Inference of hysteretic respiratory tumor motion from external surrogates: a state augmentation approach

D Ruan[1], **J A Fessler**[1], **J M Balter**[2], **R I Berbeco**[3], **S Nishioka**[4] **and H Shirato**[5]

[1] Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI, USA
[2] Department of Radiation Oncology, The University of Michigan, Ann Arbor, MI, USA
[3] Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[4] Department of Radiation Oncology, NTT East Japan Sapporo Hospital, Sapporo, Japan
[5] Department of Radiation Medicine, Hokkaido University School of Medicine, Sapporo, Japan

E-mail: druan@eecs.umich.edu

Abstract

It is important to monitor tumor movement during radiotherapy. Respiration-induced motion affects tumors in the thorax and abdomen (in particular, those located in the lung region). For image-guided radiotherapy (IGRT) systems, it is desirable to minimize imaging dose, so external surrogates are used to infer the internal tumor motion between image acquisitions. This process relies on consistent correspondence between the external surrogate signal and the internal tumor motion. Respiratory hysteresis complicates the external/internal correspondence because two distinct tumor positions during different breathing phases can yield the same external observation. Previous attempts to resolve this ambiguity often subdivided the data into inhale/exhale stages and restricted the estimation to only one of these directions. In this study, we propose a new approach to infer the internal tumor motion from external surrogate signal using state augmentation. This method resolves the hysteresis ambiguity by incorporating higher-order system dynamics. It circumvents the segmentation of the internal/external trajectory into different phases, and estimates the inference map based on all the available external/internal correspondence pairs. Optimization of the state augmentation is investigated. This method generalizes naturally to adaptive on-line algorithms.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Respiratory motion affects tumors in the thorax and abdomen. In particular, breathing is the major reason for intrafractional tumor motion for lung cancer patients. It is important to

monitor such motion during radiotherapy treatment to ensure the accurate delivery of radiation dose in motion-compensated intensity modulated radiotherapy (IMRT). Fluoroscopic imaging or portal imaging can monitor tumor motion during the treatment process. To reduce x-ray exposure, hybrid tumor tracking approaches that combine episodic radiographic imaging and continuous external surrogates have been investigated widely (Ozhasoglu and Murphy 2002, Murphy 2004, Murphy *et al* 2002, Schweikard *et al* 2000, 2004). Using external surrogates to infer internal tumor motion assumes that there is consistent relationship between the internal and external motions.

Hysteresis is typical in lung tumor movements, with the tumor taking a different path during inhale and exhale. Inhalation normally takes longer than exhalation, and the deflating lung volume exceeds the inflating volume at the same trans-pulmonary pressure (Keall *et al* 2006). Respiratory hysteresis makes inferring internal tumor locations from external surrogate signals challenging. Most of the external surrogate systems, such as thermistors, thermocouples, strain gauges, pneumotachographs (Kubo and Hill 1996) and infrared skin markers as applied in the Varian Real-time Position Management™ (RPM) system (Varian Medical Systems, Palo Alto, CA), provide one-dimensional signals, whose instantaneous amplitude (or displacement) alone does not provide sufficient information about the specific breathing stages.

Previous studies about correspondence between the internal tumor motion and external surrogates can be classified into two categories. One class of studies investigates the correlation between the two signals to justify the feasibility of using certain types of surrogates, or compare different surrogate options (including the placement mechanism) (Wade 1954, Vedam *et al* 2003, Ahn *et al* 2004, Hoisak *et al* 2004, Tsunashima *et al* 2004, Koch *et al* 2004, Mageras *et al* 2004). Alternatively, some other studies assume *a priori* the existence of a strong correlation between the internal and external signals, and aim to estimate the correspondence map (Seppenwoolde *et al* 2007). We adopt the latter perspective and study with a general setup the correspondence maps that take the external surrogate trace as the input and output estimates of the internal tumor location, including, but not restricted to linear relations as reflected by the correlation coefficient and its variants. The presence of respiratory hysteresis makes this a challenging problem, as the same external surrogate position can reflect different internal tumor locations during different phases. Existing methods address hysteresis by first separating empirically the breathing trajectories into two distinct 'directions' (inhale versus exhale), and then constructing a piecewise phase-dependent map (Seppenwoolde *et al* 2002, 2007, Low *et al* 2005, Lu *et al* 2005). However, subdividing the breathing into inhale and exhale phases often requires manual intervention, and is infeasible for real-time application, because a breathing 'peak' or 'trough' can only be identified retrospectively.

In this study, we propose to use a simple state augmentation of the external surrogate signal. Augmenting the state space with self-delayed observation bestows the model with 'memory', which is an alternative way to characterize the 'path-dependence' property of hysteretic systems. This procedure captures system dynamics, and embeds the breathing-phase information implicitly into the framework. We then provide the solution to a general class of parametric inference models with the augmented observations. As special cases, we derive optimal solutions for the parameters of linear and quadratic correspondence models. Furthermore, given a training internal/external dataset, we demonstrate a computationally efficient approach to choose a patient-specific (or fraction-dependent) augmentation scheme. Generalization to adaptive correspondence models follows naturally. We test the proposed approach on synchronized recordings of internal gold marker trajectories and external fiducial marker locations (Berbeco *et al* 2005).

Table 1. Description of study participants. Patients 1–3 were brought in for data acquisition purposes only, so there is no prescription dose. Patient 5 was treated twice at the same site, with two months between treatments. The tumor site is indicated using the common anatomical notation for lung segmentation: S1-3 is the upper lobe, S4-5 is the middle lobe and S6-10 is the lower lobe.

