Effects of Registration Regularization and Atlas Sharpness on Segmentation Accuracy

B.T. Thomas Yeo[1], Mert R. Sabuncu[1], Rahul Desikan[3], Bruce Fischl[2], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT, USA
[2] Athinoula A. Martinos Center for Biomedical Imaging, MGH/MIT/HMS, USA
[3] Boston University School of Medicine

Abstract. In this paper, we propose a unified framework for computing atlases from manually labeled data at various degrees of "sharpness" and the joint registration-segmentation of a new brain with these atlases. In non-rigid registration, the tradeoff between warp regularization and image fidelity is typically set empirically. In segmentation, this leads to a probabilistic atlas of arbitrary "sharpness": weak regularization results in well-aligned training images and a "sharp" atlas; strong regularization yields a "blurry" atlas. We study the effects of this tradeoff in the context of cortical surface parcellation by comparing three special cases of our framework, namely: progressive registration-segmentation of a new brain to increasingly "sharp" atlases with increasingly flexible warps; secondly, progressive registration to a single atlas with increasingly flexible warps; and thirdly, registration to a single atlas with fixed constrained warps. The optimal parcellation in all three cases corresponds to a unique balance of atlas "sharpness" and warp regularization that yield statistically significant improvements over the previously demonstrated parcellation results.

Key words: Registration, Segmentation, Parcellation, Multiple Atlases, Markov Random Field, Regularization.

1 Introduction

Automatic labeling of cortical brain surfaces is important for identifying regions of interests for clinical, functional and structural studies [3, 14]. Recent efforts have ranged from the identification of sulcal/gyral ridge lines [15, 17] to the segmentation of sulcal/gyral basins [3, 5, 8, 9, 11, 13, 14]. Similar to these prior studies, we are interested in parcellation of the entire cortical surface meshes, where each vertex is assigned a label, given a training set of manually-labeled cortical surfaces.

Probabilistic atlases are useful and prevalent in segmentation literature [3, 5, 12]. A typical initial step in atlas computation is the spatial normalization of the training images. Measures like the prior probability of a certain label at a location can then be computed. Spatial normalization can be achieved with different registration algorithms that can vary in the rigidity of warps, from affine [12] to fully nonrigid warps [5]. More restricted warps yield "blurrier" atlases that capture inter-subject variability of structures, enlarging the basin of attraction for the registration of a new subject. However, the trade-off between

robustness and accuracy, which is limited by the "sharpness" of the atlas, is typically ignored. Recent research [10] has shown that combining information about the warp rigidity and the residual image of the registration process can improve the classification accuracy of schizophrenic and normal subjects.

Finding the optimal warp regularization tradeoff has gathered attention in recent years. Twining *et al.* [2] and Van Leemput [7] propose frameworks to find the least complex models that explain the image intensity and segmentation labels respectively in the training images. This is useful if the goal is to analyze the training images. However, if the goal is new subject segmentation, then segmentation accuracy should drive the choice of warp regularization. An interesting question is whether outlier images require weaker regularization to warp closer to the population "average" and obtain better segmentation accuracy.

Joint registration-segmentation algorithms are generally more effective than sequential registration-segmentation as registration and segmentation benefit from additional knowledge of each other [1, 12, 18, 19]. This paper proposes a joint registration-segmentation framework with Markov Random Field (MRF) priors on both segmentation labels and registration warps that incorporates multiple atlases. Our framework is an extension of Pohl *et al.* [12] and Fischl *et al.* [5].

Here we study the effect of atlas "sharpness" and warp regularization on segmentation accuracy. In particular, we compare 3 specific cases: (1) progressive registration of a new brain to increasingly "sharp" atlases using increasingly flexible warps, by initializing each registration stage with the optimal warps from a "blurrier" atlas; (2) progressive registration to a single atlas with increasingly flexible warps; (3) registration to a single atlas with fixed constrained warps.

Another contribution of this paper is the consistency of the co-registration of labeled images when computing the atlas and the normalization of an unlabeled test image to the atlas, i.e., we treat the atlas creation and new subject registration within the same framework. Finally, to the best of our knowledge, this is the first implementation of a joint registration-segmentation algorithm applied to the labeling of the cerebral cortex.

