# Nonrigid multimodality image registration

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

University of Washington; Imaging Research Laboratory
University of Washington Medical Center, Box 356004, Seattle, WA 98195

## ABSTRACT

We have designed, implemented, and validated an algorithm capable of 3D PET-CT registration in the chest, using mutual information as a similarity criterion. Inherent differences in the imaging protocols produce significant non-linear motion between the two acquisitions. To recover this motion, local deformations modeled with cubic B-splines are incorporated into the transformation. 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, the deformation model allows for closed-form expressions of the criterion and its gradient. A limited-memory quasi-Newton optimization package is used in a hierarchical multiresolution framework to automatically align the images. To characterize the performance of the algorithm, 27 scans from patients involved in routine lung cancer screening were used in a validation study. The registrations were assessed visually by two observers in specific anatomic locations using a split window validation technique. The visually reported errors are in the 0-6mm range and the average computation time is 100 minutes.

**Keywords**: registration, nonrigid, deformation, nonlinear, multimodality, validation, multiresolution, mutual information, positron emission tomography (PET), computed tomography (CT).

## 1. INTRODUCTION

We define the problem of medical image registration as follows. Given two image sets acquired from the same patient but at different times or with different devices, we seek the geometric transformation $g$ between the two respective image-based coordinate systems that maps a point $x$ in the first image set to the point $g(x)$ 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. CT and PET) are to be registered. The idea can still be used that, if a candidate registration $g$ matches features in the first set to similar features in the second set, it is probably correct. 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 beauty of the information theoretic measures is that they make no assumptions about the actual intensity values in the images, instead they measure statistical relationships between the two images. The mutual information metric has been effective in various applications where the requisite transformations are linear [2], [3] and more recently in cases involving non-linear motion descriptions [4].

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 (it measures uptake of radioactive tracers in cells with high metabolic activity), 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 in-plane resolution. If we can register the TR and CT images, we can apply the resulting transformation to the emission or standard uptake value (SUV) image for improved PET image interpretation.

Patient and anatomic motion during acquisition blurs the images. Sharp anatomic outlines in CT scans are obtained by requesting the patient to maintain maximum inspiration during the 30 seconds required for acquisition. To avoid attenuation of the arms in the X-ray beam, the patient also holds the arms overhead if possible. Anatomically, this arm motion 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 minutes, and will 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 is not sufficient to match anatomic regions under these wildly varying conditions.

Some example images of the relevant anatomy are shown in figure 2. Why is registration useful? Generally speaking there are many answers to this question. In the context of our efforts, we have identified several areas in which our methods are directly applicable to a clinical or research setting and have divided them into four groups: improved PET image interpretation, response to therapy, tumor biology, and image-guided therapy.

## 2. METHODOLOGY

Registration is the process of aligning two datasets (a test and a reference image) according to a given metric. We use mutual information as our image similarity function. Mutual information is a statistically-based measure of image alignment derived from probabilistic measures of image intensity values and is robust in the presence of noise and varying amounts of image overlap. Our formulation requires that we express mutual information as a continuous function to allow its explicit differentiation. Doing so requires that the components used in computing mutual information also be expressed as continuous functions. This is accomplished in three ways: using a B-spline basis to represent the test image, estimating the joint probability distribution between the test and reference images with a Parzen window, and implementing deformations modeled with cubic B-splines. We draw on the work of Thevenaz and Unser [3] for the mathematical development of the first two points. We introduce our non-linear transformation model which is used to incorporate deformations into the geometric manipulation of the test image.

### 2.1 Deformations

One important aspect of our algorithm is the expression for the transformation of image coordinates. We model deformations on cubic B-splines, because of their computational efficiency (via separability in mutlidimensional expressions), smoothness, and local control. A deformation is defined by placing a sparse, regular grid of control points over the volume and varied by moving the control points. Cubic splines are used to distribute this coarse grid of deformation coefficients over the entire image. Using a spline convolution kernel to interpolate the deformation values between the control points produces a locally-controlled, globally-smooth deformation. The region of support of the 3D cubic spline convolution window means each pixel in the test image is deformed by the cube of 64 nearest deformation parameters. Using splines for a deformation model also allows simple computation of deformation derivatives with respect to both a deformation parameter (for computing the gradient of mutual information) and a spatial coordinate (useful for estimating inverse coordinates).

The resolution of the deformation is defined by the spacing of the grid, which can vary for each image dimension. The vector of gridpoint densities is a compact notation for the deformation resolution:

$$\mathbf{n}_D = [n_x, n_y, n_z]^T . \tag{1}$$

Once $\mathbf{n}_D$ is specified, the control points are defined by their coordinates in the volume; there will be $n_x \cdot n_y \cdot n_z$ such control points, each with a 3-tuple of $(x,y,z)$ coordinates and a constant spacing between them. If the image dimensions are $\mathbf{q} = (q_x, q_y, q_z)$, then the vector of gridspacings is:

$$\Delta \mathbf{x}_D = [\Delta x_D, \Delta y_D, \Delta z_D]^T = \left[ \frac{q_x - 1}{n_x - 1}, \frac{q_y - 1}{n_y - 1}, \frac{q_z - 1}{n_z - 1} \right]^T . \tag{2}$$

