DIFFEOMORPHIC TRANSFORMATIONS FOR AUTOMATIC MULTI-MODALITY IMAGE REGISTRATION

by

Ramkrishnan Narayanan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in The University of Michigan
2007

Doctoral Committee:

Professor Charles R. Meyer, Co-Chair
Professor Jeffrey A. Fessler, Co-Chair
Professor Alfred O. Hero
Professor Douglas C. Noll

ACKNOWLEDGEMENTS

The last five years at Michigan has been a huge learning experience, academically and personally. I have had the chance to meet some of the best faculty in my area of interest and have had great friends both in and out of the DIPL lab who have played a vital role in this experience.

None of this would have been possible without the continuous support of my advisor Chuck Meyer. I am indebted to the countless hours we have spent talking despite his busy schedule. I have always had the flexibility to work on problems that interested me with the freedom to consult him whenever I found myself stuck. I have also tried to learn from him the importance of research always keeping practicality in mind.

I have been fortunate to have Prof. Jeff Fessler as my co-advisor. His help in working through the math, helping me understand the problem better and stress on simulations and visualizations to get a good intuition of the theory is a lesson that will always be useful.

Jin has been my guide right since joining the lab. I am grateful for his willingness and patience in helping me understand the AVS environment initially and sharing his opinion on registration problems. Bing, has been a good friend and valuable resource. I would also like to thank Peyton for helping me tackle programming issues, running scripts and getting acquainted with AVS, Gary for his kind help with the hardware and data backup, and Tarry for helping me with the paperwork. In the process I

have also got to know them personally and I am grateful for that.

I would also like to thank Desmond, Roshni, Sarad and Jeremy in the DIPL lab and family and friends outside DIPL, Divya, Chitra, Anish, Abhishek, Vijay, Sam and Aniket for all the motivation and good times we spent together.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DIFFEOMORPHIC TRANSFORMATIONS FOR AUTOMATIC
MULTI-MODALITY IMAGE REGISTRATION

by
Ramkrishnan Narayanan

Co-Chairs: Charles R. Meyer and Jeffrey A. Fessler

Image registration is usually the first step before performing any post-processing operations such as surgical planning, volumetric measurements, diagnosis, etc. There are numerous registration algorithms that use any of several geometric interpolants to warp images. The deformation can be modeled by a suitable parameterization of the interpolant, through a uniform grid placement of control points or adaptively, where control points are only placed where images are misaligned. Nonparametric approaches do not use control points at all, e.g., fields regularized by elastic constraints.

There are two main challenges in control point based approaches: the choice of deformation model and the method of parameterization. While some transformations focus on modeling local changes, some on continuity and invertibility, there is no closed-form nonlinear parametric approach that satisfies all these properties. This dissertation presents a class of nonlinear transformations that are controllably local and continuous, and invertible under certain conditions. They are straightforward

to implement, fast to compute and can be used as alternatives to splines and radial basis functions.

The second challenge is the method of parameterization, that is, the location and scale at which control points are placed. Poor choice of parameterization results in deformations not being modeled accurately or over-parameterization, where control points may lie in homogeneous regions with low sensitivity to cost. This lowers computational efficiency due to high complexity of the search space and might also provide transformations that are not physically meaningful, and possibly folded.

This dissertation proposes a method to find mismatched locations in images and the spatial scale at which they are misregistered. Mismatch is specified based on location and smooth spatial scale (mismatch vector) at which local joint entropy is high.

First we show that mismatch vectors found by our method are in good agreement with known deformations applied to synthetic images. Next we use these attributes to parameterize our iterative registration method to demonstrate registration performance. The result is a completely automatic multimodality registration algorithm that achieves high accuracy of alignment (voxel sized errors) for the registration of brain structures in MR images.

CHAPTER I

Introduction

Medical imaging provides clinicians access to vast amounts of data that combine both functional and anatomical information of patients. While modalities such as X-Ray, Magnetic Resonance Imaging (MRI), Computed Tomography and Ultrasound (US) provide anatomical information, Positron Emission Tomography (PET), Single Positron Emission Tomography (SPECT) and functional MRI provide information on metabolic processes. Different modalities have their own advantages, e.g., MRI images help provide soft tissue contrast while CT provides information on skeletal structures, and PET is often used to study brain and heart function. A problem that commonly arises is how this data can be compared. Image intensities corresponding to the same scene cannot be compared directly because of the different sensors used, their parameters, acquisition related noise, different coordinate systems used, and tissue warps that occur due to metabolic or physical processes of the structures imaged. To be able to efficiently integrate and analyze data acquired, e.g., to perform diagnosis, treatment monitoring, dose estimation, stereotactic surgery, etc, the images must be aligned in a manner that the structures within, correspond with one another. This step is commonly referred to as registration or fusion.

Registration is also widely used outside medical image analysis. E.g., in com-

puter vision it is used to track objects or perform depth analysis, in remote sensing for surveillance, location and orientation of landmarks, and resource monitoring. In pattern recognition it could be used to authenticate signatures or character recognition. In medical imaging, several thousand papers have been published in the area of image registration depending on organs imaged, modality and type of analysis to be performed.

I will discuss all aspects of the registration strategy starting with the definition of the registration problem and will touch upon popular methods. For a broader understanding, I refer the reader to several reviews [1, 2, 3, 4, 5] that have been published over the last fifteen years. I will conclude with the desirable properties of registration algorithms, and set up the problem statement based on these challenges.

1.1 Registration problem

Following Hill's [4] definition, registration typically refers to the procedure by which one determines a transformation that relates the position of features in one image or coordinate space with the position of the corresponding feature in another image or coordinate space. Another definition also allows one to compare corresponding intensities but we will not use this because we are often interested in working with images belonging to different modalities. Mathematically, for two images A and B, sampled on grids x_A and x_B, we seek to maximize a similarity measure SIM. The geometric transformation that maximizes this similarity measure is the solution and is found as

$$\hat{T} = argmax_T \ SIM(A(x_A), \tilde{B}(T(x_A))) \tag{1.1}$$

Image A is usually referred to as the target and B is called the floating image. The transformation is found on the target grid x_A. The floating image B which is sampled

on x_B is pulled back to x_A, i.e., finding $\tilde{B}(x_A)$. We now have two images: $A(x_A)$ and $\tilde{B}(x_A)$ whose similarity can now be computed under the hypothesized transformation T. SIM is a generic similarity measure and is discussed in Section 1.4.

Recently, several papers that discuss the registration of multiple images have been published [6, 7, 8, 9]. This may be useful for atlas construction where several images are normalized to the same spatial frame. In same-patient tumor studies, it could be used to establish a consistent correspondence of voxels over time, in the region of the tumor to study the effect of therapy. The problem statement is similar to the pairwise case. For a set of N images, we define the estimated set of transformations as

$$(\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_{N-1}) = argmax_{T_1,\ldots,T_{N-1}} \ SIM(A(x_A), \tilde{B}_1(T_1(x_A)), \tilde{B}_2(T(x_A)), \ldots$$

$$, \tilde{B}_{N-1}(T_{N-1}(x_A))),$$

where A is picked as one of the images in the set and is usually chosen as the one most easy to register onto. B_i for $i = 1, \ldots, N-1$ are the other images in no particular order. It is also possible that no specific target image A is picked and all images are registered onto a common spatial frame. In this dissertation, I have only studied pairwise registration and extension of methods proposed, to multiple images will need further analysis.

Registration can be split into the following steps:

1. Transformation/Deformation model: Application of a transformation T to the target grid x_A

2. Interpolation: Compute the intensities of floating image $B(x_B)$ on the warp grid $T(x_A)$ to get $\tilde{B}(T(x_A))$

3. Similarity measure: Measure the similarity of the two images, $A(x_A)$ and $\tilde{B}(T(x_A))$

4. Optimization: Optimize transformation parameters so as to maximize the similarity measure

The specific transformation models, intensity interpolants, similarity measures and optimizers can vary widely depending on several factors. In the next few sections, I will discuss some of the commonly used approaches.

1.2 Transformation Models

The transformation that needs to be estimated through the maximization of any of the above similarity measures is rewritten below for convenience:

$$\hat{T} = argmax_T \ SIM(A(x_A), \tilde{B}(T(x_A))) \tag{1.2}$$

T above, can be estimated linearly or nonlinearly depending on the kind of deformation that the two images differ by. [10] presents a discussion on transformation functions typically used for registration. Affine transformations are global (displaces all voxels on the grid x_A), linear (with a translation part) and are usually the first step towards aligning the images.

Often parametric transformation models are initialized by specifying a correspondence of landmarks in the two images, given by $\mathbf{p}_i$ and $\mathbf{q}_i$, $i = 1, \ldots, n$. Any number (n) of such landmarks may be provided ($n \geq 3$ in 2D, $n \geq 4$ in 3D). The objective is to estimate a globally optimal solution T, i.e., maximizing the similarity measure (SIM) such that

$$\mathbf{q}_i = T(\mathbf{p}_i) \tag{1.3}$$

Nonparametric models are more recent and are typically used in large deformation problems where every voxel is allowed to move independently while constrained by an elastic or viscous fluid-based regularizer.

1.2.1   Affine

An affine transformation that relates the two sets of landmarks is given by

$$\mathbf{q} = A\mathbf{p} + \mathbf{b}$$

In the case of an over-determined system, the equality is not valid and the coefficients are optimally estimated in a least squares sense [11].

In 3D, A is a non-singular 3×3 matrix and $\mathbf{b}$ is a vector of length three. The transformation is described by twelve degrees of freedom. The new coordinates based on the affine transformation are

$$
\begin{aligned}
x' &= a_{11}x + a_{12}y + a_{13}z + b_1 \\
y' &= a_{21}x + a_{22}y + a_{23}z + b_2 \\
z' &= a_{31}x + a_{32}y + a_{33}z + b_3.
\end{aligned}
$$

Arun *et al*[12] have discussed a method of efficiently finding the rotation and translation matrices for two point sets. Such transformations are often inadequate in cases where organs may undergo local deformations and will necessitate the use of splines, radial basis functions or free form deformation models. The following sections discuss commonly used transformations for nonlinear registration.

1.2.2   Radial basis functions (RBF)

A radial basis function is a univariate continuous function that is radialized by composition with the Euclidean norm on the n-dimensional space. For a detailed

discussion on RBFs see Buhmann [13]. RBFs are essentially radially symmetric continuous functions centered about control point locations. The coefficients are weighted by a function that depends on the distance from the individual control points. Standard radial basis functions have the form

$$T(\mathbf{x}) = \sum_i^n w_i \phi(\|\mathbf{x} - p_i\|) \tag{1.4}$$

The solvability of the coefficients (w_i) for certain types of bases is not guaranteed. E.g., Thin Plate Spline (TPS) basis is not positive definite and a polynomial term is augmented to the interpolant to make it uniquely solvable yielding

$$T(\mathbf{x}) = \sum_i^n w_i \phi(\|\mathbf{x} - p_i\|) + \sum_j^m v_i P_j(\mathbf{x}). \tag{1.5}$$

Imposing the additional constraint that the coefficient vector $\mathbf{w}$ and the polynomial space are orthogonal, i.e.,

$$\sum_i^n w_i P_j(\mathbf{p}_i) = 0, j = 1 \ldots M$$

and substituting Eq. (1.5) in Eq. (1.4) we get the system

$$\begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix} = \begin{bmatrix} q_k \\ 0 \end{bmatrix}, \tag{1.6}$$

where k refers to the coordinate. K is an $n \times n$ matrix of basis functions: $K_{ij} = \phi(\|p_i - p_j\|)$. Table 1.1 lists the commonly used basis functions in registration. These

Table 1.1: Green's functions typically used for registration

Name	Dimension	$\phi(r)$
TPS	even	$r^{4-d} \log r$
TPS	odd	r^{4-d}
Multiquadratics	n	$(r^2 + c^2)^\mu, \mu \in R_+$
Inverse multiquadratics	n	$(r^2 + c^2)^{-\mu}, \mu \in R_+$
Gaussian	n	$e^{-\frac{r^2}{2\sigma^2}}$

RBFs do not have compact support. There have been several papers more recently in registration literature [14, 15, 16] that are built on theory developed by Wendland [17] and Wu [18]. These compact RBFs consist of a single polynomial piece in the unit interval $[0, 1]$. They are piecewise polynomial and positive definite which eliminates the need for the polynomial part in Eq. (1.5) and can be solved for directly as

$$Kw = q_k$$

K is a full rank matrix and the coefficients can be directly computed. Some $\psi_{d,k}(r)$(same as $\phi(r)$) functions of Wendland are listed below for dimension $(d) = 3$ and smoothness parameter k:

$$
\begin{aligned}
\psi_{3,0} &= (1-r)_+^2 \\
\psi_{3,1} &= (1-r)_+^4(4r+1) \\
\psi_{3,2} &= (1-r)_+^6(\frac{35}{3}r^2 + 6r + 1),
\end{aligned}
$$

where $(1-r)_+ = \begin{cases} (1-r), & 0 \leq r \leq 1 \\ 0, & otherwise \end{cases}$

Functions with support a having the same properties can be obtained by simply replacing r in the equation above with $\frac{r}{a}$. To the best of the author's knowledge, a result that proves the non-singularity of K for variable support at each control point location has not been reported. In the context of image registration, it is important that the transformations do not fold, i.e., preserve topology. Rohr $et\ al$ [15] discuss allowable support sizes a depending on the displacement ($\triangle$) of corresponding control points in the source and target for different ψ functions. It is not clear how the conditions would change if the supports intersected.

1.2.3 B-splines

Before use in image registration, B-splines were often used to interpolate and approximate scattered data [19, 20]. Several papers, notably [21, 22] have since appeared to model deformations in registration. In its most common form, the cubic B-spline is C^2 continuous and local, i.e., every voxel's displacement is only calculated from the position of the sixty four nearest (3D) control points around it. For a control point ϕ_{ijk} on a lattice ϕ, located at $(i, j, k), i = -1, 0, \ldots, m+1, j = -1, 0, \ldots, n+1$ and $k = -1, 0, \ldots, o+1$ transformation is calculated as

$$T(x, y, z) = \sum_{k=0}^{3} \sum_{l=0}^{3} \sum_{m=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l} \phi_{j+m} \phi_{k+n},$$

where B_l are the basis functions

$$
\begin{aligned}
B_0(u) &= \frac{(1-u)^3}{6} \\
B_1(u) &= \frac{3u^3 - 6u^2 + 4}{6} \\
B_2(u) &= \frac{-3u^3 + 3u^2 + 3u + 1}{6} \\
B_3(u) &= \frac{u^3}{6}, 0 \leq t \leq 1
\end{aligned}
$$

Indices i, j and k above are computed at each voxel as $i = [x] - 1, j = [y] - 1$ and $k = [z] - 1$. Also $s = x - [x], t = y - [y]$ and $u = z - [z]$. A grid of control points placed at different resolutions and the images are registered. Its local behavior makes it a desirable model in several applications. This property also enables the efficient computation of the pullback maps.

1.2.4 Viscous fluid flow

In nonparametric approaches, every voxel is assigned a deformation vector and is allowed to move independently, with suitable spatial regularization. Christensen

et al [23, 24] have used the Navier-Stokes equation to model viscous fluid flow:

$$\mu \bigtriangledown^2 \mathbf{v} + (\lambda + \mu) \bigtriangledown (\bigtriangledown.\mathbf{u}) + \mathbf{b}(\mathbf{v}) = 0, \tag{1.7}$$

where $\mathbf{b}(\mathbf{v})$ is the body force that drives the deformation into registration, $\mathbf{v}$ is the velocity of mass of a point passing through $\mathbf{x}$ at time t. The first term is associated with constant volume, and the second term relates to the growth or shrinkage of local regions within the target. [23] used a cost function based on a Gaussian sensor model. The expression for body force is obtained from the variation of the cost term with the displacement field. They estimated the displacement fields iteratively by solving a system of nonlinear partial differential equations through Successive Over Relaxation (SOR). These fields are constrained to be smooth, and re-gridding is performed when the transformations become singular. Such methods were limited to intra-modality registration. In a recent effort, Suetens [25] made several modifications, and extended it to inter-modality using mutual information.

1.2.5 Elastic

Elastic registration methods are also nonparametric. In [26], the authors present an elastic registration method that can be used with mutual information. For the transformation: $T(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, where $\mathbf{u}(\mathbf{x})$ is the displacement field, the matching problem is defined as

$$I(\mathbf{u}) = J(\mathbf{u}) + \alpha R(\mathbf{u}). \tag{1.8}$$

J is the similarity term that could be based on SSD, correlation or MI. R is the regularization term that applies smoothing constraints on the deformation, and is usually a function of the Jacobian of the deformation field. The displacements $\mathbf{u}$ can be determined by solving the Euler equations (setting the variational of I to zero). A gradient descent strategy was used in [26] to solve for the displacement fields.

