

# A method to map errors in the deformable registration of 4DCT images<sup>a)</sup>

Constantin Vaman,<sup>b)</sup> David Staub, Jeffrey Williamson, and Martin J. Murphy<sup>c)</sup>

Department of Radiation Oncology, Virginia Commonwealth University, Richmond, Virginia 23298

(Received 25 January 2010; revised 16 July 2010; accepted for publication 5 August 2010; published 14 October 2010)

**Purpose:** To present a new approach to the problem of estimating errors in deformable image registration (DIR) applied to sequential phases of a 4DCT data set.

**Methods:** A set of displacement vector fields (DVF) are made by registering a sequence of 4DCT phases. The DVFs are assumed to display anatomical movement, with the addition of errors due to the imaging and registration processes. The positions of physical landmarks in each CT phase are measured as ground truth for the physical movement in the DVF. Principal component analysis of the DVFs and the landmarks is used to identify and separate the eigenmodes of physical movement from the error eigenmodes. By subtracting the physical modes from the principal components of the DVFs, the registration errors are exposed and reconstructed as DIR error maps. The method is demonstrated via a simple numerical model of 4DCT DVFs that combines breathing movement with simulated maps of spatially correlated DIR errors.

**Results:** The principal components of the simulated DVFs were observed to share the basic properties of principal components for actual 4DCT data. The simulated error maps were accurately recovered by the estimation method.

**Conclusions:** Deformable image registration errors can have complex spatial distributions. Consequently, point-by-point landmark validation can give unrepresentative results that do not accurately reflect the registration uncertainties away from the landmarks. The authors are developing a method for mapping the complete spatial distribution of DIR errors using only a small number of ground truth validation landmarks. © 2010 American Association of Physicists in Medicine.

[DOI: [10.1118/1.3488983](https://doi.org/10.1118/1.3488983)]

Key words: deformable image registration, registration validation, breathing motion, 4DCT, principal components analysis

## I. INTRODUCTION

The motion of the treatment target and other structures presents complications for external beam radiotherapy treatment. Movement relative to the planning anatomy can occur day-by-day or, in the case of respiration, can proceed continuously during each treatment fraction. Respiration-correlated CTs (4DCTs) record anatomical motion due to breathing at a number of discrete time points (phases) in the breathing cycle.<sup>1-3</sup> Deformable image registration (DIR) detects and quantifies the displacement of anatomical features from one phase (i.e., time bin) to the next.<sup>4</sup> It can also map the anatomical configuration at any particular breathing phase back to some reference phase. The anatomical displacement maps produced by image registration are commonly referred to as displacement vector fields (DVFs) because they display the displacement of anatomical elements from one position to another in a sequence of images. The DVFs can be used to track the accumulation of dose over the breathing cycle<sup>4</sup> and to construct statistical models of where any particular anatomical element might be during treatment for use in probabilistic planning.<sup>5,6</sup> However, the accuracy of the DVFs directly impacts the accuracy of these applications.

At the present time, 4DCT has relatively coarse time resolution and requires the blending of motion data from several breathing cycles to fill each time bin. As a result, any particular 4DCT distorts the true thoracic movement with mo-

tion and binning artifacts. Furthermore, DIR algorithms are not perfect.<sup>7</sup> Therefore, a DVF obtained by registering pairs of images will always have two components: One that corresponds to the actual physical displacements between the two images and a second, unphysical component due to imperfections in the imaging and registration processes. The unphysical component is the *DVF error* and mathematically can be represented by a vector field defined at every voxel and added to the “real” (error-free) displacement vectors. Our present study is aimed at the problem of determining voxel-by-voxel the complete registration error field throughout the imaged volume.

If the registration errors fluctuate randomly from one voxel to another and from one DVF to the next (e.g., such as white noise), then we can sample them by locating a set of landmarks that measure the true physical motion in the DVF at specific points and subtracting their displacement from the DIR displacement vector at each landmark location, leaving the error component.<sup>8</sup> However, if the errors do not fluctuate randomly, but have patterns of spatial coherence in each DVF, then it is not trivial to extend point landmark validation elsewhere in the anatomy. For example, a DVF might show good accuracy at or near selected landmarks while having substantially greater errors elsewhere. It has therefore been suggested that a large set of landmarks (>1000) is necessary to provide an adequate picture of DVF uncertainties through-

out the anatomical volume.<sup>9</sup> However, we hypothesize that if one models mathematically the spatial correlations in the error map, it is not necessary to have a dense sampling of landmarks. We describe here a new method to determine complete DVF error maps from 4DCT registrations using a small set ( $<100$ ) of ground truth validation landmarks.

Why is it proper to assume that (in general) DVF errors are spatially correlated? The deformation model of any DIR algorithm will provide only an imperfect representation of the anatomy movement. The difference between the actual motion and the model will translate into a registration error component that changes smoothly throughout the image, i.e., that has strong spatial coherence from one DVF to the next. Local fluctuations of the anatomical motion might be overly smoothed by DIR, generating a second spatially structured component of registration error. Artifacts in the image will introduce a third source of spatial error in the DVF, and so on. Each of these error components will have spatial coherence but will be approximately independent of the others.

We can use to our advantage the fact that the DIR errors are coherent by employing principal components analysis (PCA). It can be shown mathematically that if the voxel-by-voxel displacement vector amplitudes in a set of  $N$  DVFs all vary in proportion to a single scaling function (e.g., the amplitude of a breathing waveform) from one sample to the next, PCA will reveal a single eigenmode and a single non-zero eigenvalue that describes the variance. On the other hand, if the displacement vectors vary with complete randomness, voxel-by-voxel, from one DVF to the next, PCA will produce an infinite set of eigenmodes and eigenvalues. In the latter situation, the spectrum of eigenvalues will be flat, showing that all of the modes are indistinguishable. If the DVFs are a combination of several independent sources of spatially and temporally correlated movement and errors, then PCA will produce a set of statistically uncorrelated eigenmodes with a spectrum of eigenvalues that decrease in amplitude in proportion to the amount of variance contributed by each mode. Some of these modes will be associated primarily with breathing; other modes will be primarily due to DIR errors. In particular, correlated errors will have a decreasing eigenvalue spectrum, while random errors (e.g., white noise) will produce a flat spectrum extending effectively to infinity. Thus we can use PCA to identify and characterize the coherence of the error component of the DVF.

In this paper, we describe the concept of our method and report tests of its feasibility via numerical simulations based on a simple 4DCT DVF model. We used the simulations to test approximations in our equations and to analyze the fidelity with which the method can recover known DIR error maps. We also examine the impact of the number and location of landmarks and the number of 4DCT phases on the efficacy of the method. The simulation model was based on the characteristics of breathing observed in 4DCT and DVF data obtained from a publicly available database.<sup>10</sup>

## II. METHOD AND MATERIALS

Assume that we have a 4DCT for a patient, reconstructed in  $N+1$  breathing phases. Then one phase (e.g., end-inhale)

can be designated as the reference phase, to be used as the target image for deformable registration. Deformable registration of each of the remaining breathing phases to the initial reference phase will generate a set of  $N$  DVFs that approximates voxel-by-voxel the anatomical motion. The DVFs point from each voxel in the target reference CT to the location where the associated anatomical elements have moved in each of the subsequent phases. These DVFs contain errors from 4DCT artifacts and underlying DIR inaccuracies. Assume that we also have a set of manually delineated pointlike landmarks in the CTs that track the anatomical displacements at discrete points. Setting aside for now the effects of observer errors associated with the landmarks, we can use them both as a ground truth to establish the DVF errors at those points and as a measure of respiratory motion.

For actual breathing, the anatomical movement is highly correlated in time and space, but DIR errors add unphysical variances to the apparent motion in the DVFs. We use PCA to identify and separate the true underlying modes of motion from the contributions of DIR errors to the overall DVF variance.

