Efficient parameter estimation for anatomy deformation models used in 4D-CT

Elizabeth S. Zachariah[1], Ding-Yu Fei[2] and Alen Docef[1,a]

[1]*Virginia Commonwealth University, Electrical and Computer Engineering, Richmond, Virginia 23284-3072, USA*
[2]*Virginia Commonwealth University, Biomedical Engineering, Richmond, Virginia 23284-3067, USA*

Abstract. A critical feature of radiation therapy for cancerous tumors located in the thorax and abdomen is addressing tumor motion due to breathing. To achieve this goal, a CT study (ordinarily denoted as 4D-CT) showing tumor location, size, and shape against time is essential. Several 4D-CT reconstruction methods have been proposed that employ anatomy deformation models. The proposed method estimates temporal parameters for these models using an approach that does not require markers or manual designation of landmark anatomical features. A neural network is trained to estimate the parameters based on simple statistical features of the CT projections. The proposed method achieves an average estimation error of less than 0.02 seconds, corresponding to a spatial error of less than 1.3 mm. The accuracy of the proposed method is evaluated in the presence of several limiting constraints such as computational complexity and noise.

1 Introduction

Radiation therapy is the preferred method of eradicating cancerous tumors located in the thorax and abdomen. To accurately target radiation to the tumor, its size, shape and location have to be known for both planning and delivery [1, 2, 3, 4, 5]. Image-guided radiation therapy (IGRT) techniques, such as conformal radiation therapy and intensity modulated radiation therapy, use CT imaging to optimize coverage of the target volume without affecting surrounding healthy tissue [1, 6]. A critical feature of IGRT is addressing tumor motion. The most prevalent cause for motion in the thorax and abdomen is breathing [1, 7, 8]. Tumors can move up to 4 cm during the breathing cycle [1, 2] causing the target volume to move in and out of the dose targeting window [9, 10, 11]. Techniques like controlled breath hold or respiratory gating can be uncomfortable or require long treatment sessions [8, 12, 13, 14, 15]. By contrast, in tumor tracking, the delivery field follows the tumor during the entire breathing cycle, allowing the patient to breathe freely [16].

A critical requirement for tumor tracking is a time-dependent CT study, ordinarily denoted as 4D-CT. Numerous methods have been implemented to construct a 4D-CT model of the moving anatomy. Some methods include modeling respiratory motion with cosine models, or averaging multiple breathing periods to create one model period [9]. In actuality, respiratory motion is more complex, and breathing periods change during treatment [1]. The more common approach is to combine CT data with a respiratory trace and assume that anatomical mo-

tion occurs in synchrony with the trace [7, 14, 17]. The trace can be obtained by monitoring a physical marker (such as a reflecting block attached to the patient's abdomen) or by measuring the tidal volume using a spirometer [5].

In the conventional 4D-CT method, the breathing trace is collected as CT projections are acquired [2, 17, 18, 19]. Based on their temporal locations within a breathing cycle, the projections are partitioned into several bins [5]. Commonly, 10 time bins are used. A 3D CT image is reconstructed for each time bin using filtered back-projection, resulting in ten CTs covering a typical breathing period.

Model-based 4D-CT methods do not limit the 4D-CT to one typical period, thus addressing inter-period variations. They also provide much better temporal resolution, potentially providing images of the anatomy at arbitrary time instants. These methods use a reference 3D image S which is deformed by a time-dependent Displacement Vector Field (DVF) $d(\cdot)$ such that the anatomy at any time is

$$S'(r, t) = S(r + d(r, t)) \tag{1}$$

Here, r is the 3D spatial variable. In our notation all bold-faced variables are vectors in space.

The spatio-temporal DVF model employed in this research is the one introduced in [20, 21, 22]. This model employs DVF basis vectors u_i to model spatial variation in the DVF and the breathing trace b to model temporal variation:

---

^a Corresponding author: adocef@vcu.edu

$$d(r,t) = \sum_i a_i u_i(r) b(t - \theta_i) \qquad (2)$$

Each spatial basis vector $u_i(r)$ is scaled by the amplitude a_i and modulated by the term $b(t - \theta_i)$, resulting in a DVF component that varies proportionally with breathing, with a delay θ_i. It has been shown that three basis vectors are sufficient for accurate representation of DVFs [22]. The basis vectors are obtained as follows: First, conventional 4D-CT is performed on the CT projection set. Pairs of consecutive projections are then registered, resulting in a set of 3D DVFs. Lastly, three basis vectors are computed from this set by means of Principal Component Analysis (PCA).

