Non-Rigid Image Registration - application to CT-MRI alignment

B.M. van Wyk, Dr. N.L. Muller and Prof. B.M. Herbst

Applied Mathematics
Department of Mathematical Sciences
University of Stellenbosch

bmvanwyk@gmail.com, neil@dip.sun.ac.za, herbst@sun.ac.za

Abstract
In a collaboration with iThemba labs we are developing a system for the automatic registration of Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) images of the brain. In this article we discuss the non-rigid registration component of our system. Our non-rigid registration system uses Parzen window estimation of the joint histogram, a B-spline based free-form deformation as the non-rigid transformation and a gradient descent optimisation routine. Our system is tested using synthetic images, images of a specially designed phantom and real patient data.

1. Introduction
In collaboration with iThemba Labs we are developing a system for the automatic registration of Computer Tomography Imaging (CT) and Magnetic Resonance Imaging (MRI) volumes. This system is to be used in patient treatment planning for radiotherapy procedures offered at iThemba Labs.

There is inherent distortion present in both CT and MRI volumes as a result of the image acquisition processes. These distortions can take on the form of intensity artifacts, partial volume voxels, aliasing artifacts and deformations of the anatomy. These deformations are more prevalent in MRI volumes than CT volumes and are caused by a magnetic field which is not perfectly homogeneous in the volume of interest during MRI image acquisition.

In this article we will present the non-rigid registration process implemented in our system to address the above mentioned deformation distortion. We apply Thervanez et al [9] to non-rigid registration and define expressions for the Jacobian terms of the free-form deformation cost function. A simple and efficient image model based on B-splines and ideas from the free-form deformation is also introduced. Our system uses mutual information as the similarity measure, a Parzen windowing scheme for density estimation, B-spline based free-form deformation to model our non-rigid transformation and a gradient descent optimization routine.

The evaluation of system performance we make use of simple synthetic volumes of known distortion, MRI and CT scans of a specially designed phantom.

We describe free-form deformations as our non-rigid transformation in Section 2 and move on to discuss mutual information and the Parzen windowing scheme for density estimation in Section 3. In Section 4 we introduce our image model and show our optimisation strategy in Section 5. The experiment setup is discussed in Section 6 with results displayed in Section 7. We draw a few conclusions in Section 8.

2. Free-form deformation
We use a B-spline based free-form deformation (FFD) to model our non-rigid transformation. B-spline based FFD’s are popular choices for non-rigid transformations and have been used in non-rigid registration of the brain, chest, liver and breast [2]. What makes B-splines so popular is that they are computationally efficient, only operate in a local region around each control point (finite support) and are easy to visualise and understand. The main issue with B-spline FFD’s is that they require a regularisation term to control the deformation and prevent the infliction of control points.

2.1. B-splines
We use B-splines in the free-form deformation, image model and as the Parzen window function. We therefore briefly review B-splines. The definitions are taken from [3].

The first order basis function of degree 0 is defined as

\[
B_1(x) = \begin{cases} 
1, & x \in \left[\frac{1}{2}, \frac{3}{2}\right) \\
0, & x \notin \left[\frac{1}{2}, \frac{3}{2}\right]
\end{cases}
\]  

(1)

with higher order B-spline functions recursively defined by

\[
B_m(x) = \frac{2m + x}{m - 1}B_{m-1}(x + \frac{1}{2}) + \frac{2m - x}{m - 1}B_{m-1}(x - \frac{1}{2}).
\]  

(2)

The derivative of a B-spline function is given by

\[
\frac{d}{dx}B_m(x) = B_{m-1}(x + \frac{1}{2}) - B_{m-1}(x - \frac{1}{2}).
\]  

(3)

2.2. Free-form deformation definition
We follow the definition of the FFD from Rueckert et al [8]. To define a B-spline based FFD we denote the domain of the image volume as \( \Omega = \{(x,y,z) | 0 < x < X, 0 < y < Y, 0 < z < Z\} \). Let \( \Phi \) denote a \( n_x \times n_y \times n_z \) mesh of control points with uniform spacing \( [\delta_x, \delta_y, \delta_z] \). The FFD can then be written as a 3D tensor product of 1D cubic B-splines as

\[
T(x, \Phi) = \sum_{i,m,n=-2}^{2} B_4(u-i)B_4(v-m)B_4(w-n)\Phi_{i+j+m,k+n}
\]  

