Optimal Surface Marker Locations for Tumor Motion Estimation in Lung Cancer Radiotherapy

Bin Dong

Department of Mathematics, The University of Arizona, Tucson, AZ, 85721-0089, USA, and Center for Advanced Radiotherapy Technologies, University of California San Diego, La Jolla, CA 92037-0843, USA

E-mail: dongbin@math.arizona.edu

Yan Jiang Graves, Xun Jia and Steve B. Jiang

Center for Advanced Radiotherapy Technologies and Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA 92037-0843, USA

Abstract. Using fiducial markers on patient's body surface to predict the tumor location is a widely used approach in lung cancer radiotherapy. The purpose of this work is to propose an algorithm that automatically identifies a sparse set of locations on the patient's surface with the optimal prediction power for the tumor motion. In our algorithm, it is assumed that there is a linear relationship between the surface marker motion and the tumor motion. The sparse selection of markers on the external surface and the linear relationship between the marker motion and the internal tumor motion are represented by a prediction matrix. Such a matrix is determined by solving an optimization problem, where the objective function contains a sparsity term that penalizes the number of markers chosen on the patient's surface. Bregman iteration is used to solve the proposed optimization problem. The performance of our algorithm has been tested on realistic clinical data of four lung cancer patients. Thoracic 4DCT scans with 10 phases are used for the study. On a reference phase, a grid of points are casted on the patient's surface (except for patient's back) and propagated to other phases via deformable image registration of the corresponding CT images. Tumor locations at each phase are also manually delineated. We use 9 out of 10 phases of the 4DCT images to identify a small group of surface markers that are most correlated with the motion of the tumor, and find the prediction matrix at the same time. The 10th phase is then used to test the accuracy of the prediction. It is found that on average 6 to 7 surface markers are necessary to predict tumor locations with a 3D error of about 1mm. It is also found that the selected marker locations lie closely in those areas where surface point motion has a large amplitude and a high correlation with the tumor motion. Our method can automatically select sparse locations on patient's external surface and estimate a correlation matrix based on 4DCT, so that the selected surface locations can be used to place fiducial markers to optimally predict internal tumor motions.

Keywords: tumor tracking, surface marker, sparse optimization

Submitted to: *Phys. Med. Biol.*

1. Introduction

Modern radiotherapy techniques, such as Intensity Modulated Radiation Therapy (IMRT), are capable of delivering highly conformal radiation dose to a cancerous target while sparing critical structures and normal tissues. Intra-fraction tumor motion caused by patient respiration, however, may lead to geometric miss of the target and hence potentially compromise the efficacy of these techniques while treating tumors at lung or upper abdomen area. To mitigate this problem, a number of techniques have been developed, such as gated treatment, for which accurate modeling and prompt prediction of tumor motion are necessary (Jiang, 2006*b*; Jiang, 2006*a*).

Tumor localization methods can be generally categorized according to the locations of surrogates. Methods using internal surrogates, such as gold markers implanted in or near tumor, are accurate but have issues like the risks of pneumothorax for lung cancer patients (Arslan et al., 2002; Geraghty et al., 2003), marker migration (Nelson et al., 2007), and the extra imaging radiation dose (Jiang, 2006*b*). In contrast, external surrogate based tumor localization is usually noninvasive and radiation free. In such methods, a regression model is first built between the coordinates of some empirically selected external surrogates and those of the tumor using a training data set. Such a model will be utilized to predict the tumor location using the real-time measurements of the marker locations during a treatment via, for example, Cyberknife Synchrony system (Accuray Corporate, Sunnyvale, CA, USA) (Pepin et al., 2011). Yet, the accuracy of this method usually relies on the correlation between external marker motion and internal tumor motion for a particular patient (Hoisak et al., 2004).

In fact, there are a few questions one should keep in mind while using external markers for tumor tracking. First of all, how many external markers are necessary? While using more markers may potentially provide more comprehensive information for tumor location estimation, it is evident that the motion of points on a patient surface is strongly correlated and information from many surface markers is likely to be redundant. Clinically, it is necessary and desirable to use a minimum number of markers to predict the tumor motion to a satisfactory degree. Second, given the number of markers, where shall we optimally place them? Despite a lot of studies regarding the patient breathing pattern and the selection of marker locations(Yan et al., 2006; Wu et al., 2008), markers are placed empirically in most clinical practice.

In this study, we will attempt to answer the aforementioned two questions utilizing a sparse optimization approach. Specifically, our objective is to choose a sparse set of points from all the points on the front surface of a patient, so that a linear motion model yields the smallest error in tumor location prediction. With a novel optimization model to formulate this objective in a clean and precise mathematical language, as well as an effective numerical algorithm to solve the problem, we can effectively yet efficiently identify the key surface points used to predict tumor motion. A linear regression model is also developed during the optimization process, such that those markers collaboratively predict tumor locations to a satisfactory extent.

2. Methods and Materials

We start with an introduction of some notations. Denote $Y \in \mathbb{R}^{3\times m}$ as a $3 \times m$ matrix whose column vectors are the three Cartesian coordinates of the center of the tumor at various times t_j with $j = 1, 2, \ldots, m$. Suppose there are k candidate surface points available for tumor motion prediction. We denote the coordinates of the collection of all of those surface points at a given time t_j as a column vector $X_j = [\vec{x}_1(t_j), \vec{x}_2(t_j), \ldots, \vec{x}_k(t_j)]^T$, where each vector $\vec{x}_i = [\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}]$ for $i = 1, ..., k$ contains three entries corresponding to the three Cartesian coordinates of the point i. If we assemble all the collections of markers X_j associated with different time t_j, we will have the following matrix $X := [X_1, X_2, \cdots, X_m] \in \mathbb{R}^{3k\times m}$.

