The Institute For Research In Cognitive Science

Probabilistic Matching of Brain Images

by

J.C. Gee
University of Pennsylvania
L. LeBriquer
C. Barillot
Université de Rennes I
D.R. Haynor
University of Washington

University of Pennsylvania
3401 Walnut Street, Suite 400C
Philadelphia, PA 19104-6228

April 1995

Site of the NSF Science and Technology Center for Research in Cognitive Science

University of Pennsylvania
Founded by Benjamin Franklin in 1740

IRCS Report 95-07

PROBABILISTIC MATCHING OF BRAIN IMAGES[1]

J. C. GEE
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA

L. LE BRIQUER AND C. BARILLOT
Laboratoire SIM
Faculté de Médecine, Université de Rennes I
35043 Rennes Cedex, France

AND

D. R. HAYNOR
Department of Radiology
University of Washington
Seattle, WA 98195, USA

Abstract. Image matching has emerged as an important area of investigation in medical image analysis. In particular, much attention has been focused on the atlas problem, in which a template representing the structural anatomy of the human brain is deformed to match anatomic brain images from a given individual. The problem is made difficult because there are important differences in both the gross and local morphology of the brain among normal individuals. We have formulated the image matching problem under a Bayesian framework. The Bayesian methodology facilitates a principled approach to the development of a matching model. Of special interest is its capacity to deal with uncertainty in the estimates, a potentially important but generally ignored aspect of the solution. In the construction of a reference system for the human brain, the Bayesian approach is well suited to the task of modeling variation in morphology. Statistical information about morphological variability, accumulated over past samples, can be formally introduced into the problem formulation to guide the matching or normalization of future data sets.

[1]To appear in *Proceedings of XIVth International Conference on Information Processing in Medical Imaging*, Ile de Berder, France, June 26–30, 1995.

1. Introduction

Modern imaging modalities routinely provide a variety of information—ranging from highly detailed, three-dimensional (3-D) pictures of cerebral anatomy to maps of functional activity within the brain—that have become indispensable in the investigation of the neurological integrity of an individual. These images, however, are often difficult to interpret because of ambiguity in the precise location and extent of the structure(s) of interest. This uncertainty can arise in several ways: intrinsic limitations in the resolution of the scanner; disparate angles at which image sections were obtained by the different scanners or at different times; or natural variability in both the gross and local morphology of the brain among normal individuals. The goal of our work is to ameliorate the localization of anatomical structures by developing fully automated methods based on the use of a labeled *atlas*. An early survey of atlas-based methods for cerebral localization can be found in [1], in which advantages unique to the approach are discussed. The fundamental assumption of these methods is that the topology of cerebral structures is invariant among normal individuals. To the degree that this assumption holds, the localization problem can be solved by determining the mapping from the brain image of a given individual to some labeled representation[1] or atlas of normal neuroanatomy.

The first comprehensive effort aimed at recovering the local differences between a pair of brain images was made by Broit [5]. Broit proposed modeling the atlas as an elastic object so that it could be "physically" deformed to match the brain image of an individual. A multiresolution version of this "elastic matching" technique was implemented by Bajcsy and Kovačič [6] to speed its convergence and to handle large-scale variations in anatomy. Physically-based modeling is now widely applied in image analysis to characterize shape and its deformation. A number of global object models have been proposed for recognition tasks but these are unsuited to the atlas problem because they cannot describe local changes in shape. Even the hybrid schemes [7, 8, 9], which combine lumped and distributed parameter models, and many of the distributed parameter models [10, 11] are generally not appropriate because only the surface or contour of an object is modeled, whereas fully volumetric representations are the most useful for the atlas problem.

In the present paper, we present a Bayesian decision-theoretic formulation of the volumetric image matching problem. The Bayesian approach (to regularization) represents a generalization of many physically-based methods, such as elastic matching, and is appealing for additional important rea-

[1] This representation can take many forms but is usually derived from either a brain image of an individual or a textbook (see, for example, the atlases in [2, 3, 4]).

sons. First, because we are dealing with *uncertain* quantities, such as noisy data, from which information or decisions are derived (that must themselves then be uncertain), it is natural to adopt a probabilistic approach. Second, Bayesian analysis formally embodies the use of *prior* information that we may have about the problem. In matching, the prior serves to constrain our mappings by favoring certain configurations, in effect, regularizing the problem. Among the many operational advantages of Bayesian analysis, the most relevant to our work is that the result is a *posterior distribution* for the unknown mapping, which expresses the probability of any mapping given the observed images. The existence of this distribution makes possible a range of analyses, including the estimation of the *variance* or reliability of the estimated mapping. This information about the uncertainty in the estimate better informs the investigator in his or her interpretation of the results and indicates to the user where the addition of information may improve the results (in terms of reducing its uncertainty). In the construction of a reference system for the human brain, the Bayesian approach can provide information about the variability inherent in the normalization procedure.

2. Elements of Decision Theory

Statistical decision theory is concerned with the problem of decision making in light of relevant statistical information other than that provided by the sample [12]. One important kind of such nonsample information is prior knowledge that we may have about the unknowns. We usually also have some understanding of the possible consequences of making a particular decision, and indeed its quantification will be necessary to perform estimation. This is achieved by determining a *loss function* which measures the cost incurred for making a particular decision given some value of the unknown. In our matching problem, the unknown state of nature or *parameter* θ is the "true" displacement mapping relating the pair of images; the images constitute the sample; and the decision or *action* concerns the value to which the mapping is assigned. To evaluate the effect of taking action a_1 when the true state of nature is θ_1, the loss function is used to calculate the cost $L(\theta_1, a_1)$ that is incurred.

As with knowledge about the consequences of making particular decisions, the prior information related to θ must similarly be quantified in order to be useful. This is done by considering the unknowns to be random variables. The prior knowledge is then expressed as a probability distribution π on the parameter space Θ.

In order to actually make a decision, first note that the actual incurred loss is itself a random quantity; that is, it will never be known with certainty.

A sensible approach to take under such circumstances is to base the choice of a decision on its "expected" loss:

$$E^{\pi^*} L(\theta, a) = \int_{\Theta} L(\theta, a)\, d\pi^*(\theta),$$

where $\pi^*(\theta)$ is the final (posterior) probability distribution of θ (see Section 5). Decisions can then be made by choosing the action which minimizes the expected loss.

3. Prior Information

The use of prior information is a fundamental element of Bayesian decision theory. In many problems, the introduction of prior knowledge substantially improves the solution by, for example, reducing noise or ruling out "impossible" answers. But for image matching and other ill-posed problems, prior information in the form of constraints are essential to their solution.

In the absence of specific prior information about a given problem, constraints on the smoothness of its solution, as imposed by standard regularization, have turned out to be useful for solving a wide range of ill-posed problems—see [13, 14] for examples in early vision. In the study of biological shape, the imposition of C^1 continuity has been advocated in performing morphometrics, based on the observation that spatial variation in the proportions of shape tend to be graded [15]. Nevertheless, it will be beneficial to admit mappings of C^0 continuity, which have proved satisfactory for the atlas problem when used in combination with a separate global affine registration stage [6, 16].

The task of designing priors which reflect our expectations about the smoothness of the unknown mapping is straightforward: the prior probability corresponding to a particular stabilizing functional can be modeled by the Gibbs distribution whose energy function is the same stabilizer. For the linear elastic strain energy functional we have used [17], the Gibbs energy function is quadratic: our prior distribution $\pi(\theta)$ is therefore Gaussian[2]. Szeliski [19] has observed that the spectrum of the membrane model—to which our linear elastic prior is related—coincides in form to that of a Brownian fractal. One implication is that our priors have no natural scale and hence are suitable for modeling data of any size.

