A DEFORMATION MODEL FOR IMAGES

by

Sining Chen

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Valen E. Johnson, Supervisor

_____
Dr. Merlise Clyde

_____
Dr. Dalene Stangl

_____
Dr. Timothy Turkington

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

2002

ABSTRACT

(Statistics in Imaging)

A DEFORMATION MODEL FOR IMAGES

by

Sining Chen

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Valen E. Johnson, Supervisor

_____
Dr. Merlise Clyde

_____
Dr. Dalene Stangl

_____
Dr. Timothy Turkington

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
Institute of Statistics and Decision Sciences in the Graduate School of
Duke University

2002

Abstract

Large quantities of medical images are acquired daily at nearly every medical center in the United States, but statistical models and associated softwares that would facilitate the automated analyses of these images are lacking. The goal of this research is to develop methodology that will make such automated image analysis possible.

The methodology proposed is based on an atlas-based deformation model which finds a one-on-one mapping from the atlas image to a target image in the same image class. The deformation is carried out by generalized landmarks called "facets". Knowledge about the atlas image is thus transferred to the target image through facets.

A large number of facets are placed in the volume of the atlas (often on a lattice). Each of them is then located in the target image. The model for the new location has two components: a Markov random field prior with pair-wise difference on a nearest neighborhood system, and a likelihood component based on the agreement of features in the atlas and target image. A new measure of feature difference was introduced that is robust under a variety of conditions.

The iterative conditional modes (ICM) algorithm is used to obtain the maximum *a posterior* estimate of facet locations.

The model was used to automatically segment magnetic resonance mouse brain images. It was also applied to inter-subject registration of human brain images. Both qualitative and quantitative evaluation of the results are presented.

Acknowledgements

I deeply thank my advisor Valen Johnson, whom I can always count on for support and advice. You will continue to be my mentor.

I would also like to thank Dalene and Merlise, for giving me encouragements and putting up with short notices.

Tim Turkington, who is always willing to help me with the subject matter.

Special thanks to my friend Sun, Liang, who spent so much of her time making sure that my pointers point to the right places.

I am grateful to the professors and staff at ISDS for making it a true Bayesian wonderland.

My dear friends at ISDS, I thank you for the wonderful friendship I could not have imagined.

Jin, Yun, and my parents, I am humbled by your love.

Grandma, I dedicate this thesis to you, the strongest person I ever know.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

The past decades have seen a boom in medical imaging technologies. Large quantities of images are acquired daily at nearly every medical center in the United States. Mean-while, non-medical related imaging, such as satellite imaging, radar signal processing, remote sensing, has produced valuable image data as well.

Consequently, much research effort has been devoted to facilitating the automated analyses of these images. Electrical engineers, computer scientists, physicists and statisticians are among the many who make major contributions to the advance of technologies in image processing and analyses.

Common problems of image analyses include: image restoration, segmentation, and registration, just to name a few. Haralick and Shapiro (1991) gave the definition of these terms:

Restoration is "a process by which a degraded image is restored, as clearly or as best as possible, to its ideal condition". From a Bayesian statistical point of view, restoration is the process of making inference on the unobserved ideal image given observed image data based on assumptions about the nature of degradation.

Registration is the process of bringing two similar images (images of like geometries) into spatial alignment, such that "corresponding points of the imaged scene appear in the same position on the registered images".

Segmentation is the process of segmenting an image into meaningful regions. Each region may represent an anatomical part of the subject.

Most of the literature in imaging cited in this thesis deals with the above problems.

1.2 Image Registration in General

Image registration is the primary problem this thesis aims to solve. In this section we review a sample of registration methods. There is a rich literature on image registration methods. These methods can be classified differently from several perspectives. Maintz and Viergever gave a comprehensive review in 1998.

We explain the following registration methods in the context of medical imaging, especially brain imaging, for it is where these methods have the most application in real practice. However, readers should realize that the methods may not be limited to brain and medical applications.

We also give the definition of "resolution cell", "pixel" and "voxel" in a digital image (Haralick and Shapiro, 1991) to facilitate understanding of the registration methods.

- resolution cell: the smallest most elementary areal constituent having an associated image intensity in a digital image;
- pixel, or picture element: a *resolution cell* in two dimensional images or a spatial position along with the image intensity or vector of image values associated with the spatial position;
- voxel, or volume element: same as *pixel* but in three dimensions.

With regard to the modality from which the images are acquired, registration methods can be classified into mono-modality registration and cross-modality registration.

Images of various modalities are acquired using different apparati with different objectives. One category of modalities is targeted toward studying the anatomy of the object. These include X-ray, CT (computed tomography) and MRI (magnetic resonance imaging). Specifically, we briefly introduce the mechanism of MRI facility. A MRI scanner applies a strong magnetic field to the object, causing the atomic nuclei in the matter to spin along axes parallel to the direction of the magnetic field. Then a radiofrequency (RF) pulse is broadcast in a direction perpendicular to the magnetization direction. RF pulses are applied through coils. Different coils can be used to image different body parts, such as head, neck, knee. Those nuclei precessing (spinning) about their axes at the same frequency as the imposed radiofrequency are affected (this is called "resonance"). Their precession axes will first deviate from the magnetization direction and then come back within a specific length of time called the "relaxation time." In medical applications the radiofrequency is set to the precession frequency of hydrogen nuclei. Hydrogen nuclei from different tissue types will have different "relaxation times", thus giving out different radio pulses. The MRI scanner captures these pulses and records them in an image, and we are able to distinguish between tissue types according to the image intensity.

Another general category of modalities targets the functioning of organs. These modalities include fMRI(functional magnetic resonance imaging), PET (positron emission tomography) and SPECT (single photon emission computed tomography). During fMRI, patients are injected with a contrast-enhancing agent so that MR scanners can capture the amount of oxygen in blood vessels in a very short time. PET and SPECT start by injecting the subject with radioactive substance. In PET, the

radioactive substance gives off positrons which immediately collide with electrons in the tissue then emit double gamma rays. In SPECT, the radioactive substance gives off single photons. In brain functional imaging, areas of brain activity gain a distinct level of intensity in the resulting image due to increased blood flow, which enables us to study how brains function. SPECT images are less expensive to obtain than PET, but also sacrifice some resolution and sensitivity.

Mono-modality registration refers to the registration of images acquired from the same modality. Cross-modality registration refers to the registration of images from different modalities and is usually used to integrate information from different images of the same object. One of the most common applications is matching functional brain images to a corresponding anatomic image. This allows us to better interpret the functional results and to gain insights not apparent when examining the modalities independently. Compared to mono-modality registration there are more difficulties to overcome in cross-modality registration. The assumption many mono-modality registration methods make, that the same tissues have similar intensities, is violated, thus rejecting the use of related methods. In some cases distortion may also be present in one of the modalities.

Registration methods can be extrinsic or intrinsic. Extrinsic methods are based on accurately detectable foreign objects introduced into the image space, whereas intrinsic methods are based on the image generated by the subject alone. Many intrinsic methods require identification of landmarks, which is a strenuous and complex procedure. Thus extrinsic methods are comparatively easy, fast. However, provisions must be made in pre-acquisition phase, and marker objects are usually invasive in nature, making them clinically impractical.

Registration methods can also be classified as intra-subject and inter-subject. Intra-subject registration is often used to correct for motion-induced misalignment

when several scans (not necessarily taken in the same modality) of the same subject are taken at different times. In brain imaging, intra-subject registration requires only rigid-body transformation. Inter-subject registration non-rigidly transforms the images such that their corresponding anatomical parts have the same shape, position and orientation.

Since this thesis places more emphasis on the mathematical transformation used in registration, we adopt the classification by the type of spatial transformation. Since particular types of transformation are used for particular purposes, this classification is by no means orthogonal to the previous classifications. The next section presents a sample of the pre-existing methods, organized by the type of transformation used.

1.3 Rigid-body Transformation

Transformations consisting only of rotations and translations are called rigid-body transformations. They are often applied when registering images of the same subject, even if the images are not acquired in the same modality. For example in intra-subject registration of MRI to MRI, PET to PET, or PET to MRI, usually rigid-body transformation is sufficient.

Rigid-body transformations in three-dimension can be fully described with a six-parameter model as follows: any point with coordinates (x_1, y_1, z_1) in the original space will occupy new coordinates (x_2, y_2, z_2) after transformation.

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} \quad R \quad \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T \tag{1.1}$$

Rotation is performed first, denoted by rotation matrix R. R can be decomposed into rotations around the three axes, $R = R_x R_y R_z$, where left multiplication of R_x

rotates the image around x-axis by angle θ_x and so on :

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{pmatrix}, \tag{1.2}$$

$$R_y = \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix}, \tag{1.3}$$

$$R_z = \begin{pmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.4}$$

Translation $T = (t_x, t_y, t_z)'$ is performed after rotation, where t_x, t_y, t_z are translation in the x, y, z directions, respectively.

1.3.1 Mono-modality Registration

Many mono-modality algorithms adopted image motion estimation techniques from the computer vision literature. These algorithms are based on the "intensity conservation assumption", the belief that the same point on an object should display the same intensity when imaged at different times (or under different conditions). Equivalently, with respect to digital images, it is assumed that any pair of "corresponding" voxels in two image volumes have equal intensity.

Ashburner and Friston (1993) used a straightforward approach based on this assumption: they found the parameter values in (1.1) that minimize the sum of squares for intensities of corresponding voxels. The authors also offered a more richly parameterized transformation, which will be discussed in the section of deformable models.

An approach suggested by Nestares and Heeger (2000) uses a gradient-based displacement estimation. Under the intensity conservation assumption, the intensity at coordinate (x, y, z) in image 1 is shifted by $\boldsymbol{d} = (d_x, d_y, d_z)$ to coordinate

$(x + d_x, y + d_y, z + d_z)$ in image 2, *i.e.*:

$$f_1(x, y, z) = f_2(x + d_x, y + d_y, z + d_z) \tag{1.5}$$

$f_1(.)$ and $f_2(.)$ are intensity functions of image 1 and image 2, respectively. The model further assumes that intensity functions can be well approximated by a first-order Taylor expansion, which leads to

$$\begin{aligned} f_1(x, y, z) - f_2(x, y, z) &= f_2(x + d_x, y + d_y, z + d_z) - f_2(x, y, z) && (1.6) \\ &\approx d_x * \frac{\partial f_2}{\partial x}(x, y, z) + d_y * \frac{\partial f_2}{\partial y}(x, y, z) + \\ &\quad + d_z * \frac{\partial f_2}{\partial z}(x, y, z) && (1.7) \\ &= \boldsymbol{d} \cdot \nabla f_2 && (1.8) \end{aligned}$$

After re-parameterizing (1.1), the displacement $\boldsymbol{d}$ can be written as a function of initial position $\boldsymbol{x}$ as follows,

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{x}) &= \boldsymbol{\omega} \times \boldsymbol{x} + \boldsymbol{t} && (1.9) \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 & z & -y \\ 0 & 1 & 0 & -z & 0 & x \\ 0 & 0 & 1 & y & -x & 0 \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} && (1.10) \end{aligned}$$

For each voxel, substitute the above for $\boldsymbol{d}$ in (1.8). Then for two images with N voxels each, we have an N-dimension linear system with six parameters $(t_x, t_y, t_z, \omega_x, \omega_y, \omega_z)$ to be determined. A robust regression is subsequently applied to obtain a robust estimate.

The authors also suggested that in preprocessing, the images be corrected for the uneven intensity effect resulting from the use of different coils. Then the two images

should undergo adaptive histogram equalization to meet the intensity conservation assumption.

A popular algorithm by Woods (1992) relaxed the above assumption, assuming intensities of corresponding voxels are the same up to a constant multiplicative factor. Based on this assumption he estimated the parameters by minimizing the overall variance of the intensity ratio in the image volume. The minimization is carried out by the Newton-Ralphson method. The algorithm proved to be robust for both small and large displacements and is fast enough for online positioning. This algorithm is implemented in the publicly available package AIR (automated image registration).

1.3.2 Cross-modality Registration: Maximizing Mutual Information

In cross-modality matching, the intensity of the same tissue type usually differs from modality to modality. Thus the intensity conservation assumption no longer holds. Woods (1993) came up with a new model for registration between PET and MR data. In this model, a much weaker assumption is made. It assumes that "all voxels with a particular MR intensity value represent the same tissue type," so "the intensities of all corresponding PET voxels should also be similar." In application, pixels in an MR image are grouped into 256 (grey-level 0 to 255) bins. At each iteration, a weighted average of the normalized standard deviations within each bin is calculated. The iteration ends when the weighted average converges to a minimum. The major inconvenience of this algorithm in brain registration is the necessity of editing out the scalp region in preprocessing. Both intra-modality and cross-modality algorithms by Woods are implemented in the publicly available package AIR.

The previous approach shares essentially the same idea as in Viola's PhD thesis (1995): Maximize Normalized Mutual Information . We present the idea by first

giving a basic introduction of entropy and mutual information.

The entropy of a random variable X with probability distribution function $p(.)$ is defined as follows:

$$H(X) = E[\ln(p(X))] \tag{1.11}$$

where $0 \ln(0) = 0$.

Then by Jensen's inequality, the expectation of any concave function $f(X)$ is less than the function of the expectation:

$$E[f(X)] \leq f(E[X]) \tag{1.12}$$

Conditional entropy of random variable X given random variable Y is

$$H(X|Y) = E_Y[E_X[\ln(p(X|Y))]] \tag{1.13}$$

We can also prove that for any random variables X and Y

$$H(X) \geq 0 \tag{1.14}$$

$$H(Y) \geq H(Y|X) \tag{1.15}$$

Mutual information between random variables X and Y is given by:

$$I(X,Y) \equiv H(X) - H(X|Y) \tag{1.16}$$

When the distribution of X is known, a Y that minimizes $H(X|Y)$ will maximize mutual information. Since conditional entropy is a measure of the randomness of X given Y, a Y that explains well the variability of X will result in a small $H(X|Y)$, the extreme case being that X is a deterministic function of Y.

Put image registration into the above framework. The variables X and Y can be viewed as random functions defined on the image volume. Let X be intensity function for image 1 acquired under condition 1. Let Y be the intensity function

for image 2: the same object imaged under condition 2. Then given an arbitrary alignment of image 1 and 2, $p(X|Y = y_0)$ is the distribution of the intensity of those voxels in image 1 that fall in the same positions as all voxels in image 2 with intensity y_0. Mutual information of X and Y can then be calculated using (1.13), which is maximized when X is solely determined by Y, or say, all voxels with intensity y_0 in image 2 have the same intensity in image 1 after proper alignment.

The maximization procedure is similar to that of AIR (Woods et al., 1992) described in the last subsection. Due to the weak constraint placed on similarity of the conditions under which images were acquired, it is robust in matching images from any two modalities or images taken under different lighting conditions. This approach is implemented in ANALYZE, a commercial image processing software package.

Most of the above mentioned algorithms provide the option of estimating affine transformations. In an affine transformation the matrix R in (1.1) is arbitrary.

1.4 Deformable Template Methods

Deformable template methods, generally speaking, are registration methods that require non-linear warping. Such methods are most important for inter-subject registration.

1.4.1 Curved Transformation Models

One type of deformable template methods utilize *curved transformations*. A popular example is nonlinear spatial normalization using basis functions (Ashburner and Friston, 1999). This technique is implemented in the software package SPM.

The basic idea is that after intensity transformation $f(.)$, the reference image $I_2(.)$ is essentially the target image $I_1(.)$ after spatial transformation $T(\cdot)$, plus some

error $e(.)$:

$$f(I_2(x)) = I_1(T(x)) + e(x) \tag{1.17}$$

Our interest focuses only on the spatial transformation $T(.)$. The transformation is composed of a linear combination of a discrete sine transform (DST) and a discrete cosine transform (DCT).

DCT is a widely used image compression technique. The DCT of a function is generated by left-multiplication with matrix $\boldsymbol{B}$, where the elements are defined by:

$$b_{m,1} = \frac{1}{\sqrt{M}} \qquad m = 1...M \tag{1.18}$$

$$b_{m,j} = \sqrt{\frac{2}{M}} \cos\left(\frac{\pi(2m-1)(j-1)}{2M}\right) \qquad m = 1...M, j = 2...J \tag{1.19}$$

It transforms an image into the frequency domain. Values of b with small m are basis functions representing components in low frequency, namely large scale features, while values of b with large m's represent high frequency components, namely noise and details.

In this application, it is used to decompose spatial deformations.

$$\boldsymbol{y}_i = \boldsymbol{x}_i - \boldsymbol{B}(x) \times T_i \tag{1.20}$$

Here, rows in matrix B with small m's represent global deformation and rows with large m's represent local deformation. The value T_i represents the corresponding coefficients to be estimated. When implemented in SPM, the number of basis functions can be pre-selected, depending on how local a deformation is desired. Then the algorithm iteratively estimates the coefficients in T_i.

1.4.2 Physics Models

Using results from Broit's doctoral dissertation, Bajcsy further developed brain models based on the physics of elastic objects. He posed partial differential equations

based on the equilibrium state between the external forces causing the deformation and the internal force resisting the deformation. Constants representing elastic properties of the object are crucial in the deformation behavior. These constants are pre-selected. Then the forces that determine the deformation field are derived by minimizing the cost function, *cost = cost(deformation) - cost(similarity).* Similarity is measured by normalized cross-correlation (Ferreira, 1981; Bajcsy and Kovacic, 1989; Bajcsy et al., 1983; Gee et al., 1993)

Christensen (1996) modeled the deformation field as a highly viscous fluid with changeable volume to accommodate large distance deformations. Wang (1998) modified his model to elastic solids, and added a pre-alignment step using a surface method.

