Published in final edited form as:

Phys Med Biol. 2012 January 21; 57(2): 395–413. doi:10.1088/0031-9155/57/2/395.

Classifying geometric variability by dominant eigenmodes of deformation in regressing tumours during active breath hold lung cancer radiotherapy

Ahmed M. Badawi, Elisabeth Weiss, William C. Sleeman IV, and Geoffrey D. Hugo

Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA, USA

Geoffrey D. Hugo: gdhugo@vcu.edu

Abstract

The purpose of this study is to develop and evaluate a lung tumour interfraction geometric variability classification scheme as a means to guide adaptive radiotherapy and improve measurement of treatment response. Principal component analysis (PCA) was used to generate statistical shape models of the gross tumour volume (GTV) for 12 patients with weekly breath hold CT scans. Each eigenmode of the PCA model was classified as ‘trending’ or ‘non-trending’ depending on whether its contribution to the overall GTV variability included a time trend over the treatment course. Trending eigenmodes were used to reconstruct the original semi-automatically delineated GTVs into a reduced model containing only time trends. Reduced models were compared to the original GTVs by analyzing reconstruction error in the GTV volume and position. Both retrospective (all weekly images) and prospective (only the first four weekly images) were evaluated. Average volume difference from the original GTV was $4.3\% \pm 2.4\%$ for the trending model. The positional variability of the GTV over the treatment course, as measured by the standard deviation of the GTV centroid, was 1.9 ± 1.4 mm for the original GTVs, which was reduced to 1.2 ± 0.6 mm for the trending only model. In 3/13 cases the dominant eigenmode changed class between the prospective and retrospective models. The trending only model preserved GTV volume and shape relative to the original GTVs, while reducing spurious positional variability. The classification scheme appears feasible for separating types of geometric variability by time trend.

1. Introduction

Geometric variation of the thoracic anatomy is complex, consisting of multiple sources of variability such as respiration-induced tumour motion (Sonke *et al.*, 2008); inter- and intra-fraction change in baseline lung volume (Hugo *et al.*, 2007); tumour growth and regression (Bosmans *et al.*, 2006; Britton *et al.*, 2009; Fox *et al.*, 2008; Glide-Hurst *et al.*, 2010; Juhler Nøttrup *et al.*, 2007; Kupelian *et al.*, 2005; Ramsey *et al.*, 2006; Spoelstra *et al.*, 2009; Underberg *et al.*, 2006; Woodford *et al.*, 2007); and development, progression, or resolution of associated pathology such as pleural effusion and atelectasis (Glide-Hurst *et al.*, 2010; Sonke *et al.*, 2008). These different sources of variability induce different types of anatomical change. For example, respiration and baseline changes generally cause rigid shifts of the tumour that do not trend during the treatment course, while tumour regression in response to therapy results in a decrease in gross tumour volume with a well-defined time trend.

Image guided adaptive radiotherapy (IGART) is a promising technique for managing such geometric variation over the treatment course for thoracic tumours. However, an efficient implementation of this process must select an appropriate intervention for managing a particular type of anatomical change. For example, interfraction baseline shifts of the tumour are amenable to correction with cone beam CT-mediated online guidance (Bissonnette *et al.*, 2009; Grills *et al.*, 2008; Guckenberger *et al.*, 2006; Purdie *et al.*, 2007; Sonke *et al.*, 2008; Yeung *et al.*, 2009). Tumour regression, which results in a change in tumour volume and is characterized by a trend over the treatment course, may be better addressed with periodic replanning or other adaptive techniques (Britton *et al.*, 2009; Glide-Hurst *et al.*, 2010; Fox *et al.*, 2009; Hugo *et al.*, 2007; Ramsey *et al.*, 2006; Woodford *et al.*, 2007). Ideally, optimal combinations of online position correction and adaptive replanning could be designed to maximize treatment efficiency (as online setup correction is more efficient than generating a new or adapted treatment plan).

Due to the complex nature of anatomical variation in the thorax, the direct source of a particular type of variability is not always readily apparent, and thus selection of the appropriate interventional strategy in this efficient IGART scheme may be difficult. For example, anisotropic tumour volume regression during treatment mimics change in tumour position by introducing an apparent shift in the tumour centroid relative to bony anatomy (Glide-Hurst *et al.*, 2010). This effect reduces the utility of standard metrics of breath hold reproducibility (e.g., change in tumour centroid position between subsequent breath holds) by confounding them with this additional variability from tumour shape change. In another example, measurement of tumour regression may enable adaptation to this changing tumour volume, but this measurement involves substantial uncertainty due to respiration (or breath hold reproducibility), as free-breathing imaging is known to create artefacts that can introduce error into the measurement of tumour volume. Use of 4D imaging for investigating tumour regression should improve the accuracy of tumour volume measurement (Fox *et al.*, 2009). However, adaptation to tumour regression is likely to require knowledge of the geometric dynamics of regression in addition to knowledge of tumour volume change. Means to remove spurious anatomical variability (such as breath hold variability) from a regression model would enable such modelling.

Principal component analysis (PCA) has been used to develop and optimize models for representing geometric variation during radiotherapy (Badawi *et al.*, 2010; Birkner *et al.*, 2007; Budiarto *et al.*, 2011; Li *et al.*, 2010; Price and Moore, 2007; Sohn *et al.*, 2005). Complex geometries of GTV and other structures were compressed into a set of few principal components (PCs) that change with time. The original anatomy can be reconstructed from these PCs, or the models can be used to generate new likely geometries. PCA-based models are a good option for separation of geometric variability, as PCA decomposes the position and shape change of the anatomy into these distinct principal components or eigenmodes. Previous studies of PCA geometric models qualitatively analyzed the types of shape change represented by the principal components to determine if physiologically-likely changes were captured (Birkner *et al.*, 2007; Budiarto *et al.*, 2011; Sohn *et al.*, 2005) via visualization only. However, these studies did not proceed to a quantitative analysis of the dynamics of the eigenmodes.

