

# PET-CT Image Registration in the Chest Using Free-form Deformations

David Mattes, David R. Haynor, Hubert Vesselle, Thomas K. Lewellen, and William Eubank

**Abstract**—We have implemented and validated an algorithm for three-dimensional positron emission tomography transmission-to-computed tomography registration in the chest, using mutual information as a similarity criterion. Inherent differences in the two imaging protocols produce significant nonrigid motion between the two acquisitions. A rigid body deformation combined with localized cubic B-splines is used to capture this motion. The deformation is defined on a regular grid and is parameterized by potentially several thousand coefficients. Together with a spline-based continuous representation of images and Parzen histogram estimates, our deformation model allows closed-form expressions for the criterion and its gradient. A limited-memory quasi-Newton optimization algorithm is used in a hierarchical multiresolution framework to automatically align the images. To characterize the performance of the method, 27 scans from patients involved in routine lung cancer staging were used in a validation study. The registrations were assessed visually by two expert observers in specific anatomic locations using a split window validation technique. The visually reported errors are in the 0- to 6-mm range and the average computation time is 100 min on a moderate-performance workstation.

**Index Terms**—Computed tomography (CT), deformation, multimodality, multiresolution, mutual information, nonlinear, nonrigid, positron emission tomography (PET), registration, validation.

## I. INTRODUCTION

**G**IVEN two image sets acquired from the same patient but at different times or with different devices, image registration is the process of finding a geometric transformation  $g$  between the two respective image-based coordinate systems that maps a point  $\mathbf{x}$  in the first image set to the point  $g(\mathbf{x})$  in the second set that has the same patient-based coordinates, i.e. represents the same anatomic location. This notion presupposes that the anatomy is the same in the two image sets, an assumption that may not be precisely true if, for example, the patient has had a surgical resection between the two acquisitions. The situation becomes more complicated if two image sets that reflect different tissue characteristics [e.g. computed tomography (CT) and positron emission tomography (PET)] are to be regis-

tered. The idea can still be used that, if a candidate registration  $g$  matches a set of similar features in the first image to a set of features in the second image that are also mutually similar, it is probably correct [23]. For example, according to the principle of mutual information, homogeneous regions of the first image set should generally map into homogeneous regions in the second set [1], [2]. The utility of information theoretic measures arises from the fact that they make no assumptions about the actual intensity values in the images, but instead measure statistical relationships between the two images [24]. Rather than measure correlation of intensity values in corresponding regions of the two images, which may be low or negative, mutual information measures statistical dependence of intensities in these regions. The mutual information metric has been effective in various applications where the requisite transformations are linear [2], [3] and more recently in cases involving nonlinear motion descriptions [4], [7].

We have concentrated our efforts on PET-to-CT image registration in the chest, where we attempt to fuse images from a modality with high anatomic detail (CT) with images from a modality delineating biological function (PET). Although PET is a functional imaging modality, a transmission (TR) image is acquired immediately before acquisition of the emission image and is, therefore, in near-perfect registration with the functional scan. The TR image is similar to a CT attenuation map but it uses a higher energy radiation beam, resulting in less soft-tissue detail than the CT, and detector configuration limits its spatial resolution. Once the TR and CT images are registered, the resulting transformation can be applied to the emission or standard uptake value (SUV) image for improved PET image interpretation. Our intentions are to provide a deformation that aligns the TR image with the CT image and to present the visually assessed accuracy achievable by this method.

Patient and anatomic motion during acquisition blurs the images. The sharpness of boundaries in CT scans is obtained by having the patient maintain maximum inspiration during the 30 s required for acquisition. To avoid attenuation of the X-ray beam by the arms, the patient also holds the arms overhead. Anatomically, this pose causes expansion of the arm muscles. Also, the expansion of the lungs and chest wall due to the breath hold cause descent of the diaphragm and abdominal organs. Most patients cannot endure an arms-up posture for the duration of a PET scan, which can last up to 30 min, and will, of course, be engaged in normal tidal breathing. As a result, PET scans show an average of the anatomy over the respiratory cycle. Clearly, a linear transformation model will not be sufficient to match anatomic regions under these significantly different conditions. We present a nonparametric deformation model capable

Manuscript received June 11, 2001; revised July 16, 2002. This work was supported by the National Institutes of Health (NIH) under Grant CA42045 and Grant CA80907. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was D. Townsend. *Asterisk indicates corresponding author.*

\*D. Mattes is with The Boeing Company, Phantom Works, M&CT, Advanced Systems Laboratory, PO Box 3707, MC 7L-40, Seattle, WA 98124-2207 USA (e-mail: david.mattes@boeing.com).

D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank are with the University of Washington; Imaging Research Laboratory University of Washington Medical Center, Box 356004, Seattle, WA 98195 USA.

Digital Object Identifier 10.1109/TMI.2003.809072

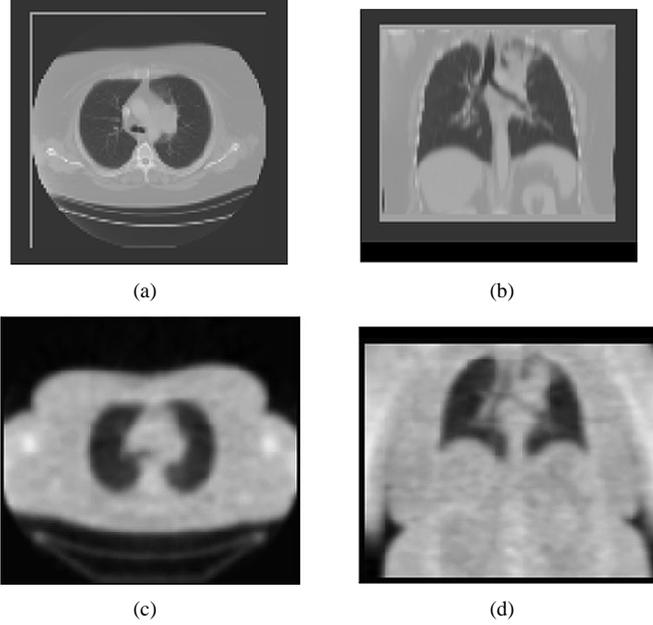


Fig. 1. Sample CT and TR images illustrating the necessity for nonrigid registration. (a) Axial and (b) coronal views of a typical CT dataset. (c) Axial and (d) coronal views of a TR scan from the same patient.

of recovering nonlinear deformations due to anatomic motion that occurs in CT images at maximum inspiration and PET TR images that are an average image of tidal breathing. Our work does not include a model of the respiratory process itself, or the resulting blurring that occurs in the TR images. Some sample images of the relevant anatomy are shown in Fig. 1.