Each DVF becomes a *feature vector* for PCA. We combine the feature vectors from all of the breathing phases into an ensemble called the *training set*. To decorrelate the motion and error variances, we compute the principal components of the training set. Thus, we obtain a matrix of PCA eigenvectors and eigenvalues that characterize the training set, rank ordered according to how much variance in the data each accounts for. The variance due to actual physical motion is very coherent in time and space and therefore contributes mainly to the first few (most significant) eigenmodes, while the variances due to imaging artifacts and DIR errors are not so coherent (especially in time) and thus are distributed over a much larger number of components, dominating in the higher level modes. In consequence, PCA can be used as a filter to separate physical from unphysical components in the DVF. Similar observations have been made by other authors interested in 4DCT breathing analysis.<sup>11,12</sup>

To identify the true physical breathing modes in the DVFs, we find the principal components of the validation landmark data for the 4DCT and we use this information to evaluate the contributions of the DIR errors in each of the DVFs principal modes. We can then reconstruct a feature vector in principal component space that contains only the variances due to DVF errors. The feature vector can then be turned back into a DVF error map that shows the entire spatial distribution of DVF errors rather than just the errors at the validation points.

### II.A. Computational details of the DVF error mapping procedure

Suppose that we have  $N$  DVFs, each of them with  $M$  voxels. We arrange the DVFs into  $N$  feature vectors  $\{\mathbf{v}_n\}, n=1, \dots, N$ , each with  $M$  components, to form the training set. We subtract from each feature vector the mean  $\langle \mathbf{v} \rangle$  of

the training set and then we arrange the  $N$  zero-mean column vectors in the  $M \times N$  matrix  $V$ . The covariance matrix of the training data is then

$$C = \frac{1}{N} \sum_{n=1}^N (\mathbf{v}_n - \langle \mathbf{v} \rangle)(\mathbf{v}_n - \langle \mathbf{v} \rangle)^T = \frac{1}{N} VV^T. \quad (1)$$

The  $M$  principal components (eigenvectors)  $\mathbf{u}_i$  of the training set and their associated eigenvalues  $\lambda_i$  are solutions of the eigenvalue problem

$$C\mathbf{u}_i = (1/N)VV^T\mathbf{u}_i = \lambda_i\mathbf{u}_i. \quad (2)$$

We collect the eigenvectors  $\mathbf{u}_i$  into an  $M \times M$  matrix  $U$  of  $M$  column vectors.

Principal components analysis uses the  $M$  orthonormal eigenvectors  $\mathbf{u}_i$  (the principal components) to rotate the zero-mean feature vectors  $\mathbf{v}_i - \langle \mathbf{v} \rangle$  into a new set of vectors  $\mathbf{p}_i$  that have most of the variance of the data vectors concentrated in the first few vector components

$$U^T(\mathbf{v}_i - \langle \mathbf{v} \rangle) = \mathbf{p}_i. \quad (3)$$

In this discussion we will call the  $\mathbf{p}_i$  the *principal coefficient (PC) vectors*. Each element  $p_{ji}$  of  $\mathbf{p}_i$  is maximally uncorrelated from all the others and has an associated eigenvalue  $\lambda_j$  that equals the variance of that component over the set of  $N$  feature vectors.<sup>13</sup>

If we write each DVF feature vector  $\mathbf{v}_i$  as the sum of a signal vector  $\mathbf{s}_i$  (representing the error-free physical motion) and an error vector  $\mathbf{e}_i$

$$\mathbf{v}_i = \mathbf{s}_i + \mathbf{e}_i,$$

then from PCA of the DVF we get  $U$  and  $\mathbf{p}_i$ , such that

$$U\mathbf{p}_i + \langle \mathbf{v} \rangle = \mathbf{v}_i = \mathbf{s}_i + \mathbf{e}_i, \quad (4)$$

$$U(\mathbf{p}_i - U^T\mathbf{s}_i + U^T\langle \mathbf{v} \rangle) = \mathbf{e}_i.$$

Equation (4) is an exact expression for the error map  $\mathbf{e}_i$ ; if we could determine  $U^T\mathbf{s}_i$  then we would recover the error exactly. We do not know how to estimate the signal component of the DVF directly. However, the form of Eq. (4) suggests that we can identify the term  $U^T\mathbf{s}_i$  as being approximately equal to the principal coefficients  $\mathbf{p}_{\text{sig},i}$  of the signal. Notice that the signal mean is not subtracted before rotation, but appears in a separate term in Eq. (4). This is because we cannot separate  $\langle \mathbf{s} \rangle$  from  $\langle \mathbf{e} \rangle$  in our DVF data to get their separate means.

The point now is to find a way to estimate the principal signal coefficients  $\mathbf{p}_{\text{sig},i}$ , then subtract them out to get the DVF errors from Eq. (4). First we assume that the signal is a quasiperiodic movement of anatomy from one breathing phase sample to the next and further assume that it can be completely described by the amplitude and relative phase of each anatomical voxel's displacement vector. We show in Appendix A that under these conditions, the covariance matrix of the breathing signal DVFs is determined mainly by the breathing cycle and that principal component analysis will arrive at only a few nonzero modes.

Let us assume that the validation landmarks are accurate samples of the signal. We can locate them in the target reference image, calculate the displacement vectors that point to their new locations in each of the subsequent CT phases, organize the landmark DVFs into feature vectors  $\mathbf{l}_i$ , and compute their PCA eigenvectors  $U_{\text{land}}$  and eigenvalues  $\lambda_{\text{land},j}$ . Then we get the landmark principal coefficients  $\boldsymbol{\pi}_{\text{land},i}$  (without subtracting the mean)

$$U_{\text{land}}^T \mathbf{l}_i \approx \boldsymbol{\pi}_{\text{land},i}.$$

If the landmarks are a good representative sample of the moving anatomy, then they will have the same principal modes as the complete signal vector and thus each landmark principal coefficient  $\boldsymbol{\pi}_{\text{land},ji}$  will be proportional to the corresponding signal principal coefficient to within a scaling term that accounts for the different lengths of the landmark and signal feature vectors. This assumption is argued mathematically in Appendix A. Because the principal coefficient vectors  $\mathbf{p}_{\text{sig}}$  and  $\boldsymbol{\pi}_{\text{land}}$  are just rotations of the feature vectors  $\mathbf{s}$  and  $\mathbf{l}$  they have the same corresponding lengths. Thus we can estimate  $\mathbf{p}_{\text{sig}}$  as

$$\mathbf{p}_{\text{sig}} \cong (\|\mathbf{s}\|/\|\mathbf{l}\|)\boldsymbol{\pi}_{\text{land}}. \quad (5)$$

We do not know  $\|\mathbf{s}\|$  but we can assume that the approximate equality

$$\|\mathbf{s}\|/\|\mathbf{l}\| \cong \|\mathbf{v}\|/\|\mathbf{l}_{\text{DVF}}\|$$

is true, where  $\mathbf{l}_{\text{DVF}}$  is the DVF vector (with errors included) at the landmark points. Then our estimate of the DVF error map for the  $i$ th phase is

$$\mathbf{e}_i = U(\mathbf{p}_i - (\|\mathbf{v}_i\|/\|\mathbf{l}_{\text{DVF},i}\|)\boldsymbol{\pi}_{\text{land},i} + U^T\langle \mathbf{v} \rangle). \quad (6)$$

The remainder of this proof of concept paper is devoted to a validation via a numerical simulation of our approximate solution given in Eq. (6).

Notice that we do not associate the first few eigenvalues and eigenvectors entirely with breathing and the higher order modes entirely with DIR error, and then simply divide the PCA spectrum into physical and nonphysical parts. Instead we recognize that all of the modes are a superposition of movement and error, then estimate and subtract the movement modes to reveal the residual error modes.

To compute the principal components of the DVF and the landmark feature vectors, we note that the number of samples  $N$  (i.e., 4DCT phases) is much less than the dimensionality  $M$  of the vectors. This is a case of the *high dimensionality* problem. In this case there are only  $N$  nonzero eigenvectors. We show in Appendix B how PCA can be done in this case.<sup>14</sup>

Notice that each element of  $[V^T V]_{ji} = \sum_k v_{jk} v_{ki}$  [see Eq. (B1)] is independent of the ordering of the feature vector components (i.e., voxels). Thus the eigenvalue spectrum is independent of how we organize the DVF elements into the feature vector (as long as we do it consistently in all of the sample feature vectors). We will show the significance of this below.