In this paper, we introduce a new framework to classify and separate geometric variability of the gross tumour volume (GTV) represented by a PCA-based model of interfraction variability. The eigenmodes are classified by change in structure volume and position, and by time trends in interfraction variability represented by the principal component coefficients trending with radiotherapy treatment time. To test feasibility, the classification scheme is then applied to the relatively straightforward case of image-guided active breath hold radiotherapy, where the major classes are either time-trending (i.e., tumour regression)

or non-trending (e.g., breath-hold reproducibility), and therefore can be separable. To assess performance, we reconstruct the trending components of the structure, and compare the volume and positional variability of the reconstructed structure to the original automatically delineated GTV structure.

2. Methods and Materials

2.1. Representing geometric variability with principal components analysis

In a recent study, we used PCA to represent and generate models of interfraction variability of the GTV in a group of lung cancer patients (Badawi *et al.*, 2010). These PCA models were used here to develop the classification scheme. Here, we briefly describe the PCA modelling process. For a more detailed description, the reader is referred to the original paper (Badawi *et al.*, 2010).

An interfraction structure surface model consists of a set of anatomical instances $\mathbf{p}$ of the surface at a number of time points t during the treatment course, where each surface $\mathbf{p}$ is represented as a vector containing a set of surface points. Such a model could be made for organs or structures (e.g., GTV, CTV, lungs, etc.) and a separate, new model is made for each patient. For PCA, the list of three-dimensional surface points is stored in a single column vector $\mathbf{p} = [x_1 \ y_1 \ z_1 \ \dots \ x_n \ y_n \ z_n]^T$, where n is the number of surface points. The surface can then be represented by the time-varying vector $\mathbf{p}(t)$. PCA decomposes and reconstructs $\mathbf{p}(t)$ from a set of linear basis vectors, where the basis vectors are eigenvectors of the covariance matrix of $\mathbf{p}(t)$. The basis vectors being eigenvectors of the covariance matrix, each capture correlated movement of the surface points over time. For example, if the observed variation of all surface points over time were perfectly correlated, only a single eigenvector would be required to reconstruct any observable state. Any observable state could then be reconstructed by multiplying the eigenvector by a constant. In clinical situations, all tissue does not move in perfect correlation, so many eigenvectors are required in practice to reconstruct the observed anatomical instances. The combination of each eigenvector and associated eigenvalue, λ_k is termed an *eigenmode*. Eigenvalues are equal to the variance of each eigenmode. Generally, eigenmodes associated with small eigenvalue do not contribute substantially to modelling of geometric variability as seen in the input geometries and can be ignored (Badawi *et al.*, 2010; Sohn *et al.*, 2005). Thus, a reduced interfraction model $\mathbf{p}'(t)$ can be formed as:

$$\mathbf{p}'(t) = \bar{\mathbf{p}} + \sum_{l=1}^L c_l(t) \mathbf{q}_l \quad (1)$$

where $\bar{\mathbf{p}}$ is the mean structure shape over the entire treatment course, $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_L\}$ is the set of eigenvectors, L is the number of principal components, or dominant eigenmodes, to keep in the reduced model, and $\{c_1(t), c_2(t), \dots, c_L(t)\}$ is the set of principal component coefficients which reconstruct each specific time instance of $\mathbf{p}(t)$ at particular treatment time, t . In this study, we set L as the minimum number of eigenmodes representing at least 95% of the total variability. The set of principal component coefficients can be generated by projecting $\mathbf{p}(t)$ onto the set of basis vectors:

$$c_l(t) = (\mathbf{p}(t) - \bar{\mathbf{p}}) \cdot \mathbf{q}_l \quad (2)$$

Note that each eigenvector $\mathbf{q}_l$ represents a displacement vector field (DVF), or direction of motion, for each surface point in $\mathbf{p}(t)$, while $c_l(t)$ represents the relative strength of this particular eigenvector necessary to reconstruct $\mathbf{p}$ at the particular instance of time t . The PC

coefficients vary with the temporal state of the anatomy. If the surface point positions were perfectly correlated, but had a time trend (for example, a shrinking sphere), then the PC coefficients would also require a time trend to reconstruct the observed surface.

2.2. Quantitative measurement of principal shapes

The general classification strategy is based on analyzing the shape and position of the structure represented by each eigenmode, and time trends in the principal component coefficients (Figure 1). Each eigenmode consists of spatially correlated changes in the surface and temporally with implementation during treatment time. For example, an eigenmode that demonstrates no volume change but some position change in the structure represents a rigid positional shift, a variation that is likely to be due to change in the position of the structure due to setup error, breath hold reproducibility, or other associated anatomical changes. As there is no volume change in the structure, such an eigenmode does not likely represent regression or other responses to therapy. Conversely, eigenmodes with volume change and a time trend in the principal component coefficients represent a time-trending change in position and volume which is likely response-related (such as regression of a tumour), and not due to breath hold variability or other random anatomical variabilities.

Shape change for each individual eigenmode was quantified by deforming the mean geometry by its respective eigenvector $\mathbf{q}_l$ to yield a set of *principal shapes* ($\mathbf{s}_l$) for each eigenmode, $l=1$ to L according to:

$$\mathbf{s}_l = \bar{\mathbf{p}} \pm X\sigma_l \mathbf{q}_l \quad (3)$$

where σ_l is the standard deviation of $\mathbf{p}(t)$ captured by the eigenmode number l (and it is equal to the square root of the eigenvalue, λ_l), and X is a constant. For this study, X was chosen to be one, and the set of (1σ) principal shapes $\{(\bar{\mathbf{p}} - \sigma_l \mathbf{q}_l), \text{ and } (\bar{\mathbf{p}} + \sigma_l \mathbf{q}_l)\}$ were reconstructed for each eigenmode. The set of principal shapes for each eigenmode provides a means to quantitatively assess the amount and type of shape change captured by the associated eigenmode, while normalization to a fixed fraction of the standard deviation of the eigenmode enables comparison between different dominant eigenmodes. For each eigenmode, the volume and position (as represented by the shape centroid) of the anatomical structure for each of the principal shapes were calculated. We term these the *modal volumes* and *modal positions* for the principal shape. The end result of this process is a set of modal volumes and modal positions for the principal shapes of the anatomical structure for each eigenmode and each subject. The modal volumes and modal positions represent characteristics of each eigenmode that can be used to compare eigenmodes quantitatively. The choice of X is arbitrary, but must be the same for each eigenmode to allow quantitative comparison between modes. Furthermore, the value of X will effect the classification threshold, as described below.