## II. ALGORITHM DESIGN AND IMPLEMENTATION

Let  $f_T(\mathbf{x})$  be a test image we want to register to a reference image  $f_R(\mathbf{x})$ . We assume that  $f_T$  and  $f_R$  are defined on the continuous domains  $V_T$  and  $V_R$ , respectively, according to the image model presented in the Section II-A. The transformation  $\mathbf{g}(\mathbf{x}|\mu)$  describes the deformation from  $V_T$  to  $V_R$ , where  $\mu$  is a set of transformation parameters to be determined. We pose the task of medical image registration as an optimization problem. To align the reference image  $f_R(\mathbf{x})$  with the transformed test image  $f_T(\mathbf{g}(\mathbf{x}|\mu))$ , we seek the set of transformation parameters  $\mu$  that minimizes an image discrepancy function  $S$

$$\hat{\mu} = \arg \min_{\mu} S(f_R, f_T \circ \mathbf{g}(\bullet|\mu)). \quad (1)$$

Values for the transformation parameters  $\mu$  are iteratively chosen to reduce the discrepancy. In our implementation, the negative of mutual information is used to measure image discrepancy. We hypothesize that the set of transformation parameters  $\hat{\mu}$  that minimizes the discrepancy function also brings the transformed test image into best registration (as defined above) with the reference image.

Mutual information is an entropy-based measure of image alignment derived from probabilistic measures of image intensity values [1], [2], [21]. Because a large number of image samples are used to estimate image statistics, the effects of image noise on the metric are attenuated. Mutual information is also robust under varying amounts of image overlap as the test image moves with respect to the reference, although

normalized mutual information may be more robust [22]. The deformation model we propose is parameterized by a large number of coefficients. Under this condition the optimization of mutual information is facilitated by a formulation that is explicitly differentiable. The requirement of differentiability in the cost function in turn means that both the deformations and the similarity criterion must be differentiable. This is accomplished by using a B-Spline basis to represent the test image and model deformations and by estimating the joint probability distribution between the test and reference images with a Parzen window. We draw on the work of Thevenaz and Unser [3] for the mathematical development. We now describe the nonlinear transformation model which is used to incorporate deformations into the geometric manipulation of the test image.

### A. Image Representation

We assume an image  $f(\mathbf{x})$  is described by a set of samples  $f_i = f(\mathbf{x}_i)$ ,  $\mathbf{x}_i \in V$  defined on a Cartesian grid with integer spacing. The calculation of  $f(\mathbf{x})$  at points not on the grid requires an interpolation method based on the samples  $f_i$  and their locations  $\mathbf{x}_i$ . We utilize an interpolation scheme that represents the underlying continuous image  $f(\mathbf{x})$  by a B-Spline basis. The expansion coefficients  $c_i$  of the basis are computed from the image samples  $f_i$  through an efficient recursive filtering algorithm [5]. Values of  $f(\mathbf{x})$  that do not lie on the lattice can then be interpolated [3]

$$f(\mathbf{x}) = \sum_i c_i \beta^{(3)}(\mathbf{x} - \mathbf{x}_i) \quad (2)$$

where  $\mathbf{x} = [x, y, z]^T$  is any real-valued voxel location in the volume,  $\mathbf{x}_i = [x_i, y_i, z_i]^T$  is the coordinate vector of a lattice point, and  $\beta^{(3)}(\mathbf{x}) = \beta^{(3)}(x)\beta^{(3)}(y)\beta^{(3)}(z)$  is a separable convolution kernel. The argument of the spline window is the sampled cubic B-Spline

$$\beta^{(3)}(x) = \begin{cases} \frac{1}{6}(4 - 6x^2 + 3|x|^3), & 0 \leq |x| < 1 \\ \frac{1}{6}(2 - |x|)^3, & 1 \leq |x| < 2 \\ 0, & 2 \leq |x| \end{cases} \quad (3)$$

The gradient of the interpolated image at any location may be calculated in a similar manner, but a derivative operator is applied to the convolution. This is simply the derivative of the spline window in the respective dimension of each gradient component

$$\frac{\partial f(\mathbf{x})}{\partial x} = \sum_i c_i \left( \frac{d\beta^{(3)}(u)}{du} \Big|_{u=x-x_i} \beta^{(3)}(y-y_i)\beta^{(3)}(z-z_i) \right) \quad (4)$$

with similar formulas for  $\partial f/\partial y$  and  $\partial f/\partial z$ . The cubic spline window  $\beta^{(3)}$  can be differentiated explicitly, and after simplification reduces to the difference of two shifted second-order splines

$$\frac{d\beta^{(3)}(u)}{du} = \beta^{(2)}\left(u + \frac{1}{2}\right) - \beta^{(2)}\left(u - \frac{1}{2}\right). \quad (5)$$

### B. Image Statistics

Let  $L_R$  and  $L_T$  be specified numbers of uniformly sized bins along the respective dimensions of the joint histogram of the reference and test images. The histogram bins are indexed by integer values  $\kappa$ ,  $0 \leq \kappa < L_R$  and  $\iota$ ,  $0 \leq \iota < L_T$  and, in

general, correspond to linearly scaled ranges in intensity values. Scaling the data provides a method of binning data with arbitrary magnitude and dynamic range such that any contribution to the histogram will fall into a valid bin. We form a histogram by determining the appropriate bin for a set of paired image values and totalling the contributions in each of the bins. When we normalize the histogram it behaves like a traditional probability distribution. A Parzen window is used to generate continuous estimates of the underlying image distributions, thereby reducing the effects of quantization from interpolation and discretization from binning the data [3].

The calculation of mutual information requires estimates of the marginal and joint probability distributions of the intensity values of the reference and test images. Let  $\beta^{(3)}$  be a cubic spline Parzen window and  $\beta^{(0)}$  be a zero-order spline Parzen window (centered unit pulse), both of which satisfy the partition of unity constraint [3]. The smoothed joint histogram of  $(f_R, f_T \circ g)$  is given by

$$p(\iota, \kappa | \mu) = \alpha \sum_{\mathbf{x} \in V} \beta^{(0)} \left( \kappa - \frac{f_R(\mathbf{x}) - f_R^\circ}{\Delta b_R} \right) \times \beta^{(3)} \left( \iota - \frac{f_T(\mathbf{g}(\mathbf{x} | \mu)) - f_T^\circ}{\Delta b_T} \right) \quad (6)$$

where  $\alpha$  is a normalization factor that ensures  $\sum p(\iota, \kappa) = 1$ , and  $f_R(\mathbf{x})$  and  $f_T(\mathbf{g}(\mathbf{x} | \mu))$  are samples of the reference and interpolated test images, respectively. Each contributing image value is normalized by the minimum intensity value,  $f_R^\circ$  or  $f_T^\circ$ , and the intensity range of each bin,  $\Delta b_R$  or  $\Delta b_T$ , to fit into the specified number of bins ( $L_R$  or  $L_T$ ) in the intensity distribution. The summation range  $V$  is over a subset of voxel locations in  $V_R$  that contribute to the distribution, i.e.  $V$  is the intersection of transformed voxels chosen from  $V_T$  that map into  $V_R$ .