Given the gridspacing, we can compute the coordinates of the control points. The control points are represented by the set of control point matrices, $\{\mathbf{x}_D\} = \{x_D, y_D, z_D\}$. Each of these components is a 3D matrix of x-, y-, or z-coordinates, respectively; i.e. each control point is specified by a $(x,y,z)$ coordinate in the volume. Any $(i,j,k)$ element of these matrices is computed as:

$$x_D(i,j,k) = i \cdot \Delta x_D$$
$$y_D(i,j,k) = j \cdot \Delta y_D \tag{3}$$
$$z_D(i,j,k) = k \cdot \Delta z_D$$

Each control point $\mathbf{x}_D(\mathbf{i}) = (x_D(\mathbf{i}), y_D(\mathbf{i}), z_D(\mathbf{i}))$ has an associated deformation term $\mathbf{c}_D(\mathbf{i}) = (c_{D,x}(\mathbf{i}), c_{D,y}(\mathbf{i}), c_{D,z}(\mathbf{i}))$ that is composed of 3 separate parameters.

We express the general form of a transformation as a locally-perturbed rigid body transformation. Given a 3x3 homogeneous rotation matrix $\mathbf{R}$, a 3-element transformation vector $\mathbf{T}$, and a deformation term $\mathbf{D}(\mathbf{c}_D;\mathbf{x})$, we can apply non-linear transformations to the test image:

$$\mathbf{g}(\mathbf{x};\mu) = \mathbf{R}(\mathbf{x} - \mathbf{x}_C) - (\mathbf{T} - \mathbf{x}_C) + \mathbf{D}(\mathbf{c}_D;\mathbf{x}), \tag{4}$$

where $\mathbf{x}_C = \left(\frac{nx}{2}, \frac{ny}{2}, \frac{nz}{2}\right)^T$ is the location of the center of the volume. With this transformation model, the set of transformation parameters becomes:

$$\mu = \left\{\gamma, \theta, \phi, t_x, t_y, t_z; \{\mathbf{c}_D\}\right\}, \tag{5}$$

where $(\gamma, \theta, \phi)$ are the Roll-Pitch-Yaw (RPY) Euler angles of the rotation matrix, $(t_x, t_y, t_z)$ are the translation components, and $\{\mathbf{c}_D\}$ is the set of deformation coefficients. In practice the set of deformation coefficients contains $\sim 10^2$-$10^3$ members.

The deformation is defined on a uniform grid by the grid control points $\{\mathbf{x}_D\}$ the spacing between the control points $\Delta\mathbf{x}_D$ and the set of deformation coefficients $\{\mathbf{c}_D\}$ and is locally distributed to neighboring voxels with cubic B-splines:

$$\mathbf{D}(\mathbf{c}_D;\mathbf{x}) = \sum_{\mathbf{h}} \mathbf{c}_D(\mathbf{h})\beta^{(3)}\left(\frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta\mathbf{x}_D} - \mathbf{h}\right), \tag{6}$$

where $\mathbf{r} = \lfloor \mathbf{x}/(\Delta\mathbf{x}_D) \rfloor$ is the central grid point, and the index of summation is determined from $\mathbf{r}$.

## 2.2 Interpolation

The test image is resampled at locations determined by applying a transformation $\mathbf{g}(\mathbf{x};\mu)$ to the voxel coordinates $\mathbf{x} \in V$ of the region of image overlap according to the transformation parameters $\mu$. The interpolation scheme relies on a cubic B-spline basis representation of the test image. The spline coefficients are determined through an efficient recursive filtering algorithm [5]. Image interpolation is accomplished via a separable convolution of the spline coefficients $\{c\}$ with the sampled cubic B-spline window $\beta^{(3)}$ [3]:

$$f_T(\mathbf{g}(\mathbf{x};\mu)) = \sum_i c(\mathbf{x}_i)\beta^{(3)}\left(\mathbf{g}(\mathbf{x};\mu) - \mathbf{x}_i\right), \tag{7}$$

where $\mathbf{x} = [x, y, z]^T$ is any voxel location in the reference image and the summation range is over the support of the cubic spline window.

The gradient of the transformed test image at each of the voxel locations is resampled in a similar manner, but a derivative operator is applied to the convolution, which is just the derivative of the spline window in the respective dimension of each gradient component:

$$\frac{df_T(\mathbf{t})}{d\mathbf{t}} = \sum_i \mathbf{c}(\mathbf{x}_i) \begin{bmatrix} \left.\dfrac{\partial \beta^{(3)}(u)}{\partial u}\right|_{u = (\mathbf{t}-\mathbf{x}_i)_x} & \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_y) & \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_z) \\[2em] \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_x) & \left.\dfrac{\partial \beta^{(3)}(u)}{\partial u}\right|_{u = (\mathbf{t}-\mathbf{x}_i)_y} & \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_z) \\[2em] \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_x) & \beta^{(3)}((\mathbf{t}-\mathbf{x}_i)_y) & \left.\dfrac{\partial \beta^{(3)}(u)}{\partial u}\right|_{u = (\mathbf{t}-\mathbf{x}_i)_z} \end{bmatrix}. \tag{8}$$

## 2.3 Mutual information

We pose the task of medical image registration as a function optimization problem. We desire the set of transformation parameters $\mu$ that maximizes an image similarity function $S$:

$$\mu_{opt} = \arg\max_{\mu} S(\mu). \tag{9}$$

In our implementation, mutual information is used as for the image similarity function. We hypothesize that the set of transformation parameters $\{\mu_{opt}\}$ that maximizes the similarity function also brings the transformed test image into best correspondence with the reference image. Having determined this set of parameters does not ensure that images are actually geometrically aligned, only that these parameters maximize (perhaps only locally) the similarity criterion.

