CHAPTER 14

Statistical Respiratory Models for Motion Estimation

Christoph Jud*, Philippe C. Cattin*, Frank Preiswerk*†

*Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland
†Department of Radiology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

Contents

14.1 Background 379
14.2 4-Dimensional MR Imaging 381
   14.2.1 Acquisition Protocol 382
   14.2.2 Retrospective Stacking 383
14.3 Motion Model Building 384
   14.3.1 Statistical Shape Model 384
   14.3.2 Statistical Motion Model 385
   14.3.3 Inference on Statistical Models
     14.3.3.1 Sparse Reconstruction 386
     14.3.3.2 Statistical Model Regression 387
14.4 Establishment of Correspondence 388
   14.4.1 Inter-Subject Correspondence 388
   14.4.2 Inter-Subject Correspondence by Registration 390
   14.4.3 Intra-Subject Correspondence 391
14.5 Statistical Motion Modeling 391
14.6 Bayesian Reconstruction from Sparse Data 392
14.7 Applications of Population-Based Statistical Motion Models to Motion Reconstruction 394
14.8 Reconstruction by Regression 401
   14.8.1 Average Breathing Cycle 401
   14.8.2 Motion Model Prediction 402
14.9 Conclusion 404
Acknowledgments 404
References 404

14.1 BACKGROUND

The quasi-perpetual motion of organs caused by the respiratory cycle is a complicating factor in many medical contexts. In particular during imaging, for example using Magnetic Resonance Imaging (MRI) or Computed Tomography (CT), respiratory motion is problematic as it is responsible for motion blur in the resulting images. But respiratory motion is not only problematic during the diagnostic stage, it also complicates surgical treatment and tumour ablation in the abdomen. There exist three approaches often used...
in these applications: (1) breath holding or high-frequency low tidal ventilation, (2) respiratory gating where the patient is only imaged/treated around exhalation or (3) target tracking with implantable markers such as gold beads or active markers. Common to these approaches is that they either prolong treatment time or they are of invasive nature, thereby complicating the procedure.

As cancer is a leading cause of death worldwide, better motion mitigation approaches are required. Cancer accounted for 13% of all deaths in 2014 [36]. The main types of lethal cancer are lung, stomach, liver, colon, and breast tumors. All these sites are subject to respiratory motion, which is a complicating factor for treatment. A well known and established method for the treatment of non-resectable tumors is radiation therapy, which uses ionizing radiation to destroy tumor cells. The art of radiation therapy is to deliver a lethal dose to all cancerous tissue while sparing as much healthy tissue as possible, in particular organs that are radiation sensitive or for which the consequences of side effects can be severe. Consequently, accurate target localization and delivery is one of the main challenges in radiation therapy.

There are a number of different factors contributing to potential inaccuracies in irradiation. Of these, respiratory motion is one of the main problems in the thorax and in the abdomen and has been shown to have a large dosimetric impact on conventional radiotherapy [41,13,21]. Besides quasi-periodic respiratory motion, which includes variation in breathing depth and speed, the organs undergo also other modes of deformation, called secondary modes [40]. The secondary modes are caused, for example, by the cardiac motion, digestive activity, gravity, muscle relaxation, or filling of the bladder. The quasi-periodic and secondary modes constitute the total organ motion seen during treatment.

Recent advances in three-dimensional planning and advanced treatment technologies, such as intensity modulated radiotherapy (IMRT), intensity modulated proton therapy (IMPT) [19] and high-frequency focused ultrasound (HIFU) [7–9], have brought about new possibilities in delivering highly conformal dose distributions. The advantage of highly localized treatment makes these techniques sensitive to organ motion, which represents a limiting factor for exploiting their full potential [40,27].

Countless works on handling organ motion, more specifically respiratory motion, have been published in recent years. A comprehensive overview including practical guidelines can be found in a report of the American Association of Physicists in Medicine [1,15].

A simple method to avoid respiratory motion is to completely interrupt breathing while the therapy beam is on [25,23,11,4,16]. In gated treatment the beam is only turned on during a certain period of the breathing cycle, for example during exhalation [13,24,18,10,12]. Although the aforementioned approaches compensate respiratory motion to some extent, they require reproducibility of the organ position for the selected breathing phase [40] and prolong the treatment time. More importantly, they
only compensate for the perpetual respiratory motion and as such are completely oblivious to all other modes of organ motion. Thus they are only accurate in a short window of a couple of minutes after patient set-up, as was show in [40]. Hence, organ motion during radiotherapy continues to be a problem [9] and much research effort is being put into understanding and addressing this issue.

It would be desirable to keep the target and the treatment beam aligned throughout the entire breathing cycle. This technique is commonly referred to as tracking. Tracking is very demanding as it requires some prediction of the target motion [35,34, 39]. Although tracking is in principle designed to follow any target motion, it profits from a possibly regular breathing pattern because this simplifies the required short-time prediction of the motion trajectory. The high sampling rate and low lag of ultrasound (US) makes it a very attractive non-ionizing imaging modality to use for tracking of the target. However, only few tumors are visible under US and it completely fails for structures inside the lung due the air. Thus, conventional ultrasound imaging alone is rather limited in its capability as a tumor tracking modality. On the other hand, theoretical investigation on the liver, reported in [39], showed that the knowledge about the position of one or more surrogate 3D points (not the tumor directly) and a statistical motion model of the organ of interest allows predicting the position of the entire organ with good spatial precision.

Recently, a theoretical analysis of a yet novel motion mitigation approach has been proposed in [22]. In this work a so-called self-scanning approach is proposed that does not try to follow the breathing induced tumor motion but rather uses a fixed treatment beam and lets the respiratory motion move the tumor through the treatment beam. It was shown that the approach works surprisingly well for high-frequency focused ultrasound (HIFU). As no active following of the tumor is required, the HIFU transducers could be simplified too, however, at the cost of computationally more complex treatment planning that requires a statistical motion model of the organ.

14.2 4-DIMENSIONAL MR IMAGING

The basis for building statistical respiratory motion models is ground truth motion data. This motion data ideally has to show the moving organs in free-breathing not only with good 3-dimensional spatial resolution but also with a sufficiently high temporal frame rate of a couple of volumes per second. In other words, to build respiratory motion models, 3-dimensional plus time (3D+t) data is needed, or in short 4-dimensional data.