2.3. Eigenmode classification

A hierarchical rule-based classification system is proposed to classify dominant eigenmodes by the principal shapes and time-trends of the surface reconstructed by the eigenmode (Figure 1). Here, we are interested not in the principal shapes themselves, but rather the change in shape captured by the eigenmode from mean shape ($\bar{\mathbf{p}}$) to the extreme shapes ($\bar{\mathbf{p}} \pm \sigma_l \mathbf{q}_l$). To measure this principal shape change, the change in modal volume and modal centroid position of the structure was calculated. For each eigenmode, the *modal volume change* was calculated as the absolute percentage difference in modal volume for the $\pm 1\sigma$ shape in relation to the mean shape volume:

$$\Delta V_m = 100 * \left| \frac{V(\bar{\mathbf{p}} - \sigma_l \mathbf{q}_l) - V(\bar{\mathbf{p}} + \sigma_l \mathbf{q}_l)}{V(\bar{\mathbf{p}})} \right| \quad (4)$$

where ΔV_m is the modal volume change (%) and $V(\mathbf{s}_l)$ is an operator returning the volume of the principal shape for eigenmode l . Similarly, the *modal position change* was calculated as the magnitude of the 3D difference between the centroid of the 1σ principal shape and the mean principal shape:

$$\Delta p_m = \|C(\bar{\mathbf{p}} + \sigma_l \mathbf{q}_l) - C(\bar{\mathbf{p}})\| \quad (5)$$

where Δp_m is the modal position change (mm), $C()$ is an operator that returns the 3D position of the centroid, and $\| \cdot \|$ is the Euclidean norm operator that returns the magnitude of the vector difference between the centroids for the 1σ shape and mean principal shape, over the mediolateral, anterior-posterior and superior-inferior directions. Since the modal position change for the 1σ and -1σ principal shapes have the same magnitude, only the value for the positive 1σ shape was calculated.

Modal volume change and modal position change were used as two out of the three input features to the proposed hierarchical rule-based classification system. The third classification rule was formulated to stratify eigenmodes as representing temporally trending or temporally non-trending. Each observed anatomical instance (the original surfaces used as input into the PCA model) was projected onto the new PCA basis using Eq. (2) to obtain the set of PC coefficients, $\{c_l\}$ for each eigenmode l . Projecting all observed states (all surfaces available during the treatment course) forms a time-varying function of coefficients, $c_l(t)$. Time trends in the PC coefficients (due to time trends in the observed states) were quantified by linear regression of $c_l(t)$, and eigenmodes sorted into time trending or non-trending. An eigenmode was considered time-trending if the slope of the regression fit of $c_l(t)$ was significantly ($p < 0.05$) different than zero using Student's t test.

Eigenmodes were then classified based on modal volume change, modal position change, and trending behaviour of the PC coefficients. A binary classification system was formed from the three input features as modal volume change ('large' or 'small'), modal position change ('large' or 'small'), and PCs ('trending' or 'random non-trending'), where the output is the class label (Table 1). No volume change and some position change represents rigid shift of the structure, while large modal volume changes imply deformation of the structure. Deformation associated with a position change implies anisotropic deformation, while no position change implies isotropic deformation. Finally, the PC coefficients represent the 'amount' of the eigenmode necessary to reconstruct the original surfaces. Randomness in the PCs implies no time trend in the change in modal volume or modal position of the structure. Time trends in the PCs are associated with trending pattern such as treatment response. For a binary classification system of three input features, the maximum number of classes is equal to eight. However, small modal volume change implies a rigid shift, so we considered only six plausible classes.

After classifying each eigenmode into one of the six general classes, classes were then stratified into two superclasses based on whether a class was likely to be associated with trending or non-trending variability type. This superclass stratification is applied, with the premise being that classes associated with time trends are likely related to tumour response and, potentially, other time trending changes such as progression or resolution of an associated pathology such as atelectasis or pleural effusion, while non-trending classes are likely related to anatomical variations such as breath hold reproducibility, baseline shifts in respiration, and setup error.

2.4. Data acquisition and delineation

12 patients with locally advanced non-small cell lung cancer (13 GTVs) undergoing curative radiotherapy gave informed consent and were enrolled on a trial to perform weekly helical CT imaging during treatment. Only GTV surface models were evaluated in this study. During simulation and once per week of treatment, CT images were acquired using an active breathing control (ABC) technique at end of normal inspiration. Three CT scans were acquired to assess breath hold reproducibility during each session; the total number of images acquired for each patient varied from 12 to 21 (corresponding to four to seven weeks of treatment) over all patients. The scanning and breath hold protocol is further described in detail in Glide-Hurst et al. (Glide-Hurst *et al.*, 2010). A physician delineated the primary GTV on the first breath hold CT image from the first week, which was defined as the reference image. Lymph node targets were not considered in this study. A small deformation, inverse consistent, linear elastic (SICLE) deformable registration algorithm was used to register each weekly image to the first CT image acquired during the first week (Christensen *et al.*, 2007). The GTV delineated contours on the reference image were automatically propagated using the deformable registration (DR) to all other images by its corresponding displacement vector field (DVF). The set of delineated contours by physician at first week and the remaining automatically delineated at later treatment weeks by the DR will all be termed delineated through the paper. The resulting automatically propagated structures were reviewed by the same physician without any modifications as the automatically propagated GTV contours were of acceptable registration accuracy compared to their manual delineation and contained structure correspondence between manual and automatic delineation generated by the deformable registration (Hugo et al., 2011). This method provided homologous point correspondence for GTV between all CT images, and a representation of the change in each structure shape and position over treatment time. The proposed method and framework rely on automatic contour propagation in order to keep a one-to-one correspondence relation for all GTV contour points in the different CT series.