2   Theory and Implementation

2.1   Joint Registration and Segmentation

We define an atlas A_α to be an atlas trained from images aligned with warp smoothness parameter $S = \alpha$. We distinguish between α and S since one can use an atlas A_α for the registration of a subject with smoothness S where $\alpha \neq S$. Let R be the registration parameters, T be the image segmentation and Y be the observed image features. We obtain the joint registration-segmentation of a new image given a set of atlases A_α and smoothness S by maximizing the Maximum-A-Posteriori (MAP) objective function:

$$(R^*, T^*, S^*, A_\alpha^*) = \underset{R,S,A_\alpha,T}{\arg\max}\, p(R, T, S, A_\alpha | Y) = \underset{R,S,A_\alpha,T}{\arg\max}\, p(R, T, S, A_\alpha, Y) \quad (1)$$

where $p(\cdot)$ denotes probability. We can optimize Eq. (1) with coordinate ascent by iteratively optimizing one of the variables while fixing the remaining variables. Unfortunately, each iteration only finds the posterior mode of a random

variable while disregarding its entire posterior distribution. Instead, we consider a variant of Eq. (1) where we marginalize over the segmentation T. Furthermore, for simplicity, we assume a uniform prior on S and A_α or equivalently treat them as non-random parameters, leading to the following formulation of the problem:

$$(R^*, S^*, A_\alpha^*) = \arg\max_{R,S,A_\alpha} \sum_T p(R,T,Y;S,A_\alpha) = \arg\max_{R,S,A_\alpha} \log \sum_T p(R,T,Y;S,A_\alpha). \quad (2)$$

This optimization can be solved by the Expectation-Maximization algorithm with the segmentation T as a hidden variable. To reduce computation time, instead of working with the continuous parameters α and S, we discretize S and A_α into a finite set $\{S, A_\alpha\} = \{S_1 > S_2 > \cdots > S_N, A_{S_1}, A_{S_2} \cdots A_{S_N}\}$, where larger values of α and S correspond to "blurrier" atlases and more restricted warps, respectively. The optimization criterion then becomes

$$\max_{S,A_\alpha} \left\{ \max_R \log \sum_T p(R,T,Y;S,A_\alpha) \right\} \quad (3)$$

With enough samples, the finite set $\{S, A_\alpha\}$ should sufficiently represent the underlying continuous space of atlases and warps. Given an unlabeled brain with image features Y, we consider the following schemes:

1. Multiple Atlas, Multiple Warp Scales (MAMS): multiscale approach where we optimize Eq. (3) w.r.t. R with "blurry" atlas A_{S_1} and warp regularization S_1, and use that to initialize the registration with sharper atlas A_{S_2} and warp regularization S_2, and so on.
2. Single Atlas, Multiple Warp Scales (SAMS): multiscale approach where we optimize Eq. (3) w.r.t. R with a fixed atlas A_{S_k} and warp regularization S_1, and use that to initialize the registration with the fixed atlas A_{S_k} and warp regularization S_2, and so on.
3. Single Atlas, Single Warp Scale (SASS): Optimize Eq. (3) w.r.t. R with a fixed atlas A_{S_k} and warp regularization S_m, where k might not be equal to m. This is most common in practice, especially when mixing and matching publicly available atlases and in-house registration algorithms.

Note that for each scheme, we do not search over values of S or α, but instead optimize the term within the curly brackets in Eq. (3), which we denote the mixed Maximum-Likelihood Maximum-A-Posteriori (ML-MAP) function:

$$\log \sum_T p(R,T,Y;S,A_\alpha) = \log p(R;S,A_\alpha) + \log \sum_T p(T,Y|R;S,A_\alpha) \quad (4)$$

$$= \log p(R;S) + \log \sum_T p(T,Y|R;A_\alpha) \quad (5)$$

where we have modeled the registration parameters R to be conditionally independent of the atlas A_α given the scale S. We also assume that both the segmentation T and the observation Y are independent of the scale S conditioned on the atlas A_α and registration R. Eq. (5) corresponds to the standard setup of the EM algorithm. In the E-step, we compute

$$Q(R;R^{(n)}) = \log p(R;S) + \sum_T p(T|Y,R^{(n)};A_\alpha) \log p(T,Y|R;A_\alpha) \quad (6)$$

where registration $R^{(n)}$ is obtained from the previous M-step. In the M-step we optimize $Q(R, R^{(n)})$ with respect to registration R:

$$R^{(n+1)} = \arg\max_{R} Q(R; R^{(n)}) \tag{7}$$

So far, the derivations have been general without any assumptions about the atlases A_α, the prior $p(R; S)$ or the image-segmentation fidelity $p(T, Y|R; A_\alpha)$.