The above is a small sample of the work on modeling non-rigid deformations. As many of the authors themselves pointed out, most of these highly deformable template models are very sensitive to the starting position. Thus, it is important to use a rigid-body method to find a good initial alignment.

1.5 Theory of Patterns and Shapes

Mathematicians and statisticians have also contributed to the theoretical background of image analyses.

Before working directly on grey level images, Grenander began by developing a theoretical framework for patterns Grenander (1994). He modeled abstract shape with a space of generators (building blocks), a connector graph describing how generators are associated with one another, and bonding relations (or regularity conditions).

Kendall (1984) also defined a space for shapes. An ensemble of points is represented by a configuration matrix X, each row of which is the spatial coordinate of one point. The "shape" of this ensemble of points is defined as "all the geometric in-

formation about the configuration matrix X that is invariant under all translations, rotations and scalings." A metric of the shape space is also defined: the distance between any two shapes X_1, X_2 is the minimum of the sum of Euclidean distance between corresponding vertices in the two configurations under all translation, rotation and scaling. Let G be the group formed by all such transforms (called *similarity transforms*),

$$d_{shape}(X_1, X_2) = \inf_{g \in G} dist(X1, g(X_2)) \tag{1.21}$$

This distance is called Procrustes' distance. The process of calculating Procruste's distance between the landmarks of two similar images is the process of finding a match between the images with respect to landmarks. This process is called a full Procrustes fit. The solution can be obtained by standard least squares.

When there are two configurations for the same k landmarks, *deformation* is defined as a mapping from one configuration to the other. The desired properties of a deformation include: continuity, smoothness, bijectivity, no gross distortions, equivariance and being an interpolant.

In two dimensions, thin-plate splines proposed by Bookstein are such a deformation. A pair of thin-plate splines maps one image space onto another in a way that minimizes bending energy (Dryden and Mardia, 1998). The total bending energy of a 2D spatial transform $\Phi(x, y) = (\Phi_1(x, y), \Phi_2(x, y))$ is given by:

$$J(\Phi) = \sum_{j=1}^{2} \int \int_{R^2} (\frac{\partial^2 \Phi_j}{\partial x^2})^2 + 2(\frac{\partial^2 \Phi_j}{\partial x \partial y})^2 + (\frac{\partial^2 \Phi_j}{\partial y^2})^2 \mathrm{d}x\mathrm{d}y \tag{1.22}$$

Thin-plate splines are the solution of the minimization of (1.22) over all possible interpolating functions mapping from one configuration to the other, and thus are a natural interpolant in two dimensions.

1.6 Classic Bayesian Image Restoration Models

Analogous to Markov chain in one dimension, Besag (1974) modeled image data in two and higher dimensions. A neighborhood system on a number of sites is specified so that the conditional density of site i's value depends only on the value of its neighbors. That is, let ∂i be the neighbors of i, then

$$p(x_i|x_{S\setminus i}) = p_i(x_i|x_{\partial i}) \tag{1.23}$$

This system is called a Markov field. However, having specified a set of conditional probability distributions $p_i(.)$, a joint distribution $P(\boldsymbol{x})$ on all sites does not necessarily exist. The Hammersley-Clifford Theorem gave the necessary and sufficient condition for the existence of a joint distribution $P(.)$:

$$Q(\boldsymbol{x}) \equiv \ln\{P(\boldsymbol{x})/P(\mathbf{0})\} \tag{1.24}$$

Here, $Q(.)$ is the *energy function* of the system (up to a constant). There exists a unique expansion of $Q(.)$ in the following form:

$$\begin{aligned} Q(\boldsymbol{x}) &= \sum_{1\le i\le n} V_i(x_i) + \sum\sum\nolimits_{1\le i\le j\le n} V_{i,j}(x_i,x_j) \\ &\quad + \sum\sum\sum\nolimits_{1\le i\le j\le k\le n} V_{i,j,k}(x_i,x_j,x_k) + ... \end{aligned} \tag{1.25}$$

where $V_{i,j,...,s}$ may be non-null *if and only if* $i,j,...,s$ form a "clique." A *clique* is a subset of sites where every site is a neighbor of every other site. The function $V_{i,j,...,s}$ is then called the *potential function* of the clique $\{i,j,...s\}$. By plugging (1.25) into (1.24) and rearranging, we get the *Gibbs form* of $P(x)$:

$$P(x) = \frac{1}{Z}\exp(-\sum_{c\in\mathcal{C}} V_C(x)) \tag{1.26}$$

where $\mathcal{C}$ denotes the set of cliques in the system.

Besag (1986) demonstrated various applications of Markov random fields in image restoration. Since it is natural to assume that the intensity of a pixel is similar to the intensity of its neighbors, Besag used a nearest neighborhood system on a lattice to model pixel values. In this neighborhood system, there exists only two types of cliques. The first type is singleton clique, the second type is pairs of two adjacent pixels.

Then by the Hammersley and Clifford theorem, the joint probability of the system can be written in the following form:

$$P(x) = P(\mathbf{0})\exp(Q(x)) \tag{1.27}$$

$$= P(\mathbf{0})\exp(\sum_{i} V_i(x_i) + \sum_{<i,j>} V_{i,j}(x_i, x_j)) \tag{1.28}$$

where $< i, j >$ denotes i and j are neighbors. $P(x)$ is called a pairwise interaction Markov random field (M.r.f.).

Under the Bayesian framework for image restoration, a M.r.f. can be used as a prior for the underlying un-degraded image, or, the "*true scene*." Posterior inference can be made about the true scene $\boldsymbol{x}$ after observing the degraded image data $\boldsymbol{y}$

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \pi(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}) \tag{1.29}$$

Geman and Geman (1993) observed that the above posterior is *Gibbsian* with approximately the same neighborhood system as the prior. They then proved the *Relaxation Theorem* that if each site is updated by making a draw from its full conditional distribution infinitely often, then the sample always converges to the equilibrium distribution $\pi(.)$, regardless of the starting state $X(\mathbf{0})$. This theorem provided theoretical foundation for the widely used *Gibbs sampler*.

To obtain a sample from the high dimensional posterior in (1.29), we begin with any configuration, then proceed by drawing a sample at each site s from its conditional

distribution

$$\pi_s(x_s|x_{\partial s}) = \frac{1}{Z_s} e^{-U_s(\boldsymbol{x})} \tag{1.30}$$

where the energy function U_s is the sum of potentials in all cliques that contain site s: $U_s(\boldsymbol{x}) = \sum_{C:s\in C} V_c(\boldsymbol{x})$.

With Gibbs sampling, we can calculate the stochastic expectation of the posterior. Nevertheless, it is often desirable to also find the m.a.p. estimate (maximum *a posteriori*), which gives the "most probable scene." Geman & Geman adopted the idea of *simulated annealing* in a physical system. Let the probability distribution be written

$$P_T(x) = \frac{1}{Z_T} \exp(Q(x)/T) \tag{1.31}$$

The parameter $T > 0$ corresponds to the "temperature" of the system. When $T \to \infty$, $P_T(.)$ approaches uniform distribution. When $T = 1$, $P_T(.)$ becomes (1.28). When $T \to 0$, $P_T(.)$ approaches a point mass concentrated on the m.a.p. estimate. Simulated annealing is the scheme of finding the m.a.p. estimate by gradually lowering the temperature according to a prescribed schedule.

However, there are some drawbacks in applying simulated annealing to image restoration. First, the cooling schedule is usually so slow that it imposes a formidable computation load for even small images. Second, due to its undesirable large scale properties, this scheme often yields a single colored image upon convergence.

Besag (1986) proposed Iterative Conditional Modes to eliminate these undesirable features of the m.a.p. estimate, yielding a more computationally feasible algorithm. ICM iterates through all sites, finds $\hat{x}_i$ that maximizes the full conditional $p_i(x_i|x_{\partial i})$, and is guaranteed to converge. This basic scheme was successfully applied to binary images, discrete color images, continuous intensity images, *etc.*

Bayesian modeling has become increasingly acknowledged as an important tool in

image analysis. A variety of knowledge can be incorporated into the prior under the Bayesian framework, such as belief about image features, prior information on the object shape, and knowledge obtained from training image sets, *etc.* In this thesis, we incorporate prior beliefs about the spatial behavior of landmarks in the Bayesian framework.

Chapter 2

Model

In this chapter, a registration method will be presented. No explicit knowledge is required of the imaged object. Furthermore, the method is not designed for one particular type of image, but can be applied generally to images acquired from a variety of apparati.

The model deforms one image to match another image in the same image class. An image class can be roughly defined as a set of images of objects with the same geometric/anatomical structure. The conditions under which the images are acquired can be very different. Throughout this thesis the image to be transformed is called the "atlas image," and the image according to which the atlas image is transformed is called "target image." "Atlas images" are also called "reference images" or "templates" by other authors. "Target images" are sometimes called "object images" or "study images." Here, the choice of which image to be the atlas is arbitrary.

In our model, the transformation is carried out by a set of imaginary labeled points called "facets," which are in facet a type of landmark(McCulloch, 1999). Thus the model is sometimes referred to as the "facet model" in this thesis.

2.1 Facets

The "facets" in our model are in fact a type of landmark.

Dryden and Mardia (1998) summarized three types of landmarks

- anatomical landmark: a point assigned by an expert that corresponds between organisms in some biologically meaningful way, for example the corner of an eye;
- mathematical landmark: a point located on an object according to some mathematical or geometrical property of the image, for example a point of maximum intensity on the cheek;
- pseudo-landmarks: also called "generalized landmarks," are constructed points located either around the outline or in between anatomical or mathematical landmarks.

"Facets" belong to the category "pseudo-landmarks" by the above classification. A generalization from biological landmarks or mathematical landmarks, a facet placed on an image doesn't usually have a specific interpretation. In the applications presented in this thesis, the underlying "true scene" is not within our concern. Thus the facets represent the pixel values of the acquired image "as is" (Laading, 2000).

Below we will describe the model in two dimensions. The description can be easily generalized into three dimensions. For example, in applying the model to digital images, we will mostly use the term "pixel" instead of "voxel," although the statements would still hold true if we replace "pixel" with "voxel".

Let Ω be the domain on which an image is defined. Normally, for a 2D image $\Omega = [0, N_1 - 1) \times [0, N_2 - 1) \subset R^2$. For gray level images, intensity, or brightness, is a non-negative function defined on Ω: $g(.) : \Omega \to R^+$. Please note that, in contrast to

digital images, in our model definition, the image domain is continuous, rather than discrete, and image intensity is a piecewise differentiable function defined on Ω.

In addition to intensity, the properties of an image at a certain location may be characterized by functions of intensity at or around that location. These properties should be intrinsic to the image. They are called "features" in this thesis. Some commonly used features include gradient $\nabla g = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y})$, gradient magnitude (edgeness) $\|\nabla g\|_2$ and Laplacian (medialness) $\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$. For more differential geometry features, see Wilson thesis 1995. A few examples of features are shown in Figure 2.3.

Let $\mathcal{F}$ be the space of features. It can be the product space of a chosen number of features: $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes ... \otimes \mathcal{F}_s$. A feature evaluated at any location on an image is a point in the feature space, *i.e.* $f = (f_1, f_2, ..., f_s) \in \mathcal{F}, f_i \in \mathcal{F}_i, i = 1, 2, ..., s$. An image A defines a mapping from image space Ω to the feature space $\mathcal{F}$: $Q_A(.) : \Omega \to \mathcal{F}$. Or say, given an image A, the corresponding feature f is a definitive function of the location. Then the feature of facet 0 at location x_0 is

$$f_0 = Q_A(x_0) \in \mathcal{F} \tag{2.1}$$

A large number of facets are placed in the volume of the atlas image A on a lattice. Each facet is then located in the target image T through the application of the model we describe in detail in the next section. For an illustration see Figure 2.4. In theory, if we are able to place a facet at any location in A and locate it in T following the model mechanism, then the model will define a mapping from A to T. Since both images are defined on Ω, this mapping is also called a "warp." However, in reality, our model is not a parametric spatial transform, and we can only place a finite number of facets in A, so our model only yields an approximation to the warp.

According to Dryden and Mardia, a warp should have the following properties:

1. continuous and smooth

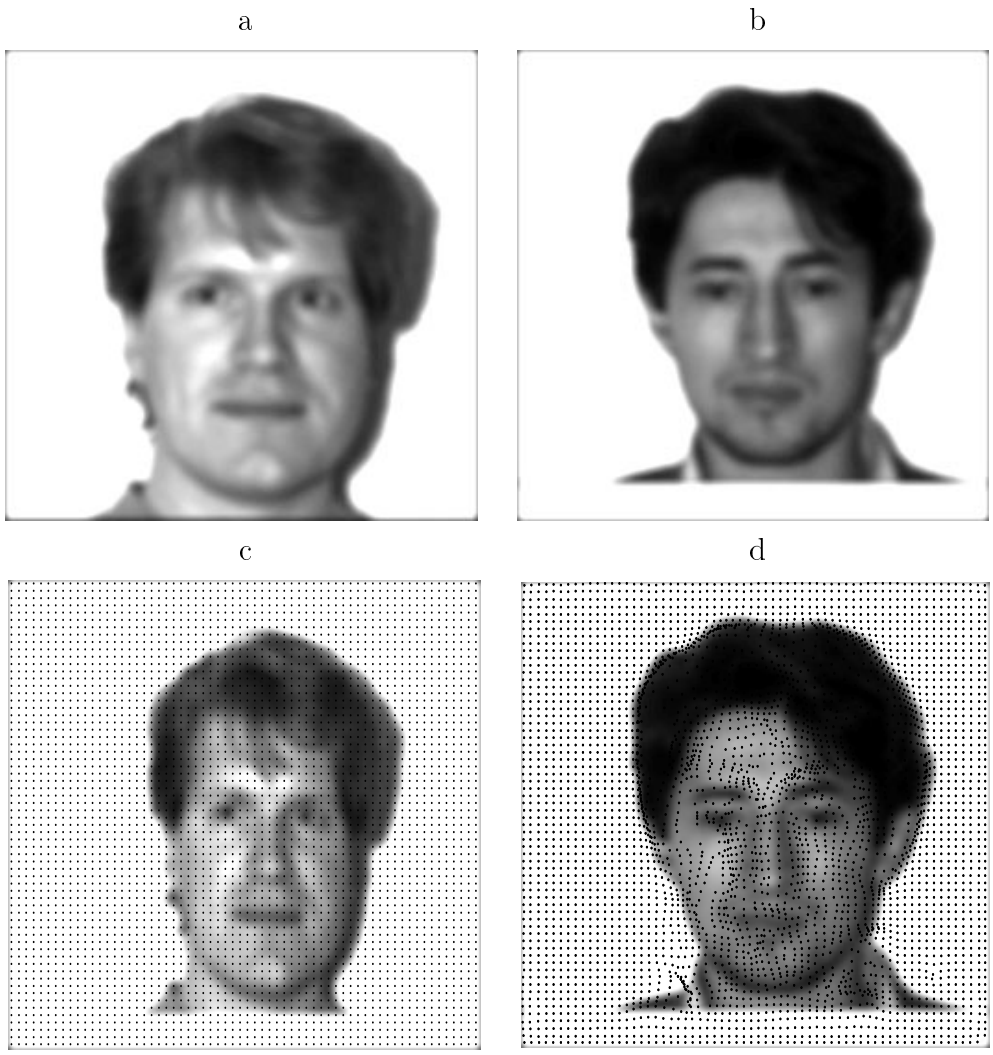


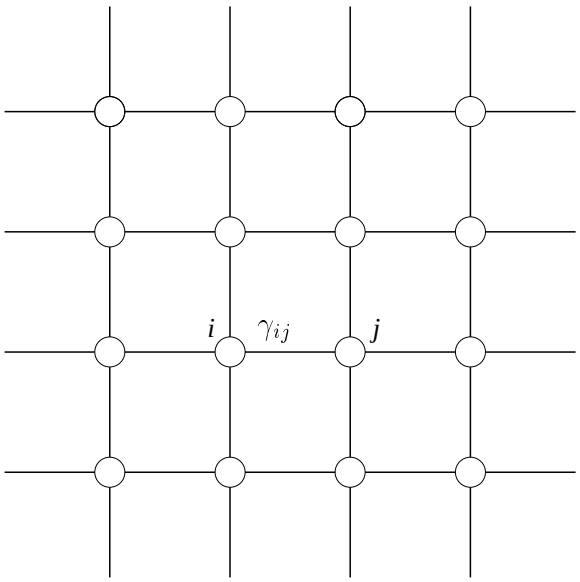
Figure 2.1: a) Atlas image; b) target image; c) atlas with facets; d) facets located on the target.

2. bijective (1-1 and onto)

3. not prone to distortions (*e.g.* not folding)

4. equivariant under relative location, scale and rotation of the objects.

We will explain how we try to incorporate these properties in the model prior.