2.5. Separate model reconstruction and analysis

The output from the proposed eigenmode classification system was used to reconstruct separate trending and non-trending components for each patient (Figure 1). In our previous work (Badawi *et al.*, 2010), only the first L dominant eigenmodes were used to reconstruct a reduced, low noise model $\mathbf{p}'(t)$ of the original GTV shapes. Similarly, one can select any subset of basis vectors and, using Eq. (1), reconstruct a reduced model $\mathbf{p}'(t)$. For the L total number of dominant eigenmodes, a specific patient may have a combination of trending and non-trending eigenmodes; have trending components only; or non-trending eigenmodes only. In the general case when there are both trending and non-trending eigenmodes, we reconstructed each reduced model (trending only and non-trending only) separately by applying Eq. (1) twice, first by reconstructing only with eigenmodes classified into the trending superclass, and next by reconstructing only with eigenmodes classified into the non-trending superclass. We term these trending only or non-trending only reduced model reconstructions the 'separate models'. Finally, we performed a reconstruction with all dominant eigenmodes, as in our previous work (Badawi *et al.*, 2010). This process resulted in three sets of GTVs for each patient, where each set contained a GTV at each imaged fraction during treatment.

In this case, as some dominant eigenmodes are not used in the reconstruction for the reduced models, the resulting reconstructed GTV shapes may not be similar to the original GTV shapes. For example, consider a patient model with two dominant eigenmodes, one that is class 4 (a time-trending change in GTV volume) and one that is class 1 (a non-trending rigid shift) (Figure 2). The class 4 eigenmode is typical of tumour regression, while the class 1 eigenmode is typical of a random, rigid shift such as setup error or breath hold variability. If

the GTV was reconstructed with only the class 4 eigenmode, it would have nearly the same volume as the delineated one and over the entire treatment, but without the positional variation represented by the class 1 eigenmode (middle row in Figure 2). Conversely, the class 1 eigenmode reconstruction would exhibit the positional variation, with almost the mean shape of the GTV at all time points (bottom row in Figure 2).

To analyze the quality of the reconstructed GTV, it is evident from the example in Figure 2 that the surface point positions from the reconstructed models cannot be directly compared to the original GTV surface point positions, as the GTV shape will not be the same for the non-trending reconstruction and the GTV centroid position will not be the same for the trending reconstruction. To evaluate the accuracy of the GTV reconstruction with the separated models, trending-only and non-trending only reconstructions were evaluated against the original GTVs (GTVs deformed to each image from the reference image) and against the reconstructions with all dominant eigenmodes (full model). For the trending-only model, we hypothesize that rigid, random positional variation will be reduced but GTV volume will be preserved to a great extent. To evaluate this hypothesis, the volume of the reconstructed GTV from the separated trending-only model was compared to that from the original GTV. Also, the standard deviation of the centroid position over all fractions for the separated trending-only model GTV was compared to that of the original GTV, under the hypothesis that removing non-trending variation would reduce the overall variability of the centroid.

2.6. Prospective modelling stability

We have previously shown that reconstruction accuracy stabilized with at least four measurements used to create a prospective model (i.e., the model is formed from four measurements and used to reconstruct new out-of-model measurements later in treatment) (Badawi *et al.*, 2010). Clinical use of the proposed classification scheme would require use of such a prospective model. We assessed the stability of the classification of dominant eigenmodes as the model grows through addition of new measurements by comparing two models, one ‘prospective model’ based on four weekly measurements and the other a ‘retrospective model’ based on all available measurements, to determine if dominant eigenmodes changed class or superclass. The prospective four weeks model was evaluated using the four weeks model data, and also the ‘testing error’ was evaluated by projecting all weekly measurements onto the model to determine if the model could be used to reconstruct new out-of-model measurements. The centroid and volume analyses of Section 2.5 were also repeated with the prospective four weeks model (both in model and out-of-model testing) and compared to the retrospective model.

3. Results

3.1. Eigenmodes classification

To select the thresholds for classifying modal volume and modal position change as small or large change, a sensitivity analysis was performed via unsupervised *k*-means clustering for modal volume change (range 7–10%) and modal position change from (1.5–2mm). Change of classification was minimal varying these thresholds. Changing the inputs boundaries from 7–10% modal volume change and from 1.5–2 mm modal position change only moved one subject to a different class (Patient 6 from class 2 to class 4) while all other patients remained in the same class. Based on these results, cluster boundaries were set to 10% for modal volume change and 2.0 mm for modal position change. Due to the complex nature of these classes, it is not fully separable and has small overlaps at the class boundaries. Note that changing the value of X used to generate the principal shapes will effect the selection of the threshold value.

Table 2 shows the classification results for both prospective and retrospective models. Eigenmode class in bold represents the trending eigenmodes in each patient, with its associated trending p -value between parentheses. For both prospective and retrospective models, each patient has at least one trending eigenmode and at least one non-trending eigenmode. Four patients had two trending eigenmodes. A total of 39 dominant eigenmodes in 13 GTVs were found for the prospective model and 46 for the retrospective model. In both models, there were 17 trending eigenmodes, while the number of non-trending eigenmodes increased from 22 to 29, prospective to retrospective models. 38% (5/13) of GTVs had a change in superclass in at least one dominant eigenmode between the prospective and retrospective models; for 23% (3/13) of these GTVs the change was in the most dominant eigenmode. From the results in Table 2, superclass stratification is applied for classes 2, 4, and 6 (trending) and classes 1, 3, and 5 (non-trending). Figure 3 shows the modal position change vs. modal volume change for the prospective (A) and retrospective (B) model classes. Clusters of modes can be observed, although the clusters were non-fully separable by a single threshold. In Figure 3c, a relation between eigenmode number and modal position change can be observed, while it appears there is little relation between eigenmode number and modal volume change. Dominant eigenmodes capture large, correlated variabilities; large rigid shifts have inherent correlated motion of the surface points and appear in the most dominant modes.

Figure 4 shows the GTV and Figure 5 shows modal volume change, modal position change, and PC coefficients for the retrospective model of Patient 1, who had a dominant trending eigenmode and two less dominant non-trending eigenmodes. Eigenmode 1, classified as class 6 (anisotropic trending regression), demonstrates a large modal volume change and corresponding large modal position change along the left/right axis, which is also apparent in Figure 4. Figure 6 and Figure 7 show the GTV and GTV modal volume change, modal position change, and PC time trends for the retrospective model of the three dominant eigenmodes for Patient 8. For this subject, eigenmodes 1, 2, and 3 were classified into classes 2, 1, and 2, respectively, and are likely representing rigid positional variations with minimal volume change of the GTV surface.