The marginal smoothed histogram for the test image is computed from the joint histogram

$$p_T(\iota | \mu) = \sum_{\kappa} p(\iota, \kappa | \mu). \quad (7)$$

The marginal probability for the reference image can be computed independently of the transformation parameters by noting that the B-Spline Parzen window satisfies the partition of unity constraint. The reference marginal smoothed histogram is computed as

$$p_R(\kappa) = \alpha \sum_{\mathbf{x} \in V} \beta^{(0)} \left( \kappa - \frac{f_R(\mathbf{x}) - f_R^\circ}{\Delta b_R} \right) \quad (8)$$

where  $\alpha$  is the constant in (6).

### C. Mutual Information

The image discrepancy measure we use is the negative of mutual information. The negative of mutual information  $S$  between the reference image and the transformed test image is expressed as a function of the transformation parameters  $\mu$  [3]

$$S(\mu) = - \sum_{\iota} \sum_{\kappa} p(\iota, \kappa | \mu) \log \frac{p(\iota, \kappa | \mu)}{p_T(\iota | \mu) p_R(\kappa)} \quad (9)$$

where  $p$ ,  $p_T$ , and  $p_R$  are the joint, marginal test, and marginal reference probability distributions, respectively.

Calculation of the gradient of the cost function is necessary for its efficient and robust minimization. For  $\mu$ —a set of  $n$ -independent parameters—the gradient of mutual information is given as

$$\nabla S = \left[ \frac{\partial S}{\partial \mu_1}, \frac{\partial S}{\partial \mu_2}, \dots, \frac{\partial S}{\partial \mu_i}, \dots, \frac{\partial S}{\partial \mu_n} \right]^T. \quad (10)$$

A single component of the gradient is found by differentiating (9) with respect to a transformation parameter (as given in [3]), and requires differentiation of the joint distribution in (6). After successive applications of the chain rule, the  $i^{\text{th}}$  partial derivative of the joint distribution is given as

$$\frac{\partial p(\iota, \kappa)}{\partial \mu_i} = \frac{1}{\Delta b_T(\#V)} \sum_{\mathbf{x} \in V} \beta^{(0)} \left( \kappa - \frac{f_R(\mathbf{x}) - f_R^\circ}{\Delta b_R} \right) \times \frac{\partial \beta^{(3)}(u)}{\partial u} \Big|_{u=\iota - \frac{f_T(\mathbf{g}(\mathbf{x} | \mu)) - f_T^\circ}{\Delta b_T}} \times \left( \frac{-\partial f_T(\mathbf{t})}{\partial \mathbf{t}} \Big|_{\mathbf{t}=\mathbf{g}(\mathbf{x} | \mu)} \right)^T \frac{\partial \mathbf{g}(\mathbf{x} | \mu)}{\partial \mu} \quad (11)$$

where  $\#V$  is the number of voxels used in the summation, the image gradient is calculated as in (4), and the various spline derivatives ( $\partial \beta^{(3)}/\partial u$ ,  $\partial f_T/\partial \mathbf{t}$ ) are computed using (5).

The final term to discuss from (11) is the expression for the partial derivatives of the transformation  $\partial \mathbf{g}(\mathbf{x} | \mu)/\partial \mu$ . This is the variation in the deformation at a point  $\mathbf{x}$  due to a variation in the transformation parameters, and depends on geometry and the transformation model. As will be described briefly in the Section II-D, the linearity of the expression of the transformation makes the differentiation straightforward.

### D. Deformations

An important aspect of our algorithm is the expression for the nonrigid transformation of image coordinates. We model deformations on cubic B-splines, because of their computational efficiency (separability in multiple dimensions, calculation via filtering), smoothness, and local control. The description of a deformation is similar to that of the images; however, the deformation is defined on a much coarser grid. A deformation is defined on a sparse, regular grid of control points  $\lambda_j$  placed over the test image and is then varied by defining the motion  $\mathbf{g}(\lambda_j)$  of each control point. Using a spline interpolation kernel to compute the deformation values between the control points produces a locally controlled, globally smooth transformation. The spline deformation model is differentiable with respect to both the deformation parameters (necessary for computing the gradient of mutual information) and spatial coordinates (useful in other applications, for example estimating inverse coordinate mappings).

The resolution  $\rho = [\rho_x, \rho_y, \rho_z]$  of the deformation determines the spacing of the grid and can be anisotropic. The set of control points is a regular grid with spacings

$$\Delta \rho = [\Delta \rho_x, \Delta \rho_y, \Delta \rho_z]^T = \left[ \frac{q_x - 1}{\rho_x - 1}, \frac{q_y - 1}{\rho_y - 1}, \frac{q_z - 1}{\rho_z - 1} \right]^T \quad (12)$$

where  $q_x$ ,  $q_y$ , and  $q_z$  are the dimensions of the test image. Given the grid spacing, we can compute the coordinates of the control points. The  $(x, y, z)$  coordinates of the control points are stored

in separate three-dimensional (3-D) matrices. Any  $j = (t, u, v)$  element of these matrices has a location in the test image given by  $[t \cdot \Delta\rho_x, u \cdot \Delta\rho_y, v \cdot \Delta\rho_z]^T$ . Each control point  $\lambda_j$  has an associated three-element deformation coefficient  $\delta_j$ , describing the x-, y-, and z-components of the deformation. For a give  $\rho$  there will be  $\|\rho\| = 3 \cdot \rho_x \cdot \rho_y \cdot \rho_z$  deformation coefficients.

The deformation at any point  $\mathbf{x} = [x, y, z]^T$  in the test image is interpolated using a cubic B-Spline convolution kernel

$$\mathbf{D}(\mathbf{x}|\delta) = \sum_j \delta_j \beta^{(3)}\left(\frac{\mathbf{x} - \lambda_j}{\Delta\rho}\right) \quad (13)$$

where  $\beta^{(3)}(\mathbf{x}) = \beta^{(3)}(x)\beta^{(3)}(y)\beta^{(3)}(z)$  is a separable cubic B-spline convolution kernel. Only those  $\delta_j$  corresponding to the 64 control points nearest  $\mathbf{x}$  contribute to this sum. When we differentiate (13) with respect to a deformation coefficient the parameter appears only once in the summation, resulting in a single nonzero term. The derivative with respect to a x component deformation coefficient is (the y and z components follow similarly)

$$\frac{\partial \mathbf{D}(\mathbf{x})}{\partial \delta_{j,x}} = \left[ \beta^{(3)}\left(\frac{x - \lambda_j}{\Delta\rho}\right), 0, 0 \right]^T. \quad (14)$$