(4)

where \( i = \lfloor \frac{x}{\delta_x} \rfloor, j = \lfloor \frac{y}{\delta_y} \rfloor, k = \lfloor \frac{z}{\delta_z} \rfloor \) and

\[
\begin{align*}
u &= \frac{x}{\delta_x} - i, \\
v &= \frac{y}{\delta_y} - j, \\
w &= \frac{z}{\delta_z} - k.
\end{align*}
\]  

The
control points $\Phi$ are the parameters of the transformation and the amount of deformation that can be modelled depends on the resolution of the control point mesh. A low resolution mesh will model global non-rigid deformations while a high resolution mesh will more accurately model local deformations.

When calculating the transformation of a point that lies close to the edge of the control point mesh the B-spline based FFD (4) runs into trouble as the calculation makes use of control points external to the defined mesh. To handle these boundary cases we make use of a set of boundary control points that stay fixed during registration. The boundary control points extend the control point mesh at the mesh borders and have the same spacing as the control points.

The resolution of the control point mesh also determines the number of degrees of freedom in the transformation. A $10 \times 10 \times 10$ control point mesh will have 3000 degrees of freedom as each control point has three translation components. We therefore need to make a trade of between the mesh resolution and computational cost.

2.3. Regularisation term

B-spline FFD’s require a regularisation term to prevent the inflection of control points and to control the smoothness of the transformation. There have been various different regularisation terms proposed ([2], [11], [8], [5]) generally based on the first and second derivatives of the FFD transformation. We follow Rueckert et al [8] and use a regularization term that is the 3D equivalent of the 2D bending energy of a thin plate of metal. This term is chosen as it will penalise non-smooth transformations and prevent the inflection of the control points.

The regularisation term is given by

$$ C = \frac{1}{V} \int_V \left[ \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right)^2 + \left( \frac{\partial^2 T(x, \Phi)}{\partial y^2} \right)^2 + 2 \left( \frac{\partial^2 T(x, \Phi)}{\partial x \partial y} \right)^2 \right] dV $$

where $x$ is the coordinate of a voxel in volume $V$. Note that this regularisation term is zero for affine transformation and therefore only penalizes non-affine transformations [8].

The calculation of the six terms required in (5) all follow a similar procedure and use of the definition of the derivative of a B-spline (3) and the FFD defined in (4). The equation for $\frac{\partial^2 T(x, \Phi)}{\partial x^2}$ is given by

$$ \frac{\partial^2 T(x, \Phi)}{\partial x^2} = \sum_{i,j,k,m,n,m=0}^2 \frac{\partial^2 A_4(u-l)}{\partial u^2} B_4(v-m) \Phi_{i,j,k+m,n} $$

where $i,j,k,m,n$ and $\Phi$ are as described in Section 2.2. For the terms that require partials with respect to the $y$ or $z$ components or combinations thereof a similar procedure is followed.

3. Mutual information

We use mutual information as the similarity measure in our system. The use of mutual information for image registration was first proposed by Viola et al [12] in 1995 and has proved to be a very effective and popular measure for multi-modal image registration [10]. We start this section with the definition of mutual information and then move on to discuss the Parzen window scheme used for density estimation.

3.1. Definition of mutual information

Mutual information is a measure of the dependence between two random variables and is based on entropy. Entropy has its roots in information theory and is an attempt to measure information content, or more accurately, the uncertainty (dispersion) of a random variable [1]. The joint entropy similarly measures the amount of information contained in a system of two random variables. The marginal and joint entropy of random variables $X$ and $Y$ are defined as

$$ H(X) = -\sum_{i=0}^n p(x_i) \log(p(x_i)), $$

$$ H(X,Y) = -\sum_{i=0}^n \sum_{j=0}^m p(x_i,y_j) \log(p(x_i,y_j)) $$

where $H(X)$ is the marginal entropy and $H(X,Y)$ the joint entropy. The MI of two random variables is defined as

$$ S(X,Y) = H(X) + H(Y) - H(X,Y) $$

or more explicitly as

$$ S(X,Y) = -\sum_{i=0}^n \sum_{j=0}^m p(i,j) \log \left( \frac{p(i,j)}{p(i)p(j)} \right). $$

We make use of the negative mutual information as we search for the minimum of the similarity measure during optimisation.