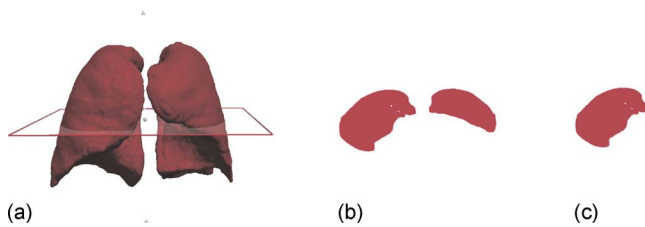


FIG. 1. The mask used to isolate the voxels of interest in the POPI data set for PCA analysis. (a) 3D view. (b) Transversal slice. (c) Only voxels in the transversal slices of the right lung contributed to the feature vectors.

## II.B. Numerical simulations of the error estimation procedure

We developed and studied our algorithm using numerical model simulations. Our model was designed to resemble the behavior of actual 4DCT DVFs. To make a plausible model, we evaluated the principal modes of breathing in a real 4DCT. We used the point-validated pixel-based (POPI) data set from the website maintained at [www.creatis.insa-lyon.fr](http://www.creatis.insa-lyon.fr).<sup>10</sup> This data set consists of a 4D fan-beam CT in which clinicians have located 40 point landmarks in each of the ten reconstructed CT breathing phases. From the same database, we also obtained two DVF sets resulting from the deformable registration of the POPI 4DCT using both the demons algorithm<sup>15</sup> and a parametric B-spline algorithm.<sup>16</sup> Each set of DVFs comes validated for accuracy at each of the point landmarks. Besides the 4DCT and the sequential DVFs for both registration methods, the website also provides the public with the set of validation landmark coordinates. To supplement the POPI validation landmarks, we measured our own independent set of landmarks in each phase of the 4DCT. We performed PCA on the POPI DVFs and on the landmarks to observe the principal components (i.e., eigenmodes) of an actual 4DCT and then used the characteristics of the principal components to construct our numerical model. Before describing our numerical model and tests, we will summarize these pertinent DVF eigenmode characteristics.

The POPI 4DCT consists of ten CT image phases. Deformable registration of these CTs will generate nine DVFs. Hence, the data sample (i.e., the PCA training set) associated with the breathing motion of this single patient consists of a set of nine observations.

We prepared our PCA feature vectors in the following way:

- (1) We used the DVFs created by registering all the breathing phases to phase 1 (full inhale), as provided by the POPI model.
- (2) We segmented out the lung tissue on the phase 1 CT image, thus generating a “mask” for the lung voxels (Fig. 1).
- (3) We further selected from the mask only one lung lobe (right).
- (4) We selected one slice only in the lung mask.
- (5) In all the DVFs, we selected the voxels of the lung mask and we considered only the vectors originating in those voxels.
- (6) For all the vectors selected by the above procedure we considered only the z component.

By selecting our feature vectors in the way described in the above, we sampled the breathing modes mainly along the z axis. Considering separately the x, y, and z components of the breathing movement greatly simplifies our PCA problem and suffices for a proof of concept application.

Principal component analysis of these feature vectors generated eight significant PC eigenvectors. In Fig. 2 we show a plot of the variances (eigenvalues) associated with these eigenvectors. The first three modes carry more than 96% of the total variance. We can compute a DVF “signature” for each PC mode in the following way:

- (1) We compute the PC eigenvectors and principal coefficient vectors [Eq. (1)].
- (2) We use the matrix  $U$  [Eq. (1)] with only the  $i$ th PC

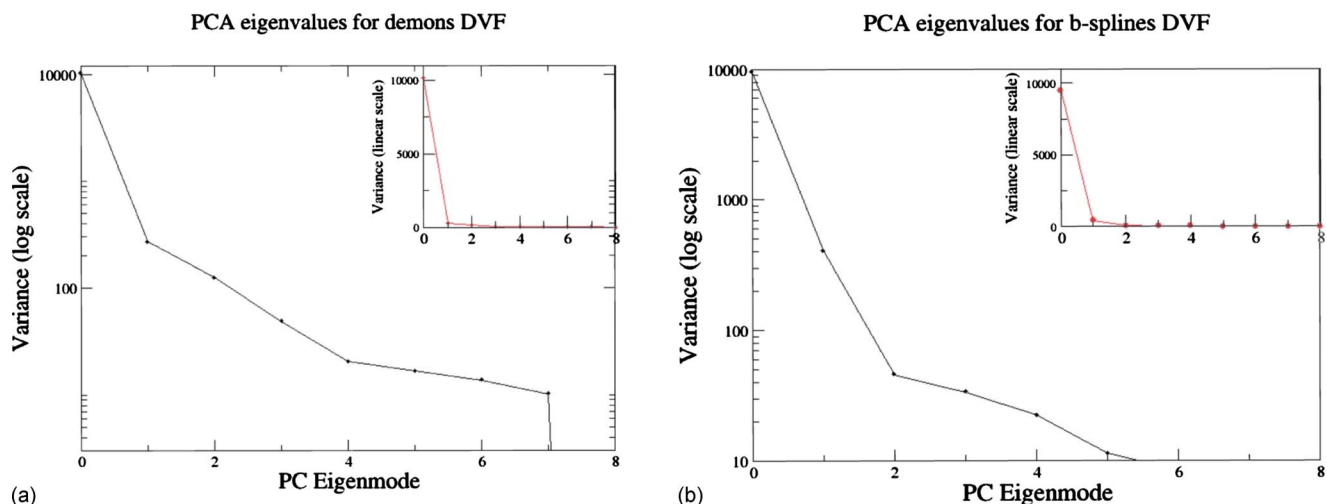
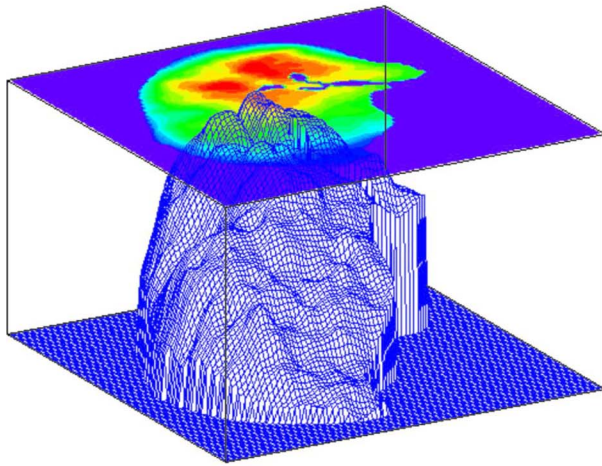


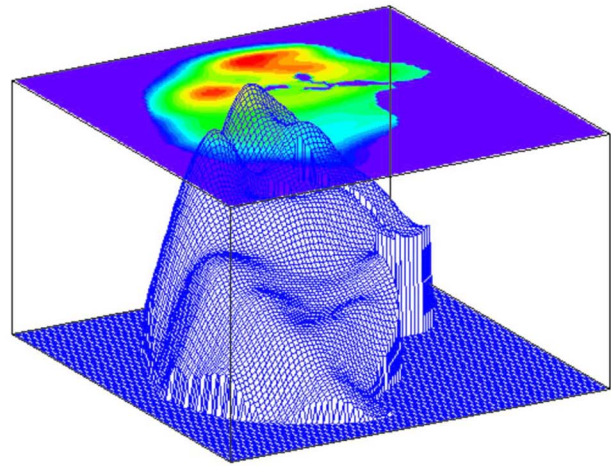
FIG. 2. The eigenvalue spectra for (a) the demons DVFs and (b) the parametric B-spline DVFs.