Patient	Gender	Age	Tumor Pathology	No of bb's	Tumor site	Prescribed dose (Gy)	No of fractions
1	F	47	Adenocarcinoma	4	R S7	N/A	1
2	F	70	Adenocarcinoma	3	L S6	N/A	1
3	F	71	Adenocarcinoma	2	R S5	N/A	1
4	F	47	Adenocarcinoma	3	R S4	48	8
5	M	81	Squamous cell carcinoma	3	R S6b	48	4
5						40	8
6	M	61	Small cell lung cancer	3	R S10	40	8
7	M	68	Squamous cell carcinoma	3	R S6	48	4
8	M	85	Adenocarcinoma	3	R S8	48	4

Section 2 describes the clinical data used for this test, discusses the challenges caused by hysteresis in converting the external surrogate position directly to the internal tumor location and presents the proposed method. A general correspondence model is formulated with polynomial models as an example. The optimal model parameters are derived and a generalization is given to accommodate adaptivity. Section 3 reports testing results followed by discussions. Section 4 concludes this study with a brief summary.

2. Methods and materials

2.1. Data description

To study the internal/external motion correspondence, we obtained synchronized recordings of internal tumor motion trajectories and external fiducial marker locations. The paired trajectories from eight lung cancer patients were collected with a Mitsubishi real-time radiation therapy (RTRT) system at the Radiation Oncology Clinic at the Nippon Telegraph and Telephone Corporation (NTT) hospital in Sapporo, Japan. Two to four 1.5 mm diameter gold ball bearings (bb's) were implanted in or near the tumor (Shirato *et al* 2003) and these internal markers were tracked in real time with diagnostic x-ray fluoroscopy (Shirato *et al* 2000). External surrogate signals were obtained with the AZ-733 V external respiratory gating system (Anzai Medical, Tokyo, Japan) integrated with the RTRT system. It uses a laser source and a detector, both attached to the treatment couch with the beam placed orthogonal to the patient's abdominal skin surface. The device calculates the change in the surface amplitude by measuring the relative position of the reflected light (Berbeco *et al* 2005) and outputs a one-dimensional relative position measurement of the abdominal surface. The data acquisition rate for the entire system is 30 frames per second. Table 1 describes the study participants. All patients included in this analysis had the peak-to-peak marker motion greater than 1 cm. The KV fluoroscopy + Anzai system took multiple readings for each fraction from several treatment field configurations to account for obscured x-ray views as the gantry rotated. The recording lengths varied between 20 s and 250 s with an average of 82 s. There are in total 128 readings, 46 of which were longer than 100 s.

2.2. A general correspondence model

To minimize diagnostic imaging dose in IGRT systems, it is important to infer the internal tumor location from external surrogates. In principle, we could use a correspondence model

that observes a trajectory $\vec{r}$ of the scalar external surrogate r up to the time instant n to infer the three-dimensional internal tumor position $\boldsymbol{p} = (x, y, z)$. We denote the collective surrogate information available at time n as $\vec{r}(n) \triangleq \{r(m) : 0 \leqslant m \leqslant n\}$. However, it is challenging to estimate such a map that estimates the internal tumor position from the complete collection of historical surrogate data, since the length of the input variable grows to infinity as the time progresses. A more practical choice is to use some much more compact quantity $\boldsymbol{r}$ that captures sufficient information from $\vec{r}$ for inference. With internal and external motions both being smooth, it is reasonable to approximate $\boldsymbol{p}(\boldsymbol{r})$ using polynomials. Therefore, we focus on estimating a class of correspondence models that are linear in their coefficients as follows:

$$\hat{\boldsymbol{p}}(\boldsymbol{r}) = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \tag{1}$$

where $\boldsymbol{f}$ is a vector function of the external surrogate $\boldsymbol{r}$; all model parameters to be optimized are contained in the coefficient matrix $\boldsymbol{A}$. In particular, two simple correspondence models, i.e. a linear model and a quadratic model introduced in Seppenwoolde *et al* (2007), are special cases of the form given in equation (1).

Linear models assume each coordinate of internal motion is affine in $\boldsymbol{r} = r(t)$. This corresponds to the case where

$$\boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r \\ 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A} = \begin{bmatrix} b_x & c_x \\ b_y & c_y \\ b_z & c_z \end{bmatrix}. \tag{2}$$

Quadratic models map the external surrogate to each coordinate of internal motion via a quadratic relation. It can be expressed in equation (1) with

$$\boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r^2 \\ r \\ 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A} = \begin{bmatrix} b_x & c_x & d_x \\ b_y & c_y & d_y \\ b_z & c_z & d_z \end{bmatrix}. \tag{3}$$

Expression equation (1) is linear in the model coefficients $\boldsymbol{A}$ and yields a closed-form optimal solution in the least-squared error (LSE) sense. Given N sample points $(\boldsymbol{r}_n, \boldsymbol{p}_n), n = 1, 2, \ldots, N$, the solution to the LSE problem

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} E(\boldsymbol{A}), \tag{4}$$

where $E(\boldsymbol{A}) = \sum_{n=1}^{N} ||\boldsymbol{p}_n - \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}_n)||^2$, is given by solving the normal equation (Luenberger 1969) and

$$\hat{\boldsymbol{A}} = \boldsymbol{P}^T \boldsymbol{F} (\boldsymbol{F}^T \boldsymbol{F})^{-1}, \tag{5}$$

where

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{r}_1)^T \\ \vdots \\ \boldsymbol{f}(\boldsymbol{r}_N)^T \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1^T \\ \vdots \\ \boldsymbol{p}_N^T \end{bmatrix}.$$