Today’s imaging methods such as US, CT or MRI have limited ability to fulfill these criteria, for different reasons. While the required chest volume to capture is too big for US, there have been approaches proposed to capture 4D CT data [14]. While 4D CT has adequate spatial as well as temporal resolution, only a limited number of breathing cycles can be observed due to the involved X-ray exposure to the patient. This lim-
its the quality and range of suitable data for motion models. MRI on the other hand, does not expose the patient to any harmful radiation allowing to capture and observe the organ motion over extended time intervals. It also shows good spatial resolution, however, this comes at the cost of increased scan time, leading to motion artifacts in the acquired images. While there exist accelerated 2d and volumetric MRI acquisition schemes (e.g. UNFOLD [42], SENSE [43], as well as combinations thereof, among other approaches), they often provide significantly inferior image resolution and signal-to-noise ratio (SNR). To overcome this trade-off between acquisition speed and image quality, learning-based methods have been proposed, among them 4D-MRI. While MRI does not natively provide a high enough spatial resolution and temporal frame rate required for 4D motion modeling, it is possible to retrospectively sort images of different planes that were acquired over tens of minutes into coherent 3-dimensional stacks of different respiratory phases, yielding a 4-dimensional data set with good spatial as well as temporal resolution. This technique is generally known as 4D-MRI [38]. Other 4D-MRI techniques, as well as hybrid ultrasound-MRI based techniques [44] have been proposed since then, but this chapter describes the original technique proposed in [38].

14.2.1 Acquisition Protocol

In contrast to some of the techniques described in the literature a sagittal slice orientation is chosen. Reason being that sagittal slicing allows tracking the vascular structures during complete breathing cycles with minimal out-of-plane motion. Moreover, the sagittal slice orientation requires the smallest field of view and therefore leads to the fastest temporal frame rate.
The volume of interest is covered by sagittal slices (Fig. 14.1A), further called data slices $D_{p,i}$ with $p$ the position of the sagittal slice and $i$ the acquisition time. The key concept of the proposed acquisition scheme is to interleave the acquisition of data slices with a dedicated so-called navigator slice $N_j$ at a fixed spatial position, with $j$ indicating the acquisition time (Fig. 14.1B). These navigator slices will be used retrospectively to derive a gating measure that determines the state of the liver on a certain data frame.

### 14.2.2 Retrospective Stacking

In order to produce sequences of 3D volumes from the acquired 2D images, one has to find those frames which show the liver in a similar state on different planes, in order to stack them together. The corresponding data frames can be determined by comparing the embracing navigator frames that were acquired immediately before and after the respective data frames in the interleaved sequence, see Fig. 14.1B. The underlying assumption is that two data frames show the same state of the liver if their embracing navigator frames are sufficiently similar. Since only navigator frames are directly compared and not the data frames themselves, data frames can be stacked even when showing the liver at distant positions. Moreover, any difference in amplitude, phase or the amount of deformation between data and navigator frames can be handled as long as the state of the liver can be unambiguously defined by the preceding and the subsequent navigator frames.

To describe the similarity measure for navigator slices in detail, a data frame $D_{p,i}$ that was acquired at location $p$ and time $i$ is considered (Fig. 14.2A). Its embracing navigator frames are $N_{i-1}$ and $N_{i+1}$. To find a temporally corresponding data frame at a different slice position $q$, the embracing navigator frames $(N_{j-1}, N_{j+1})$ are compared to the preceding and the subsequent navigator frames $(N_{j-1}, N_{j+1})$ of each acquired candidate frame $D_{q,j}$. Among the many different possible similarity measures, best results were achieved by quantifying the positions of prominent vascular structures using template matching, and using these displacements to find similar navigation frames.
The above 4D-MRI technique allows reconstructing 4D data sets that show detailed deformations of an organ during free breathing, see Fig. 14.2B for an example. While variations beyond a regular breathing cycle cannot be observed in 4D CT, they are addressed specifically with the above method. Irregularities such as drifts and deformations are recognized and handled by the retrospective image-based sorting method. Lastly, in contrast to 4D CT, volunteer studies over long acquisition sessions are possible.

14.3 MOTION MODEL BUILDING

In earlier chapters of this book, we have seen various approaches to model the variation of shapes. In this chapter, we move on to modeling the dynamics of shape changes, which goes beyond standard shape modeling. The dynamics of an object is defined as the short term changes in the shape of that object. In the case of face modeling, one can think of modeling expressions of a facial shape [5,3], which means one is interested in the changes of a facial shape with respect to a reference or neutral face shape. In our case, this would correspond to modeling the respiratory motion of abdominal organ shapes with respect to, say, the corresponding organ shape in exhalation state [37,32].

The mindset is somewhat similar to shape modeling, while a shape model is superimposed by a model of shape changes. However, there are some modeling decisions which have to be taken when modeling shape motion. If the respiratory motion over a population of organ shapes shall be modeled, there are usually a different amount of sample shape changes per subject and therefore one has to think about the underlying distribution and its estimation. Furthermore, the case application may differ from shape modeling or fitting. Usually, respiratory motion models find application in estimating and reconstructing the shape changes based on so-called surrogates, which refers to any signal that allows to infer the shape changes. In our case, surrogates might be model points which are tracked in ultrasound images, or a position signal of a magnetic tracker which is mounted on the epigastrium. We will come to such applications in detail later on.

In the following, we formalize the motion model, where we focus on the application of respiratory motion estimation of the liver, and where motion data for learning is captured from 4D-MRI data.

14.3.1 Statistical Shape Model

One distinguishes between the modeling of shape and the modeling of shape motion. In the shape model, the variation among a population of shapes originating from different individuals is considered, whereas in the motion model, the shape changes over time relative to a reference shape is investigated.

For each subject $s$, there is a finite dimensional reference shape $v^s \in \mathbb{R}^{3n}$ composed of $n$ model points. In our setting, the shape $v^s$ represents the right liver lobe in the
exhalation state and is derived from a representative exhalation MR image of \( s \). For the moment, we assume that the exhalation shapes are in correspondence. Furthermore, the shapes are assumed to be Gaussian distributed \( p(v|v_\mu, \Sigma_v) \sim \mathcal{N}(v_\mu, \Sigma_v) \) where

\[
v_\mu = \frac{1}{S} \sum_{s} v^s, \quad \Sigma_v = \frac{1}{S-1} \sum_{s} (v^s - v_\mu) \otimes (v^s - v_\mu)
\]  

(14.1)

are the maximum likelihood parameter estimates of \( p(v|v_\mu, \Sigma_v) \), \( S \) is the number of subjects and \( \otimes \) is the outer-product. Thus, a shape can be parametrized by

\[
v_\alpha = v_\mu + \sum_{i=1}^{M} \alpha_i \psi_i,
\]

where \( \psi_i \) are orthogonal basis vectors of \( \Sigma_v \) weighted by the model parameters \( \alpha_i \) and \( M \) denotes the number of basis vectors.