2.1. Optimization Model

Assume, for simplicity, there is a linear motion model that relates the external marker motion and the tumor motion. Mathematically speaking, there exist a matrix $A \in \mathbb{R}^{3\times 3k}$ such that $AX \sim Y$. Note that the columns of the matrix A can be also associated to those k surface points, each with three coordinates. If one column of the matrix A is non-zero, the corresponding coordinate for that surface point is then utilized to predict the tumor motion. As it is our purpose to select only a few surface points for tumor motion prediction, the problem can be casted as finding a matrix A with only a few non-vanishing columns, such that the motion of tumor recorded in Y can be accurately characterized by AX. Although this is simply a linear motion prediction model, our numerical experiments indicate that such an assumption is reasonable and leads to accurate tumor location estimations. We shall refer to the problem of *optimal marker selection as* the problem of finding the *linear dependence* of the motion of the internal tumor with the motion of some *sparsely* selected markers.

We propose our *optimal marker selection model* as follows:

$$\min_A \{\|A\|_{2,1} : \ AX = Y\}, \tag{1}$$

where $\|A\|_{2,1} := \sum_j \left(\sum_i a_{i,j}^2\right)^{\frac{1}{2}}$ and $A = (a_{i,j})$. In this optimization problem, the objective function is defined in such a way that it groups all the matrix elements in a column of A utilizing an ℓ_2-norm and then takes ℓ_1-norm among all columns. Minimizing such an objective function term enables us to enforce sparsity at only the level of matrix columns. This idea is inspired by that of compressed sensing (Candes et al., 2006; Candes and Tao, 2006; Candes and Tao, 2005; Donoho, 2006), which is a recent revolutionary concept in information theory. The applications of such a $\ell_{2,1}$ norm has been recently explored in many problems, such as beam orientation optimization for IMRT(Jia et al., 2011), to effectively select only a few groups of elements. Similar idea was also used in (Esser et al., 2011) where the $\ell_{1,\infty}$ norm was used for matrix factorization with applications in hyperspectral image unmixing. We remark that the model (1) not only sparsely selects markers needed to track the motion of an internal

tumor, but also provides the linear dependence of the motion of the selected markers with that of the tumor at the same time. All such information is integrated within the solution matrix A.

2.2. Fast Numerical Algorithm

To solve the proposed optimization problem (1), we use a Bregman distance-based algorithm proposed by Yin *et. al.* (Yin et al., 2008), which is proven to be efficient for ℓ_1 minimization problems. Given matrices X and Y, the fast algorithm that solves (1) can be written into an iterative form as:

$$\begin{aligned} A^{k+1} &= \arg\min_A \left\{\mu\|A\|_{2,1} + \tfrac{1}{2}\|AX - Y^k\|_F^2\right\}, \\ Y^{k+1} &= Y^k + Y - A^{k+1}X, \end{aligned} \tag{2}$$

where k is the iteration index and $\|\cdot\|_F$ is the Frobenius norm. The optimization problem in the first subproblem of (2) can be solved using the proximal forward-backward splitting algorithm (Combettes and Wajs, 2006; Hale et al., 2007), which by itself is an iterative algorithm as:

$$A^{p+1} = \mathcal{T}_\mu(A^p - \delta(A^pX - Y^k)X^T), \tag{3}$$

where p is the iteration index in this subproblem and $\mathcal{T}_\mu(B)$, for a given matrix $B = [B_1, B_2, \ldots, B_m]$, is defined as

$$\mathcal{T}_\mu(B) := \left[\max(|B_1| - \mu, 0)\frac{B_1}{|B_1|}, \cdots, \max(|B_m| - \mu, 0)\frac{B_m}{|B_m|}\right].$$

We note that (Donoho, 1995; Wang et al., 2007) $\mathcal{T}_\mu(B)$ is the closed form solution to $\min_X \left\{\mu\|X\|_{2,1} + \frac{1}{2}\|X - B\|_F^2\right\}$. For computation efficiency, we shall not solve the subproblem (2) accurately by using numerous iterations of (3), but only use one iteration instead. Now, applying (3) (with only one iteration) to (2), we have the following fast algorithm that solves (1) (also known as the Bregmanized operator splitting algorithm (Zhang et al., 2010)):

Algorithm 1 Optimal Marker Selection Algorithm

Step 0. Initialization: $k = 0$, $A^0 = 0$ and $Y^0 = \bar{Y}$ where $\bar{Y} = Y - y_m$ with $y_m \in \mathbb{R}^3$ being the mean vector of the columns of Y.

while stopping criteria is not satisfied **do**

Step 1.

$$A^{k+1} = \mathcal{T}_\mu(A^k - \delta(A^pX - Y^k)X^T)$$

Step 2.

$$Y^{k+1} = Y^k + Y - A^{k+1}X$$

end while

The proof of the mathematical properties of this algorithm, such as convergence, is beyond the scope of this paper. Interested readers can consult references for more details(Yin et al., 2008; Zhang et al., 2010).

For realistic patient data, because of the presence of noise and the fact that the motion of internal tumor is only approximately linearly dependent on the external markers, we should not expect the relative residual $\|A^kX-Y\|_F/\|Y\|_F$ decrease to 0. In fact, numerically we observe that the relative residual should have a lower bound whose value depends on X and Y and it is very difficult to estimate beforehand. Therefore, we adopt the following stopping criteria:

$$\frac{\|A^kX-Y\|_F}{\|Y\|_F}<\epsilon_1 \quad \text{or} \quad \frac{\|A^{k-1}-A^k\|_F}{\|A^k\|_F}<\epsilon_2.$$

In other words, we fix an ϵ_1 as a satisfactory amount for the residual; meanwhile, if such residual is not attainable, we will terminate the algorithm when A^k is not changing too much according to the tolerance ϵ_2.