The corresponding residual is given by

$$\begin{aligned} \Delta \boldsymbol{P} &\triangleq \boldsymbol{P} - \boldsymbol{F}\hat{\boldsymbol{A}}^T \\ &= (\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}^T \boldsymbol{F})^{-1} \boldsymbol{F}^T)\boldsymbol{P}, \end{aligned} \tag{6}$$

with the overall residual error (summed over all three dimensions) as

$$\begin{aligned} E(\hat{\boldsymbol{A}}) &= \text{trace}\{\Delta \boldsymbol{P}^T \Delta \boldsymbol{P}\} \\ &= \text{trace}\{\boldsymbol{P}^T (\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}^T \boldsymbol{F})^{-1} \boldsymbol{F}^T)\boldsymbol{P}\}. \end{aligned} \tag{7}$$

It may be preferable to have simpler models (with fewer free parameters) over more complicated models at the cost of small sacrifice in data fitting performance. This model selection preference can be incorporated into the optimization setting by modifying the objective function as

$$\tilde{E}(\boldsymbol{A}) = E(\boldsymbol{A}) + \lambda R(\#\boldsymbol{A}), \tag{8}$$

where $\#\boldsymbol{A}$ denotes the number of free parameters in the coefficient matrix $\boldsymbol{A}$ and R is a monotonically increasing function that assigns higher costs to more complicated models. The regularization weight λ controls the tradeoff between the data fitting $E(\boldsymbol{A})$ and the preference for lower-order models. A simple example of R would be the linear function $R(\#\boldsymbol{A}) = \#\boldsymbol{A}$, which directly penalizes the number of components in $\boldsymbol{A}$; this is equivalent to the Akaike Information Criterion (Akaike 1974). Using the closed-form optimal solution equation (5) and the expression for optimal residual error equation (7) for a given fixed model structure, the modified objective function can be minimized in two layers. We say two inference models have the same *model structure* if they only differ in parameter values. It follows immediately that all models with the same structure has equal number of degrees of freedom, thus the same complexity regularization $R(\#\boldsymbol{A})$ in equation (8). Therefore, to minimizer over models of different complexity, it is natural to choose the 'best' parameter setting within each model structure (with fixed degrees of freedom thus a constant complexity penalty), and then compare across structures. Within each class, the minimizer of the complexity-penalized objective $\tilde{E}(\tilde{A})$ is the same as that of $E(\tilde{A})$, and can be solved and evaluated efficiently using the closed-form optimal solution equation (5) and expression for optimal residual error equation (7). This motivates the two-layer hierarchical algorithm shown below for finding the optimal solution within K candidate model structures $\mathcal{C} = \cup_{i=1}^{K}\{C_i\}$.

Algorithm 1. Two-layer optimization routine for solving $\hat{A} = \arg\min \tilde{E}(\boldsymbol{A})$ (8).

1: $\tilde{E} \leftarrow +\infty$; $i_{\text{opt}} \leftarrow 0$; $\tilde{\boldsymbol{A}} \leftarrow []$.
2: **for** $i = 0$ to K **do**
3:     Choose model structure C_i from the collection of models $\mathcal{C}$,
4:     Compute $R_i = R(\#\boldsymbol{A})$ for structure C_i;
5:     Compute $\hat{\boldsymbol{A}}_i$ within class C_i according to (5) and its residual error $E(\hat{\boldsymbol{A}}_i)$ from (7).
6:     **if** $E(\hat{\boldsymbol{A}}_i) + R_i < \tilde{E}$ **then**
7:         $\tilde{E} \leftarrow E(\hat{\boldsymbol{A}}_i) + R_i$;
8:         $i_{\text{opt}} \leftarrow i$;
9:         $\tilde{\boldsymbol{A}} \leftarrow \hat{\boldsymbol{A}}_i$.
10:    **end if**
11: **end for**

2.3. Hysteresis and state augmentation

Conventional methods that explicitly segment the breathing process into inhale and exhale phases have their limitations, as physical phase transitions (and delays) occur continuously rather than as discrete jumps. To circumvent the intrinsic difficulty of estimating breathing phases, we study the system dynamics directly, expecting them to sufficiently convey phase information. In a discretely observed system, one usually captures the system dynamics with time-lagged samples. For the sake of simplicity and to avoid over-parameterization, we restrict this study to a single lag. The proposed method generalizes to multiple-lag models naturally.

Given a discrete-time external surrogate $r(n)$, $n = 1, 2, \ldots, N$, we augment each external surrogate state with a time τ (in discrete unit) delayed sample, i.e. $\boldsymbol{r}(n) \stackrel{\triangle}{=} (r(n), r(n-\tau))$. This augmentation captures first-order system dynamics, as the difference between $r(n)$ and $r(n-\tau)$ can be regarded as a measure of the average local velocity. As $\boldsymbol{r}$ is uniquely determined by $\vec{r}$, it fits into the general formulation equation (1). We apply the methods provided in section 2.2 to estimate the coefficients for the augmented model. To demonstrate the idea, we establish a linear model that is comparable to equation (2) and a quadratic model analogous to equation (3).

The augmented linear model (in $\boldsymbol{r}$) represents each internal coordinate as a linear combination of $r(n), r(n-\tau)$ and a constant offset, corresponding to

$$\hat{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \qquad \text{where} \quad \boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r(n) \\ r(n-\tau) \\ 1 \end{bmatrix} \tag{9}$$

with a 3×3 coefficient matrix $\boldsymbol{A}$.