### 14.3.2 Statistical Motion Model

In addition to the shape variation among a population, the relative shape changes over time is modeled. Since each subject is observed for a different amount of time \( \tau \), one assumes a mixture of Gaussian distributions over the shape changes \( x \in \mathbb{R}^{3n} \)

\[
p(x) = \sum_{s} p(s)p(x|s) = \sum_{s} \pi^s p(x|s),
\]  

(14.2)

where \( \sum_{s} \pi^s = 1, \pi^s \in (0, 1), \forall s = 1, ..., S \). Each component distribution is assumed to be Gaussian \( p(x^s) \sim \mathcal{N}(x^s_\mu, \Sigma_x^s) \) with

\[
x^s_\mu = \frac{1}{\tau^s} \sum_{t} \tau^s x^s_t, \quad \Sigma_x^s = \frac{1}{\tau^s - 1} \sum_{t} (x^s_t - x^s_\mu) \otimes (x^s_t - x^s_\mu).
\]  

(14.3)

The first two moments of the mixture \( p(x) \) are estimated by

\[
x_\mu = \sum_{s} \pi^s x^s_\mu, \quad \Sigma_x = \sum_{s} \pi^s (\Sigma_x^s + (x^s_\mu - x_\mu) \otimes (x^s_\mu - x_\mu)),
\]  

(14.4)

where \( \pi^s = \tau^s / \sum_s \tau^s \) is the weighting of the component distribution with respect to the number of temporal samples per subject (see more details about finding modes of Gaussian mixtures in [7]). If the number of samples for each subject is equal, Eqs. (14.2)–(14.4) reduce to the case of a single Gaussian.

The variation of the shape changes is finally parametrized by \( x = x_\mu + \sum_{i=1}^{N} \beta_i \phi_i \), where \( \phi_i \) are \( N \) orthogonal basis vectors of \( \Sigma_x \). With the combination of shape and
motion model, a shape to a particular time point can be synthesized by

\[ v^s_{\beta} = v_\mu + \sum_{i=1}^{M} \alpha_i \psi_i + x_\mu + \sum_{i=1}^{N} \beta_i \phi_i, \]  

(14.5)

where \( \alpha_i \) and \( \beta_i \) are coefficients of the shape and the motion model respectively. Note that for certain applications, the first term (the shape model) can be replaced by a fixed shape \( v^s \) which then merely specifies the topology of the motion model. This simpler case will be the starting point for the applications later in this chapter, before applications with shape and motion models are presented.

14.3.3 Inference on Statistical Models

A major application of motion models is the estimation of the present shape change given some observations. Suppose we are given online ultrasound images depicting parts of the changing shape of interest. The goal in this setting would be to estimate the full changing shape given these observed parts based on the already seen shape changes, i.e. the motion model. This can be tailored to observations which are even independent of the topology of the model. Suppose that only the position of a magnetic tracker placed on the epigastrium is given. If the correlation between such a position signal and the motion model is known, the present shape change can be inferred.

14.3.3.1 Sparse Reconstruction

We have seen how to derive linear models of shape changes assuming the samples are Gaussian distributed.\(^1\) In the following, we will use an important property of the Gaussian distribution to infer the shape change of a full shape given sparsely observed changing model points. Before that, we shortly recap Bayes’ theorem for applying it to Gaussian random variables.

Bayes’ Theorem

The problem of reconstructing shape change from partially observed data can be formulated in a probabilistic way. The goal is to estimate the expected shape change, having given only parts of it. Consider the two fundamental rules of probability,

\[ \text{sum rule} \quad p(X) = \sum_Y p(X, Y), \]  

(14.6)

\[ \text{product rule} \quad p(X, Y) = p(Y|X)p(X), \]  

(14.7)

\(^1\) Since only the first two moments of the Gaussian mixture distribution are estimated the resulting distribution reduces to a Gaussian distribution.
where $X$ and $Y$ are two random variables, $p(X, Y)$ is the joint probability of $X$ and $Y$ and $p(Y|X)$ is the conditional probability of $Y$ given $X$. Bayes’ theorem relates the two conditional probabilities $P(X|Y)$ and $P(Y|X)$ applying the sum and product rule

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}. \quad (14.8)$$

As we treat each displacement $x \in X$ at position $x \in v$ as a Gaussian random variable, the joint distribution $p(x)$ is a multivariate Gaussian distribution. We now partition the variables $x$ into two sets where $x_a$ are all variables which are observed and $x_b$ are the remaining variables. Since the marginal distributions $p(x_a)$ as well as $p(x_b)$ are Gaussian the conditional distribution

$$p(x_b|x_a) = \frac{p(x_a|x_b)p(x_b)}{p(x_a)} \quad (14.9)$$

is again a Gaussian and given in closed form. In Section 14.6, we will elaborate how Eq. (14.9) can be applied to the problem of reconstruction from partial information.

### 14.3.3.2 Statistical Model Regression

Let $v$ be an instance of a model according to Eq. (14.5), where the scripts for model coefficients are dropped for simplicity. The goal in model-based reconstruction is to obtain an estimation of the complete shape vector $v$ from a vector of partially observed information $r \in \mathbb{R}^l$, $l < 3n$. Finding the maximum of the posterior distribution $p(v|r)$, or in other words, the probability of $v$ given the partial information $r$, is a regression problem having the statistical model as prior information.

