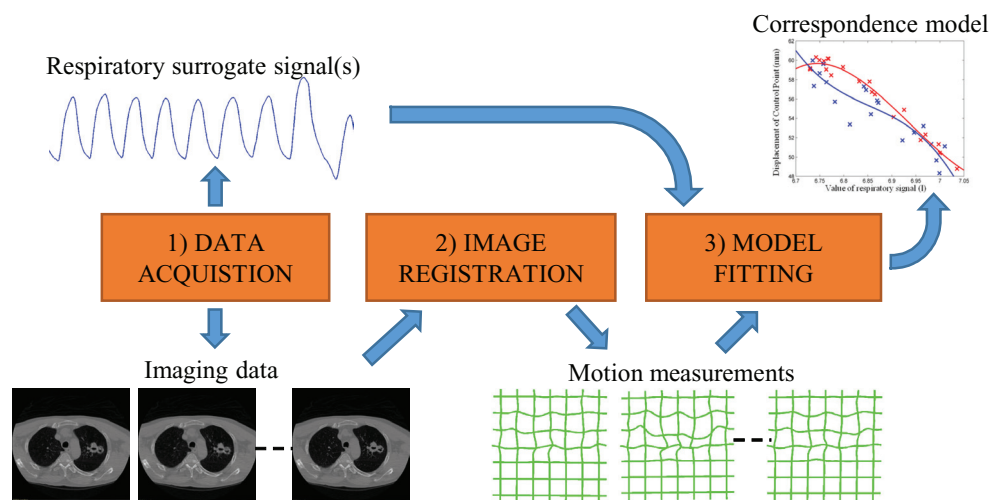
## 1. Introduction

Respiratory motion is often a problem when acquiring images and planning and guiding interventions (e.g. radiotherapy) in the abdomen and thorax. It causes artefacts in reconstructed images, limiting their utility, and can cause misalignment between the intervention plan and the moving anatomy, limiting accuracy and leading to uncertainties in the delivered treatment. If the motion is known during the acquisition/intervention then it can be corrected for, e.g. using Motion Compensated Image Reconstruction, MCIR (Batchelor *et al* 2005, Rit *et al* 2009), or using gating or tracking when delivering treatment (Keall *et al* 2006). However, it is usually very difficult to directly measure the full motion during the acquisition/intervention.

One solution, which has been proposed for a wide range of applications, is using surrogate-driven respiratory motion models (McClelland *et al* 2013). These relate the internal motion to one or more respiratory surrogate signals, such as the displacement of the skin surface (McClelland *et al* 2011) or spirometry (Low *et al* 2005), which can be easily measured during the acquisition/intervention. A correspondence model describes the mathematical relationship between the surrogate signals and the internal motion, and is fitted using data acquired before the start of the acquisition/intervention. The model can then be used to estimate the internal motion from the surrogate signals during the acquisition/intervention.

Figure 1 illustrates how a motion model is typically built. Respiratory surrogate signals are simultaneously acquired with imaging data that capture the motion of interest. Then, image registration or a similar technique is used to determine the motion from the imaging data. Finally, a correspondence model is fitted relating the motion to the surrogate signals. This process is described in detail in McClelland *et al* (2013), along with the methods that have been used for each stage in the literature.

One problem with this approach, that has so far limited its practical applicability, is the difficulty in acquiring suitable imaging data for determining the full motion of interest. For most applications it is desirable to know the 3D motion of the anatomy, thus requiring 4D data, i.e. 3D + time. Many 4D imaging methods have been presented in the



**Figure 1.** Typical approach for building a respiratory motion model. (1) Respiratory surrogate signals are acquired simultaneously with imaging data. (2) Image registration is used to determine the motion from the imaging data. (3) A correspondence model is fitted relating the motion to the surrogate signals.

literature, including 4D-CT (Pan 2005), 4D-CBCT (Sonke *et al* 2005), and 4D-MR methods (von Siebenthal *et al* 2007, Yang *et al* 2015). However, most of these are respiratory-sorted (sometimes referred to as respiratory-correlated) methods. Due to scanner technology limitations it is impossible to acquire all the data required to form a full 3D volume fast enough to image the respiratory motion. Therefore, individual frames of ‘partial image data’ are acquired over several respiratory cycles, and are then sorted using either surrogate signals or image similarity measures (Paganelli *et al* 2015) to form ‘coherent’ volumes. The partial image data could be small slabs consisting of a few slices (Pan 2005), individual slices (von Siebenthal *et al* 2007), projection data (Sonke *et al* 2005), or lines of  $k$ -space (Rank *et al* 2016). Most of these methods assume reproducible motion during every respiratory cycle, so cannot be used to study or model breath-to-breath variations, and often contain artefacts caused by these variations. Some 4D-MR methods try to image breath-to-breath variability by acquiring each frame of data over many respiratory cycles (von Siebenthal *et al* 2007). However, these methods require long acquisition times, and as the images are formed from data acquired during different breath cycles they may not give a good representation of the true motion and its variability.

Furthermore, the different frames used to form the respiratory-sorted volumes often do not have exactly the same measured surrogate signal values. Such volumes are not well suited for building motion models as they will not have unique surrogate signal values for each volume. Some researchers have derived surrogate signals from the volumes (Fayad *et al* 2011), but then the motion and surrogate signals may be determined from different parts of the image data acquired during different respiratory cycles, and may not reflect the true relationship between surrogate and motion.

‘True’ 4D-MR methods, which acquire a full volume fast enough to image the respiratory motion, have also been proposed (Yang *et al* 2015). However, using current technology these methods suffer from poor image quality due to the trade-offs between spatial resolution, spatial coverage, and temporal resolution, limiting their use for respiratory motion modelling and compensation.

In this paper we propose a new generalized framework unifying image registration and motion model fitting into a single optimization. In addition to being a more efficient and robust approach, this enables the model to be fitted directly to the frames of partial image data, and does not require the data to be first sorted into full 3D volumes as described above, which would be required if building a surrogate-driven respiratory motion model using the typical approach. As the model is fitted to all of the frames of partial image data simultaneously the full 3D motion can be estimated, which can be challenging or impossible when using individual frames to estimate the motion. MCIR can also be easily incorporated into this framework using an iterative scheme. Consequently, the framework is particularly well suited to MCIR applications, as the motion model can be fitted directly to the unreconstructed partial image data, rather than requiring full reconstructed 3D images for fitting the model.

Some papers have proposed similar approaches to the framework presented here: Odille *et al* (2008) presented a method for fitting a linear motion model to MR  $k$ -space data; Hinkle *et al* (2012) presented a method for fitting a piece-wise linear model to CT projection or slice data; and Martin *et al* (2013) and Liu *et al* (2015) have presented methods for fitting linear models to CBCT projection data (modelling just tumour translations and deformations of the full anatomy respectively). However, these methods have been for specific types of data, registration algorithms, and motion models, whereas the framework presented here is generalized and applicable to a wide range of data, registration algorithms, and motion models. This work extends a previous conference publication (McClelland *et al* 2014) by giving a more detailed and rigorous presentation of the theory, using a more efficient B-spline based implementation, incorporating a super-resolution MCIR method, and presenting results from several real clinical datasets as well as phantom data.

## 2. Theory and methodology

### 2.1. Respiratory motion models

To build a respiratory motion model image data are simultaneously acquired with one or more surrogate signals (figure 1). The image data consist of  $N_i$  dynamic images,  $\mathbf{I}_1 \cdots \mathbf{I}_{N_i}$ , each acquired at a different time-point,  $t$ , and representing a different respiratory state. These should cover at least one respiratory cycle, although sometimes several cycles will be imaged so that inter-cycle variations can be sampled. A reference-state image,  $\mathbf{I}_0$ , is also required. This could be one of the dynamic images or it could be another image, e.g. a high quality breath-hold image.