3.2. Parzen window density estimation

We use a Parzen windowing scheme to estimate the densities from which we calculate the mutual information similarity measure. This is done to help us more accurately model the underlying intensity distributions from which our volume intensities (samples) are obtained.

3.2.1. Parzen window definition

The Parzen window estimate of a density distribution $P(x)$ given a window function $w(x)$ with unit integral and a set of samples $X = [x_0, x_1, ..., x_N]$ is defined as

$$ P_{\text{parzen}}(x) = \frac{1}{N} \sum_{i=1}^N w \left( \frac{x-x_i}{\varepsilon(N)} \right) $$

where $N$ is the number of samples available to us and $\varepsilon(N)$ a strictly positive scaling factor that controls the width of the window function $w(x)$ [9].

We follow Thevenaz et al [9] and use cubic B-spline functions for the Parzen window. This is done as B-splines have very useful properties such as being smooth, having explicit derivatives and finite support. B-splines are also positive and obey the unit integral constraint thereby being valid Parzen windows.
3.3. Density estimation

During registration we wish to align a source volume, \(V_s(x)\), to a destination volume, \(V_d(x)\) by optimising the parameters \(t = [t_0, t_1, ..., t_n]\) of some geometric transformation \(T(x, t)\). The volumes \(V_s(x)\) and \(V_d(x)\) are defined over a continuous domain \(x \in V^c\) that has three dimensions. The coordinates \(x_i\) are samples of \(V^c\) and the discrete set of these samples is called \(V\).

The evaluation of our similarity measure (4) requires the estimation of the joint and partial density distributions of the intensities of \(V_s(x)\) and \(V_d(x)\). Let \(I_s\) and \(I_d\) be the sets of intensities (samples) associated with \(V_s(x)\) and \(V_d(x)\). Let \(w(x)\) be a separable Parzen window defined in Section 3.2.1. Following Thevanez et al [9] the joint histogram can then be defined as

\[
h(i, k, t) = \sum_{x \in V} w \left( \frac{i-V_s(T(x, t))}{\epsilon_d} \right) w \left( \frac{k-V_d(x_i)}{\epsilon_s} \right)
\]

where \(i \in I_d\) and \(k \in I_s\). \(\epsilon_s\) and \(\epsilon_d\) are related to the size of the sample sets \(I_s\) and \(I_d\). To obtain the required joint intensity distribution, \(p(i, k, t)\), a normalisation factor

\[
\alpha(t) = \frac{1}{\sum_{i \in I_d} \sum_{k \in I_s} h(i, k, t)}
\]

is introduced such that

\[
p(i, k, t) = \alpha(t)h(i, k, t).
\]

This normalisation factor ensures that \(p(i, k, t)\) has a unit integral and replaces the term \(\frac{1}{\alpha}\) in (11). We can then extract the required marginal densities using

\[
p_s(k, t) = \sum_{i \in I_d} p(i, k, t)
\]

and

\[
p_d(i, t) = \sum_{k \in I_s} p(i, k, t)
\]

It is important to point out that even though the intensity distribution associated with the source volume, \(p_s(k, t)\), does not change during registration it depends on the transformation parameters \(t\) in (15). To overcome this Thevanez et al [9] use cubic B-splines as the Parzen window function. Cubic B-splines obey the partition of unity constraint and are shown to decouple this dependence [9].

Using (14-16) the negative of the mutual information similarity measure (10) can be written as

\[
S(t) = \sum_{i \in I_d, k \in I_s} p(i, k, t) \log \left( \frac{p(i, k, t)}{p_d(i, t)p_s(k, t)} \right)
\]

4. Image model

We require an image model that will allow us to interpolate the intensity values at coordinates in our image volumes as well as estimate the image gradients. Our model is based on B-splines of Section 2.1.