3.2. Trending components separation

Table 3 shows the percentage difference in GTV volume between the reconstructed GTVs and the associated delineated ones, averaged over the population, for reconstruction by all dominant eigenmodes (L) and by trending eigenmodes only. Using all dominant eigenmodes, the volume of the reconstructed GTVs was approximately 1% more accurate than those GTVs reconstructed with trending only modes. Table 4 shows the variability of the centroid position, as measured by the standard deviation. The trending only reconstruction achieved lower centroid variability, on average, than the original delineated GTVs, demonstrating that trending only modal reconstruction reduced the positional variation of the reconstructed GTV while largely preserving structure volume. Modelling with a prospective model and applying this model to all future data produced similar centroid variability as with the retrospective model.

Table 5 shows the centroid variability results for Patient 1 in Figure 4. The centroid variability in the mediolateral direction for the trending-only reconstruction is similar to that for the original delineated GTV, as it is the major direction of tumour regression (Figure 4). In the other directions, where regression is less prevalent, the centroid variability is reduced. In these other directions, the original large centroid variability is likely then due to other non-trending positional variations such as breath hold reproducibility.

Patients were grouped into two cohorts based on whether a superclass change in at least one eigenmode was observed from the prospective to the retrospective model, to evaluate

stability of the prospective model over future fractions. Table 3 and Table 4 also show the results of statistics for grouped patients who had changes in their principal eigenmodes observed from prospective to retrospective models. Volume differences and centroid position standard deviation for the “N” group had smaller volume and centroid position errors (compared to the original delineated GTVs) than the “Y” group.

4. Discussion and conclusions

A new framework for classifying and separating geometric variability by type based on principal components was developed and evaluated. In a classification scheme with binary criteria, it is important to verify that threshold selection does not impact the classification. For classifiers for which the threshold can be variable, unless a threshold value is based on some intuitive principle of the data, sensitivity of the classification to the threshold is an important property to evaluate. In section 2.3, good clustering and insensitivity within a reasonable range of thresholds were demonstrated. We selected 10% and 2 mm as modal volume change and modal position change thresholds, respectively, from the k-means clustering step of these two features, although an independent sensitivity analysis is recommended for any new dataset or classification scheme, or a change in the principal shape scaling parameter (X).

We assigned classes to the superclasses based on assessment of what might cause the type of variability represented by the class based on its trending nature. Classes 4, and 6 are intuitively related to tumor regression as these represent time-trending change in GTV volume. Class 1 is also intuitive, as it represents a rigid, random variation of the GTV position, as might be observed due to respiration, setup error, or breath hold variability. It is more difficult to postulate a source for the few resulted classes 3 and 5; non-trending volume change could be due random deformation due to chest wall attachment, for example, or delineation and registration error. We chose to separate classes by time trend as this presents the most obvious difference between random, anatomical variations and trending (progression or regression) changes in tumor volume, atelectasis, and pleural effusion. Although one could plausibly argue with this labeling, the impact on the results of changing the label would be minimal as only five eigenmodes were classified as either class 3 or 5 (1 in four weeks models and 4 in all weeks models) in four patients, and four out of these five eigenmodes are not the first dominant eigenmode.

The accuracy of the model can be expected to be no better than the quality of the input data to the modeling process. One potentially important error source is the accuracy of deformable registration used to propagate the surfaces across treatment images. For this study, we previously characterized the accuracy of the registration near the GTV surface using manually-defined landmarks. (Hugo *et al.*, 2011), and found the accuracy for surface modeling to be within the voxel size (2 mm). Secondly, we try to reduce the impact of random errors (such as noise) small by keeping only the dominant modes with the largest 95% of variability in the PCA representation. There are several potential uses for the classification scheme. Current image-based (Ramsey *et al.*, 2006) and radiobiological (Chvetsov *et al.*, 2008) models for predicting tumor volume deterministically only produce a predicted volume. Incorporating a classification PCA scheme into such models could help enable prediction of tumor geometry as well. Including a trending-only PCA model could make incorporation of geometric regression into these predictive models less complex by reducing unrelated geometric variability from setup and respiration. The non-trending only PCA model may enable measurement of random, rigid type variabilities without the confounding effects of tumor regression (Glide-Hurst *et al.*, 2010) to, for example, aid in margin-finding studies for active breathing control in locally-advanced lung cancer.

The classification scheme could also support coverage probability or probabilistic treatment planning. Non-trending eigenmodes could be identified and planned using existing techniques developed for non-trending random distributions (Gordon *et al.*, 2010; Price and Moore, 2007). Trending eigenmodes could be incorporated using a predictive deterministic model.

The classification scheme could potentially be applied to free-breathing patients through the use of 4DCT imaging rather than repeat breath hold imaging. However, the much larger intra-fraction variations, plus the added complexity of baseline and random variation in the respiratory cycle, would likely require larger number of dominant eigenmodes for accurate reconstruction and a larger number of classes (composite spatial and temporal) and its stratifications (both in time and space). Essentially, due to increase in the complexity of variability, a larger number of samples (i.e., more frequent imaging) would be necessary to adequately characterize free-breathing in this manner.

Most patients could be grouped into sets of a few generic types (not to be confused with the ‘superclasses’ of trending regression and non-trending types with breath hold). The most common set of classes was a trending eigenmode as the most dominant eigenmode, followed by several much less dominant non-trending eigenmodes. However, a study of a larger patient population is required to determine if these generic types of eigenmodes combinations and dominance order are common or unique to the study population.

The prospective and retrospective models showed good agreement on average and for most patients, which is necessary for clinical implementation of a classification scheme in IGART. Several patients showed a change in superclass between the models, and also had larger errors in GTV reconstruction than the remaining patients. Three of the five patients with superclass changes had resolving atelectasis or had a developing or progressing pleural effusion. These types of large, dramatic change in anatomy may affect the ability to model anatomical variation with a statistical method such as PCA. In this study, it appears that these types of change affected the class of the dominant modes, and therefore changed the ‘type’ of variability observed for the patient. If such patients and/or the occurrence of such events could be identified, the PCA model could be updated, and classification adjusted. For example, a new model could be generated at a successive fraction or week, and the classification of the models compared for changes in the principal modes classification to determine if the model is stable or if the new model should be used. As this type of patients are likely to benefit the most from IGART (Sonke and Belderbos, 2010), such identification and the effect of remodeling is a subject of our future work.