Following the typical approach to building motion models shown in figure 1, the respiratory motion is first determined from the images, and then a correspondence model is fitted relating the motion and the surrogate signals. To determine the motion  $\mathbf{I}_0$  is registered to each of the dynamic images,  $\mathbf{I}_t$ . Each registration is usually performed independently, and results in a spatial transformation parameterized by the motion parameters,  $\mathbf{M}_t = [m_{t,1}, \cdots, m_{t,N_m}]$ , where  $N_m$  is the number of motion parameters. e.g.  $\mathbf{M}_t$  could be the parameters of an affine transform, a B-spline control point grid, or a voxel-wise deformation field.

The correspondence model relates  $\mathbf{M}_t$  to the surrogate signals,  $\mathbf{S}_t = [s_{t,1}, \cdots, s_{t,N_s}]$ , where  $N_s$  is the number of surrogate signals measured at each time-point,  $t$ . The model is parameterized by the model parameters,  $\mathbf{R} = \mathbf{R}_1 \cdots \mathbf{R}_{N_r}$ , where  $N_r$  is the number of model parameters per motion parameter, and  $\mathbf{R}_n = [r_{1,n}, \cdots, r_{N_m,n}]$ , i.e. the total number of model parameters is  $N_r \times N_m$ . The correspondence model can be represented as a function,  $F$ , of  $\mathbf{S}_t$  and  $\mathbf{R}$  that results in an estimate of the motion,  $\mathbf{M}_{F_t}$ :

$$\mathbf{M}_{F_t} = F(\mathbf{S}_t, \mathbf{R})$$

The correspondence model is fitted using a standard technique such as ordinary least squares, which minimizes the squared difference between the motion estimated by the model and the registration results over all time-points:

$$\min_{\mathbf{R}} \sum_t \|\mathbf{M}_t - \mathbf{M}_{F_t}\|_2^2$$

## 2.2. Unifying image registration and respiratory motion modelling

The framework presented in this paper does not follow the typical approach described above. Instead, it directly optimizes the correspondence model parameters,  $\mathbf{R}$ , on all the dynamic images simultaneously, such that the motion calculated using the correspondence model transforms  $\mathbf{I}_0$  to best match the dynamic image data, i.e.  $\mathbf{M}_t = \mathbf{M}_{F_t}$ .

Most image registration algorithms try to find the optimal values of  $\mathbf{M}_t$  by minimising a cost function,  $C_t$ , which consists of a (dis)similarity term, *Sim*, and an (optional) constraint term, *Con*.

$$C_t(\mathbf{I}_t, \mathbf{I}_0, \mathbf{M}_t) = \text{Sim}(\mathbf{I}_t, T(\mathbf{I}_0, \mathbf{M}_t)) + \text{Con}(\mathbf{M}_t)$$

where  $\mathbf{I}_t$  is the fixed (target) image,  $\mathbf{I}_0$  the moving (source) image, and  $T$  the function that applies the transformation parameterized by  $\mathbf{M}_t$  to image  $\mathbf{I}_0$ . Note,  $\mathbf{I}_0$  is used as the moving image here as this is more convenient when using partial image data (see section 2.3). A common approach for optimizing the values of  $\mathbf{M}_t$  is to estimate the gradient of  $C_t$  with respect to  $\mathbf{M}_t$ , i.e.  $\frac{\partial C_t}{\partial \mathbf{M}_t}$ , and then use a gradient based optimization such as gradient descent or conjugate gradient.

To directly optimize the correspondence model parameters on the image data the gradient of the cost function with respect to the model parameters,  $\frac{\partial C_t}{\partial \mathbf{R}}$ , is required. If the correspondence model can be differentiated to give the gradient of the motion parameters with respect to the model parameters,  $\frac{\partial \mathbf{M}_t}{\partial \mathbf{R}}$ , the chain rule can be applied, i.e.:

$$\frac{\partial C_t}{\partial \mathbf{R}} = \frac{\partial \mathbf{M}_t}{\partial \mathbf{R}} \frac{\partial C_t}{\partial \mathbf{M}_t}$$

Therefore, any correspondence model can be used in this framework as long as it is possible to calculate  $\frac{\partial \mathbf{M}_t}{\partial \mathbf{R}}$ . This includes all of the popular correspondence models that have been used in the literature, such as linear, polynomial, and B-spline models (McClelland *et al* 2013). e.g. for a 2nd order polynomial model with one surrogate signal:

$$\mathbf{M}_t = \mathbf{R}_1 s_t^2 + \mathbf{R}_2 s_t + \mathbf{R}_3$$

$$\frac{\partial \mathbf{M}_t}{\partial \mathbf{R}_1} = s_t^2, \quad \frac{\partial \mathbf{M}_t}{\partial \mathbf{R}_2} = s_t, \quad \frac{\partial \mathbf{M}_t}{\partial \mathbf{R}_3} = 1$$

The correspondence model parameters are optimized on all of the image data by summing the cost function and gradient from each of the individual images:

$$C_{\text{total}} = \sum_{t=1}^{N_i} C_t, \quad \frac{\partial C_{\text{total}}}{\partial \mathbf{R}} = \sum_{t=1}^{N_i} \frac{\partial C_t}{\partial \mathbf{R}}$$

$C_{\text{total}}$  and  $\frac{\partial C_{\text{total}}}{\partial \mathbf{R}}$  can then be used to find the optimal values of  $\mathbf{R}$  using the same gradient based optimization methods as used for standard registration between two images.

### 2.3. Using partial image data

With this new framework it is possible to fit the model directly to the partial image data, removing the need for full images, by modelling the image acquisition process:

$$\mathbf{P}_t = A_t(\mathbf{I}_t) + \varepsilon_t$$

where  $\mathbf{P}_t$  is the measured partial image data at time  $t$ ,  $\mathbf{I}_t$  the (unknown) full image, and  $\varepsilon_t$  the imaging noise.  $A_t$  is the function which simulates the image acquisition, e.g. for projection data  $A_t$  would be the forward-projection operator and for slice data  $A_t$  would be the slice selection profile.

When fitting the correspondence model the partial image data is simulated from the transformed reference-state image, and the similarity term,  $Sim$ , measures the (dis)similarity between the measured and simulated partial image data, i.e.

$$C_t(\mathbf{P}_t, \mathbf{I}_0, \mathbf{M}_t) = Sim(\mathbf{P}_t, A_t(T(\mathbf{I}_0, \mathbf{M}_t))) + Con(\mathbf{M}_t)$$

However,  $Sim$ , is now defined in the space of the partial image data but the transformation parameterized by  $\mathbf{M}_t$  is defined in the space of the full images. Therefore, when calculating  $\frac{\partial C_t}{\partial \mathbf{M}_t}$  it is necessary to transform the gradient of  $Sim$  from partial image space into full image space. This is done using the adjoint of the image acquisition function,  $A_t^*$ , e.g. the adjoint of the forward projection operator is the back projection operator:

$$\frac{\partial C_t}{\partial \mathbf{M}_t} = \frac{\partial Sim}{\partial \mathbf{M}_t} + \frac{\partial Con}{\partial \mathbf{M}_t}$$

$$\frac{\partial Sim}{\partial \mathbf{M}_t} = \frac{\partial \mathbf{I}_t}{\partial \mathbf{M}_t} \frac{\partial Sim}{\partial \mathbf{I}_t}$$

$$\frac{\partial Sim}{\partial \mathbf{I}_t} = A_t^* \left( \frac{\partial Sim}{\partial \mathbf{P}_{A_t}} \right)$$

where  $\mathbf{I}_t = T(\mathbf{I}_0, \mathbf{M}_t)$ , and  $\mathbf{P}_{A_t} = A_t(\mathbf{I}_t)$ .  $\frac{\partial Sim}{\partial \mathbf{P}_{A_t}}$  is the gradient of the similarity measure with respect to the simulated partial image data,  $\frac{\partial Sim}{\partial \mathbf{I}_t}$  is the gradient of the similarity measure with respect to the transformed reference-state image, and  $\frac{\partial \mathbf{I}_t}{\partial \mathbf{M}_t}$  is the gradient of the transformed reference-state image with respect to the motion parameters, e.g. for B-spline transformations this is the spatial gradient of the reference-state image transformed by  $\mathbf{M}_t$  and convolved with the B-spline kernel.