This concept of inferring the shape changes given surrogate data can be drawn even further to surrogates which are not bound to the topology of the motion model. Suppose we are given a position signal of a magnetic tracker which is placed on the epigastrium whose dynamics are naturally correlated to respiration. Let an attribute vector $a \in \mathbb{R}^d$ be such a signal at a particular time point. Consider an observed finite set $A = \{(a_0, \beta_0), \ldots, (a_n, \beta_n)\} \subset \mathbb{R}^d \times \mathbb{R}^N$ of $n$ pairs of i.i.d. vectors $a_i$ and motion model coefficient vectors $\beta_i$. Let further assume that there exists a function $f: \mathbb{R}^d \to \mathbb{R}^N$ which maps the attribute vectors to the coefficient vectors, while we only observe noisy instances of $\beta$ such that $\beta \sim N(f(a), \sigma I)$. The derivation of such a function $f$ is a regression problem as well. However, no statistical model is available as prior and therefore some generic smoothness assumptions are usually taken.

### Gaussian Process Regression

Gaussian process regression is a non-linear regression method where such a generic smoothness prior can be applied. Let $f \in \mathcal{GP}(0, k)$ be a Gaussian process with the
covariance function \( k: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \). Assuming a Gaussian likelihood, the posterior distribution \( p(f|A) \) is given in closed form \cite{33} and is again a Gaussian process \( GP(\mu_A, k_A) \) with

\[
\mu_A(a) = K_{a,A}(K_{A,A} + \sigma^2 I)^{-1} B,
\]

(14.10)

\[
k_A(a, a') = k(a, a') - K_{a,A}(K_{A,A} + \sigma^2 I)^{-1} K_{a',A},
\]

(14.11)

where \( K_{a,A} = (k(a, a'))_{a=1}^n \in \mathbb{R}^n \), \( K_{A,A} = (k(a_i, a_j))_{i,j=1}^n \in \mathbb{R}^{n \times n} \) and \( B = (\beta_0, \ldots, \beta_n)^T \in \mathbb{R}^{n \times N} \). The expectation of an unseen output \( \beta^* \), given an attribute \( a' \), yields Eq. (14.10).

### 14.4 ESTABLISHMENT OF CORRESPONDENCE

Before a population-based statistical model can be built from the 4D-MRI registration results, correspondence must be established between all subjects. In other words, a common topology must be defined for the data to be modeled. Let the vector \( v_s(t) \) describe the liver of subject \( s \) at time \( t \), with points located on the surface as well as within the liver, obtained from image registration of the 4D-MRI data. This can be achieved for example using a B-spline based non-rigid registration algorithm \cite{45, 46}. The dependency on \( s \) and \( t \) hints that a distinction must be made between two types of correspondence: *intra-subject* and *inter-subject* correspondence. Intra-subject correspondence is associated with all time steps \( t \) for a single subject, while inter-subject correspondence concerns the establishment of a common data format, or topology, among all subjects. The concept of correspondence is closely related to image and surface registration, and the latter is typically used to establish correspondence between images or shapes. Automatic registration techniques constitute a large area of research in medical image processing. In fact, automatic registration is of highest importance to the clinical success of statistical models, with potential time and cost savings both for building patient-specific models as well as applying population-based models to unseen subjects. However, fully automatic registration is also a very challenging problem, which is the reason why many approaches rely on manual or semi-automatic techniques.

#### 14.4.1 Inter-Subject Correspondence

The goal for inter-subject correspondence in motion modeling is to establish mechanical correspondence \cite{37}, which means that the correspondence criterion should be based on how similar points move during respiration. As elaborated in the introduction, this motion is driven by the diaphragm and guided by the abdominal wall as a “tube” defining the direction where the organ can move. The process starts at selecting a reference image at full expiration for each subject, here called exhalation master. Manual establishment of correspondence amounts to selecting a number of landmarks that correspond according to the correspondence criterion, on the exhalation master of each subject.
subject. Through further processing of a relatively small number of manually selected points, a dense set of corresponding surface points can be obtained. A binary mask of the organ of interest can be incorporated into the process, as depicted in Fig. 14.3. For volumetric data, this procedure is performed on each slice. This yields a set of surface points $\mathbf{v}_s$ for each exhalation master.

In a next step, all surface shapes have to be aligned, for example using generalized procrustes analysis (GPA) [8], and the mean of the aligned points, $\mathbf{\mu}^c$, can be used to define a regular grid of interior points. In GPA, a rigid transformation $T_s, R_s$ is found for each subject, to move it into a common coordinate system. For each subject $s$, GPA finds a translation and rotation matrix according to

$$
\arg \min_{T_s, R_s} \sum_{i=1}^n \| T_s R_s \mathbf{v}^i_s - \mathbf{\mu}^c \|_2,
$$

(14.12)

to move all samples into a common coordinate system. The “goodness-of-fit” criterion is usually the $l_2$-norm. The procedure is iterative and repeated until convergence. The algorithm outline is the following:

1. Arbitrarily choose a reference shape among all shapes.
2. Superimpose all shapes $\mathbf{v}^i$ to current reference shape.
3. Compute the mean shape $\mathbf{\mu}^c$ of the current set of superimposed shapes.
4. Compute the $l_2$-norm between the reference and all shapes, set reference to mean shape $\mathbf{\mu}^c$ and continue to step 2.
This interior grid can finally be warped to each $\hat{\mathbf{v}}^s$. This way, the mechanical correspondence is transferred from the surface to the inside of the liver, and an exhalation master shape $\hat{\mathbf{v}}$ with points on the surface as well as in the interior is obtained for each subject. Fig. 14.4 depicts the final result of such an approach, on the mean shape as well as on two subjects.

14.4.2 Inter-Subject Correspondence by Registration

The present approach can be refined in a sense that the correspondence among the individual shapes is established automatically by non-rigid shape registration. As we have given for each volunteer a manually segmented label map of the liver structure (cf. Fig. 14.3B) a dense shape $\hat{\mathbf{v}}_i \in \mathbb{R}^{3n}$ can be obtained using standard marching cubes. Correspondence is established by an iterative group-wise registration of the shapes in order to reduce a bias of the mean shape $\bar{\mathbf{v}}$ to a specific exhalation master shape $\hat{\mathbf{v}}$:

$$\bar{\mathbf{v}}_i = \hat{\mathbf{v}}_{\text{mean}}, \quad \bar{\mathbf{v}}_i^{i+1} = \frac{1}{S} \sum_i \bar{\mathbf{v}}_i^j + \Delta \bar{\mathbf{v}}_i^j,$$

(14.13)
where \( m \) is randomly chosen and \( \Delta \mathbf{v}_i \) is obtained by the Gaussian process registration method of [20] such that \( \mathbf{v} + \Delta \mathbf{v}_i \approx \mathbf{ar{v}}_i \). In the following, by \( \mathbf{ar{v}}_i \) we refer to the registered shapes \( \mathbf{v} + \Delta \mathbf{v}_i \), if nothing else is mentioned.

