

A Marginalized MAP Approach and EM Optimization for Pair-Wise Registration

Lilla Zöllei¹, Mark Jenkinson², Samson Timoner³, and William Wells⁴

¹ A. A. Martinos Center, MGH, USA
lzollei@nmr.mgh.harvard.edu

² FMRIB Centre, University of Oxford, UK
mark@fmrib.ox.ac.uk

³ Scalable Display Technologies, USA
samson@ai.mit.edu

⁴ Radiology Department, BWH, USA
sw@bwh.harvard.edu

Abstract. We formalize the pair-wise registration problem in a *maximum a posteriori* (MAP) framework that employs a multinomial model of joint intensities with parameters for which we only have a prior distribution. To obtain an MAP estimate of the aligning transformation alone, we treat the multinomial parameters as nuisance parameters, and marginalize them out. If the prior on those is uninformative, the marginalization leads to registration by minimization of joint entropy. With an informative prior, the marginalization leads to minimization of the entropy of the data pooled with pseudo observations from the prior. In addition, we show that the marginalized objective function can be optimized by the Expectation-Maximization (EM) algorithm, which yields a simple and effective iteration for solving entropy-based registration problems. Experimentally, we demonstrate the effectiveness of the resulting EM iteration for rapidly solving a challenging intra-operative registration problem.

1 Introduction

The field of medical image registration has been very active in the past two decades. Although numerous alignment methods have been introduced, only limited attention has been devoted to study the relationship among the available methods and the justification for their preference, implicit and explicit hypothesis and their performance[1,2]. Currently, a significant number of methods build upon the maximum likelihood framework because of its intuitive nature and ease of implementation. According to this approach, correct alignment is obtained if we find the transformation that makes the the current observations most probable. Recently, with the abundance of available data sets, using prior information to guide registration has received significant attention. Such techniques allow for more robustness and a larger capture range in the implementation, and with respect to the transformation domain, they facilitate the introduction of smoothness constraints on the deformation models.

When relying on prior models, however, it is challenging to find the appropriate balance between previous and current observations. For example, it is not obvious how the level of confidence in the model can be automatically encoded and / or changed over time. Preliminary efforts using both model- and data-related terms have introduced both sequential and unified algorithms with arbitrary weighing terms [3,4]. In this work, we focus on the establishment of a *maximum a posteriori* (MAP) framework that allows for making use of prior information about both the transformation and the joint statistics of the observed intensity distributions. By treating the former as nuisance parameters, we can marginalize them out and define our registration goal as a posterior on the transformation components. Depending on how informative the prior is on the joint statistics, we demonstrate implicit relationships with currently used methods. We also introduce an attractive optimization framework over our marginalized formulation – the Expectation-Maximization (EM) algorithm allows us to compute the registration update in a simple and elegant way.

2 Problem Formulation

2.1 Marginalized MAP Formulation of Registration

Given multi-modal input data sets, u and v , our goal is to find transformation T that brings those into correct alignment. In addition to the unknown transformation parameters, we introduce another set of parameters, Θ , that encode a discrete joint probability distribution. This model is used in such a way that, if the intensities in the two images are discretized to take values in $\{I_1, I_2, \dots, I_K\}$ then the probability of corresponding voxels (at location j) having discrete intensity values of $T(u)_j = I_a$ and $v_j = I_b$ is $p(u, v|T) = p(T(u)_j = I_a, v_j = I_b) = \Theta(a, b)$. If we use a simplified, one-dimensional indexing for the parameters of the discrete joint probability model, $\Theta = \{\theta_1, \dots, \theta_g\}$ encodes information about intensity joint value occurrences as parameters of an unknown multinomial model, where $\forall i \theta_i \geq 0$ and $\sum_{i=1}^g \theta_i = 1$. This model can also incorporate any additive noise.