[2]It should be emphasized that smoothness constraints compose only one kind of prior information useful for image matching and that the particular prior we have chosen is not essential to the present Bayesian formulation of the matching problem—see [18] for additional discussion on prior models suitable for atlas matching.

4. Likelihood Functions

The likelihood $f(\boldsymbol{z}|\theta)$ expresses the probability of observing the data $\boldsymbol{z}$ given any particular value of the unknowns θ. Its definition requires a model for predicting the observations from given values of the unknowns as well as knowledge about the statistical characteristics of the measurement noise. The likelihood function therefore provides us with the means to explicity model the process by which our observations are degraded.

The observations encountered in image matching may be of several kinds, each corresponding to a different feature space. This is different from the usual case in which the observations would be degraded samples of the unknowns. Rather, the likelihood of a mapping is determined by measuring the similarity between pairs of features. The uncertainty in these measurements can be modeled by imagining that the similarity values are obtained from a virtual sensor. The procedure is described in [18] but the general idea is as follows: an error surface is first estimated locally about the current displacement estimate of the given point; this surface is approximated by a quadratic that we then use to define a Gaussian representing the likelihood function for the point.

5. Posterior Analysis

We began by requiring that, in the absence of additional information, a mapping should be smooth and that the likelihood of its occurrence increased with its smoothness. When data is presented to us in the form of "observations" made of the unknown mapping $\theta = \boldsymbol{\delta}$, the likelihood of an arbitrary mapping no longer depends only on its smoothness but must take into account its probability of actually having produced the observations. Our revised beliefs about the mappings $\boldsymbol{\delta}$ in view of the observed data $\boldsymbol{z}$ is simply the conditional distribution of $\boldsymbol{\delta}$ given $\boldsymbol{z}$:

$$\pi(\boldsymbol{\delta}|\boldsymbol{z}) = \frac{f(\boldsymbol{z}|\boldsymbol{\delta})\pi(\boldsymbol{\delta})}{\int_{\Theta} f(\boldsymbol{z}|\boldsymbol{\delta})\, d\pi(\boldsymbol{\delta})},$$

where f and π are the likelihood and prior, respectively. We determined earlier that an optimal action is one which minimizes the expected loss of the problem. We recognize now that the expectation should be taken with respect to the posterior distribution $\pi(\boldsymbol{\delta}|\boldsymbol{z})$.

5.1. DISPLACEMENT FIELD ESTIMATION

The most common estimate used in Bayesian analysis is the value of the unknown which maximizes the posterior distribution. This *maximum a pos-*

teriori (MAP) estimate is of interest because it represents the maximum likelihood estimate derived from the posterior distribution.

Consider the situation in which the similarity of an image pair is determined by some similarity measure S and the prior is the Gibbs distribution with energy U_{prior} equal to the linear elastic strain energy. Because the likelihood function is similarly a Gibbs distribution, the posterior is Gibbsian with energy $U(\boldsymbol{\delta}, \{I_{\text{subject}}, I_{\text{atlas}}\})$ equal to:

$$-\int_{\Omega_{\text{atlas}}} S(I_{\text{atlas}}(\mathbf{x}+\boldsymbol{\delta}), I_{\text{subject}}(\mathbf{x}))\, d\mathbf{x} + U_{\text{prior}}(\boldsymbol{\delta}), \tag{1}$$

where I_{atlas} and I_{subject} represent the images of the atlas and brain of the individual, respectively, and S is computed over all points $\mathbf{x}$ in the domain of the atlas Ω_{atlas}. For Gibbs distributions, MAP estimation reduces to the problem of minimizing energy U. Note that the energy functional in (1) is the same one that appears in the variational formulation of elastic matching. Thus, the current MAP estimate and the standard-regularized solution to the elastic matching problem are equivalent.

Our optimization algorithm to estimate the MAP solution is described in [20], and its performance along with that of the *minimum mean squared error* (MMSE) estimator were recently examined in [18]. The MMSE estimate is the optimal action for the *squared error loss* function: $L(\theta, a) = (\theta - a)^2$. It is equal to the mean of the posterior and can be estimated through Monte Carlo integration using the Gibbs sampler described next.

5.2. VARIANCE ESTIMATION

Although Bayesian estimates represent "optimal" interpretations of the information contained in the prior and the sample data, a potential fault with their use—and with all other methods in the brain image matching literature—is that there is no indication of the uncertainty in their values. The uncertainty encodes information about the influence of the prior on the estimate. In addition, the uncertainty in the sample data is reflected in the final reliability of the solution.

Consider a Gibbs distribution p with quadratic energy:

$$-\log p(\boldsymbol{Z} = \boldsymbol{z}) = \frac{1}{2}(\boldsymbol{z} - \boldsymbol{z}^*)^{\mathrm{T}}\mathbf{K}(\boldsymbol{z} - \boldsymbol{z}^*) + k,$$

where $\boldsymbol{z}^*$ is the minimum energy value and k a constant. This is Gaussian and, therefore, the mean and covariance completely characterize the distribution. To obtain the covariance matrix, it is only necessary to invert the matrix $\mathbf{K}$ associated with the Gibbs energy function. For energies that are not convex, the definition of the corresponding Gibbs distribution requires

higher order statistics. We can nonetheless approximate the variance near an estimate by fitting a quadratic to the posterior energy function at the estimate. The matrix inversion needed to obtain the covariance matrix $\mathbf{K}^{-1}$ can be performed using a number of techniques. However, for the problem domains that are normally encountered in image matching, the covariance matrix is very large. We therefore considered only the diagonal elements of the covariance matrix, which represent the variance of the estimate at each site.

A finite element approximation to $\boldsymbol{\delta}$ was used. In the finite element method [21], the problem domain is subdivided into regions or *elements*. Neighboring elements share discrete *nodal* points along their interface. The only unknowns are those at the element nodes. The displacement $\boldsymbol{\delta}$ at any other point within an element is interpolated from the nodal values: $\delta^j = \sum_{n \in \mathcal{N}(e)} \delta_n^j N_n$, where $\mathcal{N}(e)$ is the set of nodes which delimit element e, N_i are the interpolating functions, and we write δ_n^i for the value of the ith component of the displacement at node n. It is straightforward to verify that the finite element representation of an arbitrary Gibbs energy function leads to a Markov random field (MRF) model. Since our energy functions are quadratic, the local characteristics of the MRF are Gaussian. It is thus possible to build an efficient continuous state space Gibbs sampler for our problem [22]. To illustrate its construction, consider the following quadratic (finite element) energy:

$$\begin{aligned} U(\boldsymbol{\delta}) &= \frac{1}{2}[\boldsymbol{\delta}]^{\mathrm{T}}\mathbf{K}\boldsymbol{\delta} + [\boldsymbol{\delta}]^{\mathrm{T}}\mathbf{f} + k \\ &= \frac{1}{2}\sum_{nm}[\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{K}_{nm}\boldsymbol{\delta}_m + \sum_n [\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{f}_n + k, \end{aligned}$$

where $\mathbf{K}$ is a positive definite matrix, $\mathbf{f}$ a constant vector, and k a constant. The indices n and m range over the set of all element nodes. Define the lattice of sites $\mathcal{S}$ to be the set of finite element nodes. The neighborhood system $\mathcal{G}$ at each node n comprises those nodes m which potentially exert a "force" on it: $\mathcal{G}_n = \{m | m \neq n \text{ and } m \in \mathcal{N}(\mathcal{E}(n))\}$, where $\mathcal{E}(n)$ is the set of elements containing node n. Finally, define the clique potentials as follows:

$$\begin{array}{ll} V_C(\boldsymbol{\delta}) = \frac{1}{2}[\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{K}_{nn}\boldsymbol{\delta}_n + [\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{f}_n & \text{if } C = \{n\}, \\ V_C(\boldsymbol{\delta}) = [\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{K}_{nm}\boldsymbol{\delta}_m & \text{if } C = \{n, m\}, \text{ where } n \neq m, \\ V_C(\boldsymbol{\delta}) = 0 & \text{otherwise.} \end{array} \tag{2}$$