The transformation of the test image is specified by mapping reference image coordinates according to a locally perturbed rigid body transformation. Given a  $3 \times 3$  homogeneous rotation matrix  $\mathbf{R}$ , a three-element transformation vector  $\mathbf{T}$ , and a deformation term  $\mathbf{D}(\mathbf{x}|\delta)$ , we can apply nonlinear transformations to the test image

$$\mathbf{g}(\mathbf{x}|\mu) = \mathbf{R}(\mathbf{x} - \mathbf{x}_C) - (\mathbf{T} - \mathbf{x}_C) + \mathbf{D}(\mathbf{x}|\delta) \quad (15)$$

where  $\mathbf{x}_C = [q_x/2, q_y/2, q_z/2]^T$  is the location of the center of the test volume and  $\mathbf{x} = [x, y, z]^T$  is any voxel location in the reference image. With this transformation model, the set of transformation parameters becomes

$$\mu = \{\gamma, \theta, \phi, t_x, t_y, t_z; \delta_j\} \quad (16)$$

where  $\{\gamma, \theta, \phi\}$  are the Roll-Pitch-Yaw (RPY) Euler angles of the rotation matrix  $\mathbf{R}$ ,  $\mathbf{T} = [t_x, t_y, t_z]^T$ , is the translation vector, and  $\delta_j$  is the set of deformation coefficients. In the application described here, the set of deformation coefficients contained up to  $\sim 2200$  members.

### E. Multiresolution Optimization Strategy

The registration process is automated by varying the deformation in the test image until the discrepancy between the two images is minimized. The alignment process is divided into two registrations by separating (15) into a term containing only rigid body parameters and a second with only deformation parameters. We use L-BFGS-B [6], a limited-memory, quasi-Newton minimization package, to reduce the cost function in (9) until termination criteria are satisfied. The limited-memory method is useful here because of the high dimensionality of the parameter space. Instead of estimating the entire Hessian during minimization, only a low-rank approximation is calculated, allowing linear or super-linear convergence rates. L-BFGS-B provides an additional advantage in that it allows bound constraints on the independent variables.

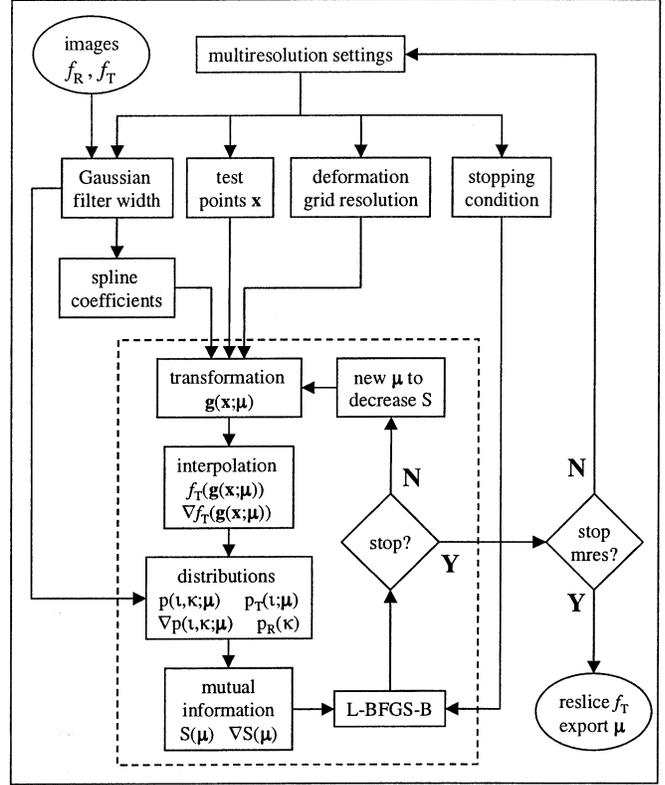


Fig. 2. The registration algorithm in a multiresolution framework. The area enclosed in a dashed line is the algorithmic core for computing and optimizing mutual information for a single resolution step.  $f_T$  and  $f_R$  are the test and reference images, respectively.  $\mathbf{x}$  is any geometric location in the reference image,  $\mu$  is the set of transformation parameters.  $p$ ,  $p_T$ ,  $p_R$ , and  $\nabla p$  are the joint, marginal test, marginal reference, and derivative of the joint distribution, respectively.  $S$  and  $\nabla S$  are the values of mutual information and its gradient, respectively, between the two images. L-BFGS-B is an optimization package that searches the parameter space of  $\mu$ .

In order to avoid local minima, and to decrease computation time, we use a hierarchical multiresolution optimization scheme. A flowchart of the algorithm in the multiresolution context is shown in Fig. 2. We initially optimize for a deformation to recover the gross motion of the patient and large anatomic structures. As we increase the resolution, we recover increasingly fine misalignments. To simplify the transition between multiresolution steps, the size of the image volume remains the same while other parameters vary. According to an empirically determined schedule, we maintain control over the resolution of the deformation, the number of image samples used to measure mutual information, the degree of Gaussian image blurring, and the optimizer's termination criterion. The resolution steps are denoted as  $L_1 \dots L_M$ , where  $M$  is the number of multiresolution steps (typically  $M = 4$ ). The optimum set of transformation parameters at a coarser resolution level is upsampled to become the starting point for the next resolution level.

If  $\rho = [\rho_x, \rho_y, \rho_z]$  specifies the resolution of the deformation, and  $\rho_1 \dots \rho_M$  are the deformation resolutions for each of the multiresolution steps, then we proceed from resolution  $m$  to resolution  $m + 1$  as follows. Starting at current deformation resolution  $\rho_m$  with deformation coefficients  $\delta_{j,m}$  and grid spacing  $\Delta\rho_m$ , we first place a deformation grid over the volume at higher resolution  $\rho_{m+1}$ . The deformation at resolution  $m$  is

TABLE I  
MULTIRESOLUTION PARAMETERS USED FOR REGISTERING THE 27 DATASETS

Parameter	$L_1$	$L_2$	$L_3$	$L_4$
Gaussian kernel $\sigma$ (pixels)	16	8	2	0
Isotropic deformation resolution (# variables $\ \mathbf{n}_{D,m}\ $ )	3 (87)	5 (381)	7 (1035)	9 (2193)
Percentage voxels used [ $\eta=10^{-4}$ ]	0.8	3.4	9.3	19.7
Termination criterion $\epsilon$	$10^{-5}$	$10^{-6}$	$10^{-7}$	$10^{-8}$

computed for each of the control point locations  $\lambda_{j,m+1}$  for resolution  $m+1$

$$\mathbf{D}(\lambda_{j,m+1}) = \sum_j \delta_{j,m} \beta^{(3)} \left( \frac{\lambda_{j,m+1} - \lambda_{j,m}}{\Delta \rho_m} \right). \quad (17)$$

Then the deformation coefficients are the spline coefficients of the new grid of deformation values

$$\delta_{j,m+1} = \Gamma(\mathbf{D}(\lambda_{j,m+1})) \quad (18)$$

where  $\Gamma$  is the recursive filter used to compute a spline basis of the test image [5].