2.2 Prior

Facets labeled $1, 2, ..., M$ are placed on atlas image A at locations $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_M)$. The atlas features are denoted by $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_M) = (Q_A(\mu_1), Q_A(\mu_1), ...Q_A(\mu_M))$. In the deformation process, facets move from atlas locations $\boldsymbol{\mu}$ to new locations $\boldsymbol{x} = (x_1, x_2, ..., x_M)$ in target image T and thus take on new features $\boldsymbol{f} = (f_1, f_2, ..., f_M) = (Q_T(x_1), Q_T(x_2), ..., Q_T(x_M))$.

Inference shall be drawn on target locations $\boldsymbol{x}$.

Assume that M facets are laid down in a lattice on the atlas image, such that the distance between adjacent facets is small compared to the scale of the image. In the facet model context, the continuity and smoothness property (of a desirable deformation described by Dryden and Mardia) can be translated as: if facet i has deformed from its atlas location μ_i by dx, then the deformations of i's adjacent facets from their respective atlas locations should not be much different from dx. This heuristic gives rise to a pair-wise difference Markov random field as the prior for $\boldsymbol{x}$.

A first degree neighborhood system on a rectangular lattice is used (see Figure

2.2). Only two types of cliques exist for this system: singleton cliques and pairs of nearest neighbors. By the Hammersley-Clifford Theorem, for the joint distribution of $\boldsymbol{x}$ to exist, only the potential of these two types of cliques can be non-null. Letting the former be null and the latter be based on pair-wise difference, we can write the prior in Gibbs form

$$p(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\gamma})} \exp\{-\frac{1}{2} \sum_{\langle i,j \rangle} \gamma_{ij} ((x_i - \mu_i) - (x_j - \mu_j))^a\} \tag{2.2}$$

Here, $\langle i, j \rangle$ indicates that i and j are neighbors. The parameter $\boldsymbol{\gamma}$ is a hyper-parameter that describes how "strong" the link is between i and j. The link is symmetric, so $\gamma_{ij} = \gamma_{ji}, \forall i, j$. The pair-wise differences are raised to the power $a > 0$. As a gets larger, more smoothness in the deformation is encouraged, *i.e.*, more gridness is preserved in the deformed lattice.

In particular, when $a = 2$, all conditionals are Gaussian and we have an auto-normal model. Then we can re-write (2.2) as follows:

$$p(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\gamma})} \exp\{-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T Q (\boldsymbol{x} - \boldsymbol{\mu})\} \tag{2.3}$$

where Q is a M by M matrix whose diagonal is $\sum_{j \in \partial i} \gamma_{ij}$ for the i-th element. Its off-diagonal element is $-\gamma_{ij}$ where i and j are neighbors and 0 elsewhere. It is easy to see that all the row sums of Q are zeros. So $(1, 1, ..., 1)$ is an eigenvector of Q, and the corresponding eigenvalue is 0. Thus Q is not a positive-definite matrix.

In addition, it is obvious from (2.2) that the prior is translational invariant, namely:

$$p(\boldsymbol{x}) = p(x_1, x_2, ..., x_M) = p(x_1 + c, x_2 + c, ..., x_M + c) \tag{2.4}$$

for any real number c.

Translational invariance is a desirable property. However, it also renders the prior improper. The prior cannot be sampled unless a singleton potential is added.

Now we look at the conditional representation of the prior when $a = 2$.

In the full conditional specification, the expected deformation of a facet is a weighted average of its neighbors' deformations:

$$p(x_i|.) = p(x_i|x_{\partial i}) \propto \exp\{-\frac{1}{2}\sum_{j\in\partial i}\gamma_{ij}((x_i-\mu_i)-(x_j-\mu_j))^2\} \tag{2.5}$$

or

$$E(x_i|.) = E(x_i|x_{\partial i}) = \mu_i + \frac{\sum_{j\in\partial i}\gamma_{ij}(x_j-\mu_j)}{\sum_{j\in\partial i}\gamma_{ij}} \tag{2.6}$$

Thus, the spatial relationship among sibling facets is preserved as shown in Figure 2.2.

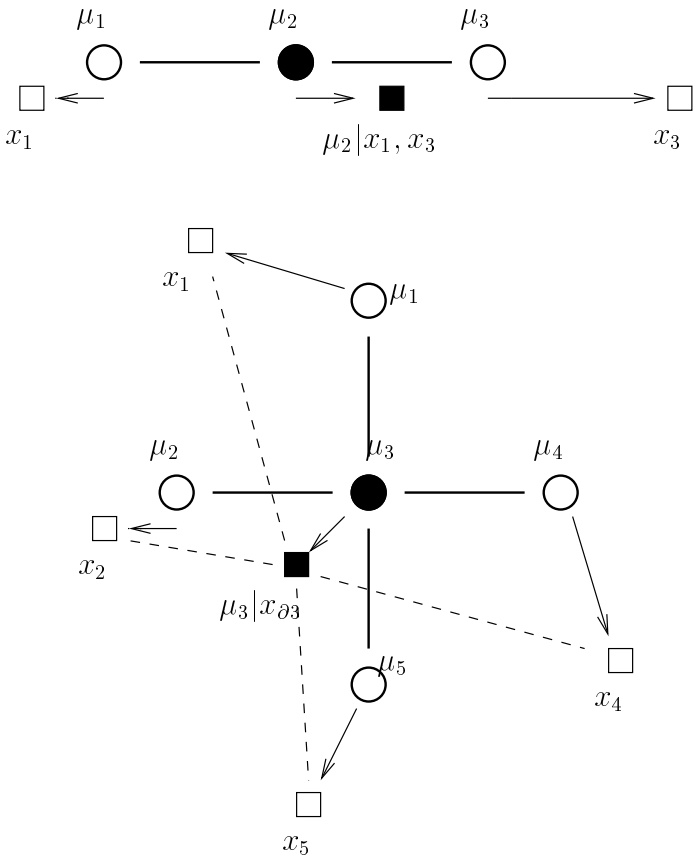


Figure 2.2: Preserving relative position within siblings.

Markov random fields have been traditionally used to model image features at fixed locations. In this thesis, it is used to model location itself. The positivity rule still applies, which requires a positive probability be assigned to any configuration of the facets. The majority of all possible configurations would not give rise to a smooth deformation without any folding. However, we can carefully choose the range of model parameters so that the probability of yielding a configuration that represents gross distortions is negligible.

We discuss in Section 2.6 how we determine $\boldsymbol{\gamma}$.

The computational detail are detailed in Section 2.5.

2.3 Likelihood: Measure of Similarity

In the likelihood we model our belief that the feature of a facet in the target image should be similar to that in the atlas image. How similar is "similar"? We propose a measure of "similarity" later in this section. For now we use only $-sim(f_0, \phi_0)$ to represent similarity between two features f_0 and ϕ_0.

Suppose facet i has atlas feature ϕ_i. We let the probability of seeing facet i taking on a target feature f_i be proportional to the similarity between f_i and ϕ_i exponentiated, *i.e.*.

$$p(f_i|\phi_i) \propto \exp\{sim(f_i, \phi_i)\} \tag{2.7}$$

Remember that a feature is a function of location, $f_i = Q_T(x_i)$. In fact we have just defined a probability density function for x_i:

$$p(x_i|T, \phi_i) = \frac{1}{C_i(T, \phi_i)} \exp\{sim(Q_T(x_i), \phi_i)\} \tag{2.8}$$

The normalizing constant of this distribution is

$$C_i(T, \phi_i) = \int_{x \in \Omega} \exp\{sim(Q_T(x), \phi_i)\} dx \tag{2.9}$$

Note that (2.9) may not be integrable when $sim(.)$ does not have an upper bound.

Before observing target T, we assume that facet features are independent given their locations. Then the likelihood can be written as a product:

$$p(\mathrm{T}|\boldsymbol{x}) \propto \exp\{\sum_i sim(Q_T(x_i), \phi_i)\} \tag{2.10}$$

Now we come back to the problem of choosing a measure of similarity. First, let's look at some of the commonly used measures.

Many pre-existing registration methods are based on the "intensity preservation assumption." These methods directly take the absolute difference of intensities between corresponding locations as a measure of difference. Furthermore, if intensity is preserved, gradient and other differential invariants should be preserved as well, so the difference of these invariants can also be a difference measure. Later methods relaxed the assumption so that the intensities could be equal up to a multiplicative constant. Then correlation of intensities or other features in a small neighborhood may be used.

People have also explored the use of other difference measures that are not explicitly based on such assumptions, for example the difference between local ranking of some features.

Figure 2.3 shows examples of features used in the above mentioned measures

Below we summarize the pre-existing measures and list the properties we are most concerned with

- robustness under weak assumptions on intensity. Absolute difference measures have very limited application because the intensity conservation assumption usually does not hold in reality. Even within the same modality, if scans are taken under slightly different settings (*e.g.* using different surface coils), the

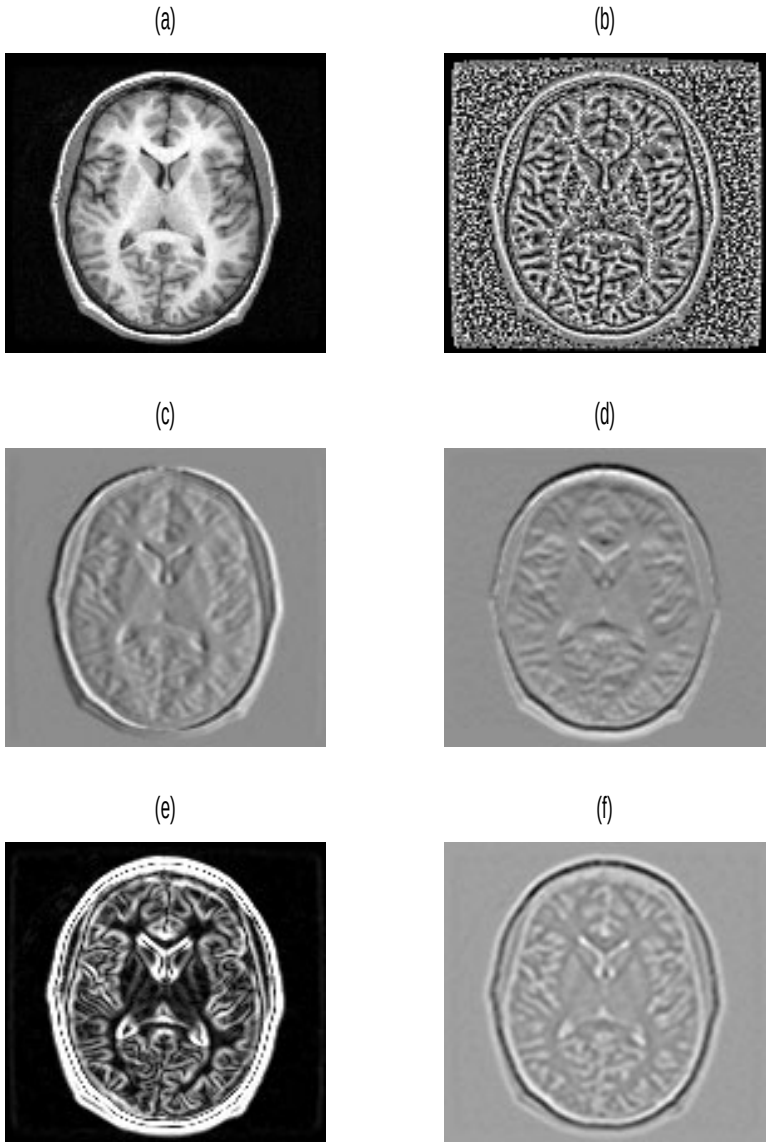


Figure 2.3: Examples of image features: a) Original image; b) local rank of intensity; c and d) directional derivatives; e) gradient magnitude; f) Laplacian.

intensity of the same tissue can vary greatly from one scan to another. Usually there is a difference in the overall intensity level among scans. Using absolute differences in these cases will lead to erroneous matches. Preprocessing steps such as scaling the overall intensity level linearly according to quantiles or histogram equalization can alleviate this problem to some extent. Measures invariant under linear scaling of the intensity such as ranking and correlation

can avoid this problem.

- ability to distinguish between the "meaningful" and "meaningless." A major disadvantage of many scale-invariant features is their locality. For example, local ranking works well in areas where image intensity changes fast. But when intensities are ranked in regions where intensities are approximately constant, the rank will fail to represent the actual lack of information and may mislead the match. See Figure 2.3(b).

- ability to accommodate non-existent features. Sometimes both images do *not* contain all the same structures. We would like to incorporate a mechanism so that the missingness of one structure does not mislead the registration.

A measure that has all the above properties is proposed as follows:

$$sim(f_i, \phi_i) = b(r_i \cdot \rho_i)^\alpha \cdot \cos(u_i - \mu_i), \qquad \alpha > 0 \tag{2.11}$$

where r_i is the quantile of the gradient magnitude at facet i in the target image, *i.e.* $(100r)\%$ of all pixels have a smaller gradient magnitude than facet i. The value u_i is the direction of the gradient. The values ρ_i and μ_i are those of the atlas, respectively. Equation 2.11 represents the difference in the directions of gradients, weighted by the quantile of gradient magnitudes.

The heuristic is that, when looking at an image, the human visual system would capture salient features first (edges, corners, intensity maxima). In our registration process, we want facets associated with these features to guide the match even if a large deformation is required. The rest of the facets are less influential in the spatial deformation.

The multiplier in (2.11) is $b(r_i \cdot \rho_i)^\alpha$, which controls how much weight is given to the feature. A facet with a larger r has a more significant gradient, and thus would

be more likely to be picked up by human eyes. In the deformation, higher priority will be given to facets with larger r to achieve a close match, which is exactly what our visual system tends to do.

This measure is invariant under all linear transformation of the intensity. Because it uses only the direction of gradients, and the weight is a function of global quantiles, which are always between 0 and 1. It is also robust under nonlinear but monotonicly increasing transformations of the intensity.

Furthermore, we can slightly change the measure to fit images with different orderings of intensities according to tissue types. For example, suppose that under modality A, the relationship between three tissue types is: tissue 1 > tissue 2 > tissue 3; under modality B, the relationship is tissue 2 > tissue 1 > tissue 3. Then if we use $|\cos(u_i - \mu_i)|$ instead of $\cos(u_i - \mu_i)$ in (2.11), the reversal of gradients will not affect the match.

In background areas or regions in the image where intensities are approximately constant, r and ρ should both be small, and then the feature difference will be given little weight. The deformation of facets in those areas would be mostly guided by the prior.

Sometimes one of the images is missing a structure, which means that that region is replaced with a featureless patch of near-constant intensity. In this case the corresponding quantile (either r or ρ) will be small, causing the whole multiplier to be small. Again the feature would be given small weight. The deformation will mostly retain gridness instead of trying to look for a non-existent feature.

Figure 2.4 is an example demonstrating the performance of this measure in different situations. The atlas and target we used are both images of human faces. The atlas image is significantly brighter than the target, and some details are also very different. In the atlas image to the left, we highlight facets on the nose blade (edge),

Figure 2.4: Facets highlighted on the atlas image (left) and located on the target image.

pupils (intensity minima), tip of the nose (intensity maximum), both corners of the mouth (corner) and the eyebrow bone (missing feature because the subject's eyebrow is hardly visible). Despite the difference, the facets seem to have found reasonable locations on the target image in each situation.

To summarize, compared to traditional measures of similarity, the measure proposed in (2.11) does not make strong assumptions about the images and seems to deliver good performance.

2.4 Posterior

We have specified a prior on the facet locations and a likelihood based on "similarity" of facet features. It follows that the posterior is proportional to the product of the prior and the likelihood component:

$$\begin{aligned} p(\boldsymbol{x}|A,T,\boldsymbol{\gamma}) \propto \exp\{ \quad & -\textstyle\sum_{\langle i,j\rangle} \gamma_{ij}((x_i-\mu_i)-(x_j-\mu_j))^2 \\ & +\textstyle\sum_i b(r(x_i)\cdot\rho_i)^\alpha \cdot cos(u(x_i)-\mu_i)\} \end{aligned} \tag{2.12}$$

Note that $r(x_i)$ and $u(x_i)$ are written as functions of facet location.

In fact the posterior in (2.12) is a Gibbs representation of a new random field. The energy function is the sum of shape potentials and feature potentials. The resulting field is a balance between the shape and feature.

Since feature potentials exist only in singleton cliques, the resulting random field still has a first-order neighborhood system as the prior does. The nice Markovian property of this field makes inference by Markov chain Monte Carlo straightforward. However, slow mixing is exhibited because of the extremely high dimensionality and large number of local maxima. The following literature deals with this common problem in M.r.f. inference: Higdon and Lee (2001), Lee et al. (2000), Besag et al. (1995), Rue (2001), Higdon et al. (1997), *etc.*

In cases where we are interested in the m.a.p. estimate, most maximization algorithms will be very inefficient because of the huge amount of local maxima.

To facilitate computation, we introduce a hierarchical design on the shape portion of the model. Details are described in next section.

2.5 Multi-resolution Maximization

The iterative conditional modes (ICM) method proposed by Besag (1986) is used to find a local maximum of (2.12). It maximizes, one by one, the full conditional of each facet location, then repeats until convergence is achieved. ICM converges relatively fast, especially compared to annealing. However, it does not guarantee convergence to the global maximum. Due to the dimensionality of the problem, the posterior has a large number of modes, making it nearly impossible to obtain the global maximum. Fortunately, we can take advantage of the scale-space of images to find a good mode.