The problem as presented in equation 9 is stated as a maximization problem, but we will actually minimize the *negative* of the function $S$. Let $L_T$ and $L_R$ be discrete sets of intensities associated to the test and reference image, respectively. 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]:

$$S(\mu) = -\sum_{\iota \in L_T} \sum_{\kappa \in L_R} p(\iota, \kappa;\mu) \log \frac{p(\iota, \kappa;\mu)}{p_T(\iota;\mu) p_R(\kappa)}, \tag{10}$$

where $p$, $p_T$, and $p_R$ are the joint, marginal test, and marginal reference probability distributions, and will be derived shortly.

Due to the high dimensionality of the space of transformation parameters, the gradient of the criterion facilitates the search for its maximum. The gradient of mutual information is given as

$$\nabla S = [\frac{\partial S}{\partial \mu_1}, \frac{\partial S}{\partial \mu_2}, ..., \frac{\partial S}{\partial \mu_i}, ..., \frac{\partial S}{\partial \mu_n}]^T. \tag{11}$$

A single component of the gradient is found by differentiating equation 10 with respect to a transformation parameter:

$$\frac{\partial S}{\partial \mu_i} = -\sum_{\iota \in L_T} \sum_{\kappa \in L_R} \frac{\partial p(\iota, \kappa;\mu)}{\partial \mu_i} \log \frac{p(\iota, \kappa;\mu)}{p_T(\iota;\mu)}, \tag{12}$$

where $\partial p(\iota, \kappa;\mu)/\partial \mu_i$ is the $i^{th}$ partial derivative of the joint distribution.

The probability distributions used to compute mutual information are based on marginal and joint histograms of the reference and test images. Parzen windowing is used to form continuous estimates of the underlying image histograms, also reducing the effects of quantization from interpolation and discretization from binning the data [3]. The joint distribution is therefore an explicitly differentiable function. 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 joint discrete probability is given by:

$$p(\iota, \kappa; \boldsymbol{\mu}) = \alpha \sum_{\mathbf{x} \in V} \beta^{(0)}\left(\kappa - \frac{f_R(\mathbf{x}) - f_R'}{\Delta b_R}\right) \beta^{(3)}\left(\iota - \frac{f_T(\mathbf{g}(\mathbf{x};\boldsymbol{\mu})) - f_T'}{\Delta b_T}\right), \tag{13}$$

where $\alpha$ is a normalization factor that ensures $\sum p(\iota, \kappa) = 1$, $\iota \in L_T$ and $\kappa \in L_R$, and $f_R(\mathbf{x})$ and $f_T(\mathbf{g}(\mathbf{x};\boldsymbol{\mu}))$ are samples of the reference and interpolated test images, respectively. Each contribution is normalized by the minimum intensity value, $f_R'$ or $f_T'$, and the intensity range of each bin, $\Delta b_R$ or $\Delta b_T$, to fit into a specified number of bins in the intensity distribution. The summation range $V$ is the set of voxel pairs that contribute to the distribution.

The marginal discrete probability for the test image is computed from the joint distribution:

$$p_T(\iota; \boldsymbol{\mu}) = \sum_{\kappa \in L_R} p(\iota, \kappa; \boldsymbol{\mu}). \tag{14}$$

The marginal discrete 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 distribution is computed as:

$$p_R(\kappa) = \alpha \sum_{\mathbf{x} \in V} \beta^{(0)}\left(\kappa - \frac{f_R(\mathbf{x}) - f_R'}{\Delta b_R}\right). \tag{15}$$

The derivative of the joint distribution with respect to one of the transformation parameters is:

$$\frac{\partial p(\iota, \kappa)}{\partial \boldsymbol{\mu}} = \frac{1}{\Delta b_T (\#V)} \sum \beta^{(0)}\left(\kappa - \frac{f_R(\mathbf{x}) - f_R'}{\Delta b_R}\right) \left.\frac{\partial \beta^{(3)}(\xi)}{\partial \xi}\right|_{\xi = \iota - \frac{f_T(\mathbf{g}(\mathbf{x};\boldsymbol{\mu})) - f_T'}{\Delta b_T}} \left(\left.\frac{-df_T(\mathbf{t})}{d\mathbf{t}}\right|_{\mathbf{t} = \mathbf{g}(\mathbf{x};\boldsymbol{\mu})}\right)^{\mathrm{T}} \frac{\partial \mathbf{g}(\mathbf{x};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}, \tag{16}$$

where $\#V$ is the number of voxels used in the summation.

The final term to discuss from equation 16 is the expression for the partial derivatives of the transformation $\partial \mathbf{g}(\mathbf{x};\boldsymbol{\mu})/\partial \boldsymbol{\mu}$. This is the variation in position due to a variation in transformation parameter, and depends on geometry and the transformation model. The linearity of the expression of the transformation makes the differentiation of equation 4 straightforward, i.e. derivatives with respect to the rotation angles will depend only on the term $\mathbf{R}(\mathbf{x} - \mathbf{x}_C)$, while the derivatives with respect to the transformation parameters are even simpler. The partial derivatives with respect to the deformation parameters are more complicated and merit additional attention.