1.3 Intensity interpolation

For the two images A and B sampled on x_A and x_B to be compared by a similarity measure, they need to first be normalized on the same grid. The deformation is applied to the target image grid x_A and the intensities in B are computed at the deformed locations $T(x_A)$ based on the knowledge of intensities $B(x_B)$. These intensities $\tilde{B}(T(x_A))$ are computed via interpolation. Several intensity interpolants have been discussed in the literature, e.g., nearest neighbor, linear, splines, etc. While these methods rely on finding the interpolated image $\tilde{B}(T(x_A))$, Partial Volume Interpolation (PVI) methods [27, 28] do not directly interpolate intensities. Instead they update the corresponding histogram positions fractionally. In 2D for a point (x, y) transformed to $(T_X(x, y), T_Y(x, y))$, four locations are fractionally updated in the 2D histogram corresponding to its four nearest neighbors based on the distance of $(T_X(x, y), T_Y(x, y))$ from the respective grid positions. See Pluim *et al* [29] for more information. Due to the discrete nature of the sampled data, these interpolants are known to induce artifacts in the registration that affect the smoothness of the similarity measure. Tsao [30] performed a comparison of commonly used interpolants for registration.

1.4 Similarity measures

Voxel similarity measures differ based on the intensity relationship of the two images.

1.4.1 Sum of Squared Differences (SSD)

SSD is one of the simplest similarity measures and has been shown to be the optimal measure when the two images differ only by Gaussian noise [31]. It is calculated

as

$$SSD = \frac{1}{N} \sum_{x_A} |A(x_A) - \tilde{B}(x_A)|^2, \tag{1.9}$$

where N is the number of voxels in the target image. This method is sensitive to even small contrast differences and effect of outliers. Sum of Absolute Differences (SAD) is less sensitive to the presence of outliers on the image boundary and is given by

$$SAD = \frac{1}{N} \sum_{x_A} |A(x_A) - \tilde{B}(x_A)|. \tag{1.10}$$

1.4.2 Correlation coefficient

Correlation coefficient is more robust measure in that it can be used when the images have a linear relationship. It is also sensitive to varying illumination, has a flatter similarity maxima and higher complexity compared to SSD or SAD. The most commonly used form is the normalized cross correlation given by

$$CC = \frac{\sum_{x_A}(A(x_A) - \bar{A})(B(x_B) - \bar{B})}{\sqrt{\sum_{x_A}(A(x_A) - \bar{A})^2 \sum_{x_B}(\tilde{B}(x_A) - \bar{B})^2}} \tag{1.11}$$

See [32] for a review on CC and its modifications. Recently correlation ratio [33] has been proposed to register multi-modality images assuming a known function that relates intensities.

1.4.3 Information theoretic approaches

Information theoretic approaches have now become standard to register images belonging to different modalities. They measure the statistical dependency between the two images. A study of Mutual Information (MI) - the most commonly used information theoretic method, and its relationship with other methods (correlation coefficient, correlation ratio) can be found in [34]. Entropic graph methods, have also

emerged more recently due to their direct extension to higher dimensions (multiple images).

Mutual information

Mutual information is now the most popular objective function for inter-modality registration. Pluim [35] surveyed mutual information based registration methods recently. In the paper, history of MI based registration is discussed that I will briefly repeat here to give the reader an idea of the evolution of measures for multi-modality registration. One of the first papers proposed was by Woods [36, 37] called the Variance of Intensity Ratios. The method was based on the assumption that the same grey value in images corresponded to the same tissue type. The ratio of corresponding voxels in image pairs was found and its variance was minimized. Hill [38] measured the extent of clustering of corresponding grey values in a feature space and later proposed [39] using third order moments of the histogram to measure the skewness. This further led to Collignon [40] and Studholme's [41] work on minimizing joint entropy which paved the way for mutual information based registration.

MI is a histogram method, i.e., it is measured from the density functions estimated from the joint histogram of the image pair. Analyzing structure and function has been made possible through the use of Mutual Information (MI) and joint entropy. Registration by maximizing MI proposed by Viola [42] and is given by

$$MI = H(A) + H(\tilde{B}) - H(A, \tilde{B}) = H(A) - H(A|\tilde{B}), \qquad (1.12)$$

where $H(A) = E_A(\log(p(A)))$ is the entropy of the target image A and $p(A)$ is its probability distribution. It can be thought of as a measure of how well $\tilde{B}$ explains A. If the images are well registered, then $\tilde{B}$ reduces the conditional entropy $H(A|\tilde{B})$ in the equation above resulting in high MI. Although joint entropy [41] can be used

directly, MI was found to vary more smoothly with mis-registration [4].

More compactly it can be rewritten as

$$MI(A, \tilde{B}) = \sum_a \sum_b p(a, b) \log_2(\frac{p(a, b)}{p(a)p(b)}), \tag{1.13}$$

where $p(a)$ and $p(b)$ are the marginal densities and $p(a, b)$ is the joint density respectively of A and B. Observe that $p(a)$ remains the same throughout the registration process and need not be recomputed each time. The marginal densities can be estimated empirically from the normalized histograms of the images or through Parzen windowing [43, 42].

MI was found to be more robust to poor image overlap compared to joint entropy. Studholme *et al* proposed Normalized Mutual Information (NMI)[44] to overcome this dependence on overlap by normalizing MI with the joint entropy estimated from the joint histogram of the overlap volume as

$$NMI(A, B) = \frac{H(A) + H(\tilde{B})}{H(A, \tilde{B})} = \frac{I(A, \tilde{B})}{H(A, \tilde{B})} + 1 \tag{1.14}$$

Graph methods

Graph methods [45, 46] for registration were proposed as an alternative to histogram methods because they eliminated the need to estimate the densities of underlying images, which are unknown to begin with. They can also be directly extended to higher dimensions, i.e., more than two images, unlike histogram based methods which requires large memory and run much slower in higher dimensions.

Given a set $Z_n = \{z_1, z_2, \ldots, z_n\}$ where z_i is defined as pixel pair of intensity levels (or projection of the images on a basis) associated with images A and $\tilde{B}$, or a longer vector depending on the number of images to be registered, a minimal spanning tree is an acyclic graph that passes through all coordinates associated with Z_n and is one

that minimizes the total edge weight:

$$L(Z_n) = min_e \sum_e \|e\|^\gamma \qquad (1.15)$$

All points on the graph are connected by $n-1$ edges e_i. $\|e\|$ is the Euclidian (L2) norm of the edge and γ is a real weight exponent $\gamma \in (0,1)$. It has been shown [46] that the length when normalized by $\sqrt{n}$ produces sequences that converge within a constant factor to the alpha entropies with $\alpha = 0.5$ when bivariate pixel intensity coincidences are implemented.

These measures have higher compute times for cost function evaluations compared to histogram methods in $2D$. To the best of the author's knowledge no studies have been performed comparing them with MI or joint entropy.

1.5 Optimization strategies

Different applications are characterized by different types of parameterizations, images sizes, similarity measures and deformation complexities. Broadly speaking, optimizers are either function value based, e.g., simplex, Powell's method, or gradient based, e.g., steepest descent and conjugate gradient. Suetens *et al* [47] have done a detailed study of the optimizers, multi-resolution strategies and how this effects the number of function evaluations.

1.6 General concerns

Before picking the "optimal "registration strategy for the application targeted, some factors to keep in mind are as follows

- Modality type: same or different

- Transformation: parametric or non-parametric, topology preserving or not, smoothness, locality

- Computational speed: real time, minutes or hours

- Initialization mechanism: multi-resolution grid refinement or adaptive control point placement

- Intensity interpolation methods: linear, spline based, partial volume

- Degrees Of Freedom: affects time taken and complexity of the search space

- Computation of the inverse transform, inverse consistency

1.7 Problem statement

The broader goal of this thesis is to provide a framework for completely automatic multimodality registration. Due to its applicability in multimodality registration, I have used mutual information or joint entropy to measure similarity or mismatch based on standard techniques. Optimization and intensity interpolation methods have also been well studied. One of the big challenges is the selection of a suitable transformation model and its parameterization throughout the image.

To run without user intervention, geometric parameterization that involves the placement of control points throughout the image must be done automatically. A straightforward solution would be to place a grid of control points uniformly over the entire image. However, control points placed in locations where the images are already well registered are unnecessarily optimized over. This increases the time taken and the complexity of the optimization. Additionally, if the images are relatively homogenous, moving these control points has little effect on the similarity measure and may provide geometrical deformations in that region that are not physically

meaningful and possibly folded. Researchers have suggested adaptive approaches where control points may be adaptively placed or optimized over, depending on local information content or image mismatch. No analysis has been made on the nature of the mismatch in terms of its location and spatial extent where control points, or in our case seed points must be initialized. Such an analysis would enable parameterization at the appropriate location and scale where the images are mismatched which could potentially reduce the number of degrees of freedom and possibly speed up computation.

The choice of the transformation model is also crucial. Several transformations are global in nature and may not be appropriate in applications where local changes such as tumor growths may need to be modeled. In such cases models whose support can be changed depending on the nature of the warps would be useful. The flexibility of being able to pick the smoothness of the transformation is a plus. In certain applications where deformations are not very large, smoother models could be used as opposed to large deformation situations where we would prefer a less smooth requirement to allow for greater flexibility of the transformation. The definition of a family of transformations that possess similar properties with variable degrees of smoothness and function profile under a common theoretical framework would be desirable. Finally transformations that are topology preserving, i.e., that yield solutions that are always invertible would be very applicable in registration problems where such models do not violate physically improbable tissue deformations. Many commonly used parametric transformations do not implicitly possess this property and often times a penalty must be imposed that discourages these solutions.

The challenges that this dissertation deals with are these two important aspects of registration, i.e., the initialization through an appropriate spatial parametrization

of control points and a transformation model that possesses the desirable qualities discussed in the previous paragraph. This would then enable us to combine the two strategies into a completely automatic registration framework for multimodality registration.

1.8 Dissertation overview

The dissertation consists of two parts: a parametric deformation model that is smooth and invertible, and an automatic method of parameterization for this model based on image mismatch through location and scale. The first part stemmed out of a need to replace Thin Plate Splines with methods that could model local changes without affecting distant regions already well registered, guarantee invertibility and provide variable degree of smoothness in the deformation. To eliminate user intervention, a means of efficient parameterization of this model was developed in the second part. Note however that the second part of the paper only discusses image mismatch location and scale without any special consideration for the type of geometrical model, i.e., B-spline knots could be initialized using this information as well. Although the second part is the first step of registration, I will discuss about it after discussing deformation models because it is essentially an independent procedure.

In chapter 2, I will discuss the geometric deformation model, its properties and derive conditions for invertibility and show some results. In chapter 3, I will present a method to estimate mismatch by computing local joint entropy through location and scale simultaneously and provide a visual and quantitative analysis to validate it. Chapter 4 combines the theory in chapters 2 and 3 to demonstrate a completely automatic multi-modality registration algorithm. Validation is performed using known warps on brain images to register tumors and other brain structures to arrive at voxel

sized errors. The conclusion in chapter 5 summarizes the contributions and future work.

1.9   MIAMI fuse

Our group has compiled a set of registration software packages built on Advanced Visualization System 5 (AVS 5). The implementation is a MI based similarity measure using TPS as the geometric interpolant. MIAMI fuse [48, 49, 50] stands for Mutual Information based Automatic Multimodality Image fusion. Due to its modular nature, MIAMI fuse in several modified forms was used for testing and validation of theory in this dissertation, specifically the modules that computed the similarity measure, deformation model parameters, warping and interpolation.

CHAPTER II

Diffeomorphic transformations: A local parametric approach

Methods for image registration have three main components: the geometric transformation used to model deformations, the objective function, and the optimization algorithm. While using Mutual Information (MI) as the objective function has been successfully explored and validated [42], finding a simple transformation possessing the useful qualities of smoothness, compact support, and the existence of an inverse has been an ongoing effort. Rigid or affine transformations cannot be used to recover local warps. Deformation fields that are solutions to Ordinary Differential Equations (ODEs) [51, 52] have been proposed because of their ability to recover large deformations while still being invertible. These methods have large number of degrees of freedom except Arsigny's Polyrigid transforms [53] and geodesic spline representations of diffeomorphisms [54]. In contrast the parametric transformation proposed here is a more parsimonious approach in that it can be applied only in regions that need correction.

Different types of radial basis functions with global e.g., Thin Plate Splines [48], and local support [14] are used as well. This is because they have fewer degrees of freedom and can be used to recover local warps. But these types of deformations are

not invertible in general. In addition, functions with global support change distant regions of the image that may not require correction while attempting to change local regions that do. B-Splines [21] have been used successfully because of their C^2 continuity and local support, but injectivity conditions are non-trivial [55]. In Arsigny's Polyrigid transforms [53] using ODEs, the deformation vector is obtained by integrating the velocity vector that is a distance weighted sum of individual vectors corresponding to 'action' points whose solution is the trajectory equation. This method always ensures that the transform is continuous and invertible. However these weights are normalized, so the transform is global. Furthermore, methods that use ODEs do not have a closed form and the deformation is computed by integrating the velocity vector in a finite number of time steps to obtain the transformation. This paper was motivated by the ideas discussed in [53]. We introduce a nonlinear transformation that possesses the properties discussed by modifying the affine transformation, so that at the center of the region that needs correction we have an affine transform described by all the parameters of the transform, and gradual convergence to identity as we move away from the center. This convergence can be controlled using our transform model. Also our transform has a closed form and is easy and fast to compute because it is characterized by few parameters and always ensures that an inverse exists under certain trivial conditions.

We show some preliminary results using a multi-scale approach to image registration by applying corrections starting from the coarsest level of scale to the finest. We applied synthetic B-Spline based deformations to images and then corrections were applied at three levels of scale using only one seed point at each. A seed point is the center of the region that we are trying to correct. They are picked based on finding high gradients of local MI with respect to local affine transformation parame-

ters. Results show that using these transforms could be a good alternative to current methods used in image registration.

We used normalized mutual information (NMI), first proposed by Studholme [44] as the objective function and a simultaneous perturbation based gradient optimizer [56] to maximize NMI.

2.1 The Locally Affine Transformation Model

This section introduces a locally affine transformation model for image registration. Such warps can be applied in specific regions of the image through appropriate parameterizations so as to model deformations locally while still being spatially continuous. They are derived through a modification of the commonly used global affine transformation model. In order to motivate the specific form of the model, let us first examine the affine transformation of vector $\mathbf{x}$ about the center $\mathbf{x}_0$ in $\mathbb{R}^n$ given by

$$T(\mathbf{x}) = e^{sA}e^{sS}(\mathbf{x} - \mathbf{x}_0) + s\mathbf{t} + \mathbf{x}_0, \tag{2.1}$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$, $\mathbf{x}_0 = \begin{bmatrix} x_{01} & x_{02} & \dots & x_{0n} \end{bmatrix}^T$, $\mathbf{t} = \begin{bmatrix} t_1 & t_2 & \dots & t_n \end{bmatrix}^T$ (translation), A is the skew symmetric matrix corresponding to the rotation matrix, S is the symmetric matrix corresponding to the scale matrix and $s \in [0, 1]$.

At $s = 0$ we have the identity transformation while at $s = 1$ we get the complete affine transformation about the center $\mathbf{x}_0$. All intermediate values of s render affine transformations of varying degrees between the two extents. One possibility of producing warps that appear locally affine would be to parameterize s in the transformation above spatially in the form of a continuous function, say $\lambda(r)$, where $r = \|\mathbf{x} - \mathbf{x}_0\|$ so that at $r = 0$, $s = 1$ and as r increases $s \to 0$ to get

$$T(\mathbf{x}) = e^{\lambda(r)A}e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0) + \lambda(r)\mathbf{t} + \mathbf{x}_0. \tag{2.2}$$

The locality of the transformation can now be easily controlled by the scale of the support function λ and we now have with warps that appear affine near $\mathbf{x}_0$ and approaching identity with increasing distance from $\mathbf{x}_0$. Although this appears to be a good model, proving that an inverse exists for a choice of affine parameters and support extents is difficult. An alternate possibility to model locally affine warps is writing this in the form of a composition of local warps namely rotate and scale, and translation to get the transformation $(T : \mathbb{R}^n \to \mathbb{R}^n)$

$$T(\mathbf{x}) = e^{\lambda(r)A}e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0) + \lambda(r')\mathbf{t} + \mathbf{x}_0 \tag{2.3}$$

where

$$r' = \|e^{\lambda(r)A}e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0)\| = \|e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0)\|$$

and

$$s \equiv \lambda(r) \tag{2.4}$$

One can also write T in Eq. (2.3) as

$$T(\mathbf{x}) = (T_T \circ T_{RS})(\mathbf{x}) + \mathbf{x}_0 \tag{2.5}$$

where

$$T_{RS}(\mathbf{x}) = e^{\lambda(r)A}e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0) \tag{2.6}$$

and

$$T_T(\mathbf{x}) = \mathbf{x} + \lambda(\|\mathbf{x}\|)\mathbf{t}. \tag{2.7}$$

Eq. (2.5) above is a composition of a local translation applied on a local rotate and scale about $\mathbf{x}_0$. We can show that the individual transformations are invertible under certain conditions making it easy to prove that the transformation is in general topology preserving as long as translations and scale are within certain bounds. Any support function satisfying the properties mentioned earlier can be used, e.g.,

Gaussians, piecewise polynomial functions, etc. In the case of the Gaussian the function $\lambda(r) = e^{-\frac{r^2}{2\sigma^2}}$. The simulations in this chapter are registrations applied to head images. Due to the milder nature of warps associated with head registration we chose the Gaussian support function although compactly supported functions could also be picked to model more complex deformations depending on the application involved, e.g., inter-patient registration for atlas construction where the inherent variability across the population needs to be modeled as opposed to physical or metabolic changes in same patient studies.