2.2   Model Instantiation

We now apply the above framework to the joint registration-parcellation of cortical brain surfaces, represented by triangular meshes with a spherical coordinate system that minimizes metric distortion [4]. The aim is to register an unlabeled cortical surface to a set of manually labeled surfaces and classify each vertex of the triangular mesh into anatomical units.

We model the warp regularization with a MRF parameterized by S:

$$p(R; S) = \frac{F(R)}{Z_1(S)} \exp\left\{ -S\left[\sum_i \sum_{j \in \mathcal{N}_i} \left(\frac{d_{ij}^R - d_{ij}^0}{d_{ij}^0} \right)^2 \right] \right\} \tag{8}$$

where d_{ij}^R is the distance between vertices i and j under registration R, d_{ij}^0 is the original distance, $\mathcal{N}_i$ is a neighborhood of vertex i and $Z_1(S)$ is the partition function. Our regularization penalizes metric distortion weighted by a scalar S that reflects the amount of smoothness (rigidity) of the final warp. Function $F(\cdot)$ ensures invertibility and is zero if any triangle is folded by warp R and one otherwise.   Similarly, we impose a MRF prior on the parcellation labels:

$$p(T, Y|R; A_\alpha) = p(T|R; A_\alpha)p(Y|T, R; A_\alpha)$$
$$= \frac{1}{Z_2(A_\alpha)} \exp\left\{ \sum_i U_i(T_i; A_\alpha) + \sum_i \sum_{j \in \mathcal{N}_i} V(T_i, T_j; A_\alpha) \right\} \prod_i p(Y_i|T_i, R; A_\alpha) \tag{9}$$

where T_i and Y_i are vertex i's parcellation label and observation respectively. The local potential $U_i(T_i; A_\alpha)$ captures the frequency of label T_i at vertex i. The compatibility function $V(T_i, T_j; A_\alpha)$ reflects the likelihood of labels T_i and T_j being neighbors. $Z_2(A_\alpha)$ is the partition function dependent on the atlas A_α.

Incorporating Eq. (8, 9) into $Q(R; R^{(n)})$ of Eq. (6) and discarding all terms independent of R yields (with some work!)

$$Q(R; R^{(n)}) = \log F(R) - S \sum_i \sum_{j \in \mathcal{N}_i} \left(\frac{d_{ij}^R - d_{ij}^0}{d_{ij}^0} \right)^2 + \sum_i \sum_{T_i} \left\{ p(T_i|Y, R^{(n)}; A_\alpha)U_i(T_i; A_\alpha) \right.$$

$$\left. + \left[\sum_{j \in \mathcal{N}_i} p(T_i, T_j|Y, R^{(n)}; A_\alpha)V(T_i, T_j; A_\alpha) \right] + p(T_i|Y, R^{(n)}; A_\alpha)\log p(Y_i|T_i, R; A_\alpha) \right\} \tag{10}$$

where the first term prevents folding triangles, the second term penalizes metric distortion, the third and four terms are the Markov prior on the labels and the last term is the likelihood of the surface geometries given the segmentation.

2.3   Atlas Building

In our framework, the atlas construction is consistent with the registration of a new brain. Consider registering a training subject to the atlas A_{S_k} under the

smoothness S_k. Using the same objective function from before, the E-step is trivial since we know the ground truth segmentation T^*, and hence $p(T|Y, R^{(n)}; A_\alpha)$ is the delta function $\delta(T^* - T)$. Simplifying the M-step gives (for each subject):

$$R^* = \arg\max_R \log p(R; S_k) + \log p(T^*, Y|R; A_{S_k}). \qquad (11)$$

We can then warp each subject to a common coordinate system via registration R^*, and create atlas A_{S_k}. However, since A_{S_k} is unknown and yet appears on the right hand side of Eq. (11), we solve Eq. (11) using a fixed-point iterative method, where we initialize A_{S_k}, solve for the best R^*, create a new atlas A_{S_k} with R^* and repeat until convergence. In practice, we first create an atlas A_∞ after simple rigid-body registration and use it to initialize the creation of atlas A_{S_1}, where S_1 corresponds to an almost rigid warp. We then use atlas A_{S_1} to initialize the creation of atlas A_{S_2} where $S_1 > S_2$, and so on.

2.4   Implementation

The atlas A_S is defined by the local potential U, compatibility potential V and observation model $p(Y_i|T_i, R; A_\alpha)$. Features Y (sulcal depth and mean curvature) for the MRF and training follow that of Fischl *et al.* [5], except for simplicity, we set V to be spatially stationary and isotropic.