The posterior parameter distribution of the transformation variable T and the joint statistics model Θ with respect to the image observations is $p(T, \Theta|u, v)$. In order to align the data sets, this quantity is to be maximized. For our registration purposes though, it is more attractive to compute the posterior on just the transformation parameters. This is manageable given a prior model on the joint statistics parameters that can be marginalized out. Assuming independence between the transformation and the joint distribution models, the posterior distribution of the transformation variable T with respect to solely the image observations is:

$$p(T|u, v) \propto \int P(u, v|T, \Theta)P(T)P(\Theta)d\theta \quad . \quad (1)$$

In Eq.(1), the $P(u, v|T, \Theta)$ term indicates the likelihood function of the transformation and the model parameters given the input observations, and the

right-most two terms are prior distributions over the T and Θ parameters respectively. Optimizing over such a quantity provides a convenient formulation of the registration objective, if we assume that we have prior knowledge about both the transformation and the joint statistics. The most optimal set of transformation parameters \hat{T} are then the ones that optimize the following objective function:

$$\hat{T} = \arg \max_T \log p(T|u, v) \quad . \quad (2)$$

2.2 The Likelihood Term

Using the multinomial model for ordered data, the likelihood of the image observations given the offsetting transformation and an unknown joint probability model is

$$p(u, v|T, \Theta) \propto \prod_j \theta_j^{n_j(T)} \quad , \quad (3)$$

where $n_j(T)$ is the number of voxel pairs that map to the intensity bin associated with θ_j (i.e. $T(u)_j = I_a$ and $v_j = I_b$) and $\sum_{i=1}^g n(T)_i = N$ is the total number of observations. Note that the values of n_j depend on the transformation T , and this dependency is made explicit in this notation.

2.3 Prior on the Joint Statistics Model

For the registration problem, the multinomial distribution is a convenient choice of representation as the θ_i parameters naturally correspond to the widely used histogram encoding of the joint statistics of images given g number of components. Additionally, prior information about such parameters can be expressed by using the Dirichlet distribution, the conjugate prior to a multinomial distribution. We choose the following encoding

$$\text{Dirichlet}(\Theta; w) = \frac{1}{Z(w)} \prod_{i=1}^g \theta_i^{(w_i-1)} = \Gamma(w_0) \prod_j \frac{\theta_j^{(w_j-1)}}{\Gamma(w_j)} \quad , \quad (4)$$

where $\forall i, w_i > 0$ and $\sum_{i=1}^g w_i = w_0$. We may interpret w_0 as the total number of pseudo measurements observed to obtain information about the model and w_i as the individual frequency counts for the joint distribution parameters. The higher the former number is, the greater our confidence becomes in the values of the prior observations.

2.4 The Proposed Objective Function

Using Eq.(3) and Eq.(4), we write the posterior probability on T from Eq.(1) as

$$p(T|u, v, w) \propto P(T) \int \left(\prod_i \theta_i^{n_i(T)} \right) \Gamma(w_0) \left(\prod_j \frac{\theta_j^{(w_j-1)}}{\Gamma(w_j)} \right) d\theta \quad (5)$$

$$\propto P(T) \frac{\Gamma(w_0)}{\Gamma(N + w_0)} \left(\prod_j \frac{\Gamma(n_j(T) + w_j)}{\Gamma(w_j)} \right) \quad (6)$$

where we derived Eq.(6) by identifying a Dirichlet distribution with parameters $n_j(T) + w_j$ and using the fact that the integral of the distribution over its domain is equal to one. Therefore, the objective function of our proposed marginalized MAP registration method becomes

$$\hat{T} = \arg \max_T \log \left[P(T) \frac{\Gamma(w_0)}{\Gamma(N + w_0)} \left(\prod_{j=1}^g \frac{\Gamma(n_j(T) + w_j)}{\Gamma(w_j)} \right) \right] \quad (7)$$

$$= \arg \max_T \left[\log P(T) + \sum_{j=1}^g \log \Gamma(n_j(T) + w_j) + C \right] \quad (8)$$

2.5 Special Cases: Strength of Priors

In the following, we demonstrate how our objective function changes with respect to the nature of the prior, and we also examine the equivalence of our novel objective function of Eq.(8) with some widely used methods. These relationships help us better explain why these registration techniques are expected to converge to the correct alignment configuration.