The augmented quadratic model (in $\boldsymbol{r}$) estimates each internal coordinate as a linear combination of $r^2(n), r(n)r(n-\tau), r^2(n-\tau), r(n), r(n-\tau), 1$, corresponding to

$$\hat{\boldsymbol{p}} = \boldsymbol{A}\boldsymbol{f}(\boldsymbol{r}), \qquad \text{where} \quad \boldsymbol{f}(\boldsymbol{r}) = \begin{bmatrix} r^2(n) \\ r^2(n-\tau) \\ r(n)r(n-\tau) \\ r(n) \\ r(n-\tau) \\ 1 \end{bmatrix} \tag{10}$$

with a 3×6 coefficient matrix $\boldsymbol{A}$.

In both cases, linearity in $\boldsymbol{A}$ results in the closed-form solution given by equation (5) with the corresponding $\boldsymbol{F}$, respectively.

2.4. Choice of lag length

The delay τ should be chosen properly, since too long a lag provides minimal local dynamic information and too short a lag makes the estimation sensitive to observation noise. For inference purposes, we desire a lag that maximally resolves the ambiguity in the estimated correspondence map. We choose the lag that minimizes the fitting error for training data:

$$\hat{\tau} = \arg\min_{\tau} E(\hat{\boldsymbol{A}}(\tau)), \tag{11}$$

with the objective function E defined in equation (4). The coefficients $\hat{\boldsymbol{A}}$ and the error E depend on τ because $\boldsymbol{f}$ contains both the current external surrogate displacement $r(n)$ and its lagged state $r(n-\tau)$.

Equations equation (6) and equation (7) provide a closed-form expression for $E(\hat{A}(\tau))$ for each given τ. The optimization problem equation (11) simplifies to a simple one-dimensional line search that we solve by searching over an interval with the corresponding delay time between 0 (no lag) and about half of an average breathing period.

2.5. Adaptivity of the correspondence map

Adaptivity may be useful to accommodate gradual changes in the correspondence models, due to drifting or variations in patients' breathing. In the case of linear and quadratic models, the operation in equation (5) involves inverting fairly small matrices (3×3 and 6×6, respectively),
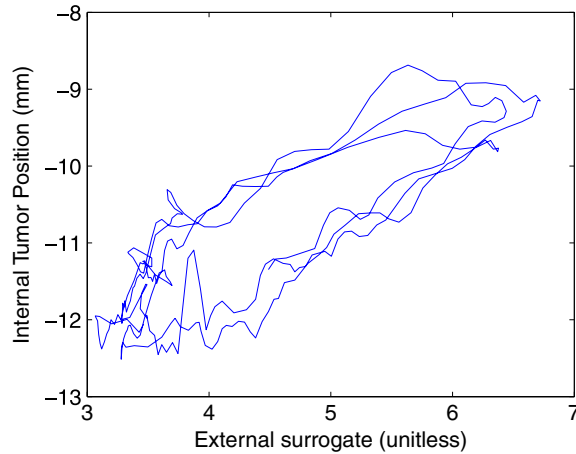
Figure 1. Example of a breathing trajectory with respiratory hysteresis.
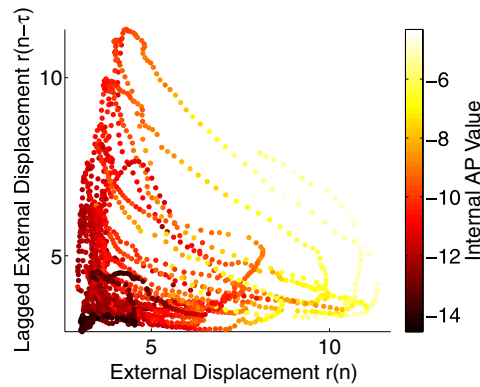

Figure 2. Scatter plot showing the data samples in augmented external state space with the colors indicating internal AP value. Locally consistent colored samples suggests the potential of resolving hysteretic ambiguity by distinguishing among different respiratory phases implicitly with state augmentation.

so direct inversion is numerically feasible. However, when more complicated models with higher degrees of freedom are used, it is desirable to reduce computation by applying recursive algorithms that modify current estimates based on newly available data. The key to recursively updating equation (5) is to avoid recomputing $(\boldsymbol{F}^T \boldsymbol{F})^{-1}$ from scratch every time. This is effectively the inversion of empirical correlation matrix with the observation $\boldsymbol{f}_i$. (Ruan *et al* 2008) provide rank-one update equations for sliding window and exponential discount adaptivities.