If we expand the sum in equation 6, keeping the vector notation but omitting the vector permutations of the index of summation $\mathbf{h}$, we have:

$$\mathbf{D}(\mathbf{x}) = \mathbf{c}_D(\mathbf{r}-1)\beta^{(3)}\left(\frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta \mathbf{x}_D} - (\mathbf{r} - 1)\right) + \mathbf{c}_D(\mathbf{r})\beta^{(3)}\left(\frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta \mathbf{x}_D} - \mathbf{r}\right)$$
$$+ \mathbf{c}_D(\mathbf{r}+1)\beta^{(3)}\left(\frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta \mathbf{x}_D} - (\mathbf{r} + 1)\right) + \mathbf{c}_D(\mathbf{r}+2)\beta^{(3)}\left(\frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta \mathbf{x}_D} - (\mathbf{r} + 2)\right). \tag{17}$$

When we take the partial derivative of equation 17 with respect to a 3-tuple of deformation coefficients $\mathbf{c}_D(\mathbf{r})$, we have a single non-zero term:

$$\frac{\partial \mathbf{D}(\mathbf{x})}{\partial \mathbf{c}_D(\mathbf{r})} = \beta^{(3)} \left( \frac{\mathbf{x} - \mathbf{x}_D(\mathbf{r})}{\Delta \mathbf{x}_D} - \mathbf{r} \right). \tag{18}$$

This is the 3-tuple of deformation derivatives:

$$\frac{\partial \mathbf{D}(\mathbf{x})}{\partial \mathbf{c}_D(\mathbf{r})} = \left[ \frac{\partial \mathbf{D}(\mathbf{x})}{\partial c_{D,x}(r_x)}, \frac{\partial \mathbf{D}(\mathbf{x})}{\partial c_{D,y}(r_y)}, \frac{\partial \mathbf{D}(\mathbf{x})}{\partial c_{D,z}(r_z)} \right]^{\mathrm{T}}. \tag{19}$$

We will only be interested in the deformation derivative with respect to a single parameter. However, the various deformation derivatives will still be 3-element vectors. The derivative with respect to a x-component deformation coefficient is (the y- and z-components follow similarly):

$$\frac{\partial \mathbf{D}(\mathbf{x})}{\partial c_{D,x}(r_x)} = \left[ \beta^{(3)} \left( \frac{x - x_D(r_x)}{\Delta x_D} - r_x \right), 0, 0 \right]^{\mathrm{T}}. \tag{20}$$

## 2.4 Multiresolution optimization strategy

Minimization problems for large sample-based datasets are often benefited by a multi-resolution approach. Instead of constructing an image pyramid, we approach this technique in a somewhat different fashion. The registration process is automated by varying the deformations in the test image such that the mutual information of the two images is maximized, at which point corresponding structures in the reference image are brought into geometric alignment. We use L-BFGS-B [6], a limited-memory, quasi-Newton minimization package, to descend the expression for mutual information in equation 10 until termination criteria are satisfied. In order to avoid local minima, and to decrease computation time, the algorithm has been cast into the framework of a hierarchical multiresolution optimization scheme. A flowchart of the algorithm in the multiresolution context is shown in figure 1.

We keep the image size the same for each resolution step, but vary several other parameters as the minimization proceeds from coarser to finer resolution: the resolution of the deformation, the number of image samples used in computing mutual information, the degree of Gaussian image blurring, and the optimizer's termination criteria. We vary the parameters according to an empirically determined schedule, but following a general rule that the number of image samples is proportional to the deformation resolution. The resolution steps are denoted as $n_1 \ldots n_m$, where m is the number of multiresolution steps (typically m=4). The solution vectors of transformation parameters at coarser resolution levels become the starting vectors for the finer resolution levels.

One aspect of our multiresolution strategy is the hierarchical recovery of the deformation. In our approach, we initially recover a global, overall deformation in an attempt to match up the patient motion and lung boundaries. As we increase the resolution (in a multiresolution sense), we try to recover increasingly fine deformations. The resolution of the deformation grid is specified in terms of the number of control points placed over the image and is specified separately for each dimension of the deformation. As the number of control points increases, so does the deformation resolution, and the spacing between the control points decreases. If $\mathbf{n}_D = [n_x, n_y, n_z]$ is a 3-tuple of numbers of control points in the deformation, and $\mathbf{n}_{D,1} \ldots \mathbf{n}_{D,4}$ are the control points density vectors for each of the four multiresolution steps, then we proceed from resolution 1 to resolution 4 as follows. A deformation at resolution level $r$ is positioned on the volume at the control point locations $\{\mathbf{x}_{D,r}\}$ referenced by its spline deformation coefficients $\{\mathbf{c}_{D,r}\}$, i.e. $\{\mathbf{x}_{D,3}\}$ and $\{\mathbf{c}_{D,3}\}$ represent the deformation at resolution step 3.