Images can be viewed at different scales. A higher scale version of an image can be obtained by re-sampling the original image at a lower-resolution, or by convolving the original image with a Gaussian kernel. In one dimension, there are always fewer

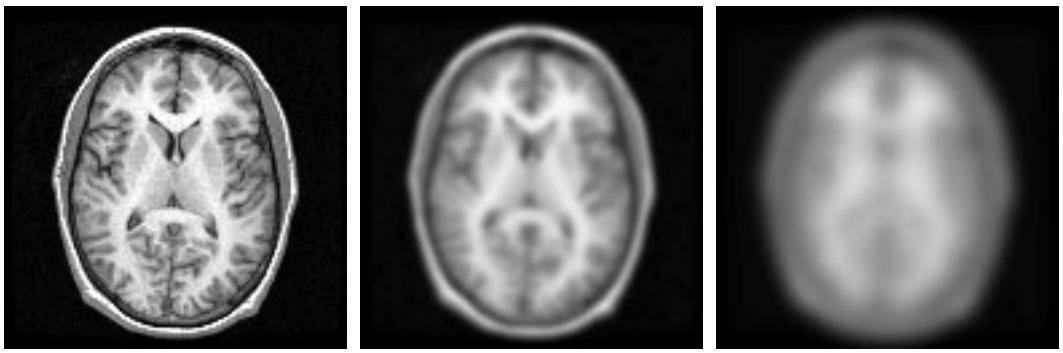


Figure 2.5: Original image on the left; image blurred with Gaussian kernels on the right.

local maxima and minima in the higher scale of a signal, although counter-examples exist in higher dimensions (Lifshitz and Pizer, 1987). Generally speaking, details and textures are lost when we compute a high scale version of an image; only a few of the most salient features are left. See, for example, Figure 2.5.

Similarly, any spatial deformation can be seen as the additive result of global deformations and local deformations. Global deformations happen only in large scale while the details keep relatively unchanged. For example, a large organ is shifted by a certain distance in the target image. Such a deformation would be somewhat hard to achieve under our current model mechanism since it needs to separately move each of the many facets belonging to this organ by the same distance.

To overcome this difficulty, we place facets in different scales of an image, letting facets at higher scales represent large scale features, those at lower scales represent small scale features. Then we have several levels of facets. From top level down, each level of facets are laid out in a finer grid according to the image resolution. Each facet (except for the ones at the top level) is linked to a facet one level above through a parent-child link. Figure 2.5 is an example of this hierarchical structure in one dimension.

If the parent facet p deviates from μ_p by δx_p, all its children are expected to

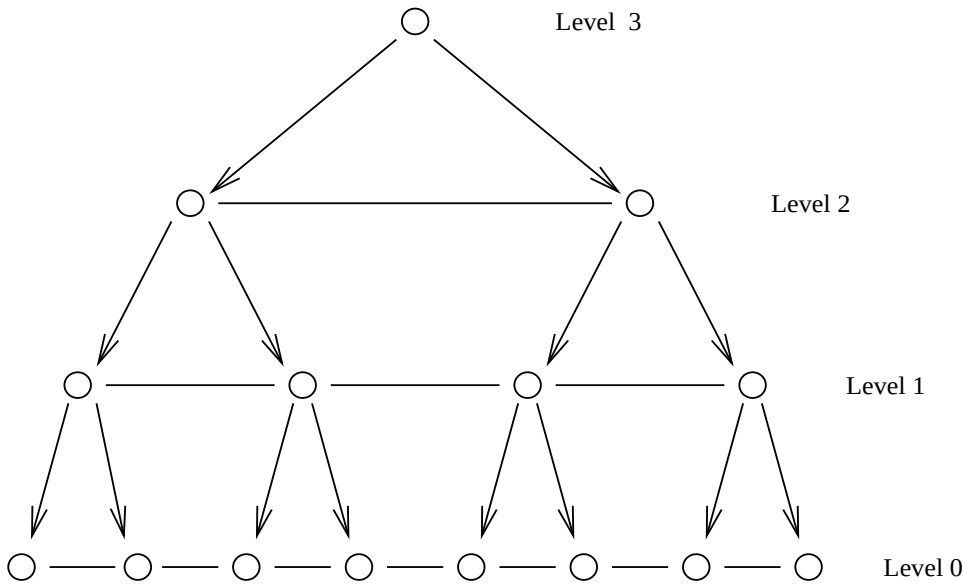


Figure 2.6: The hierarchy of facets in 1D with all neighborhood links.

deviate in the same direction but with a smaller penalty (Figure 2.7).

In our maximization scheme, we take advantage of the hierarchical structure of facets to achieve close to global maximums faster. For each facet, we modify the full conditional to include the deformation of its parent. So, on the basis of (2.5), we have

$$p(x_j|x_s, s \in S_j, x_p) \propto \exp\{-\gamma_p((x_j-\mu_j)-(x_p-\mu_p))^2 + \sum_{s\in S_j} \gamma_{js}((x_j-\mu_j)-(x_i-\mu_i))^2\} \tag{2.13}$$

We begin by optimizing the location of the facet(s) in the highest level one by one using (2.5). Then we move to the next lower level, maximizing all facets in that level. The algorithm proceeds until it reaches the bottom level, using (2.13) for each facet. This can be repeated several times until convergence is achieved.

To further increase the possibility of finding a good maximum, we designed a variation of ICM for the customized need of our setup. In each level the prior at a higher temperature is used first to encourage facet movements. The temperature

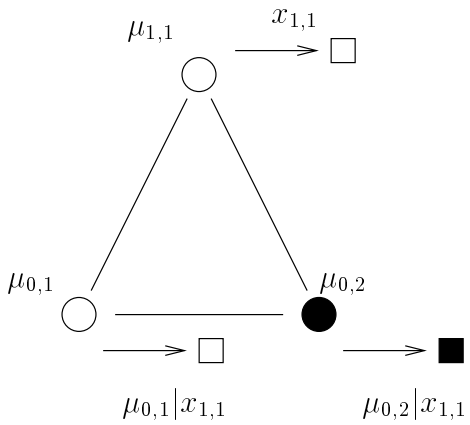


Figure 2.7: Deformation of the parent propagated down the tree.

then gradually cools down in the next few iterations to 1.

One thing to note is the maximization algorithm used in each update. For all pre-existing numerical algorithms, even in a small neighborhood, if multiple modes exist, the maximum achieved is not necessarily the best one among them. But we would like to get as close to the global maximum as possible. Based on the strength of the spatial constraint, we can usually obtain a good maximum by doing a grid search in a calculated neighborhood.

2.6 Determining Model Parameters

Since we have more model parameters than facets in a single image, it is unrealistic to estimate them from one pair of images. More importantly, we want to have control over how the model performs. So below we explain how we choose the values of the parameters beforehand.

First of all, for the sake of convenience, we assume that the deformation is homogeneous within each level of the scale-space, which means we assign only one value to all sibling link γ's within each level.

Now we consider each level in isolation. Within the level, as mentioned in Section 2.2, γ's control the smoothness of the deformation. Values of γ's that are too big do

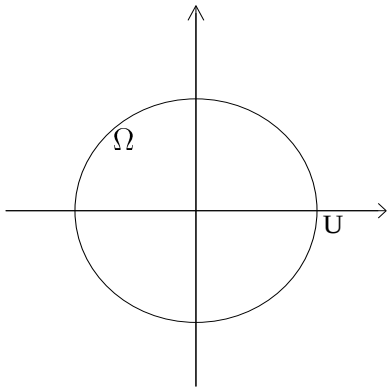


Figure 2.8: Illustration for (2.16) and (2.17): the probability of each facet falling in Ω is 0.95.

not allow necessary deformation, while values of the γ's that are too small produce gross local distortions such as folding. Assume then that the facet is originally placed on an evenly-spaced quadrangular lattice, and the distance between first-order neighbors is D. Then we choose γ according to the following principle: given that all other facets remain in their atlas locations, the probability of any interior facet deforming more than $U = \frac{D}{2}$ from its atlas location is exactly 5%:

$$\Pr(\|x_i - \mu_i\| > U) = 0.05 \tag{2.14}$$

Let σ^2 denote the conditional variance of x_i.

Then in two dimensions (2.14) can be re-written as the integral of a bi-variate normal density within a circular region Ω as shown in Figure 2.8, which can be easily calculated under polar coordinates:

$$\begin{aligned} 0.05 &= \Pr(\|x_i - \mu_i\| > U) && (2.15)\\ &= 1 - \Pr(\|x_i - \mu_i\| \leq U) && (2.16)\\ &= 1 - \int\int_{\Omega} N(\boldsymbol{x}|\boldsymbol{\mu}, \sigma^2 * I) dx_1 dx_2 && (2.17)\\ &= \exp(-\frac{U^2}{2\sigma^2}) && (2.18) \end{aligned}$$

Then σ^2 can be obtained according to the value of U. From (2.5), $\sigma^2 = (\sum_{j\in\partial i}\gamma_{ij})^{-1} = (4\gamma)^{-1}$ in a homogeneous two dimensional field. Finally we get

$$\gamma = \frac{-\log(0.05)}{2U^2} \tag{2.19}$$

In three dimensions, this calculation can be repeated by integrating over a sphere S of radius U:

$$\begin{aligned} 0.05 &= 1 - \int\int\int_S N(\boldsymbol{x}|\boldsymbol{\mu}, \sigma^2 * I)dx_1dx_2dx_3 && (2.20)\\ &= 1 + 2(U\exp(-\frac{U^2}{2\sigma^2}) - \Phi(\frac{U}{\sigma}) + \frac{1}{2}) && (2.21)\end{aligned}$$

where $\Phi(.)$ is the cumulative distribution function of standard normal. For each U, the root can be found numerically. In three dimensions, $\sigma^2 = (6\gamma)^{-1}$. Listed below are the values of γ corresponding to different values of U:

$$\gamma = \begin{cases} 2.17, & U = 0.5 \\ 0.643, & U = 1 \\ 0.185, & U = 2 \\ 0.0565, & U = 4 \end{cases} \tag{2.22}$$

In contrast to two dimensions, the parameter specification in three dimensions is not independent of the scale of U. If we let $\alpha = U/\sigma$, then rewrite (2.21), we can see that as U approaches 0, σ has an order of magnitude $O(U)$. As $U \to \infty$, σ is of order $o(U)$.

Now we put the current level back into the multi-resolution context. For a facet, besides the influence of its siblings, we let the parent guide a portion of the total deformation. The proportion can be chosen at the modeler's discretion. We consider two extremes: the first being parents contribute 0 to its children's deformations. Then the levels are completely un-connected. The maximization algorithm would

proceed as if maximization is done on each resolution in isolation. The other extreme would be the parent determines 100% of its children's deformation, then the facet structure would behave like a tree structure where there is very high correlation between children of a common parent (or grandparent, *etc.*), but no correlation between facets without a common ancestor. This destroys the homogeneity of the field.

Based on this heuristic, we chose 50% as the parent contribution. Note that the parent-child link is directed, which means the child-parent link is not equal to the parent-child link. In our maximization scheme, the child to parent contribution will be 0.

Chapter 3

Application to Segmentation

3.1 Segmentation of *in vivo* Mouse Brain MRIs

In a study conducted on live mice that investigates the relationship of Alzheimer's disease to changes in the volume of the hippocampus, it was necessary to obtain hippocampus volumes for a large number of mouse brains. *In vivo* scans of two groups of mice were collected. One group was a strain of genetically altered mice that had developed the mice's version of Alzheimer's disease. The other group was the control group consisting of normal mice. To obtain measurements of hippocampi, researchers usually viewed the brain images with imaging softwares such as Adobe Photoshop, NIH Image, or specter, manually mark the region they believe to belong to hippocampi on each 2D slice, and measure that area on each slice. Finally a simple program could be used to calculate the overall volume of the hippocampus in three dimensions.

Shown in the top panel of Figure 3.2 is a horizontal view of a mouse brain numbered 16283 from the control group. It is one of the 28 slices in a 3D MR scan of resolution $256 \times 512 \times 28$. The hippocampus regions have been carefully marked by an expert, as the highlighted areas shown in the lower panel of Figure 3.2.

Next look at the right hippocampus. Starting counterclockwise from the right-most: the hippocampus stops at the boundary of the brain; then it neighbors the right ventricle (hollow regions in the brain, corresponds to low intensities in MR scans); the mid-line between hemispheres clearly separates the right hippocampus from the left hippocampus; finally the border between hippocampus and superior colliculus (the tissue below hippocampi) is very vague as the intensities of the tissues are indistinguishable from each other and the border has been blurred by white noise.

Several difficulties encountered in the segmentation of hippocampi are illustrated in this example: first, there is a high level of high frequency noise in the entire image, leaving a large gray area when specifying borders; second, there is an absence of a clear border of hippocampi from the rest of the brain tissue, most of it is left to the segmentor's experience and imagination; thirdly, a small angle between the horizontal plane of the mouse brain and the scanning plain caused the left ventricle to be almost completely missing in the illustrated slice. Each of these factors reduces the consistency of human segmentation.

It is obvious that none of the intrinsic segmentation methods would be able to segment such a poorly defined area.

3.2 Applying the Facet Model

Now we apply the facet model to mouse brain images to automate hippocampus segmentation. For illustration purposes we use brain 16283 as an atlas, wherein each slice that contains hippocampus has been manually segmented. The target brain to be segmented is number 16333 from the genetically altered group. Since the distance between slices are about 10 times pixel width, the matching algorithm was applied in a 2-D fashion slice by slice.

In the model a hierarchy of facets are placed in the scale space of the atlas image.

Figure 3.1: Manual segmentation of hippocampi in one slice from mouse brain number 16283. Left: original image. Right: manual segmentation of hippocampi highlighted.

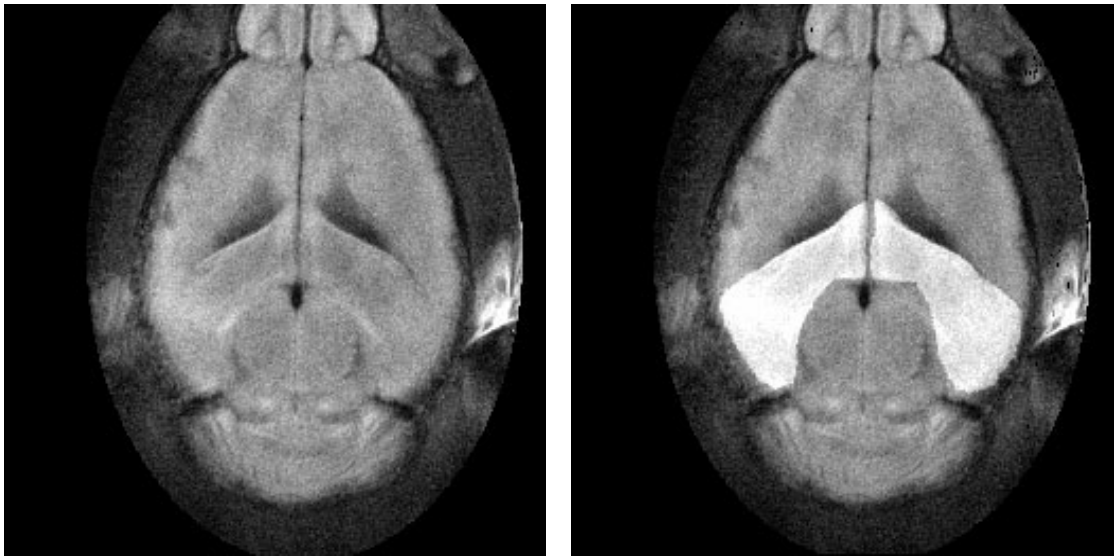


Figure 3.2: Automated segmentation of corresponding slice from mouse brain number 16333. Left: original image. Right: automated segmentation (based on manual segmentation of 16283) highlighted.

The bottom level facets are placed at the center of each pixel, so we have as many facets as pixels in the images. After running the deformation model, each facet finds its new location in the target image. If we give a label to all facets that belong to the segmented region in the atlas, then in the target the region occupied by the facets with that label is the automated segmentation. Interpolation is performed when pixels overlap with more than one facet. The label of each of those pixels is the weighted average of the labels of all the facets it overlaps with, the weights determined by the overlapped area.

Figure 3.2 shows the automated segmentation of a slice from brain 16333 by matching it to slice 23 of brain 16283. We can see that the auto-segmentation correctly captured the hippocampus region.