In addition to varying the deformation resolution, Gaussian blurring is applied to the images with a kernel that narrows as multiresolution proceeds. Smoothing the images tends to smooth the mutual information cost function, thereby avoiding local maxima. Because the image (and the cost function) is smoothed we can use fewer samples to estimate the joint distributions, which is an alternative to constructing an image pyramid. We follow the general rule that the number of data points  $N_m$  at resolution level  $m$  should be proportional to the number of independent parameters for which we are trying to solve

$$N_m = \eta \|\rho_m\| (\#V) \quad (19)$$

where  $\#V$  is the number of voxels in the volume,  $\|\rho_m\| = 3 \cdot \rho_{x,m} \cdot \rho_{y,m} \cdot \rho_{z,m}$  is the number of deformation coefficients, and  $\eta$  is a proportionality constant. Finally, the optimizer's termination criterion is varied. The iteration terminates when the change in the objective function between consecutive iterations falls below a specified value  $\epsilon$ . At coarse resolutions, this criterion can be relaxed, however, it is set more strictly with increasing resolution.

The alignment process is divided into two registrations: a rigid body (6 parameters,  $\eta = 0.05$ ) followed by a deformation ( $\|\rho_m\|$  parameters,  $\eta = 10^{-4}$ ) recovery. Each registration proceeds from coarse to fine resolution in four steps. The rigid body parameters are recovered first and then fixed for the deformation recovery; the flowchart in Fig. 2 is used for both recoveries. The separation in recovery stages is needed because redundancy in the independent variables can cause competition during minimization. For the registrations presented in this paper, the multiresolution parameter settings for the deformation recovery are summarized in Table I. For rigid body recovery,  $\epsilon$  and  $\sigma$  follow the same schedules given in Table I, while the number of parameters and  $\eta$  (and, consequently,  $N_m$ ) are fixed at 6 and 0.05,

respectively. In all cases, the number of bins  $L_R$  and  $L_T$  used to compute the joint histogram are both fixed at 50, a number established empirically by ourselves and others [22]. Computation time is most strongly affected by the number of transformation parameters  $\|\rho_m\|$  and the number of voxels used to estimate mutual information  $N_m$  (and, consequently, the voxel sampling proportionality constant  $\eta$ ).

### III. VALIDATION

Validating the performance of an image registration algorithm with real images is not straightforward. The lack of a gold standard complicates matters and prevents any automated assessment of registration accuracy. Even if individuals trained to interpret medical images are involved in a validation experiment, providing a method for consistently assessing individual images is difficult. There is a tradeoff between the number of images that can be assessed and the time required to assess each one, a situation that often forces researchers to validate algorithms based on a limited sample size.

For algorithms with linear/global transformation models, retrospective validation can be performed if fiducial markers are in place at the time the scan is performed [9]. This method would not be appropriate here as the imaged anatomy is subject to non-rigid motion between examinations or due to physiological effects such as breathing. In cases where placing fiducial markers is impractical or impossible, researchers often conduct validation experiments using simulated data, generated with a physical or simulated [29] phantom. Such studies provide insight into the performance range of the algorithm but still mean little in terms of performance on real data. Landmark identification is another common method of validation, and was considered for our validation experiments. A difficulty with this technique is that there are few (if any) point landmarks in the anatomy. Structures like the carina, which shows the split in the trachea and is well-defined in the test and reference images, do not terminate at a single three-dimensional (3-D) coordinate, but are distributed over a surface. There will always be some degree of error in the selection of corresponding points.

We would like to avoid the pitfalls in landmark identification, but the capacity of the human eye to rapidly and accurately (albeit only qualitatively) determine the quality of a registration should not be underestimated. While validation methods involving human assessment will be prone to error, bias, and inconsistencies, if two images are presented in a user-friendly

manner, experienced physicians can rapidly assess the “gestalt” quality of the registration in a short amount of time. Furthermore, they have *a priori* knowledge of where in the anatomy the quality of the registration should be high and where these requirements can be relaxed. Our aim with the experiment was to assess the algorithm only in anatomically relevant areas. We took seven images from each patient set: five axial slices (four regularly spaced through the lungs, and one in the upper abdomen), one coronal slice at the carina, and one sagittal midline slice. Additionally, we included 10% duplicate slices to measure intraobserver consistency.

We created a user interface for validation that allows rapid visual assessment of registered images. To provide a measure of consistency between multiple observers, we allowed only two-dimensional (2-D) image navigation, and presented only specific images for assessment. The interface is based on a split window, a display technique that fixes one image over another and allows the user to vary the lines of transition from the top image to the image below. The images are stationary, but with the mouse button held the vertical and horizontal lines of transition move with the mouse cursor. To quantify the perceived error, a ruler is placed over the image and fixed to the mouse cursor. A set of error bars physically sized according to the ruler divisions is used to assess the registration accuracy for the given image pair. The observer makes two error assessments for each image pair: overall, or “representative,” error and maximum error. Using this method, the two strongest visual tests will be the presence/absence of similar anatomic structures and the magnitude of discontinuities in tissue boundaries. The validation technique will measure errors in translation (both locally and globally) but will not assess rotation errors. Since both images contain similar anatomic information, we found the split window display technique preferable to a pseudocolor image overlay for validation.

Since the intensity values for CT and TR scans are not in the 8-bit display range, window and level functionality is necessary for rendering specific anatomic structure. Image interpretation is sensitive to the display parameters, in that if the anatomy is not rendered at similar window and level settings, feature boundaries can significantly vary. Therefore, some form of display normalization is required to provide consistency in the images presented in the validation experiment. Window and level settings for CT images are standardized in Hounsfield units so we estimated the intensity transformation for attenuation values to Hounsfield units using the joint histogram from a single registered image set. Averaging along the CT Hounsfield (reference) and PET attenuation (test) dimensions produces several datapoints that are fit with the third-order polynomial shown in Fig. 3. As shown in Tai *et al.* the intensity transformation from CT to TR values is nonlinear due to scattering, and they generate a figure similar to Fig. 3 [8].