These warps can in general be applied as a composition of several warps while still preserving topology with local behavior and is discussed in a later section of this chapter. The location of seed points $(\mathbf{x}_0)$ about which the warps are centered can be picked so that they lie in mismatched regions. The scale of the support function can also be fine tuned in the optimization optionally along with the affine parameters to improve alignment.

The transform also satisfies many desirable properties discussed in the following subsections. These properties depend on the choice of λ which can be chosen to be local and smooth. The properties discussed here are for $\lambda(r)$ chosen to be gaussian in Eq. (2.4). One may select the function λ based on what properties one seeks to satisfy. The formulation for both the 2D and 3D case is conceptually the same. However, in 3D it is easier to represent rotation components by a rotation vector. A rotation about each coordinate is equivalent to a single rotation about the axis of the rotation vector whose magnitude specifies angle, and individual components specify orientation (see appendix).

2.1.1 Continuity and Locality

Continuity is determined by the choice of the function λ. The Gaussian ensures C^∞ continuity. Locality also depends on λ. The Gaussian function has "nearly" local support. Functions with strictly local support may also be used to arrive at different conditions for an inverse to exist.

2.1.2 Existence of Inverse

Current methods using spline-based deformation models have either difficult conditions to incorporate in the optimizer to prevent folding, i.e not invertible or use regularization methods that discourage folding by adding an additional smoothness term in the objective function [21]. In our method we derive loose bounds for the transformation parameters which are straightforward to implement and always ensure invertibility.

The Jacobian matrix for a transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ must be positive definite everywhere to ensure invertibility. We have found the conditions for which the determinant of the Jacobian of the transformation is positive to always guarantee an inverse (see Appendix). We picked λ to be Gaussian because of its loose bounds, infinite continuity and an easily controllable region of influence. Other functions like inverse multi-quadratics or differentiable local support functions of the type proposed by Wendland [17] may also be used and lead to similar conditions.

As λ Gaussian, shown in Eq. (2.3) the transformation T has an inverse as long as

$$\|\mathbf{t}\| < \frac{\sigma e}{\sqrt{2}}$$

and

$$0 < a < e^{e^{0.5}} \approx 5.2003,$$

where $\mathbf{t}$ is the translation vector and $a = max(a_x, a_y)$, the larger of the two anisotropic scales in the x and y direction. In our previous work we reported the bound for translation for the Gaussian as $\sigma e^{0.5}$ instead, which is incorrect and the correct bound is in fact $\frac{\sigma e}{\sqrt{2}}$ as reported in this dissertation.

It is worth mentioning at this point that the bounds above are sufficient for invertibility. A larger parameter space could exist where the transformation would still be invertible. However practically these bounds were found to be loose and we rarely experienced folding in our simulations. In the rare occasion that this occurred, the cost was reset to the original cost thereby discouraging these solutions. Table 2.1 summarizes bounds derived for two local support functions to compare with the Gaussian. The corresponding plots are shown in Fig. (2.1).

Table 2.1: Bounds for inverse

	$N(\mathbf{x}_0, \sigma)$	$\psi_a(r) = (a - r)_+^4(4r + a)/a^5$	$\psi_a(r) = (a - r)_+^2/a^2$
Translation	$\frac{\sigma e}{\sqrt{2}} \approx 1.9221\sigma$	$\frac{64}{135}a \approx 0.4741a$	$\frac{a}{2}$
Scale	$e^{e^{0.5}} \approx 5.2003$	$e^{\frac{625}{432}} \approx 4.2493$	$e^2 \approx 7.3891$

In general these bounds appear to be rather tight for the compactly supported Wendland functions. In order to allow a fair comparison of the bounds, equivalent support sizes were used as in [15] where $a = 3\sqrt{\frac{\pi}{2}}\sigma$ so that the integrals over the functions were equal. See Table 2.2 for the new comparison.

Table 2.2: Bounds for inverse under equal area

	$N(\mathbf{x}_0, \sigma)$	$\psi_a(r) = (a - r)_+^4(4r + a)/a^5$	$\psi_a(r) = (a - r)_+^2/a^2$
Translation	1.9221σ	1.7826σ	1.88σ
Scale	5.2003	4.2493	7.3891
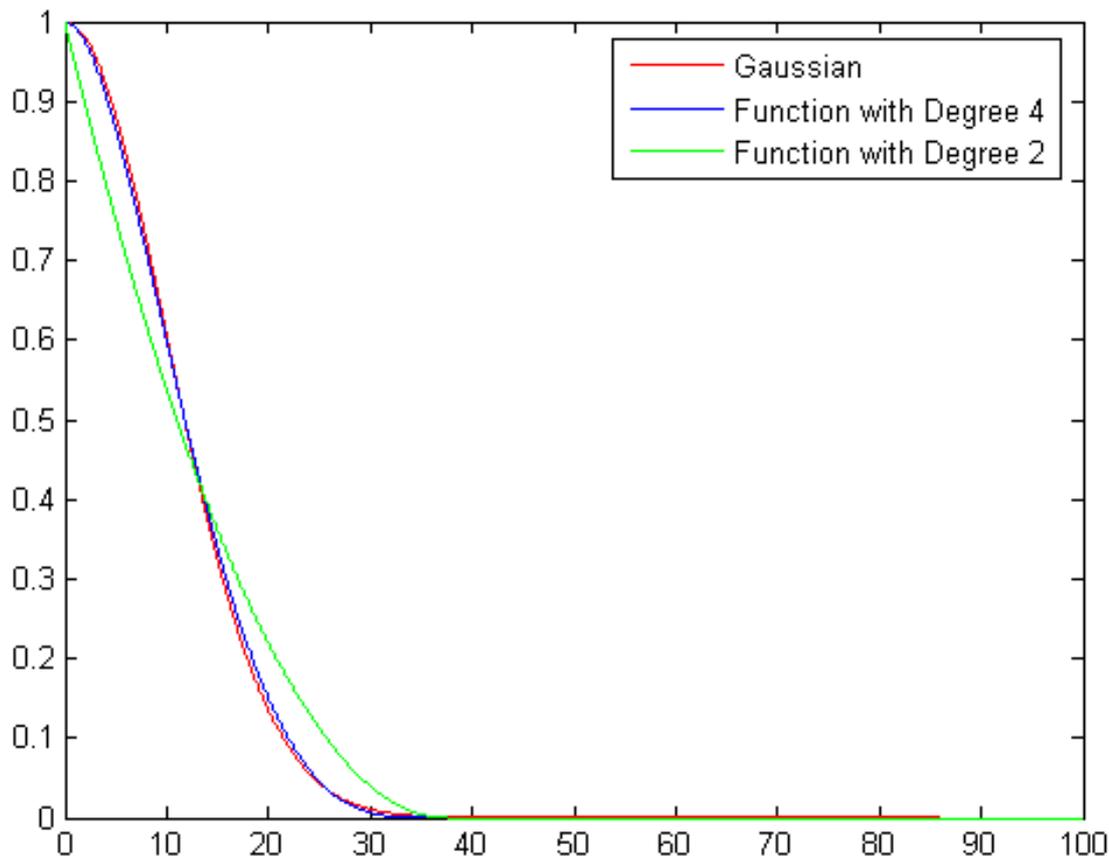
Figure 2.1: Gaussian compared with other functions with compact support

2.2 Initialization and Registration

2.2.1 Initialization

We implement a multi-scale approach to image registration starting from the coarsest level of scale and proceeding to the finest.

At each level of scale we pick only regions that are mis-registered and apply the algorithm. Rohde et al. [14] picked regions with large gradient of cost function with respect to radial basis function coefficients while Park et al. [57] used a mismatch measure to quantify mis-registration.

Here we pick regions based on its sensitivity to local affine deformations. Since

we apply corrections based on a locally affine transformation model, the gradients computed give us a meaningful estimate on the extent of mis-registration. The way these gradients are computed is as follows. A rectangular window is picked with dimensions in correspondence with the scale and three control points are placed in a triangular fashion spanning the area of the window. The window is placed in the reference and the floating image and the control points in the floating image are perturbed and the gradient of NMI with respect to the affine coefficients is found. This window is moved over the complete reference and floating image in an overlapping fashion. If the gradients of the cost in a region is not small, then it is likely that this region is mis-registered. Regions with large magnitude of gradient norm above a selected threshold are picked and the centers of these regions denoted as seed points are used in the global registration step. If p_i are the parameters that define our affine transformation, the gradient of local NMI is computed as

$$\hat{\mathbf{g}} = \left[\frac{\partial NMI}{\partial p_1} \frac{\partial NMI}{\partial p_2} \cdots \frac{\partial NMI}{\partial p_6} \right]^T$$

We apply transformations about these points and correct for them locally using the transformation model at different levels of scale. Since these points are also fed as parameters to the optimizer they will also be allowed to move to model the deformation better.

2.2.2 Multi-scale Nonrigid Registration

The final deformation is computed iteratively across different levels of scale. Since the spatial support of the deformation can be constrained to be local, seed points are picked in the initialization step at these different scales, and they serve as the centers for our locally affine transformation model. Global registration is then initiated at the coarsest level of scale (large σ) and optimization is performed over all the seed

points with large to smallest σ. The final transform is computed as a composition of individual transformations.

After optimizing over each region, the geometric maps are stored and this is repeated over other regions of the image. Since each of these transforms correct for only one region at a time, they have very few parameters and high local sensitivity yielding their ability to model local changes accurately. Also, only regions that are mis-registered are picked and corrected instead of placing a grid of control points and picking which ones are active (needing optimization) and inactive. This gives us a finer control over the region we are trying to correct. E.g. if we have N seed points, the final transformation is

$$T(\mathbf{x}) = (T_N \circ T_{N-1} \ldots T_2 \circ T_1)(\mathbf{x}), \tag{2.8}$$

where each seed point 'i' is associated with a transformation T_i.

The reference and the floating image are assumed to be already affine registered with each other before we begin the algorithm. The individual transformation parameters are computed for each seed point. There may be several seed points identified at a level of scale. Global normalized mutual information was used as the objective function and a simultaneous perturbation based gradient optimizer proposed by Spall [56] was used to arrive at the final solution. All eight parameters corresponding to the transformation were optimized: i.e. two translation parameters (t_x and t_y), rotation angle (θ), two anisotropic scale parameters in the scale matrix (a_x and a_y), two center coordinates (C_x and C_y) and a variance parameter (σ from the Gaussian function).

Algorithm

1: Initialize reference (A) and floating images (B) and set T_0 to an identity map

2: **for** $i = 1\ to\ Levels\ of\ Scale$ **do**

3:     $M = \#$ of seed points picked based on high local gradients

4:     **for** $k = 1\ to\ M$ **do**

5:         $\hat{T}_{i,k} = argmax\ _{T_{i,k}}\ NMI(A(\bullet), B((T_{i,k} \circ \hat{T}_{i,k-1} \ldots \hat{T}_{i,2} \circ \hat{T}_{i,1} \circ \hat{T}_{i-1} \circ \hat{T}_{i-2} \ldots \hat{T}_0)(\bullet)))$

6:     **end for**

7:     $\hat{T}_i = \hat{T}_{i,k}$

8: **end for**

9: $\hat{T} = \hat{T}_i$

(a)	(b)	(c)
(d)	(e)	(f)
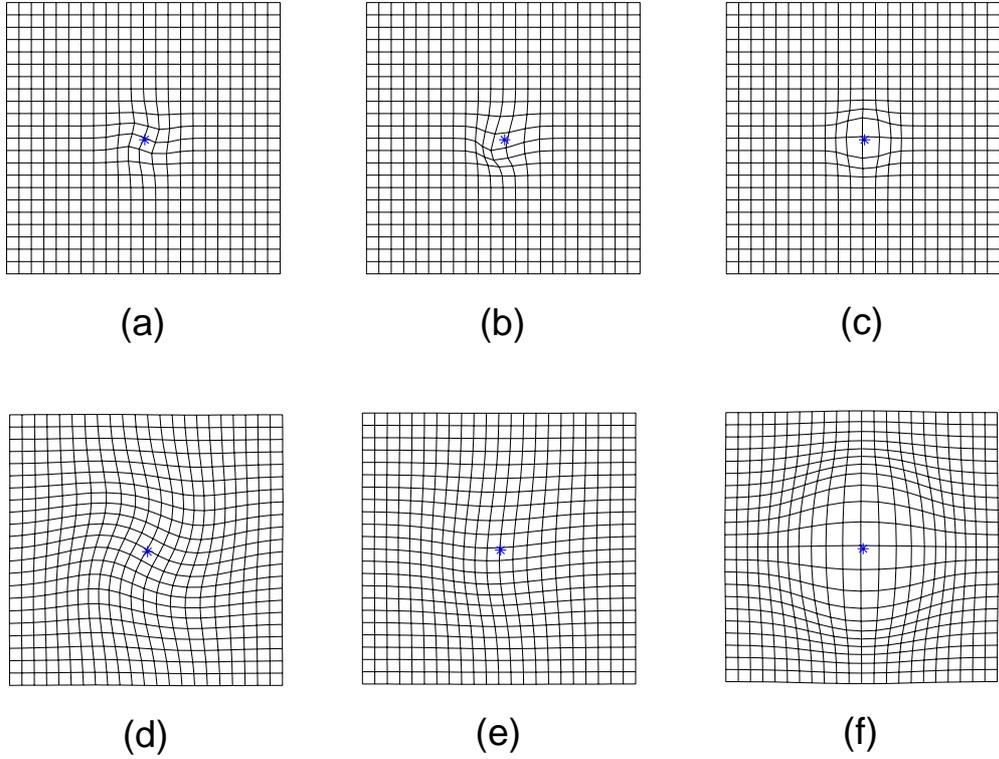
Figure 2.2: Deformations applied to a uniform grid at two different levels of scale (σ). The figure shows the same amount of rotation (a and d), translation (b and e) and scale (c and f) applied individually for a small and large σ respectively

2.3 Results

2.3.1 Examples of Locally Affine deformations

Fig. (2.2) shows examples of rotate, translate and scale applied individually about one seed point for two different σ. This is to show that we can model all kinds of local and global changes using a combination of these parameters.