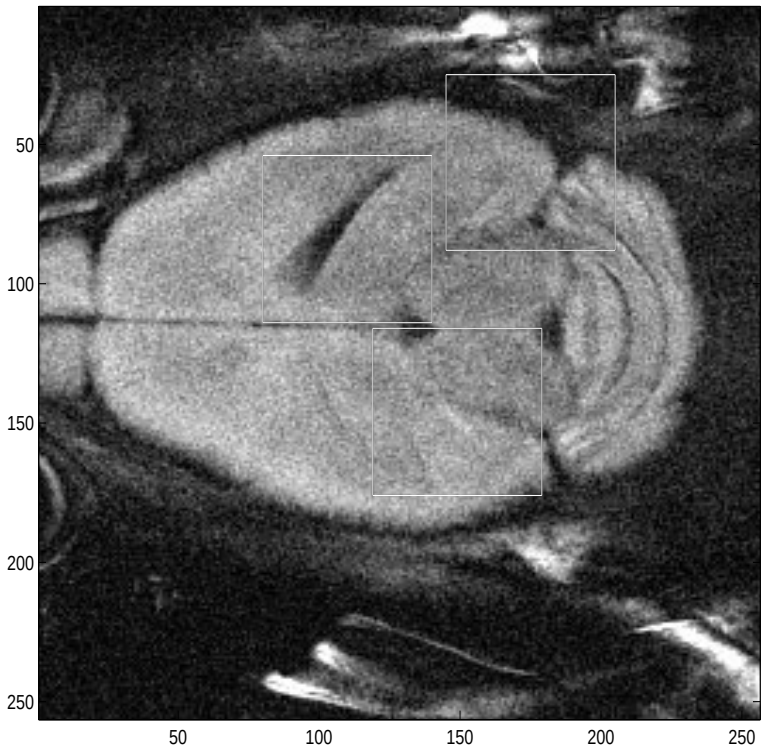


Figure 3.3: Three regions of interest are marked on the atlas image (16283) with rectangles. Relocation of facets is shown in Figure 3.4.

Especially encouraging is the segmentation of the left hippocampus. The model

found the border between hippocampus and ventricle correctly although the left ventricle was missing from the atlas image.

To examine in close details how the facets moved, we boxed three regions of interest on the atlas in Figure 3.2 and show the details of facet deformation in each of these boxes in Figure 3.4.

In particular we show the trace of the movement of facets in the last box, most facets traveled quite a distance to find the location that gives the best match.

We next applied the model to all slices of brain 16333. The resulting segmentation is illustrated in Figure 3.7.

We can directly use the output from the facet model to segment any part of the target image, given a manual segmentation of that part in the atlas.

3.3 Evaluation of Auto-segmentation

We wish to evaluate the uncertainty of the auto-segmentation procedure described in the above section. This uncertainty is a relative measure, because no segmentation can be made until an atlas segmentation is given. The subsequent auto-segmentation is then based on the manual segmentation; errors in the manual segmentation are thus propagated to the auto-segmentation.

Segmentation error is comprised of two components: the error of the facet model and the error of the initial segmentation. For example, let us assume the underlying (unobserved) true area of a region is t_i in the atlas, and t_I in the target, but the manual segmentation measured the atlas region to be m_i. Let $c_{iI}^{(k)}$ denote the area yielded by auto-segmentation of the target region i using the k-th manual segmentation of region i as the atlas. Proportionally, $c_{iI}^{(k)}$ should be expected to deviate from the true target area by the same ratio, yielding an area of $t_I \frac{m_i}{t_i}$. On the log scale, we can write

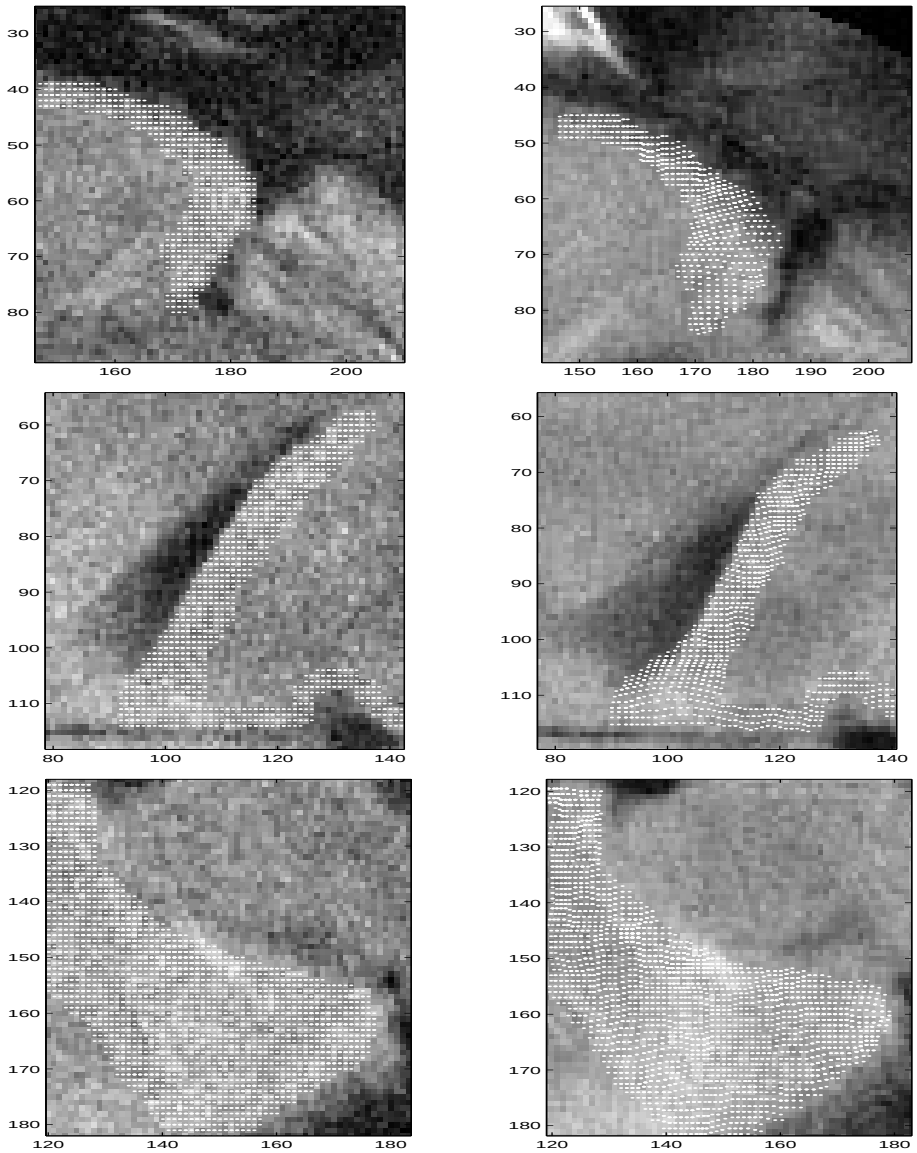


Figure 3.4: Facets before and after deformation. On the left, facets of interest in three regions are highlighted on the atlas image. On the right, the same facets have deformed from their original locations, their new locations are highlighted on the target image.

the above model as follows:

$$\log c_{iI}^{(k)} = \log t_I + (\log m_i^{(k)} - \log t_i) + e_{iI}^{(k)}, e \sim N(0, \delta^2), \tag{3.1}$$

$$i, I = 1, 2, ..., N \times R, \quad i \sim I, \quad k = 1, 2, ..., K.$$

The organization of the data is as follows: there are N images in all, each contains

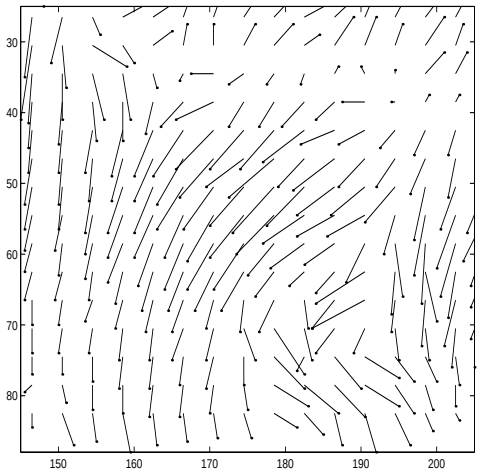


Figure 3.5: Deformation of facets in the rightmost box.

R regions of interest, each region of interest has been manually segmented by K experts. We pool all the regions of interest of all images together and label them 1, 2, ..., $N \times R$ (to give enough strength to the estimation here we do not distinguish which region belongs to which image. Otherwise a random-effects model could be fitted where we have image-specific and/or region-specific effects). The notation $i \sim I$ means that i and I are the same anatomical region in two images, so that each image can be the atlas of the other. The uncertainty of auto-segmentation is represented by δ.

In addition, we assume the measurements are normally distributed around the true value (on the log scale):

$$\log m_i^{(k)} = \log t_i + \epsilon_i^{(k)}, \qquad \epsilon \sim N(0, \Delta^2), \tag{3.2}$$
$$i = 1, 2, ..., N \times R, \quad k = 1, 2, ..., K$$

where Δ represents the magnitude of measurement inconsistency among experts.

Our dataset consists of: $N = 18$ images of mouse brains, $R = 4$ regions of interest measured by $K = 4$ experts. A correspondence is established among the images and regions so that there are 144 atlas-target pairs. With four sets of measurements we

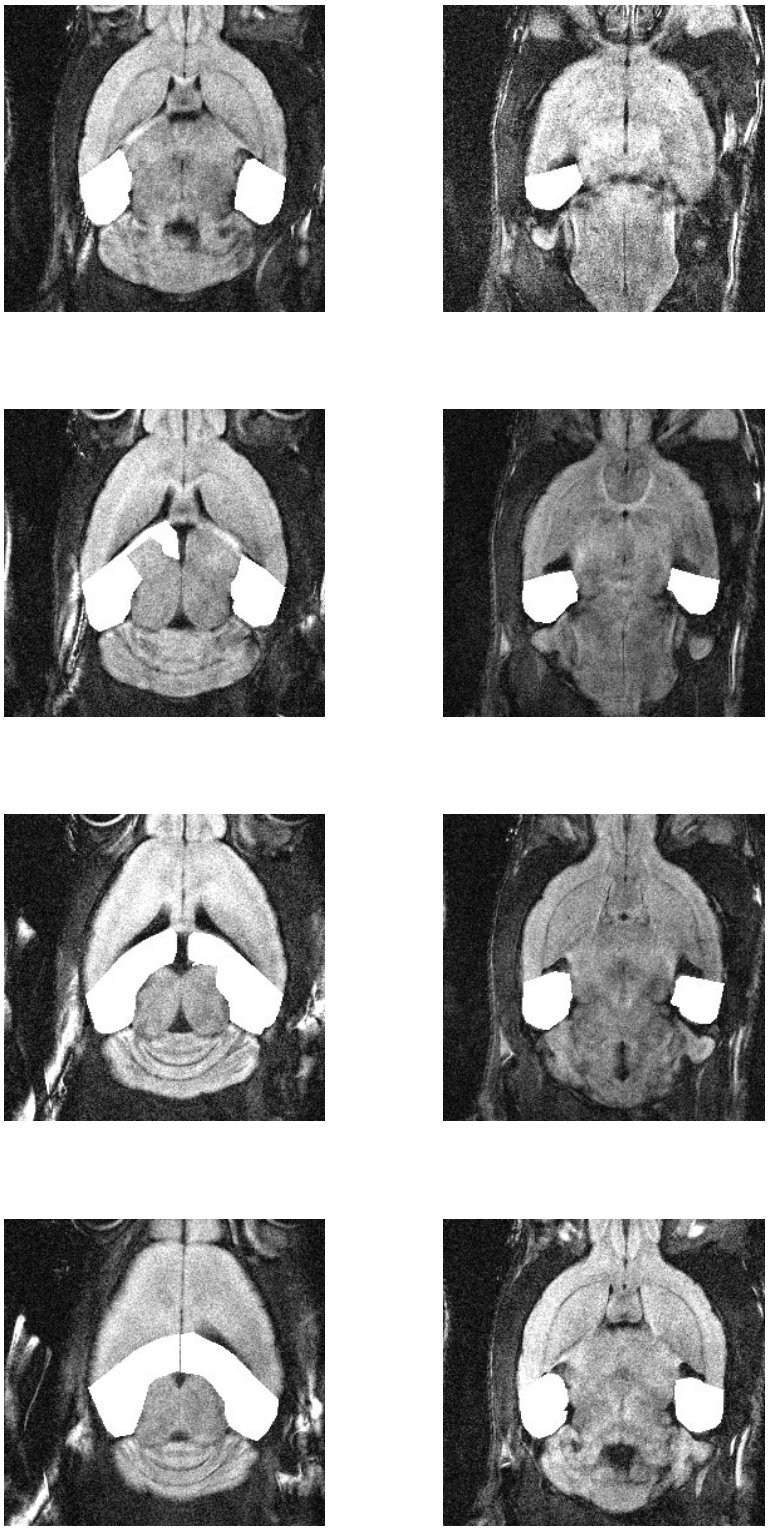


Figure 3.6: Manual segmentation of all eight slices in atlas brain 16283.

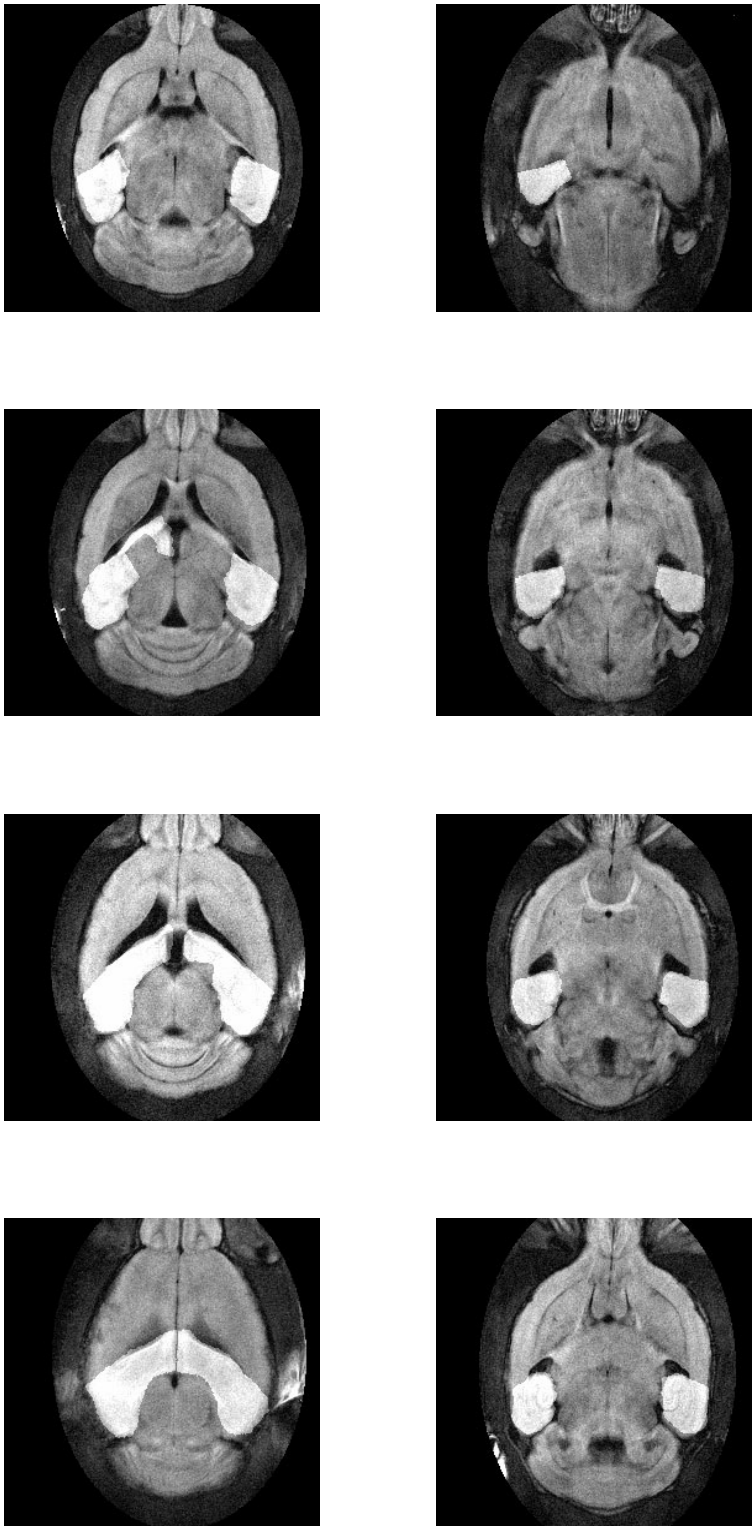


Figure 3.7: Auto-segmentation on all eight slices of target brain 16333 based on the manual segmentation of atlas shown in Figure 3.6 using facet model.

then have 576 values of c to estimate δ from. Similarly, there are $N \times R \times K = 288$ values of m to estimate Δ from. However, since t's are unobservable,we pool strength from (3.2) and (3.3) to account for the uncertainty there.