During validation the intensity values of the TR slices are transformed using a lookup table generated from the polynomial fit. We should note that the intensity transformation applied to the TR slices discussed here is used for validation only and not during the registration process. Standard window and level settings are applied to both the TR and CT slices for improved consistency during image interpretation. The physicians were

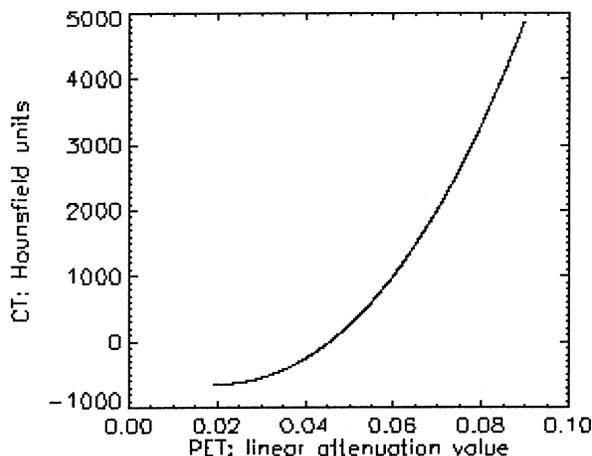


Fig. 3. A third-order polynomial is used to map PET attenuation values to CT Hounsfield units.

requested to use only the provided presets when measuring registration accuracy.

#### IV. DATA

The image sets used for the validation study consist of FDG-PET and CT scans of the chest in 28 patients. Each set contains an FDG-PET scan and a chest CT scan performed on a patient undergoing initial staging for nonsmall cell lung cancer. These scans were all performed at the University of Washington Medical Center. The chest CT scan was performed with spiral technique and bolus intravenous contrast administration, using a variety of GE CT scanners (and with varying collimation and pitch) with submillimeter intrinsic spatial resolution. The imaging protocol specified an arms-up position and breath-hold at maximum inspiration. PET imaging was performed on a GE Advance Scanner with measured isotropic resolution of approximately 5 mm in air (which decreases due to scattering) [31]. For the FDG PET scan, emission imaging was performed from the neck to the pelvis. Transmission imaging (15 min with a 511-keV energy beam) of the thorax and upper abdomen was also performed for the purpose of attenuation correction. Therefore, emission images spanning from the neck to pelvis, transmission and attenuation-corrected images of the thorax and upper abdomen were available for testing our image fusion approach.

All patients received both scans within at most two months of each other, and did not have surgical resection or major tumor growth between acquisitions. The TR scans are exported from the PET station as a single 3-D floating-point data set of size  $128 \times 128 \times 103$ . The TR voxels have a physical size of  $4.29 \times 4.29 \times 4.25 \text{ mm}^3$ ; therefore, each TR scan axially images 437.75 mm in its field-of-view (FOV). The CT scans are acquired axially with slice thicknesses that vary from 2–7 mm. Once the patient scan is acquired, the in-plane voxel size is set to magnify the patient anatomy as much as possible, within the range of 0.7–1.1 mm. The CT FOV varies across patients and is generally smaller than that of the TR scan (a typical CT FOV covers 350 mm of axial anatomy). For the data presented here, the PET TR scan is the test image  $f_T$  and the CT scan is the reference

$f_R$ . We should note that theoretically either scan can be the test or reference, but we have consistently used the CT scan as the reference image. In the one case we did juxtapose the relationship, a poor registration resulted.

Several preprocessing steps are performed before registering the images. The images are filtered with a Gaussian blur ( $\sigma = 1$  voxel), the TR image is resliced to have isotropic voxels, and using cubic B-Spline interpolation the spatial resolution of the CT is reduced to match that of the TR. All the images presented in the validation experiment have an isotropic voxel size of 4.29 mm for the registration. As explained in the introduction the arms are present in the PET images and are extraneous features for which there is no correspondence in the CT. With the aid of a simple graphical user interface, manually placed polygons in several axial slices are extended through the PET volume to exclude the arms. The voxels within this volume [V in (11)] represent the set of image samples from which mutual information is computed. No other cropping or masking is performed on the images. A final postprocessing step zooms the selected slices of the TR and CT volumes for validation, producing 2-D images of size  $384 \times 384$  with  $1.43 \times 1.43 \text{ mm}^2$  pixel dimensions. In some cases, the rigid body registration required a manual initialization along the z axis.

## V. RESULTS

Twenty-eight patient sets were registered for the validation experiment, with one failing due to a limited CT FOV. All data presented is for the 27 patients successfully registered, with no gross misregistrations being the only criteria for a successful registration. Alignment was assessed by two radiologists with experience interpreting both PET and CT images. Registration was performed on a DEC Alpha RS6000 (dual 375-MHz processors, 1.2-GB RAM) workstation running Compaq Tru64 Unix. The algorithm is implemented as a library of C modules and compiled on both the Windows 98 and Unix platforms using Microsoft Visual C++ 6.0 (Microsoft Corp., Redmond, WA) and gcc (Gnu C compiler), respectively. L-BFGS-B is a library of routines written in Fortran 77 [20]. The Interactive Data Language (IDL, Research Systems Inc., Boulder, CO) was used to manage program execution by allocating memory, loading and calling library modules, and interfacing with L-BFGS-B. IDL is also used for the graphical user interfaces for manual segmentation, dual image viewing, and validation.

Using the DEC Alpha workstation, an average of  $\sim 100$  min ( $\sim 10$  min for rigid body,  $\sim 90$  min for deformable) was required to register each dataset. Example images from a registered dataset are shown in Fig. 4 with a locked cursor, while Fig. 5 shows sample images from the validation user interface. The validation experiment was conducted on a Macintosh G4 400-MHz desktop computer. The 205 slices used in the experiment were divided into four sets, each of which required  $\sim 1.5$  h for an observer to assess. The observers scored the accuracy of the registration by selecting the appropriate error range (Table II) based on the on-screen error bars, and the observers' selections were recorded by the validation software. At the time the validation experiments were conducted, we did not feel it

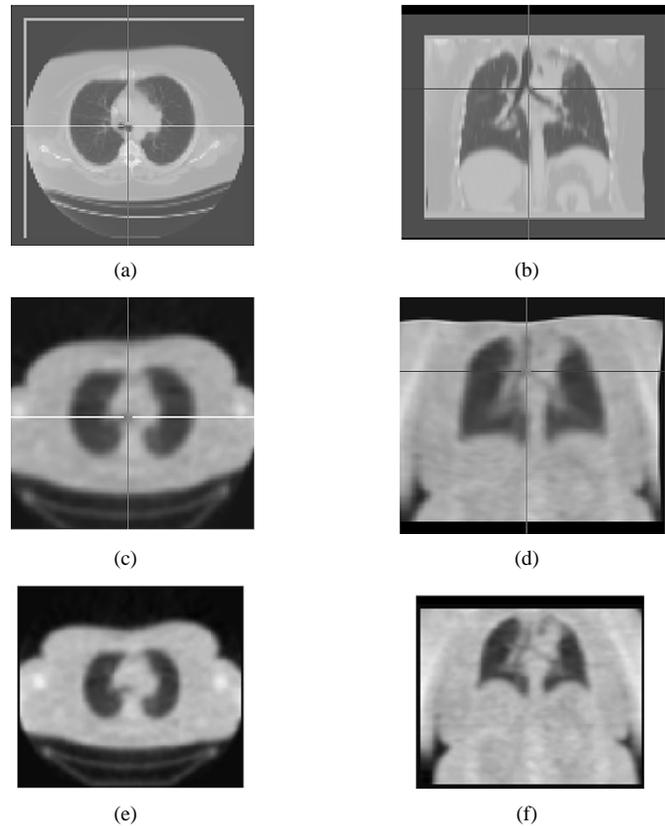


Fig. 4. Sample images from a registered TR-CT pair. (a) Axial and (b) coronal views of the CT. (c) Axial and (d) coronal views of the transformed TR image. The crosshairs are at the same image coordinates in both CT and TR volumes. Approximately corresponding (e) axial and (f) coronal (f) slices from the unregistered TR volume are shown for comparison.