(a)　　　　　　(b)　　　　　　(c)

(d)　　　　　　(e)　　　　　　(f)
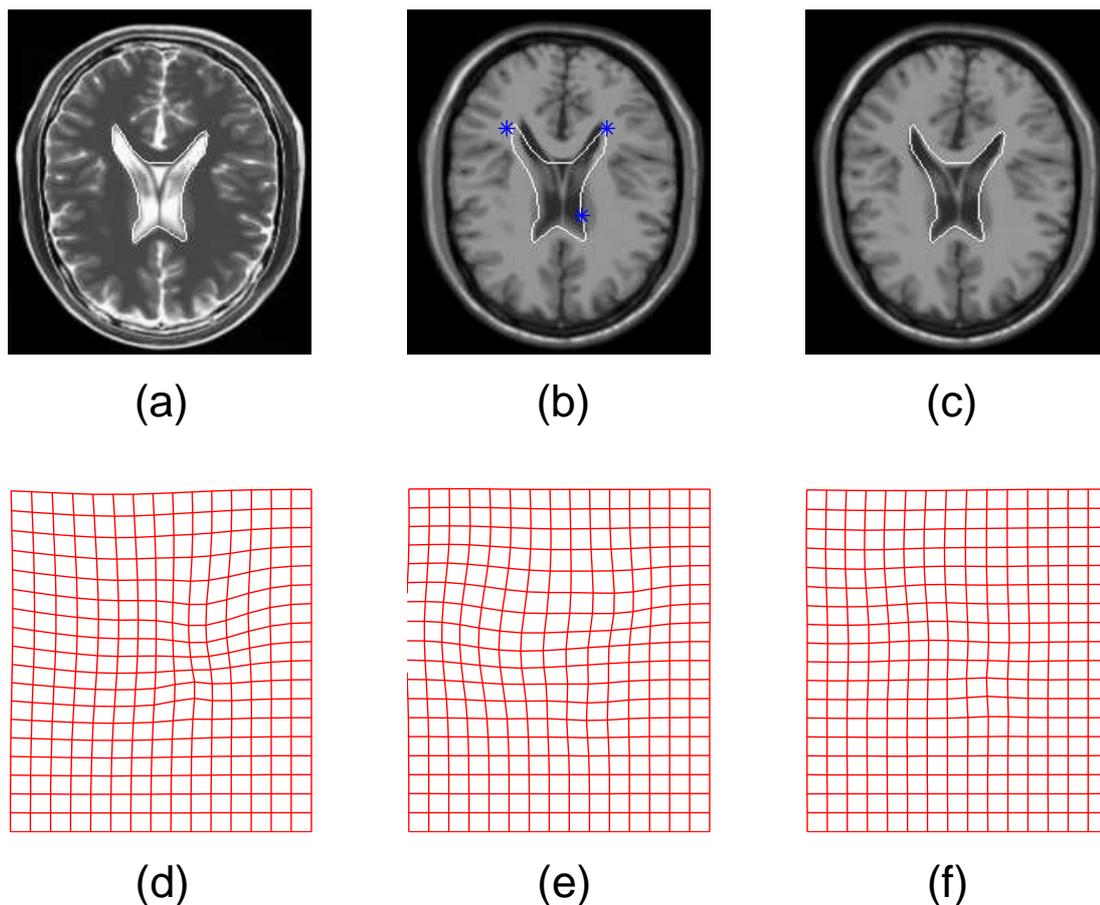
Figure 2.3: Registration of T1 and T2 weighted slices using three seed points (a) Original T2 weighted reference image. (b) artificially deformed T1 weighted floating image. (c) T1 weighted floating image after registration. (d) Applied Deformation (e) Estimated inverse after registration (f) Estimated inverse applied to the induced deformation
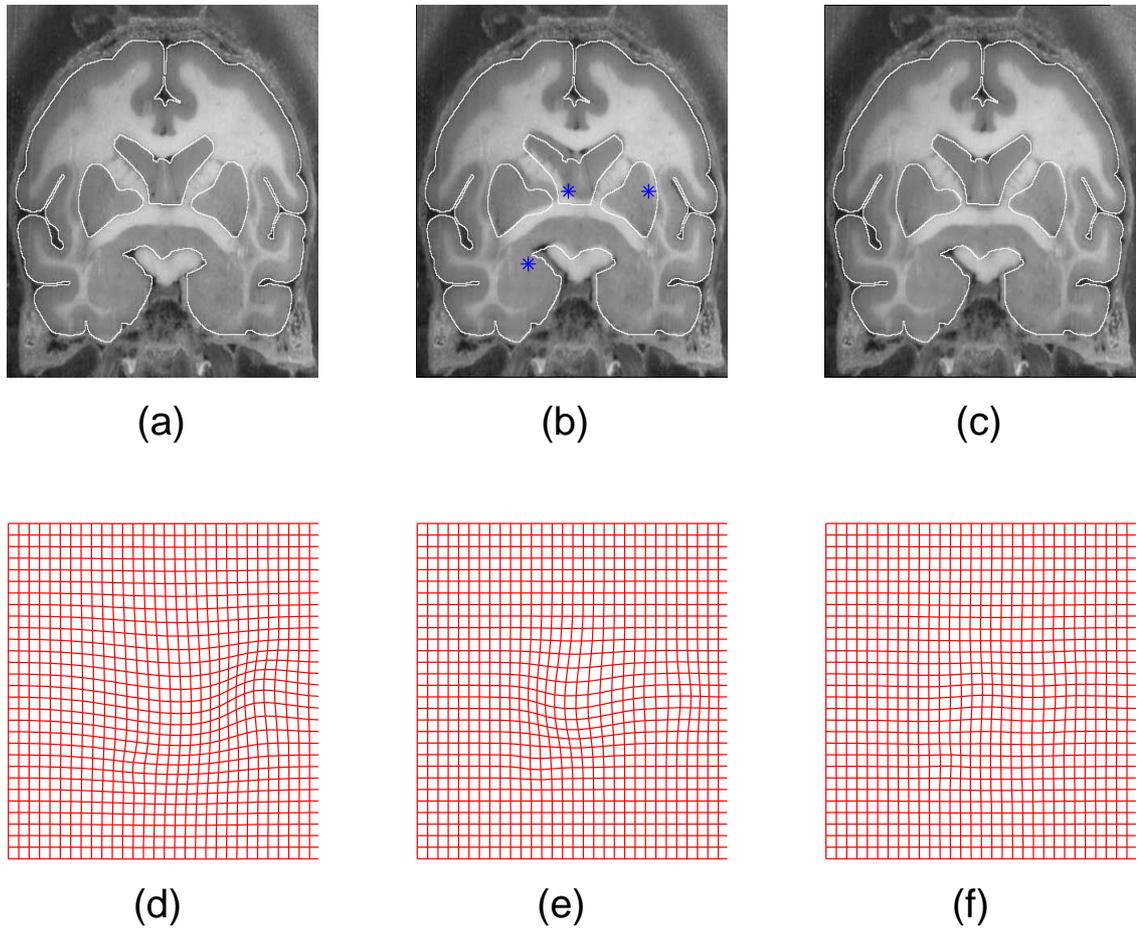
Figure 2.4: Registration of a coronal slice of a vervet monkey using three seed points. (a) Original reference image. (b) artificially deformed floating image. (c) Reconstructed floating image after registration. (d) Applied Deformation (e) Estimated inverse after registration (f) Estimated inverse applied to the induced deformation
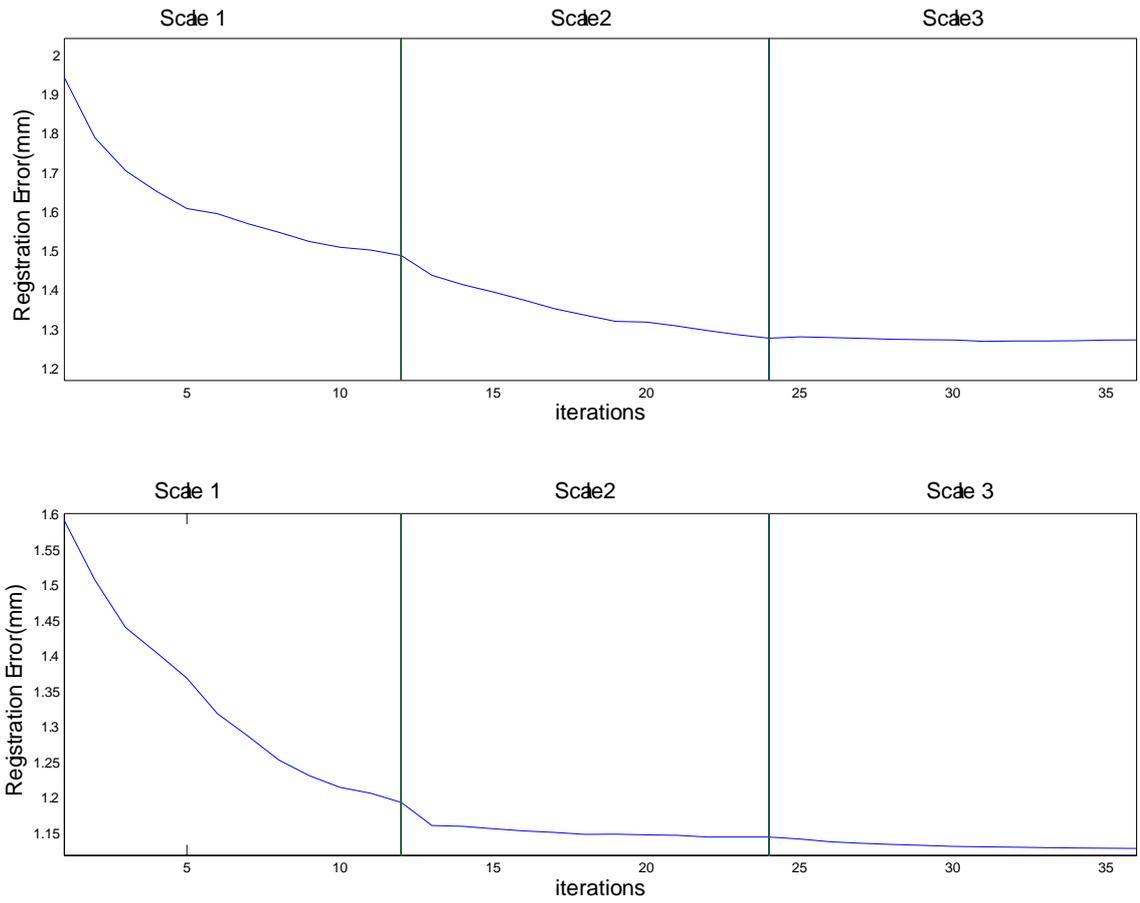
Figure 2.5: Average norm of registration error vs. iterations at three different levels of scale (σ). (a) Human brain - T1, T2 weighted registration. (b) Monkey brain registration

2.3.2 Registration Experiments

Fig. (2.3) shows a head registration example using an axial slice from T1 and T2 weighted images from Brainweb [58]. They were artificially deformed using B-Splines at three different levels of scale. This was done by moving one knot in the B-Spline grid by a known amount at each scale, refining the grid and repeating the procedure at the next level. Registration was performed using three seed points each working at a different scale. i.e different σ. The outline for the ventricle was marked manually so that the registration performance could be visually assessed. In Fig. (2.3), (a) is the original T2-weighted reference image and (b) is the deformed floating image with three seed points marked. The seed points from left to right are in decreasing levels of scale (σ), i.e. three optimizations were performed, one at each scale. The registered T1 image in (c) shows that the ventricles follow the contours more tightly after registration. (d) shows the applied deformation, (e) is the estimated inverse obtained via registration and (f) shows the deformation computed(e) applied to the induced deformation(d) which should resemble a uniform grid as best as possible.

Fig. (2.4) shows a coronal slice from a vervet monkey atlas developed at UCLA's Laboratory of Neuroimaging [59]. The slice was deformed using B-Splines similar to the procedure described in the previous paragraph. Seed points were placed at exactly three locations each with a different variance ($\sigma_{center} > \sigma_{right} > \sigma_{left}$) for the Gaussian function that controls the support of the transformation. The contours in (c) shows that the boundaries of the caudate and putamen hug the manually segmented boundaries more tightly, a marked improvement from (b).

Fig. (2.5) shows the registration error versus iterations at three different levels of scale for the first and second example respectively. Only twelve iterations were performed at each scale and each experiment took less than three minutes to run on

a 3.2 Ghz PC with 2 Gb memory running MATLAB 7.

2.4 Appendix

Here we derive conditions under which an inverse exists for a Gaussian weighting function. The conditions have been derived for a vector $\mathbf{x} \in \mathbb{R}^n$ (i.e. $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$) for rotation and scale about $\mathbf{x}_0 = \begin{bmatrix} x_{01} & x_{02} & \dots & x_{0n} \end{bmatrix}^T$. The bounds for 2D and 3D that we are interested in will turn out to be the same as the N dimensional case.

We will derive the conditions for translation, rotation and scale each treated individually and will show that these are sufficient conditions for an inverse to always exist.

For e.g., for λ gaussian we get,

$$\frac{\partial \lambda(r)}{\partial x_i} = -\frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \frac{(x_i - x_{0i})}{r} = -d\bar{x}_i,$$

where $d = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$ and $\bar{x}_i = \frac{x_i - x_{0i}}{r}$. Let $\bar{\mathbf{x}}$ be the direction cosine vector so that $\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_n \end{bmatrix}^T$ and $\|\bar{\mathbf{x}}\| = 1$.

2.4.1 Translation

Consider the case where the vector $\mathbf{x}$ is subjected to pure translation with no rotation or scaling (i.e. $T(\mathbf{x}) = \mathbf{x} + \lambda(\|\mathbf{x} - \mathbf{x}_0\|)\mathbf{t}$). One can show that the Jacobian of this transformation can be derived as

$$J = I + \mathbf{t}\nabla_{\mathbf{x}}\lambda(r). \tag{2.9}$$

$T(\mathbf{x})$ always has an inverse as long as $det(J) > 0$. Applying this condition we get

$$det(J) = 1 + \sum_i t_i \frac{\partial \lambda(r)}{\partial x_i} \quad > \quad 0 \tag{2.10}$$

$$1 + <\mathbf{t}, \nabla_{\mathbf{x}}\lambda> \quad > \quad 0 \tag{2.11}$$

$$1 + \frac{\partial \lambda}{\partial r} <\mathbf{t}, \bar{\mathbf{x}}> \quad > \quad 0 \tag{2.12}$$

Applying the Cauchy-Schwarz inequality to the inner product and substituting $\|\bar{\mathbf{x}}\| = 1$ we get the sufficient condition

$$\|\mathbf{t}\| < -\frac{1}{\frac{\partial \lambda}{\partial r}} \tag{2.13}$$

This condition can be further simplified based on the worst case scenario to get simpler conditions by choosing the smallest possible value the right hand side of the above equation can assume. For example, in the Gaussian case we get

$$\|\mathbf{t}\| < \frac{\sigma e}{\sqrt{2}}, \tag{2.14}$$

where $\sigma e^{\frac{1}{2}}$ is the smallest value possible for $-\frac{1}{\frac{\partial \lambda}{\partial r}}$.

Note that although the bounds may be loose practically speaking for registration, in that sufficiently large local translations can be accommodated for a given scale, there might still be a much larger parameter space of possible translations where the inverse may still exist. Such situations are possible because the conditions derived above are only sufficient. Additionally one can imagine more complex situations where two seed points in close proximity can cause equal and opposite translations that may be too large to satisfy the above condition and yet invertible.

2.4.2 Rotation

For $\mathbf{x} \in \mathbb{R}^n$, the rotation matrix can be constructed as the composition of elementary rotations in planar subspaces. Each of these matrices is a Jacobi rotation matrix. The rotation matrix is invertible as long as each of these matrices has an inverse. The Jacobian for a Jacobi matrix corresponding to the planar subspace containing axes 'i' and 'j' is given by

$$J = Q_{ij} \begin{bmatrix} p_1 & p_2 & & \\ & & 0 & \\ p_3 & p_4 & & \\ & & & \\ 0 & & I_{n-2} \end{bmatrix} Q_{ij}$$

where

Q_{ij} is the permutation matrix

$p_1 = \cos(\lambda\theta) + r\theta d\bar{x}_i^2 \sin(\lambda\theta) + r\theta d\bar{x}_i\bar{x}_j \cos(\lambda\theta),$

$p_2 = -\sin(\lambda\theta) + r\theta\bar{x}_i\bar{x}_j d \sin(\lambda\theta) + r\bar{x}_j^2\theta d \cos(\lambda\theta),$

$p_3 = \sin(\lambda\theta) - r\bar{x}_i^2\theta d \cos(\lambda\theta) + r\bar{x}_i\bar{x}_j\theta d \sin(\lambda\theta)$ and

$p_4 = \cos(\lambda\theta) - r\bar{x}_i\bar{x}_j\theta d \cos(\lambda\theta) + r\bar{x}_j^2\theta d \sin(\lambda\theta).$

The determinant of this matrix is always 1. The volume is always preserved under rotation. So the transformation $(T(\mathbf{x}) = e^{\lambda(r)A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0)$ always has an inverse

2.4.3 Scale

Finally consider the case when the transformation consists of only scaling. (i.e. $T(\mathbf{x}) = e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0$). Similar to the translation case, one can find the determinant of the Jacobian for scale and apply the condition $det(J) > 0$ to get

$$det(J) = 1 + r\frac{\partial\lambda}{\partial r} < \bar{\mathbf{x}}, \mathbf{s} >> 0 \qquad (2.15)$$

where $\mathbf{s} = \begin{bmatrix} \log a_1 & \log a_2 & \ldots & \log a_n \end{bmatrix}^T$ and a_i are the anisotropic scales in each dimension.

Let $s_{max} = \|\mathbf{s}\|_\infty = \log a$, i.e., the logarithm of the largest anisotropic scale $a = max(a_1, a_2, \ldots, a_n)$. Eq. (2.15) above can be simplified ($\|\bar{\mathbf{x}}\|^2 = 1$) to get the sufficient condition

$$1 + s_{max}r\frac{\partial\lambda}{\partial r} > 0 \qquad (2.16)$$

or

$$s_{max} < -\frac{1}{r\frac{\partial\lambda}{\partial r}} \qquad (2.17)$$

Depending on the choice of the support function, the right hand side can be replaced further by the smallest value it can assume. For e.g., in the case of the Gaussian we

get the condition

$$0 < a < e^{e^{0.5}} \approx 5.2003. \tag{2.18}$$

Typically the largest scale for existence of inverse was found to be sufficiently large to accommodate large local volume changes.