Uninformative Prior: First we choose an uninformative prior. This means that the prior does not favor any solutions *a priori* and the current observations are going to be solely responsible for the solution to be computed. As an uninformative prior (whose use achieves an equivalence with the maximum likelihood (ML) solution of the problem), we choose to use Jeffreys’ prior [5]. These are not affected by any transformations applied to the variables. Accordingly, we set, $w_0 = 1$ and $\forall i, w_i = \frac{1}{g}$, or in words, we assume to have only one prior observation, and it is distributed over all the existing bins (where g represents the total number of bins used in the multinomial model). Consequently,

$$\log p(T|u, v, w) \rightarrow \log P(T) + \sum_j \log \Gamma \left(n_j(T) + \frac{1}{g} \right) + C \quad (9)$$

The information theoretic joint entropy [6], measuring uncertainty related to the joint occurrence of the input random variables, is defined by :

$$H \left(\frac{n}{N} \right) \equiv - \sum_j \frac{n_j}{N} \log \frac{n_j}{N} = - \frac{1}{N} \sum_j n_j \log n_j + \log N \quad (10)$$

where $\sum_j n_j = N$. For the most part, the differences between expressions $\left(\sum_j n_j \log(n_j) \right)$ and $\left(\sum_j \log \Gamma(n_j) \right)$ are small (“first order” version of the Stir-

ling approximation). Therefore, we may re-write the registration objective function using an uninformative prior as

$$\hat{T} = \arg \max_T \log p(T|u, v, w) \approx \arg \min_T \left[cH \left(\frac{n(T) + \frac{1}{g}}{N + 1} \right) - \log P(T) \right] \quad . \quad (11)$$

This approximation, ignoring the transformation prior, is equivalent to the widely used minimization of joint entropy method [7]. Although widely used, the justification of using that metric in registration has long been debated. Formally, joint entropy is maximum likelihood method. While others have discussed a similar equivalence between ML and joint entropy [8], the present marginalization approach provides a rigorous demonstration of its validity.

Informative Prior: When the priors are informative, or in words, when we have access to a sufficient number of pseudo observations (w_j) from, for example, training data sets, they allow for a more certain belief in their information content. In such a case, the objective function can be approximated by

$$\hat{T} = \arg \max_T \log p(T|u, v, w) \approx \arg \min_T \left[H \left(\frac{n(T) + w}{N + w_0} \right) - \log P(T) \right] \quad . \quad (12)$$

According to Eq.(12), prior information is introduced into the framework by *pooling* together corresponding samples from the previously aligned (prior distribution model) and from the current, to-be-aligned cases. Throughout the optimization, the model observations remain fixed and act as anchor points to bring the other samples into a more likely configuration. Interestingly, this formulation is closely related to another type of entropy-based registration algorithm. Sabuncu et al. introduced a registration technique based on minimizing Renyi entropy, where the entropy measure is computed via a non-plug-in entropy estimator on pooled data [9]. This estimator is based upon constructing the EMST (Euclidean Minimum Spanning Tree) and using the edge length in that tree to approximate the entropy. The reason that such an arrangement would provide a favorable solution has not been previously theoretically justified.

Strong Prior: Lastly, we briefly mention the scenario, where the prior information is very strong.

$$\log p(T|u, v, w) \approx \log P(T) + \sum_j \left(\log \Gamma(w_j) + n_j(T) \frac{\Gamma'(w_j)}{\Gamma(w_j)} \right) + C \quad (13)$$

$$\approx \log P(T) + \sum_j (w_j \log(w_j) + n_j(T) \log(w_j)) + C \quad . \quad (14)$$