### 2.4. Incorporating motion compensated image reconstruction

The framework as described above assumes a reference-state image,  $\mathbf{I}_0$ , is available. However, if the motion is known MCIR can be used to reconstruct  $\mathbf{I}_0$  from the partial image data (Batchelor *et al* 2005, Rit *et al* 2009, Van Reeth *et al* 2012), so a separate reference-state image is not required. MCIR can be incorporated into this framework using an iterative approach. Firstly, a MCIR is performed from the partial image data using an initial estimate of no motion (or using



a previous model if one is available, e.g. from an earlier scan). The result will contain blurring and/or other motion artefacts, but can be used as an initial estimate of  $\mathbf{I}_0$  to fit the model parameters,  $\mathbf{R}$ , as described above. The fitted model is then used to perform another MCIR, updating the estimate of  $\mathbf{I}_0$ . The process continues to iterate between fitting  $\mathbf{R}$  and reconstructing  $\mathbf{I}_0$  until there is no more improvement or the maximum number of iterations is reached.

### 3. Experiments

#### 3.1. Implementation details

The framework has been implemented using the open source NiftyReg<sup>7</sup> deformable image registration software (Modat *et al* 2010). This uses a B-spline transformation model and conjugate gradient optimisation in a multi-resolution approach. As all the experiments use intra-modality data, sum of squared differences (SSD) was used as the similarity measure. The use of a constraint term was found to be unnecessary.

Any correspondence model can be used which can be expressed as a linear combination of terms pre-calculated from the surrogate signal(s),  $\varphi_n(\mathbf{S}_t)$ , i.e.

$$\mathbf{M}_t = \sum_{n=1}^{N_r} \mathbf{R}_n \varphi_n(\mathbf{S}_t)$$

then

$$\frac{\partial \mathbf{M}_t}{\partial \mathbf{R}_n} = \varphi_n(\mathbf{S}_t)$$

This includes linear, polynomial, and B-spline models. e.g. for a linear model with two surrogate signals,  $\mathbf{S}_t = [s_{t,1}, s_{t,2}]$ :

$$\varphi_1(\mathbf{S}_t) = s_{t,1}, \quad \varphi_2(\mathbf{S}_t) = s_{t,2}, \quad \varphi_3(\mathbf{S}_t) = 1$$

for a 2nd order polynomial model with one surrogate signal,  $\mathbf{S}_t = s_t$ :

$$\varphi_1(\mathbf{S}_t) = s_t^2, \quad \varphi_2(\mathbf{S}_t) = s_t, \quad \varphi_3(\mathbf{S}_t) = 1$$

and for a periodic B-spline model with 4 control points ( $N_r = 4$ ) and respiratory phase as the surrogate signal,  $\mathbf{S}_t = \vartheta_t$  (where  $\vartheta_t$  is the respiratory phase and has a value between 0 and 1):

$$\varphi_n(\mathbf{S}_t) = B_i(j)$$

Where  $i = (\lfloor 4\vartheta_t \rfloor + n - 2) \bmod 4$ ,  $j = 4\vartheta_t - \lfloor 4\vartheta_t \rfloor$ , and  $B_i$  is the  $i$ th cubic B-spline basis function:

$$B_1(j) = (1 - j)^3/6$$

$$B_2(j) = (3j^3 - 6j^2 + 4)/6$$

$$B_3(j) = (-3j^3 + 3j^2 + 3j + 1)/6$$

$$B_4(j) = j^3/6$$

The dynamic images can be full images, ‘slab’ images (consisting of a few slices), or individual slices. If the dynamic images are not full images the image acquisition function,  $A_t$ ,

<sup>7</sup> <http://cmictig.cs.ucl.ac.uk/research/software/niftyreg>



samples the deformed image,  $\mathbf{S}_{T_i}$ , at the slice(s) corresponding to the dynamic image,  $\mathbf{P}_i$ . The dynamic images can have a different resolution to the reference-state image,  $\mathbf{I}_0$ , enabling the use of a high resolution reference-state image with low resolution dynamic images, and permitting the use of super-resolution MCIR methods. If the dynamic images are lower resolution than  $\mathbf{I}_0$  the function  $A_i$  first convolves the deformed image,  $\mathbf{I}_{T_i}$  (which has the same resolution as  $\mathbf{I}_0$ ), with an appropriate Gaussian kernel to account for the different image resolutions (Cardoso *et al* 2015), before sampling the slices corresponding to  $\mathbf{P}_i$ . Likewise, the adjoint function,  $A_i^*$ , first resamples  $\frac{\partial \text{Sim}}{\partial \mathbf{P}_{A_i}}$  from the space of  $\mathbf{P}_i$  into the space of the full deformed image,  $\mathbf{I}_{T_i}$ . If the dynamic images are lower resolution, the result is then convolved with the same Gaussian kernel as above.

Two MCIR methods have been implemented: a simple averaging of the deformed dynamic images, and a super-resolution reconstruction using the iterative back-projection method (Irani and Peleg 1991). When performing a MCIR the dynamic image data (after applying  $A_i^*$ , i.e.  $A_i^*(\mathbf{P}_i)$ ) must be deformed into the space of  $\mathbf{I}_0$ . This could be done by estimating the inverse of the transformations parameterized by  $\mathbf{M}_i$ , but this can be computationally expensive. Instead the adjoint function  $T^*$  is used, i.e. ‘push-interpolation’ is used instead of the usual ‘pull-interpolation’. This can potentially lead to ‘holes’ (voxels with no intensity information) in the deformed images, but as long as the holes are not at the same location for all dynamic images they will not be present in the final MCIR. If holes are present in the final MCIR,  $A_i^*(\mathbf{P}_i)$  can be resampled at a higher resolution before applying  $T^*$ . However, sampling  $A_i^*(\mathbf{P}_i)$  at the same resolution as  $\mathbf{I}_0$  and using linear interpolation resulted in no holes in the MCIRs for all the experiments below.

### 3.2. Real datasets

Four real data sets have been used: full images (4DCT), slab images (Cine CT), thin slices (helical CT), and thick slices (multi-slice MR).

The full images dataset is one of the freely available DIR-Lab 4DCT datasets<sup>8</sup> (Castillo *et al* 2009). This consists of 10 CT volumes with a resolution of  $1.1 \times 1.1 \times 2.5 \text{ mm}^3$ , representing different phases of the respiratory cycle.

The slab image dataset uses Cine CT volumes (unsorted 4DCT data) acquired at University College Hospital, London, as the dynamic images. Each Cine CT volume consists of  $8 \times 2.5 \text{ mm}$  slices ( $0.98 \text{ mm} \times 0.98 \text{ mm}$  in-slice resolution), i.e. each volume only covers 2 cm, but they are acquired from 17 contiguous couch-positions such that the entire lungs are covered. Each couch-position is imaged 11 times, sampling just over one breath cycle, giving a total of 187 dynamic images. The surrogate signal was acquired using the Varian RPM system.

The thin slice dataset consists of individual CT slices from fast helical scans acquired at UCLA (Thomas *et al* 2014). Each helical scan consists of 301 slices and takes  $\sim 5 \text{ s}$  to acquire. The images were resampled with a resolution was  $1 \times 1 \times 1 \text{ mm}^3$ . 25 helical scans were acquired in alternating directions while the patient was freely breathing, although only 10 scans were used here. Therefore, each individual slice is imaged 10 times, with the images corresponding to arbitrary points in different breath cycles, such that both the intra- and inter-cycle variations are sampled for each slice. The surrogate signal was acquired using a respiratory belt.