2.4.4 Conditions for Inverse

We have derived the bounds so that an inverse always exists for rotation, translation and scale each individually applied. Let T_R be the isomorphism for pure rotation (i.e. no scale or translation) so that $T_R(\mathbf{x}) = e^{\lambda(r)A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0$ and let $T_{R'}(\mathbf{x}) = T_R(\mathbf{x}) - \mathbf{x}_0$. We need to show that T in Eq. (2.5) has an inverse. Let us first show that T_{RS} in Eq. (2.6) is invertible. The transformation T_{RS} is

$$
\begin{aligned}
T_{RS}(\mathbf{x}) &= e^{\lambda(r)A}e^{\lambda(r)S}(\mathbf{x} - \mathbf{x}_0) \\
&= P\, T_{R'}(\mathbf{x})
\end{aligned}
$$

where

$$P = e^{\lambda(r)A}\begin{bmatrix} a_x^{\lambda(r)} & 0 \\ 0 & a_y^{\lambda(r)} \end{bmatrix}e^{-\lambda(r)A}.$$

Being similar to a diagonal matrix P is invertible. Also $T_{R'}$ is always invertible since T_R is. So T_{RS} always has an inverse as long as $a < e^{e^{0.5}} \approx 5.2003$.

We have already proved that T_T in Eq. (2.5) has an inverse as long as $\|\mathbf{t}\| < \sigma e^{\frac{1}{2}}$.(See Eq. (2.14)). So the transformation $T(\mathbf{x}) = (T_T T_{RS})(\mathbf{x}) + \mathbf{x}_0$ has an inverse as long as

$$\|\mathbf{t}\| < \frac{\sigma e}{\sqrt{2}} \tag{2.19}$$

and

$$0 < a < e^{e^{0.5}} \approx 5.2003. \tag{2.20}$$

2.4.5 Computation of Jacobian

The chain rule is used to find the derivatives of composed functions. In higher dimensions it may not be as obvious. For a detailed discussion see [60]. I have only discussed the notations and the final expression here.

Let $F(x_1, \ldots, x_n)$ and $G_1(x_1, \ldots, x_m), \ldots, G_n(x_1, \ldots, x_m)$ be differentiable functions of several variables, and let

$$H(x_1, \ldots, x_m) = F(G_1(x_1, \ldots, x_m), \ldots, G_n(x_1, \ldots, x_m))$$

be the function determined by the composition of F with $G_1, \ldots, G_n$

Let, $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ be open domains and let $F : Y \to \mathbb{R}^l$, $\quad G : X \to Y$ be differentiable mappings. In essence, the symbol F represents l functions of n variables each:

$$F = (F_1, \ldots, F_l), \qquad F_i = F_i(x_1, \ldots, x_n),$$

whereas $G = (G_1, \ldots, G_n)$ represents n functions of m variables each. The Jacobian determinant of G composed with F is the product of the Jacobian detorminant of F at $(x1, x2, \ldots, x_m)$ and that of G at $(G_1, G_2, \ldots, G_n)$. For a detailed discussion see [60]

The derivative of such mappings not a function but a matrix of partial derivatives called the Jacobian matrix.

Thus

$$DF = \begin{pmatrix} D_1 F_1 & \ldots & D_n F_1 \\ \vdots & \ddots & \vdots \\ D_1 F_l & \ldots & D_n F_l \end{pmatrix} \qquad DG = \begin{pmatrix} D_1 G_1 & \ldots & D_m G_1 \\ \vdots & \ddots & \vdots \\ D_1 G_n & \ldots & D_m G_n \end{pmatrix}$$

The chain rule now takes the same form as it did for functions of one variable:

$$D(F \circ G) = ((DF) \circ G)(DG),$$

In our locally affine transformation the function F refers to the translation component T_T and G is the Rotation-Scale component T_{RS}. See equations 2.6 and 2.7. We can then find the Jacobian individually(G and F) and estimate the local volume change as

$$det(D(F \circ G)) = det((DF) \circ G)\, det(DG)$$

or in this case

$$det(D(T_T \circ T_{RS})) = det((DT_T) \circ T_{RS})\, det(DT_{RS}) \tag{2.21}$$

We have already derived in Section 2.4.2 that local rotations are always volume preserving so that $det(DT_{RS}) = det(DT_S)$, i.e new volume depends only on scale. We can now easily find the Jacobian as

$$D(T_T \circ T_{RS}) = \left(\begin{pmatrix} 1 - d\bar{x}_1 t_1 & \dots & -d\bar{x}_n t_1 \\ \vdots & \ddots & \vdots \\ -d\bar{x}_1 t_n & \dots & 1 - d\bar{x}_n t_n \end{pmatrix} \circ T_{RS} \right) \cdot$$

$$\begin{pmatrix} a_1^{\lambda(r)}(1 - rd\bar{x}_1{}^2 \log(a_1)) & \dots & -a_1^{\lambda(r)} rd\bar{x}_1 \bar{x}_n \log(a_1) \\ \vdots & \ddots & \vdots \\ -a_n^{\lambda(r)} rd\bar{x}_n \bar{x}_1 \log(a_n) & \dots & a_n^{\lambda(r)}(1 - rd\bar{x}_n{}^2 \log(a_n)) \end{pmatrix} \tag{2.22}$$

The local volume at $\mathbf{x}$ can be estimated by computing the determinant of the matrix product as

$$det(D(T_T \circ T_{RS})) = (1 - d\bar{\mathbf{x}}^T \mathbf{t})|_{x=T_{RS}(x)} (\prod_{j=1}^{n} a_j^{\lambda(r)})(1 - rd \sum_{j=1}^{n} \bar{x}_j^2 \log(a_j)) \tag{2.23}$$

2.4.6   Rotation vector and the Rodriguez formula

Doing infinitesimal rotations about each axes could be very tedious. Since the rotation of a vector about each of the three axes can be done sequentially, one followed by the other, we can represent the rotation as a single rotation matrix. The Rodriguez formula[15] is a useful technique by which rotations about each axis in 3D can be performed by a single rotation about an arbitrary axis called the rotation vector whose coordinates specify the position of the axis, magnitude corresponds to the angle and sign refers to the direction of rotation. Consider a vector $\bar{x}$ that is to be rotated about an axes $\bar{a}$. The vector x can be split up into two components $\bar{x}_{\parallel}$ and $\bar{x}_{\perp}$. $\bar{w}$ is a vector normal to both $\bar{a}$ and $\bar{x}_{\perp}$ so that $\bar{w} = \bar{a} \times \bar{x}_{\perp}$. Let the operator 'R' be the rotation operator. So

$$R(\bar{x}_{\perp}) = \bar{x}_{\perp} \cos(\theta) + \sin(\theta)\bar{w} \tag{2.24}$$

$$R(x) = R(\bar{x}_{\perp}) + R(\bar{x}_{\parallel}) \tag{2.25}$$

$$= \cos(\theta)\bar{x}_{\perp} + \sin(\theta)\bar{w} + R(\bar{x}_{\parallel}) \tag{2.26}$$

$$= \cos(\theta)\bar{x}_{\perp} + \sin(\theta)\bar{w} + (\bar{a}.\bar{x})\bar{a} \tag{2.27}$$

$$= \cos(\theta)(\bar{x} - (\bar{a}.\bar{x})\bar{a}) + \sin(\theta)\bar{w} + (\bar{a}.\bar{x})\bar{a} \tag{2.28}$$

So,

$$R(x) = \cos(\theta)\bar{x} + (1 - \cos(\theta))(\bar{a}.\bar{x})\bar{a} + \sin(\theta)(\bar{a} \times \bar{x}) \tag{2.29}$$

Note that,

$$Symmetric \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} a_x^2 & a_x a_y & a_x a_z \\ a_x a_y & a_y^2 & a_y a_z \\ a_x a_z & a_y a_z & a_z^2 \end{bmatrix} = \bar{a}.\bar{a}^T \tag{2.30}$$

So,

$$Symmetric \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \bar{a}(\bar{a}.\bar{x}) \qquad (2.31)$$

And for any two vectors $\bar{a}$ and $\bar{x}$ the cross product in matrix form is given by,

$$\bar{a} \times \bar{x} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \qquad (2.32)$$

So,

$$Skew \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \bar{a} \times \bar{x} \qquad (2.33)$$

Substituting (2.31) and (2.33) in equation(2.29) we get the final expression for the Rodriguez formula for a rotation about an arbitrary axes.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} Symmetric \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} + skew \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \sin(\theta) + I \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (2.34)$$

CHAPTER III

Mismatch location and scale selection

Image registration provides geometric correspondences of structures using the same or different modalities. Over the last decade several innovative registration methods have been proposed and used successfully in medical imaging. For a survey see [4, 61]. Often an image pair is aligned initially based on an affine transformation before a nonlinear warp is applied. The nonlinear warp may be parametric (e.g., thin plate splines [48], B-splines [22], compact radial basis functions [15] or other parameterizations [53, 21, 14, 54, 62, 22, 63]), or non-parameterized free-form deformation fields based on an elastic [26, 64] or viscous fluid flow [23, 25] regularization.

Regardless of the specific deformation model used, the time needed to complete registration is related to the extent of geometrical misalignment, the time taken for each cost computation, and the total number of DOF used to describe the deformation. In certain applications the images may not belong to the same modality, necessitating the use of Mutual Information (MI) based methods [42] that maybe more time consuming than Sum of Squared Differences (SSD) or correlation. Even for intra-modality registration, MI is often used because of its robustness to differences in scanner parameters. Also registration times tend to be high if the images are large and deformations are complex with the added requirement that user in-

tervention is impractical. However often we can use fewer DOF if we place control points at locations having the most image mismatch.

There have been some papers that discuss spatially preferential parameterizations prior to registration. Rohlfing *et al.* [65] used multi-level B-splines for registration where a control point at a particular scale was held fixed (i.e., not optimized) if its local marginal entropies on the target and floating images were below 50% of the respective maximum entropies. Rohde *et al.* [14, 66], in their compactly supported radial basis function strategy, used only those control points that produced large gradients of cost. Schnabel *et al.* [67] also picked active control points based on cost gradients for their B-spline method. Park *et al.* [68, 57] added thin plate spline control points at each iteration at locations with large local mismatch. As far as we are aware, thus far no thorough analysis has been made to find mismatched regions and their degree of mismatch through an analysis of a wide range of spatial scales at all locations in the image pair prior to registration. For any of the several deformation models used, the knowledge of location and spatial extent of deformation could help drastically reduce compute times by avoiding parameterization of local regions already well aligned. This enables algorithms to work at several distinct locations at specific scales and yet contain the dimensionality of the optimization search space.

The choice of deformation model depends on the specific application. In our simulations we demonstrated the performance of this initialization approach of location and scale selection in conjunction with a parametric deformation model we have demonstrated previously [62].

The scale of a deformation model relates to the notion of scale space. Koenderink [69] discussed three aspects of scale in images: pixellation, i.e., inner scale, extent of

the Regions Of Interest (ROI), i.e., outer scale and the resolution of the histogram used to describe the intensity distribution of the image, i.e., tonal scale. In this paper we focus on the spatial extent of the mismatch between images so we will only use outer scale. The estimated mismatch vector contains the location and outer scale at which the images are mismatched. Any reference further to scale in the paper is to outer scale.

The ideas discussed here are extensions to commonly used approaches in the vision community to quantify local complexity in images using information theoretic measures. Gilles [70] used Shannon entropy to quantify local image complexity. Salient patches identified as peaks in scale space were used to estimate a global transformation between aerial reconnaissance images. Jagersand [71] used Kullback contrast between successive scales as a measure of differential information gain at a particular scale. Kadir [72] made several extensions to Gilles' idea notably the search through a wide range of scales in color images containing complex scenes and the use of a new saliency measure that weighs the entropy by a measure of self-dissimilarity in scale space. In this paper we use a similar approach to estimate feature mismatch for a *pair* of images using *joint entropy*. This makes it possible to study the mismatch of image pairs through feature space and scale space simultaneously. Further, we have computed mismatch through continuous outer scale (Gaussian) giving smooth joint entropy estimates. Our registration algorithm also optimizes over this support in addition to the transformation model parameters to provide fine control over scale. Building on our previous work from [62] and [73] we demonstrate a completely automatic multimodality registration algorithm.

The paper is organized as follows. We begin Section 3.1 by defining the mismatch vector. Sections 3.2 and 3.3 define some of the tools that we will use to find mismatch.

In Section 3.4 we show the accuracy and precision in location and scale to find mismatch using our method. Finally in Section 3.5 we outline the algorithm for mismatch vector selection. These mismatch vectors contain attributes that provide useful information to parameterize the deformation model we use for registration in the next chapter. We have included several examples to show that the mismatch vectors are in excellent agreement with the deformation errors in the mis-registered images.

3.1 Mismatch Measure

We propose to locate feature mismatch for a pair of images using joint entropy that is computed from local histograms. In three dimensions each mismatch vector found is made of individual location components C_x, C_y and C_z, and scale component α corresponding to mismatch and is of the form $[C_x \ C_y \ C_z \ \alpha]^T$, or $[C_x \ C_y \ \alpha]^T$ in two dimensions. In the next two sections we will discuss some of the tools we have used to find mismatch.

3.2 Local Histogram

The local joint histogram for a pair of images captures the distribution of intensities within a locally specified support. Koenderink [69] used histograms to represent local information content in images parameterized based on the "inner scale" (gaussian blurring window), "outer scale" (gaussian region of interest) and histogram resolution. We define the local joint histogram of two images, $I_1(\mathbf{r})$ and $I_2(\mathbf{r})$ as

$$h(i_1, i_2; \alpha, \beta, \mathbf{r}_0) = \frac{1}{2\pi\alpha^2} \sum_{\mathbf{r}} A(\frac{\mathbf{r} - \mathbf{r}_0}{\alpha}) e^{-\frac{(I_1(\mathbf{r},\sigma) - i_1)^2}{2\beta^2}} e^{-\frac{(I_2(\mathbf{r},\sigma) - i_2)^2}{2\beta^2}} \qquad (3.1)$$

where $A(\frac{\mathbf{r} - \mathbf{r}_0}{\alpha}) = e^{\frac{-(\mathbf{r} - \mathbf{r}_0)^2}{2\alpha^2}}$ denotes a gaussian window function with "outer scale" parameter α centered at $\mathbf{r}_0$, and $I_k(\mathbf{r}, \sigma) = I_k(\mathbf{r}) \otimes G(\mathbf{r}, \sigma)$ where $k = 1, 2$. i_1 and i_2

are variables that span the respective image intensities, G is the gaussian kernel of the scale space filter with kernel width σ, and β is the Parzen window width. We are interested in finding the location($\mathbf{r}_0$) and region of interest (α, i.e. outer scale) over which the two images are misaligned. The $2\pi\alpha^2$ factor is the normalization applied to the spatial gaussian window so that the histogram sums to one. The gaussian window parameterized by α prevents abrupt truncation of features and gives much smoother entropy profiles across scale compared to using a circular window.

3.3 Local joint entropy

The joint entropy is measured from the normalized local histogram described above and is computed as

$$H(\alpha; \mathbf{r}_0) = -\frac{1}{N_0} \sum_{i_1} \sum_{i_2} h(i_1, i_2; \alpha, \beta, \mathbf{r}_0) \log_2(\frac{h(i_1, i_2; \alpha, \beta, \mathbf{r}_0)}{N_0}) \qquad (3.2)$$

where N_0 is the sum of the histogram h. (N_0 is always 1 except near image boundaries). The spatial scale (α) at which the target image is misaligned with the floating image at location $\mathbf{r}_0$ is then estimated over a range of scales using joint entropy. If the two images are locally mis-registered at $\mathbf{r}_0$ and scale α^*, then we expect a peak in joint entropy at that scale.