It follows that the likelihood function for the segmented volume data can be written:

$$\begin{aligned} p(\boldsymbol{c}, \boldsymbol{m}|\boldsymbol{t}, \Delta^2, \delta^2) &\propto \prod_{k=1}^{K}\prod_{i=1}^{N*R} N(m_i^{(k)} - t_i|0, \Delta^2) \\ &\quad \cdot \prod_{k=1}^{K}\prod_{i\sim I} N(c_{iI}^{(k)} - t_I - (m_i^{(k)} - t_i)|0, \delta^2) \qquad (3.3) \\ &\propto (2\pi\Delta^2)^{-\frac{KNR}{2}}(2\pi\delta^2)^{-\frac{K\cdot\#\{i\sim I\}}{2}} \exp(-\frac{1}{2\Delta^2}\sum_{k=1}^{K}\sum_{i=1}^{NR}(m_i^{(k)} - t_i)^2) \\ &\quad \exp(-\frac{1}{2\delta^2}\sum_{k=1}^{K}\sum_{i\sim I}(c_{iI}^{(k)} - t_I - (m_i^{(k)} - t_i))^2) \qquad (3.4) \end{aligned}$$

where $M \equiv \log m, T \equiv \log t, C \equiv \log c$

If we use non-informative priors for the parameters, *i.e.*

$$\pi(\Delta^2) \propto 1, \quad \pi(\delta^2) \propto 1, \quad \pi(\boldsymbol{t}) \propto 1,$$

the posterior will have the same form as the likelihood. It is then quite easy to write the full conditionals of the parameters:

$$\Delta^2|. \sim inv - \Gamma(\frac{KNR}{2} - 1, \sum_{k=1}^{K}\sum_{i=1}^{N*R}(m_i^{(k)} - t_i)^2) \qquad (3.5)$$

$$\delta^2|. \sim inv - \Gamma(\frac{K\cdot\#\{i\sim j\}}{2} - 1, \sum_{k=1}^{K}\sum_{i\sim I}(c_{iI}^{(k)} - t_I - (m_i^{(k)} - t_i))^2) \qquad (3.6)$$

$$t_i|. \sim N(\mu_i, \sigma_i^2) \qquad (3.7)$$

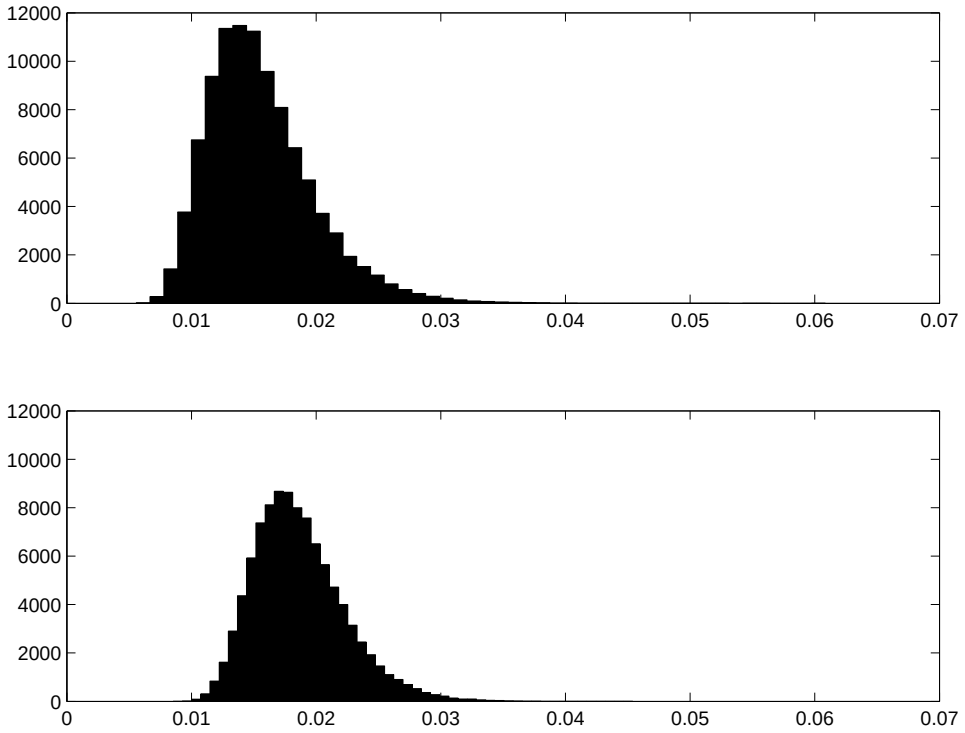


Figure 3.8: Top: posterior sample of Δ^2 – between-expert variance for manual segmentation. Bottom: posterior sample of δ^2 – variance of auto-segmentation using facet model.

where

$$\sigma_i^2 = (\frac{K}{\Delta^2} + \frac{2K\#\{j : j \sim i\}}{\delta^2})^{-1}, \tag{3.8}$$

$$\mu_i = \sigma_i \cdot \{\sum_{k=1}^{K} m_i^{(k)} + \frac{1}{2\delta^2} \sum_{k=1}^{K} \sum_{j:j \text{ is atlas of} i} (c_{ji} - (m_j^{(k)} - t_j)) +$$

$$\frac{1}{2\delta^2} \sum_{k=1}^{K} \sum_{l:l \text{ is target of} i} (t_l + m_l(k) - c_{il})\} \tag{3.9}$$

Sampling from the posterior using a Gibbs sampler is straightforward. The chain converges rapidly. We ran the sampler for $100,000$ iterations, discarded the first 1000 iterations as burn-in. Histograms of the posterior of Δ^2 and δ^2 are shown in Figure 3.8. The posterior distribution of Δ^2 and δ^2 look very similar.

This result can be summarized by noting that the ratio of the mean value of Δ to

δ is 0.96. For both manual and automated segmentations, there is approximately a 90% chance that the area of either segmentation whil fall with 25% of the true area.

As a by-product from the Gibbs sampling, we also get estimates of the true areas $\boldsymbol{t}$.

It can be concluded that in our dataset of mouse brain segmentations, the magnitude of uncertainty of auto-segmentation based on the facet model is similar to the inconsistency of the human experts. The result that the facet model performs about as well as human experts in this aspect is encouraging. It is worth noting that in this particular study, we are only concerned with the absolute volume of hippocampi, thus one segmentation is deemed equivalent to another as long as they give the same volume. In other applications, shape differences between segmentations may be important for evaluation.

Chapter 4

Inter-subject Registration of Human Brain MRIs

A major portion of medical imaging research is devoted to the processing and analysis of human brain images. Today a full range of brain imaging apparati are available for both the study of brain anatomy (MRI) and for monitoring of brain functional activities (PET, SPECT, fMRI). Brains are being imaged in increasing resolution and precision as well as quantities, yet the corresponding processing and analysis have not kept pace. Automated tools are in greater demand than ever.

Among the variety of problems the brain imaging community regularly encounters, some are considered solved, for example intra-subject alignment (intra-modality and multi-modality), some problems have never had a unique solution that everyone agrees on, for example inter-subject registration, and tissue classification (*i.e.* segmentation). In this chapter we make an attempt to tackle some of these problems using the facet model.

4.1 Issues on Evaluating Registration Results

Below we address the issue of evaluating the performance of our registration.

There does not exist a standard criterion against which all inter-subject registration can be validated. This lack of evaluation frameworks results from the absence of a ground truth.

There have long been debates on whether all human brains are alike. Advocates of this view argue that if evidence is lacking, the only reason would be the absence of a powerful registration tool capable of registering all areas of a brain properly. Antagonists deny the existence of a unique one-to-one correspondence between the brains of any two individuals. They believe that if there is an algorithm that matches two persons' brains perfectly, then it must have a mechanism that creates extra structures that do not really exist. The following quote describes well the stand of the thesis on this issue:

> [...] *all normal brains, at least at a certain level of representation, have the same topological structure, but may differ in shape details.*
>
> –Bajcsy and Kovacic (1989)

We illustrate with some brain images below.

Figure 4.1 are a few transaxial slices extracted from six images. By visual inspection, they appear quite different. It can be easily noticed that although the overall structure of the brains are alike, the details of the cortex vary greatly from individual to individual, and increasingly so towards the top of the head.

Some methodological papers use an invasive procedure (for example inserting markers) or synthetic images to inform about the truth. An obvious disadvantage of the former is its invasiveness, especially when the imaged object is a human brain *in vivo*. The disadvantage of the latter is the bias generated by the nature of the synthetic deformation. Many other papers employ a variety of non-standard evaluation techniques. Few address the issue of evaluation of inter-subject registration. Informal methods include visual inspection of mosaic images (illustrated in Figure

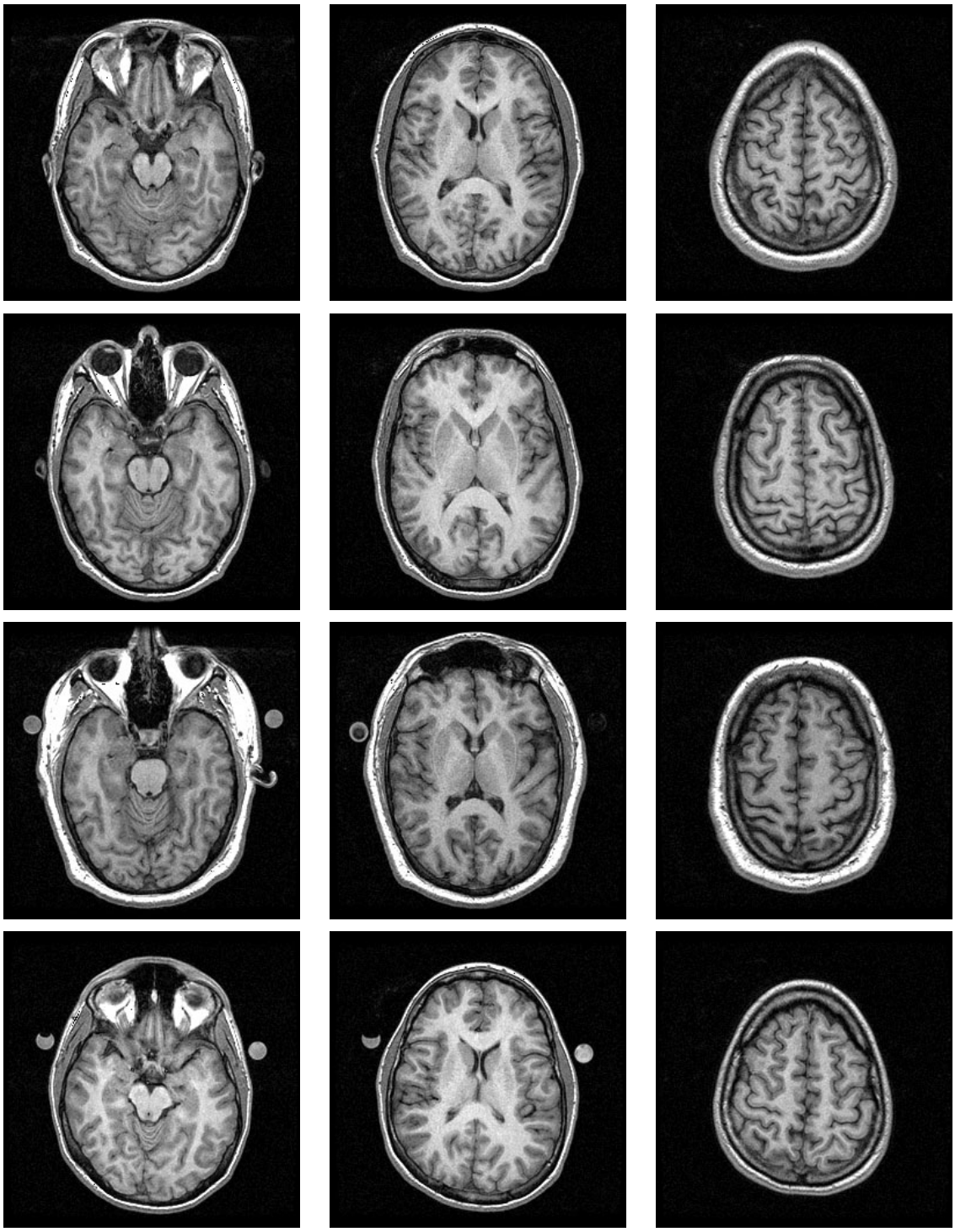


Figure 4.1: Comparison of three corresponding transaxial slices from four brains at random. Each row corresponds to one brain. Slices at the right are closer to the top of the brain.

4.2) and sum of deformed images, *etc.* More quantitative methods include calculating the difference between landmarks and comparing segmentation results, as in Hellier et al. (2001), Rogelj et al. (2002), Lane et al. (1999)

We use several of these techniques to evaluate our registration results below. For comparison, we also registered our population of brain images using the popular software SPM (Statistical Parametric Mapping)(Friston et al., 1995; Ashburner and Friston, 1997, 1999). SPM is developed by members and collaborators of the Wellcome Department of Cognitive Neurology and has become widely used due to its ease of use. We compare the results from the facet model to those from SPM.

4.2 Registration Results

We extend our algorithm to process 3D images. The extension is straightforward from the 2D version. The parameters are fixed as explained in Section 2.6.

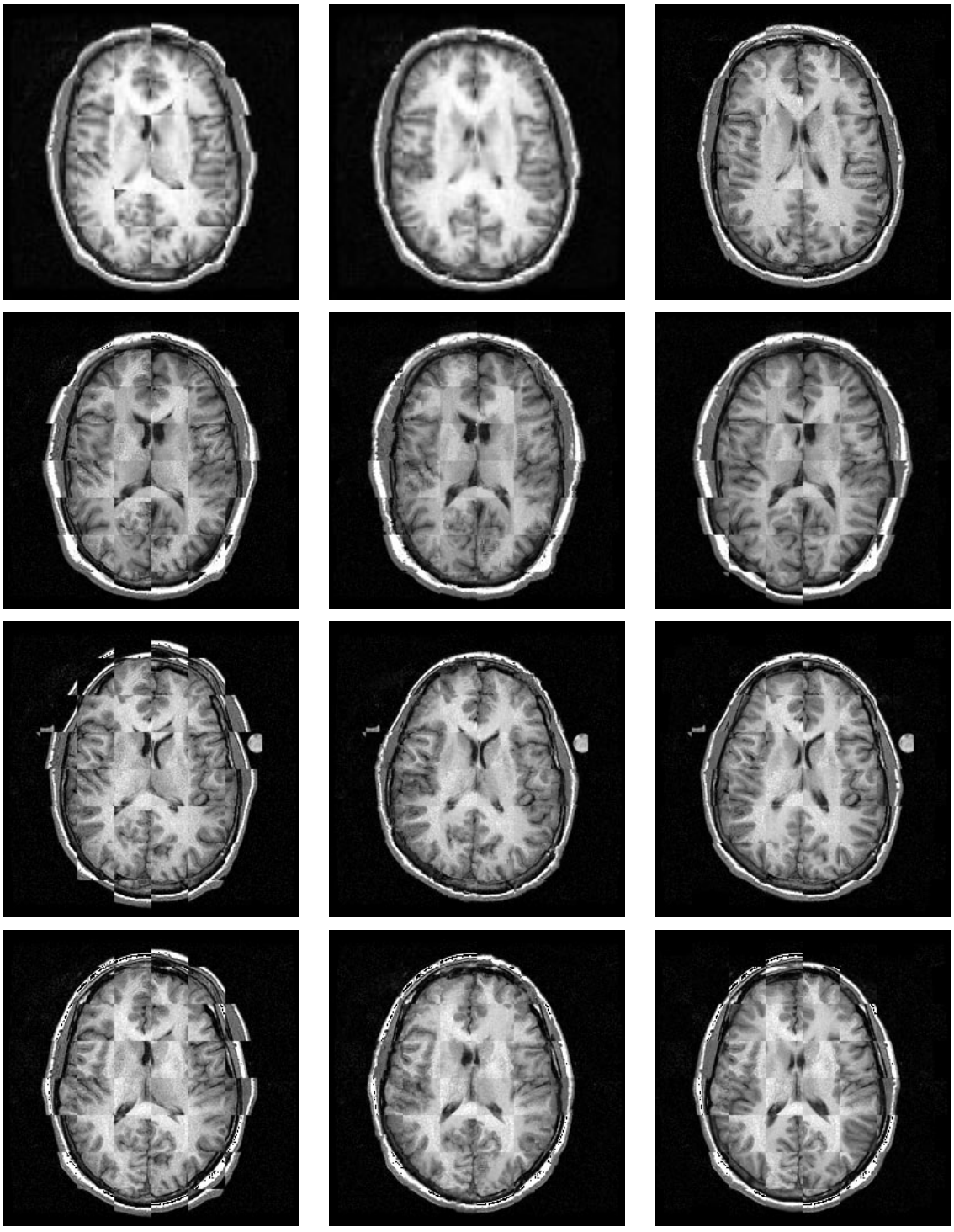
We have MR images from six patients. Each is a T1-weighted image of dimension $256 \times 256 \times 60$. The voxels are each of size $0.859 \times 0.859 \times 3.00 mm^3$. We arbitrarily assign one of the images to be the atlas image. We apply our algorithm to obtain the MAP estimate of the deformation fields that deform the atlas to match the targets. Then we "warp" the atlas according to each deformation field to obtain the "warped" version that is "similar" to the target brain. We will show the result of registering the atlas to the five target images.

4.2.1 Qualitative – Mosaic

Shown below are mosaics (*a.k.a.* "chessboard" images) for visualizing the mismatch between two images. This technique is particularly effective in examining how well the tissue boundaries line up. Poorly matched boundaries will appear as broken lines in a mosaic image. Below we overlay images from two brains before registration,

then re-do the mosaic for the same images after facet registration, and for after SPM registration.

SPM results are obtained by using $6 \times 7 \times 6$ basis functions, 12 nonlinear iterations (with SPM options "medium regularization, no brain mask, do not mask object image, bilinear interpolation").