<sup>8</sup> Dataset 5: [www.dir-lab.com/4DCT5.html](http://www.dir-lab.com/4DCT5.html)

**Table 1.** Details of the different experiments performed on the real datasets.

Dataset	Exp. num.	Correspondence model	B-spline CPG spacing	Reference-state image	Segmented lungs?
Full images	1	Periodic B-spline	10 mm	End-exh. 4DCT	Yes
Slab images	1	Linear	10 mm	End-exh. 4DCT	Only in ref.
Slab images	2	Linear	5 voxels	MCIR: averaging	No
Thin slices	1	Linear	5 mm	Helical scan	No
Thin slices	2	Linear	5 mm	MCIR: averaging	No
Thick slices	1	Linear	30 mm	MCIR: averaging	Yes
Thick slices	2	Linear	30 mm	MCIR: super-res.	Yes

The thick slice dataset consists of multi-slice MRI data acquired at the Institute of Cancer Research, London. Sagittal multi-slice MR data were acquired from a volunteer during free-breathing, from 30 adjacent slices encompassing both lungs. Each slice was imaged 10 times. The in-slice resolution was  $1.77 \text{ mm} \times 1.77 \text{ mm}$  and the slices thickness was 10 mm. The acquisition time for each slice was approximately 0.2 s, so the full acquisition took approximately 1 min. This acquisition was then repeated 4 more times, with a 2 mm left-right shift between each acquisition, giving a total of 150 overlapping slices. The surrogate signal was acquired from an MR-compatible ABC device working passively as a spirometer (Kaza *et al* 2015).

The details of the experiments performed on the different datasets are given in table 1. For the full images dataset a periodic B-spline correspondence model was used with respiratory phase as the surrogate signal. This can model intra-cycle variation (hysteresis), but not inter-cycle variation. For the other datasets a linear correspondence model was used, relating the motion to both the value and temporal derivative of the surrogate signal, i.e.  $\mathbf{S}_t = [s_t, \dot{s}_t]$ . This correspondence model has been popular in the literature due to its simplicity and ability to model both intra- and inter-cycle variations. The control point grid (CPG) spacing of the B-spline transformation was empirically set for each experiment, such that reasonable results were obtained. In all cases, the same spacing was used in all directions.

For experiment 1 on the full images and slab images datasets the end-exhalation 4DCT volumes was used as  $\mathbf{I}_0$ , as this is considered the most reproducible position and so the images usually contain the fewest artefacts. For experiment 1 on the thin slice dataset one of the helical scans was used as  $\mathbf{I}_0$ . As this was acquired while the patient was freely breathing the anatomy appears distorted, but due to the fast scan time there were no obvious missing or repeated structures. For the experiments that used the simple averaging MCIR method  $\mathbf{I}_0$  was reconstructed with the same resolution as the dynamic images. For experiment 2 on the thick slice dataset the super-resolution MCIR method was used to reconstruct  $\mathbf{I}_0$  with a voxel size of  $2 \times 1.77 \times 1.77 \text{ mm}^3$ .

Note, as the surrogate signals have been normalized so that their mean values equal 0 the reference-state images reconstructed by MCIR will represent the average position of the anatomy during the acquisition. For the experiments that did not perform MCIR,  $\mathbf{I}_0$  does not represent the average position of the anatomy. Therefore, an extra constant offset model

parameter was included in the linear models to account for the difference between the  $\mathbf{I}_0$  and the average position of the anatomy (this is not necessary for the B-spline model).

One problem encountered with several of the datasets was sliding motion between the lungs and the chest wall, which cannot be correctly modelled by a standard B-spline transformation. To alleviate this problem, the lungs were automatically segmented in some datasets using thresholding and morphological operations, as indicated in table 1. Voxels outside the lungs were set to the same intensity as soft tissue. For the slab images automatic lung segmentation was challenging in the dynamic images. Therefore, for experiment 1 the lungs were only segmented in  $\mathbf{I}_0$ , and for experiment 2, which used MCIR, they were not segmented at all. For the thin slice data good results were obtained with unsegmented images, so segmentation was not used.

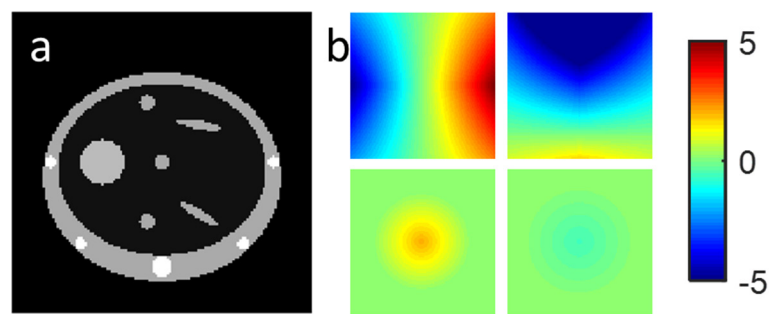
For the full images manually located landmark points are available which can be used to quantitatively assess the results. These consist of 300 landmarks well distributed over both lungs, which have been located in the end-inhalation and end-exhalation images. A subset of 75 of the landmarks have also been located on all of the exhalation images. Both sets of landmarks were used to assess the fitted motion model using the ‘snap-to-voxel’ approach described in Castillo *et al* (2009). For comparison the landmark errors have also been calculated using an estimate of no motion, i.e. measuring the magnitude of the motion in the original images.

For the other datasets it is difficult to accurately locate landmarks in the images, therefore only qualitative assessments of their results have been possible.

### 3.3. Phantom datasets

As it is difficult to estimate the ground truth motion for most of the real datasets, additional experiments were performed using datasets generated from a computer phantom. For these experiments the ground truth motion and reference-state image,  $\mathbf{I}_{\text{true}}$ , are known, so the results of the motion model fitting and the MCIR can be quantitatively assessed. A simple 2D ‘lung-like’ phantom was used for all the experiments (figure 2(a)). The phantom was animated using displacement fields generated from a linear correspondence model relating the motion to a surrogate signal and its temporal gradient. For the slab images, thin slices, and thick slices, the surrogate signals from the real datasets were used to animate the phantom. For the full images a real surrogate signal is not available, so an idealized surrogate signal from a  $\cos^4$  function was used. For all of the phantom image datasets Gaussian noise with a standard deviation of 3% of the maximum image intensity (1500) was added to the simulated images. When animating the phantom, the displacement field itself was used as the motion parameters,  $\mathbf{M}_t$ , whereas the B-spline control point displacements are used as  $\mathbf{M}_t$  when fitting the model. This means there are two motion parameters at each pixel, the displacement in the x and y directions, and therefore four correspondence model parameters for each pixel,  $r_{x,1}$ ,  $r_{x,2}$ ,  $r_{y,1}$ , and  $r_{y,2}$ . The values of these model parameters were manually defined such that they were smoothly varying over the phantom (figure 2(b)), and produced plausible looking motion that included deforming anatomy and intra- and inter-cycle variations.

The phantom datasets and the experiments performed with them mimicked the real datasets and experiments as closely as possible. The same number of images/slabs/slices was used as for the real data, but as the phantom was only 2D, images were 2D, and slices were 1D. Different sized phantoms were required for the different datasets:  $128 \times 128$  pixels for the full images,  $136 \times 136$  pixels for the slab images,  $301 \times 301$  pixels for the thin slices, and  $150 \times 150$  pixels for the thick slices. However, the phantom was resized so that its proportions were always the same. As the phantom used for the thin slices had approximately twice as many pixels as for the other datasets, the magnitude of motion was twice as large (in pixels) for this dataset. The thick slices were simulated by convolving the deformed images with a



**Figure 2.** (a) ‘Lung-like’ 2D computer phantom. (b) The values of the correspondence model parameters used to animate the phantom: top-left =  $r_{x,1}$ , top-right =  $r_{y,1}$ , bottom-left =  $r_{x,2}$ , bottom-right =  $r_{y,2}$ .

Gaussian kernel representing a slice thickness of 5 pixels. The surrogate signals from the real datasets were also used for the phantom experiments, and the same correspondence models and MCIR methods were applied as for the real data. The B-spline CPG spacing was empirically set to 10 pixels for the thin slice dataset, and to 5 pixels for all the others. As the phantom datasets did not contain sliding motion there was no need for segmentation.

The results of all the phantom experiments were assessed by calculating the Displacement Field Error, DFE, defined as the 2D Euclidean distance at each pixel between the true displacement field and the displacement field resulting from the fitted motion model. The DFE was calculated at every pixel inside the deformed phantom for every time-point. Pixels outside the phantom were ignored as they do not contain any image data to guide the model fitting. The DFE was also calculated for each experiment using an estimate of no motion, to quantify the amount of motion present in the original data. The experiments that performed MCIRs were also assessed by comparing the reconstructed  $\mathbf{I}_0$  to  $\mathbf{I}_{\text{true}}$ . This was accomplished by calculating the absolute difference in the image intensities at every pixel and the correlation coefficient between the images. Similar calculations were performed for an image reconstructed using the same method (averaging or super-resolution) but assuming no motion.