3. Results and discussions

To illustrate the challenges caused by hysteresis, figure 1 shows an example of the relationship between the internal tumor location obtained by fluoroscopic imaging and an external surrogate from an abdominal surface measurement as described in section 2.1. We depict only the anterior–posterior (AP) coordinate against the surrogate signal, as this axis demonstrates
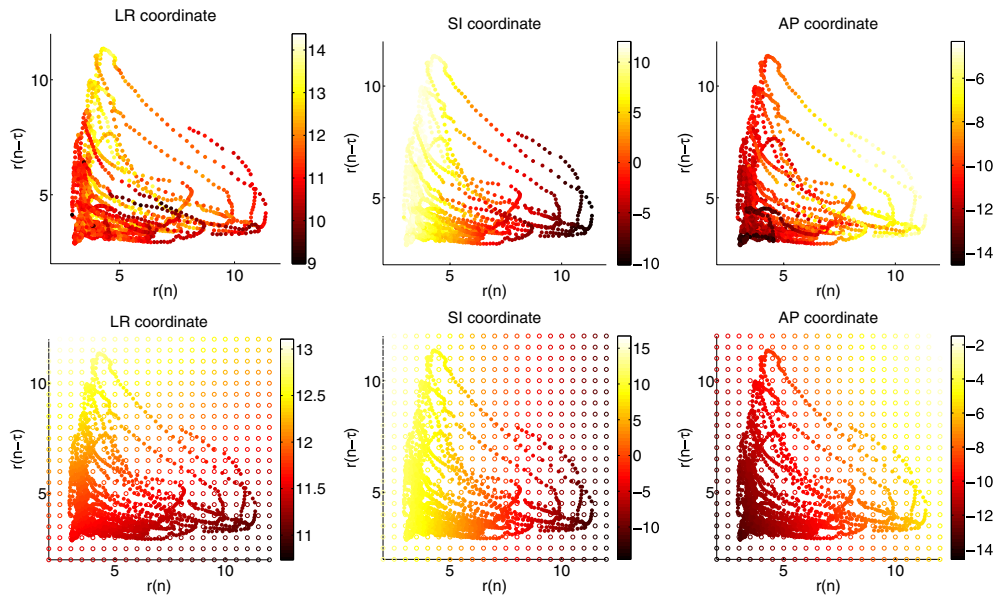
Figure 3. Correspondence relations in augmented state space and their linear fittings. Upper row: the internal tumor coordinate versus augmented state for observed samples with colors indicating internal AP value; bottom row: estimates of the tumor coordinate via linear fit with hollow circles depicting modeled hypersurface evaluated at regular grid points and solid circles for the evaluation at the sample locations, with colors indicating the estimated AP value.
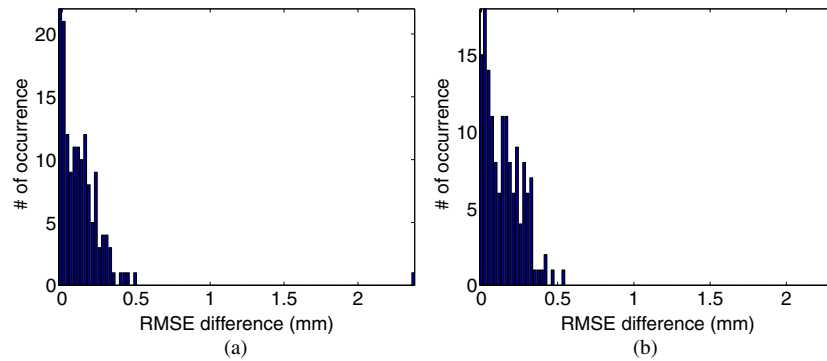

Figure 4. Histogram of paired differences between the RMS errors of the direct and augmented methods: (a) difference between the RMSE of the direct linear approach and augmented linear approach; (b) difference between the RMSE of the direct quadratic approach and augmented quadratic approach.

the strongest hysteresis for this test subject. The optimal linear and quadratic correspondence maps (Seppenwoolde *et al* 2007) provide a reasonable inference of the internal tumor motion from external surrogates, yet they fail to describe the breathing-phase dependency of an ideal correspondence map. In fact, any function that tries to map the scalar $r(n)$ to p would experience the same problem, since this is a one-to-multiple relation with hysteresis.

Figure 2 illustrates the internal tumor location in the anterior–posterior (AP) direction versus the state augmented external surrogates for $\tau = 45$, which corresponds to a 1.5 s delay
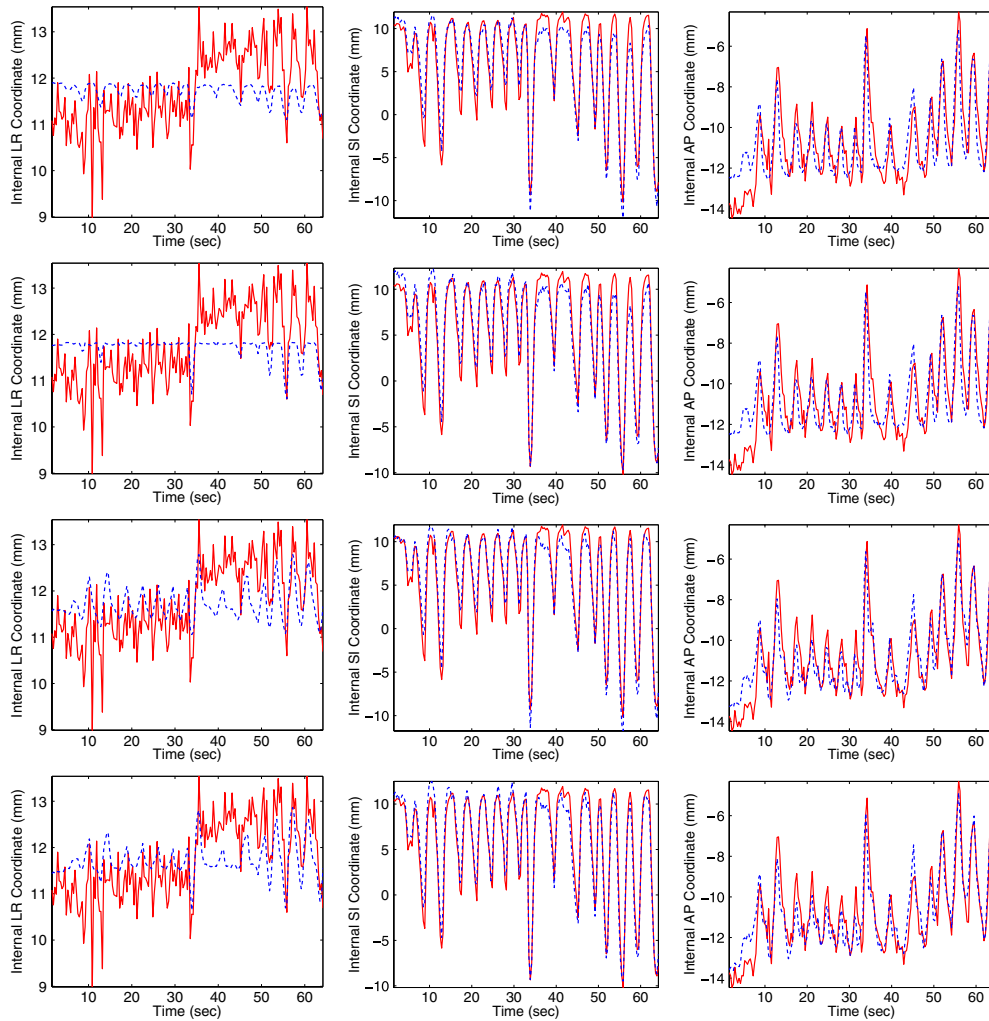
Figure 5. Estimation performance comparison among different methods. Red-solid line depicts the internal tumor position obtained from fluoroscopic imaging, and dashed-blue line provides estimated quantities from external surrogates. Each column represents one internal motion coordinate. Each row indicates the time series generated with one estimation method: (1st row) direct linear; (2nd row) direct polynomial; (3rd row) augmented linear; (bottom row) augmented polynomial.