The image model maps the intensity samples \(f_i\) at coordinates \(x_i = [x_i, y_i, z_i]\) to a continuous function \(f(x)\). Our model follows similar thinking to the FFD transformation (4) where we consider the image volume as a equally spaced grid of intensity values with spacing \([\delta_x, \delta_y, \delta_z]\). The image model is defined as

\[
f(x) = \sum_{i, n, m=-l} C_f B_i(u-l)B_k(v-m)B_n(w-n)
\]

where \(i = \lfloor \frac{x}{\delta_x} \rfloor\), \(j = \lfloor \frac{y}{\delta_y} \rfloor\) and \(k = \lfloor \frac{z}{\delta_z} \rfloor\), and \(u = \frac{x}{\delta_x} - i\), \(v = \frac{y}{\delta_y} - j\) and \(w = \frac{z}{\delta_z} - k\). The \(C_f\) terms are the intensities associated with the voxels at coordinates \([i+l, j+n, k+m]\).

The components of the image gradient can be calculated using

\[
\frac{\partial f(x)}{\partial x} = \sum_{i, n, m=-l} C_f \frac{dB_i(u-l)}{du} B_k(v-m)B_n(w-n),
\]

\[
\frac{\partial f(x)}{\partial y} = \sum_{i, n, m=-l} C_f \frac{dB_i(u-l)}{dv} B_k(v-m)B_n(w-n),
\]

\[
\frac{\partial f(x)}{\partial z} = \sum_{i, n, m=-l} C_f \frac{dB_i(u-l)}{dw} B_k(v-m)B_n(w-n).
\]

5. Optimisation

We use a gradient descent optimisation routine in our system. For this we need to estimate the first order partial derivatives of our similarity measure and regularisation term. In this section we will discuss the calculation of these terms.

5.1. Optimisation of mutual information

We use Thevanez et al [9] for the calculation of our first order partial derivatives of the mutual information similarity measure. Thevanez et al define closed form expressions for the Jacobian and Hessian terms of the similarity measure and implement a Levenberg-Marquardt type optimisation approach for rigid registration. They are able to define a closed form expressions of these terms by taking advantage of their definition of the image model and Parzen estimate of the joint intensity distribution as continuous differentiable functions [4]. A similar method has been proposed by Mattes et al [6] for non-rigid registration of PET-CT chest images.

The gradient vector of our similarity measure (10) is defined as

\[
\nabla S = \begin{bmatrix}
\frac{\partial S}{\partial t_0} & \frac{\partial S}{\partial t_1} & \cdots & \frac{\partial S}{\partial t_n}
\end{bmatrix}
\]

where \([t_0, t_1, ..., t_n]\) are the parameters of the transformation. Each component of this gradient vector is calculated using

\[
\frac{\partial S}{\partial t} = -\sum_{i \in I_d, k \in I_s} \frac{\partial p(i, k, t)}{\partial t} \log \left( \frac{p(i, k, t)}{p_d(i, t)p_s(k, t)} \right).
\]
The partial derivative of the joint distribution can be expanded as
\[
\frac{\partial p(i, k, t)}{\partial t} = \frac{1}{N} \sum_{x \in C} B_i \left( \frac{k - V_t(x_i)}{\varepsilon_s} \right) \frac{\partial B_i(x)}{\partial \xi} \times \frac{1}{\varepsilon_d} \left( \frac{df(\mu)}{d\mu} \right) \cdot \frac{\partial T(x_i, t)}{\partial t} \tag{24}
\]
where \(\xi = \left( \frac{i - V_d(T(x_i, t))}{\varepsilon_s} \right)\), \(\mu = V_d(T(x_i, t))\), \(\frac{df(\mu)}{d\mu}\) is the image gradient, and \(\frac{\partial T(x_i, t)}{\partial t}\) is the geometric transformation gradient. The image gradient can be calculated using (19-21), we estimate the transformation gradient using a finite difference approximation. The derivative of the B-spline in (24) can be calculated using (3).