The model parameters   and   could be computed using a curve-fitting approach for anatomical landmarks (either fiducials or manually identified features); however, these are not available in realistic scenarios. Instead, projection-matching methods have been proposed in [21, 22]. These methods deform the reference image using a DVF model, compute simulated projections, and iteratively adjust the DVF model parameters to minimize the difference between simulated and actual projections. Each iteration involves a large number of computations, and the algorithm may converge to a local minimum.

A method is proposed for computationally efficient marker-less estimation of the temporal parameters. A neural network is trained to estimate the parameters based on simple statistical features of the CT projections. For projection-based parameter estimation, this method reduces the number of computations and the likelihood of convergence to local minima. For binning-based DVF interpolation, the method becomes a two-step non-iterative algorithm.

Details on the proposed method are presented in Section 2. The experimental results are discussed in Section 3 and the findings are summarized in Section 4.

2 Proposed Estimation Method

In a typical clinical scenario, a 4D-CT image set must be reconstructed from the following available data: a complete set of CT projections collected as the patient breathes freely, a digitized breathing trace, and an a priori 3D reference CT scan acquired for treatment planning. From these inputs, a complete 4D volumetric image must be computed. Usually, the reference is a fan-beam scan while the free-breathing scan is a cone-beam CT.

The temporal parameters for the model are estimated in this work using neural networks. Neural networks have been used successfully for prediction of breathing amplitudes and irregular breathing patterns [23, 24]. Given their proven ability to model temporal behavior, they are a good candidate for temporal parameter estimation. We next describe the method as used in two different scenarios.

Case 1: Use in projection matching 4D-CT. In this scenario, the spatial basis vectors   in Eq. 2 are known;   and   are sought. A feature vector is first computed from the projection set $\{P_i, i = 1, \dots\}$. The features are then fed to a neural network trained to estimate the delay parameters   . Once   are known, the amplitudes   are

found using the projection matching approach. An initial guess of   is assumed and a DVF is computed using Eq. 2. The DVF is used to deform the reference image at time instants $t_1, \dots$. Simulated CT projections   { are computed through the moving anatomy and compared to the actual projections   . The parameters   are iteratively adjusted to minimize the difference between the simulated and actual projections.

The features used to estimate   are simple statistical features of the CT projections. In this paper, the mean, standard deviation, and centroid of each projection are computed and grouped in a vector of size   . The neural network is trained using a large number of simulated projection sets corresponding to known   values.

In contrast with existing projection matching methods [21, 20], the proposed method estimates   and   separately. Since the number of free parameters is reduced by half, projection matching is less likely to converge to a local minimum, requires fewer iterations, and fewer computations in each iteration. The single-step estimation of   requires a comparatively insignificant amount of computation.

Case 2: Use in basis DVF computation. The proposed method can be used to simplify the process by which the DVF basis functions are obtained. The PCA method of [22] assumes no knowledge of the delays   . The computations can be significantly simplified if   are first estimated using the method described in Case 1. We next describe the method for computing   and   , when   are known.

The method assumes that   3D DVFs are known at time instants $t_1, \dots$. These have been computed from a conventional 4D-CT. Eq. 2 is applied   times, resulting in a set of equations that can be written in matrix form as:

$$D = BAU, \qquad (3)$$

where each row of matrix D is a DVF, each row of matrix U is a row basis vector, $A = \mathrm{diag}(a_1, a_2, a_3)$ is a diagonal matrix of amplitude parameters, the matrix

$$B = \begin{bmatrix} b(t_1 - \theta_1) & b(t_1 - \theta_2) & b(t_1 - \theta_3) \\ b(t_2 - \theta_1) & b(t_2 - \theta_2) & b(t_2 - \theta_3) \\ \vdots & \vdots & \vdots \\ b(t_N - \theta_1) & b(t_N - \theta_2) & b(t_N - \theta_3) \end{bmatrix} \qquad (4)$$

contains values of the breathing trace, and θ_1, θ_2, θ_3 are the temporal parameters. In Eq. 3, the unknowns are the matrices A and U. This is an over-determined system, and the least-square solution for the product AU is given by

$$AU = B^+ D, \qquad (5)$$

where B^+ is the Moore-Penrose pseudo-inverse of B [25]. Since u_i are unit vectors, they, together with a_i, can be computed by normalization of $a_i u_i$, the rows of AU:

$$u_i = \frac{a_i u_i}{\|a_i u_i\|}, \quad a_i = \frac{a_i u_i}{\|u_i\|} \qquad (6)$$

The method is non-iterative and involves computation of the pseudo-inverse of a small matrix (typically 10). By comparison, the PCA method of [22] processes vectors of size equal to the number of voxels in the image. More details about the pseudo-inverse method and its performance can be found in [26]. Note that in this context the proposed method can also be seen as a model-based DVF interpolation method.