When Algorithm 1 is implemented we choose $\delta=\frac{2}{\|X^\top X\|_2^2}$ which makes sure the stability of the algorithm (see for example (Yin et al., 2008)). The parameter μ has an influence mainly on the speed of the convergence, but not much on the quality of the results. In our numerical experiments, we fix it to be 0.1. The parameters ϵ_1 and ϵ_2 in the stopping criteria controls the quality of the results. ϵ_1 has a major effect on the results, while ϵ_2 is to make sure the algorithm terminates after a reasonable time when ϵ_1 is too small. In practice, since there is always noise present in the data, we do not want $AX=Y$ to be satisfied exactly. Otherwise, we would be end up with choosing much more markers than necessary, just to compensate on apparent motion caused by noise. Therefore, ϵ_1 and ϵ_2 cannot be too small. However, if ϵ_1 and ϵ_2 are too large, the algorithm may select fewer markers, but the prediction errors will be large as well. Therefore, ϵ_1 and ϵ_2 (mostly ϵ_1) are the parameters that balance between number of markers and prediction accuracy. In our numerical experiments, we choose $\epsilon_2=10^{-7}$ and $\epsilon_1=0.11$, 0.08, 0.03 and 0.0555 for patient 1-4. The values of ϵ_1 are specifically chosen for optimal prediction errors. In practice, it is not possible to optimize ϵ_1 for each patient to minimize the prediction error. One potential approach to overcome this difficulty is to let the algorithm run till a fixed number of markers are selected. The targeting number of markers, e.g. $5\sim7$ should be clinically practical and yet still be able to predict tumor motion to a satisfactory degree. Another approach is to pre-set the ϵ_1 value to the mean optimal value determined in this or any following patient studies. The resulting predicted motion may not be optimal for a particular patient but it should be accurate enough for clinical practice. It is our future work to investigate these practical issues.

2.3. Patient Data

To validate our algorithm with realistic clinical cases, 4DCT scan data of four lung cancer patients is used. For those patients, a four-slice GE LightSpeed CT scanner

(GE Medical Systems, Milwaukee, WI, USA) was used to acquire the 4DCT data for treatment simulation. Each axial CT slice has a thickness of $2.5mm$ and the 4DCT was obtained using respiratory signals from the Varian RPM system (Varian Medical Systems, Inc., Palo Alto, CA, USA). The 4DCT scan consists of ten different phases of one breathing cycle; and the CT volume at each respiratory phase consists of 100 to 144 slices of CT images covering the most of thorax area depending on patients. Each slice of CT image has 512×512 pixels, with a pixel size of $0.977 \times 0.977mm^2$. For each patient, tumor GTV was manually contoured on 4DCT scan images of ten respiratory phases by an expert observer and the 3D tumor center coordinates were identified.

Meanwhile, the external surfaces of each patient, excluding the patient's back, at each phase are extracted by segmenting the CT images using a simple threshold method. For each patient, the CT image volume at the end of inhale is set as the target image; the other nine CT image volumes, corresponding to the other nine different respiratory phases, are set as moving images. The correspondence between surfaces at different phases is established by deformable image registration (Thirion, 1998; Gu et al., 2010). When surface points are available on the external surfaces of each patient, we further sub-sampled the point sets uniformly to reduce the total number of candidate points for a better computational efficiency. In our experiments, we choose approximately 200 candidate surface points for each patient.

2.4. Validation

To validate our marker selection algorithm, we employ an leave-one-out cross validation (LOOCV) method. Specifically, 10 tests are performed for a patient, and for each test, we single out one of the 10 respiratory phases and use the other 9 to form the matrix Y and solve for the matrix A using Algorithm 1. We then validate our method by using the matrix A to predict the location of the tumor at the phase that has been singled out. The deviation of the predicted tumor location from the actual tumor location is characterized by the 3D Euclidean distance between them in mm.

The patients' 4DCT image volumes cover a complete breathing cycle, hence contain information of external surface motion. We could in principle identify regions of interest (ROIs) on the patient surface that strongly correlate with tumor motions. It is expected that the marker locations selected by Algorithm 1 should fall closely into those ROIs. This also serves as a criterion for the justification of the correctness of our marker selection algorithm. To select the ROI, we consider the following two metrics. First, from the deformation vector fields between different respiratory phases, the motion trajectory for all surface points were extracted. The correlation function between the internal tumor motion in the S-I direction and the motion vector of each point on the external surface was employed as a metric. However, only part of the external surface has considerable motion amplitude and those points with small motion amplitude should not be considered for predicting tumor motions despite their possible high correlations with tumor S-I motion. Therefore, we only focus on the surface region with large

motion amplitudes. Combining the two criteria, we define the ROI as the areas on the surface in which the motion amplitude is larger than 80% of the maximum value and the correlation is above 0.85. Although those threshold values for the two criteria are chosen empirically, the general conclusions presented in the rest of this paper are found not sensitive to them.

3. Results

3.1. Marker selection

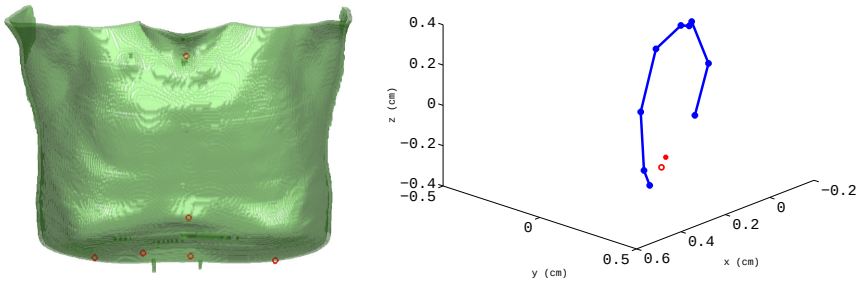

Figure 1. *Left:* Markers selected by our algorithm are shown as red circles on one of the patient's surface. *Right*: the LOOCV results for the same patient using the phases 1 through 9 as training data (blue dots) and the phase 10 as the testing data (red dot). The red circle is the predicted tumor location.

We have studied the validation of our surface marker selection algorithm on 4 lung cancer patients. The selected 6 surface markers in one typical patient (patient No. 4) are drawn in 3D space on the patient surface, as shown in the left panel of Fig. 1. Meanwhile, in the right panel of Fig. 1, we demonstrate the LOOCV results for the same patient using the phases 1 through 9 as training data and the phase 10 as the testing data. Specifically, the blue dots are the locations of the tumor in the training phases and the red dot is the location of the tumor at the phase 10. The red circle is the predicted location using the selected surface markers and the matrix A. The 3D distance between the true tumor location and the predicted location is $0.83mm$, indicating the great capability for tumor motion prediction of our algorithm.