for 30 Hz sampling rate. The scatter-plot in figure 2(a) represents each data sample in the $(r(n), r(n − \tau))$ space with a circle, and uses color (or intensity if viewed in the gray scale) to depict the internal AP coordinate values (in mm) from fluoroscopic readout. The one-to-multiple discrepancy appears largely resolved as different colored circles are not overlaid on each other, suggesting the existence of a single-valued inference map.

To illustrate the idea of model fitting in augmented state space, we first apply the simple linear model in equation (9) to the dataset shown in figure 1 with a lag length of 1.5 s (which may not be optimal), and illustrate the results in figure 3. Even though there are still noticeable differences between the observed internal coordinates in the upper row of figure 3 and their linear fit in the bottom row, the aggregated estimation error (across all patients and fractions)
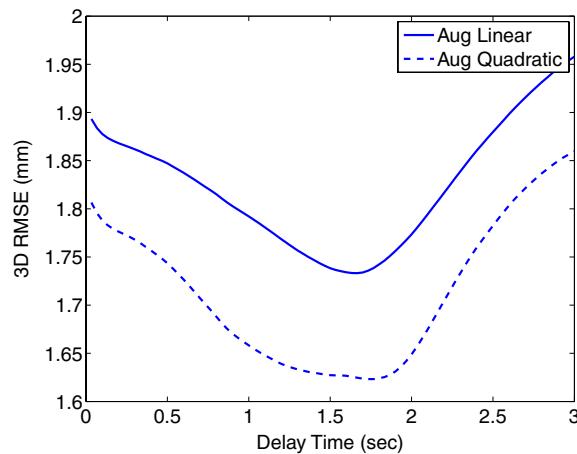
Figure 6. Estimation error as a function of lag length for state augmentation: linear fit (solid line); quadratic fit (dashed line).

reduced to 1.74 mm from 2.01 mm with direct linear fitting as in equation (2) and 1.93 mm with direct quadratic fitting as in equation (3). In particular, we observe noticeable decreases in estimation error in the AP direction, where hysteretic ambiguity is the most significant. Table 2 reports the root mean-squared error (RMSE) in each direction, respectively, for the linear and quadratic models, with and without state augmentation[6]. Figure 4 reports the paired (across patient/fraction) differences between the RMS error of the direct methods and the augmented methods. The RMSE difference between the direct linear and augmented linear methods has a mean of 0.14 mm and a median of 0.11 mm; the RMSE difference between the direct quadratic and augmented quadratic methods has a mean of 0.17 mm and a median of 0.15 mm. To assess statistical significance, we performed a paired Student's *t*-test with the null hypothesis that the performance of the direct and augmented methods do not differ. The *p* values for the linear method and the quadratic method are 4.96×10^{-13} and 4.08×10^{-18}, respectively, demonstrating that the error reductions were statistically significant.

Figure 5 shows the estimated time series of these four approaches for converting external surrogates to internal tumor locations. The higher-order models were more descriptive with the extra degrees of freedom, as demonstrated by the relative performance of quadratic models and linear models within each class respectively. State augmentation enables varying response patterns during different stages of breathing as indicated implicitly by the system dynamics.

As discussed in section 2.4, to properly choose the lag length , we use a short training set with internal–external pairs to compute offline the estimation performance $E(\hat{A}(\tau))$ defined in equation (4) as a function of the lag length τ. In practice, the lag length does not have to be the exact optimum in equation (11); values near that optimum should sufficiently convey system dynamics. Reasonable insensitivity in the choice of lag length τ is desirable as this value is determined prior to the treatment and remains fixed subsequently. Figure 6 illustrates that the estimation error is a smooth function of the lag length, which suggests the desired robustness.