It is evident that $U(\boldsymbol{\delta}) = \sum_C V_C(\boldsymbol{\delta})$. We now establish that the local characteristics of this MRF are Gaussian or, equivalently, show that the local energies are quadratic. By definition,

$$p(\boldsymbol{\delta}_n | \boldsymbol{\delta}_m, m \neq n) \propto \exp - \sum_{\{C | n \in C\}} V_C(\boldsymbol{\delta}). \tag{3}$$

Substituting the clique potentials defined in (2) into (3), the local energy $U_n(\boldsymbol{\delta}) = \sum_{\{C|n\in C\}} V_C(\boldsymbol{\delta})$ can be rewritten as:

$$U_n(\boldsymbol{\delta}) = \frac{1}{2}[\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{K}_{nn}\boldsymbol{\delta}_n + [\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{f}_n + \sum_{m\in\mathcal{G}_n}[\boldsymbol{\delta}_n]^{\mathrm{T}}\mathbf{K}_{nm}\boldsymbol{\delta}_m.$$

This expression can be factored into the following quadratic form:

$$U_n(\boldsymbol{\delta}) = \frac{1}{2}(\boldsymbol{\delta}_n - \mu_n)^{\mathrm{T}}\mathbf{K}_{nn}(\boldsymbol{\delta}_n - \mu_n) - \frac{1}{2}[\mu_n]^{\mathrm{T}}\mathbf{K}_{nn}\mu_n,$$

where

$$\mu_n = -\mathbf{K}_{nn}^{-1}\left[\mathbf{f}_n + \sum_{m\in\mathcal{G}_n}\mathbf{K}_{nm}\boldsymbol{\delta}_m\right].$$

The implementation of our continuous state space Gibbs sampler at iteration k, where site n is being updated, can be summarized as follows:

- Calculate the mean vector $\mu_n^{(k)} = -\mathbf{K}_{nn}^{-1}\left[\mathbf{f}_n + \sum_{m\in\mathcal{G}_n}\mathbf{K}_{nm}\boldsymbol{\delta}_m^{(k)}\right]$.
- Set the new displacement $\boldsymbol{\delta}_n^{(k+1)} = \mu_n^{(k)} + \mathbf{b}_n$, where $\mathbf{b}_n$ is Gaussian noise with zero mean and covariance equal to $\mathbf{K}_{nn}^{-1}$.

6. Experimental Results

To demonstrate some of the manifold features of the Bayesian approach, it will be enough to use the two-dimensional data set displayed in Figure 1. The left image in the figure represents an MRI section, extracted at the level of the AC-PC plane. Its slice dimensions are 128×128 with pixel size equal to 1.87mm $\times$ 1.87mm. This image was then distorted with a known thin-plate spline to produce the image shown in the center of Figure 1. The object of the experiment was simply to recover the mapping from the original brain section to its distorted version. To make the problem more challenging, the test image depicted in the right of the figure was created by adding white Gaussian noise ($\sigma = 15$) to the spline-distorted image after first blurring it with a Gaussian ($\sigma = 1$ pixel).

Figure 2 illustrates some of the results obtained when the match between a pair of points was rated according to the similarity of their underlying tissue constituents [18]. In the current case, the MRI volume from which our section was extracted had previously been segmented into its gray and white matter components [23]. To interpet the displacement mappings in the figure, focus on their central portion and ignore those parts which displace only the background pixels—the discrepancy in the background motion is partially due to the zero displacement boundary condition that we imposed

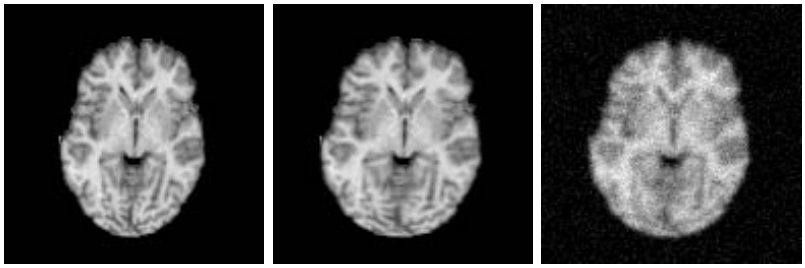

Figure 1. Original MRI section (left) and its distorted versions.

at the image periphery in lieu of first performing a global affine registration of the images. The MMSE estimate has a smoother appearance than that of the MAP solution and, as such, more closely resembles the truth in this case. In spite of the perceived disparity between their motion fields, the two estimates produce similar registrations of the gray and white matter components in terms of the amount of relative overlap between their matched and distorted versions. The specific values are listed in the second and third columns of Table 1 along with the overlap values prior to matching. Note that the MAP estimate actually results in a slightly better match of the tissue components.

Displayed to the right of the estimated mappings in Figure 2 are images of the magnitude of the variance estimates. Compare the variance of the MAP estimate (maximum = 1.1 pixels) with the original MRI section: the underlying structural anatomy is visible, reflecting the fact that the uncertainty is reduced at the gray-white interfaces but is relatively large within those portions of the image that lack informative features. In contrast, the posterior variance (maximum = 2.7 pixels) plainly depicts the complexity of the entire posterior distribution. In this case, for example, it reveals information that would otherwise be unavailable about portions of the mapping which have an exceptionally large variance—in the present experiment where the truth is known, these large uncertainties are also manifest as mismatches by the mean estimate. Unfortunately, it is not possible to visually qualify the extent of the differences between the two variance estimates in the figure because their gray-level values were not scaled in the same way.

The results for the noise-corrupted test image are presented in Figure 3. In this experiment, the normalized cross-correlation between two given points was used to measure their similarity [18]. The good visual agreement between the estimates and the truth is supported by the accurate registration that the estimates produce for the gray and white matter as indicated in Table 1. The motion fields of the two estimates are remarkably similar. Even more striking is the resemblance between their variance