3.3.1 Example1: Sensitivity to scale

Fig(3.1) shows the target images (left column) and corresponding floating images (middle column). The images have the same intensities within the shapes (circle and square). Clearly the images in the top row are mismatched at a lower scale in the local region centered at the circle compared to the images in the bottom row. The right column shows their entropy vs outer scale(α). The peaks show the scale at which the corresponding images are mismatched. Circles are drawn on the
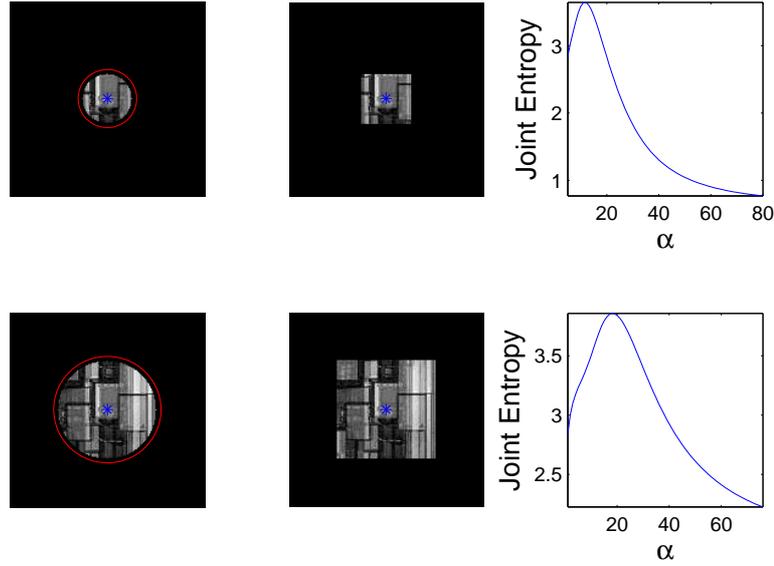
Figure 3.1: Entropy profile in outer scale. Left column: target images; Middle column: floating images; Right column: joint entropy vs. outer scale for center pixel denoted by *, i.e., $H(\alpha; \mathbf{r}_{center})$.

images in the left column with radius α^* that correspond to peaks in joint entropy. Observe that this mismatch in scale can be distinguished easily in outer scale. This is discussed in greater detail in following sections. Additionally the magnitude of entropy at this scale gives us a quantitative estimate of the relative importance of this location and scale. Later in this paper we show examples that demonstrate this and our ability to pick highly deformed regions for placing control points. The joint entropy is computed for the image pair at all locations through a broad range of scales and only peaks in joint entropy centered at each pixel through outer scale are picked as being mis-aligned. It is possible that multiple peaks may be found corresponding to more than one spatial scale being significant at a pixel location. In these cases we choose the scale corresponding to the larger magnitude of entropy.
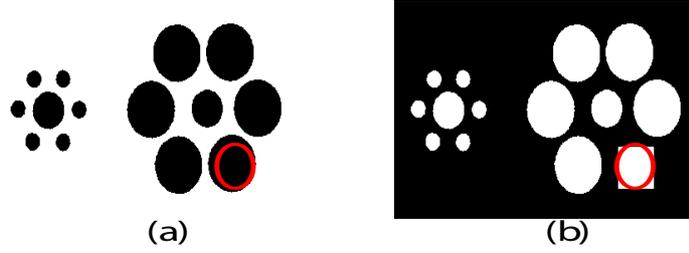
Figure 3.2: Search through feature space and scale space a) Target Image, b) Floating image. Center and radius of the red circle shows location and scale of mismatch.

3.3.2 Example2: Largest joint entropy location and scale

Fig. (3.2) shows several ellipses (3.2a) and a very similar pattern in (3.2b) except for the square in one position replacing the ellipse and the inversion of colors. We computed joint entropy over the entire image and found the peaks in scale at every pixel location. The red circle plotted on the first image shows the center of the region and the scale (circle radius) at which the joint entropy was highest in the entire image. Observe that it tracks the boundary of the ellipse in (3.2a) closely.

3.3.3 Example3: Spatial variation of joint entropy

Fig. (3.3) shows a local deformation applied to (a) to get image (b). The applied deformation error norm is shown in (c) and (d) is the absolute difference image ((b) and (a)). All calculations are made only along the horizontal line superimposed on the figures to get an idea of the linear spatial variation of peaks and corresponding joint entropy. (e) shows the computed scale at which peaks were found on this line and (f) shows the joint entropy computed at the peaks. The entropy is clearly maximum near the region of the deformation and decreasing with increasing distance from mismatch. The scale that corresponds to the maximum entropy is estimated from (e) which looks bowl shaped because different points on the line find mismatch
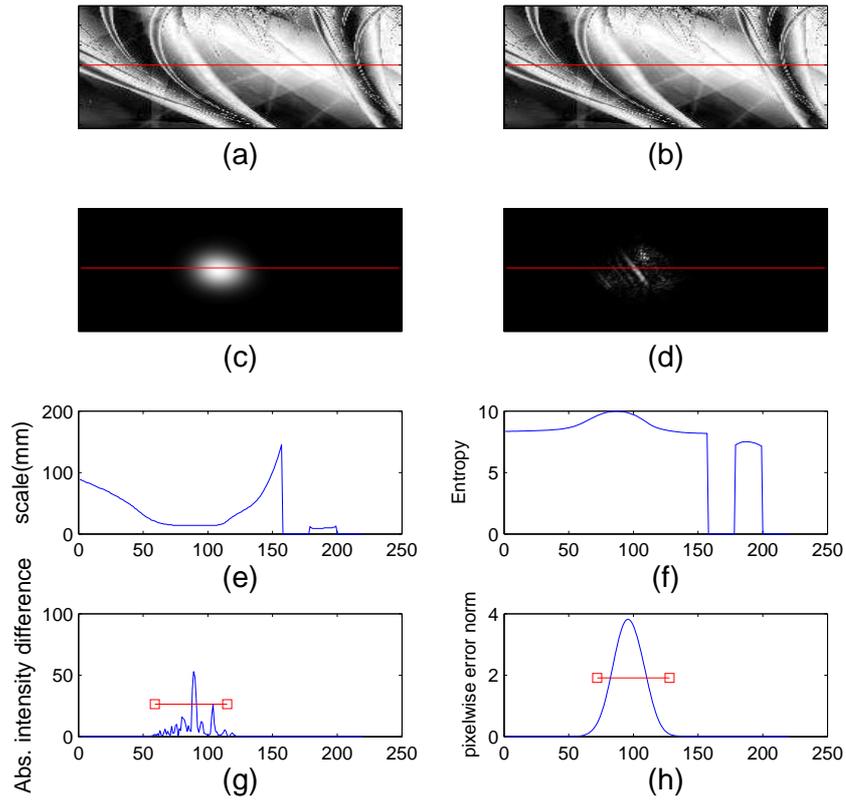
Figure 3.3: a) Target image, b) Floating image , c) Registration error norm, d) Absolute difference between images in a) and b), e) Significant spatial scale peaks, f) Joint entropy at scale peaks, g) Outer scale found(line)on intensity difference along line profile and h) Outer scale(line) on geometric error norm

at different scales. If no peaks were found over the discrete range of scales searched, the scale estimated was set to zero and entropy was not computed (seen as zero in the plot). While points near the deformation estimate scale correctly, distant points find peaks at higher scales. This is dealt with by varying scale of a gaussian low-pass filter as spatial scale increases (discussed in the next subsection). Additionally this also enables us to discriminate local deformations better as will be seen in a following example. The computed scale location and its magnitude are plotted in red (width of two standard deviations of outer scale peak estimated) in (g) and (h) over the intensity difference plot and error norm plot respectively showing how well the estimated location (center of the red line) and scale correspond with the existing mismatch.

3.3.4   Inner and Outer scale

Eq.(3.1) contains three components of scale. While the Parzen window width β is fixed at all spatial scales (α), the degree of smoothing (σ) should be related to the spatial scale (α). Jagersand [71] used different patch sizes at different resolution. In our simulations we varied the scale σ with α as $\sigma = \alpha/3$ i.e., when we are looking at larger regions of interest (α) we are also interested in the variation of coarser features (higher σ).

Example4: Mismatch location and scale identification

The importance of varying σ with α is shown in Fig. (3.4) for an image of size 213×291. (a) is the original image, (b) the deformed image obtained by warping (a) using B-splines, (c) pixel-wise deformation error norm, and (d) shows the positions where peaks in joint entropy were found, where its magnitude corresponds to the scale of deformation. (e) and (f) are respectively the magnitudes of joint entropy
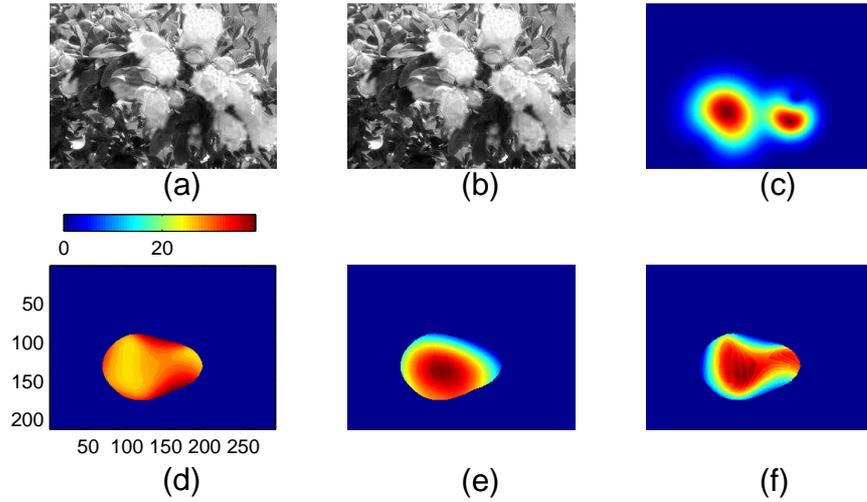
Figure 3.4: a) Target image, b) Floating image, c) geometric error norm, d) peaks in scale, e) joint entropy at peaks without gaussian scale space low pass filter and f) joint entropy at peaks with gaussian scale space filter

obtained by keeping σ fixed, and varying σ with α. In both cases the scale is found to be $\alpha \approx 25$ at the pixel with highest entropy but (f) tracks the deformation error(c) much more closely than (e). The range of scales searched was limited from 6 to 40, i.e., it was assumed that the maximum scale of the deformation was below 40.

3.4 Precision and Accuracy of location and scale

In the next experiment, we deformed an image (dimensions 236×330) using a locally affine geometric deformation [62] at five different locations and four different scales. We applied each deformation independently (20 cases, one location and scale at a time) and the computed the mismatch vector. A typical deformation is shown in Fig. (3.5a) centered at '+' with a deformation spread width [62] = 27 pixel units. The circle drawn has a radius equal to twice this width. The corresponding warped image is Fig. (3.5b) and the absolute intensity difference before and after warp is shown in Fig. (3.5c).
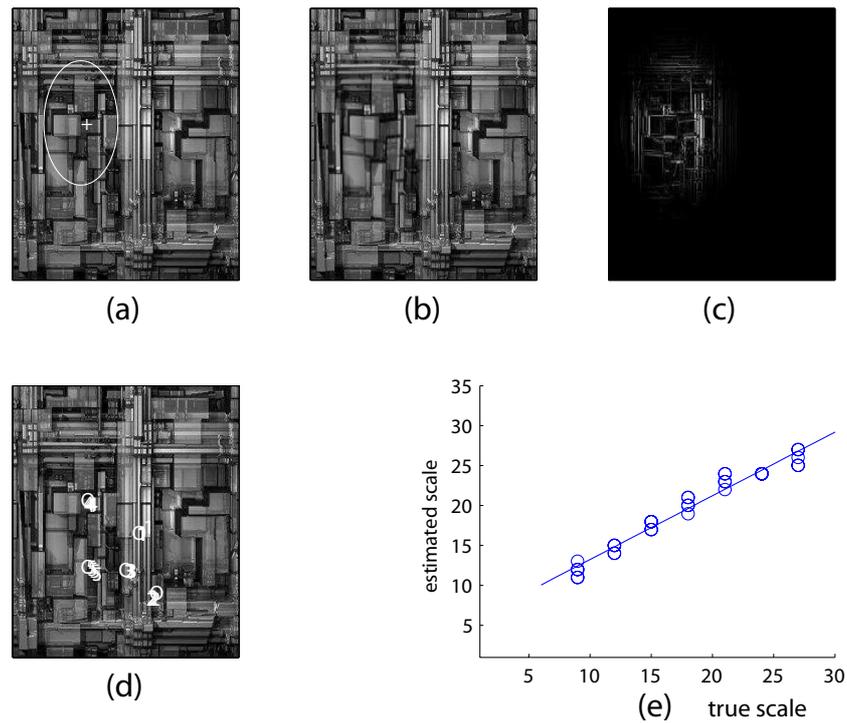
Figure 3.5: a) Typical target Image with deformation center and scale, b) Warped floating image, c) Absolute intensity difference, d) Clusters of identified centers, f) Estimated vs. true deformation scale

The circles in Fig. (3.5d) show the original location of the deformation center and the numbers show the estimated centers for each location. Note that the five numbered markings each are well clustered. Accuracy can be seen visually by the proximity of the markings to the circles. However it possible that regions that are maximally mismatched in intensity may not lie exactly at circle locations, i.e, locations of maximum geometric deformation error norms may not coincide with maximum intensity mismatch so accuracy measurement quantitatively is not possible here but the combined vector precision was found to be small (1.71 pixels) and was computed as the mean deviation of estimated locations from their respective cluster mean vector. Fig. (3.5e) compares the estimated outer scale with the true scale(kernel width) of the deformation. The scale estimation error was found to be 0.87 pixel units.

3.4.1   Example5: Multimodality case

An artificial multi-modality case is considered where we picked an image, (Fig. (3.6(a))), inverted its intensities, affine transformed and warped it to get (b). (c) is got after registering (b) with (a) through an affine transformation as before. The geometrical error norm (only due to warps) is shown in (d). (e) shows the locations where peaks in entropy were found and (f) shows their joint entropy at the respective scale. The highest peak in (f) corresponds well with the location having the largest deformation error in (d).

3.5   Mismatch vector selection

For automatic registration, several mismatch vectors are picked from the image. The number of vectors picked is variable although we found 16-24 vectors at each iteration gave the best trade-off in terms of time and registration accuracy taking
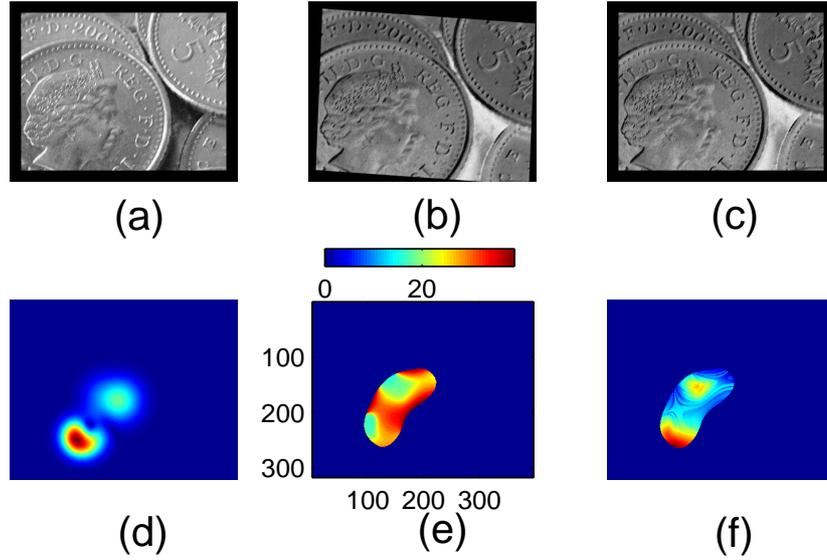
Figure 3.6: a) Target image, b) Affine transformed and warped floating image, c) Floating after affine registration, d) Known warp error norm, e) Estimated peaks in scale, f) Joint entropy at peaks in scale

into consideration both this initialization step and the registration step to be discussed in the next chapter. A clustering procedure was implemented to space them sufficiently apart so that they have non-overlapping supports. The support parameter for registration is set as the scale(α) from the corresponding mismatch vector. Therefore we define the selection of mismatch vectors as follows:

3.5.1 Mismatch vector selection procedure

1: Initialize number of vectors to pick in MV

2: Compute all mismatch locations and scale as

$$(\mathbf{r}_0^*, \alpha^*) = \left\{ \alpha : \frac{\partial^2 H(\alpha; \mathbf{r}_0)}{\partial \alpha^2} < 0 \big|_{\frac{\partial H(\bullet)}{\partial \alpha} = 0} \right\}$$

3: **while** $i = 1 \, to \, MV$ **do**

4:   Pick i'th mismatch vector components$(\mathbf{r}_0, \alpha)$ as

$$(\mathbf{r}_{0i}^*, \alpha_i^*) = argmax_{(\mathbf{r}_0^*, \alpha^*)} H(\mathbf{r}_0^*, \alpha^*)$$

5:    **if** $is\_empty(\mathbf{r}_{0i}^*, \alpha_i^*)$ **then**

6:       break

7:    **end if**

8:    $no\_overlap = \begin{cases} 0, \ D_j < S_j \ \ \forall j = 1 \ldots i-1 \\ \\ 1, \ \ otherwise \end{cases}$

     where $D_j = \|\mathbf{r}_{0i}^* - \mathbf{r}_{0j}^*\|, S_j = \alpha_i^* + \alpha_j^*$ and $j = 1 \ldots i-1$

9:    **if** $no\_overlap$ **then**

10:       Accept $(\mathbf{r}_{0i}^*, \alpha_i^*)$. i = i + 1.

11:    **end if**

12: **end while**

In the actual implementation, the joint entropy was computed at all voxels, and a wide range of α. The locations $\mathbf{r}_0^*$ estimated and the scale α^* were used to initialize a locally affine deformation model (described in the previous chapter). Potentially, other deformation models like B-splines or Radial basis functions could be used as well based on this initialization. This method of initialization followed by registration can be repeated iteratively until no further improvement occurs. The complete registration framework is described in the next chapter.