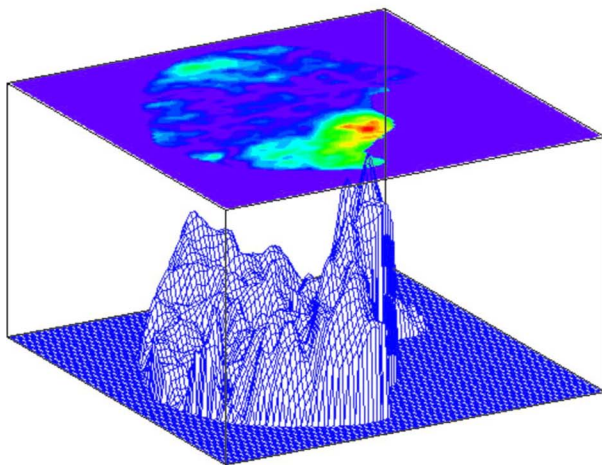
PCA mode 1 demons



PCA mode 1 b-splines



PCA mode 2 demons



PCA mode 2 b-splines

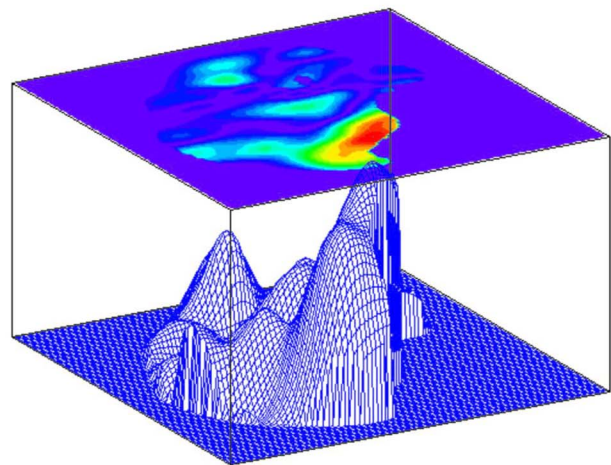


FIG. 3. The PC signatures for the first two PC modes. In the right column we show the first two modes in the demons method DVFs; in the left we show the same two modes in the B-splines DVFs.

eigenvector to rotate the principal coefficient vectors back into DVF space, which projects out the  $i$ th eigenmode of the DVF.

- (3) For every voxel, we combine the absolute values of the variation of the projected eigenmode across the different phases, obtaining a map of the variability associated with the  $i$ th PC eigenmode across the breathing phases. We call this map a “PC signature of the  $i$ th mode.”

Obviously, we would expect the PC signatures of the modes associated with real physical breathing motion to be approximately the same for the two different sets of DVFs. In Figs. 3 and 4 we show the signature maps for the first three modes and one example of the signature maps for the higher order modes. From the figures it can be seen that the two sets of signatures show equivalent, easy to recognize features for the first two PC modes. The third mode contains more of the registration error components and it starts already to show significant differences between the two sets of

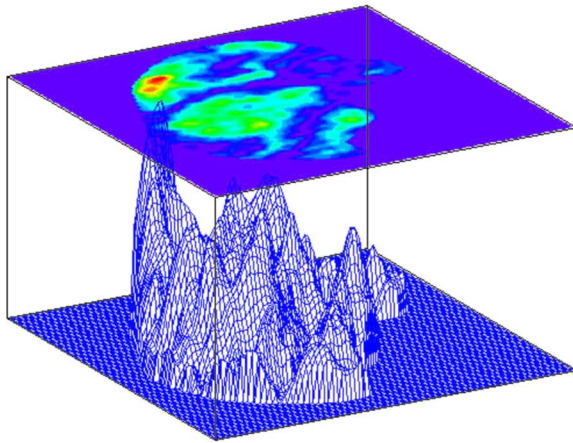
DVFs. Finally, the higher order modes, as illustrated in Fig. 6, have basically no recognizable shared features in the two DVFs. As the DVFs are describing the same movement of the same anatomy, the only conclusion can be that the higher PC eigenmodes mainly contain noise.

Suppose that there are regions in the lung where the anatomy moves approximately in proportion to a simple amplitude scaling function, (i.e., the breathing amplitude), such that

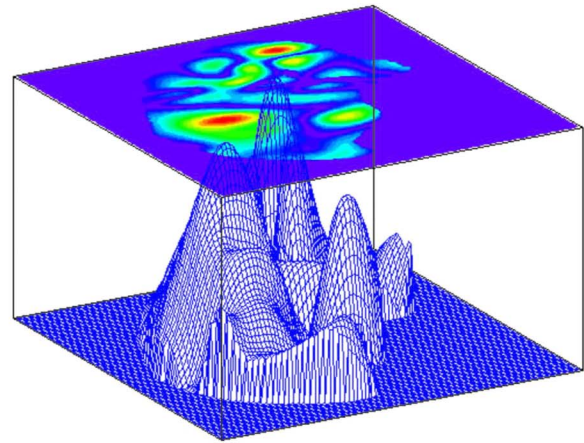
$$u(\mathbf{x}, t_i) \approx \alpha(t) u(\mathbf{x}, t_{i-1}). \quad (7)$$

In these regions, we would expect that the ratio of displacement vectors from one phase to the next would be approximately uniform. We demonstrate this for the POPI data by taking the ratios, voxel-by-voxel, of “consecutive” DVF phases (e.g., the ratio of the DVF connecting the 4DCT phases 1–5 to the DVF connecting the 4DCT phases 1–4), as shown in Figs. 5 and 6. One noticeable feature in most of

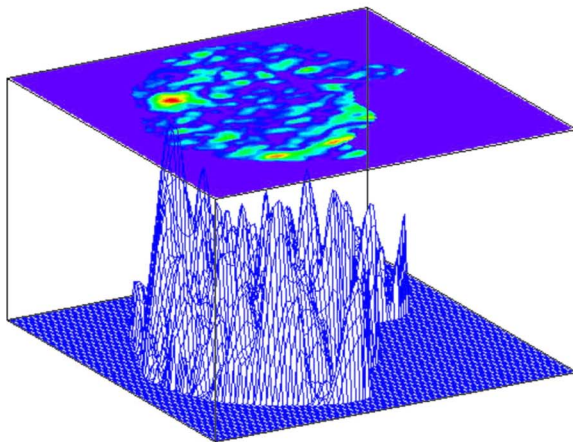
PCA mode 3 demons



PCA mode 3 b-splines



PCA mode 7 demons



PCA mode 7 b-splines

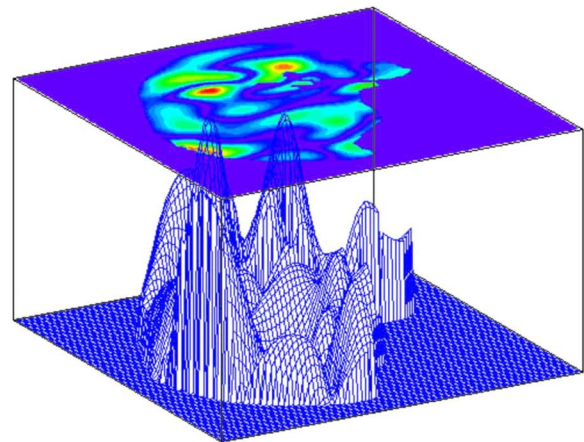


FIG. 4. The PC signatures for modes 3 and 7. In the right column we show the two modes in the demons method DVFs; in the left we show the same two modes in the B-splines DVFs.

these ratios is the presence of large patches in which the ratio of the z components has an almost constant value. This quasi-uniformity of the successive DVF ratios across the slice indicates that the movement between two successive breathing phases can be well approximated as an amplitude scaling, as in Eq. (7), and that all the voxels within a uniform patch are moving in phase.

If the sequence of DVFs was strictly governed by a relation such as Eq. (7), with all voxels moving in phase, then the covariance matrix would be proportional to the outer product of one feature vector and its transpose, which has rank 1. There would then be only one principal mode (see Appendix A). If the voxels move with different phases and the amplitude function  $\alpha(t)$  is a sine or cosine function, then the covariance matrix can be separated into a sine and cosine component, each with rank 1; the complete matrix then has rank 2 and there are two principal modes that are exactly  $90^\circ$  out of phase (see Appendix A). As the amplitude function becomes more complex, then voxel-by-voxel phase differences cause higher order modes to appear. Because normal

Voxel-by-voxel successive DVFs ratio

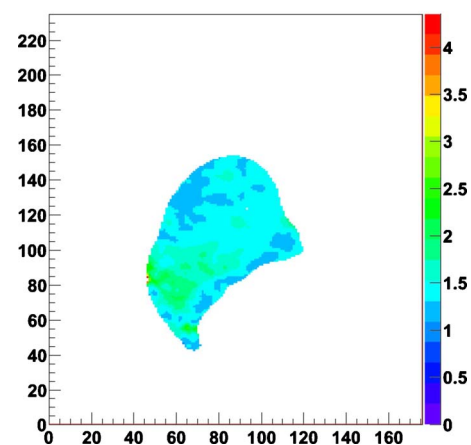


FIG. 5. Example of ratio of the z components of two POPI DVFs for the voxels in one CT slice as shown in Fig. 2(c). For this particular figure we considered the ratios of the DVF between phases 1 and 5 and between phases 1 and 4.