## 4. Results

### 4.1. Real datasets

Videos showing the results of the experiments on the real datasets are available as supplementary files. These show the original dynamic images,  $\mathbf{I}_i$  or  $\mathbf{P}_i$ , the simulated dynamic images,  $\mathbf{I}_{T_i}$  or  $\mathbf{P}_{A_i}$ , and a colour overlay of these. In addition, for the datasets using partial image data (i.e. the slab and slice datasets), the videos also show one or more views through the full deformed reference-state image,  $\mathbf{I}_{T_i}$ .

**4.1.1. Full images.** As shown in the supplementary movie (real\_data\_full\_images.mp4) ([stacks.iop.org/PMB/62/4273/mmedia](https://stacks.iop.org/PMB/62/4273/mmedia)), there is generally a good match between  $\mathbf{I}_i$  and  $\mathbf{I}_{T_i}$ . Some mismatches are noticeable near the back of the lungs due to sliding motion, even though the lungs had been segmented. There are also some smaller mismatches throughout the lungs, but some of these are due to artefacts present in the original dynamic 4DCT images. Some images contain blurred and repeated structures (due to the CT rotation time being too slow to ‘freeze’ the motion) but the  $\mathbf{I}_{T_i}$  images do not contain these artefacts as they are not present in  $\mathbf{I}_0$ . Also,

**Table 2.** Landmark errors, in mm, for the full images dataset using an estimate of no motion and using the fitted motion model to estimate the motion (model). Landmark set 1 are the 300 landmarks identified in the end-exhalation and end-inhalation images only. Landmark set 2 are the 75 landmarks that have been identified in all exhalation images.

	Landmark set 1		
	Mean	Std. dev.	95th percentile
No motion	7.48	5.51	17.67
Model	1.88	2.03	5.24
	Landmark set 2		
	Mean	Std. dev.	95th percentile
No motion	3.57	4.49	15.04
Model	1.72	2.25	4.28

the motion seen in the original 4DCT images appears ‘wobbly’ due to binning errors and inter-cycle variations, but the motion model estimates smooth and continuous motion over the whole breath cycle. The landmark errors given in table 2 support the qualitative results observed in the movie. Landmark errors are greatly reduced when using the fitted motion model, and are comparable to those reported on the DIR-Lab website<sup>9</sup>, where the mean errors range from 1.07 mm to 3.59 mm. Some of the more recent methods have lower landmark errors than achieved here, but this may be partly due to most other methods registering the images individually, which can result in lower landmark errors but will also reproduce the ‘wobbly’ motion seen in the 4DCT.

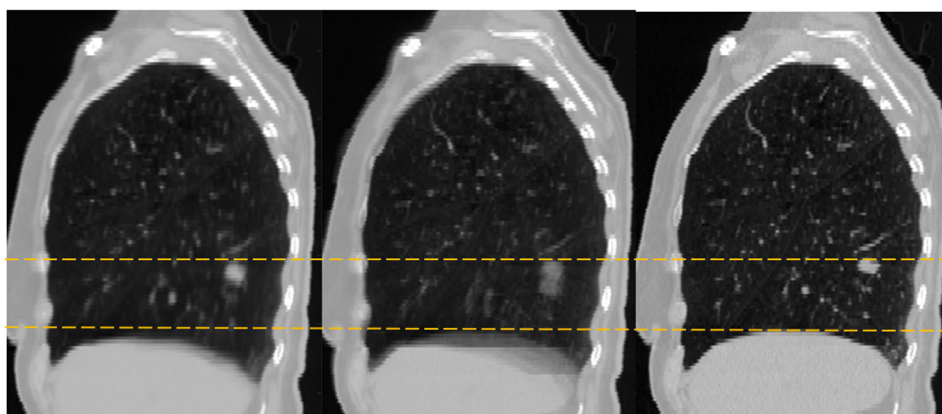
**4.1.2. Slab images.** There was a fairly good match between  $\mathbf{P}_i$  and  $\mathbf{P}_{A_i}$  for both experiments, as can be seen in the supplementary movies (real\_data\_slab\_images\_exp1.mp4 and real\_data\_slab\_images\_exp2.mp4). As  $\mathbf{P}_i$  only sample one respiratory cycle at each couch-position it is difficult to assess how accurately the models estimate the inter-cycle variations, but the images of  $\mathbf{I}_{T_i}$  show plausible motion over the entire field of view, exhibiting both intra- and inter-cycle variations. As with the 4DCT data, there are some noticeable mismatches at the back of the lungs due to the sliding motion, and some smaller mismatches in other regions, but again these are partly due to blurred and repeated structures in some of the Cine CT volumes.

Figure 3 compares the end exhalation 4DCT volume with the MCIR produced in experiment 2, and a similar image formed without applying any motion correction (i.e. the result of averaging all of the Cine CT volumes together). It is evident that blurred structures are greatly reduced in the MCIR compared to the result with no motion correction, indicating that most of the motion has been well recovered. Some structures still appear slightly blurred compared to the 4DCT volume, particularly near the back of the lungs where sliding motion occurs. This blurring is partly caused by errors in the estimated motion, but also due to the structures appearing blurred in some of the Cine CT volumes.

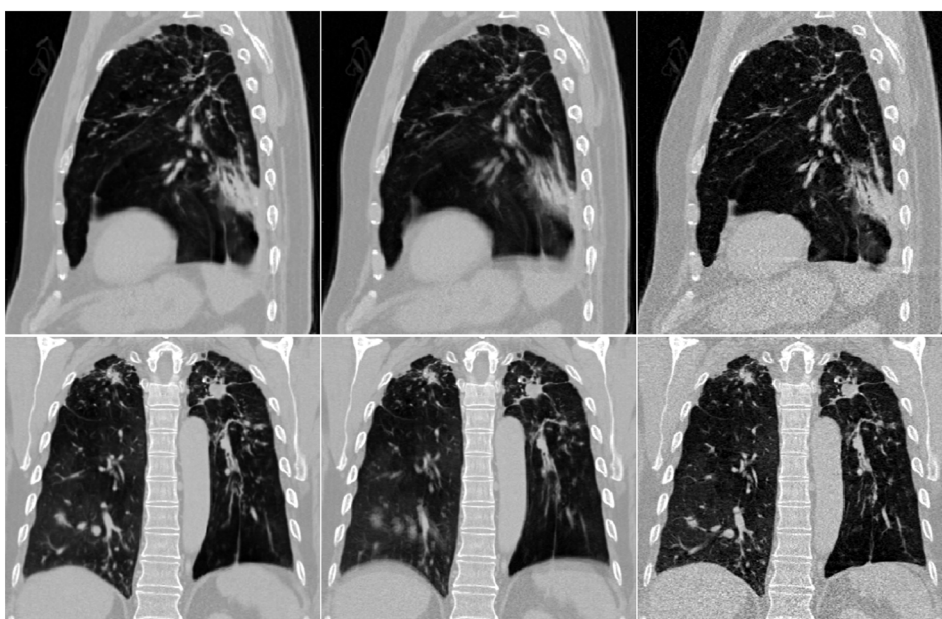
**4.1.3. Thin slices.** For both experiments there is a very good match between  $\mathbf{P}_i$  and  $\mathbf{P}_{A_i}$ , as can be seen in the supplementary movies (real\_data\_thin\_slices\_exp1.mp4 and real\_data\_thin\_slices\_exp2.mp4). The images of  $\mathbf{I}_{T_i}$  show plausible motion, including intra- and inter-cycle variations, over the entire field of view, and even seem to reproduce most of the sliding motion seen at the back of the lung. Since  $\mathbf{P}_i$  sample both the intra- and

<sup>9</sup>Dataset 5: [www.dir-lab.com/Results.html](http://www.dir-lab.com/Results.html)



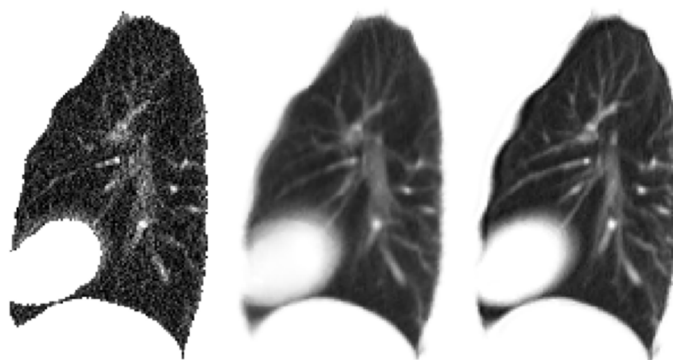


**Figure 3.** The MCIR produced from the real slab images dataset (left), a similar image formed without applying any motion correction (middle), and the end-exhalation 4DCT volume for the same subject (right). The dashed lines represent the height of the tumour and diaphragm in the end-exhalation 4DCT volume, and show that these are both lower in the MCIR. This is expected as the MCIR corresponds to the average position of the anatomy.