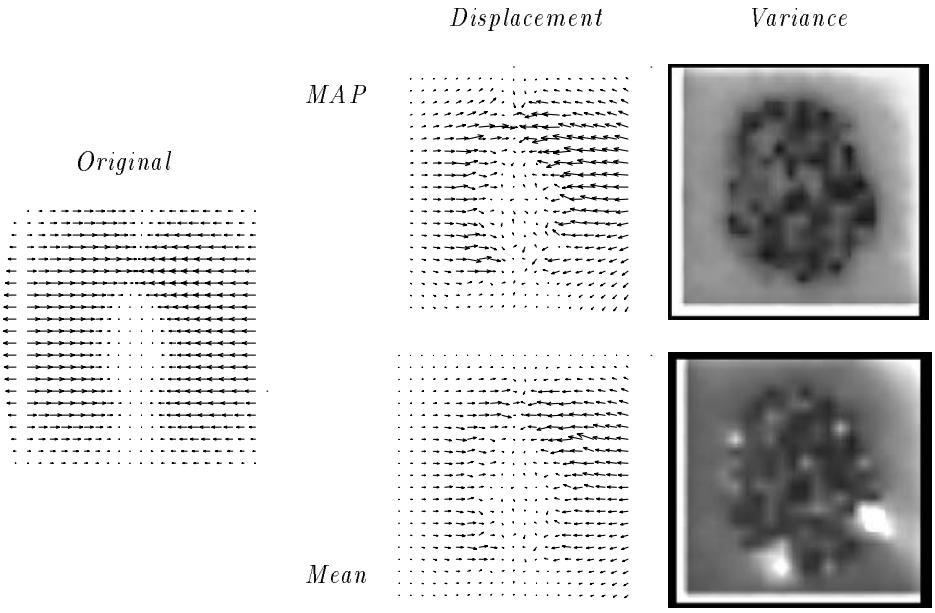

Figure 2. Original and estimated motion fields for the distorted but noise-free test image, where the match between a pair of points was rated according to the similarity of their underlying tissue constituents. Displayed to the right of each displacement estimate is the estimated (magnitude of the) variance of its values.

estimates—both have maximum values equal to 2.1 pixels. The gray-white interfaces are not only readily discernible in the variance estimate about the MAP solution but are equally apparent in the posterior variance. The implication is that the matching problem for the current experimental set-up based on cross-correlation is actually fairly "well posed." Reducing the

TABLE 1. Relative overlap between the estimated and true regions corresponding to the gray and white matter components. The overlap was defined as the ratio between the intersecting area of the two regions under comparison and the area of their union.

		Classification		Correlation	
	Original	MAP	Mean	MAP	Mean
Gray Matter	0.66	0.91	0.87	0.89	0.83
White Matter	0.68	0.91	0.87	0.90	0.85

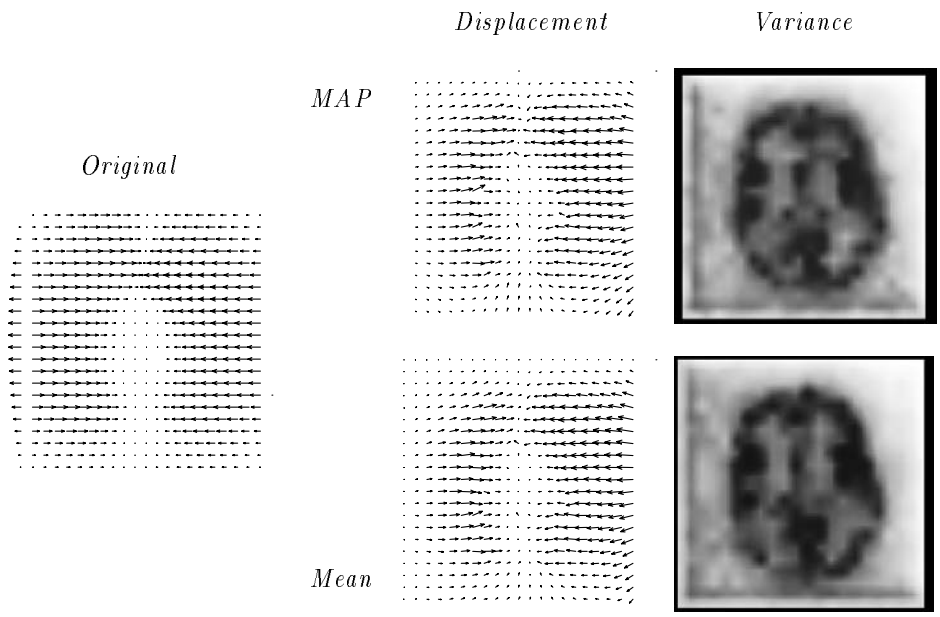

Figure 3. Original and estimated motion fields for the distorted and noise-corrupted test image, where the normalized cross-correlation between two given points was used to measure their similarity. Displayed to the right of each displacement estimate is the estimated (magnitude of the) variance of its values.

"strength" of the prior by an order of magnitude results in a decrease of only approximately 4 percentage points in the final overlap values for the gray and white matter registration. However, as the influence of the prior is further reduced, the solution begins to deteriorate and eventually becomes singular.

7. Discussion

The basic elements of a Bayesian approach to the problem of matching brain images has been described. The method has important advantages that are particularly useful in the construction and application of a probabilistic reference system for the human brain. Since the method is inherently probabilistic, it is well suited to the task of modeling variation in morphology, one of the principal features of such a reference system. Prior statistical information about morphological variability, accumulated over past samples, can be formally introduced into the problem formulation to guide the matching of future data sets. Equally important is that the Bayesian methodology facilitates a principled derivation of the likeli-

hood model, which together with the prior make up a particular matching (posterior) model. In the evaluation of any matching algorithm then, its essential assumptions can be made explicit by determining the models and the estimator effectively in use by the method. For example, the solutions obtained using variational regularization correspond to the MAP estimate. In contrast, the specification of the loss function is *part* of the problem definition in Bayesian estimation.

The task of properly building a Bayesian model is a difficult one, but the process encourages careful deliberation of the problem at hand to arrive at the assumptions used to form the components of the model. For brain image matching, a number of effective similarity measures exist, whereas there is comparatively little known about the kinds of prior models that are specifically suited to the problem (and its applications) or how their influence on the solution, relative to the likelihood, should be determined. The resolution of these issues directly impacts the viability of any matching procedure, Bayesian or otherwise, but is best addressed within a Bayesian setting. The potential benefits that follow from the Bayesian approach, including the capability to estimate and explore the uncertainty of a Bayesian solution [24], are compelling reasons for adopting the method in brain image matching.

8. Acknowledgments

Much of the work reported in this paper was performed while the first author was a post-doctoral fellow at the Laboratoire SIM. The authors are grateful to the Region Council of Brittany for their financial support of this work. We thank Pr. J. M. Scarabin and the Radiology Department of the University Hospital of Rennes for their assistance in obtaining the MRI data sets.

References

1. Evans AC, Beil C, Marrett S, Thompson CJ, Hakim A. Anatomical-functional correlation using an adjustable MRI-based region of interest atlas with positron emission tomography. *J Cereb Blood Flow Metab* 1988;8:513–530.
2. Talairach J, Tournoux P. *Referentially oriented cerebral MRI anatomy: an atlas of stereotaxic anatomical correlations for gray and white matter.* Stuttgart, Germany:Thieme, 1993.
3. Tiede U, Bomans M, Höhne KH, Pommert A, Riemer M, Schiemann T, Schubert R, Lierse W. A computerized three-dimensional atlas of the human skull and brain. *Amer J Neuroradiology* 1993;14(3):551–559.
4. Nowinski WL, Fang A, Fang S, Nguyen BT, Bryan RN, Raghavan R. Three-dimensional electronic atlas of human cerebral deep structures. To appear: Bizais Y, Barillot C, Di Paola R, eds. *Information Processing in Medical Imaging.* Dordrecht:Kluwer Academic Publishers, 1995.