Voxel-by-voxel successive DVFs ratio

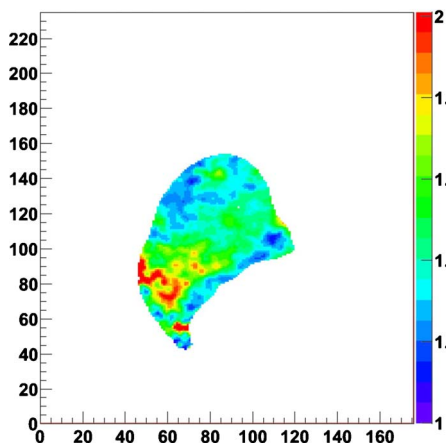


FIG. 6. The figure with a zoom on amplitude.

breathing is regular and periodic, this supports our expectation that a DVF dominated by breathing motion should have only a few strong modes and that higher order modes will be mainly due to noise.

Figure 7 shows the eigenvalue spectra of the POPI landmarks and our own landmarks. This shows that random subsamples of vector components from a training set of feature vectors will have the same eigenvalue spectrum as the original training set. This is a necessary condition for our assumption that the DVF breathing signal PCA spectrum can be approximated by the landmark samples. Figure 8 shows the phase-to-phase temporal variation of the first two principal coefficients of the landmarks. Notice how the landmark principal coefficients track the breathing cycle but with a relative phase shift of approximately  $90^\circ$ . This conforms to our expectations based on the above discussion.

In order to test our error reconstruction method, we constructed a simple 2D numerical simulation to emulate as much as possible the POPI 4DCT and the associated DVFs. Our typical model consists of a  $128 \times 128$  pixel square DVF describing motion in the  $(x, y)$  plane. We assumed that during breathing, the displacement of each anatomical point was

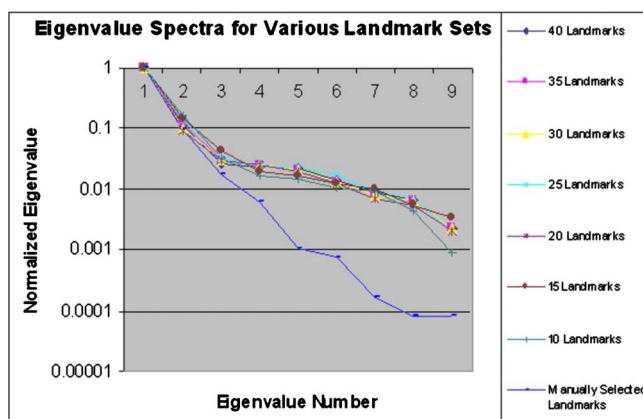


FIG. 7. The eigenvalue spectra for subsets of the POPI point landmarks and for our manually selected landmarks.

First Two Principal Coefficients for Manually Selected Landmarks

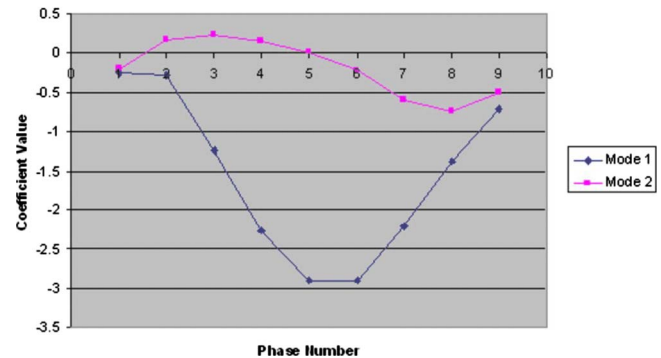


FIG. 8. The time dependence of the two strongest principal components of the DVFs computed from the POPI data.

completely described by its relative amplitude and phase. (NB: The relative movement phases in the breathing model are not the same as the phase bins used to describe the individual time steps of a 4DCT.) We made a reference DVF that varied spatially as a simple half-sine function in  $x$  and  $y$ . The DVF amplitude at each pixel was then varied in time in proportion to a measured (i.e., nonsinusoidal) breathing signal to produce a temporal sequence of anywhere from  $N = 5$ –200 DVFs spanning multiple breathing cycles. This way we could simulate a temporal sequence of thoracic DVFs from a 4DCT of arbitrary duration, which allowed us to see the effects of sample size  $N$  on the eigenmode and DVF error analysis. To simulate relative phase differences within the DVF, we shifted the relative breathing phases of individual pixels according to various spatial patterns. As the results will show, this model emulates the essential features of the breathing eigenmodes observed in the POPI 4DCT.

We modeled four different relative phase scenarios: (1) All the pixels in the DVF moved in phase; (2) alternating pixels had a phase shift; (3) pixels were randomly assigned one of three different phases; and (4) the  $(x, y)$  plane was divided in three sectors and each sector moved with a different phase. These DVFs represented the signal feature vectors [i.e.,  $s_i$  in Eq. (4)]. We randomly selected anywhere from 5 to 1000 pixels from the signal vector as ground truth landmarks to fill the landmark feature vectors. Then we generated simulated maps of spatially correlated errors. The error maps were created by generating a white-noise image with a Gaussian noise distribution of width  $\sigma$  (expressed as a percent fraction of the mean image intensity), taking its Fourier transform, multiplying the real and imaginary Fourier coefficients by the noise power spectrum (NPS) in Eq. (8),

$$P(k) = e^{-k\lambda}(1 - e^{-k\lambda}), \quad (8)$$

and then taking the inverse Fourier transform to recover a spatial noise image. The error map created in this way had the NPS spectrum of Eq. (8). The factor  $\lambda$  determined the correlation length (i.e., the degree of spatial correlation) of the errors in each image. The error map was then added to the simulated breathing DVF. The process was repeated

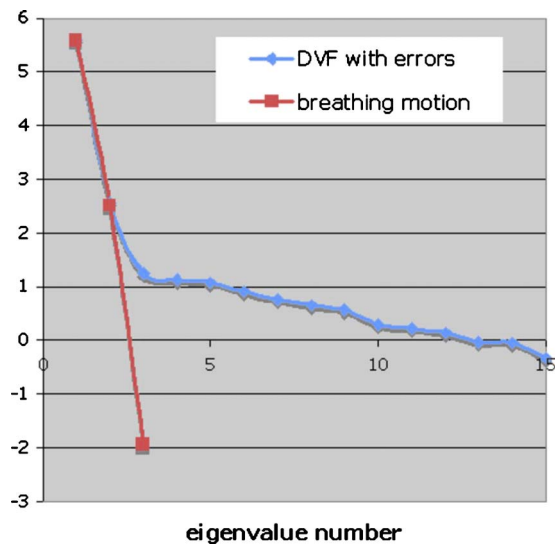


FIG. 9. The eigenvalue spectra for the simulated three-phase breathing motion and for the DVF with errors. The landmark spectrum is identical to the breathing spectrum.

phase-by-phase, each time generating a new random 2D noise image.

The error maps were spatially structured but varied randomly in amplitude from one DVF to the next. The error amplitudes were generated independently of the DVF amplitude in a given voxel. On average, the error magnitude was 10%–15% of the maximum DVF amplitude. The  $N$  noisy DVFs made up our set of feature vectors. We then applied our technique of PCA error reconstruction to the simulated data and estimated the error maps for comparison to the known errors. To evaluate the similarity in the spatial structure of the estimated and actual error maps, we calculated the normalized cross-correlation between the two maps.

### III. RESULTS

When PCA is applied to our hypothetical breathing model (the simulated DVFs without errors), it identifies only as many principal modes as there are breathing phases present in the motion. When all anatomical points move in one phase there is exactly one eigenmode (as explained above); when they move with two relative phases there are two eigenmodes; etc. The higher modes are the combined effect of multiple phases and a nonsinusoidal breathing waveform. Each new mode adds progressively less variance to the cumulative motion, so that even if there are more than three or four modes present, the higher order modes are negligible. (Recall that in this context, “phase” refers to the relative motion of different voxels in the DVF, not to the time sequence of the DVFs.) When noise is added, additional higher order modes appear in the eigenvalue spectrum. We see this in Fig. 9, which shows the eigenvalue spectra of the simulated DVFs, with and without added noise, for three relative phases of breathing motion. The eigenvalue spectrum of our simulated DVFs shares similar characteristics with the POPI spectrum.