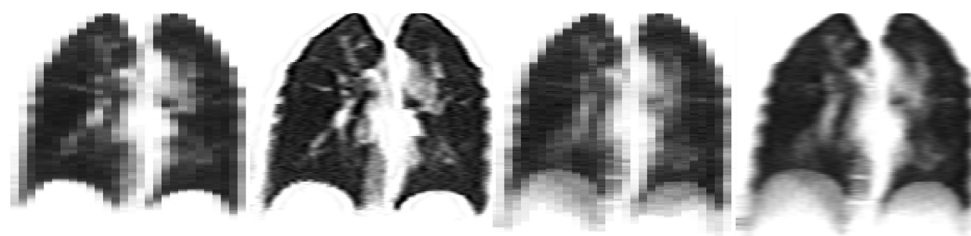


**Figure 4.** Sagittal (top) and Coronal (bottom) views through the MCIR produced from the real thin slices dataset (left), a similar image formed without applying any motion correction (middle), and one of the fast helical scans from the same subject (right).

inter-cycle variations at each location, this implies the true variations have been modelled well. Figure 4 shows the helical scan used as  $I_0$  for experiment 1, the MCIR produced by experiment 2, and a similar image formed without applying any motion correction. The blurring has been almost completely removed in the MCIR, with the majority of structures appearing as



**Figure 5.** One of the original dynamic images from the real thick slice dataset (left), and the corresponding simulated dynamic images from experiment 1 (middle) and experiment 2 (right).



**Figure 6.** The MCIRs produced from the real thick slice dataset in experiment 1 using the averaging method (far-left) and experiment 2 using the super-resolution method (middle-left), and similar images formed using the averaging method (middle-right) and the super-resolution method (far-right) without applying motion correction.

sharp as in the helical scan, indicating the motion has been modelled very well. In addition, the signal-to-noise ratio was noticeably improved in the MCIR due to combining the image data from all 10 slices at each location.

**4.1.4. Thick slices.**  $\mathbf{P}_{A_i}$  match  $\mathbf{P}_i$  well for both experiments, as can be seen in the supplementary movies (real\_data\_thick\_slices\_exp1.mp4 and real\_data\_thick\_slices\_exp2.mp4). Similar to the thin slices dataset,  $\mathbf{P}_i$  sample both the intra- and inter-cycle variations at each location, which indicates these have been modelled well. The images of  $\mathbf{I}_{T_i}$  show plausible motion over the entire field of view, although the motion at the edge of the lungs looks questionable, particularly for experiment 2. This is due to the sliding motion between the lungs and the ribs, and the ribs sometimes being included in the segmented lungs as they have similar intensity in the original MR images.

Figure 5 shows an example  $\mathbf{P}_i$  (after lung segmentation), together with the corresponding  $\mathbf{P}_{A_i}$  from both experiments. The signal-to-noise ratio has greatly improved in  $\mathbf{P}_{A_i}$  from experiment 1, due to averaging the dynamic images, and  $\mathbf{P}_{A_i}$  from experiment 2 shows even finer structure inside the lung due to the super-resolution reconstruction and an increased number of dynamic images. Figure 6 shows coronal views through the MCIRs produced by both experiments, together with similar images formed without applying any motion correction. The blurring has been removed in the MCIRs, indicating that motion has been recovered



**Table 3.** Displacement field error (DFE), in pixels, for the different datasets and experiments, assuming no motion (no motion) and using the fitted models to estimate the motion (model).

Dataset	Experiment	DFE—no motion			DFE—model		
		Mean	Std. dev.	95th percentile	Mean	Std. dev.	95th percentile
Full Images	1	3.56	3.79	11.13	0.42	0.35	1.08
Slab Images	1	3.53	3.61	11.09	0.49	0.47	1.26
	2	2.82	2.19	7.15	0.53	0.66	1.94
Thin Slices	1	8.37	6.82	20.99	1.52	1.57	4.97
	2	5.98	3.82	12.92	0.49	0.64	1.87
Thick Slices	1	2.83	2.02	6.91	0.84	0.80	2.47
	2	2.98	1.93	6.56	0.39	0.48	1.38

**Table 4.** Absolute intensity differences and correlation coefficients between  $\mathbf{I}_0$  and  $\mathbf{I}_{\text{true}}$ . Avg. = MCIR performed by averaging the deformed dynamic images, S-R = MCIR performed using the super-resolution method, model = motion estimated using fitted motion model, no motion = MCIR performed assuming no motion.

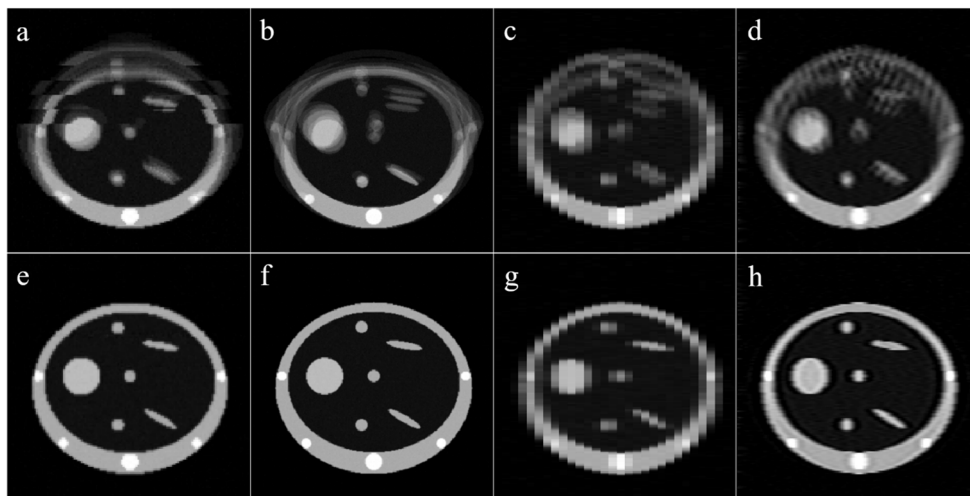
Dataset	MCIR	Absolute intensity difference			Correlation coefficient
		Mean	Std. dev.	95th percentile	
Slab images	Avg.—model	23.78	55.93	156.03	0.99
	Avg.—no motion	97.85	180.73	538.29	0.83
Thin slices	Avg.—model	15.20	39.00	28.83	0.99
	Avg.—no motion	94.67	187.22	563.21	0.82
Thick slices	Avg.—model	63.66	120.42	365.63	0.93
	Avg.—no motion	101.49	170.14	484.51	0.85
	S-R—model	45.37	76.04	209.13	0.97
	S-R—no motion	101.65	175.56	515.81	0.83

well. Much more detail can be seen in the MCIR from experiment 2, indicating that the super-resolution method has been applied successfully and that the motion has been modelled well.