If we are at current deformation resolution $\mathbf{n}_{D,1}$, and have a deformation defined on this grid, given by the deformation coefficients at this resolution $\{\mathbf{c}_{D,1}\}$ then we first place a new deformation grid over the volume at resolution $\mathbf{n}_{D,2}$. Next we compute what the deformation coefficients $\{\mathbf{c}_{D,2}\}$ need to be in order to have the same deformation at this higher resolution (this
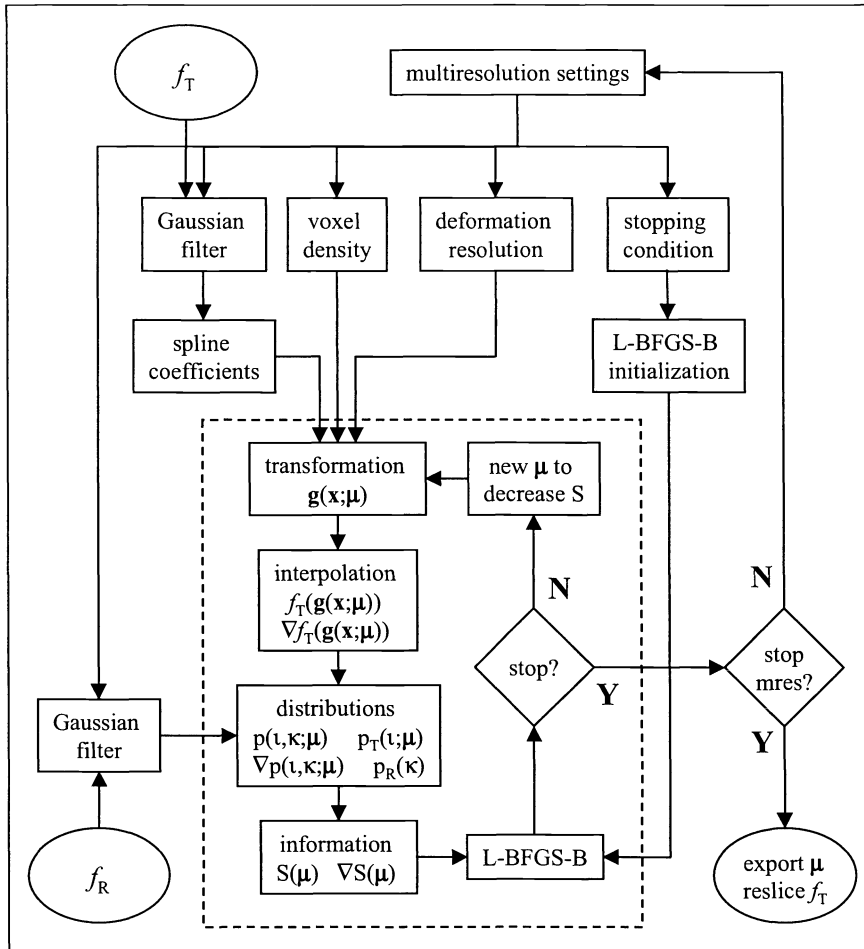
**Figure 1:** The registration algorithm in a multiresolution framework. The area enclosed in dashed lines is the algorithmic core for computing and optimizing mutual information for a single resolution step. $f_T$ and $f_R$ are the test and reference images, respectively. $\mathbf{x}$ is any geometric location in the images, $\mu$ is the set of transformation parameters. $p, p_T, p_R, \nabla p$ are the joint, marginal test, marginal reference, and joint derivative distributions, respectively. $S$ and $\nabla S$ are the estimates of mutual information and its gradient, respectively, between the two images. L-BFGS-B is an optimization package that searches the parameter space of $\mu$.

becomes our starting point for the second resolution step). We do this by calculating the deformation $\{c_{D,1}\}$ at the first resolution level for each of the control point locations $\{\mathbf{x}_{D,2}\}$ of the second:

$$\mathbf{D}(\mathbf{x}_{D,2}) = \sum_i \mathbf{c}_{D,1}(i)\beta^{(3)}\left(\frac{\mathbf{x}_{D,2} - \mathbf{x}_{D,1}(i)}{\Delta\mathbf{x}_{D,1}}\right). \tag{21}$$

Then we compute the spline coefficients of the new grid of deformation values:

$$\mathbf{c}_{D,2} = \Gamma\left(\mathbf{D}(\mathbf{x}_{D,2})\right), \tag{22}$$

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

We want the lower resolution steps to converge quickly and bring the independent variables (that define the deformation) close to maximizing mutual information. Besides starting with low resolution deformations, Gaussian blurring is applied to the images with a kernel that narrows as multiresolution proceeds. Smoothing the images tends to smooth the mutual information critereion, thereby avoiding local maxima.

## 2.5 Validation

Validating the performance of an image registration algorithm presents a host of challenges, and the poor scores for literature searches on this topic highlight the difficulties. The lack of a gold standard complicates matters further, preventing any quantitative 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, and time-consuming. There is a trade-off 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 based on a limited sample size.

For algorithms with linear transformation models, retrospective validation can be performed if fiducial markers are in place at the time the scan is performed, a method that will in general fail as the imaged anatomy is subject to non-linear motion either as the patient is moved between scanners 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 phantom or an image simulation algorithm. Such studies provide insight into the performance range of the algorithm but still mean little in terms of absolute performance once the algorithm is applied to 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 our images, do not terminate at a single 3D 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 underscored. While validation methods involving human assessment will be prone to error, bias, and inconsistencies, if two images are presented in a conducive manner, experienced radiologists 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 registration should be precise and where the required accuracy can be relaxed.