A summary of the results of all 10 tests for each of the 4 patients is given in Table 1. For each patient, we compute the mean and the standard deviation of the 3D errors for the predicted tumor locations and the number of selected markers over all the 10 tests in the LOOCV. It is found that, on average, our algorithm can automatically select about 6 surface markers that collaboratively predict tumor motion with an 3D error about $1mm$.

Algorithm 1 is implemented using MATLAB on a laptop with Intel Core i7 (1.73 GHz) CPU and 8.0G RAM. As for the computation time, it is found that the average time required to perform one optimization is about $14sec$. We emphasize that the

	Error (mm)		#Markers		Time (sec)	
Patient	mean	std	mean	std	mean	std
1	1.85	1.15	5.5	0.85	10.6	4.5
2	1.22	1.06	5.5	1.58	4.6	1.9
3	0.44	0.28	5.4	1.84	10.8	3.0
4	0.83	0.29	7.5	1.35	30.5	11.6
Average	1.08	0.69	5.98	1.04	14.1	5.2

Table 1. Summary of tumor location prediction errors, the numbers of markers selected, and the computation time.

time reported here is the one for marker selection. Once the markers are selected, the matrix A becomes available. The prediction of tumor motion using selected markers only requires a simple matrix multiplication and hence the prediction can be achieved in a negligible amount of computation time. From Table 1, it is also found that the computational time for marker selection varies from case to case, which is possibly ascribed to the different patient sizes.

In order to see how much the selected markers are correlated with each other, we remove the least important marker at each time and calculate the prediction errors using the markers left. The importance of a marker is determined by the ℓ_2-norm of the corresponding column of the matrix A from Algorithm 1. If the marker being removed is correlated with one of the remaining makers, the prediction error should not be impacted significantly. The results are presented in Figure 2. It shows that the 6th marker is correlated with the 5th marker in the sense that it does not provide a significant decrease in prediction error. This can be easily fixed by decreasing the tolerance ϵ_1 in the stopping criteria. In fact, one may as well add the maximum allowable number of markers in the stopping criteria, which is very easy to implement. Other than the 6th marker, the rest of the markers are uncorrelated, except for patient 3, which shows that each of the selected marker has significant contribution to the prediction and prediction error will noticeably increase if some of them are removed. For patient 3 (red curve in Figure 2), although there is a 50% increase in error when the number of markers is reduced to one, the magnitudes of the errors are already small, and thus we may only need to choose 1 or 2 markers. The reason that our algorithm picked multiple markers is because relative error is used in the stopping criteria, and less markers will be selected if we replace it by the absolute error $\|A^k X - Y\|_F$.

When only one marker is to be used, it is usually placed in the area that has the largest motion amplitude and highest correlation with tumor motion. This is actually the typical way of placing a marker for tumor localization in clinical practice. We use the marker location found by our algorithm and then resolve the problem $AX = Y$ by restricting A, X and Y to this single location. Note that resolving for A this way can further reduce the prediction error comparing to the errors (using single marker) shown in Figure 2. The results are shown in Table 2. The relatively large prediction

errors (especially for patient 1,2 and 4) in comparison with the situation using optimally placed markers indicates the advantages of our method. In practice, the gain of using our method to select multiple markers over the conventional single marker method should be patient dependent. Although for some patients, e.g. patient 3, the gain could be very limited, we believe for most patients, multiple markers are indeed needed and our method should be able to noticeably reduce the prediction error in comparison to the traditional single marker method, as supported by the results of patient 1, 2 and 4.

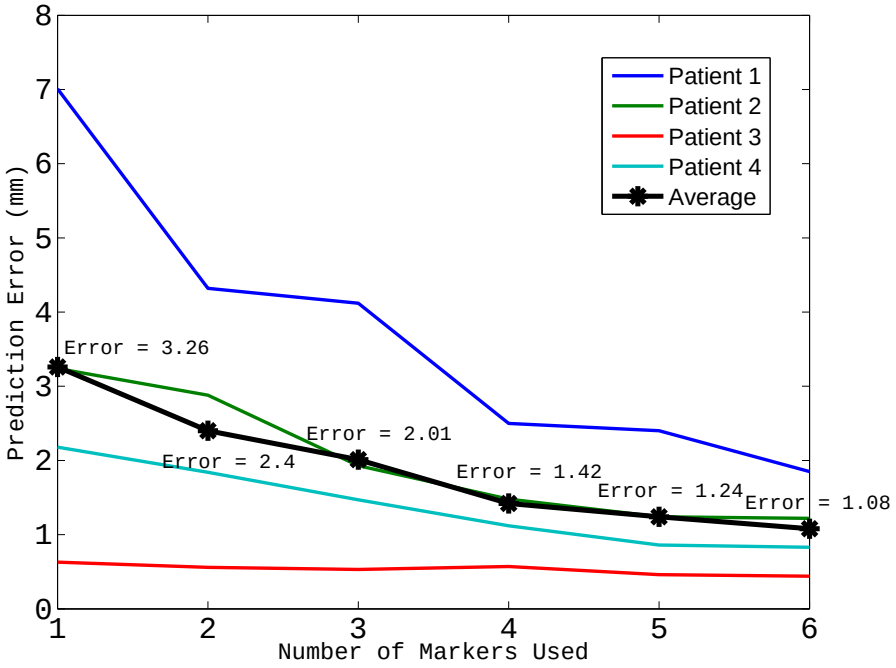

Figure 2. The above curves shows the (average) prediction errors v.s. (average) number of markers used (averaged among all 10 leave-one-out simulations). Errors are measured in mm.

Prediction Errors (mm)	Patient 1	Patient 2	Patient 3	Patient 4	Average
Our Method	1.85	1.22	0.44	0.83	1.08
Single Marker	6.24	2.49	0.47	3.33	3.13

Table 2. Comparison between traditional single marker selection and our optimal multiple markers selection.