5. Broit C. *Optimal registration of deformed images.* Doctoral dissertation, Department of Computer and Information Science, University of Pennsylvania, PA, 1981.
6. Bajcsy R, Kovačič S. Multiresolution elastic matching. *Comput Vision, Graphics, Image Process* 1989;46:1–21.
7. Nastar C, Ayache N. Non-rigid motion analysis in medical images: a physically based approach. In: Barrett HH, Gmitro AF, Eds. *Information Processing in Medical Imaging: Proceedings of 13th International Conference, IPMI '93.* Berlin:Springer-Verlag, 1993;LNCS 687:15–32.
8. Worring M, Smeulders AWM, Staib LH, Duncan JS. Parameterized feasible boundaries in gradient vector fields. In: Barrett HH, Gmitro AF, Eds. *Information Processing in Medical Imaging: Proceedings of 13th International Conference, IPMI '93.* Berlin:Springer-Verlag, 1993;LNCS 687:48–61.
9. Vemuri BC, Radisavljevic A, Leonard CM. In: Barrett HH, Gmitro AF, Eds. *Information Processing in Medical Imaging: Proceedings of 13th International Conference, IPMI '93.* Berlin:Springer-Verlag, 1993;LNCS 687:62–76.
10. Amini AA, Owen RL, Anandan P, Duncan JS. Non-rigid motion models for tracking the left ventricular wall. In: Colchester ACF, Hawkes DJ, eds. *Information Processing in Medical Imaging.* Berlin:Springer-Verlag, 1991;LNCS 511:343–357.
11. Cohen LD, Cohen I. Finite element methods for active contour models and ballons for 2-D and 3-D images. *IEEE PAMI* 1993;15:1131–1147.
12. Berger JO. *Statistical Decision Theory and Bayesian Analysis.* New York:Springer-Verlag, 1985.
13. Poggio T, Torre V, Koch C. Computational vision and regularization theory. *Nature* 1985;317(26):314–319.
14. Terzopoulos D. Regularization of inverse visual problems involving discontinuities. *IEEE PAMI* 1986;8(4):413–424.
15. Bookstein FL. Principal warps: Thin-plate splines and the decompostion of deformations. *IEEE PAMI* 1989;11(6):567–585.
16. Dann R, Hoford J, Kovačič S, Reivich M, Bajcsy R. Evaluation of elastic matching system for anatomic (CT, MR) and functional (PET) cerebral images. *J Comput Assist Tomogr* 1989;13:603–611.
17. Timoshenko S, Goodier JN. *Theory of elasticity.* New York:McGraw-Hill, 1951.
18. Gee JC, Le Briquer L, Barillot C, Haynor DR, Bajcsy R. Bayesian approach to the brain image matching problem. To appear: Loew MH, ed. *Medical Imaging 1995: Image Processing* Bellingham, WA:SPIE, 1995;2434.
19. Szeliski R. *Bayesian Modeling of Uncertainty in Low-level Vision.* Norwell, MA:Kluwer Academic Publishers, 1989.
20. Gee JC, Haynor DR, Reivich M, Bajcsy R. Finite element approach to warping of brain images. In: Loew MH, ed. *Medical Imaging 1994: Image Processing* Bellingham, WA:SPIE, 1994;2167.
21. Zienkiewicz OC. *The Finite Element Method in Engineering Science.* New York:McGraw-Hill, 1971.
22. Geman S, Geman D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE PAMI* 1984;6(6):721–741.
23. Lachmann F, Barillot C. Brain tissue classification from MRI data by means of texture analysis. In: Loew MH, ed. *Medical Imaging VI: Image Processing* Bellingham, WA:SPIE, 1992;1652:72–83.
24. Hanson KM, Cunningham GS. Exploring the reliability of Bayesian reconstructions. To appear: Loew MH, ed. *Medical Imaging 1995: Image Processing* Bellingham, WA:SPIE, 1995;2434.

A. Maximum A Posteriori Estimation

This appendix is extracted from [20] and does not appear in the proceedings version of this manuscript. Its inclusion here is to help make the present paper as self-contained as possible.

Recall (1) and suppose for concreteness that U_{prior} is equal to the linear elastic strain energy—we will also show how to implement the related membrane energy functional. In addition, we are provided with a set of landmarks. In [20], the likelihood of observing the intensity at a point in one image I_T, given the intensity at the corresponding point in the second image I_R, was represented by an unknown conditional probability density and its estimation formed part of the problem. The two different sources of information about the unknown mapping—landmarks and the images themselves—are combined in a logarithmic opinion pool [12]. The problem of finding the MAP estimate was expressed as follows: find $\mathbf{\Phi}$ and p to minimize

$$\begin{aligned} U &= -\int_{\Omega_R} \log p(I_T(\mathbf{\Phi}(\mathbf{x}))|I_R(\mathbf{x}))\, dV \\ &\quad + \alpha \sum_{i=1}^{N} \frac{(\mathbf{\Phi}(\mathbf{p}_i) - \mathbf{p}_i^{'})^2}{2\eta_i^2} \\ &\quad + \frac{1}{2}\int_{\Omega_R} [\boldsymbol{\sigma}]^{\mathrm{T}} \boldsymbol{\varepsilon}\, dV \\ &= U_{\mathrm{I}} + U_{\mathrm{II}} + U_{\mathrm{III}}, \end{aligned} \tag{4}$$

where

- $\mathbf{x}$ is a generic point in R^d, d being equal to the number of spatial dimensions,
- $\boldsymbol{\delta}(\mathbf{x})$ is a vector-valued deformation or displacement,
- $\mathbf{\Phi}(\mathbf{x}) = \mathbf{R}(\mathbf{x} + \boldsymbol{\delta}(\mathbf{x}))$ is the correspondence induced by the local deformation $\boldsymbol{\delta}$ and the global affine map $\mathbf{R}$; $R(\mathbf{x})_i = \sum_j R_{ij} x_j + C_i$,
- $(\mathbf{p}_i, \mathbf{p}_i^{'})$ is a set of user-defined (approximately) corresponding points, where the true location of $\mathbf{\Phi}(\mathbf{p}_i)$ is assumed to have a distribution centered at $\mathbf{p}_i^{'}$, with uncertainty measured by the Gaussian with width η_i, and
- $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ are the stress and strain vectors respectively.

The first term in the expression above represents the logarithm of the unknown likelihood, integrated over the domain Ω_R of I_R. The second term represents the sum over the user-specified landmarks $\mathbf{p}_i$, which must be mapped by $\mathbf{\Phi}$ into the points $\mathbf{p}_i^{'}$; the value of α is adjusted so that $\|\mathbf{\Phi}(\mathbf{p}_i) - \mathbf{p}_i^{'}\|$ is of order η_i. The last term represents the internal strain energy of the deformation $\boldsymbol{\delta}$. It penalizes deviations from a smooth (globally affine) map $\mathbf{\Phi}$. For simplicity, we assume that the atlas, in addition to

being linear elastic, is isotropic and homogeneous; thus, only Lame's elastic constants λ and μ remain to define its material properties.

The minimization problem implied by MAP estimation is made difficult in image matching because the expressions involved, as exemplified by (4), are highly nonlinear. Iterative techniques must therefore be used to search for the solution. We begin by computing a spline approximation to $\mathbf{\Phi}$ if landmarks are available; otherwise, only an initial registration $\mathbf{R}$ is inferred by aligning the centroids and principal axes of the brain volumes. The joint histogram between $I_R(\mathbf{x})$ and $I_T(\mathbf{\Phi}(\mathbf{x}))$ then provides us with an estimate of p, which is smoothed. We use a finite element approximation to $\boldsymbol{\delta}$. In the finite element method, the domain of the problem is first divided into regions called elements. The elements are connected at discrete nodal points along their periphery. The only unknowns are those at the element nodes. We define $\mathcal{N}$ to be the set of all nodes and $\mathcal{E}$ the set of all elements. The jth component of the displacement $\boldsymbol{\delta}$ at any other point within the element is interpolated from its nodal values:

$$\delta^j = \sum_{n \in \mathcal{N}(e)} \delta_n^j N_n, \tag{5}$$

where $\mathcal{N}(e)$ represents the set of nodes in element e and we write δ_n^i for the value of δ^i at node n. In our preliminary investigations, the interpolating or shape functions N_i were those for the two-dimensional (2-D) parabolic isoparametric element or the bilinear quadrilateral element.