3 Experimental Results

3.1 Experimental Setup

The proposed method is evaluated in the context of Case 1 described above. The CT images in real clinical settings are 3D, and cone-beam CT projections are 2D. In this paper, the feasibility of the proposed method is assessed in a simplified setting using 2D CT images. The projections are therefore 1D and the DVF in Eq. 2 is 3D (2D + time).

Since ground truth values are not available for real clinical data, a combination of real anatomy, real breathing, synthetic motion, and simulated projections was used, as follows:

- The reference static CT image S used in all experiments is an axial fan-beam CT slice of size 256×256. A cropped version of the image is shown in Figure 1.
- The real trace $b(t)$ previously used in [23] was chosen to model breathing and is shown in Figure 2. The trace was normalized to have zero mean and a maximum absolute value of 1.0. The scan time instants $t_1, \ldots, t_N$ were chosen to uniformly cover two full breathing periods. The period was estimated to be $T = 3.1$ s based on the average interval between peaks.
- The synthetic motion corresponds to a DVF computed using Eq. 2 with two basis vectors: u_1 being an antero-posterior expansion and u_2 a lateral contraction of the anatomy. In both directions the maximum displacement was chosen to be 3.5 cm. For clarity, we hereafter denote the two vectors by u_x and u_y and the corresponding delays by θ_x and θ_y.
- For a given (θ_x, θ_y) pair, the DVF is computed and used to animate the reference image. A set of N simulated parallel projections is then computed at time instants $t_1, \ldots, t_N$. The N projection angles uniformly cover a full gantry rotation. The algorithm is evaluated for N ranging from 8 to 64.
- For each projection, the features (mean, standard deviation, and centroid) are computed. The N triplets are merged into a single feature vector, then fed to the neural network, which estimates the two delays.

This approach made possible the comparison between the estimated and actual temporal parameters.

A feed-forward neural network with one hidden layer of 20 neurons was used. The network takes a vector of projection features as input and outputs the estimated delay parameters. The network is trained to minimize the estimation error in the Mean Square Error (MSE) sense. The testing data sets, consisting of representative input feature vectors and target delay parameters, were created as follows: A set of random (θ_x, θ_y) delays is first generated with values ranging from $-T$ to $+T$ where is the period of the breathing trace. For each of these, a DVF is computed and used to animate the reference image, simulated projections are computed, and a feature vector is calculated. A total of 2200 samples were used to train and test the network, with 10-fold cross-validation. In all the experiments described below, the neural network is trained for 50 Levenberg-Marquardt iterations.

To evaluate the accuracy and robustness of the proposed method, multiple scenarios have been considered, with varying numbers of projections, projection resolutions, and levels of projection noise. The method's accuracy was quantified by means of the root MSE (rMSE) error.
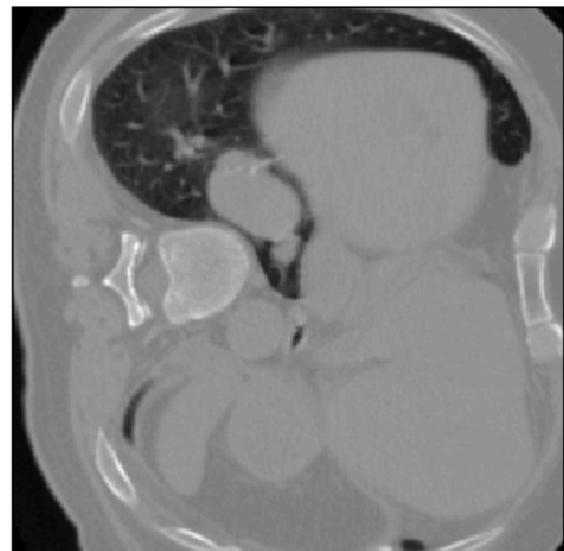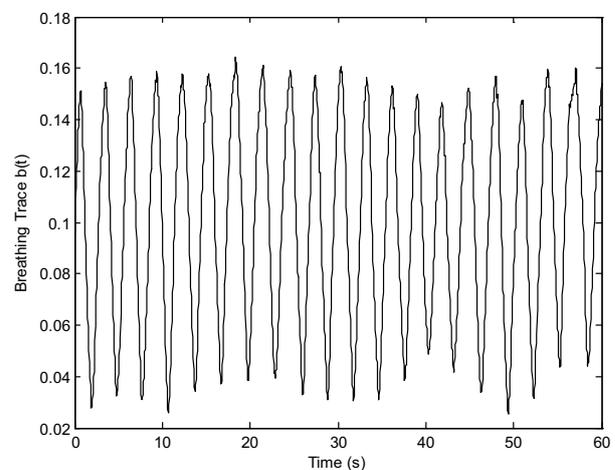