#### 4.2. Phantom datasets

Videos showing the results of the experiments on the phantom datasets are available as supplementary files. Each of these show the original dynamic images,  $\mathbf{I}_i$  or  $\mathbf{P}_i$ , the simulated dynamic images,  $\mathbf{I}_{T_i}$  or  $\mathbf{P}_{A_i}$ , and a colour overlay of these. Also, for the phantom datasets using partial image data the true deformed reference-state images,  $\mathbf{I}_i$ , are shown together with the reference-state image deformed using the fitted model,  $\mathbf{I}_{T_i}$ , and a colour overlay of these. The videos show that there were good matches between  $\mathbf{P}_i$  and  $\mathbf{P}_{A_i}$  and between  $\mathbf{I}_i$  and  $\mathbf{I}_{T_i}$  for all phantom datasets experiments. In addition, for the experiments that performed MCIRs,  $\mathbf{I}_0$  closely resemble  $\mathbf{I}_{\text{true}}$ , although for experiment 1 on the thick slice data  $\mathbf{I}_0$  has a noticeably lower through-slice resolution, as expected.

The qualitative results shown in the videos are supported by the quantitative results in tables 3 and 4. Table 3 gives the Displacement Field Error (DFE) for each experiment. When using the fitted model the DFE is greatly reduced for all experiments, in most cases to below



**Figure 7.** MCIRs generated from the phantom datasets assuming no motion (a)–(d) and using the fitted motion models to estimate the motion (e)–(h). The MCIRs were constructed by averaging the deformed dynamic images for the slab datasets (a) and (e), the thin slice dataset (b) and (f), and the thick slice dataset (c) and (g), and using the super-resolution method for the thick slice dataset (d) and (h).

1 pixel, indicating the motion has been recovered very well. The DFE is larger for experiment 1 on the thin slice data, but the magnitude of the motion was twice as large for this dataset (as it represents higher resolution data), and the relative reduction in DFE when using the model is comparable to the other datasets. Interestingly, the use of MCIR (i.e. experiment 2) leads to a considerably lower DFE for the thin slice dataset but not for the slab images dataset. For the thick slice data, experiment 2, which uses a super-resolution MCIR, yields a much lower DFE than experiment 1, which uses the averaging MCIR. This result is expected, as the super-resolution MCIR has a higher resolution and hence can recover the motion more accurately.

Table 4 displays the results of comparing  $I_0$  to  $I_{\text{true}}$ , and figure 7 shows images of  $I_0$  for each experiment performing MCIR. These results indicate that the motion has been recovered well by the fitted motion models and that  $I_0$  closely resembles  $I_{\text{true}}$  for all experiments. As expected, the results for the thick slice dataset are not as good as for the thin slice and slab image datasets, because the dynamic images have a lower resolution than  $I_{\text{true}}$ . However, the use of the super-resolution method improves the results, and produces an  $I_0$  more similar to  $I_{\text{true}}$ .

## 5. Discussion and conclusions

This paper has presented a generalized framework that unifies image registration and fitting a respiratory correspondence model into a single optimisation, and can incorporate motion compensated image reconstruction (MCIR) using an iterative scheme. This framework allows motion models to be fitted directly to the unsorted partial image data, as well as to respiratory-sorted 3D volumes. This overcomes the need for using respiratory-sorted volumes when building the surrogate-driven motion models, which are required by most other methods, and we believe has been a major factor limiting the accuracy and applicability of such methods to date.

There are a few other methods from the literature that could potentially be directly applied to the unsorted slab and slice data used in this paper. The method of Hinkle *et al* (2012) could be applied to any of the slab and slice datasets, however, their method uses a piece-wise linear motion model, and so cannot model inter-cycle variations. In McClelland *et al* (2011) a motion model is fitted directly to Cine CT slab data using the typical approach of first performing the image registrations and then fitting the motion model. However, the Cine CT data used in McClelland *et al* (2011) had 12 slices, whereas the data used here only has 8 slices, and the image registration results were often found to be unsatisfactory on the 8-slice Cine CT data. In addition, the method in McClelland *et al* (2011) is not appropriate for single slice data. In Thomas *et al* (2014) a motion model is fitted to the same thin slice CT data used in this paper, and the final results are similar to the results from this paper. However, the method of Thomas *et al* (2014) requires the helical scans to be acquired very quickly, so as to minimise the distortions and artefacts in the full 3D helical images, and hence allow them to be registered to each other. The method presented here does not require the 3D helical images, so can use data acquired more slowly on a standard CT scanner, as long as the individual slices are acquired fast enough to be considered motion-artefact free (a 0.5 s rotation time is usually considered fast enough, although even then there can still be artefacts, especially for slices acquired at mid in/exhale).

This framework has been implemented using the NiftyReg registration software, and has been applied to phantom and real datasets consisting of full images, small slabs, and individual slices. Good quantitative results were obtained on the phantom data, where the ground truth motion and true reference-state image were known, and on the real 4DCT dataset, where manually annotated landmark points are available. Promising results were also obtained on the other real datasets, but these could only be assessed qualitatively as it was difficult to accurately estimate the ground truth motion. Intra-cycle and inter-cycle variations appear to have been successfully modelled when present in the data. The results indicate that MCIR can be successfully incorporated into the framework, and that a super-resolution method can be used.

The results presented in this paper represent a general proof of principle. Future work is required to tailor the framework for specific applications. This will include investigating the best choices for the surrogate data, correspondence model, imaging data, registration algorithm, and MCIR method (if used), as well as tuning the various parameters settings. The choice of surrogate signals will depend on which signals can be easily acquired during the acquisition/intervention and how well they relate to the respiratory motion and its variation. The choice of correspondence model will depend on the type of motion and variation to be modelled, and on the surrogate signals and imaging data available. See McClelland *et al* (2013) for a detailed discussion of the choices available. The imaging data used in this paper had already been reconstructed into 2D slices or 3D volumes. Future work will investigate using 'raw' imaging data, such as CT projections or MR *k*-space data, instead of reconstructed images. Work will also be required to determine how much data to acquire and the exact protocol to use, so as to best capture the motion and variation to be modelled whilst limiting the scanning time and (for CT) dose.

The popular B-spline registration algorithm was used in this paper, as it has previously been shown to provide good results in a wide range of applications, and an efficient open source implementation was available. However, one of the drawbacks of this algorithm (and many others) is its inability to properly account for sliding motion which often occurs between the lungs and the chest wall during respiration. A number of approaches to handling sliding motion have been proposed (Delmon *et al* 2013) which could be incorporated into the framework presented in this paper. Some of these approaches require the sliding regions to be segmented in the images. Segmenting the sliding regions in the dynamic image data may be

impossible if they are partial or raw image data, but it should be possible in the reference-state image. However, if this is formed using MCIR the segmentation will need repeating every time a new MCIR is performed. Future work will also investigate the use of a stationary velocity field transformation model, which is guaranteed to give diffeomorphic transformations (Wilms *et al* 2014), as well as different similarity measures, constraint terms, and optimisation methods.

Super-resolution reconstruction has previously been proposed for lung MR data (Van Reeth *et al* 2015), but this required 3D volumes to determine the motion so was based on the assumption that there was no inter-cycle variation. To the best of our knowledge, this is the first time that super-resolution reconstruction has been used directly on the slice data so that inter-cycle variations can be accounted for. The iterative back-projection method was used as this was the simplest to implement. Future work will investigate the use of different super-resolution methods (Van Reeth *et al* 2012), as well as other iterative and non-iterative MCIR methods applicable to ‘raw’ imaging data. Additionally, the possibility of jointly optimising the image and model parameters simultaneously, rather than iteratively, will be explored.

In addition to tailoring the framework to the specific applications, the results will need to be thoroughly validated. As highlighted in this paper, this can be challenging for partial image data, and will be even more challenging if raw imaging data are used in the future. It will require a combination of software and hardware phantoms, manually annotated clinical data (where possible), manually and automatically graded results, and extra data specifically for validation (e.g. motion traces from implanted markers).