CHAPTER IV

Automatic multimodality registration

In this chapter we will discuss the complete registration framework using the deformation model and mismatch measure discussed in Chapters 2 and 3 respectively. We begin Section 4.1 by discussing the specific support function used in our model followed by the registration strategy in Section 4.2. For our validation, brain images were warped with known B-spline and compactly supported Radial Basis Function(RBF) based deformations and registration was attempted. An automatic iterative strategy via initialization (finding mismatch vectors) and registration resulted in voxel sized errors.

4.1   Deformation model

Once the mismatch vector is estimated, control point or seed point parameterizations are then be applied at these locations that are relatively more mismatched (using additional scale information available depending on the model initialization). Any deformation model may be used to recover these warps during registration. Here we have demonstrated its effectiveness using a deformation model proposed by us earlier to do non-rigid registration [62] using warps that are locally affine. Functions with local support can be used to control the degree of spatial warp with a scale

parameter that can be fined tuned during registration. We have used compactly supported Wendland functions [17] to define the locality of the deformation. Specifically we used $\psi(\tilde{r}) = (1 - \tilde{r})^4_+(4\tilde{r} + 1)$ for $\tilde{r} = r/a$ where r is the distance from the deformation center and a is the spatial scale. This function was found to work well for estimated scale to correct deformations while providing loose bounds under trivial conditions for an inverse to always exist (See appendix).

4.1.1 Registration deformation model invertibility conditions

In our previous work[62] we proved the existence of simple conditions that allowed for loose bounds in the deformation that always guaranteed an inverse. However the bounds were derived for a gaussian support function. Following similar steps, conditions can be derived for any support function. In this paper we used a the compactly supported function $\psi(\tilde{r}) = (1-\tilde{r})^4_+(4\tilde{r}+1)$ and in a similar manner found that the transformation(T) for any seed point always has an inverse as long as

$$\|\mathbf{t}\| < 0.4741\,b \tag{4.1}$$

and

$$0 < a < \lessapprox 4.2493, \tag{4.2}$$

where $\mathbf{t}$ is the translation vector and $a = max(a_x, a_y, a_z)$, the largest of the three anisotropic scales in the x, y and z direction(or only x and y in 2D), and b is the support of the piecewise polynomial function for the translation part at each seed point. These bounds in practice were found to be loose and we rarely experienced any folding in our simulations.

Each deformation is specified via a seed point location, scale and its associated set of affine parameters. Registration for successive seed points is performed through a composition of the transformation over all previously optimized parameters. E.g.

if we have N seed points, the final transformation is

$$T(\mathbf{x}) = (T_N \circ T_{N-1} \ldots T_2 \circ T_1)(\mathbf{x}), \qquad (4.3)$$

where each seed point 'i' is associated with a transformation T_i. The individual transformation parameters are computed for each seed point. Global Normalized Mutual Information (NMI) [44] was used as the objective function and a simplex based optimizer was used to optimize over seed point parameters, i.e. for 2D, two translation parameters (t_x, t_y), rotation angle (θ), two anisotropic scale parameters (a_x, a_y), two center coordinates (C_x, C_y) and a support parameter (s) where (C_x, C_y) and s were set equal mismatch vector components $\mathbf{r}_0$ and α respectively.

4.2 Registration framework

In the previous chapter we discussed mismatch vectors and how they were computed. This is the initialization step. The initializer returns a set of mismatch vectors sorted in descending order of joint entropy at the respective scale where entropy was peaked. To improve computational speed the joint entropy over scale was computed only at edge pixels along the target image. Edges were computed using a Laplacian Of Gaussian(LoG) filter with a threshold set to 0.03 and kernel width$(\sigma) = 3.5$ for the byte ranged images. This reduced the number of pixels over which entropy (through outer scale) was calculated ten fold. The components of the mismatch vector and affine parameters $(t_x, t_y, \theta, a_x, a_y)$ that define the model parameters are optimized one seed point at a time. The support parameter was independently varied for translation and rotation-scale, i.e., two different supports although initialized to the same value (as specified by the mismatch vector component α, i.e., outer scale) were optimized individually. This is different from our previous work [62] where a single support parameter was used. Although this increases the number of parameters by

one it was found to give a much greater degree of flexibility while still preserving invertibility, compact support and smoothness of the transformation.

Several seed points may be optimized simultaneously in Eq. (4.3) if their supports do not overlap, allowing a certain degree of parallelization. Specifically for the case of our registration experiments and deformation model used, we found that using 25 seed points at a time to do registration worked effectively although the parallelization option was not exploited. After registration, the floating image is warped onto the target and the initialization procedure is repeated. This is performed iteratively until $\triangle$ NMI $<$ *thresh*. Some seed points picked may not lead to improvement in cost ($\triangle$ NMI $< 3e^{-4}$). Mismatch vector selection may further be influenced by a distance filter that restricts it to locations (and scale) with high joint entropy and non-overlapping support with these seed points in subsequent iterations.

4.3   Algorithm

A is the target image. After affine registration with the target, the interpolated floating image(B) is calculated from the pullback map.

1: Set $\hat{T}$ to identity.

2: Compute NMI$(A(\bullet), B(\hat{T}(\bullet)))$. $\triangle NMI = thresh + \epsilon$. Set i $= 1$.

3: **while** $\triangle$ NMI $>$ *thresh* **do**

4:     Pick M mismatch vectors sorted in descending order of joint entropy

5:     **for** $k = 1 \, to \, M$ **do**

6:         $\hat{T}_{i,k} = argmax_{\ T_{i,k}}$ NMI$(A(\bullet), B((T_{i,k} \circ \hat{T})(\bullet)))$

7:         Update: $\hat{T} = \hat{T}_{i,k} \circ \hat{T}$

8:     **end for**

9:     Compute NMI$(A(\bullet), B(\hat{T}(\bullet)))$ and $\triangle$ NMI

Figure 4.1: Automatic multimodality registration

10:    $i = i + 1$

11: **end while**

Note that $\hat{T}(\bullet)$ always stores the most current deformation map in the form of a 2D/3D array and is not computed from the previously estimated transformation parameters each time, speeding up computation. Fig. (4.1) shows the flowchart of the registration framework.

4.3.1 Implementation details

Table 4.1: Parameter specifications for simulations

Procedure	*Step*	*Parameter selection*
Preprocessing	Laplacian of Gaussian filter	kernel width = 3.5
		threshold = 0.03
Initialization	Gaussian Histogram weighting function	Outer scale = σ
	Locations search range	Image pixel locations
	Outer scale search range	6-33 pixels
	Mismatch vector distance filter	Non-overlapping supports
Registration	Support function	$\psi(\tilde{r}) = (1-\tilde{r})^4_+(4\tilde{r}+1)$
	Seed points/iteration	25
	Termination criteria	$\triangle$ NMI < 0.01

4.4 Results

4.4.1 Tumor growth modeling

The nature of tumor growth is emulated for an axial T1 weighted real 256×256 image containing a real tumor using a simplistic model of a pure geometrical deformation with no intensity variation (see Fig. (4.2)) and no physical constraints to prevent warping at the boundary of the head. Fig. (4.2(a)) shows the target image with contours delineating tumor and ventricles, Fig. (4.2(b)) shows the target image and a 5×5 grid of control points displaced randomly (N(0,3.5) in X and Y directions. A compactly supported (support = 40mm) radial basis function ($\psi(\tilde{r}) = (1-\tilde{r})^6_+(\frac{35}{3}\tilde{r}^2 + 6\tilde{r} + 1)$) was used to warp the target image, creating a warped image Fig. (4.2(d)) using the pullback map. Fig. (4.2(d)) also shows the original and new locations of the contours superposed on the warped image and Fig. (4.2(c)) shows the deformation map. Fig. (4.2(e)) is the outer scale image where joint entropy peaks through scale were found, intensity(α) mapped to color and Fig. (4.2(f)) shows joint entropy at these locations. Both Fig. (4.2(e)) and Fig. (4.2(f)) were estimates at the end of the first iteration (of initialization). The mismatch locations and scale
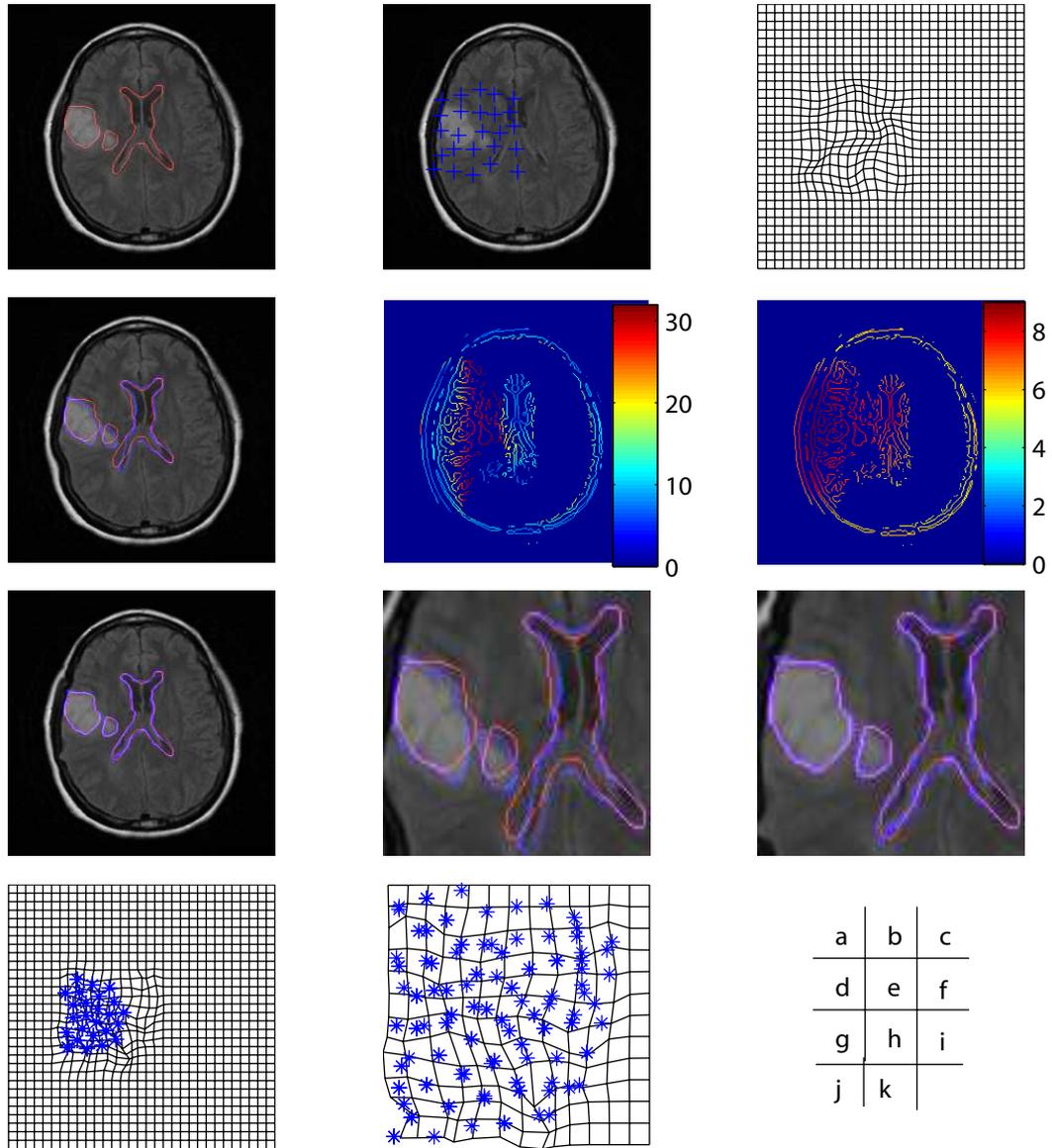
Figure 4.2: a) Target image with ventricles delineated, b) Target image with control points, c) Applied known deformation, d) Warped image with outlines before and after warp showing mismatch e) Locations with peaks in scale, scale mapped to color f) Joint entropy at peaks in scale, g) Floating image on target after registration, h) magnified outline before registration near ventricles, i) After registration, j) Seed point locations on estimated deformation, k) magnified seed point locations on estimated deformation

seem to agree with true deformations. Initialization and registration were iterated, and Fig. (4.2(g)) shows the result after only three iterations (Normally iterated until $\triangle$ NMI $< thresh$). The average original error norm at warped locations was 1.6616mm and the average final error norm after three iterations was 0.8404mm. Two iterations were enough to get voxel sized errors (1.0619mm). Fig. (4.2(g)) and Fig. (4.2(h)) show the contour mismatch before and after registration close to brain structures, and Fig. (4.2(j)) and Fig. (4.2(k)) show the locations of seed points on the registration deformation map for the whole image and in the vicinity of the tumor respectively. Note how well clustered the seed points are in accordance with the applied deformation error (as opposed to being scattered over the entire image). Although seventy five seed points were used in all (25 per iteration), each optimization was performed over a strictly local region of the image consisting of only 9 parameters making it very fast. Picking anywhere between 16 and 32 seed points per iteration seemed to work well in terms of time taken in our registration experiments for comparable registration accuracy. See Table 4.3. We used 25 seed points in all our simulations. It is possible that some redundant parameterization maybe performed in later iterations (see head boundaries in Fig. (4.2(e))) that do not get picked until much later due to their lower joint entropy (see Fig. (4.2(f))). These seed points will not lead to improvement in the objective function and registration will automatically terminate($\triangle$ NMI < 0.01).

4.4.2 Registration of brain structures

In this example we demonstrate registration of brain structures for a single axial slice of a 181×217 T1 weighted MR image (in-plane spacing $1mm$) from Brainweb [58]. The slice (see Fig (4.4)top row left) is warped using a 11×11 grid B-spline
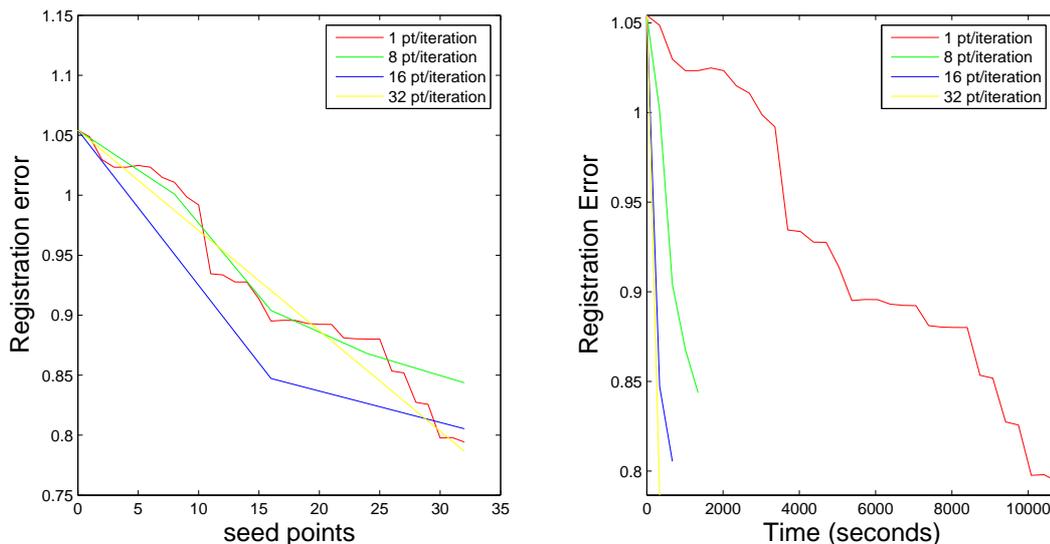
Figure 4.3: Left: Registration error in mm vs. seed points per iteration. Right: Registration error in mm vs. time in seconds

deformation (top row center) where only the 16 knots at the center marked distinctly are displaced randomly (N(0,4)). Eight synthetically warped images were generated based on independent deformations applied each time. The target image was delineated based on important structures in the brain namely the lenticular nucleus, ventricles and the caudate nucleus shown in the same figure (top right). On warping, each of their contours are displaced and original contour displacement errors are computed. The bottom row in the same figure shows one typical example (of the eight warped slices). The original contour plotted on the target image (bottom left), original and warped contour superposed on the target image (bottom center) and the original and final contours on the target (right center) are shown.