The mean shape \( \mathbf{ar{v}} \) is equidistantly sampled in the inside to add several interior points to the topology. With thin-plate-spline interpolation these points can be transferred to each shape \( \mathbf{v}_i \). To recap, for each subject we have now an exhalation master shape \( \mathbf{ar{v}}_s \), which is in correspondence with the population mean shape \( \mathbf{ar{v}} \). Having given the registration results of the 4DMR images, we will see that for each time point a shape \( \mathbf{v}_i(t) \) can be derived.

### 14.4.3 Intra-Subject Correspondence

Recall that correspondence is linked to both the subjects \( s \) and to time \( t \). The latter means that temporal correspondence must be established along the motion sequence for each subject. However, given the inter-subject correspondence from the previous sections, this is now as simple as warping the points of the master shape \( \mathbf{ar{v}}_s \) to all individual time steps using the deformation field obtained from the non-rigid registration. Using this approach, the location of every point in the master shape is known over time. In other words, dense motion information \( \mathbf{v}_s(t) \) of the liver is obtained for each subject.

The final result of the correspondence step is an exhalation master vector

\[
\mathbf{ar{v}}_s = (\mathbf{ar{x}}_{s,1}, \mathbf{ar{y}}_{s,1}, \mathbf{ar{z}}_{s,1}, \ldots \mathbf{ar{x}}_{s,m}, \mathbf{ar{y}}_{s,m}, \mathbf{ar{z}}_{s,m})^T \in \mathbb{R}^{3m}
\]  

(14.14)

for each subject, as well as all \( n_t \) respiratory states \( \mathbf{v}_s(t) \in \mathbb{R}^{3m}, t \in [1, \ldots, n_t] \) per subject in dense correspondence.

### 14.5 STATISTICAL MOTION MODELING

With all data samples in dense correspondence, principal component analysis (PCA) is used to transform the 4D-MRI registration results into a representation where each dimension is independent and distributed according to a normal distribution. First, we discuss the most common case, where only motion is modeled [39,28–30,32,31]. This amounts to Eq. (14.5) where the shape model term is replaced with a fixed exhalation master \( \mathbf{ar{v}} \),

\[
\mathbf{v}_\beta = \mathbf{\bar{v}} + \mathbf{x}_\mu + \sum_{i=1}^{N} \beta_i \phi_i. 
\]  

(14.15)

Furthermore, often just a single Gaussian component is used for the motion model.

The data used for computing the motion model is the offset between each sample and the exhalation master,

\[
\mathbf{x}_s(t) = \mathbf{v}_s(t) - \mathbf{\bar{v}}_s. 
\]  

(14.16)
The arithmetic mean is estimated using
\[
\mu_s = \frac{1}{n} \sum_{i=1}^{n} x_s(t), \tag{14.17}
\]
and the mean-free data of all subjects is gathered in the data matrix,
\[
X = [x_1(1) - \mu_1, \ldots, x_1(n) - \mu_1, \ldots, x_n(1) - \mu_n, \ldots, x_n(n) - \mu_n] \in \mathbb{R}^{3m \times (n_s \cdot n_t)}. \tag{14.18}
\]

Standard Principal Component Analysis is performed on \(X\), as elaborated in much detail in earlier chapters, to obtain the matrix of eigenvectors \(\Phi = [\phi_1^T, \ldots, \phi_N^T]\), the coefficient vector \(b = (\beta_1, \ldots, \beta_N)^T\) for each sample as well as the mean displacement \(x_\mu\) for the motion model in Eq. (14.15).

### 14.6 BAYESIAN RECONSTRUCTION FROM SPARSE DATA

The probabilistic framework of PCA provides an efficient tool to reconstruct missing information based on the statistics of the modeled data. In machine learning language, this is called a regression problem. Recall Bayes’ rule,
\[
p(Y) = \sum_X p(Y|X)p(X). \tag{14.19}
\]

Let \(X\) be the observation vector and \(Y\) the set of model parameters \(\Theta\). The conditional probability
\[
p(x(\Theta)) \tag{14.20}
\]
is called the likelihood of \(x\) given the set of model parameters \(\Theta\). It describes how likely an observed data vector is for different settings of the parameter vector \(\Theta\). The value \(p(\Theta)\) gives the probability of the model parameters before a sample has yet been drawn and is called prior probability of \(\Theta\). Accordingly, the quantity \(p(\Theta|x)\) gives the probability of a set of model parameters after a sample \(x\) has been drawn and is therefore called posterior probability of \(\Theta\).

Here, the goal is to obtain an estimation of the complete data vector for the organ at time \(t\), \(x_s(t)\), from a vector of partially observed information \(r(t) \in \mathbb{R}^l\), \(l < 3m\). For the ease of notation, the subscripts are dropped and \(x\) and \(r\) are used instead. In [6], a Bayesian approach for model-based reconstruction from partial information is derived as follows. Let \(L: \mathbb{R}^{3m} \mapsto \mathbb{R}^l\), \(l < < 3m\) be a linear transformation matrix that governs the mapping of a complete data vector to its partial observation,
\[
r = Lx. \tag{14.21}
\]
In general, $L$ is not an injective mapping and thus the solution of Eq. (14.21) with respect to $x$ is not uniquely defined. Likewise, we can define the reduced version of the model matrix $\Phi_1$, scaled with the eigenvalues,

$$Q = (q_1, \ldots, q_n) = \mathrm{diag}(\sigma_i) \cdot L \Phi_1 \in \mathbb{R}^{l \times (n-1-k)}.$$  (14.22)

In terms of the unknown model parameter vector $b = (\beta_1, \ldots, \beta_N)^T$, Bayes’ rule becomes

$$p(b | r) = \frac{p(r | b) \cdot p(b)}{p(r)}.$$  (14.23)

If we assume that the measurement is subject to uncorrelated Gaussian noise of variance $\sigma_N^2$, the posterior probability becomes

$$p(b | r) \propto \exp \left( -\frac{1}{2\sigma_N^2} \| Qb - r \|^2 \right).$$  (14.24)

We are looking for the model parameters $b$ that maximize this posterior, which amounts to minimizing the following energy function,

$$E(b) = \| Qb - r \|^2 + \eta \cdot \| b \|^2,$$  (14.25)

where $\eta$ is a regularization parameter that allows to trade-off between the data fitting error and the prior probability of $c$.