In Eq.(13) a first order approximation of $\log \Gamma(n_j(T)+w_j)$ around w_j was written out, with Γ' indicating the derivative of the Gamma function. In Eq.(14), the approximation for $\sum_j \log \Gamma(w_j)$ from Eq.(11) and $\frac{\Gamma'(n)}{\Gamma(n)} \approx \log(n)$ are used, where

the latter is valid for a large range of n . The only term that depends on $n_j(T)$ is $\sum_j n_j(T) \log(w_j)$, thus

$$\hat{T} = \arg \max_T \log p(T|u, v, w) \approx \arg \max_T \left[\log P(T) + \sum_j (n_j(T) \log(w_j)) \right] \quad (15)$$

If we re-express the sum in Eq.(15) as a sum over data points, we can see that this formulation is equivalent to an approximate maximum likelihood approach, where the current samples are, indeed, evaluated under the previously constructed model distribution [10]. Chung et al. experimentally showed that the performance of this similarity measure is not always sufficiently accurate[11]. That finding is now explained by the fact that this objective function considers the model with such a high level of confidence that might not be justified by the number of previous observations.

3 EM Optimization for the Marginalized MAP Registration Problem

The above described objective functions could be optimized in several ways. In general, we can differentiate between gradient- and non-gradient-based techniques. Using gradient-based information in the optimization procedure often results in significant computational advantages, however, they might also require challenging approximation of some terms. For example, when using the multinomial model, update terms need to be estimated from a discrete distribution and also it is required to calculate the partial derivative of *changing* (non-fixed) joint statistics parameters with respect to the to-be-optimized transformation components. Although close approximations exist, one might worry about their accuracy. To escape such computational difficulties, we might optimize both of the above marginalized formulations by using the Expectation-Maximization (EM) algorithm [12]. This framework, from the statistical computing literature, is known to have good properties as an optimizer and the resulting iteration is attractive from a practical standpoint. If we consider the input images as observations and the Θ parameters as hidden information, the EM algorithm defines the solution to $\hat{T} = \arg \max_T \log p(T|u, v)$ as iteratively obtaining the best current solution (\hat{T}_{next}) based upon:

$$\hat{T}_{\text{next}} = \arg \max_T E_{\Theta|u,v,\hat{T}_{\text{old}}} [\log P(T, \Theta|u, v)] \quad (16)$$

$$= \arg \max_T E_{\Theta|u,v,\hat{T}_{\text{old}}} [\log P(T|u, v, \Theta) + \log P(\Theta)] \quad (17)$$

$$= \arg \max_T E_{\Theta|u,v,\hat{T}_{\text{old}}} \left[\log(P(u, v|T, \Theta)P(T)) + \sum_{j=1}^g (w_j - 1) \log \theta_j \right] \quad (18)$$

where Bayes rule was applied in order to express the conditional probability term on the registering transformation. Continuing with the manipulation

$$\hat{T}_{\text{next}} = \arg \max_T E_{\Theta|u,v,\hat{T}_{\text{old}}} \left[\log \prod_j \theta_j^{n_j(T)} + \log P(T) + \sum_{j=1}^g (w_j - 1) \log \theta_j \right] \quad (19)$$

$$= \arg \max_T \left[\log P(T) + E_{\Theta|u,v,\hat{T}_{\text{old}}} \left[\sum_{j=1}^g (n(T)_j + w_j - 1) \log \theta_j \right] \right] \quad (20)$$

$$= \arg \max_T \left[\log P(T) + \sum_{j=1}^g (n(T)_j + w_j - 1) E_{\Theta|u,v,\hat{T}_{\text{old}}} [\log \theta_j] \right] \quad (21)$$

If we define $l_j \equiv \log \theta_j$, then we can define the two steps of the EM algorithm: maximization (M) and expectation (E) as follows:

$$\mathbf{M - step} : \hat{T}_{\text{next}} = \arg \max_T \left[\log P(T) + \sum_{j=1}^g (n(T)_j + w_j - 1) l_{j_{\text{old}}} \right] \quad (22)$$

$$\mathbf{E - step} : l_{j_{\text{next}}} = E_{\Theta|u,v,\hat{T}_{\text{old}}} [l_j], \quad (23)$$

In words, the M-step searches for transformation T that optimizes the sum of expectation over the log model parameters and a transformation prior; and the E-step calculates the expectation of these log parameters given the current best estimate of the transformation parameters. Note that in this framework, we need to pay special attention to the scenario where the θ_j parameters tend to zero and we also have to enforce the property that the parameters add to one. More details on these special cases are discussed below.