Of the three terms in (4), the second and third are immediately seen to be quadratic in $\boldsymbol{\delta}$ —because of the local character of finite element basis functions, the middle term in (4) involves only a few of the δ_n^i's. If the first term in (4) were also quadratic, then the problem of minimizing U would reduce to solving a large but sparse system of linear equations. Our strategy is to replace the first term by a linearized or quadratic approximation. The global minimum $\boldsymbol{\delta}^\star$ of the resulting quadratic is found. This is used to form the next estimate of $\mathbf{\Phi}$ and p is recomputed again. We then develop a new approximation valid near $\boldsymbol{\delta}^\star$ and repeat the process until a stable solution is reached.

The algorithm is illustrated in Figure 4. It is has an outer loop over different resolution levels and an inner optimization algorithm. The inner algorithm, given a starting approximation to $\boldsymbol{\delta}$, fixed values for α, elastic constants λ and μ, and $p(I_T|I_R)$, finds a local optimum for U given by (4). The outer loop defines a series of image sets I_R^k, I_T^k, Lagrange multipliers α^k, and elastic constants λ^k and μ^k designed so that if $\boldsymbol{\delta}^k$ minimizes $U^k(\boldsymbol{\delta}) = U(\boldsymbol{\delta}; I_R^k, I_T^k, \{(\mathbf{p}_i, \mathbf{p}_i^{'})\}, \alpha^k, p^k, \mathbf{R}^k, \lambda^k, \mu^k)$; then:

(i) $\boldsymbol{\delta}^k \rightarrow \boldsymbol{\delta}^*$, the value that globally minimizes U;

for $(k = 0;\ k < N;\ k{+}{+})$ { Loop over each of N resolution levels.

if $(k = 0)$ {

Calculate $\mathbf{\Phi}^{(0,0)}$ initial spline approximation to $\mathbf{\Phi}$ if tie points, $(\mathbf{p}_i,\ \mathbf{p}_i^{'})$, are available; otherwise, $\mathbf{\Phi}^{(0,0)} = \mathbf{R}^{(0,0)}\mathbf{x}$, where $\mathbf{R}^{(0,0)}$ is obtained by aligning centroids and principal axes of the brain volumes;

Calculate initial displacement field $\delta^{(0,0)}(\mathbf{x}) = (\mathbf{R}^{(0,0)})^{-1}\mathbf{\Phi}^{(0,0)}(\mathbf{x}) - \mathbf{x}$ for $\mathbf{x} \in \Omega_R$;

} else {

$\mathbf{R}^{(k,0)} = \mathbf{R}^{(k-1,m_{k-1})}$;

$\delta^{(k,0)} = \delta^{(k-1,m_{k-1})}$;

$\alpha^{(k,0)} = \alpha^{(k-1,m_{k-1})}$;

If $\mathcal{N}^{k-1}$, $\mathcal{E}^{k-1}$ are different from $\mathcal{N}^k$, $\mathcal{E}^k$, then $\delta^{(k,0)}$ is interpolated from $\delta^{(k-1,m_{k-1})}$;

}

for $(l = 0;\ l < m_k;\ l{+}{+})$ { Optimize $\mathbf{\Phi}$ over m_k iterations, updating α, p at intermediate steps.

if $(l > 0)$ {

$\mathbf{R}^{(k,l)} = \mathbf{R}^{(k,l-1)}$;

$\delta^{(k,l)} = \delta^{(k,l-1)}$;

$\alpha^{(k,l)} = \alpha^{(k,l-1)}$;

}

Calculate $p^{(k,l)}$, kernel-smoothed estimate of p, based on the map $\mathbf{\Phi}^{(k,l)}(\mathbf{x}) = \mathbf{R}^{(k,l)}(\mathbf{x} + \delta^{(k,l)}(\mathbf{x}))$;

Optimize local deformation

$\delta(\delta^{(k,l)}, I_R^k, I_T^k, \{(\mathbf{p}_i, \mathbf{p}_i^{'})\}, \alpha^{(k,l)}, p^{(k,l)}, \mathbf{R}^{(k,l)}, \lambda^k, \mu^k)$;

Optimize global affine map

$\mathbf{R}(\delta^{(k,l)}, I_R^k, I_T^k, \{(\mathbf{p}_i, \mathbf{p}_i^{'})\}, \alpha^{(k,l)}, p^{(k,l)}, \mathbf{R}^{(k,l)}, \lambda^k, \mu^k)$; Optional.

Update $\alpha^{(k,l)}$;

}

}

Figure 4. Algorithm to iteratively estimate the mapping $\mathbf{\Phi}$ and the conditional probability density p by minimizing the negative logarithm of the posterior over multiple resolution levels—see (4).

(ii) $\alpha^k \to \alpha^*$, which is such that $\mathbf{\Phi}(\mathbf{p}_i) \approx \mathbf{p}_i^{'}$ to within η_i;

(iii) $p^k \to p^*$, which is an estimate for the joint probability density of I_R and I_T;

(iv) $I_R^k \to I_R$ and $I_T^k \to I_T$.

I_R^k and I_T^k represent smoothed versions of I_R and I_T that have been reduced in resolution under some fixed filtering schedule. For example, the Gaussian pyramid is a suitable representation for multiscale applications and one that has been successfully applied in elastic matching. We wish in general the elastic constants λ^k and μ^k to be decreasing but experimental investigation will be required to determine their schedules. Initially, assuming $\lambda^k \equiv \lambda, \mu^k \equiv \mu$ for all k is reasonable. The element and node sets $\mathcal{E}^k$, $\mathcal{N}^k$ are normally different at each iteration k, with the size of the elements decreasing as the value of k increases. α is adjusted within the optimization loop so that $\mathbf{\Phi}(\mathbf{p}_i)$ agrees with $\mathbf{p}_i^{'}$ to within approximately η_i. Thus, α is decreased if $\sum_i(\mathbf{\Phi}(\mathbf{p}_i) - \mathbf{p}_i^{'})^2/2\eta_i^2$ is too small and increased if it is too large. As with the elastic constants, an initial estimate of α will be obtained from prior experience. In order to calculate $p^{(k,l)}$ and $\nabla p^{(k,l)}$ we construct the joint empirical probability histogram of I_R and I_T by determining $(I_R^k(\mathbf{x}), I_T^k(\mathbf{\Phi}^{(k,l)}(\mathbf{x})))$ for a large number of points $\mathbf{x}$. This empirical estimate is smoothed. $\nabla p(I_T|I_R)$ can then be estimated at a given point $\mathbf{x}$ in I_R by first determining the values of $p(I_T|I_R)$ for a small neighborhood centered at $\mathbf{x} + \boldsymbol{\delta}^\star$ in I_T, where $\boldsymbol{\delta}^\star$ is the current displacement estimate. A quadratic surface is fit to these likelihood values, from which our gradient is finally derived. Following the density estimation step the algorithm attempts to find a new local minimum of U given the current estimate of $\boldsymbol{\delta}$. We proceed by differentiating each of the three terms in (4) separately.