After substituting $Q$ with its singular value decomposition $Q = \hat{U} \hat{W} \hat{V}^T$, we obtain

$$\hat{V} \hat{W}^2 \hat{V}^T b + \eta b = \hat{V} \hat{W} \hat{U}^T r,$$  (14.26)

and the following series of transformations leads to the final solution:

$$\hat{W}^2 \hat{V}^T b + \eta \hat{V}^T b = \hat{W} \hat{U}^T r,$$  (14.27)

$$\mathrm{diag}(\eta) \cdot \hat{V}^T b = \hat{W} \hat{U}^T r.$$  (14.28)

$$\hat{V}^T b = \mathrm{diag} \left( \frac{w_i}{\eta \cdot w_i^2 + \eta} \right) \hat{U}^T r.$$  (14.29)

$$b = \hat{V} \mathrm{diag} \left( \frac{w_i}{\eta \cdot w_i^2 + \eta} \right) \hat{U}^T r.$$  (14.30)

The final estimation of the complete data vector is computed according to

$$x = Ub + \mu.$$  (14.31)

The exhalation master shape is added to obtain the absolute position of the organ, $v = x + \tilde{v}$. 

The matrix $Q$ is of size $l \times \hat{N}$, with $\hat{N} \ll N$ the number of principal components used for reconstruction. Therefore, Eq. (14.30) can usually be solved in real time.

### 14.7 APPLICATIONS OF POPULATION-BASED STATISTICAL MOTION MODELS TO MOTION RECONSTRUCTION

In the remainder of this chapter, a number of papers are discussed that leverage the described reconstruction method on 4D-MRI and ultrasound data.

In [28], the Bayesian approach to population-based statistical respiratory motion modeling is described. 4D-MRI liver images were acquired of 12 healthy subjects over roughly one hour on 22 to 30 sagittal slices, at a temporal resolution of 2.6–2.8 Hz. Deformation fields are extracted using non-rigid registration, and a topology for both inter- and intra-subject correspondence are defined using a set of manually selected landmarks on one exhalation master image per subject. The left liver lobe was kept out of the analysis because of the influence of heart motion. From the resulting 12 topologically equivalent 3D liver volumes and their dense spatio-temporal motion data, nine respiratory cycles from the beginning, the middle and the end of the acquisition session are used, each consisting of 8–20 time points. Using this data, 12 leave-one-out population-based models were built. Each of the models was used to evaluate the performance on the left-out subject. Fig. 14.5 shows the first two principal components of the model, and Fig. 14.6A shows a plot of the principal component's cumulative variance. To measure how well the model is able to describe the most extreme respiratory state (full inhalation), the vertices of a typical unseen full inhalation state are projected into each of the leave-one-out models, as a measure of model expressiveness (Fig. 14.6B).

Partially observed data, for example from ultrasound imaging or implanted electromagnetic beacons, is simulated by selecting a point at the inferior tip of Couinaud segment VI, at the diaphragm and near the center of the right liver lobe in each subject. Eq. (14.30) is used to estimate the model parameters of a 20 min sequence for each subject, and the corresponding shape is reconstructed using Eq. (14.31). An average error of 1.2 mm is reported.

Besides quasi-periodic superior/inferior (SI) motion, some data sets further contain secondary modes of motion due to cardiac cycle motion, digestive activity, gravity and muscle relaxation. Some of them cause a drift of the organ. It is shown that including organ drift in the model leads to a significant decrease of the reconstruction error in the presence of such motion, as depicted in Fig. 14.7.

In a clinically realistic scenario, where the goal is to destroy tumor cells under free breathing using motion-compensated radiotherapy, 3D fiducial information is not generally available. More commonly, 2D planar projections of fiducials from a beam-eye view X-ray, portal imager or fluoroscope are acquired in an image-guided therapy setting. Furthermore, there may be a considerable amount of measurement noise which
Figure 14.5 Coronal and sagittal views of the mean liver shape deformed in direction of the first principal component $\phi_1$ (left) and the second principal component $\phi_2$ (right) of our respiratory model. The white and gray surfaces represent deformations of plus and minus 3 standard deviations, respectively. (From [28].)

Figure 14.6 (A) Cumulative variance plot of the statistical population model for the liver in [28]. (B) Median, 25th and 75th percentiles, and maximum (diamond) of the projection error at inhalation. (From [28].)

Figure 14.7 Mean reconstruction error for a 60 min sequence subject to organ drift of up to $d_{avg} = 5.6$ mm (averaged over the entire liver). (A) With drift model that only contains states from the beginning of the acquisition session. (B) Without drift information. (From [28].)
Statistical Shape and Deformation Analysis

Figure 14.8 In [29], three close model vertices were selected. The motion of all grid points within a sphere of radius $r = 15$ mm around their centroid was reconstructed, based on the planar projection of the fiducials. (From [29].)

Figure 14.9 Mean reconstruction error as a function of the target volume radius. Both errors increase only very little, however, the model-based reconstruction is significantly more precise for any size. (From [29].)

directly impacts the reconstruction accuracy of the statistical model. In [29], these constraints are further studied (Fig. 14.8). Leave-one-out models were generated from a superset of the previously studied data, containing a total of 20 subjects. On each exhalation master shape, three close landmarks were manually selected, to serve as fiducials in the experiments. For each set of landmarks, two-dimensional planar projections onto a sagittal plane were generated from the motion data over 5 min. Additionally, Gaussian noise ($\sigma = 2$ mm) was superimposed on the projections, to simulate the uncertainties in a more realistic setting. From these 2D input points, the motion of target volumes of 10 to 40 mm radius around the centroid of the three fiducials was reconstructed over 5 min of respiratory motion, for each subject. To show that the non-rigid motion from the statistical model reconstruction is an improvement over a simple model, the same sequence was also reconstructed using a rigid body motion model, as depicted in Fig. 14.9. On average, the statistical model outperformed the rigid model by more than 1 mm. The next experiment investigated the influence of the camera projection angle with respect to the patient, as shown in Fig. 14.10A. It was found that projections
Figure 14.10 (A) In [29], a large number of camera positions was simulated to study its importance for motion reconstruction. (B) This surface shows the mean error over all subjects as a function of the camera position. It can be nicely seen that the reconstruction accuracy drops significantly when the elevation \( \epsilon_l \) approaches the values \( +90^\circ \) and \( -90^\circ \). (From [29].)

perpendicular to the main direction of respiratory motion yield better reconstructions. Conversely, if the elevation angle gets closer to \(+90^\circ\) or \(-90^\circ\) degrees, the observable motion almost disappears in the projected image, and leads to much higher reconstruction errors. Fig. 14.10B depicts the error as a function of the azimuth and elevation angle.