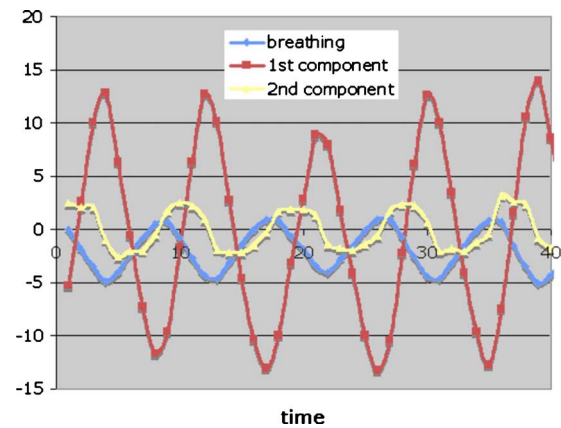


FIG. 10. The two nonzero principal components of a DVF moving with two breathing phases, versus time, together with the breathing amplitude.

When one uses the PCA eigenvector matrix to rotate the temporal sequence of DVF feature vectors into a sequence of principal coefficient vectors  $\mathbf{p}_i$ , via Eq. (1), the components of each  $\mathbf{p}_i$  decrease in amplitude in proportion to the amount of residual variance that they describe. Each component of  $\mathbf{p}$  fluctuates in time with the breathing amplitude. Figure 10 shows this for the simulated case of two modes associated with two phases of alternating pixels. The two principal components follow the breathing amplitude and are approximately  $90^\circ$  out of phase with each other, as expected from the discussion above. This is what we also saw in the PCA of the POPI data (cf. Fig. 8).

As we remarked earlier while discussing Eq. (B1), the eigenvalue spectrum is independent of the ordering of the DVF voxels in the PCA feature vector. This means that it is independent of the spatial pattern of phase variations. We saw numerically that this is the case: Arranging the three phases in our model randomly or by sector in the  $(x, y)$  plane gave the same result. This leads us to suspect that true breathing motion should have a comparatively simple PCA spectrum with only a few dominant modes associated mainly with phase differences. Any additional modes in the DVFs will be due to image noise and artifacts rather than physical motion. Again, this is what we saw in the POPI data. This is the basis for our hypothesis that PCA can be used as a filter to separate motion from noise.

To test the importance of sample size  $N$  on the convergence of the principal component solution, we downsampled the DVFs to 450 and then to 200 pixels and calculated the eigenvalues of the simulated data with three phases while increasing the number of DVF samples from  $N=5$  to  $N=200$ . Figure 11 shows the first eigenvalue for each size vector versus training set size  $N$ . In both cases the eigenvalue approached a reasonably stable value once there were about  $N \geq 10$  samples in the training set. For the subsampled DVFs, the error map converged to an accurate representation of the actual errors when there were approximately 20 or more landmarks.

We then made a matrix of calculations of the estimated  $128 \times 128$  pixel error map while varying the sample size



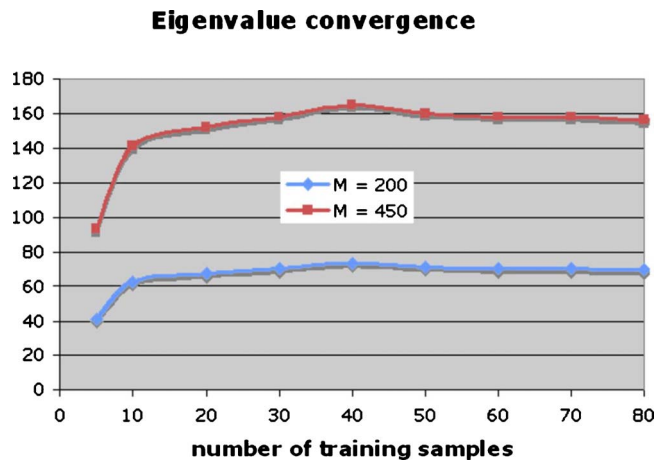


FIG. 11. Convergence of the first eigenvalue of  $10 \times 10$  and  $15 \times 15$  pixel 2D DVFs as the number of training samples increases.

from  $N=5$  to  $N=100$  and the number of landmarks from 5 to 1000. For each test case in the matrix, we calculated the normalized cross-correlation of the estimated and actual error maps for all breathing phases. Figure 12(a) shows one of the simulated DVFs with the added errors. The magnitude of the displacement at each pixel is given by the color scale. Figure 12(b) shows the actual map of errors in Fig. 12; Fig. 12(c) is the map estimated from 50 samples and 40 landmarks using our PCA method. The cross-correlation between these two maps is 0.959. Table I summarizes the cross-correlations for the entire matrix of test cases while Fig. 13 displays the results graphically. We observed that, for 40 or more landmarks and 15 or more time samples, our estimated error map consistently showed a strong spatial correlation with the known error map ( $C > 0.90$ ), while the error amplitude was generally underestimated by 10%–30%. We also observed that the number of landmarks required for good accuracy increased very slowly (from 20 to 40) when the size of the DVF increased from 200 to 16 384 pixels.

#### IV. DISCUSSION

The validation of deformable image registration results must contend with limited ground truth data and registration errors that can have complex spatial distributions. Consequently, point-by-point landmark validation can give unrepresentative results that do not realistically reflect the global pattern of registration errors.<sup>17</sup> We have tested a new concept for DIR error evaluation, which we hope will allow us to map the complete spatial distribution of DVF errors, when only a small number of validation landmarks are available. The new method applies PCA to the landmark movement to estimate the fundamental eigenmodes of breathing and then subtracts these modes from the principal modes of each complete DVF to expose the underlying errors. This makes possible a new and more realistic picture of deformable image registration errors than is customarily obtained via simple point landmark validation.

Our method requires a set of DVFs generated by a particular DIR method and assumes that the DVFs record ana-

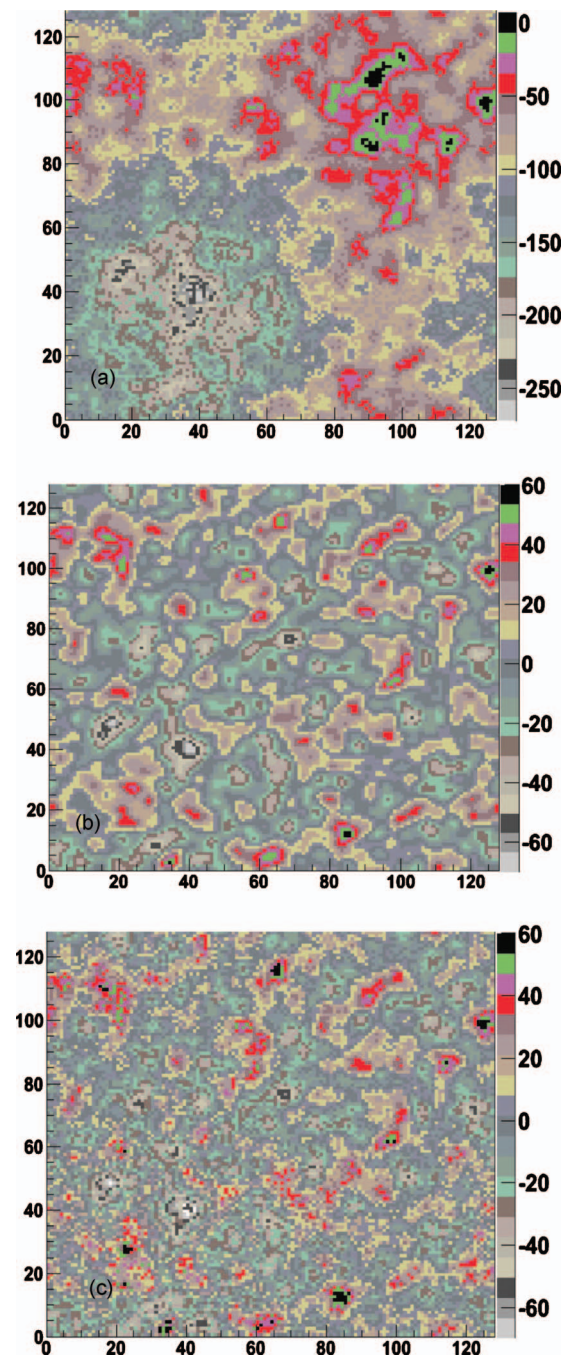


FIG. 12. (a) The DVF with errors. (b) The actual error map. (c) The estimated error map. The color scale for the error maps is centered with yellow at zero to show the sign of the errors and is magnified six times relative to the DVF displacements.

tomical motion that has spatial coherence and some degree of continuity from one DVF to the next. A good example of that (and the problem that motivated us to develop the method) is a set of DVFs obtained by registering the phases of a 4DCT. Using a simple numerical simulation of such a data set, for which we could define an absolute ground truth for the entire spatial error map, we have shown that our method is capable of recovering a reasonable approximation of the known errors. In particular, we have shown that it can accurately recover the spatial structure in the error map.