Our aim with the experiment is to assess the algorithm only in anatomically relevant areas. We are unsure of where the algorithm will be successful and where it will fail, and certain regions are of more clinical importance to researchers interested in utilizing the registration algorithm. We will take 7 images from each patient set: 5 axial slices (4 regularly spaced through the lungs, and 1 in the upper abdomen), 1 coronal slice at the carina, and 1 sagittal midline slice. Additionally, we will include 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 allow only 2D image navigation, and present only specific images for assessment. The interface is based on the 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 gradations is used to assess the registration accuracy for the given image pair. The observer makes two error assessments for each image pair: overall and maximum. 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. In the case absence of anatomy out of the plane of the image, the observer will have to make the best assessment possible based on his/her knowledge of the anatomy.

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. Window and level settings for CT images are standardized, but there is no comparable parameterization for attenuation values. Instead, we apply an empirically determined intensity transformation to the PET attenuation values and display the TR image in converted Hounsfield units. Slider bars are available for adjusting the window and level settings for both images and there are presets for common values to render anatomy such as the mediastinum and the lungs.

# 3. DATA

The images acquired for the validation studies presented here are part of a lung study at the University of Washington Medical Center. They are scans from patients who have been screened for lung nodules, lymphoma, etc. All patients had PET scans on the UWMC's GE Advance Scanner; a 15 minute TR scan with an energy beam of 511 keV was acquired immediately before the emission data was recorded. A diagnostic CT scan was also performed at the UWMC, however using a variety of GE scanners. The patients received the scans within at most 2 months from one another, and did not have surgical resection or a high degree cancer development between acquisitions. The TR scans occupy 3 fields-of-view (FOV) and are exported from the PET station as a single 3D floating-point binary data set of size *128x128x103*. The TR voxels have a physical size of *4.29x4.29x4.25* mm; therefore each TR scan axially images 437.75mm. The CT scans are axial images with varying slice thicknesses that in practice vary from 2-7mm. Once the patient scan is acquired, the in-plane resolution is set to magnify the patient anatomy as much as possible, within the range of 0.7-1.1mm. The portion of anatomy imaged in the CT scans varies from patient to patient, and in general the CT FOV is smaller than that of the TR scan (a typical CT FOV covers 350mm of axial anatomy).
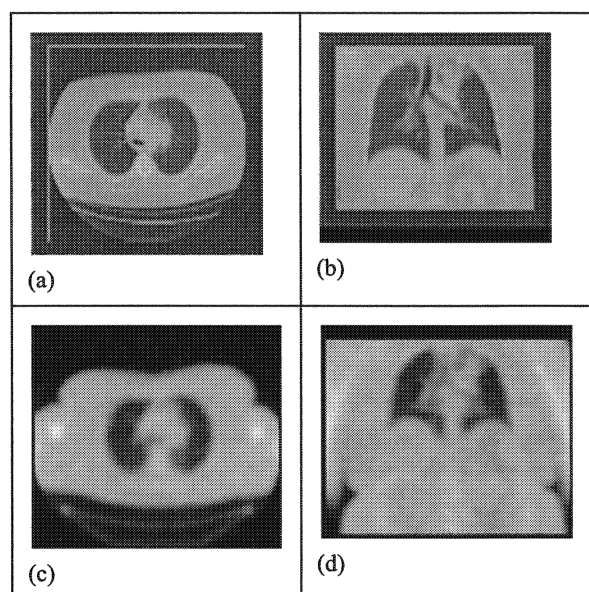


**Figure 2:** Sample CT and TR images illustrating the necessity for non-rigid registration. The top row shows (a) axial and (b) coronal views of a typical CT dataset while the bottom row shows (c) axial and (d) coronal views of a typical TR scan.
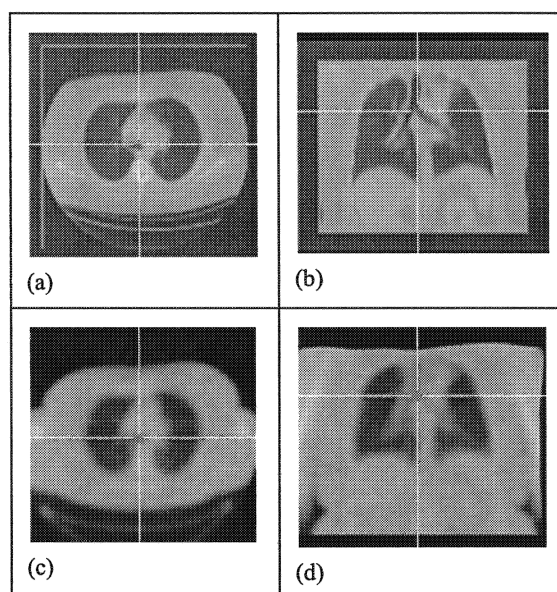


**Figure 3:** Sample images from a registered TR-CT pair. The top row shows (a) axial and (b) coronal views of the CT while the bottom row shows (c) axial and (d) coronal views of the transformed TR image. The crosshairs show are at the same image coordinates in both CT and TR volumes.

Preprocessing steps are performed before registering the images. The images are filtered with a Gaussian blur, the TR image is resliced to have isotropic voxels, and the spatial resolution of the CT is reduced to match that of the TR. 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. Manually-placed polygons in several axial slices are extended through the volume. The voxels within this volume *(V* in equation 16) represent an optimal set of image samples with which to compute mutual information.