Another study investigated the difference between 3D, 2D or only 1D surrogate information available to the reconstruction algorithm [30]. While the ability to obtain motion information in 3D would be the preferred case, it is also the most unrealistic of all, since implanted electromagnetic beacons are not a clinically established technique. More commonly, T1-weighted MR images are acquired in an MR-guided high intensity focused ultrasound (MRgHIFU) setting for example, since such images are necessary for temperature monitoring, and are as such readily available for tracking purposes. In addition to 2D images, a 1D pencil beam signal of the diaphragm might additionally be acquired at a fast rate, and small cost in scan time, to increase the temporal resolution of the surrogates. In other scenarios where temperature monitoring might not be needed, such as conventional radiotherapy, 2D ultrasound imaging might be a more cost-efficient method for feature tracking. All these possible settings were simulated in a number of experiments where 3D, 2D or only 1D fiducial information is made available to the reconstruction algorithm, and the location of all model vertices was estimated based on this information [30]. A single fiducial at the diaphragm was manually selected on the 4D-MRI navigator image for each of the 20 subjects. Normalized cross-correlation was used to track this position in 20 min worth of navigator images. Fig. 14.11 depicts the navigator slice of one subject together with the tracked positions of the diaphragm. The results are depicted in Fig. 14.12 and show that even when only 1D surrogate information is available, the statistical model is able to recon-
Figure 14.11 Selection of tracked organ position using template matching that was used as input for reconstruction (A). Close-up view (B). (From [30].)

Figure 14.12 Average errors for reconstruction using a single point near the diaphragm in 1D, 2D and 3D. In the 3D case, the actual 3D position of the fiducial was used instead of the tracking results. (From [30].)

struct all model vertices with an average error of less than 3 mm for most subjects. The study also considers a gated scenario. In gating, a radiation or imaging device is only triggered when the respiratory state is close to a known reference position, typically full expiration, to minimize the influence of respiratory motion. Fig. 14.13 shows the reconstruction errors as a function of the gating window size. It can be seen that even for a smaller gating window size, the difference between gating only and gating combined with model-based reconstruction is significant.

All previously described approaches rely on synthetic fiducials ultimately generated from the acquired MR images, in one form or another. To address the question how well such a population-based statistical motion model performs on surrogates from 2D ultrasound images, a modified, MR-compatible phased array US transducer [26] is used in [32] to acquire images of the liver under free breathing inside the MR bore, while 4D-MRI is acquired simultaneously (Figs. 14.14,14.15). The ultrasound images are used
Figure 14.13 Comparison of reconstruction error with amplitude gating. Model-based gating outperforms traditional amplitude gating for any gating window size. (From [30]).

Figure 14.14 Ultrasound tracking of five cycles for two subjects. (A) Tracked points of subject 1 are distributed mainly along the liver surface, as the available acoustic window did not expose many vessel structures inside the organ. (B) Trajectories for subject 4 could be extracted at the diaphragm and also for vessels within the liver. The acoustic shadow occluding about half of the liver is a typical artefact due to absorption by the ribs. (From [32]).

for feature tracking, providing the surrogate signal for the reconstruction algorithm, whereas 4D-MRI provides the ground-truth motion for validation.

A mean error of 2.4 mm (standard deviation of 1.7 mm) is reported over all 8 subjects, validated on the 4D-MRI ground-truth data acquired simultaneously. In addition, two observers manually annotated three vessel locations in the MR images, and the prediction error was additionally validated using this manual ground-truth. The reason for this is that the 4D-MRI registration results might itself be biased due to the inaccuracy of image registration. An average error of 1.85 mm (standard deviation 1.2 mm) is reported using the manual ground-truth. In addition, it is shown how a statistical mo-
Figure 14.15 Overview of the experimental setup. (A) US probe accommodated in a gel-filled bag and its dedicated holder (EM shielded). (B) US probe attached to an MR-compatible orbital ring using an articulated handler. (C) US probe accommodated in a gel-filled bag and its dedicated holder (EM shielded). (D) Subject with US probe orbital ring inside the MR scanner. (From [32].)

A motion model can be combined with a temporal predictor, in order to estimate the entire organ from the same small number of surrogates at time $t + t_\Delta$, where $t_\Delta$ considers the typical system lag of a radiosurgery device, which is reported to be at least in the order of 100 ms. A neural network is used to learn the mapping of the US surrogates from $t$ to $t + t_\Delta$ from a short sequence of tracking data, and the forward-predicted surrogates are used to reconstruct the entire organ at $t + t_\Delta$. Errors between 2.3 mm ($t_\Delta = 50$ ms) and 2.7 mm ($t_\Delta = 400$ ms) are reported.

All previous approaches model respiratory motion independent of the temporal evolution of the signal. The advantage is that the reconstruction is independent of the specific breathing pattern, such as frequency and amplitude. On the other hand, the information about signal evolution can be incorporated into the reconstruction, as it is done in [31], where the presented Bayesian reconstruction concept is incorporated into a bilinear model that models and reconstructs entire respiratory cycles, rather than individual respiratory states.
14.8 RECONSTRUCTION BY REGRESSION

We now move on to topologically independent surrogates for reconstructing the full shape changes [17]. This has the advantage that one can reduce the involving ultrasound imaging and interest point tracking to a lighter setup. Given an attribute signal which is correlated with the organ motion, the goal is to estimate the motion model parameters $\beta$ of the current respiratory state with Eq. (14.10). Based on that, a shape change can be synthesized using Eq. (14.5).

Following [17], first a high-temporal resolution respiratory cycle is synthesized in order to give a feeling about the capabilities of motion model regression. This is followed by the evaluation of the estimation performance of this method.