During treatment, the physical locations of the markers placed by a therapist may not exactly match the digital locations suggested by the proposed method. In order to show robustness of the proposed method against the error during maker placement in a real treatment, we have conducted the following experiments. For each patient and each of the 10 leave-one-out tests, instead of using the computed marker locations by Algorithm 1, we randomly perturb these computed markers within a disk of radius 1cm-5cm on the patient surface. For each leave-one-out test, we repeat the experiment

for 50 times at each perturbation level. We plot the mean errors calculated over all the 50 experiments and the 10 leave-one-out tests in Figure 3. As one can see, although the deviations of the physical marker locations lead to larger prediction errors, the errors are still very well controlled. In particular, if the actual markers are placed within a 1cm disk of the identified position (which indeed can be achieved in practice), the prediction errors will only be moderately increased. This fact may be ascribed to the smoothness of the motion at the patient surface. As such, the motions at nearby points are very similar. Hence a little deviation of the marker placement would not lead to large changes to the predicted tumor locations.

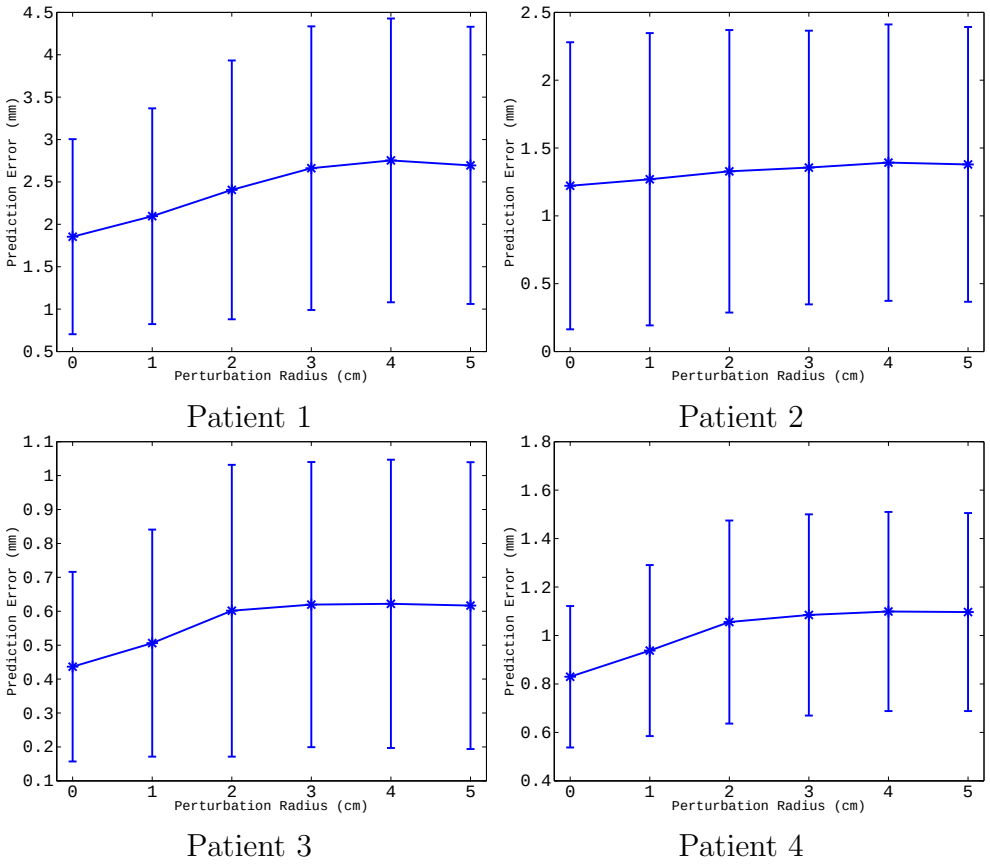

Figure 3. This figure shows the robustness of our proposed method in terms of marker perturbation. Each of the 4 graphs shows the predictions errors v.s. radius of perturbation, and the vertical bars illustrate the standard deviations of each case among all 50 tests and 10 leave-one-out tests. Note that for the case $\sigma = 0$, we assume all markers are placed at the exact locations selected by the algorithm. Thus, the results here are the same as in Table 1, and the variance is only caused by the 10 leave-one-out tests.

3.2. Comparison with ROI

The correlation between the internal tumor motion in the S-I direction and the external surface motion is shown on Fig. 4. In Fig. 5, we also present the amplitude of external

surface motion. Combining the correlation map and the motion amplitude map, the ROIs for each patient can be identified, shown as red regions in Fig. 6, where the ROIs have correlation coefficients larger than 0.85 and surface motion amplitude greater than 80% of the maximum value. Apparently, the ROIs are highly dependent on different breathing motion patterns among patients. We also plot in Fig. 6 the locations of markers selected with our algorithm for one of the leave-one-out tests. We can see that most of the marker positions selected by our algorithm fall inside or close to the ROIs, which indicate the robustness of our algorithm.