5.2. Optimisation of regularisation term
Let \(\Phi(a, b, c)\) be a function that maps the control point coordinate \((a, b, c)\) to a control point vector such that
\[
\Phi(a, b, c) = [a, b, c]^T = \Phi_{a, b, c} \tag{25}
\]
where \(\Phi_{a, b, c}\) is as defined in Section 2.2. Define the gradient vector of the regularisation term (5) as
\[
\nabla C = \left[ \frac{\partial C}{\partial \xi_1}, \frac{\partial C}{\partial \xi_2}, \ldots, \frac{\partial C}{\partial \xi_n} \right] \tag{26}
\]
where \([\xi_0, \xi_1, \ldots, \xi_n]\) are the parameters of our transformation.

The general form of a component of \(\nabla C\) is given by
\[
\frac{\partial C}{\partial \xi_i} = \frac{1}{V} \int_V \left[ \frac{\partial}{\partial \xi_1} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right)^2 \right] + \frac{\partial}{\partial \xi_1} \left( \frac{\partial^2 T(x, \Phi)}{\partial y^2} \right)^2 + \frac{\partial}{\partial \xi_1} \left( \frac{\partial^2 T(x, \Phi)}{\partial z^2} \right)^2 \tag{27}
\]
\[
+ \frac{\partial}{\partial \xi_i} \left( 2 \frac{\partial^2 T(x, \Phi)}{\partial x \partial y} \right) + \frac{\partial}{\partial \xi_i} \left( 2 \frac{\partial^2 T(x, \Phi)}{\partial x \partial z} \right) + \frac{\partial}{\partial \xi_i} \left( 2 \frac{\partial^2 T(x, \Phi)}{\partial y \partial z} \right) \right] dV. \tag{28}
\]
The calculation of each of the six terms in (27) follows a similar procedure. We work through the calculation of \(\frac{\partial}{\partial \xi_i} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right)^2\) as an example. We start with
\[
\frac{\partial}{\partial \xi_i} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right)^2 = 2 \frac{\partial^2 T(x, \Phi)}{\partial x^2} \frac{\partial}{\partial \xi_i} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right) \tag{29}
\]
and using (6) expand \(\frac{\partial}{\partial \xi_i} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right)\) to
\[
\frac{\partial}{\partial \xi_i} \left( \frac{\partial^2 T(x, \Phi)}{\partial x^2} \right) = \sum_{l,m,n=-2}^2 \frac{\partial^2 B_4(u-l)}{\partial u^2} B_4(v-m) B_4(w-n) \times \frac{\partial}{\partial \xi_i} \left( \Phi_{i+j+m+k+n} \right) \tag{30}
\]
where \(i, j, k, u, v, w\) are as described in Section 2.2. For the FFD transformation the parameters \([\xi_0, \xi_1, \ldots, \xi_n]\) are the control points in our mesh and with our definition of the control point function (25) we can rewrite \(\frac{\partial}{\partial \xi_i} \left( \Phi_{i+j+m+k+n} \right)\) as
\[
\frac{\partial}{\partial \xi_i} \left( \Phi_{i+j+m+k+n} \right) = \frac{\partial}{\partial \Phi_{a, b, c}} \left( \Phi(i + l, j + m, k + n) \right) \tag{31}
\]
where \(\Phi_{a, b, c}\) is the control point associated with parameter \(\xi_i\).

5.3. Optimisation algorithm
To obtain the transformation that registers our volumes we optimise over a cost function \(S_{total}\) that combines the similarity measure \(S\) and regularisation term \(C\) such that
\[
S_{total} = S + \lambda C. \tag{32}
\]
The scaling factor \(\lambda\) determines the effect of the regularisation term on the optimisation process. From testing it is found that a lambda value of \(1e^{-2}\) works well.

We make use of a gradient descent optimisation routine for registration. The algorithm proceeds as follows:

1. Initialise control point mesh \(T\)
2. Calculate initial cost \(S_{total}\)
3. while \((\beta > \epsilon)\) do
   - calculate gradient vector \(\nabla S_{total}\)
   - while \((S_{total} < S_{new})\) and \((\beta > \epsilon)\) do
     - \(T_{new} = T + \beta \nabla S_{total}\)
     - calculate \(S_{new}\) at the updated transformation
     - if \((S_{total} < S_{new})\) then \(\beta = \frac{\beta}{2}\)
     - else \(\beta = 2 \cdot \beta, T = T_{new}, S_{total} = S_{new}\)
4. return \(T\).