was possible for the observers to determine an exact error measure, especially for the maximum assessed error. In future work, we will experiment with a method of estimating the error on a continuous domain.

The means, medians, and standard deviations of the overall and maximum assessed errors for both observers are given in Table III. Although the median error index is zero for the “overall” assessments, remember that the error index of zero corresponds to a mm-equivalent as given in Table II. We can also group the results by anatomic region and compare errors in all patients for a given region and for all the anatomy. The assessed error by anatomy is shown in Fig. 6. The errors in axial slices in the upper lung regions are small, while the poorest performance is in the abdomen. As shown in Fig. 6 the sagittal and coronal slices were also assessed as being less well registered than the lung regions shown in axial slices. There is some tendency for observer 2 to grade more harshly than observer 1 in the “overall” category, but their choices agree fairly well ( $\kappa = 0.61$ ). Perhaps because the maximum error can be assessed more objectively, the observers tend to agree more in this error category ( $\kappa = 0.68$ ). Intraobserver consistency was measured from the assessments of 21 duplicate slice presentations. Observer 1 ( $\kappa = 0.94$  for “overall,”  $\kappa = 0.81$  for “maximum”) tended to be more consistent than observer 2 ( $\kappa = 0.60$  for “overall,”  $\kappa = 0.69$  for “maximum”). Given the nature of the validation experiment, only translation errors

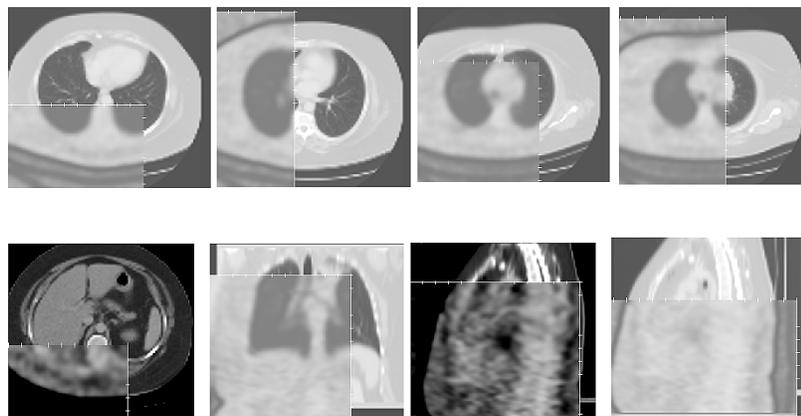


Fig. 5. Sample images from the seven anatomic locations considered in the validation experiment, as presented to the observers with the split window viewer. Images are rendered using lung window/level unless noted otherwise. (top row from left to right) Lower lung, mid-lower lung, mid-upper lung, and upper lung. (bottom row from left to right) Abdomen (mediastinal window), carina, midline (mediastinal window), and midline (repeat in lung window).

TABLE II  
ERROR INDEX RECORDED BY THE VALIDATION USER INTERFACE CORRESPOND TO ERROR RANGES IN PIXELS AND MILLIMETERS

recorded error index	pixels	mm
0	0-4	0-5.7
1	4-8	5.7-11.4
2	8-12	11.4-17.2
3	12-16	17.2-22.9
4	16-20	22.9-28.6

are measurable and we can make no claims about the rotational accuracy of the results.

## VI. DISCUSSION AND CONCLUSION

The results of the validation study indicate that the algorithm is capable of accurate registrations in the thorax, and the radiologists who validated the results felt the errors were generally within clinically acceptable ranges. We have identified regions in the anatomy for which the algorithm succeeds in varying degrees. According to visual assessments of 27 patient datasets by two observers, the overall error in the considered anatomy has an error index of 0.54, which is in the 0- to 6-mm error range (recall that the original PET voxel size is  $\sim 4.3$  mm on a side). The mid to upper lung regions are registered the most accurately with a mean overall assessed error index of 0.24, which is also in the 0-6 mm error range. The assessed performance in the abdomen is the worst in general with a mean overall assessed error index of 1.37, which is in the 6- to 11-mm error range.

The algorithm produces registrations with relatively large assessed error in the coronal and sagittal planes. This is particularly true at the lung-diaphragm interface since more surface area of the diaphragm is imaged on coronal or sagittal slices compared to the axial plane, and the diaphragm is not expected to be in the same location due to the differences in image acquisition (breath-hold vs non breath-hold). Additionally, larger

TABLE III  
GLOBAL STATISTICS FOR VALIDATION EXPERIMENT

Statistic	Error index for “overall” assessment observer 1 / observer 2	Error index for “maximum” assessment observer 1 / observer 2
mean	0.36 / 0.74	1.16 / 1.26
median	0 / 0	1 / 1
standard deviation	0.54 / 1.12	1.10 / 1.39

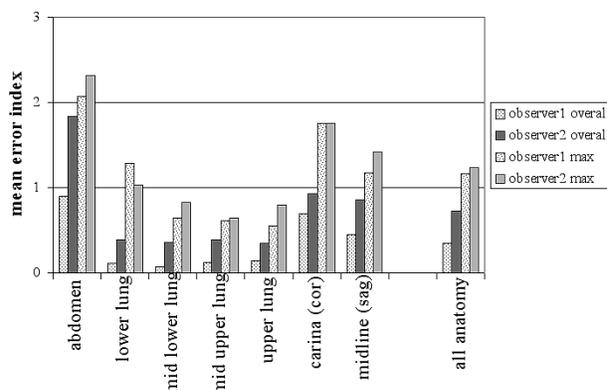


Fig. 6. Mean overall and maximum assessed errors by anatomy for both observers for the 27 patient datasets. The average performance for all the images is given in the last data series.

lung and soft tissue regions are perhaps driving the deformation, and the smaller airway anatomy is moved along with them. Increasing the deformation resolution in these dimensions may yield better results. Using a mathematical phantom such as the MCAT phantom would allow us to characterize the algorithm when a respiratory model is included in the image data [29]. The registration accuracy in the smaller anatomic regions (such as along the trachea) can likely benefit from some form of clustering of deformation control points. This highlights the principal limitation of the algorithm: the inability to use a nonuniform grid of sampling points in the test image. Future work will address this limitation.