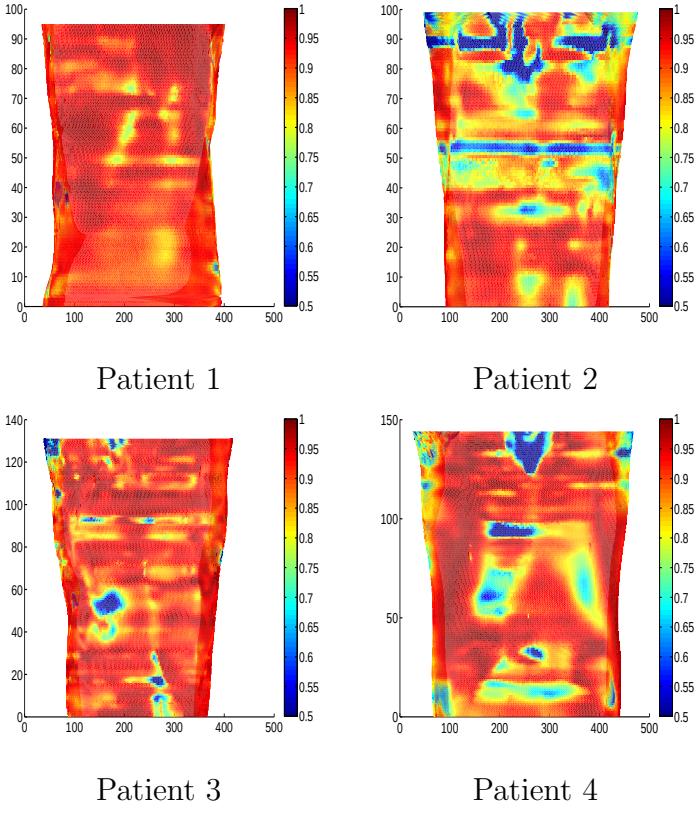

Figure 4. Color maps showing the correlation coefficients between the external surface motion and the internal tumor motion for 4 patients.

4. Discussions

4.1. Limitations and future work

The major limitation of this work is the lack of data to further validate the proposed model and algorithm. The results would be much more convincing, if the model is validated in the subsequent respiratory cycles, rather than in the same 4D-CT data via the leave-one-out scheme. Yet, it is very difficult to obtain such data from real patients treated at our clinic. It is our objective in future work to collect more data to validate our method in a comprehensive manner.

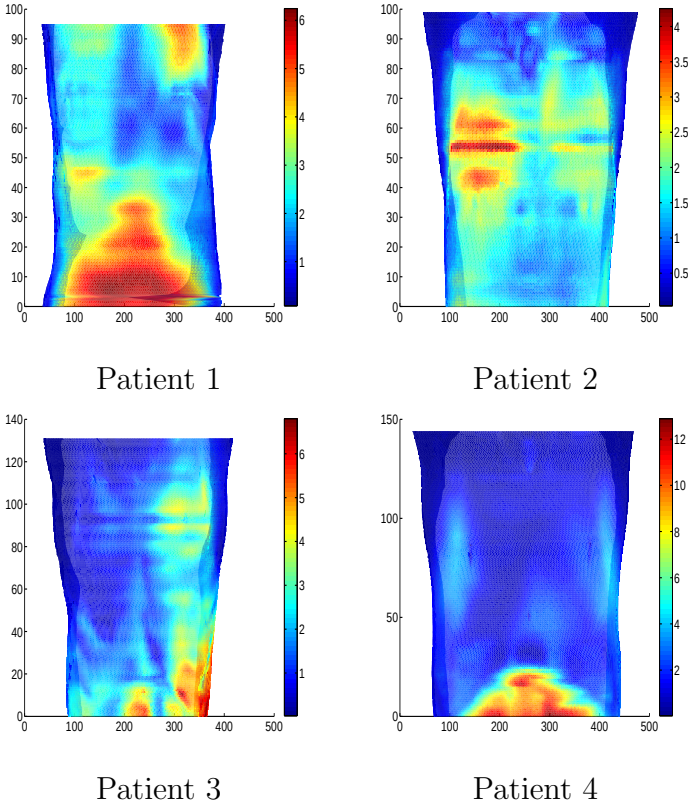

Figure 5. Color maps showing the amplitude of external surface motion for 4 patients.

One underlying assumption of 4DCT is the periodicity of breathing motion and we believe that the 4DCT images can represent the average patient breathing pattern. Therefore, the motion prediction model built on 4DCT images should be valid to a good extent. Yet, we would like to point out that 4DCT data may not be ideal for building motion prediction model. First, the 4DCT data only represent one breathing cycle and cannot account for variations among different breathing cycles. Second, the 4DCT images at each phase do not come from the same breathing cycle. Moreover, 4DCT images contains artifacts caused by the reconstruction process, e.g. unphysical horizontal bands seen in the patient surface images in Figs. 4 and 5. These artifacts will cause errors when determining surface motions. In considering all of these facts, the accuracy of the predicted tumor motion will be inherently limited.

Additionally, the 4DCT data set is usually acquired during the treatment planning stage. Yet, it has been observed that the patient motion pattern may change among treatment fractions(McClelland et al., 2011) as well as the correlation between external surrogates and the tumor motion, which may certainly degrade the validity of our model. However, the validation of the tumor prediction accuracy before each treatment fraction is a difficult problem not unique to our method. The comparison results in Table 2 are valid only under the assumption that the internal/external correlation is consistent between model construction and model application and thus give the upper limit of our

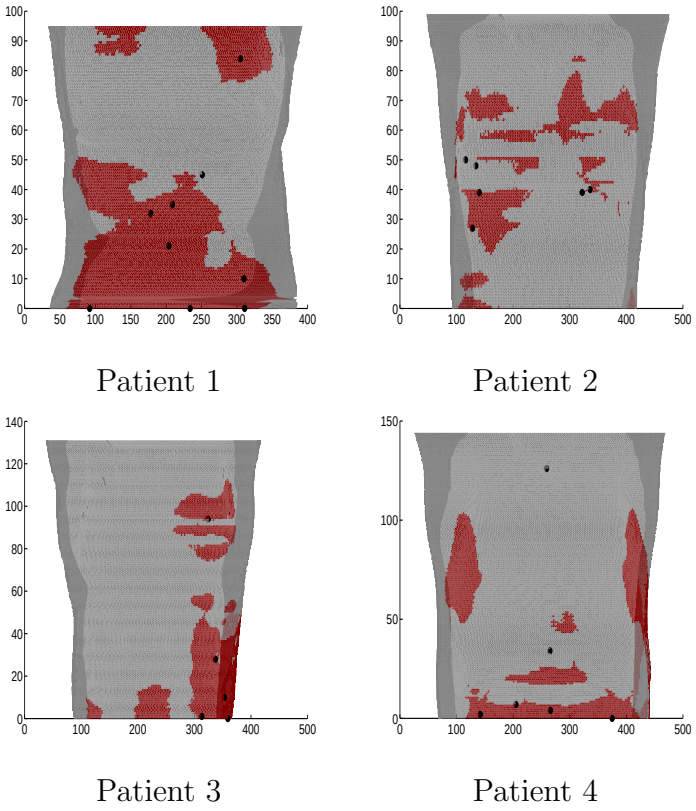

Figure 6. Color maps showing the regions of interest (where the motion amplitudes are relatively large and the correlation coefficients are relatively high) and the locations of the selected markers for one of the leave-one-out tests.

model.