TABLE I. Correlation coefficients between true and estimated error maps for simulation data.

Number of samples	Number of landmarks							
	1000	50	40	30	20	15	10	5
100	0.965 249	0.970 613	0.971 227	0.882 481	0.845 844	0.805 168	0.693 698	0.554 951
50	0.954 346	0.958 393	0.958 901	0.959 96	0.960 524	0.904 191	0.844 078	0.730 723
40	0.952 898	0.955 26	0.955 717	0.956 624	0.957 441	0.958 186	0.862 409	0.783 023
30	0.946 151	0.948 507	0.949 305	0.949 683	0.950 201	0.950 848	0.947 608	0.725 655
20	0.931 103	0.931 153	0.931 782	0.932 206	0.933 414	0.933 758	0.933 944	0.802 656
15	0.896 587	0.895 343	0.895 296	0.8968	0.896 428	0.896 163	0.896 609	0.845 068
10	0.880 022	0.882 914	0.882 637	0.883 397	0.881 115	0.881 791	0.881 509	0.8737
5	0.652 757	0.650 382	0.666 681	0.669 438	0.664 782	0.655 753	0.654 319	0.650 606

Principal components analysis is a logical solution to the problem of analyzing spatially correlated data because it projects the data onto a set of orthogonal eigenvectors that maximally decorrelate each projected mode of variance from all of the others. This allows us to identify the error modes and the breathing modes regardless of their patterns of spatial coherence.

It might be tempting to simplify the problem by assuming that the first few principal DVF modes are associated entirely with breathing while the noise and registration errors are associated entirely with the higher order modes (cf. Ref. 11). If this would be the case, then we could simply cut the eigenvalue spectrum at the appropriate point and reconstruct the errors from the higher modes. Then we would avoid the need to estimate the breathing components from the landmarks. However, as Eq. (5) shows, each principal coefficient combines contributions from breathing and registration errors, i.e., although the modes are statistically decorrelated, they are not mathematically separated. [When we attempted to reconstruct the error map just from the higher modes without following the methodology of Eq. (6), the resulting estimated error map was very inaccurate.]

Within our simplifying assumptions, we have found that our method gives a robust reconstruction of the error map when there are ten or more DVF samples (i.e., breathing phases). This makes it practical for a typical 4DCT of 10–25 phases.<sup>18</sup> We also found that the method gave good results with as few as 20–40 landmarks. Both of these results were very nearly independent of the number of pixels in the DVF, which offers some assurance that the reconstruction of a full 3D volumetric map will not require substantially more time bins or landmarks. However, it is important to ensure that the landmarks provide a reasonable sampling of the breathing modes. For example, if the landmarks cluster in one part of the thorax, but the breathing modes are asymmetrically distributed in the thorax, the landmarks might provide not only an incomplete picture of the DVF error distribution but also an incomplete picture of respiratory motion. By making DVF difference maps as in Figs. 5 and 6, one can select landmarks such that there are at least a few in each of the regions associated with a principal breathing mode.

Our method is in an early stage of development. So far, we have only used numerical simulations of 4DCT DIR results to test the general principles of the method and to validate our basic assumptions leading up to Eq. (6). These simulations were designed to be simple approximations to the basic behavior one would expect in 4DCT DVFs, so that we could perform basic proof of concept tests of the potential merit of the method. We used the POPI data to show that the basic properties of our simulation, as well as the basic properties of its PCA eigenmodes, are roughly consistent with a real 4DCT. The simulation was a long way from a realistic model of human respiratory motion as it is captured in a 4DCT, but it did show that our method has the basic capability of recovering spatially correlated error maps.

We are far from a complete validation of this technique. Our simulation assumed that the landmarks were a perfect ground truth when in reality they will have observer uncertainties that will introduce their own errors into the analysis. Already in our PCA analysis of the landmark movement, we found different eigenvalue spectra for different sets of landmarks, each with tails that are quite likely associated with observer inaccuracies. We must now analyze how those observer errors might confound our DIR error map. Further-

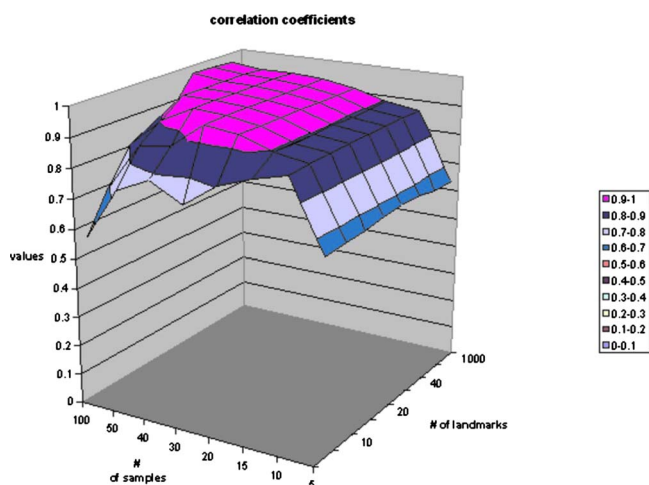


FIG. 13. The cross-correlation of the estimated and actual error maps as a function of the number of breathing phases (i.e., DVF samples) and the number of validation landmarks.

more, our simulation of a sequence of 4DCT DVFs was very simplistic when compared to real breathing motion. The next stage of testing will require physically realistic 4DCT and DVF data, which we intend to acquire with a 3D deformable imaging phantom rather than via more complicated simulation models. Realistic tests must also investigate the effects of irregularities in breathing (e.g., hiccups) that could affect the PCA breathing spectrum. Finally, it will be important to enlarge the PCA training set. A single 4DCT typically has only 10–15 phases, which is at the limit for which our method can work effectively. We can enlarge our training set in two ways: Either by obtaining multiple 4DCTs for a single patient, or by developing the means to assimilate 4DCTs from a cohort of patients. In both cases, we will try to exploit the assumption that breathing motion has the basic elements of regularity for individuals as well as across a population. However, because the use of DIR in the radiation therapy clinic will often involve patients with lung disease, the development of a 4DCT PCA training set across a cohort of patients must deal with the likelihood of substantially different breathing behaviors for some cases.

## V. SUMMARY

We have presented here a proof of concept study for an improved method to analyze deformable image registration errors when the only ground truth is provided by a small set of measured landmarks. The method explicitly accounts for the spatial coherence of registration errors and is thus more physically realistic than an analysis that assumes that the errors are randomly distributed.

## ACKNOWLEDGMENTS

This work was supported in part by NIH Grant No. P01-CA116602.