---

[6] For comparison purposes, we have also computed estimate from the fifth-order polynomial model with direct method, which has the same degrees of freedom (18 parameters) as the augmented quadratic model. Its estimation error is 0.75, 1.25 and 1.11 (mm) in LR, SI and AP direction respectively, with a 3D RMSE equals 1.83 mm. A paired Student's *t*-test between the RMSE for the fifth-order polynomial model and the augmented quadratic model yields a *p*-value of 1.06×10^{-10}, which indicates statistically significant error reduction by the augmented quadratic model. This shows that the improved performance of the proposed method is not a direct consequence of increased degrees of freedom, but should rather be attributed to its capability of resolving hysteretic ambiguity via state augmentation.
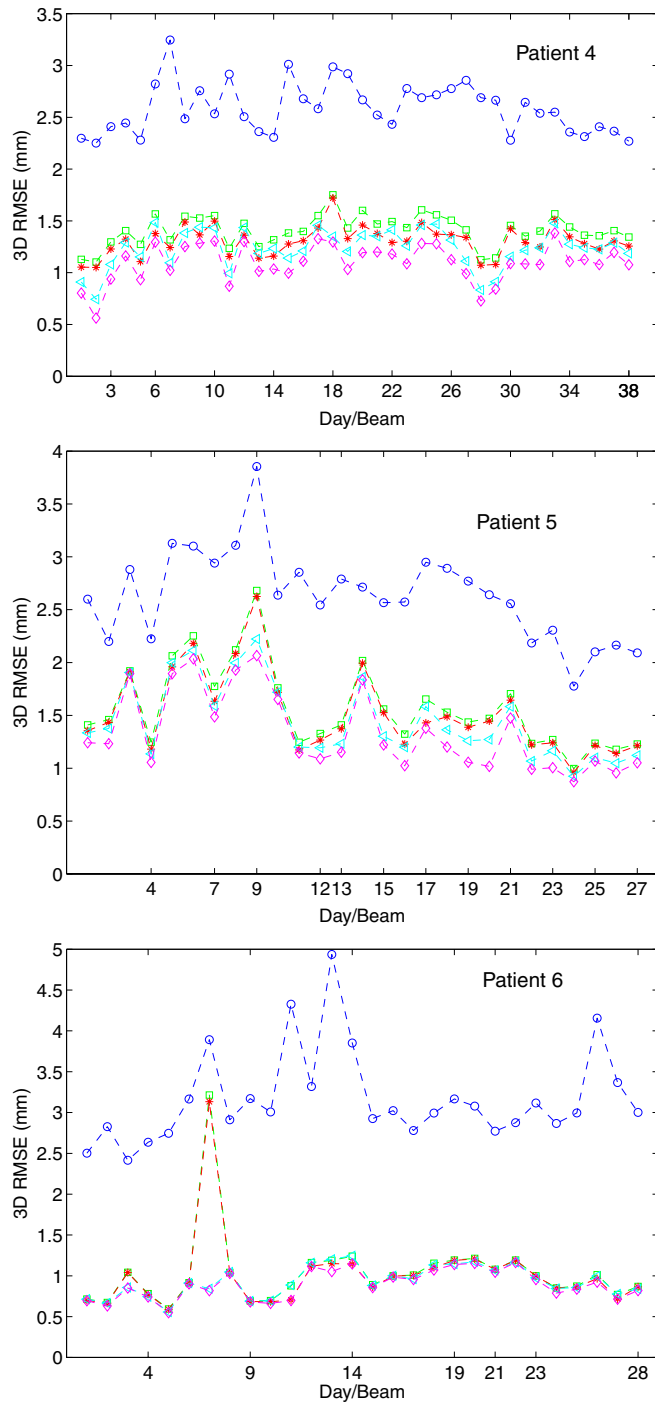
Figure 7. Beam-wise 3D RMSE (mm) for patients 4–6: minimum non-surrogate (blue circle-dashed); linear inference (green square-dashed); polynomial inference (red star-dashed); augmented linear inference (cyan triangle-dashed); augmented polynomial (magenta diamond-dashed). Non-uniform tick locations along the *x*-axis indicate the number of beams applied to each individual on the treatment day.

Table 2. Estimation error table.

	LR (mm)	SI (mm)	AP (mm)	3D (mm)
Direct linear	0.80	1.45	1.13	2.01
Direct quadratic	0.79	1.35	1.13	1.93
Aug. linear	0.75	1.30	0.87	1.74
Aug. quadratic	0.74	1.18	0.84	1.63

For both the linear correspondence model equation (9) and the second-order polynomial model equation (10) with state augmentation, the optimal τ corresponds to about 1.7–1.8 s delay. Without this knowledge, our previous experiments used 1.5 s delay to augment the state space (figures 2–5), and still yielded plausible results. The asymmetric slopes in figure (11) around the optimal $\hat{\tau}$ suggest that it may be preferable to use a relatively small time delay in the absence of precise information.

Assuming that the choice of lag length is robust to inter-patient and inter-fraction variations, we used a fixed lag length equivalent to 1.5 s delay for simplicity, and illustrate in figure 7 the beam-wise 3D RMSE for patients 4, 5 and 6, whose treatment extended over multiple days. The minimum RMS error for non-compensated treatment, which corresponds to a constant estimate at the retrospective mean value, is also shown for reference purposes. These results confirm that the augmented methods consistently exhibit lower error.

Adaptivity is most beneficial for irregular respiration traces. Our test data had relatively regular breathing patterns, so the inclusion of adaptivity improved the estimation accuracy only slightly.

4. Conclusion and future work

We have proposed a method to map external surrogate signals to internal tumor positions. Breathing-phase-dependent response patterns due to hysteresis are incorporated implicitly by using a simple state augmentation technique to capture system dynamics. We introduced a general class of correspondence models that are linear in model parameters, with the linear and quadratic (in external surrogate) models as special cases. We described closed-form expressions for both the optimal model parameters and the corresponding error value. Based on the latter, we further investigated the proper choice of lag length in state augmentation, and argued its relative robustness. Test results on clinical data demonstrated reduced inference error over the direct linear and polynomial models.