Although the deformable image registration used to generate the surface models captures GTV rotation, the classification framework as developed here will not classify rotational changes of the structure. We chose not to model rotations in this study as it was unclear how much of the deformation, if any, was due to rotation of the GTV. However, modal rotation change (analogous to modal position change) could be also considered via calculating principal eigen axes rotation (calculated from all point coordinates in PCA model such that the eigenvectors are the principal axes), and can be added as an input feature to the classification system. In future work, we can expand the number of classes including modal rotation change into the automatic classification system, if necessary (for example, when considering rotation of the GTV during respiration).

A deficiency of the study is the use of surface models to construct a statistical shape model rather than using a statistical deformation model (a voxel-based model), incorporating deformations inside the structures into the model. Patient 8 in Figure 6 is a good example of where use of a full deformation field would be interesting, as this patient had a tumor whose

overall volume (based on the surface) did not change, but whose tumor demonstrated some local regression that re-expanded an airway and the esophagus. Such complex deformations are not classified in the current model, but could be by including more complex classifiers based on a PCA model of the full deformation vector field. A second issue with the study is that it is possible that a source of variability could be distributed into multiple eigenmodes, or a single eigenmode could contain variability from multiple sources. This effect may be particularly important moving from a surface-based to a voxel-based model. Careful analysis of the types of deformations observed will be necessary to enable selection of appropriate features to group deformation types in this complex case.

In conclusion, a new eigenmode classification system was developed to classify geometric variability by type using PCA. Selecting eigenmodes classified as trending in time, PCA reconstruction of the GTV with these eigenmodes preserved the delineated GTV volume, but with reduced random positional variability of the GTV. Modelling with the initial four weeks of treatment and prospectively projecting future weeks achieved similar error levels as a retrospective model including all data, demonstrating preliminary feasibility for use with adaptive radiotherapy. Finally, the proposed classification scheme appears feasible for separating types of geometric variability based on its trending nature, preserving volumes and reducing standard deviation of positional variability.

Acknowledgments

This work has been supported by the NIH grants No. R01CA116249 and P01CA116602. We thank Gary Christensen for helpful discussions on deformable image registration and use of the SICLE deformable image registration algorithm.

References

- Badawi A, Weiss E, Sleeman WC, Yan C, Hugo GD. Optimizing principal component models for representing interfraction variation in lung cancer radiotherapy. *Med Phys*. 2010; 37:5080–5091. [PubMed: 20964228]
- Birkner M, Thorwarth D, Poser A, Ammazalorso F, Alber M. Analysis of the rigid and deformable component of setup inaccuracies on portal images. *Phys Med Biol*. 2007; 52:5721–5733. [PubMed: 17804891]
- Bissonnette JP, Purdie TG, Higgins JA, Li W, Bezjak A. Cone-beam computed tomographic image guidance for lung cancer radiation therapy. *Int J Radiat Oncol Biol Phys*. 2009; 73:927–934. [PubMed: 19095368]
- Bosmans G, van Baardwijk A, Dekker A, Öllers M, Boersma L, Minken A, Lambin P, De Ruyscher D. Intra-patient variability of tumor volume and tumor motion during conventionally fractionated radiotherapy for locally advanced non-small cell lung cancer. A Prospective Clinical Study. *Int J Radiat Oncol Biol Phys*. 2006; 66:748–753. [PubMed: 17011450]
- Britton KR, Starkschall G, Liu H, Chang JY, Bilton S, Ezhil M, John-Baptiste S, Kantor M, Cox JD, Komaki R, Mohan R. Consequences of anatomic changes and respiratory motion on radiation dose distributions in conformal radiotherapy for locally advanced non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys*. 2009; 73:94–102. [PubMed: 18929448]
- Budiarto E, Keijzer M, Storchi P, Hoogeman M, Bondar L, Mutanga T, de Boer H, Heemink A. A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases. *Physics in Medicine and Biology*. 2011; 56:1045. [PubMed: 21258137]
- Christensen GE, Song JH, Lu W, El Naqa I, Low DA. Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Med Phys*. 2007; 34:2155. [PubMed: 17654918]
- Chvetsov AV, Palta JJ, Nagata Y. Time-dependent cell disintegration kinetics in lung tumors after irradiation. *Phys Med Biol*. 2008; 53:2413–2423. [PubMed: 18421118]