3.1 Evaluating the E-Step

In order to evaluate the expression in the E-step, we may use the following form for the posterior on Θ given the input images and the transform:

$$P(\Theta|u, v, T) = \frac{P(u, v|\Theta, T)P(\Theta)}{\int_{\Theta} P(u, v|\Theta, T)P(\Theta)d\theta} \quad (24)$$

Computing the expectation term thus becomes:

$$E_{\Theta|u,v,\hat{T}} [\log \theta_l] = \int_{\Theta} \log \theta_l P(\Theta|u, v, \hat{T})d\theta = \int_{\Theta} \log \theta_l \frac{P(u, v|\Theta, \hat{T})P(\Theta)}{\int_{\Theta} P(u, v|\Theta, \hat{T})P(\Theta)d\theta} d\theta. \quad (25)$$

Using our multinomial model and the Dirichlet distribution as its prior, the numerator in Eq.(25) has the form of another unnormalized Dirichlet distribution.

Therefore, if we define a new Dirichlet distribution with parameters β , where $\beta_i = (n(\hat{T})_i + w_i)$, we may write

$$E_{\Theta|u,v,\hat{T}} [\log \theta_l] = \int_{\Theta} \log \theta_l \frac{\prod_{j=1}^g \theta_j^{n(\hat{T})_j} \frac{1}{Z(w_0)} \prod_{k=1}^g \theta_k^{(w_k-1)}}{\int_{\Theta} \prod_{j=1}^g \theta_j^{n(\hat{T})_j} \frac{1}{Z(w_0)} \prod_{k=1}^g \theta_k^{(w_k-1)} d\theta} d\theta \quad (26)$$

$$= E_{\text{Dir}(\Theta;\beta)} [\log \theta_l] \quad . \quad (27)$$

The expression in Eq.(27) is the expected value of the $(\log \theta_l)$ parameters given a Dirichlet distribution parameterized by β . As the Dirichlet distribution belongs to the family of exponential functions, the expectation over its sufficient statistics can be computed as the derivative of the logarithm of the normalization factor with respect to its natural parameters[13]. Writing the Dirichlet distribution in its exponential form results in

$$\text{Dir}(\Theta; w) = \exp\left(\sum_{i=1}^g (w_i - 1) \log \theta_i + \log \Gamma(w_0) - \sum_{i=1}^g \log \Gamma(w_i)\right) \quad . \quad (28)$$

Given this form, we see that the sufficient statistics are indeed the $(\log \theta_l)$ parameters. Therefore, the expectation term we are concerned with can be expressed by using the Digamma (or Psi) function which is the first derivative of the log Gamma functions [14].

$$E_{\text{Dir}(\Theta;\beta)} [\log \theta_l] = \Psi(\beta_l) - \Psi(\beta_0) = \Psi(n(\hat{T})_l + w_l) - \Psi(N + w_0) \quad . \quad (29)$$

In the following, we approximate the Digamma expressions in Eq.(29) by using $\Psi(x) \approx \log(x - .5)$ [15]. This approximation is very accurate in the positive real domain, except for values in the range of $[0, 1]$. A value in that particular range would correspond to the extreme scenario of an “empty bucket”, where there are neither current nor prior observations associated with a particular model parameter. In order to make the approximation hold even in such a scenario, we differentiate between two cases: uninformative and informative priors. In the case of uninformative priors, we propose to use Laplace priors [5] instead of zero counts for the Dirichlet parameters. Consequently, when no prior / pseudo information is available for the model, we initialize all parameters uniformly as one. The case of informative priors requires more attention. Although the total sum of pseudo information in this case is significant, we cannot directly assume that it holds for all individual Dirichlet parameters. Thus in that case we explicitly require that each pseudo count should hold at least one count. With such specifications we can ensure that our approximation holds regardless of the nature of the prior information and that the argument of the log function, in the case of the parameter update, does not approach zero (which could be a concern as mentioned while defining the EM framework). Thus the E-step of the EM algorithm is expressed as