6. Experiment setup
For our experiments we use of three sets of volumes. One set is a simple synthetic pair of images related by a simple known deformation. The second set consisting of correspond CT and MRI brain volumes. Third is a set of CT and MRI volumes of our CT-MRI phantom.

The phantom is specially designed for CT-MRI quality assurance purposes [7] and has various compartments that can be filled with liquids that have tissue mimicking properties, a brain insert that provides complex structure in the head cavity and fiducial markers on the exterior surface that can be used to evaluate the accuracy of registration.

The rigid transformation that aligns the volumes for the non-rigid registration process is estimated using an implementation of the Levenberg-Marquardt registration process proposed by Thevenez et al [9].
7. Test results

7.1. Synthetic tests

The synthetic tests involve the non-rigid registration of the source and destination volume in Figure 1 and are used to show the correct working of the system. The deformation used in this test deforms the sphere in the source volume Figure 1(a) to an ellipsoid by stretching the sphere along the $x$-axis and compressing it along the $y$-axis.

On inspection of the registered volumes in Figure 2 we see a significant improvement in registration quality between the rigid and non-rigid transformation. Looking more closely at the non-rigid transformation in Figure 2(b) we see the deformation modelled by our system closely matches that of the distortion present in the synthetic volumes. Figure 3 shows the improvement in the difference of the synthetic volumes before and after the non-rigid registration process.

7.2. Phantom tests

An MRI and CT image of our phantom is shown in Figure 4 with the results of the non-rigid registration process displayed in Figure 5. Figure 5 (a) shows the MRI and CT volumes overlaid in a checker-board pattern after registration. In this image we see that the structure in the brain compartment of the phantom as well as the boundaries of the various compartments line up closely.

The perspex in the phantom has a high intensity in the CT volume and a low intensity in the MRI volume. The difference image, Figure 5 (b), has bright sharp edges at the perspex boundaries indicating the phantom volumes are aligned well. The cost of the deformation modelled by the non-rigid transformation was close to zero, this indicates there is very little distortion present in this volume pair.

7.3. Patient data tests

The MRI and CT volumes used in this test are displayed in Figure 6 with the results of the non-rigid registration process shown in Figure 7. On inspection of the overlay image Figure 7(a) we see edges of the head and the various air filled cavities, and the structure in the nasal area smoothly transitioning between the adjacent MRI and CT blocks of the checker-board pattern.

Bone shows up brightly in the CT volume and at a low intensity in the MRI volume. The difference image, Figure 7(b), shows well defined edges at bone structure indicating a good alignment.

The smoothness cost of the non-rigid deformation of this test is significantly higher then that of the phantom data tests, a close up of the deformation field can be seen in Figure 8. We notice that the deformation seems larger as we move further away from the centre of the image, this is consistent with the effect an inhomogeneous magnetic field would have during MRI image acquisition.
8. Conclusion

In this article we presented our system for the non-rigid registration of CT and MRI head volumes. We use the formulation of the mutual information similarity measure as a continuous differentiable function proposed by Thevenaz et al [9] and apply it to non-rigid registration. We further developed expressions for the required partial derivatives of the FFD cost function and introduce a simple and efficient image model based on B-splines and ideas from the FFD.

We tested our system with synthetic data, MRI and CT images of our phantom, and real patient data. The successful synthetic test of Section 7.1 show that our system is working correctly and that it is able to model simple non-rigid distortions. The tests using the phantom data (Section 7.2) indicate our system is working accurately modelling the small distortion present in the phantom volumes and not introducing unwanted noise. The final set of tests performed on real patient data (Section 7.3) again gave good results and showed a larger amount of distortion is present in the real data test volumes.

Future work includes developing routines to use the phantom to determine a geometric measure of the accuracy of our system. We will also investigate the effect on registration accuracy caused by the distortion and noise introduced by the various CT and MRI modalities and image acquisition routines.

9. References