Another potential problem of the current experimental setting is the possibility of over-fitting the data. When data is sufficient, the amount of data we choose for the learning and test sets should be comparable. We need to conduct comprehensive experiments to make sure our proposed model not only fits the data well, but also provides accurate predictions. At the currently stage, we are not able to acquire that many data to facilitate such experiments, whereas it is definitely worth studying in the future when more data is available.

Although our entire procedure (marker selection and prediction) needs further validation using more patient data in a variety of different clinical scenarios, we would like to point out that our method still serves as an innovative way of choosing optimal marker numbers and locations utilizing the concept of sparsity. Such choice made by our method is not only automated and requires minimal human interaction, but is also more reliable than the traditional method using only one marker (see Table 2 which shows that choosing multiple markers can significantly reduce prediction errors).

Another way to validate the results is to see if the locations of the markers selected from each of the 10 leave-one-out tests are consistent. Based on our observations, however, the sets of selected markers are not exactly the same as each other. This can

be ascribed to the numerical error in the algorithm, such as the early termination of the iteration, and in the data, such as the deformable registration error and the presence of noise. These errors will be well controlled if we have more training data. However, due to current restrictions, only limited number of training data (only 10 phases in a 4DCT) can be obtained. Nonetheless, a few marker locations repeatedly appear in all the tests, which indicate some key area on the patient surface for predicting tumor motion. The capability of finding those key areas using our algorithm in different leave-one-out tests validates our algorithm to a certain extent. We believe if more data is available, then the sets of markers should be very similar for different tests. We shall investigate this in the future when more data is available.

4.2. Further improvement of the proposed model (1)

The major assumption of the proposed model is the linear dependence of the coordinates of the tumor Y and the coordinates of markers X. However, the actual dependence of X and Y may not be linear, and a nonlinear model may produce better results (less markers or/and smaller prediction errors).

We note that the major novelty of the proposed model is to utilize sparsity and the $\ell_{2,1}$ norm, which are rather general concepts. Therefore, linearity is not crucial for our model, and it can be modified to facilitate nonlinearity. Taking the nonlinear model of (Ruan et al., 2008) as an example, if the dependence of X and Y is quadratic $AX^2 + BX = Y$, where the squares are taken on each entry of X, then our model (1) can be easily modified, for example, to the following one

$$\min_{A,B} \left\{ \|A\|_{2,1} + \|B\|_{2,1} : \ AX^2 + BX = Y \right\}$$

However, if a nonlinear model is used, more unknowns need to be solved than using a linear model, which means more data is required for reliable results. Therefore, we did not consider nonlinear dependency of X and Y in this paper due to the limited data.

Another assumption hidden in our model is that the location of tumor only depends on the locations of the surface markers at the same time point. However, in reality, due to hysterisis, the tumor location may as well depend on the marker positions at previous time points. There have been some work alone this line in the literature (Low et al., 2005; Gao et al., 2008). Here we briefly describe how we can extend our current model to properly consider temporal information of the data.

Suppose the current location of the tumor depends on the current and several previous locations of the surface markers. Let n be the total number of time point that will affect the locations of the tumor. We let $\mathcal{A}$ be a 3-dimensional tensor with $\mathcal{A} \in \mathbb{R}^{3\times 3k\times n}$ and denote each of the n slices of $\mathcal{A}$ as $A_j \in \mathbb{R}^{3\times 3k}$. We let $\mathcal{X}$ be another 3-dimensional tensor with $\mathcal{X} \in \mathbb{R}^{3k\times m\times n}$ with each of the n slices denoted as $X_j \in \mathbb{R}^{3k\times m}$. Each of the X_j are formed by shifting the columns (which correspond to time point) of X (defined at the beginning of Section 2) to the right by j periodically. Now, we define

the linear operation $\mathcal{A} \otimes \mathcal{X}$ as

$$\mathcal{A} \otimes \mathcal{X} := \sum_{j=1}^{n} A_j X_j;$$

and define the generalized $\ell_{2,1}$-norm of $\mathcal{A}$ as

$$\|\mathcal{A}\|_{2,1} := \sum_{i_2} \left(\sum_{i_1, i_3} a_{i_1,i_2,i_3}^2 \right)^{\frac{1}{2}}.$$

Then, our model (1) can now be generalized to

$$\min_{\mathcal{A}} \{ \|\mathcal{A}\|_{2,1} : \ \mathcal{A} \otimes \mathcal{X} = Y \}.$$

This model can be solved via a similar algorithm as Algorithm 1 proposed in this paper and will be investigated in our future work.

5. Conclusions

In this paper, we proposed a novel mathematical model to automatically determine the optimal number and locations of fiducial markers on patient's surface for predicting lung tumor motion. We also introduced an efficient numerical algorithm for solving the proposed model. Experiments on the 4DCT data of 4 lung cancer patients have shown that, by using our method, usually 6-7 markers are selected on patient's external surface. Most of these markers are in the regions where the surface motion is relatively large and the correction between the surface motion and the internal tumor motion is relatively high. Using these markers, the lung tumor positions can be predicted with an average 3D error of approximately $1mm$. Both the number of markers and the prediction accuracy are clinically acceptable, indicating that our method can be used in clinical practice.

Acknowledgements

This work is supported in part by the Master Research Agreement from Varian Medical Systems, Inc.. We would also like to thank the anonymous reviewers for their suggestions and comments that helped us improve the presentation of our work.

References

Arslan, Sulhattin, Adnan Yilmaz, Birol Bayramgurler, Ozlem Uzman, Edhem Nver and Esen Akkaya (2002), 'CT- guided transthoracic fine needle aspiration of pulmonary lesions: accuracy and complications in 294 patients', *Medical science monitor : international medical journal of experimental and clinical research* **8**(7), CR493–7.