# 4. RESULTS

28 patient sets were registered for the validation experiment, with 1 failing due to a limited CT FOV; all data presented is for the 27 patients successfully registered, which was validated by 2 radiologists. Example images from a registered dataset are shown in figure 2 with a locked cursor, while figure 4 shows sample images from the validation user interface. The errors in ranges are recorded by the user interface as error points, with 0 points representing the lowest error and 4 points being the largest possible error selection; i.e. each point corresponds to a range of error. The means, medians, and standard deviations of the overall and

maximum assessed errors for both observers are given in table 1. The error statistics fall into the first and second smallest error ranges, respectively. We can also group the results by anatomic region and compare assessed errors in all patients for a given region and for all the anatomy. The assessed error by anatomy is shown in figure 5. The assessed errors in axial slices in the upper lung regions are small, with the poorest performance in the abdomen. The transaxial slices (sagittal and coronal) are also assessed poorly compared to the lung regions. There is some tendency for observer 2 to grade more harshly than observer 1, but their choices are fairly well correlated ($\rho$=0.68). Also, the intraobserver consistency is well within acceptable ranges.

**Table 1:** Global statistics for validation experiment.

| Statistic | Overall error in points/ (range in mm) | Maximum error in points/ (range in mm) |
|---|---|---|
| mean | 0.54 (0-5.72) | 1.21 (5.72-11.44) |
| median | 0.39 (0-5.72) | 1.10 (5.72-11.44) |
| standard deviation | 0.48 (0-5.72) | 0.58 (0-5.72) |

The fully automated registration process takes an average of ~100 minutes on a moderate workstation to determine the optimal set of transformation parameters that maps coordinates in the test image to coordinates in the reference image. The process is divided into two registrations: a rigid body (~10 minutes, 6 parameters) followed by a deformation (~90 minutes, ~2000 parameters) recovery. Each registration is performed within a multiresolution framework that proceeds from "coarse" to "fine" resolutions in 4 steps. Computation time is most strongly effected by the number of transformation parameters and the number of voxels used to estimate mutual information. The multiresolution parameters used for the registrations presented here are summarized in table 2.
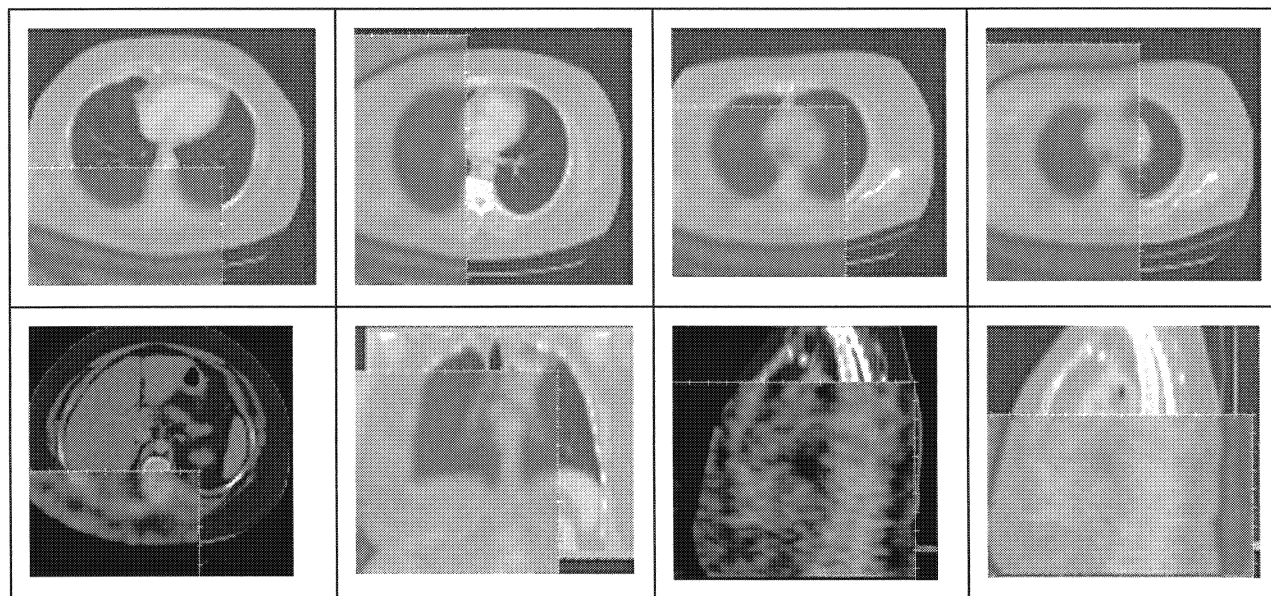


**Figure 4:** Example images from all 7 anatomic locations considered in the validation experiment, as presented to the observers with the split window viewer. Images are rendered using lung window/level unless noted otherwise. Top row from left to right: lower lung, mid-lower lung, mid-upper lung, upper lung. Bottom row from left to right: abdomen (mediastinal window), carina, midline (mediastinal window), midline (repeat in lung window).