$$(\bar{l}_l)_{\text{next}} = \Psi\left(n(\hat{T}_{\text{old}})_l + w_l\right) - \Psi(N + w_0) \quad (30)$$

$$\approx \log \left(n(\hat{T}_{\text{old}})_l + w_l - .5 \right) - \log(N + w_0 - .5) \tag{31}$$

$$= \log \left(\frac{n(\hat{T}_{\text{old}})_l + w_l - .5}{N + w_0 - .5} \right) . \tag{32}$$

This rule assigns the logarithm of normalized sum of the pseudo and observed counts minus a constant to the most current $\log(\theta)$ parameters. This pooling of current and pseudo counts for describing joint intensity statistics has been already discussed in Sec.2.5. We point out that, in order to enforce the relationship $\sum_{i=1}^g \theta_i = 1$, we compute the E-step for all $1 \leq j \leq (g - 1)$ and we assign the last remaining parameter as $(\bar{l}_g)_{\text{next}} = \log \left[1 - \sum_{i=1}^{g-1} \exp((\bar{l}_i)_{\text{next}}) \right]$.

3.2 Evaluating the M-Step

Using the results from the E-step, we may now express the M-step (the most current estimate of the registration parameter), which indeed results in a simple formulation of the problem:

$$\hat{T}_{\text{next}} = \arg \max_T E_{\Theta|u,v,\hat{T}_{\text{old}}} [\log P(u, v, \Theta|T)] \tag{33}$$

$$= \arg \max_T \left[\log P(T) + \sum_{j=1}^g (n(T)_j + w_j - 1) E_{\Theta|u,v,\hat{T}_{\text{old}}} [l_j] \right] \tag{34}$$

$$= \arg \max_T \left[\log P(T) + \sum_{j=1}^g (n(T)_j + w_j - 1) \log \left(\frac{n(\hat{T}_{\text{old}})_j + w_j - .5}{N + w_0 - .5} \right) \right] \tag{35}$$

$$= \arg \max_T \left[\log P(T) + \sum_{j=1}^g n(T)_j \log \left(n(\hat{T}_{\text{old}})_j + w_j - .5 \right) \right] \tag{36}$$

Therefore, ignoring the prior on transformation T , our objective function in the M-step is the maximization of a log likelihood term. When solving the optimization problem with a non-informative set of priors, the solution approaches an *iterated* ML framework. This result underlies the fact that an MAP framework with a non-informative prior produces results that are equivalent to the ML solution. The best transformation parameters therefore are calculated by using old, currently the best, model estimates. In the scenario where we have significant amount of prior information with respect to the model, again, the pseudo and the current observations are pooled together.

We note that the expression in Eq.(36) can be further simplified to result in an information theoretic framework that comprises of a data- and model-related term. The former measures the KL-divergence between two categorical distributions over the parameters describing the current joint statistics and that

including *pseudo* observations from the prior model and the latter computes the entropy of the estimated joint statistics of the observations.