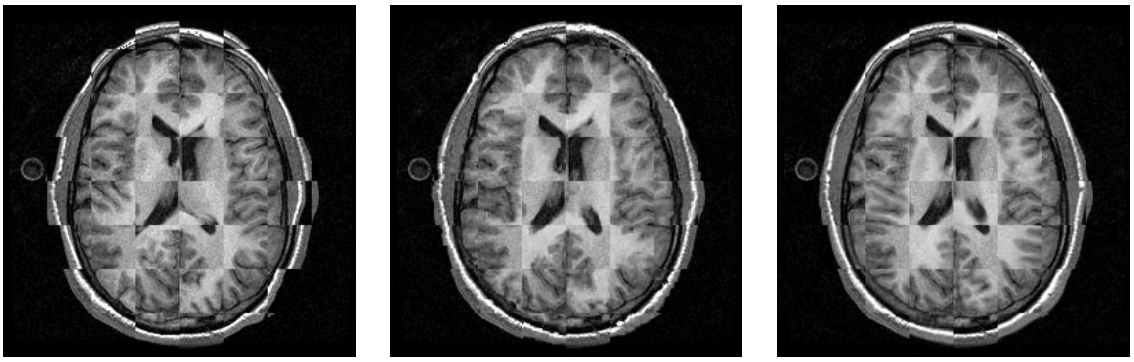


Figure 4.2: Mosaic images of one atlas image and five target images before and after registration. Left column: slice 32 from original atlas brain overlaid with that from target. Middle column: warped atlas slice (after facet model registration) and target. Right column: warped atlas slice (after SPM registration) and target.

By comparison facet model did a better job on the overall mapping.

4.2.2 Qualitative – Facet Movement

Here, we examine the movement of some facets when the atlas is deformed to match the targets. In Figure 4.3, several facets in the CSF (ventricle) and in the subcortical region were highlighted in the atlas. Their estimated locations in the targets spread through three consecutive slices in the target brain. From the distribution of the facets we can see that the target brain has a slight angle of rotation around the saggital axis.

In Figure 4.4 we can see how the facet model stretched a small ventricle in the atlas brain to match a much larger one in the new brain. If we look back at the second row of Figure 4.2 and compare the facet warping and the SPM warping, we can see that SPM lacks the local elasticity that the facet model has to accommodate dramatic volume change in a small area.

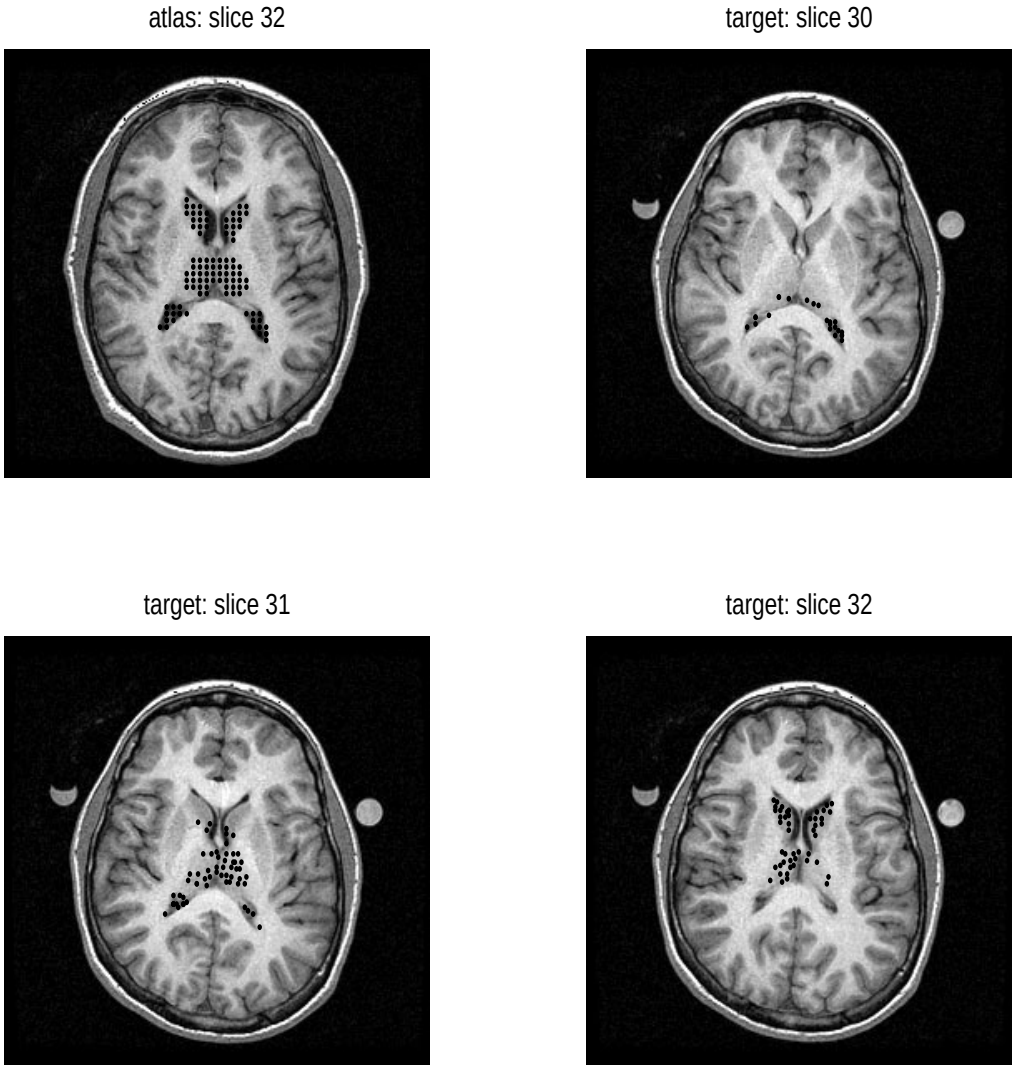


Figure 4.3: Top left: facets of interest are shown on slice 32 of the atlas volume. Top right, bottom left, bottom right: the same facets are located in consecutive slices in target volume. The distribution of the new locations of the facets suggests that the target brain has a slight angle of rotation around the saggital axis compared to the atlas brain.

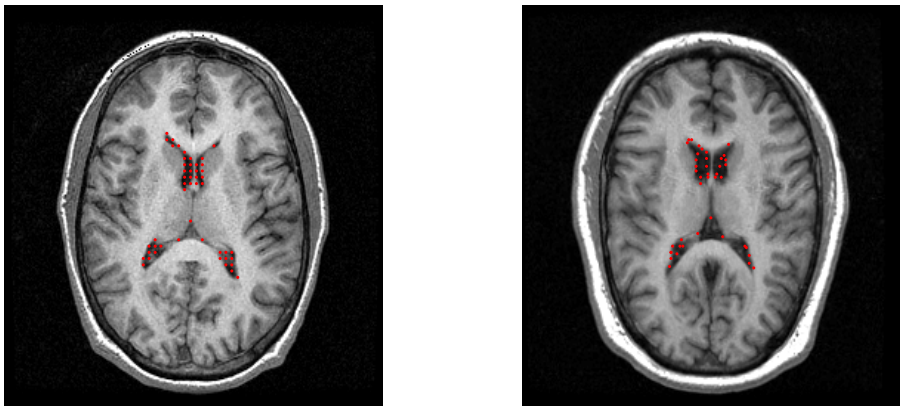


Figure 4.4: Left: facets of interest are highlighted in the atlas volume. Right: the same facets are located in the corresponding slice of the target volume. Facets have deformed to accommodate the expansion of CSF.

4.2.3 Quantitative – Segmentation Based

To evaluate brain registration from a practical aspect, some people look at the segmentation results they produce. The basic idea is: if an algorithm can warp brain A into the same shape as brain B, then the segmentation of the warped brain A should be the same as segmentation of brain B.

However, this approach is flawed because there does not exist a unique true segmentation of brains. Segmentations done by human observers are inconsistent. Segmentations done by computer tools are themselves high-level operations that have yet to be validated.

Despite these problems, we will attempt a segmentation-based-validation as a starting point for quantitative evaluations.

First, an exploratory analysis was carried out. We used an automated brain segmentation tool provided in the SPM package to segment a brain into three complimentary tissue types: gray matter, white matter and cerebrospinal fluid (CSF). We were mostly interested in the first two tissue types. Figure 4.2.3 illustrates the segmentation of brain into white matter and gray matter regions.

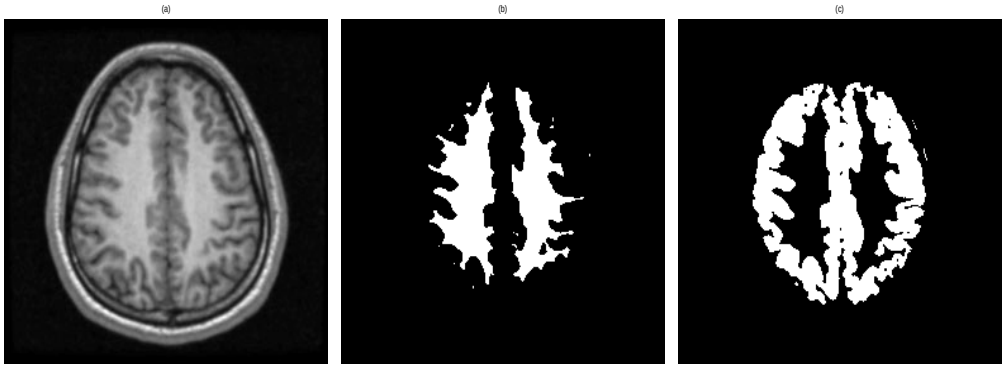


Figure 4.5: Segmentation of brain into white matter and gray matter. Left: original slice. Middle: white matter. Right: gray matter.

We refer to the segmentation of the original target brain as the "true" segmentation, and that of the warped atlas as the "warped" segmentation. Two sets of quantities were calculated for each pair of target and warped atlases: the percentage of overlapped voxels and the percentage of misclassified voxels. These quantities are defined as follows:

$$\text{overlap}\% = \text{overlap}(\text{true}, \text{warped})/\text{true} \times 100\% \tag{4.1}$$

$$\text{misclassified}\% = (\text{warped} - \text{overlap})/\text{true} \times 100\% \tag{4.2}$$

In the small exploratory study, we showed that the facet model yields a higher percentage of overlap and lower percentage of misclassification compared to SPM registration.

Second, we conducted a similar analysis using regions of interest (ROIs) specified by a human brain imaging expert. The ROIs we use are shown in Figure 4.6.

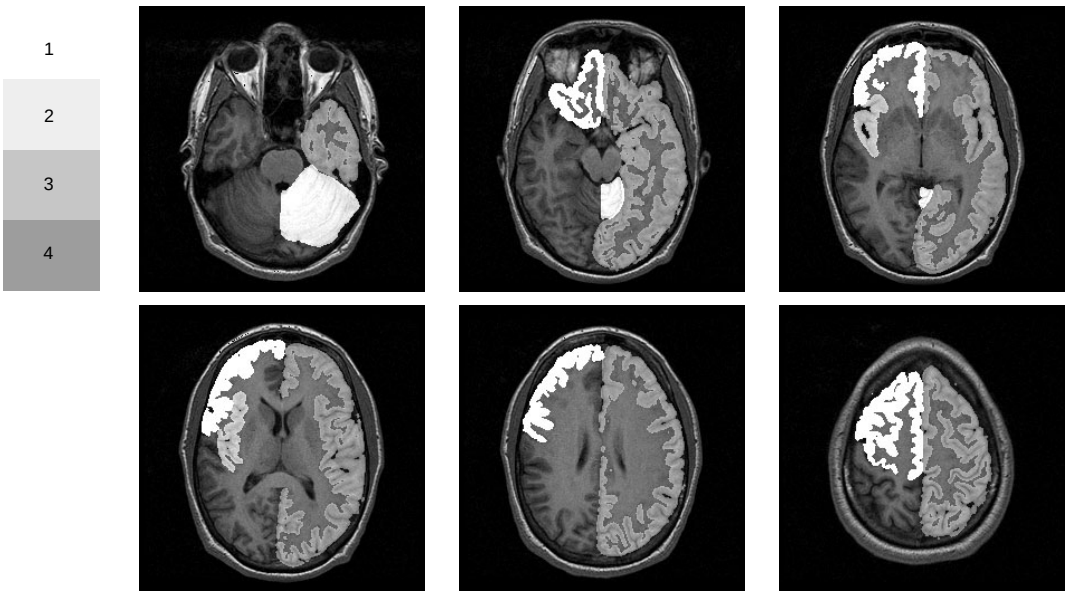


Figure 4.6: Regions of interest across transaxial slices marked on the atlas by an expert. On each slice, regions are highlighted with different shades. Shade 1: right frontal gray matter. Shade 2: left cerebellum. Shade 3: right insula gray matter. Shade 4: left cortical gray matter.

We have expert segmentations for all six brains. Since one brain is used as the atlas and the remaining five as targets, there were a total of five pairs of targets and warped atlases. We show the discrepancy between target ROIs and warped atlas ROIs for one of the targets in Figure 4.7. The study in Figure 4.7 is a comparison to the target segmentation in Figure 4.6, and we see that the discrepancies mostly occured as a result of differences in shape details between atlas and target.

Once again, it should be noted that the segmentations obtained from experts should not be regarded as the ground truth. In other words, a warped ROI can be good even though it is slightly different from the target ROI.

Percentages of overlap and of misclassification are tabulated for each ROI for each target/facet model against SPM.

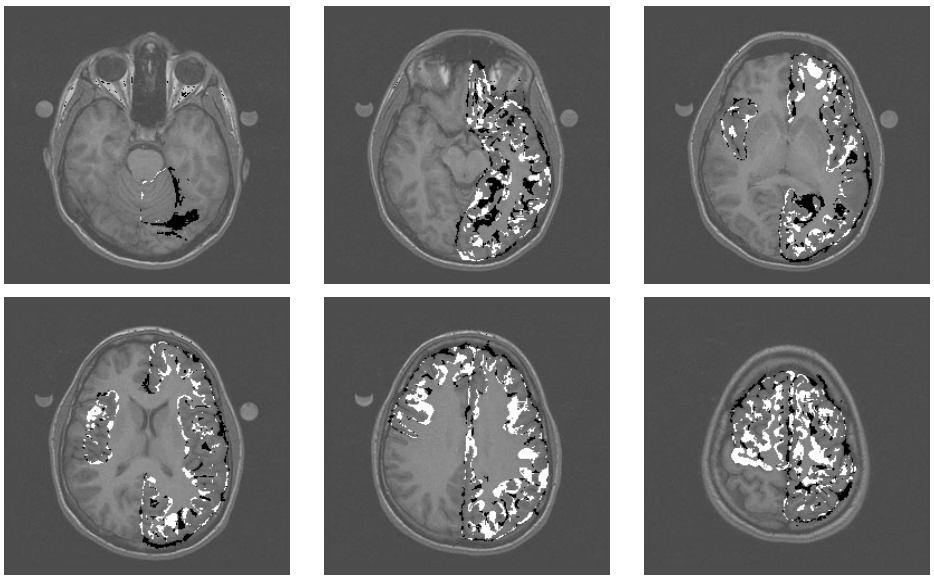


Figure 4.7: Discrepancy between target ROIs and warped atlas ROIs. Bright voxels belong to target ROIs but not to warped atlas ROIs. Dark voxels are the opposite (misclassified voxels). ROIs involved in each plot (from left to right, top to bottom): left cerebellum; left cortical gray matter; left cortical and right insula; left cortical and right insula; left cortical and right frontal; left cortical and right frontal.

target	Facet Model		SPM	
	overlap (%)	misclassify (%)	overlap (%)	misclassify (%)
1	81.2	24.6	76.3	30.8
2	77.9	38.7	73.2	40.5
3	74.5	18.5	73.6	21.1
4	75.4	23.0	71.3	31.9
5	75.7	20.7	74.8	24.6
mean	76.9	25.1	73.8	29.8

Table 4.1: Left cortical gray matter.

We have underlined the cases where the facet model did not perform as well as SPM. Note that it is not necessarily undesirable when the overlap percentage is slightly smaller while the misclassification percentage shows a significant improvement, and vice versa. While it is clearly more desirable to achieve both, that can be hard because, when the percentage of overlap increases, the corresponding percentage of misclassification tends to increase as well.

target	Facet Model		SPM	
	overlap (%)	misclassify (%)	overlap (%)	misclassify (%)
1	76.1	27.5	71.2	32.9
2	71.8	39.5	70.3	40.3
3	64.5	18.6	68.2	26.7
4	68.7	29.4	68.6	37.8
5	66.8	22.6	67.8	32.5
mean	69.6	27.5	69.2	34.0

Table 4.2: Right frontal gray matter.

target	Facet Model		SPM	
	overlap (%)	misclassify (%)	overlap (%)	misclassify (%)
1	73.2	16.4	68.7	21.2
2	82.8	28.2	75.1	32.4
3	73.5	8.4	66.0	20.3
4	75.3	16.1	68.1	23.4
5	73.1	20.3	65.7	25.9
mean	75.6	17.9	68.7	24.6

Table 4.3: Right insula gray matter.

target	Facet Model		SPM	
	overlap (%)	misclassify (%)	overlap (%)	misclassify (%)
1	95.6	14.9	88.6	12.9
2	83.9	21.9	81.0	36.9
3	94.6	17.0	91.0	13.2
4	92.7	26.8	82.4	31.5
5	87.8	42.0	94.8	28.1
mean	90.9	24.5	87.6	24.5

Table 4.4: Left cerebellum.