## 5. CONCLUSIONS

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 feel the errors are 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
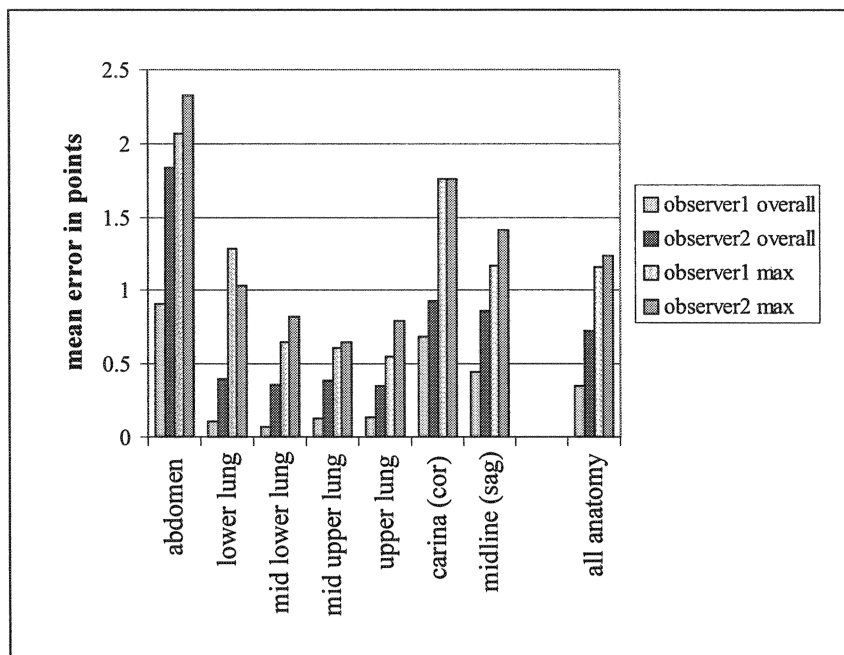
**Figure 5:** Mean overall and maximum assessed errors by anatomy for both observers for the 27 patient datasets. The average performance in all the images is given in the last data series.

**Table 2:** Multiresolution parameters used for registering the 27 datasets

| Parameter | Multiresolution schedule $n_1 \ldots n_4$ |
|---|---|
| Gaussian kernel $\sigma$ | 16, 8, 2, 0 |
| Isotropic deformation resolution (# variables) | 3, 5, 7, 9  (87, 381, 1035, 2193) |
| Percentage voxels used (constant reduction factor) | 0.8, 3.4, 9.3, 19.7 ($9 \times 10^{-5}$) |

observers, the overall error in the considered anatomy is 0.54 error points, which is in the 0-6mm error range. Recall that the PET voxel size is ~4.3mm on a side. Divided into 5 regimes, the possible range of error is 0-29mm. The mid to upper lung regions are registered the most accurately with a mean overall assessed error of 0.24 error points, which is also in the 0-6mm error range. The assessed performance in the abdomen is the worst in general with a mean overall assessed error of 1.37 error points, which is in the 6-11mm error range.

The algorithm produces registrations with relatively large assessed error in the coronal and sagittal planes. Anatomically, however, the chest varies much more *transaxially* than in any given axial slice and we may have a situation in which increasing the deformation resolution in this dimension may yield better results. Additionally, larger lung and soft tissue regions are perhaps driving the deformation, and the smaller airway anatomy is moved along with them. This highlights the principal limitation of the algorithm: the inability to define deformations on a non-uniform grid. We do however have ideas on how to approach this extension.

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 attenuation and emission values simultaneously, which may possible improve the overall accuracy in the torso.

In addition to the more general question of how fine a deformation grid do we *need*, we often face the question of how fine a deformation grid *can* we use. If we notice troublespots in the registration, and we do not have the capability to use an irregular grid, why not increase the resolution of the deformation until enough control points are clustered in the difficult region? We believe that the robustness of the deformation is due in part to the coarseness of the grid, in that mutual information is adjusted such that large features are moved into alignment. When the grid density is increased, smaller and more local anatomic variations will be driving the deformation, and we might expect some rather violent local deformations. In this sense the coarseness of the grid somewhat regularizes the deformation as a whole. The registration accuracy in certain anatomic regions such as along the trachea can likely benefit from some form of clustering of deformation points.

We have validated the algorithm only for 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 (except perhaps SPECT) of the 3D 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.

## ACKNOWLEDGEMENTS

## REFERENCES

[1]     C. Studholme, D.L.G. Hill, D.J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment", *Pattern Recognition*, 32, pp. 71-86, 1999.

[2]     F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, "Multimodality image registration by maximization of mutual information", *IEEE Transactions on Medical Imaging*, 16, No. 2, pp. 187-198, 1997.

[3]     P. Thevenaz, M. Unser, "Spline pyramids for inter-modal image registration using mutual information", *Proceedings of the SPIE*, 3169, pp. 236-247, 1997.

[4]     D. Rueckert, L. I. Sonada, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images", *IEEE Transactions on Medical Imaging*, 18, No. 8, pp. 712-722, 1999.

[5]     M. Unser, A. Aldroubi, M. Eden, "Fast B-spline transforms for continuous image represetation and interpolation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, No. 3, pp. 277-285, March 1991.

[6]     J. Nocedal, S. J. Wright, *Numerical Optimization*, Chapters 8-9, Springer-Verlag Inc., New York, 2000.

[7]     J. Kybic, P. Thevenaz, A. Nirkko, M. Unser, "Unwarping of unidirectionally distorted EPI images", *IEEE Transactions on Medical Imaging*, 19, No. 2, pp. 80-93, 2000.