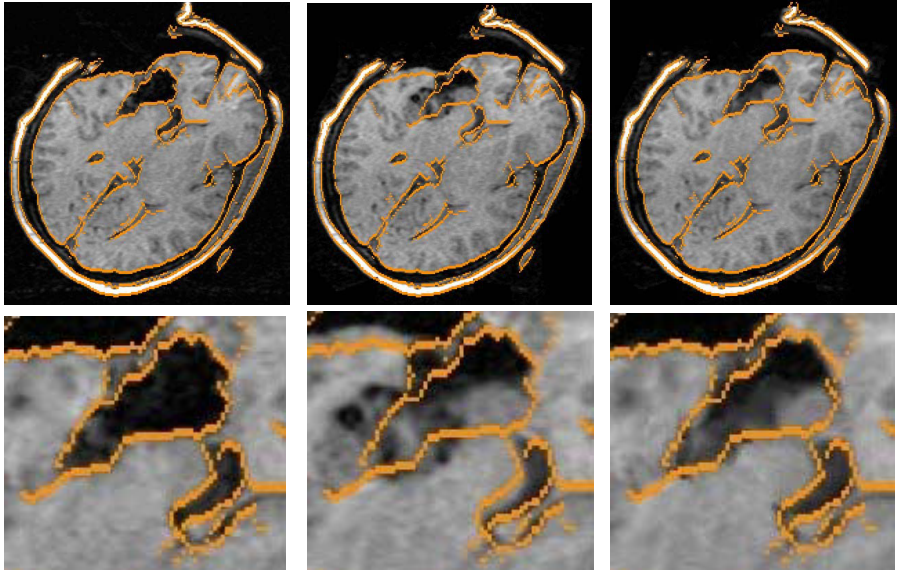
$$\begin{aligned} \hat{T}_{\text{next}} &\approx \arg \max_T \left[\log P(T) - D_{\text{KL}} \left(n(T) \parallel n(\hat{T}_{\text{old}}) + w_j - .5 \right) - H(n(T)) \right] \quad (37) \\ &\approx \arg \min_T \left[D_{\text{KL}} \left(p_T \parallel p_{\hat{T}_{\text{old}}, w} \right) + H(p_T) - \log P(T) \right] \quad (38) \end{aligned}$$

In summary, the utilization of the EM framework provides an iterative method for the transformation parameter estimation where the model parameter updates are computed in an efficient and principled manner.

4 Experimental Results

In this section, we present our experimental results from an iterated registration algorithm that uses uninformative priors. The resulting EM solution to the ML (in θ) formulation is demonstrated on a complex intra-operative non-rigid brain registration problem. We introduced a prior on the deformation field that approximates a fluid flow model. The algorithm also facilitates (on the EM iterations) a quadratic approximation to the search problem that may be quickly solved. That would not be justified if the multinomial model were not constant within the iterations. Preliminary experimental results for the scenario where we do have access to informative prior information was presented in [16]. At each iteration, our non-rigid registration algorithm defines the prior probability on configurations phenomenologically, using linear elastic deformation energy and Boltzmann’s Equation, as $P(T) \propto \exp(-E/\tau)$. Configurations of the deforming anatomy are parameterized by the coordinates of the vertices of an adaptive tetrahedral mesh. Using standard methodology of the Finite Element Method (FEM) [17], the stress-strain integral is linearized about the current configuration, yielding a quadratic approximation for the energy in terms of displacements from the current configuration, $E \approx \Delta_T^T K \Delta_T$, where K is a banded elasticity matrix. At each iteration, the log likelihood, $\log P(n(T)_{\text{old}} | \Delta_T)$, is approximated by a second order Taylor expansion. This is centered on a nearby local maximum of the likelihood function that we locate using a greedy method. The resulting quadratic expression is combined with the quadratic prior term, and the resulting approximation to the posterior probability is easily solved as a linear system. This iterative method resets the elastic energy to zero at each iteration. Papademetris et al. [18] call this “history free” approach the *incremental approach*, and point out that, in a limiting case, it is equivalent to the fluid model [19].

Figures 1 (a) and (b) show two MRI images of the same brain taken during brain surgery. The images contain 256x256x124 voxels with 0.9375x0.9375x1.5 mm spacing. According to visual comparison, deformation between the two images mostly occurs near the incision. The result of the non-rigid registration algorithm is shown in Fig. 1 (c). The warped image is very similar to the target image. In particular, the small displacement of the left and right ventricles was captured and the boundaries near the incision appear to be in appropriate



(a) Intra-operative slice at t_1 (b) Intra-operative slice at t_2 (c) Deformed acquisition

Fig. 1. MRI acquisition of the brain: (a) taken during surgery with edges highlighted; (b) taken later in the surgery, with the edges from (a) overlaid; (c) deformed the former image onto the latter, with the edges of (a) overlaid

places as well. The final match used 8500 mesh nodes, and the entire matching process was finished in less than 6 minutes on a 2 Ghz desktop computer. To validate our registration results, we compared manually located landmarks in the initial and the deformed images. The majority of the motion was captured by our non-rigid registration algorithm, with disagreements typically not more than the voxel size. More details on the experimental setup and the validation results is described in dissertation work [20].