U_{II}: For each i, $\mathbf{p}_i \in e_i \in \mathcal{E}$, where e_i is some uniquely determined element of $\mathcal{E}$. For $\mathbf{x} \in e_i$,

$$\mathbf{\Phi}(\mathbf{x}) = \mathbf{R}(\mathbf{x} + \sum_{n \in \mathcal{N}(e_i)} \boldsymbol{\delta}_n N_n(\mathbf{x})).$$

It follows that

$$\frac{\partial}{\partial \delta_n^j}\left\{\frac{(\mathbf{\Phi}(\mathbf{p}_i) - \mathbf{p}_i^{'})^2}{2\eta_i^2}\right\} = \begin{cases} \sum_k \frac{1}{\eta_i^2}[\mathbf{\Phi}(\mathbf{p}_i) - \mathbf{p}_i^{'}]_k R_{kj} N_n(\mathbf{p}_i) & \text{if } n \in \mathcal{N}(e_i), \\ 0 & \text{otherwise.} \end{cases}$$

In the finite element implementation it is convenient to develop an equivalent matrix form of the above derivation. Let

$$\boldsymbol{\delta}^e = [\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_n]^{\mathrm{T}},$$

the vector of n nodal displacements of the element e and

$$\mathbf{N} = [\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_n],$$

where $\mathbf{N}_i = N_i\mathbf{I}$ and $\mathbf{I}$ is the $r \times r$ identity matrix with r the number of unknowns per node. Then (5), which defines the displacement at a point within the element, can be written as

$$\boldsymbol{\delta} = \mathbf{N}\boldsymbol{\delta}^e.$$

For pedagogic clarity, we will ignore the affine component $\mathbf{R}$ in the matrix derivations. The minimization of U_{II} with respect to the nodal displacement vector $\boldsymbol{\delta}^e$ for each i becomes in matrix notation:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\delta}^e}\left\{\frac{(\boldsymbol{\Phi}(\mathbf{p}_i)-\mathbf{p}_i^{'})^2}{2\eta_i^2}\right\} &= \frac{1}{2\eta_i^2}\frac{\partial}{\partial \boldsymbol{\delta}^e}\left\{(\mathbf{p}_i+\mathbf{N}\boldsymbol{\delta}^e)-\mathbf{p}_i^{'}\right\}^2 \\
&= \frac{1}{2\eta_i^2}\frac{\partial}{\partial \boldsymbol{\delta}^e}\left\{(\mathbf{p}_i-\mathbf{p}_i^{'})^2+2[\boldsymbol{\delta}^e]^{\mathrm{T}}[\mathbf{N}]^{\mathrm{T}}(\mathbf{p}_i-\mathbf{p}_i^{'})\right. \\
&\quad \left.+[\boldsymbol{\delta}^e]^{\mathrm{T}}[\mathbf{N}]^{\mathrm{T}}\mathbf{N}\boldsymbol{\delta}^e\right\} \\
&= \frac{1}{\eta_i^2}[\mathbf{N}]^{\mathrm{T}}(\mathbf{p}_i-\mathbf{p}_i^{'})+\frac{1}{\eta_i^2}[\mathbf{N}]^{\mathrm{T}}\mathbf{N}\boldsymbol{\delta}^e \\
&= \mathbf{f}_{\mathrm{II}}^e+\mathbf{K}_{\mathrm{II}}^e\boldsymbol{\delta}^e,
\end{aligned}$$

where

$$\mathbf{K}_{\mathrm{II}}^e=\frac{1}{\eta_i^2}[\mathbf{N}]^{\mathrm{T}}\mathbf{N}$$

is the contribution of the point matching term to the element "stiffness" matrix $\mathbf{K}^e$ of the element e and

$$\mathbf{f}_{\mathrm{II}}^e=\frac{1}{\eta_i^2}[\mathbf{N}]^{\mathrm{T}}(\mathbf{p}_i-\mathbf{p}_i^{'})$$

is its contribution to the element "load" vector $\mathbf{f}^e$.

U_{III}: In the finite element method integrals, such as the one representing the internal strain energy in U_{III}, are calculated on an element-by-element basis and then summed. Thus, we write

$$U_{\mathrm{III}}=\sum_e U_{\mathrm{III}}^e=\sum_e\int_{V^e}[\boldsymbol{\sigma}]^{\mathrm{T}}\boldsymbol{\varepsilon}\,dV, \tag{6}$$

where it is understood that the summation is over the elements comprising Ω_R. In our linear elastic problem the strains at a point within the element can be interpolated from the nodal displacements as follows:

$$\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_{x_1}\\ \varepsilon_{x_2}\\ \gamma_{x_1x_2}\end{bmatrix}=\begin{bmatrix}\frac{\partial\delta^1}{\partial x_1}\\ \frac{\partial\delta^2}{\partial x_2}\\ \frac{\partial\delta^1}{\partial x_2}+\frac{\partial\delta^2}{\partial x_1}\end{bmatrix}=\mathbf{B}\boldsymbol{\delta}^e, \tag{7}$$

where

$$\mathbf{B}=[\mathbf{B}_1,\mathbf{B}_2,\ldots,\mathbf{B}_n]$$

and

$$\mathbf{B}_i = \begin{bmatrix} \frac{\partial N_i}{\partial x_1} & 0 \\ 0 & \frac{\partial N_i}{\partial x_2} \\ \frac{\partial N_i}{\partial x_2} & \frac{\partial N_i}{\partial x_1} \end{bmatrix}.$$

The stresses are in turn linearly related to the strains:

$$\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}, \tag{8}$$

where $\mathbf{D}$ is the elasticity matrix. In isotropic plane stress, for example, the elasticity matrix is equal to

$$\mathbf{D} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix},$$

where E and ν are Young's modulus and Poisson's ratio, respectively—these can be directly related to Lame's elastic constants λ and μ. Substituting (7) and (8) into (6), we obtain for an arbitrary element e

$$U^e_{\mathrm{III}} = \frac{1}{2} \int_{V^e} [\boldsymbol{\delta}^e]^{\mathrm{T}} [\mathbf{B}]^{\mathrm{T}} \mathbf{D}\mathbf{B}\boldsymbol{\delta}^e \, dV.$$

Minimizing with respect to $\boldsymbol{\delta}^e$, as was done in U_{II}, we have

$$\begin{aligned} \frac{\partial U^e_{\mathrm{III}}}{\partial \boldsymbol{\delta}^e} &= \int_{V^e} [\mathbf{B}]^{\mathrm{T}} \mathbf{D}\mathbf{B}\boldsymbol{\delta}^e \, dV \\ &= \mathbf{K}^e_{\mathrm{III}} \boldsymbol{\delta}^e, \end{aligned}$$

where

$$\mathbf{K}^e_{\mathrm{III}} = \int_{V^e} [\mathbf{B}]^{\mathrm{T}} \mathbf{D}\mathbf{B} \, dV$$

is the contribution of the linear elastic prior to the element stiffness matrix $\mathbf{K}^e$.