The method presented here also has some inherent assumptions and limitations. Like all surrogate-driven motion modelling approaches, there is an assumption that the motion is related to some surrogate signal(s), and that the signal(s) can be easily acquired during the image acquisition (and if the models are to be used to guide treatment, during the treatment delivery). However, there is good evidence (e.g. from the Cyberknife system (Hoogeman *et al* 2009)) that this assumption holds well, at least for short time frames. If partial imaging data is used, then the method will also be limited by how accurately the image acquisition function (and its adjoint) represents the real image acquisition, but this is a limitation for all image reconstruction methods. This is also related to the issue of computation time, as more accurate image acquisition functions may be more computationally expensive. But computation time is also affected by other factors, in particular the type and amount of data and the type of MCIR used (if any). e.g. experiments 1 and 2 on the real thick slice (MR) dataset took 6 and 107 min to run respectively, and experiments 1 and 2 on the real thin slice (CT) dataset took 297 and 386 min to run respectively. It should be noted that what is considered an acceptable computation time depends very much on the specific application that the model is being used for, and ensuring that the computation time is acceptable will be an important part of tailoring the framework for different applications.

In conclusion, this paper has presented a general purpose and versatile framework that can be used for building respiratory motion models from partial imaging data, as well as full images. Much work is still required to optimize and fully exploit the framework for different types of data and clinical applications, but very promising results have been obtained on several phantom and real datasets, and the framework has potential uses in a wide range of clinical applications.

## Acknowledgments

This work has been funded by an EPSRC programme grant (EP/H046410/1), the CRUK/EPSRC KCL-UCL Comprehensive Cancer Imaging Centre (MOETAJR, NS/A000001/1), a CRUK Multi-disciplinary grant (CRC 521), and the CRUK Centres’ Network Accelerator

Award Grant (A21993). We acknowledge CRUK and EPSRC support to the Cancer Imaging Centre at ICR and RMH in association with MRC and Department of Health C1060/A10334, C1060/A16464 and NHS funding to the NIHR Biomedical Research Centre and the Clinical Research Facility in Imaging. MOL is an NIHR Emeritus Senior Investigator.

## References

- Batchelor P G, Atkinson D, Irarrazaval P, Hill D L, Hajnal J and Larkman D 2005 Matrix description of general motion correction applied to multishot images *Magn. Reson. Med.* **54** 1273–80
- Cardoso M J, Modat M, Vercauteren T and Ourselin S 2015 Scale factor point spread function matching: beyond aliasing in image resampling *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* vol 9350, ed N Navab, J Hornegger, W Wells and A Frangi (Cham: Springer) ([https://doi.org/10.1007/978-3-319-24571-3\\_81](https://doi.org/10.1007/978-3-319-24571-3_81))
- Castillo R, Castillo E, Guerra R, Johnson V E, McPhail T, Garg A K and Guerrero T 2009 A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets *Phys. Med. Biol.* **54** 1849–70
- Delmon V, Rit S, Pinho R and Sarrut D 2013 Registration of sliding objects using direction dependent b-splines decomposition *Phys. Med. Biol.* **58** 1303–14
- Fayad H, Pan T, Clement J F and Visvikis D 2011 Technical note: correlation of respiratory motion between external patient surface and internal anatomical landmarks *Med. Phys.* **38** 3157–64
- Hinkle J, Szegedi M, Wang B, Salter B and Joshi S 2012 4d ct image reconstruction with diffeomorphic motion model *Med. Image Anal.* **16** 1307–16
- Hoogeman M, Prevost J B, Nuytens J, Poll J, Levendag P and Heijmen B 2009 Clinical accuracy of the respiratory tumor tracking system of the cyberknife: assessment by analysis of log files *Int. J. Radiat. Oncol. Biol. Phys.* **74** 297–303
- Irani M and Peleg S 1991 Improving resolution by image registration *Graph Models Image Process* **53** 231–9
- Kaza E, Symonds-Taylor R, Collins D J, McDonald F, McNair H A, Scurr E, Koh D M and Leach M O 2015 First MRI application of an active breathing coordinator *Phys. Med. Biol.* **60** 1681–96
- Keall P J *et al* 2006 The management of respiratory motion in radiation oncology report of aapm task group 76 *Med. Phys.* **33** 3874–900
- Liu J L, Zhang X, Zhang X Q, Zhao H K, Gao Y, Thomas D, Low D A and Gao H 2015 5d respiratory motion model based image reconstruction algorithm for 4d cone-beam computed tomography *Inverse Problems* **31** 115007
- Low D A, Parikh P J, Lu W, Dempsey J F, Wahab S H, Hubenschmidt J P, Nystrom M M, Handoko M and Bradley J D 2005 Novel breathing motion model for radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **63** 921–9
- Martin J *et al* 2013 Building motion models of lung tumours from cone-beam ct for radiotherapy applications *Phys. Med. Biol.* **58** 1809–22
- McClelland J R, Champion B A S and Hawkes D J 2014 Combining image registration, respiratory motion modelling, and motion compensated image reconstruction *Workshop on Biomedical Image Registration (WBIR)* vol 8545, ed S Ourselin and M Modat (Cham: Springer) ([https://doi.org/10.1007/978-3-319-08554-8\\_11](https://doi.org/10.1007/978-3-319-08554-8_11))
- McClelland J R, Hawkes D J, Schaeffter T and King A P 2013 Respiratory motion models: a review *Med. Image Anal.* **17** 19–42
- McClelland J R, Hughes S, Modat M, Qureshi A, Ahmad S, Landau D B, Ourselin S and Hawkes D J 2011 Inter-fraction variations in respiratory motion models *Phys. Med. Biol.* **56** 251–72
- Modat M, Ridgway G R, Taylor Z A, Lehmann M, Barnes J, Hawkes D J, Fox N C and Ourselin S 2010 Fast free-form deformation using graphics processing units *Comput. Methods Prog. Biomed.* **98** 278–84
- Odille F, Vuissoz P A, Marie P Y and Felblinger J 2008 Generalized reconstruction by inversion of coupled systems (grics) applied to free-breathing MRI *Magn. Reson. Med.* **60** 146–57
- Paganelli C, Summers P, Bellomi M, Baroni G and Riboldi M 2015 Liver 4dmri: a retrospective image-based sorting method *Med. Phys.* **42** 4814–21
- Pan T 2005 Comparison of helical and cine acquisitions for 4d-ct imaging with multislice ct *Med. Phys.* **32** 627–34

- Rank C M, Heusser T, Buzan M T, Wetscherek A, Freitag M T, Dinkel J and Kachelriess M 2016 4d respiratory motion-compensated image reconstruction of free-breathing radial mr data with very high undersampling *Magn. Reson. Med.* **77** 1170–83
- Rit S, Wolthaus J W H, van Herk M and Sonke J J 2009 On-the-fly motion-compensated cone-beam ct using an *a priori* model of the respiratory motion *Med. Phys.* **36** 2283–96
- Sonke J J, Zijp L, Remeijer P and van Herk M 2005 Respiratory correlated cone beam ct *Med. Phys.* **32** 1176–86
- Thomas D, Lamb J, White B, Jani S, Gaudio S, Lee P, Ruan D, McNitt-Gray M and Low D 2014 A novel fast helical 4d-ct acquisition technique to generate low-noise sorting artifact-free images at user-selected breathing phases *Int. J. Radiat. Oncol. Biol. Phys.* **89** 191–8
- Van Reeth E, Tan C H, Tham I W and Poh C L 2015 Isotropic reconstruction of a 4-d mri thoracic sequence using super-resolution *Magn. Reson. Med.* **73** 784–93
- Van Reeth E, Tham I W K, Tan C H and Poh C L 2012 Super-resolution in magnetic resonance imaging: a review *Concept Magn. Reson. A* **40a** 306–25
- von Siebenthal M, Szekely G, Gamper U, Boesiger P, Lomax A and Cattin P 2007 4d MR imaging of respiratory organ motion and its variability *Phys. Med. Biol.* **52** 1547–64
- Wilms M, Werner R, Ehrhardt J, Schmidt-Richberg A, Schlemmer H P and Handels H 2014 Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy *Phys. Med. Biol.* **59** 1147–64
- Yang Y X, Teo S-K, Van Reeth E, Tan C H, Tham I W K and Poh C L 2015 A hybrid approach for fusing 4d-mri temporal information with 3d-ct for the study of lung and lung tumor motion *Med. Phys.* **42** 4484–96