Figure 1. Reference CT image

Figure 2. Breathing trace, in arbitrary dimensionless units.

3.2 Results

The *reference* scenario, against which all scenarios are evaluated, assumes that 32 full-resolution noise-free projections are available. The input to the neural network is the feature vector described previously. The rMSE values for θ_x and θ_y are 0.0188 and 0.0178 seconds, respectively.

The effect of the *number of projections* on accuracy is evaluated by changing the reference setting of $N =$ up to 64 and down to 16 and 8. The results are shown in Figure 3(a). As expected, as the number of projections decrease, the accuracy worsens. While the algorithm does not completely fail when fewer than 16 projections are used, such settings are probably of little practical use.

Next, to assess the necessary *projection resolution*, the CT projections are downsampled by a factor of 2, then 4. The resolution of the parallel projections in the reference setting is equal to the spatial resolution of the reference CT image. The results shown in Figure 3(b) indicate that downsampling by 2 is acceptable but the accuracy decreases significantly for higher ratios. The plot also suggests that higher resolution projections would have little benefit.

The method's robustness in the presence of *projection noise* is also evaluated. Zero-mean white Gaussian noise is added to the projections in increasing amounts (5%, 10%, 20%, and 40%). These noise levels represent the ratio between the standard deviations of noise and projection. The results in Figure 3(c) show, unsurprisingly, higher estimation errors for increased noise. Noise levels higher than 10% are unacceptable.

In an attempt to assess the *suitability of the features* chosen for temporal parameter estimation, a final experiment was carried out, comparing feature-based estimation to estimation based on the entire projection sequence. The dimension of a full projection set is prohibitive (for both statistical and computational reasons), therefore we chose to downsample the reference image to $32 \times$ and to use at most 8 noise-free projections. This is clearly not a practical setting, however, the results in Table 1 indicate that a larger feature set may achieve improved accuracy or allow the use of fewer projections.

**Table 1.** Suitability of the feature set

|	Full Projections		Features	
Number of Projections	rMSE (θ_x)	rMSE (θ_y)	rMSE (θ_x)	rMSE (θ_y)
8	0.0042	0.0050	0.0849	0.0603
4	0.0109	0.0108	0.1178	0.1136