The priors corresponding to the Tikhonov stabilizers similarly form a portion of the stiffness matrix when their (finite element) energies are differentiated with respect to the nodal unknowns. To illustrate this result, consider the first order Tikhonov stabilizer, also known as the membrane energy functional:

$$\begin{aligned} U_{\mathrm{III}} &= \int_{\Omega_R} \sum_{ij} \left\{ \frac{\partial \delta^j(\mathbf{x})}{\partial x_i} \right\}^2 dV \\ &= \sum_e \int_{V^e} \sum_{ij} \left\{ \frac{\partial \delta^j(\mathbf{x})}{\partial x_i} \right\}^2 dV \\ &= \sum_e \int_{V^e} \sum_j \|\nabla \delta^j\|^2 \, dV, \end{aligned}$$

where its finite element representation is as follows:

$$\begin{aligned}\int_{V^e} \sum_j \|\nabla \delta^j\|^2 \, dV &= \sum_j \int_{V^e} (\sum_{n \in \mathcal{N}(e)} \delta_n^j \nabla N_n)^2 \, dV \\ &= \sum_{jnn'} \delta_n^j \delta_{n'}^j \int_{V^e} \nabla N_n \nabla N_{n'} \, dV.\end{aligned}$$

Differentiating this expression with respect to $\boldsymbol{\delta}^e$, we have

$$\frac{\partial}{\partial \delta_n^j} \left\{ \int_{V^e} \sum_j \|\nabla \delta^j\|^2 \, dV \right\} = 2 \sum_{n' \in \mathcal{N}(e)} \delta_{n'}^j \int_{V^e} \nabla N_n \nabla N_{n'} \, dV \qquad \text{for } n \in \mathcal{N}(e).$$

In matrix form, we write

$$\int_{V^e} \sum_j \|\nabla \delta^j\|^2 \, dV = \int_{V^e} [\boldsymbol{\delta}^e]^{\mathrm{T}} [\nabla \mathbf{N}]^{\mathrm{T}} \nabla \mathbf{N} \boldsymbol{\delta}^e \, dV,$$

where $[\nabla \mathbf{N}]_{ij} = \partial N_j / \partial x_i \mathbf{I}$ and $\mathbf{I}$ is the $r \times r$ identity matrix. The derivative with respect to $\boldsymbol{\delta}^e$ can then be expressed as

$$\begin{aligned}\frac{\partial U_{\mathrm{III}}^e}{\partial \boldsymbol{\delta}^e} &= \int_{V^e} 2 \, [\nabla \mathbf{N}]^{\mathrm{T}} \nabla \mathbf{N} \boldsymbol{\delta}^e \, dV \\ &= \mathbf{K}_{\mathrm{III}}^e \boldsymbol{\delta}^e,\end{aligned}$$

where

$$\mathbf{K}_{\mathrm{III}}^e = \int_{V^e} 2 \, [\nabla \mathbf{N}]^{\mathrm{T}} \nabla \mathbf{N} \, dV$$

is the element stiffness contribution of the first order Tikhonov stabilizer.

We can express the sum of U_{II} and U_{III} as follows:

$$\frac{\partial (U_{\mathrm{II}} + U_{\mathrm{III}})}{\partial \boldsymbol{\delta}} = \mathbf{K} \boldsymbol{\delta} + \mathbf{f}_0 \tag{9}$$

with $\mathbf{K}$ a constant (global stiffness) matrix and $\mathbf{f}_0$ a constant (global load) vector.

U_{I}: To differentiate U_{I} with respect to $\boldsymbol{\delta}$, the integral is again divided into the contributions of individual elements:

$$\begin{aligned}U_{\mathrm{I}} &= -\sum_e \int_{V_e} \log p \left(I_T(\mathbf{R}(\mathbf{x} + \boldsymbol{\delta}(\mathbf{x}))) \,|\, I_R(\mathbf{x}) \right) \, dV \\ &= -\sum_e \int_{V_e} \log p \left(I_T(\mathbf{R}(\mathbf{x} + \sum_{n \in \mathcal{N}(e)} \boldsymbol{\delta}_n N_n(\mathbf{x}))) \,|\, I_R(\mathbf{x}) \right) dV. \qquad (10)\end{aligned}$$

The derivative of U_{I} with respect to δ_n^j arises only from the terms in (10) for which $n \in \mathcal{N}(e)$, i.e. for $e \in \mathcal{E}(n)$ where $\mathcal{E}(n)$ is the set of elements containing n. We have, for such an e,

$$\frac{\partial U_{\mathrm{I}}^e}{\partial \delta_n^j} = \frac{\partial}{\partial \delta_n^j} \left\{ \int_{V_e} -\log p(I_T(\,\mathbf{R}(\mathbf{x} + \sum_{n \in \mathcal{N}(e)} \boldsymbol{\delta}_n\, N_n(\mathbf{x}))) \,|\, I_R(\mathbf{x}))\, dV \right\} \tag{11}$$

$$= \int_{V_e} \left\{ \frac{-1}{p(I_T(\mathbf{\Phi}(\mathbf{x})) \,|\, I_R(\mathbf{x}))} \cdot \sum_k \{\nabla p(I_T(\mathbf{\Phi}(\mathbf{x})), I_R(\mathbf{x}))\}_k\, R_{kj} N_n(\mathbf{x})\, dV \right\}, \tag{12}$$

where the summation is over the number of nodal unknowns. The key approximation is to treat this as a constant, say $\mathbf{f}_0^{'}$: we then have, with $\mathbf{f} = \mathbf{f}_0 + \mathbf{f}_0^{'}$,

$$\frac{\partial U}{\partial \boldsymbol{\delta}} \approx \mathbf{K}\boldsymbol{\delta} + \mathbf{f}; \tag{13}$$

a minimum is sought for U by solving (13). In the case a linear approximation $\mathbf{K}_0^{'}\boldsymbol{\delta} + \mathbf{f}_0^{'}$ is used in place of (12), we would have

$$\frac{\partial U}{\partial \boldsymbol{\delta}} \approx (\mathbf{K} + \mathbf{K}_0^{'})\boldsymbol{\delta} + (\mathbf{f}_0 + \mathbf{f}_0^{'}). \tag{14}$$

The result with either approximation is a large sparse matrix equation. Moreover, the stiffness matrix is symmetric positive definite—this is ensured in (14) or, more specifically, in the construction of $\mathbf{K}_0^{'}$ by discarding points that lead to negative definite element stiffnesses. We solve (13) or (14) using the ITPACK iterative linear system solver[3] and call the solution $\boldsymbol{\delta}^{\star}$. We then seek the true minimum for U along the line segment between the previous estimate $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^{\star}$. The algorithm is presented in Figure 5, where a constant approximation to (12) has been assumed.

The procedure to optimize the global affine map is similar to, but simpler than, the one to optimize $\boldsymbol{\delta}$. First, we can ignore U_{III} since it does not depend on $\mathbf{R}$. The gradient of U_{II} with respect to $\mathbf{R}$ is easily calculated. The gradient of U_{I} is computationally intensive, and is similar to the calculations for optimizing $\boldsymbol{\delta}$. However, $\mathbf{R}$ can be updated much less frequently than $\boldsymbol{\delta}$ since the initial solution is likely to be quite close to the truth.

[3]Kincaid DR, Respess JR, Young DM. *ITPACK 2C: A FORTRAN Package for Solving Large Sparse Linear Systems by Adaptive Accelerated Iterative Methods.* Center for Numerical Analysis, University of Texas, Austin, TX.

```
Calculate K, f_0;                                          See (9)
for (it = 0; it < N; it++) {                               Optimize over N iterations
    Calculate f_0' at δ, the current estimate of δ*;
    Solve K δ* = f_0 + f_0';
    Find η ∈ [0,1] such that U(η δ + (1 − η) δ*) is minimal;
    Set δ = η δ + (1 − η) δ*;
    If δ agrees with its previous value to within a predefined tolerance, then
        exit;
}
```

Figure 5. Subroutine to minimize the negative logarithm of the posterior with respect to the local deformation—see Figure 4.