- Fox J, Ford E, Redmond K, Zhou J, Wong J, Song DY. Quantification of tumor volume changes during radiotherapy for non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys.* 2009; 74:341–348. [PubMed: 19038504]
- Glide-Hurst CK, Gopan E, Hugo GD. Anatomical and pathological variability during radiation therapy for a hybrid active breath hold gating technique. *Int J Radiat Oncol Biol Phys.* 2010; 77:910–917. [PubMed: 20510201]
- Gordon JJ, Sayah N, Weiss E, Siebers JV. Coverage optimized planning: probabilistic treatment planning based on dose coverage histogram criteria. *Med Phys.* 2010; 37:550–563. [PubMed: 20229863]
- Grills IS, Hugo G, Kestin LL, Galerani AP, Chao KK, Wloch J, Yan D. Image-guided radiotherapy via daily online cone-beam CT substantially reduces margin requirements for stereotactic lung radiotherapy. *Int J Radiat Oncol Biol Phys.* 2008; 70:1045–1056. [PubMed: 18029110]
- Guckenberger M, Meyer J, Wilbert J, Baier K, Mueller G, Wulf J, Flentje M. Cone-beam CT based image-guidance for extracranial stereotactic radiotherapy of intrapulmonary tumors. *Acta Oncol.* 2006; 45:897–906. [PubMed: 16982556]
- Hugo GD, Weiss E, Badawi A, Orton M. Localization Accuracy of the Clinical Target Volume during Image-guided Radiotherapy of Lung Cancer. *Int J Radiat Oncol Biol Phys.* 2011; 81:560–567. [PubMed: 21277096]
- Hugo GD, Yan D, Liang J. Population and patient-specific target margins for 4D adaptive radiotherapy to account for intra-and inter-fraction variation in lung tumour position. *Phys Med Biol.* 2007; 52:257. [PubMed: 17183140]
- Juhler Nøttrup T, Korreman SS, Pedersen AN, Aarup LR, Nyström H, Olsen M, Specht L. Intra-and interfraction breathing variations during curative radiotherapy for lung cancer. *Radiother Oncol.* 2007; 84:40–48. [PubMed: 17588697]
- Kupelian PA, Ramsey C, Meeks SL, Willoughby TR, Forbes A, Wagner TH, Langen KM. Serial megavoltage CT imaging during external beam radiotherapy for non-small-cell lung cancer: observations on tumor regression during treatment. *Int J Radiat Oncol Biol Phys.* 2005; 63:1024–1028. [PubMed: 16005575]
- Li R, Jia X, Lewis JH, Gu X, Folkerts M, Men C, Jiang SB. Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy. *Med Phys.* 2010; 37:2822–2826. [PubMed: 20632593]
- Price GJ, Moore CJ. A method to calculate coverage probability from uncertainties in radiotherapy via a statistical shape model. *Phys Med Biol.* 2007; 52:1947–1965. [PubMed: 17374921]
- Purdie TG, Bissonnette JP, Franks K, Bezjak A, Payne D, Sie F, Sharpe MB, Jaffray DA. Cone-beam computed tomography for on-line image guidance of lung stereotactic radiotherapy: localization, verification, and intrafraction tumor position. *Int J Radiat Oncol Biol Phys.* 2007; 68:243–252. [PubMed: 17331671]
- Ramsey CR, Langen KM, Kupelian PA, Scaperth DD, Meeks SL, Mahan SL, Seibert RM. A technique for adaptive image-guided helical tomotherapy for lung cancer. *Int J Radiat Oncol Biol Phys.* 2006; 64:1237–1244. [PubMed: 16446055]
- Sohn M, Birkner M, Yan D, Alber M. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. *Phys Med Biol.* 2005; 50:5893–5908. [PubMed: 16333162]
- Sonke J-J, Lebesque J, Herk MV. Variability of four-dimensional computed tomography patient models. *Int J Radiat Oncol Biol Phys.* 2008; 70:590–598. [PubMed: 18037579]
- Sonke JJ, Belderbos J. Adaptive radiotherapy for lung cancer. *Semin Radiat Oncol.* 2010; 20:94–106. [PubMed: 20219547]
- Spoelstra FOB, Pantarotto JR, Van Sörnsen de Koste JR, Slotman BJ, Senan S. Role of adaptive radiotherapy during concomitant chemoradiotherapy for lung cancer: analysis of data from a prospective clinical trial. *Int J Radiat Oncol Biol Phys.* 2009; 75:1092–1097. [PubMed: 19327915]
- Underberg RWM, Lagerwaard FJ, van Tinteren H, Cuijpers JP, Slotman BJ, Senan S. Time trends in target volumes for stage I non-small-cell lung cancer after stereotactic radiotherapy. *Int J Radiat Oncol Biol Phys.* 2006; 64:1221–1228. [PubMed: 16442240]

- Woodford C, Yartsev S, Dar AR, Bauman G, Van Dyk J. Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images. *Int J Radiat Oncol Biol Phys.* 2007; 69:1316–1322. [PubMed: 17967322]
- Yeung AR, Li JG, Shi W, Newlin HE, Chvetsov A, Liu C, Palta JR, Olivier K. Tumor localization using cone-beam CT reduces setup margins in conventionally fractionated radiotherapy for lung tumors. *Int J Radiat Oncol Biol Phys.* 2009; 74:1100–1107. [PubMed: 19395197]

Figure 1.

Flowchart of the proposed modal classification and separation scheme. Conventional PCA analysis of interfraction surfaces is performed for a patient. Modal volume change and modal position change of each dominant eigenmode are calculated, and the PC coefficients required to reconstruct the original surface contours with each eigenmode are calculated and evaluated for time trends. Each dominant eigenmode is then classified using these three characteristics as features. Finally, separated models are reconstructed of the surface by selecting only certain eigenmodes by superclass, and reconstructing the original surface models using only these eigenmodes.

Figure 2.

Illustration of a simple patient GTV model consisting of two dominant eigenmodes (class 1 and class 4 from Table 1). Top row: The delineated GTV, exhibiting both a rigid positional shift and a (uncorrelated) time-trending loss of volume during treatment. Middle row: Reconstructed GTV using only the time-trending mode representing volume change (class 4). The positional variation has been removed from the reconstructed GTV through removal of the class 1 mode. Bottom row: Reconstructed GTV using only the non-trending mode representing rigid shift (class 1). The positional variation is captured by the reconstruction, but the GTV shape remains constant as the mean GTV shape over the treatment course.

Figure 3. Modal position change (mm) vs. modal volume change (%), for all patients, all modes. (A) prospective model, (B) retrospective model, (C) all data, colored by dominant mode number, with mode 1 being the most dominant.

Figure 4. Patient 1 CT images with delineated GTV contours. Left to right, Axial and coronal images of GTV week 1 (blue), GTV week 6 (orange), and both week 1 and week 6 GTVs overlaid on week 1 image.

Figure 5.

Example modal volume change (MVC), modal position change (MPC) and PC coefficient time trend for Patient 1. (A) MVC (%) and MPC (mm), (B) PC coefficients versus scan number, in temporal order. Trendlines are only displayed for modes with trends significant at the $p=0.05$ level. Modes 1, 2, and 3 were classified as class 6, 1, and 1. LR: left-right, AP: anterior-posterior, SI: superior-inferior.

Figure 6. Patient 8 CT images with delineated GTV contours. Left to right, Axial and coronal images of GTV week 1 (blue), GTV week 7 (orange), and both week 1 and week 7 GTVs overlaid on week 1 image.

Figure 7.

Example modal volume change (MVC), modal position change (MPC), and PC coefficient time trend for Patient 8. (A) MVC (%) and MPC (mm), (B) PC coefficients versus scan number, in temporal order. Trendlines are only displayed for modes with trends significant at the $p=0.05$ level. Modes 1,2, and 3 were classified as class 2,1, and 2. LR: left-right, AP: anterior-posterior, SI: superior-inferior.