From the mean overlap and misclassification fraction we can see that the facet model on average gave more accurate segmentations for all the above ROIs. We say "more accurate" in the sense that both the overlap fractions are larger and the misclassification fractions are smaller (except they are equal in the left cerebellum (Table 4.4)). Pairwise t tests for all 20 cases also indicated that the overlap fractions

are significantly higher for the facet model than for SPM, while the misclassification fractions are significantly lower (both on a level of 0.005). On average, there is a 5% improvement in the overlap fraction from SPM to the facet model. The improvement is 17% for the misclassification fraction.

The above results are evaluated by the two fractions separately. Since they are correlated with each other, we propose an *ad hoc* measure for evaluating the consistency between two sets of segmentations: the overlap fraction minus the misclassification fraction. It follows that the larger this quantity is, the more consistent the two segmentations. A perfect match would result in a 1, while a bad match could result in a negative value. Under this combined measure, there are only two cases out of 20 (targets 3 and 5 in Table 4.4) where the facet model model did not outperform SPM. A pairwise t test showed that the facet model scored significantly higher than SPM. There is a 8% improvement from SPM to the facet model on average in terms of this measure.

4.2.4 Quantitative – Cross-Covariance

A commonly used measure of match is given by normalized cross-covariance (Cox, 1995):

$$C(f,g) = \frac{n \cdot \sum_A f_{ij} \cdot g_{ij} - \sum_A f_{ij} \cdot \sum_A g_{ij}}{n\sqrt{(\sum_A f^2 + \frac{n-2}{n^2}(\sum_A f)^2)(\sum_A g^2 + \frac{n-2}{n^2}(\sum_A g)^2)}} \tag{4.3}$$

where f and g are from two matched images. We apply the measure to our registration results, which means calculating the normalized cross-covariance of voxel intensities between pairs of "warped atlas" and "target". A 3D cross-covariance map shows where the two images match well locally in terms of correlation.

We can see that for both registrations, the area close to the scalp was matched

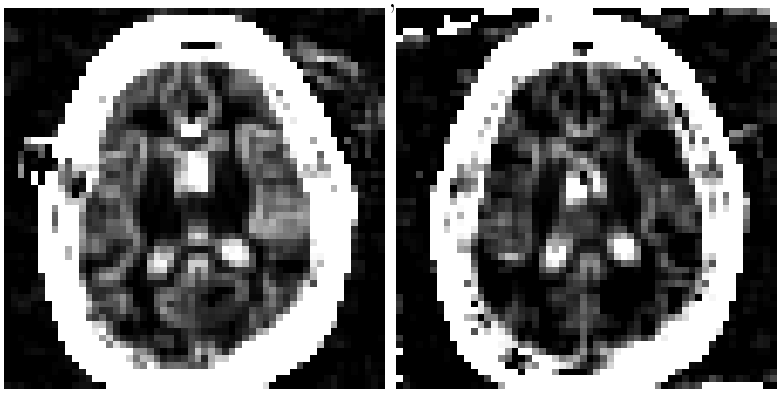


Figure 4.8: Cross-covariance maps that correspond to slice 32 of the brains. On the left is facet registration, on the right is SPM registration. Bright areas corresponds to a large covariance, thus a good local feature match.

well. Areas close to the ventricle were also matched comparatively well, although the match appears better with the facet model. Overall, the cross-correlation of the facet model was higher than it was for the SPM registration.

4.3 Significance of Results

In this chapter, we presented a tool that non-rigidly registers 3D brain images of one individual to that of another. The tool registered randomly picked MR brain images successfully. We have also shown that the tool out-performs the "industrial standard" software package SPM in several ways. First, the spatial transform of the tool offers more local elasticity and so is suitable for the non-rigid nature of cross-subject brain variations. As a consequence, the warp is more accurate by visual inspection. Second, it yields a segmentation closer to that made by an expert. The drawback of this algorithm is its computational burden.

Chapter 5

Discussion

5.1 Multi-resolution Modeling

Going back to the multi-resolution maximization scheme described in Section 2.5, let us imagine the case where we have $L+1$ levels: $x^{(0)}$ is the finest level, $x^{(1)}$ is the next coarse level, $x^{(L)}$ denotes the coarsest level. Facets in level l, $x^{(l)}$, are linked to their parent facet(s) in $x^{(l+1)}$. Recall the conditional distribution of each child facet on its parent facet from the coarser level was given in (2.13). In matrix form, if $x^{(l)}$ is composed of n_l facets, then the parent-child relationship between level l and level $l+1$ can be represented by an n_l by n_{l+1} matrix A, where the (i_l, i_{l+1})-th entry $A_{i_l,i_{l+1}} = \gamma_{i_l i_{l+1}}$ (as defined in (2.2)) if i_{l+1} is the parent of i_l, and zero otherwise. The value of each entry can be determined as stated in Section 2.6. Then we can re-write the coarse to fine conditional as

$$p(x^{(l)}|x^{(l+1)}) = N(A^{(l)}x^{(l+1)}, \Gamma^{(l)}) \tag{5.1}$$

where $\Gamma^{(l)} = \tau^{(l)}I$ assuming independent errors.

We also claimed that at each resolution, the probability distribution on the facets form a M.r.f. as specified in (2.3). We slightly modify the original setting to give a

positive potential to singleton cliques, such that

$$p(x^{(l)}) = N(\mu^{(l)}, (Q^{(l)})^{-1}), l = 0, 1, ..., L \tag{5.2}$$

where each $Q^{(l)}$ is similarly defined in level l (*e.g.* Section 2.2), except that its diagonals are now non-zero to avoid impropriety.

It is obvious that the marginals in (5.2) are incompatible with the conditional specification in (5.1).

Below we derive marginals of finer levels based on the coarse level and the conditional distribution (5.1):

$$\begin{aligned} p^*(x^{(l)}) &= \int p(x^{(l)}|x^{(l+1)}) \cdot p*(x^{(l+1)}) dx^{(l+1)} && (5.3) \\ &= N(\mu^{(l)}, \Sigma^{(l)}) && (5.4) \\ &= N(\mu^{(l)}, \Gamma^{(l)} + A^T \Sigma^{(l+1)} A) && (5.5) \end{aligned}$$

$\Sigma^{(L)} = (Q^{(L)})^{-1}$. All subsequent Σ's can be derived recursively.

If we look at the n_l by n_l matrix $A^T \Sigma^{(l+1)} A$, an entry (i, j) is non-zero if i's parent and j's parent are siblings. This has an intuitive explanation. If two facets, say 1 and 2, are correlated in a coarse level, then in the next finer level, all child facets of 1 will be correlated with all child facets of 2. Considering the whole hierarchy, a pair of facets cannot be conditionally independent as long as there exists a sibling pair between their ancestors.

Thus, the more levels we go in the fine direction, the larger a facet's neighborhood. Such a correlation structure encourages smooth and elastic deformation field more than a single-level nearest-neighbor system does. However, it also exhibits a somewhat blocky behavior in finer levels. See Figure **??** for an illustration.

In his recent thesis, Ferreira proposed the Hidden Resolution Model (Ferreira, 2002). Using Jeffrey's rule of conditioning (Jeffrey, 1992), he revised the marginal distribution for each level following Diaconis and Zabell (1982) as $p^*(H) = \sum p(H|E_i) p^*(E_i)$.

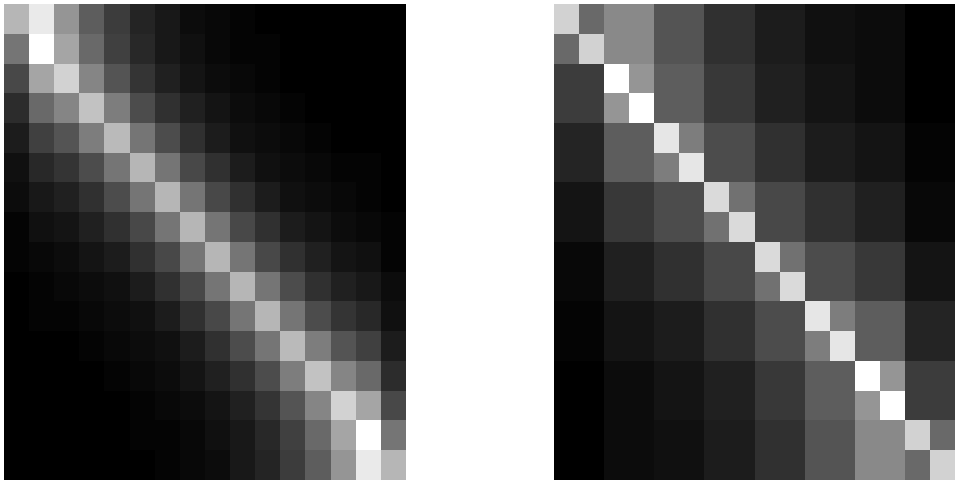


Figure 5.1: Left: correlation structure of a single-level nearest-neighbor system in one dimension; right: adding a coarse level.

The revised probability distribution $p^*(.)$ is called the Hidden Resolution Model. It revises the marginal distribution at each level (except the coarsest) while preserving the conditionals. Ferreira has shown that HRM exhibits many nice properties in spatial applications.

Other multi-resolution modeling that can be considered for image applications include Goutsias (1994), Luettgen et al. (1993), Higdon and Lee (2001), *etc.*

5.2 Parameter Estimation

Throughout an image population, we should be able to characterize the statistical behavior of the deformation of each facet. Such a characterization can be done with an inhomogeneous model. In such a model, the values of the γ's would be allowed to vary across the image. In theory, the re-parameterization of Aykroyd and Zimeras (1999) can be adopted, without any training, to place an equivalent of a Markov random field hyper-prior on these γ's). The Geman-McClure method (Geman and McClure, 1987) could also be used on a set of toy images to estimate γ simultaneously with the optimization of x. Preliminary results indicate that the estimate of γ so

obtained is visibly smaller in areas where greater deformation is expected. With training images, a calibrated prior can be estimated for each image class. In image regions that remain relatively constant throughout the image class, the calibrated prior gives more weight to the location component, so that facets in that region are constrained to be more "grid-like". In regions that undergo greater variation across images, the calibrated prior allows the corresponding facets more deformation, thus resulting in a more adaptive match than a homogeneous prior would.

We can also use MCMC to sample γ. However, the normalizing constants in Gibbs form (2.2) also contain γ, which complicates exact sampling. To obtain a reasonable sample, we may adopt sampling techniques from Aykroyd (2000), Meng and Wong (1996), Gelman and Meng (1998), Johnson (1998), Geyer and A.Thompson (1992). A final option to avoid sampling normalizing constant is to sample γ from a pseudolikelihood (Besag, 1975).

Bibliography

Ashburner, J. and Friston, K. J. (1997). Multimodal image coregistration and partitioning - a unified framework. *Neuroimage* **6(3)** 209–217.

Ashburner, J. and Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping* **7(4)** 254–266.

Aykroyd, R. G. (2000). Approximations for gibbs distribution normalising constants. Tech. rep., University of Leeds.

Aykroyd, R. G. and Zimeras, S. (1999). Inhomogeneous prior models for image reconstruction. *Journal of the American Statistical Association* **94** 934–946.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing* **46** 1–21.

Bajcsy, R., Lieberson, R. and Reivich, M. (1983). A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *J. Comput. Assisted Tomogr.* **7** 618–625.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (disc: P41-66). *Statistical Science* **10** 3–41.

Besag, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B, Methodological* **34** 75–83.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* **36** 192–236.

Besag, J. E. (1975). Statistical analysis of non-lattice data. *The Statistician* **24** 179–196.

Besag, J. E. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B, Methodological* **48** 259–279.

Christensen, G., Rabbitt, R. and Miller, M. (1996). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing* **5(10)**.

Cox, G. (1995). Template matching and measures of match in image processing.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.

Diaconis, P. and Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association* **77** 822–830.

Dryden, I. L. (2001). *Statistical shape analysis in high-level vision*. Springer-Verlag.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley.

Ferreira, B. (1981). *Optimal Registration of Deformed Images*. Ph.D. thesis, University of Pennsylvania.

Ferreira, M. A. R. (2002). *Hidden Resolution Model*. Ph.D. thesis, Duke University.

Friston, K. J., Ashburner, J., Poline, J. B., Frith, C. D., Heather, J. D. and Frackowiak, R. S. J. (1995). Spatial registration and normalisation of images. *Human Brain Mapping* **2** 165–189.

Fritsch, D. (1993). *Registration of Radiotherapy Images using Multiscale Medial Description of Image Structure*. Ph.D. thesis, Department of Computer Science, University of North Carolina at Chapel Hill.

Gee, J., Reivich, M. and Bajcsy, R. (1993). Elastically deforming 3d atlas to match anatomical brain. *Journal of Computer Assisted Tomography* **17(2)**.

Gelman, A. and Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* **13** 163–185.

Geman, S. and Geman, D. (1993). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Journal of Applied Statistics* **20** 25–62.

Geman, S. and McClure, D. E. (1987). Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute* **52** 5–21.

Geyer, C. J. and A.Thompson, E. (1992). Constrained Monte Carlo maximum likelihood for dependent data (Disc: p683-699). *Journal of the Royal Statistical Society, Series B, Methodological* **54** 657–683.

Goutsias, J. (1994). Stochastic pyramids for multiscale signal synthesis and analysis. *International Conference on Image Processing* **(1)**.

Grenander, U. (1994). *General pattern theory. A mathematical study of regular structures.* Oxford University Press.

Haralick, R. M. and Shapiro, L. G. (1991). Glossary of computer vision terms. *Pattern Recognition* **24** 69–93.

Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Goualher, G. L., Collins, L., Evans, A., Malandain, G. and Ayache, N. (2001). Retrospective evaluation of inter-subject brain registration. *Lecture Notes in Computer Science* **2208**.
URL `http://citeseer.nj.nec.com/hellier01retrospective.html`

Higdon, D., Bowsher, J. E., Johnson, V. E., Turkington, T., Gilland, D. and Jaszczak, R. (1997). Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging* **16**.
URL `http://citeseer.nj.nec.com/higdon97fully.html`

Higdon, D. and Lee, H. (2001). A bayesian approach to characterizing uncertainty in inverse problems using coarse and fine scale information. Tech. rep., Duke University.
URL `http://citeseer.nj.nec.com/415836.html`

Jeffrey, R. C. (1992). *Probability and the art of judgment.* Cambridge University Press.

Johnson, V. E. (1998). Posterior distributions on normalizing constants. Tech. Rep. 98-26, Duke University.

Kendall, D. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society* **16** 81–121.

Laading, J. K. (2000). *Practical Methodology for Inclusion of Modality-specific Modifications in a Hierarchical Bayesian Deformation Model.* Ph.D. thesis, Duke University.

Lane, C., Atkins, S. and Liddle, P. (1999). Evaluation of the spatial normalization of brain images. http://www.cs.sfu.ca/people/Faculty/Atkins/papers/wbir.doc.

Lee, H., Higdon, D., Bi, Z., Ferreira, M. and West, M. (2000). Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media. Tech. Rep. 00-35, Duke University, ISDS.
URL http://citeseer.nj.nec.com/lee00markov.html

Lifshitz, L. M. and Pizer, S. M. (1987). A multiresolution hierarchical approach to image segmentation based on intensity extrema. Tech. rep., Dept. of Computer Science and Radiology, University of North Carolina, Chapel Hill.

Luettgen, M., Karl, W., Willsky, A. and Tenney, R. (1993). Multiscale representations of markov random fields. *IEEE Trans. Signal Process.* **41(12)**.
URL http://citeseer.nj.nec.com/luettgen93multiscale.html

Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis* **2(1)**.
URL http://citeseer.nj.nec.com/maintz98survey.html

McCulloch, C. (1999). *High-level Image Understanding Through Bayesian Hierarchical Models*. Ph.D. thesis, Duke University.

Meng, X. and Wong, W. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica* **6** 831–860.

Nestares, O. and Heeger, D. J. (2000). Robust multiresolution alignment of mri brain volumes. *Magnetic Resonance in Medicine* **43** 705–715.
URL http://citeseer.nj.nec.com/nestares00robust.html

Rogelj, P., Kovacic, S. and Gee, J. C. (2002). Validation of a non-rigid registration algorithm for multi-modal data. http://citeseer.nj.nec.com/507648.html.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B, Methodological* **63** 325–338.

Viola, P. A. and III, W. W. (1995). Alignment by maximization of mutual information. *International Conference on Computer Vision* .
URL http://citeseer.nj.nec.com/viola95alignment.html

Wang, Y. and Staib, L. H. (1998). Elastic model based non-rigid registration incorporation statistical shape information. In *MICCAI*.
URL http://citeseer.nj.nec.com/wang98elastic.html

Wells, W. M., Viola, P. A., Atsumi, H., Nakajima, S. and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis* **1(1)** 35–51.

Wilson, A. G. (1995). *Statistical models for shapes and deformations*. Ph.D. thesis, Institute of Statistics and Decision Sciences, Duke University.

Woods, R., Mazziotta, J. C. and Cherry, S. R. (1993). Mri-pet registration with automated algorithm. *Journal of Computer Assisted Tomography* **17(4)** 536–546.

Woods, R. P., Cherry, S. R. and Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing pet images. *Journal of Computer Assisted Tomography* **16** 620–633.

Biography

Sining Chen was born in Beijing, China on December 2, 1978. She received her B.S. in applied mathematics from Tsinghua University, Beijing, China in 1997. She joined the graduate program of statistics at Duke University, Durham, North Carolina, USA that year.