3.3 Discussion

To put these rMSE numbers in context, recall that the delays take values from $-T/2$ to $+T/2$, where $T = 3.1$. Thus the rMSE values for the reference setting are 0.6% of the breathing period, or 6% of the interval between the time bins of a typical conventional 4D-CT. An even more illuminating interpretation of these results is to consider the effect of the estimation error on the DVF accuracy.
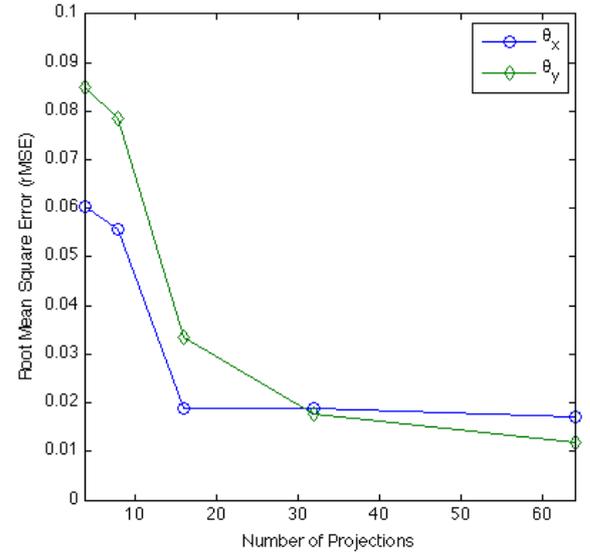
(a)
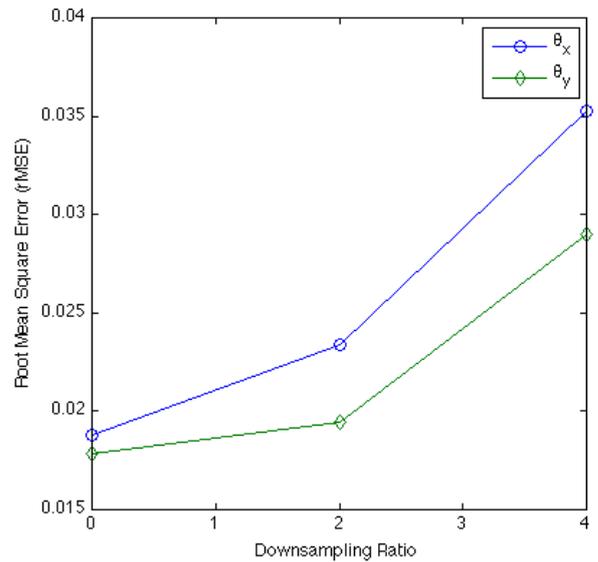
(b)
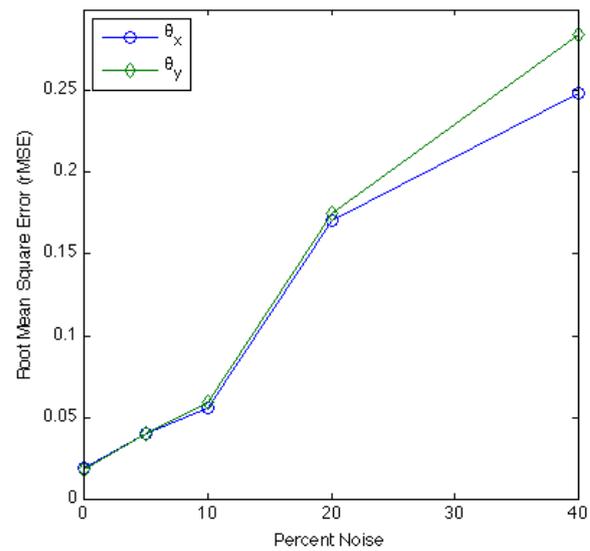
(c)

Figure 3. Estimation error as a function of (a) the number of projections, (b) the projection resolution and (c) noise level.

In a worst-case scenario, considering the maximum slope of the trace $\dot{z}$ in Eq. 2 and a maximum deformation of   cm, the maximum DVF error is found to be   mm. This value is well within the accuracy of DVF estimation methods.

One limitation of the experimental setup is that the neural network is trained for a particular reference image and set of DVF basis vectors. When these change, it may therefore fail to accurately estimate   values. Addressing this issue is beyond the scope of this feasibility study. However, the rather generic nature of the features and the substantial amount of a priori information that can be exploited (reference image, lengthy breathing trace, conventional 4D-CT) are grounds for optimism that the issue could be successfully addressed in future work.

Another minor limitation is that the test DVF is perfectly modeled by Eq. 2. However, it has been shown previously that the equation accurately models realistic breathing motion [22]. Real anatomical motion will be used in future work.

4 Conclusions

A neural network-based method for estimating temporal parameters of a DVF model has been proposed. The method's average estimation error is less than 0.02 seconds, and its worst-case DVF error is 1.3 mm. The accuracy of the proposed method has been evaluated in the presence of computational complexity constraints and CT projection noise.

References

[1] L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T.-F. Li, G.-Y. Kim, G. Luxton and A. Koong, "Overview of Image-Guided Radiation Therapy," *Medical Dosimetry,* vol. 31, no. 2, pp. 91-112, 2006.

[2] L. A. Dawson and M. B. Sharpe, "Image-guided radiotherapy: rationale, benefits, and limitations.," *Lancet Oncology,* no. 7, pp. 848-858, 2006.

[3] L. A. Dawson and D. A. Jaffray, "Advances in image-guided radiation therapy," *J Cin Oncol,* vol. 25, pp. 938-946, 2007.

[4] M. Baumann, T. Holscher and D. Zips, "The future of IGRT - cost benefit analysis," *Acta Oncologica,* vol. 47, no. 7, pp. 1188-1192, 2008.

[5] M. Lecchi, P. Fossati, F. Elisei, R. Orecchia and G. Lucignani, "Current concepts on imaging in radiotherapy," *European Journal of Nuclear Medicine and Molecular Imaging,* vol. 35, pp. 821-837, 2008.