For each of the warped images, an initialization routine was used to find the mismatch vectors followed by registration using 25 seed points each time. In order to compare registration performance all images were iterated exactly 4 times although practically registration could be performed until the improvement is above a threshold ($\triangle$ NMI < 0.01 seemed to work well). The registration accuracy was found to
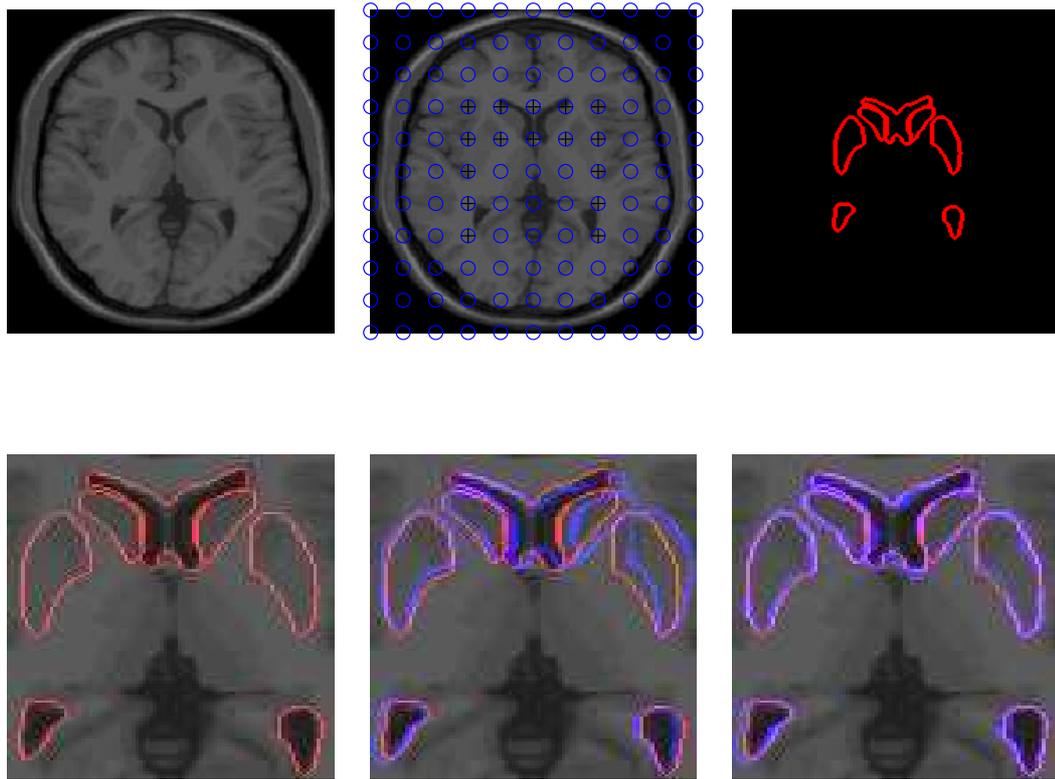
Figure 4.4: a) Target image, b) Target image with 16 control points displaced marked ⊕, c) Ventricles, caudate and the lenticular nucleus delineated, d) Outlines on magnified target, e) Outline mismatch before registration, f) After registration

be close to voxel size (see Table 4.2 showing OCDE - original contour displacement error and FCDE - final contour displacement errors both in mm) on completion.

Finally Fig. (4.5) shows the original contours of the brain structures plotted on a 2D histogram of all seed point locations picked(100 in each image, 800 in all) for registration. The histogram was blurred with a narrow windowed gaussian for easier visual inspection. Also shown('+') are the locations of knot points on the B-spline grid that were displaced. Observe that seed points tend to be picked at locations close to where the original deformations were applied.
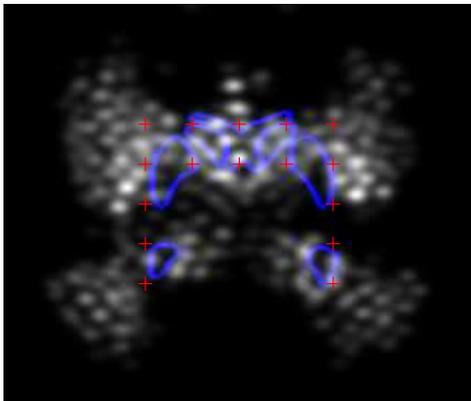
Figure 4.5: Histogram of seed point locations for all eight experiments on the contours of structures. Original locations of B-spline control points displaced are given by a +.

Table 4.2: Contour registration errors. Top row shows original average contour displacement errors for each sample and bottom row shows final average contour displacement errors

Samples	1	2	3	4	5	6	7	8
OCDE	2.5192	2.2425	1.9142	2.5769	2.7295	1.6745	2.5227	2.8405
FCDE	1.2348	1.1329	1.3864	1.5513	1.3915	1.1463	1.3112	1.1725

CHAPTER V

Conclusion

5.1 Contributions

We have demonstrated and tested a new local nonlinear transformation for multi-modality image registration. This transformation is local, smooth and has affine behavior near the region of interest. The parameters controlling the support of this transform can be initialized and changed (by the optimizer) during the course of the registration to match the level of scale of the induced deformation. Furthermore, the transformation has a closed form and there is no need to integrate the velocity vector over time as in the case of methods using ODEs, making it very fast. Since each region is optimized one at a time, only eight parameters are used in the optimization which makes it very fast. Although we can always guarantee that folding does not occur, finding a direct inverse is not straightforward. A numerical inverse could be found finally using the optimized parameters after registration if required. In Chapter 2, I have also derived the expression for the Jacobian that can be used to measure local volume change.

This dissertation also focussed on identifying mismatch in images through location and scale. Mismatch vectors are picked based on peaks in joint entropy through outer scale computed at every pixel location, and location-scale pairs with high

entropy were used to parameterize a geometric deformation model. This type of initialization allowed for finer correction and parameterization only in regions that were significantly warped instead of a uniform grid placement with control point placement in regions with no deformation or low mismatch.

Combining the proposed parameterization method based on mismatch with the deformation model we have demonstrated a completely automatic registration algorithm. Results show that they can be used to reliably register structures in the brain to within voxel sized errors.

5.2 Transformation model selection

Different local support functions can be used. In Chapter 2, I used the Gaussian function for the brain registration examples, while in Chapter 4, I used a local support function. In addition to their different profiles, ability to use different support functions give some flexibility in smoothness and invertibility bounds. Their support parameter can also be fine tuned during optimization.

5.3 Future work

5.3.1 Time considerations

The deformation model has a closed form and consists of only nine parameters in 2D. Computation time for the transformation, intensity evaluation and computation of MI for a given set of parameters was of the order of a few hundred milliseconds on a 3.6 GHz, Pentium 4 CPU with 2 Gb memory running MATLAB (for images sizes 256×256). Each seed point initialization led to an optimization cycle consisting of about 30-40 function evaluations. The number of seed points depends on how misaligned the images are. Typically we needed around 100 seed points to achieve

voxel sized errors to register brain structures. Since the registration is iterative with the initialization step, where mismatch vectors are computed, seed points could be picked in a non-overlapping fashion (within the registration step) each time, i.e., deformations are allowed to overlap in general. This way several optimizations can be performed at once.

The initialization step was computationally more expensive. Finding joint entropy throughout the image and through a range of spatial extents took several minutes to compute for the same machine configuration as above for 2D images. Due to its coarse grained nature, mismatch can be computed over image sub-blocks independently on several machines at a time. Parallelization of this routine by running on 16 machines simultaneously (Pentium 4, 3.06 GHz, 4 Gb memory each) took approximately 1.5 minutes (image size 256×256). Potential speed ups can be achieved by converting the code to C discussed next.

5.3.2 Extensions to 3D

All simulations in this dissertation were done for 2D images. Conceptually the methods can be directly extended to 3D. In order to register 3D data sets, the implementation would have to be done in C.

5.3.3 Comparison with B-splines and radial basis functions

A comparison of registration accuracy and time taken using the transformation model proposed here for different support functions with B-splines and radial basis functions would be useful to decide the best transformation model for the targeted application.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Lisa G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.

[2] Calvin R. Maurer and Michael Fitzpatrick. A review of medical image registration. In *Interactive Image-guided Neurosurgery*, pages 17–44, 1993.

[3] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[4] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.

[5] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[6] J Zhang and A Rangarajan. Multimodality image registration using an extensible information metric and high dimensional histogramming. In *Information Processing in Medical Imaging (IPMI)*, 2005.

[7] K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. *IEEE Symposium on Biomedical Imaging (ISBI)*, pages 908–911, 2004.

[8] Carole Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic, Roy Schestowitz, and Chris Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In *IPMI*, pages 1–14, 2005.

[9] H. Neemuchwala and A. O. Hero. *Multi-Sensor Image Fusion and its Applications*, chapter Entropic Graphs for Registration, pages 185–235. Marcel Dekker, 2005.

[10] Transformation functions for image registration. *Image Fusion Systems Research*.

[11] T. J. Rivlin. *Least-squares approximation*, pages 48–65. Blaisdell publishing, 1969.

[12] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, 1987.

[13] M. D. Buhmann. *Radial basis functions*, pages 1–38. Cambridge Unirersity Press, 2000.

[14] Gustavo K. Rohde, Akram Aldroubi, and Benoit M. Dawant. The adaptive bases algorithm for intensity based nonrigid image registration. *IEEE Trans. Med. Imaging*, 22(11):1470–1479, 2003.

[15] M Fornefett, K Rohr, and H S Stiehl. Radial basis functions with compact support for elastic registration of medical images. In *WBIR'99*, pages 173–185, 1999.

[16] Mark P. Wachowiak, Xiaogang Wang, Aaron Fenster, and Terry M. Peters. Compact support radial basis functions for soft tissue deformation. In *ISBI*, pages 1259–1262, 2004.

[17] H. Wendland. Piecewise polynomial positive definite and compactly supported radial basis functions of minimal degree. *AICM*, 4:389–396, 1995.

[18] Z. Wu. Multivariate compactly supported positive definite radial functions. *Adv. Comput. Math.*, 4, 1995.

[19] S. Lee, K. Y. Chwa G. Wolberg, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Trans. Vis. and Comp. graphics*, 2, 1996.

[20] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Trans. Vis. and Comp. graphics*, 3(3):228–244, 1997.

[21] 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. Imaging*, 18(8):712–721, 1999.

[22] Yan Kybic and Michael Unser. Fast parametric elastic image registration. *IEEE Trans. Medical Imaging*, 12(11), 2003.

[23] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10):1435–1447, 1996.

[24] G. E. Christensen. Volumetric transformation of brain anatomy. *IEEE Trans. Med. Imaging*, 16(6):864–877, 1997.

[25] E. D'Agostino, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. In *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, pages 541–548, London, UK, 2002. Springer-Verlag.

[26] Christophe Chefd'Hotel, Gerardo Hermosillo, and Olivier Faugeras. A variational approach to multi-modal image matching. In *VLSM '01: Proceedings of the IEEE Workshop on Variational and Level Set Methods (VLSM'01)*, page 21, Washington, DC, USA, 2001. IEEE Computer Society.

[27] F. Maes, A. Collignon, D. Vandermuelen, G. Marchal, and P. Suetens. Multimodal volume registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16:187–198, 1997.

[28] A. Collignon, F. Maes, D. Delaere, D. Vandermuelen, P. Suetens, and G. Marchal. Automated multimodality medical image registration using information theory. pages 263–274, 1995.

[29] Josien P. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Mutual information matching and interpolation artifacts. volume 3661, pages 56–65. SPIE, 1999.

[30] Jeffrey Tsao. Interpolation artifacts in multimodality image registratino based on maximization of mutual information. *IEEE Trans. Med. Imag.*, 22(7):854–864, 2003.

[31] P A Viola. Alignment by maximization of mutual information. *Ph.D. thesis, Massachussetts Institute of Technology.*, 1995.

[32] Pratt W. K. In *Digital Image Processing, second edition*. Wiley, 1991.

[33] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. The correlation ratio as a new similarity measure for multimodal image registration. *Lecture Notes in Computer Science*, 1496:1115–1124, 1998.

[34] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. 11:71–80, 2000.

[35] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Mutual information based registration of medical images: a survey. *IEEE Trans. Medical Imaging*, 22(8):986–1004, 2003.

[36] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. Mri-pet registration with automated algorithm. *Journal of computer assisted tomography*, 17(4):536–546, 1993.

[37] R. P. Woods, S. R. Cherry, and J. C. Mazziotta. Rapid automated algorithm for aligning and reslicing pet images. *Journal of computer assisted tomography*, 16(4):620–633, 1992.

[38] D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff. A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics. *Lecture notes in compute science*, 687:182–196, 1993.

[39] D. L. G. Hill, C. Studholme, and D. J. Hawkes. Voxel similarity measures for automated image registration. *Visualization in biomedical computing, Proc. of SPIE*, 2359:205–216, 1994.

[40] A. Collignon, D. Vandermuelen, P. Suetens, and G. Marchal. 3d multi-modality medical image registration using feature space clustering. *Computer Vision, Virtual Reality and Robotics in Medicine*, 905:195–204, 1995.

[41] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Multiresolution voxel similarity measures for mr-pet registration. In *IPMI*, pages 287–298, 1995.

[42] Paul Viola and William M. Wells. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, 1997.

[43] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information, 1996.

[44] C. Studholme, D. Hill, and D. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.

[45] Neemuchwala H, Hero A, Carson P, and Meyer C. R. Local feature matching using entropic graphs. *IEEE Symposium on Biomedical Imaging (ISBI)*, pages 704–707, 2004.

[46] Hero A. O., Bing M., Michael O. J. J., and Gorman J. Applications of entropic spanning graphs. *IEEE signal processing magazine*, pages 85–95, 2002.

[47] Maes F., Vandermeulen D., and Suetens P. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386, 1999.

[48] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Med. Image. Anal.*, 1:195–206, 1997.

[49] C. R. Meyer, J. L. Boes, B. Kim, and P. H. Bland. Probabilistic brain atlas construction: Thin plate spline warping via maximization of mutual information. In *MICCAI*, 1999.

[50] C. R. Meyer, J. L. Boes, B. Kim, and P. H. Bland. Warping normal patients onto the icbm atlas by maximizing mi. In *Proc. 4th Int. Conf. of functional mapping of the human brain*, volume 7, 1998.

[51] S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Medical Imaging*, 9(8), 2000.

[52] C. Chef'd Hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. In *ISBE*, pages 753–756, 2002.

[53] Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Polyrigid and polyaffine transformations: A new class of diffeomorphisms for locally rigid or affine registration. In *MICCAI (2)*, pages 829–837, 2003.

[54] Stephen Marsland and Carole J. Twining. Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images. *IEEE Trans. Medical Imaging*, 23(8), 2004.

[55] Yongchoel Choi and Seungyong Lee. Injectivity conditions of 2D and 3D uniform cubic B-Spline functions. *Graphical Models*, 62(6).

[56] H. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control*, 37:332–341, 1992.

[57] Hyunjin Park, Peyton H. Bland, Kristy K. Brock, and Charles R. Meyer. Adaptive registration using local information measures. *Medical Image Analysis*, 8(4):465–473, 2004.

[58] C. A. Cocosco, V. Kollokian, R. K. S. Kwan, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4), 1997.

[59] Daniel J. Rubins, Karen Ambach, Arthur W. Toga, William P. Melega, and Simon R. Cherry. Development of a digital brain atlas of the vervet monkey. *BrainPET*, 4, 1999.

[60] Richard Williamson, Richard Crowell, and Hale Trotter. Caluclus of vector functions. *Prentice-Hall, Inc*, pages 290–297, 1962.

[61] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.

[62] Ramkrishnan Narayanan, Jeffrey A Fessler, Hyunjin Park, and Charles R Meyer. Diffeomorphic nonlinear transformations: A local parametric approach for image registration. In *IPMI*, pages 174–185, 2005.

[63] Xavier Pennec Vincent Arsigny, Olivier Commowick and Nicholas Ayache. A log-euclidean polyaffine framework for locally rigid or affine registration. In *WBIR'06*, pages 120–127, 2006.

[64] R Bajcsy and S Kovacic. Multiresolution elastic matching. volume 46, pages 1–21, 1989.

[65] Torsten Rohlfing and Calvin R. Maurer Jr. Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint. In *MICCAI*, pages 111–119, 2001.

[66] Gustavo K. Rohde, Akram Aldroubi, and Benoit M. Dawant. Adaptive free-form deformation for interpatient medical image registration. volume 4322, pages 1578–1587. SPIE, 2001.

[67] Julia A. Schnabel, Daniel Rueckert, Marcel Quist, Jane M. Blackall, Andy D. Castellano-Smith, Thomas Hartkens, Graeme P. Penney, Walter A. Hall, Haiying Liu, Charles L. Truwit, Frans A. Gerritsen, Derek L. G. Hill, and David J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–581, London, UK, 2001. Springer-Verlag.

[68] Hyunjin Park and Charles Meyer. Grid refinement in adaptive Non-Rigid registration. In *MICCAI*, pages 796–803, 2003.

[69] Jan J. Koenderink and Andrea J. Van Doorn. The structure of locally orderless images. *International Journal of Computer Vision*, 31(2/3):159–168, 1999.

[70] S Gilles. Robust description and matching of images. *Ph.D. thesis, University of Oxford.*, 1998.

[71] Martin Jagersand. Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach.

[72] Timor Kadir and Michael Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[73] R. Narayanan, J. A. Fessler, B. Ma, and C. R. Meyer. Local mismatch location and spatial scale detection in image registration. SPIE, Medical Imaging, to appear, 2007.