Candes, E.J., J. Romberg and T. Tao (2006), 'Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information', *IEEE Transactions on Information Theory* **52**(2), 489–509.

Candes, E.J. and T. Tao (2005), 'Decoding by linear programming', *IEEE Transactions on Information Theory* **51**(12), 4203–4215.

Candes, E.J. and T. Tao (2006), 'Near-optimal signal recovery from random projections: Universal encoding strategies?', *IEEE Transactions on Information Theory* **52**(12), 5406–5425.

Combettes, P.L. and V.R. Wajs (2006), 'Signal recovery by proximal forward-backward splitting', *Multiscale Modeling and Simulation* **4**(4), 1168–1200.

Donoho, D.L. (1995), 'De-noising by soft-thresholding', *IEEE transactions on information theory* **41**(3), 613–627.

Donoho, D.L. (2006), 'Compressed sensing', *IEEE Trans. Inform. Theory* **52**, 1289–1306.

Esser, E., M. Möller, S. Osher, G. Sapiro and J. Xin (2011), 'A convex model for non-negative matrix factorization and dimensionality reduction on physical space', *Arxiv preprint arXiv:1102.0844* .

Gao, G., J. McClelland, S. Tarte, J. Blackall and D. Hawkes (2008), Modelling the respiratory motion of the internal organs by using canonical correlation analysis and dynamic mri, *in* 'The First International Workshop on Pulmonary Image Analysis', Lulu. com.

Geraghty, PR, ST Kee, G McFarlane, MK Razavi, DY Sze and MD Dake (2003), 'CT-guided transthoracic needle aspiration biopsy of pulmonary nodules: Needle size and pneumothorax rate', *Radiology* **229**(2), 475–481.

Gu, Xuejun, Hubert Pan, Yun Liang, Richard Castillo, Deshan Yang, Dongju Choi, Edward Castillo, Amitava Majumdar, Thomas Guerrero and Steve B. Jiang (2010), 'Implementation and evaluation of various demons deformable image registration algorithms on a GPU', *Physics in Medicine and Biology* **55**(1), 207–219.

Hale, E., W. Yin and Y. Zhang (2007), 'A fixed-point continuation method for ℓ_1-regularization with application to compressed sensing', *CAAM Technical Report TR, Rice University, Houston, TX* pp. 07–07. CAAM Technical Report TR07-07, Rice University, Houston, TX.

Hoisak, JDP, KE Sixel, R Tirona, PCF Cheung and JP Pignol (2004), 'Correlation of lung tumor motion with external surrogate indicators of respiration', *International Journal of Radiation Oncology Biology Physics* **60**(4), 1298–1306.

Jia, Xun, Chunhua Men, Yifei Lou and Steve B. Jiang (2011), 'Beam orientation optimization for intensity modulated radiation therapy using adaptive l(2,1)-minimization', *Physics in Medicine and Biology* **56**(19), 6205–6222.

Jiang, Steve B. (2006*a*), 'Radiotherapy of mobile tumors', *Seminars in radiation oncology* **16**(4), 239–248.

Jiang, Steve B. (2006*b*), 'Technical aspects of image-guided respiration-gated radiation therapy', *Medical Dosimetry* **31**(2), 141–151.

Low, D.A., P.J. Parikh, W. Lu, J.F. Dempsey, S.H. Wahab, J.P. Hubenschmidt, M.M. Nystrom, M. Handoko and J.D. Bradley (2005), 'Novel breathing motion model for radiotherapy', *International Journal of Radiation Oncology* Biology* Physics* **63**(3), 921–929.

McClelland, JR, S. Hughes, M. Modat, A. Qureshi, S. Ahmad, DB Landau, S. Ourselin and DJ Hawkes (2011), 'Inter-fraction variations in respiratory motion models', *Physics in Medicine and Biology* **56**, 251–272.

Nelson, Christopher, George Starkschall, Peter Balter, Rodolfo C. Morice, Craig W. Stevens and Joe Y. Chang (2007), 'Assessment of lung tumor motion and setup uncertainties using implanted fiducials', *International Journal of Radiation Oncology Biology Physics* **67**(3), 915–923.

Pepin, EW, H. Wu, Y. Zhang and B. Lord (2011), 'Correlation and prediction uncertainties in the cyberknife synchrony respiratory tracking system.', *Medical physics* **38**(7), 4036.

Ruan, D., J.A. Fessler, JM Balter, RI Berbeco, S. Nishioka and H. Shirato (2008), 'Inference of hysteretic respiratory tumor motion from external surrogates: a state augmentation approach', *Physics in medicine and biology* **53**, 2923–2936.

Thirion, J.P. (1998), 'Image matching as a diffusion process: an analogy with maxwell's demons', *Medical image analysis* **2**(3), 243–260.

Wang, Y., W. Yin and Y. Zhang (2007), 'A fast algorithm for image deblurring with total variation

regularization', *Rice University CAAM Technical Report TR07-10* .

Wu, Huanmei, Qingya Zhao, Ross I. Berbeco, Seiko Nishioka, Hiroki Shirato and Steve B. Jiang (2008), 'Gating based on internal/external signals with dynamic correlation updates', *Physics in Medicine and Biology* **53**(24), 7137–7150.

Yan, Hui, Fang-Fang Yin, Guo-Pei Zhu, Munther Ajlouni and Jae Ho Kim (2006), 'The correlation evaluation of a tumor tracking system using multiple external markers', *Medical Phisics* **33**(11), 4073–4084.

Yin, W., S. Osher, D. Goldfarb and J. Darbon (2008), 'Bregman iterative algorithms for l 1-minimization with applications to compressed sensing', *SIAM J. Imaging Sci* **1**(1), 143–168.

Zhang, X., M. Burger, X. Bresson and S. Osher (2010), 'Bregmanized nonlocal regularization for deconvolution and sparse reconstruction', *SIAM Journal on Imaging Sciences* **3**, 253.