5 Conclusion

We introduced an MAP framework for the pair-wise registration problem that allowed us to include prior information about not only the transformation parameter space but also the joint intensity statistics of the inputs. Treating the latter as nuisance parameters and marginalizing them out allowed us to establish a close relationship between our method and certain entropy-based objective functions. We also demonstrated that using an EM-based optimization framework the aligning transformations can be computed in a principled and computationally attractive manner.

Acknowledgements

This work was supported by grants NIH U41-RR019703, NCRR NAC P41-RR13218, NSF CISST/ERC EEC9731748, NCRR P41-RR14075, NCRR R01 RR16594-01A1, NINDS R01 NS052585-01 and by the UK BBSRC.

References

1. Zöllei, L., Fisher, J., Wells, W.: A unified statistical and information theoretic framework for multi-modal image registration. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 366–377. Springer, Heidelberg (2003)
2. Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology* 11(7180), 71–80 (2000)
3. Chung, A., Gan, R., Wells, W.: Robust multi-modal image registration based on prior joint intensity distributions and minimization of kullback-leibler distance. Technical Report HKUST-CS07-01, HKUST (2007)
4. Gütter, C., Xu, C., Sauer, F., Hornegger, J.: Learning based non-rigid multi-modal image registration using kullback-leibler divergence. In: Gerig, G., Duncan, J. (eds.) MICCAI. LNCS, vol. 2, pp. 255–263. Springer, Heidelberg (1973)
5. Gelman, A., Carlin, J., Stern, H., Rubin, D.: *Bayesian Data Analysis*. CRC Press LLC, Boca Raton, FL (2003)
6. Cover, T., Thomas, J.: *Elements of Information Theory*. John Wiley & Sons, Inc, New York (1991)
7. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition* 32(1), 71–86 (1999)
8. Zhu, Y., Cochoff, S.: Likelihood maximization approach to image registration. *IEEE Transactions on Image Processing* 11(12), 1417–1426 (2002)
9. Sabuncu, M., Ramadge, P.: Graph theoretic image registration using prior examples. In: *Proceedings of European Signal Processing Conference* (Sept 2005)
10. Leventon, M., Grimson, W.: Multi-modal Volume Registration Using Joint Intensity Distributions. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 1057–1066. Springer, Heidelberg (1998)
11. Chung, A., Wells, I.W., Norbash, A., Grimson, W.: Multi-modal Image Registration by Minimizing Kullback-Leibler Distance. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 525–532. Springer, Heidelberg (2002)
12. Dempster, A.P., Laird, N.M., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society* 39(1), 1–38 (1977)
13. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022 (2003)
14. Minka, T.: *Estimating a dirichlet distribution*. Notes (2003)
15. Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions*. Dover, New York (1964)
16. Zöllei, L., Wells, W.: Multi-modal image registration using dirichlet-encoded prior information. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 34–42. Springer, Heidelberg (2006)
17. Zenkiewicz, O., Taylor, R.: *The Finite Element Method*, 4th edn. McGraw-Hill, Berkshire, England (1989)

18. Papademetris, X., Onat, E.T., Sinusas, A.J., Dione, D.P., Constable, R.T., Duncan, J.S.: Specification and Compositional Verification of Real-Time Systems. LNCS, vol. 0558, pp. 36–49. Springer, Heidelberg (2001)
19. Christensen, G., Rabbit, R., Miller, M.: Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing* 5(10), 1435–1447 (1996)
20. Timoner, S.: Compact Representations for Fast Nonrigid Registration of Medical Images. PhD thesis, Massachusetts Institute of Technology (July 2003)