## APPENDIX A: THE PRINCIPAL MODES OF QUASIPERIODIC BREATHING

Assume that voxel-by-voxel breathing motion can be completely described by the amplitude and temporal phase of the DVF vector field. In the simplest possible situation, suppose that our signal in Eq. (4) is a DVF for which every voxel moves in phase in proportion to some time-dependent breathing amplitude function  $a(t)$ . Let the error component be zero, so that  $V$  just contains the true DVF breathing signal. Let our  $N$  time samples of the DVF be taken at the breathing amplitude steps  $a(t_j)=a_j$ . If  $v_0$  is the feature vector representing the DVF at the reference time  $t=0$ , then each DVF feature vector at later time  $t_j$  is given by

$$v_j = v_0 a_j.$$

If we write the initial DVF  $v_0$  and the breathing amplitude samples  $a_j$  as column vectors, then our matrix of DVF feature vectors from Eq. (2) can be written as

$$V = v_0 a^T.$$

From this it follows that

$$VV^T = v_0(a^T a)v_0^T = \|a\|v_0v_0^T.$$

Similarly,

$$V^T V = \|v_0\|aa^T. \quad (A1)$$

The column vectors  $v_0$  and  $a$  have rank 1.

$$\text{rank}(AA^T) = \text{rank}(A^T A) = \text{rank}(A).$$

Thus  $VV^T$  and  $V^T V$  have rank 1 and Eqs. (2) and (B1) have exactly one nonzero eigenvalue and eigenvector. From Eqs. (A1) and (B1), we see that all of the covariance in the DVF, and thus in the PCA solution, comes from the time dependence of  $a$ . Also, from Eqs. (A1) and (B1), it is apparent that the one nonzero PCA eigenvalue depends only on the magnitude of the reference DVF feature vector. This feature vector could contain all of the voxels or only selected landmarks. Thus, for this simplest of all cases, we hypothesize that Eq. (5) will hold

$$p_{\text{sig}} \cong (\|s\|/\|l\|)\pi_{\text{land}}.$$

More realistically, suppose that the breathing amplitude  $a(t_j)$  is approximately periodic but each voxel  $i$  can move with its own relative phase  $\phi_i$ . For example, approximate the breathing as a sine wave

$$a_i(t_j) = \sin(t_j + \phi_i) = \sin t_j \cos \phi_i + \cos t_j \sin \phi_i,$$

$$V_{ij} = (v_{0i} \cos \phi_i) \sin t_j + (v_{0i} \sin \phi_i) \cos t_j.$$

As above, we write

$$v_{1i} = v_{0i} \cos \phi_i \quad \text{and} \quad v_{2i} = v_{0i} \sin \phi_i$$

as initial DVF vector amplitudes, let

$$a_j = \sin t_j \quad \text{and} \quad b_j = \cos t_j$$

and write

$$V = v_1 a^T + v_2 b^T.$$

We observe that

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$

and conclude that  $\text{rank}(V)=2$  for this case. If so, then  $\text{rank}(VV^T)=2$ . This predicts that for a sinusoidal breathing amplitude and any arbitrary combination of voxel-by-voxel phase differences, the DVF will have exactly two nonzero eigenvalues and eigenvectors. If one sets up a numerical computation of this situation, then that is in fact what PCA shows.

For this case we now have

$$V^T V = \|v_1\|aa^T + \|v_2\|bb^T + (v_1^T v_2)ab^T + (v_2^T v_1)ba^T. \quad (A2)$$

Again the covariance is due to the time variation of the DVF due to the breathing amplitude represented by  $a$  and  $b$ . The DVF signal (landmark) reference vector enters as a vector magnitude. However, if we are only going to use selected landmarks to estimate our signal modes, we see from Eq. (A2) that they should be representative of the different phases. From this, Eq. (5) arises naturally. We rely on the numerical simulation to validate these assumptions.



If the breathing function is more complicated than a simple sine wave, then more modes will appear in the principal components of the DVF, but the PCA spectrum will still be dominated by just two or three modes for all but the most irregular breathing patterns. That is in fact what we see in the POPI data.

## APPENDIX B: PRINCIPAL COMPONENTS CALCULATION FOR HIGH-DIMENSIONALITY DATA

Multiply both sides of Eq. (2) by  $V^T$

$$(1/N)V^T V V^T \mathbf{u}_i = \lambda_i V^T \mathbf{u}_i.$$

Let  $\mathbf{w}_i = V^T \mathbf{u}_i$  so that we get the new eigenvalue equation for the  $N \times N$  matrix  $V^T V$

$$(1/N)V^T V \mathbf{w}_i = \lambda_i \mathbf{w}_i, \quad (\text{B1})$$

which has the same first  $N-1$  eigenvalues as the original  $M \times M$  matrix  $V V^T$  (which has  $M-N-1$  additional eigenvalues). To get the eigenvectors of Eq. (2), multiply by  $V$

$$(1/N)(V V^T) V \mathbf{w}_i = \lambda_i V \mathbf{w}_i,$$

from which we see that  $V \mathbf{w}_i = \mathbf{u}_i$  to within a normalization factor. To complete the derivation we normalize

$$\mathbf{u}_i = (1/(N\lambda_i)^{1/2}) V \mathbf{w}_i. \quad (\text{B2})$$

<sup>a)</sup> The authors report no conflicts of interest in conducting the research.

<sup>b)</sup> Electronic mail: cvaman@vcu.edu

<sup>c)</sup> Author to whom correspondence should be addressed. Electronic mail: mjmurphy@vcu.edu

<sup>1</sup> G. S. Mageras, "Respiration correlated CT techniques for gated treatment of lung cancer," *Radiother. Oncol.* **64** (Supplement 1), S75–S75 (2002).

<sup>2</sup> D. A. Low *et al.*, "A method for the reconstruction of four-dimensional synchronized T scans acquired during free breathing," *Med. Phys.* **30**(6), 1254–1263 (2003).

<sup>3</sup> P. G. Keall *et al.*, "Acquiring 4D thoracic CT scans using a multislice helical method," *Phys. Med. Biol.* **49**(10), 2053–2067 (2004).

<sup>4</sup> J. Orban de Xivry *et al.*, "Tumour delineation and cumulative dose computation in radiotherapy based on deformable registration of respiratory correlated CT images of lung cancer patients," *Radiother. Oncol.* **85**, 232–238 (2007).

<sup>5</sup> M. Birkner *et al.*, "Adapting inverse planning to patient and organ geometrical variation: Algorithm and implementation," *Med. Phys.* **30**, 2822–2831 (2003).

<sup>6</sup> P. Zhang, G. D. Hugo, and D. Yan, "Planning study comparison of real-time target tracking and four-dimensional inverse planning for managing patient respiratory motion," *Int. J. Radiat. Oncol., Biol., Phys.* **72**(4), 1221–1227 (2008).

<sup>7</sup> K. K. Brock *et al.*, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," *Int. J. Radiat. Oncol., Biol., Phys.* **76** (2), 583–596 (2010).

<sup>8</sup> 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," *Med. Phys.* **35** (3), 1008–1018 (2008).

<sup>9</sup> R. Castillo *et al.*, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**, 1849–1870 (2009).

<sup>10</sup> J. Vandemeulebroucke, D. Sarrut, and P. Clarysse, "The POPI-model, a point-validated pixel-based breathing thorax model," in the Proceedings of the 15th International Conference on the Use of Computers in Radiation Therapy (ICCR), Toronto, Canada, 2007.

<sup>11</sup> Q. Zhang, A. Pevsner, A. Hertanto, Y. C. Hu, K. E. Rosenzweig, C. C. Ling, and G. S. Mageras, "A patient-specific respiratory model of anatomical motion for radiation treatment planning," *Med. Phys.* **34**(12), 4772–4781 (2007).

<sup>12</sup> T. Klinder, C. Lorenz, and J. Ostermann, *Free-Breathing Intra- and Inter-Subject Respiratory Motion Capturing, Modeling, and Prediction* (SPIE Medical Imaging, Orlando, FL, 2009).

<sup>13</sup> J. Kittler and P. C. Young, "A new approach to feature selection based on the Karhunen-Loeve expansion," *Pattern Recogn.* **5**, 335–352 (1973).

<sup>14</sup> C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).

<sup>15</sup> X. Pennec, P. Cachier, and N. Ayache, "Understanding the demon's algorithm: 3D non-rigid registration by gradient descent," in *Medical Image Computing and Computer-Assisted Intervention MICCAI '99*, Cambridge, United Kingdom, edited by C. Taylor and A. Colchester (Springer-Verlag, Berlin, 1999) [Lect. Notes Comput. Sci. **1679**, 597–605 (2009)].

<sup>16</sup> J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.* **12**(11), 1427–1442 (2003).

<sup>17</sup> S. Kabus, T. Klinder, K. Murphy, B. van Ginneken, C. Lorenz, and J. P. W. Pluim, "Evaluation of 4D-CT lung registration," in Proceedings of MICCAI 2009, 2009, edited by G. Z. Yang *et al.*, pp. 747–754.

<sup>18</sup> P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. E. Murray, C. R. Ramsey, M. B. Van Herk, S. S. Vedam, J. W. Wong, and E. Yorke, "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med. Phys.* **33**, 3874–3900 (2006).