In both experiments, a straightforward Gaussian process model is used, applying a Gaussian kernel as covariance function

$$k_\text{G}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\theta_1^2}\right),$$

(14.32)

where $\theta_0$ is a scaling parameter and $\theta_1$ is a length scale or smoothness parameter.

14.8.1 Average Breathing Cycle

We first show a study presented in [17] where respiratory motion of the liver in general was analyzed. A motion model out of the motion samples among $V = 9$ volunteers was built, while 99.9% of the variance was kept. For each temporal sample shape $S_t$ at time point $t$, an attribute $\epsilon \in [0, 2\pi]$ was considered which indicates the cycle state of $t$ within a respiratory cycle. Such a rather abstract attribute can be applied to synthesize an average respiratory cycle of the liver shape. For the regression, 8000 pairs of cycle attributes resp. motion model coefficients were randomly picked among all volunteers.

In Fig. 14.16, an example right liver lobe and its displacements within this average respiratory cycle is visualized. Note that here a semantic and non-linearly captured respiratory cycle of a shape is synthesized, where not simply the most dominant principal component of the motion model is varied. Thus, for each patient an average respiratory cycle can be generated e.g. for planning [22]. While the source 4DMRI has a framerate of 2.8 Hz the temporal resolution can now be upsampled to an arbitrary high framerate. In this example, 100 samples were synthesized which corresponds to approximately 25 Hz.

---

2 This cycle attribute was computed using a greedy cycle detection algorithm which is based on the average vertical coordinates of the shape changes $\mathbf{x}_t$. 
14.8.2 Motion Model Prediction

In the motion estimation experiment of [17] a surrogate signal was simulated which indicates the depth of the diaphragm measured from the abdominal skin for example by a 1D US sensor. A 1D signal was defined which is generated by a ground truth model point in the region of the diaphragm. Let \( s : [0, \tau] \rightarrow \mathbb{R}^3 \) be the 3D signal of absolute coordinates of this point at time point \( t \in [0, \tau] \). To get invariant to the absolute positioning of the subject, let us project the signal into its dominant mode of variation

\[
\mathcal{F}[s] = (s - \mu_s) \psi_0 + \epsilon, \tag{14.33}
\]

where \( \mu_s = \frac{1}{\tau} \int_0^\tau s(t) \, dt \) is the signal mean value, \( \epsilon \sim \mathcal{N}(0, \sigma_\epsilon) \) is additive noise, which was set to \( \sigma_\epsilon = 2 \) mm and \( \psi_0 \) is the orthonormal eigenfunction corresponding to the largest eigenvalue \( \lambda_0 \) of the equation \( \int_0^\tau \text{cov}(s_i, s_j) \psi_0(s_i) \psi_0(s_j) \, ds_i = \lambda_0 \psi_0(s_j) \). Here, \( \text{cov} \) is the covariance function of the signal \( s \).

In this evaluation, the motion estimation performance of the method can be shown given the simulated signal \( \mathcal{F}[s] \). For each volunteer, a leave-one-out (L1O) motion model was generated, where only motion samples from the other volunteers were considered. 99.9% of the variance was kept yielding L1O models of 20 to 22 principal modes. Note that for each temporal sample, additionally \( \mathcal{F}[s] \) was computed for the
Figure 14.17 For each L1O experiment, the average corresponding point difference between the ground truth and the predicted shape is visualized. We compare our method Attribute Regression with the Conditional Model \([2,28]\). The upper x-axis indicates how many time points the motion has been predicted.

later usage as an attribute (Eq. (14.33)). As a shape model, the model with 2571 surface and 368 interior points was used which had been constructed as described in Section 14.4.2. To derive the corresponding model parameters, the shapes of the volunteers are projected into the shape model.

For the Gaussian process regression, 8000 \(F[s]\)-attributes resp. ground truth coefficient vector pairs were randomly picked, again only from the other volunteers. The kernel parameters were set to \(\theta_1 = 3\), while the exact value of \(\theta_0 = 5000\), had only minor influence to the estimation performance. In Fig. 14.17, for each volunteer the average estimation error is plotted. The estimation error is robustly kept below 5 mm, whereas the median stays around 2 to 3 mm. For radiotherapy these are reasonable error bounds.

Let us compare the topology-independent method to the sparse reconstruction method presented above where the simulated 3D point signal \(s\) serves as surrogate data. The estimation is performed by estimating the mean of a statistical motion model which is conditioned on \(s\), cf. Eq. (14.9) and \([2]\). For a fair comparison, we added to \(s\) an isotropic Gaussian noise \(\mathcal{N}(0, \sigma^2/3)\). The conditional model performs equally well, while it better generalizes in the experiment with volunteer 7 and 9. Certainly, the topology-independent model with only 9 volunteers is not capable to generalize to the respiratory motion of these two subjects. This can be confirmed when comparing to the results with patient specific models (Fig. 14.18). Here, for each volunteer, a motion model was built where only samples from the volunteer of interest were considered. For the regression, 700 attribute/coefficient pairs were randomly picked and the kernel parameter was adjusted to \(\theta_1 = 5\). For all volunteers including for volunteer 7 and 9, the average error was considerably improved to less than 1 mm. In Fig. 14.16 on the right,
a patient specific average respiratory cycle is plotted for a comparison to the population mean cycle.

14.9 CONCLUSION

In this chapter, an overview of population-based statistical motion models for respiratory motion was laid out. The 4D-MRI imaging technique was introduced first, along with registration techniques for extracting the motion information from subjects over several minutes. This data is the foundation for the discussed motion models that nicely represent large amounts of respiratory motion data in a compact mathematical framework. The statistical nature of these models allows to reconstruct the entire organ from only a small number of surrogate signals, be it from 1D signals such as a respiratory belt or a pencil beam, from 2D signals such as single slice imaging or projections, or from 3D signals such as implanted electromagnetic beacons. Both manual and automatic techniques for establishing correspondence among a number of subjects were discussed, as well as the Bayesian framework for sparse reconstruction, and a number of applications from the literature were presented. Finally, a generalization based on Gaussian processes was presented that eliminates the need to obtain a direct mapping between surrogates and their corresponding model vertices, allowing for topology-independent reconstruction.

ACKNOWLEDGMENTS

This work was funded by the Swiss National Science Foundation projects P300P2-164647 and CRSII2-127549.

REFERENCES