The number of degrees of freedom in a correspondence model determines the trade-off between flexibility and robustness. We seek a model that is descriptive enough to fit the data without undesired sensitivity to observation noise, also known as 'overfitting'. The proposed method may have more degrees of freedom than previous methods due to state augmentation. On the other hand, because it incorporates the breathing-stage information implicitly, it can use all the available internal–external correspondence pairs, without subdividing the training data as required for piecewise models (Seppenwoolde *et al* 2007, Lu *et al* 2005, Low *et al* 2005). In principle, using all the data may compensate for the possible increased sensitivity caused by the extra flexibility. The choice among different complexity levels in augmented models is still open. Both the number of augmentations and the model degree contribute to the overall complexity. Further studies should investigate methods for properly penalizing model complexity based on information criteria as explained in section 2.2.

Many research groups have observed phase shifts between external surrogate signals and internal tumor motions (Chi *et al* 2006, Ford *et al* 2003). Typically, this phase shift was to be

avoided to obtain higher internal–external correlation. However, it is possible to compensate for consistent phase shift, to simplify and improve the correspondence map estimation. In particular, we can artificially synchronize the internal–external phase by shifting one of them according to a constant offset estimated from training data. We will further study phase-offset estimation and its use in external–internal inference in the future.

This work is a preliminary study to validate the existence of a reasonably simple correspondence map and the possibility to estimate it with high accuracy. In practice, internal–external pairs are obtained at a much slower rate. Correspondence maps must be extracted from sparse imaging data and applied to continuously obtained external surrogate signals to estimate the internal tumor locations. Our method can serve as a critical module in this overall framework, yet intensive simulations and validations are further required.

Even though our test data did not exhibit dramatic improvements when using adaptive model estimation, model updates in response to changes are necessary in general. Pursuing this direction requires more thorough analysis of breathing motion variations, change detection and model adaptive rate.

Acknowledgments

References

Ahn S, Yi B, Suh Y, Kim J, Lee S, Shin S, Shin S and Choi E 2004 A feasibility study on the prediction of tumor location in the lung from skin motion *Br. J. Radiol.* **77** 588–96

Akaike H 1974 A new look at the statistical model identification *IEEE Trans. Autom. Control* **19** 716–23

Berbeco R I, Nishioka S, Shirato H, Chen G T and Jiang S B 2005 Residual motion of lung tumors in gated radiotherapy with external respiratory surrogates *Phys. Med. Biol.* **50** 3655–67

Chi P M, Balter P, Luo D, Mohan R and Pan T 2006 Relation of external surface to internal tumor motion studies with cine CT *Med. Phys.* **33** 3116–23

Ford E C, Mageras G S, Yorke E and Ling C C 2003 Respiration-correlated spiral CT: a method of measuring respiratory-induced anatomic motion for radiation treatment planning *Med. Phys.* **30** 88–97

Hoisak J D, Sixel K E, Tirona R, Cheung P C and Pignol J P 2004 Correlation of lung tumor motion with external surrogate indicators of respiration *Int. J. Radiat. Oncol. Biol. Phys.* **60** 1298–306

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

Koch N, Liu H H, Starkschall G, Jacobson M, Forster K, Liao Z, Komaki R and Stevens C W 2004 Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: part I. Correlating internal lung motion with skin fiducial motion *Int. J. Radiat. Oncol. Biol. Phys.* **60** 1459–72

Kubo H D and Hill B C 1996 Respiration gated radiotherapy treatment: a technical study *Phys. Med. Biol.* **41** 83–91

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

Lu W *et al* 2005 Quantitation of the reconstruction quality of a four-dimensional computed tomography process for lung cancer patients *Med. Phys.* **32** 890–901

Luenberger D G 1969 *Optimization by Vector Space Methods* (New York: Wiley)

Mageras G S *et al* 2004 Measurement of lung tumor motion using respiration-correlated CT *Int. J. Radiat. Oncol. Biol. Phys.* **60** 933–41

Murphy M J 2004 Tracking moving organs in real time *Semin. Radiat. Oncol.* **14** 91–100

Murphy M J, Jalden J and Isaksson M 2002 Adaptive filtering to predict lung tumor breathing motion during image-guided radiation therapy *Proc. 16th Int. Congr. on Computer-assisted Radiology and Surgery* pp 539–44

Ozhasoglu C and Murphy M J 2002 Issues in respiratory motion compensation during external-beam radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1389–99

Ruan D, Fessler J A and Balter J M 2008 Mean position tracking of respiratory motion *Med. Phys.* **35** 782–92

Schweikard A, Glosser G, Bodduluri M, Murphy M J and Adler J R 2000 Robotic motion compensation for respiratory movement during radiosurgery *Comput. Aided Surg.* **5** 263–77

Schweikard A, Shiomi H and Adler J 2004 Respiration tracking in radiosurgery *Med. Phys.* **31** 2738–41

Seppenwoolde Y, Berbeco R I, Nishioka S, Shirato H and Heijmen B 2007 Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study *Med. Phys.* **34** 2774–84

Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque J V and Miyasaka K 2002 Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 822–34

Shirato H *et al* 2000 Physical aspects of a real-time tumor-tracking system for gated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **48** 1187–95

Shirato H *et al* 2003 Feasibility of insertion/implantation of 2.0 mm diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **56** 240–7

Tsunashima Y, Sakae T, Shioyama Y, Kagei K, Terunuma T, Nohtomi A and Akine Y 2004 Correlation between the respiratory waveform measured using a respiratory sensor and 3D tumor motion in gated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **60** 951–8

Vedam S S, Kini V R, Keall P J, Ramakrishnan V, Mostafavi H and Mohan R 2003 Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker *Med. Phys.* **30** 505–13

Wade O L 1954 Movement of the thoracic cage and diaphragm in respiration *J. Physiol.* 193–212