Hierarchical rule-based modal classification scheme. Shown are types of structure change represented by the class, the 3 input features used (modal volume change, modal position change, and PC coefficients trending), classes labels, and superclass type. 'PC time trend?' represents whether the slope of the linear regression fit of the PC coefficients, $\{c_i\}$ for the mode was significantly different from zero.

Table 1

Change type	Modal volume change	Modal position change	PC time trend?	Class label	Superclass type
Rigid shift	Small	Small or Large	No	1	Non-trending
Rigid shift, trending	Small	Small or Large	Yes	2	Trending
Isotropic deformation	Large	Small	No	3	Non-trending
Isotropic deformation, trending	Large	Small	Yes	4	Trending
Anisotropic deformation	Large	Large	No	5	Non-trending
Anisotropic deformation, trending	Large	Large	Yes	6	Trending

Table 2

Classification, all patients (dominant eigenmodes only), for prospective and retrospective models. Trending only superclass (grouped class labels 2, 4, and 6) are shown in bold with its associated trending p-value in parentheses. 'Superclass change?': If a dominant mode changed superclass between the prospective and retrospective models, this column is listed as 'Y'. 'First mode superclass change?': If the most dominant (first) mode changed superclass, this column is listed as 'Y'.

Patient #	Prospective model class labels				Retrospective model class labels				First mode superclass change?
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 1	Mode 2	Mode 3	Mode 4	
1	4 (0.00)	1 (0.26)			6 (0.00)	1 (0.67)	1 (0.19)	N	N
2-1	1 (0.27)	1 (0.18)	2 (0.00)		1 (0.76)	1 (0.77)	2 (0.00)	1 (0.71)	N
2-2	2 (0.00)	1 (0.70)	4 (0.05)		1 (0.22)	2 (0.02)	3 (0.33)	Y	Y
3	6 (0.00)	1 (0.84)			5 (0.49)	6 (0.00)	1 (0.23)	Y	Y
4	4 (0.00)	1 (0.90)			4 (0.02)	1 (0.40)	3 (0.12)	N	N
5	6 (0.04)	4 (0.00)	1 (0.79)	1 (0.83)	6 (0.00)	4 (0.00)	1 (0.95)	1 (0.81)	N
6	2 (0.00)	1 (0.82)	1 (0.74)	1 (0.19)	2 (0.00)	1 (0.82)	1 (0.75)	1 (0.19)	N
7	1 (0.47)	2 (0.00)	1 (0.82)	3 (0.19)	1 (0.94)	2 (0.00)	1 (0.91)	3 (0.31)	N
8	6 (0.00)	1 (0.31)			2 (0.00)	1 (0.96)	2 (0.01)	Y	N
9	1 (0.10)	1 (0.92)	4 (0.00)		6 (0.02)	2 (0.01)	1 (0.23)	Y	Y
10	2 (0.01)	1 (0.36)	2 (0.02)		4 (0.00)	1 (0.47)	1 (0.09)	1 (0.45)	N
11	2 (0.05)	1 (0.07)	2 (0.05)	1 (0.28)	2 (0.00)	1 (0.75)	2 (0.00)	1 (0.22)	N
12	4 (0.00)	1 (0.24)	1 (0.60)		6 (0.00)	1 (0.09)	1 (0.35)	1 (0.67)	N

Volume difference, reconstructed vs. original GTV, averaged over all patients. Mean (standard deviation) is displayed for volumes reconstructed from prospective models, prospective models with testing projection of all weeks data onto the model, and retrospective models. Results are shown for all patients, the subgroup that did have (Y), and the subgroup that did not have (N) change in classification for the first principal eigenmode from prospective to retrospective model.

Table 3

Model	Volume difference (%), PCA-reconstructed vs. original GTV					
	All dominant modes			Trending only modes		
	All patients	Y ^a	N ^a	All patients	Y	N
Prospective model, evaluated with out of model data only	2.4 (3.0)	3.0 (3.3)	2.1 (2.7)	3.6 (2.5)	4.7 (2.8)	3.0 (2.0)
Prospective model, evaluated with all weeks data	3.4 (2.8)	4.2 (2.8)	2.8 (2.7)	4.4 (2.7)	5.6 (2.9)	3.6 (2.2)
Retrospective model	3.3 (3.4)	4.0 (1.8)	2.8 (4.1)	4.3 (2.4)	5.2 (2.5)	3.8 (2.2)

^aY: patients whose first principal eigenmode changed superclass from prospective to retrospective model, N: patients with no change in superclass.

GTV positional variability, as measured by the magnitude of the standard deviation of the centroid position (mm) for prospective models, prospective models with testing projection of all weeks data onto the model, and retrospective models. Values shown are mean (standard deviation) of the centroid standard deviation over the population. Results are shown for all patients, the subgroup that did have (Y), and the subgroup that did not have (N) change in classification for the first principal eigenmode from prospective to retrospective model.

Table 4

Model	Magnitude, SD centroid (mm)									
	Original GTV			All dominant modes			Trending only modes			
	All patients	Y ^a	N ^a	All patients	Y	N	All patients	Y	N	
Prospective model, evaluated with out of model data only	1.8 (1.5)	2.6 (0.2)	1.3 (0.1)	1.8 (1.5)	2.6 (0.2)	1.3 (0.1)	1.2 (1.6)	2.0 (0.3)	0.8 (0.1)	
Prospective model, evaluated with all weeks data	1.9 (1.4)	2.5 (0.2)	1.5 (0.1)	1.7 (1.2)	2.3 (0.2)	1.4 (0.1)	1.3 (1.3)	1.8 (0.2)	0.9 (0.1)	
Retrospective model	1.9 (1.4)	2.5 (0.2)	1.5 (0.1)	1.9 (1.4)	2.5 (0.2)	1.5 (0.1)	1.2 (0.6)	1.5 (0.1)	1.0 (0.1)	

^aY: first principal eigenmode changed superclass from prospective to retrospective model, N: no change in superclass.

Table 5

Standard deviation (over all weeks) of the GTV centroid position (mm) for Patient 1. L/R: left/right, A/P: anterior/posterior, S/I: superior/inferior.

Direction	SD centroid (mm), mean over all weeks		
	Delineated	All dominant modes	Trending only modes
L/R	2.0	2.0	2.0
A/P	0.6	0.6	0.1
S/I	0.7	0.7	0.2