From the validation study, we notice a poor assessment in the abdomen. At 511 keV, the TR scan shows little intensity differences between fat, muscle, and soft tissue, resulting in a uniform image in the abdomen (except for gas bubbles). At 140 keV, the CT scan of the abdomen, on the other hand, shows clear delineation of the anatomy. Since there are few shared structures in the TR and CT scan of the abdomen on which the algorithm can anchor, a poor registration results. In this case, the PET emission or SUV scan may be a better candidate for alignment with CT. An alternative approach includes using attenuation and emission values simultaneously, which may improve the overall accuracy in the torso.

In addition to the question of how fine a grid do we need to represent a given set of deformations, we must also decide how fine a deformation grid we can effectively use. The coarseness of the grid effectively regularizes the deformation and prevents violent local deformations. This may impose an upper limit on grid fineness, unless additional regularization terms are included in the cost function.

We have validated the algorithm only for the TR to CT matching problem. It is natural to question if it will work well for other inter- or intra-modality image pairs. PET images are among the worst (after single photon emission computed tomography) of the 3-D imaging modalities in terms of tissue delineation and signal to noise ratio. We believe that given the success of the TR-CT registration problem, the other modality combinations will be no more difficult or significantly easier. Clinical studies are currently underway for assessing the utility of the registration method described here for improved PET image interpretation, assessment of response to therapy, studying tumor biology, and image-guided therapy applications.

#### ACKNOWLEDGMENT

The authors wish to thank S. Kohlmyer for his comments throughout this research project, and B. Lewellen and S. Santoni for their assistance with acquiring the images.

#### REFERENCES

- [1] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recogn.*, vol. 32, pp. 71–86, 1999.
- [2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.
- [3] P. Thevenaz and M. Unser, "Spline pyramids for inter-modal image registration using mutual information," in *Proc. SPIE*, vol. 3169, 1997, pp. 236–247.
- [4] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, pp. 712–722, Aug. 1999.
- [5] M. Unser, A. Aldroubi, and M. Eden, "Fast B-spline transforms for continuous image representation and interpolation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 277–285, Mar. 1991.
- [6] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 2000, ch. 8–9.
- [7] J. Kybic, P. Thevenaz, A. Nirkko, and M. Unser, "Unwarping of unidirectionally distorted EPI images," *IEEE Trans. Med. Imag.*, vol. 19, pp. 80–93, Feb. 2000.
- [8] Y. Tai, K. P. Lin, C. K. Hoh, S. C. H. Huang, and E. J. Hoffman, "Utilization of 3-D elastic transformation in the registration of chest X-ray CT and whole body PET," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1606–1612, Aug. 1997.
- [9] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer Jr., R. M. Kessler, and R. J. Maciunas, "Comparison and evaluation of retrospective intermodality image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, pp. 554–566, July/Aug. 1997.
- [10] J. M. Fitzpatrick, D. L. G. Hill, Y. Shyr, J. B. West, C. Studholme, and C. R. Maurer Jr., "Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain," *IEEE Trans. Med. Imag.*, vol. 17, pp. 571–585, Aug. 1998.
- [11] J. B. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer Jr., R. M. Kessler, and R. J. Maciunas, "Retrospective intermodality registration techniques: surface-based versus volume-based," *IEEE Trans. Med. Imag.*, vol. 18, pp. 144–150, Feb. 1999.
- [12] R. Szeliski and S. Lavalée, "Matching 3-D anatomical surfaces with nonrigid deformations using octree-splines," *Int. J. Comput. Vis.*, vol. 18, no. 2, pp. 171–186, 1996.
- [13] H. Lester and S. R. Arridge, "A survey of hierarchical nonlinear medical image registration," *Pattern Recogn.*, vol. 32, pp. 129–149, 1999.
- [14] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: part I—theory," *IEEE Trans. Signal Processing*, vol. 41, pp. 821–833, Feb. 1993.
- [15] ———, "B-spline signal processing: part II—efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, pp. 834–848, Feb. 1993.
- [16] P. J. Kostelec, J. B. Weaver, and D. M. Healy Jr., "Multiresolution elastic image registration," *Med. Phys.*, vol. 25, no. 9, pp. 1593–1604, Sept. 1998.
- [17] B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard, "A level-set based approach to image registration," in *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis*, June 2000, pp. 86–93.
- [18] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, pp. 809–814, Aug. 1999.
- [19] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey, and D. J. Hawkes, "Voxel similarity measures for 3-D serial MR brain image registration," *IEEE Trans. Med. Imag.*, vol. 19, pp. 94–102, Feb. 2000.
- [20] "Tech. Rep.," EECs Dept., Northwestern Univ., Evanston, IL, 1994.
- [21] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.
- [22] C. Studholme, "Measures of 3D image alignment," Ph.D. dissertation, Univ. London, London, U.K., Aug. 1997.
- [23] P. Thevenaz and M. Unser, "A pyramid approach to sub-pixel image fusion based on mutual information," in *Proc. 1996 IEEE Int. Conf. Image Processing (ICIP'96)*, vol. I, Lausanne, Switzerland, Sept. 16–19, 1996, pp. 265–268.
- [24] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.
- [25] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "Multimodal image registration by maximization of the correlation ratio," Paris, France, INRIA Res. Rep. 3378, 1998.
- [26] G. Golub and C. van Loan, *Matrix Computation*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.
- [27] R. P. Woods, S. T. Grafton, J. D. G. Watson, N. L. Sicotte, and J. C. Mazziotta, "Automated image registration: II. Intersubject validation of linear and nonlinear methods," *J. Comput. Assist. Tomogr.*, vol. 22, no. 1, pp. 153–165, 1998.
- [28] M. van Herk, J. C. de Munck, J. V. Lebesque, S. Muller, C. Rasch, and A. Touw, "Automatic registration of pelvic computed tomography data and magnetic resonance scans including a full circle method for quantitative accuracy evaluation," *Med. Phys.*, vol. 25, no. 10, pp. 2054–2066, 1998.
- [29] W. P. Segars, D. S. Lalush, and B. W. Tsui, "Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms," *IEEE Trans. Nucl. Sci.*, vol. 48, pp. 89–97, Jan. 2001.
- [30] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Processing*, vol. 9, Dec. 2000.
- [31] B. Bendriem and D. W. Townsend, *The theory and practice of 3D PET*. Dordrecht, The Netherlands: Kluwer Academic, 1998.