[6] D. Verellen, M. De Ridder and G. Storme, "A (short) history of image-guided radiotherapy," *Radiotherapy and Oncology,* vol. 86, pp. 4-13, 2008.

[7] G. S. Mageras, A. Pevsner and E. D. Yorke, "Measurement of lung tumor motion using

respiration-correlated CT," *Int J Radiat Oncol Biol Phys,* no. 60, pp. 933-941, 2004.

[8] P. Keall, "4-dimensional computed tomography imaging and treatment planning," *Seminars in Radiation Oncology,* vol. 14, pp. 81-90, 2004.

[9] V. Boldea, G. C. Sharp, S. B. Jiang and D. Sarrut, "4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis," *Medical Physics,* vol. 35, pp. 1008-1018, 2008.

[10] Y. Sun, S. Dieterich, B. Cho and P. J. Keall, "An analysis of thoracic and abdominal tumour motion for stereotactic body radiotherapy patients," *Physics in Medicine and Biology,* vol. 53, pp. 3623-3640, 2008.

[11] S. Mori, M. Endo, S. Komatsu, T. Yashiro, S. Kandatsu and M. Baba, "Four-dimensional measurement of lung tumor displacement using 256-multi-slice CT-scanner," *Lung Cancer,* vol. 56, pp. 59-67, 2007.

[12] A. Trofimov, E. Rietzel and H. Lu, "Temporo-spatial IMRT optimization: Concepts, implementation and initial results," *Phys Med Biol,* no. 50, pp. 2779-2798, 2005.

[13] E. M. Leter, F. Cademartiri, P. C. Levendag, H. Stam and P. J. Nowak, "Four-dimensional multislice computed tomography for determination of respiratory lung tumor motion in conformal radiotherapy," *International Journal of Radiation Oncology, Biology, Physics,* vol. 62, no. 3, pp. 888-892, 1 July 2005.

[14] J. Wong, M. B. Sharpe and D. A. Jaffray, "The use of active breathing control (ABC) to reduce margin for breathing motion," *Int J Radiat Oncol Biol Phys,* no. 44, pp. 911-919, 1999.

[15] R. Zeng, J. A. Fessler and J. M. Balter, "Respiratory motion estimation from slowly rotating x-ray projections: theory and simulation," *Medical Physics,* vol. 32, no. 4, pp. 984-991, April 2005.

[16] F. Khan, G. Bell, J. Antony, M. Palmer, P. Balter, K. Bucci and M. J. Chapman, "The use of 4DCT to reduce lung dose: a dosimetric analysis," *Medical Dosimetry,* vol. 34, no. 4, pp. 273-278, 2009.

[17] G. Chen, J. Tang and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics,* vol. 35, pp. 660-663, 2008.

[18] P. J. Keall, G. Starkschall and H. Shukla, "Acquiring 4D thoracic CT scans using a multislice helical method," *Phys Med Biol,* no. 49, pp. 2053-2067, 2004.

[19] E. Yorke, K. E. Rosenzweig, R. Wagman and G. S. Mageras, "Interfractional anatomic variation in patients treated with respiration-gated radiotherapy," *Journal of Applied Clinical Medical Physics,* vol. 6, pp. 19-32, 2005.

[20] A. Docef and M. J. Murphy, "Reconstruction of 4D

deformed CT for moving anatomy," *International Journal of Computer Assisted Radiology and Surgery,* no. 3, pp. 591-598, 2008.

[21] R. S. Brock, A. Docef and M. J. Murphy, "Reconstruction of a cone-beam CT via forward iterative projection matching," *Medical Physics,* vol. 37, 2010.

[22] D. Staub, A. Docef, R. S. Brock and C. Vaman, "4D Cone-beam CT reconstruction using a motion model based on principal component analysis," *Medical Physics,* vol. 38, no. 12, pp. 6212-6220, December 2011.

[23] M. J. Murphy and S. Dieterich, "Comparative performance of linear and nonlinear neural networks to predict irregular breathing," *Physics in Medicine and Biology,* vol. 51, no. 22, p. 5903, 21 Nov 2006.

[24] M. J. Murphy and D. Pokhrel, "Optimization of an adaptive neural network to predict breathing," *Medical Physics,* vol. 36, no. 1, pp. 40-47, January 2009.

[25] A. G. T. Ben-Israel, Generalized Inverses: Theory And Applications, Springer, 2003.

[26] E. Zachariah, "4-D Modeling of Displacement Vector Fields for Improved Radiation Therapy," 2010. [Online]. Available: http://scholarscompass.vcu.edu/etd/2323.