Computing $p(T_i|Y, R'; A_\alpha)$ and $p(T_i, T_j|Y, R'; A_\alpha)$ is NP-hard. Mean field approximation [6, 16] yields the following fixed-point iterative solution:

$$b_i(k) \propto e^{U_i(k) + \log p(Y_i|T_i = k, R; A_\alpha) + \sum_{j \in \mathcal{N}_i} \sum_{T_j} b_j(T_j)[V_{ij}(k, T_j) + V_{ji}(T_j, k)]} \qquad (12)$$

where $b_i(k)$ approximates the true belief $p(T_i = k|Y, R'; A_\alpha)$ and must be normalized at the end of each step. Estimating $p(T_i, T_j|Y, R'; A_\alpha)$ requires a variation of mean field [16]. To maximize over registration R in Eq. (10, 11), we warp each vertex of the mesh individually, and use conjugate gradient ascent with parabolic line search on a coarse to fine grid pyramid. The final segmentation is obtained by selecting the label with the highest posterior probability $p(T_i|Y, R^*; A_\alpha)$ for each vertex i and given atlas A_α.

3   Experiments

We consider 39 left hemispheres manually parcellated by a neuroanatomical expert, consisting of 35 labels (see Fig. 1). We use dice measure to evaluate segmentation quality and compare our results to the algorithm demonstrated by Fischl *et al.* [5] and extensively validated by Desikan *et al.* [3]. The benchmark algorithm is essentially "Single Atlas, Single Warp Scale" (SASS), but with sequential registration-segmentation and a more complex MRF model.

We perform cross-validation 4 times by leaving out subjects 1 to 10 in the atlas construction, followed by joint registration-segmentation of subjects 1 to 10. We repeat with subjects 11 to 20, 21 to 30 and finally 31 to 39. We select S to be the set $\{100, 50, 25, 12.5, 8, 5, 2.5, 1, 0.5, 0.25, 0.1, 0.05, 0.01\}$, where we find that in practice, $S = 100$ corresponds to allowing minimal metric distortion and $S = 0.01$ corresponds to allowing almost any distortion.

Fig. 2(a) shows a plot of average dice (defined as the ratio of cortical surface area with correct labels to the total surface area averaged over the test set) for

Lateral View   Medial View
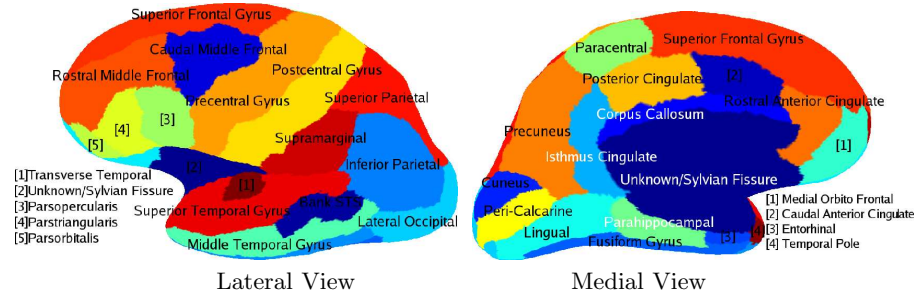
Fig. 1. Example of manual parcellation shown on a partially inflated cortical surface. Note that our neuroanatomist prefer gyri labels to sulci labels. There are also regions where sulci and gyri are grouped together as one label, such as the superior and inferior parietal complexes. Some labels in the images are of different colors for contrast.

SAMS ($A_\alpha = A_1$) and MAMS as we vary S. The average dice peaks at $S = 1$ for all cross-validation trials, although individual variation exists (not shown). Smaller values of regularization parameter S allow larger warps for the outlier subjects because the tradeoff in the cost function is skewed towards data-fidelity over regularization. However, it is surprising to find that the optimal S is mostly constant across subjects.

In general, the mixed ML-MAP objective function of Eq. (3, 4) is not a good measure of the optimal (S, α): the ML-MAP function continues to improve as S decreases due to overfitting. Instead, we use dice as an independent measure of selecting the optimal (S, α). Empirically, finding the best S for each individual subject only improves the average dice minimally and hence, we are contented with choosing $S = 1$.

Fig. 2(b) shows a plot of dice averaged over all 39 subjects. SAMS performs the best for $\alpha = 1$. For illustration, we see that $SAMS$ with $\alpha = 0.01$ starts off well, but eventually overfits with a worse peak at $S = 1$ ($p < 10^{-5}$ for one-sided paired-sampled t-test). Similarly for SASS, the best α and S are both 1. We also show SASS with $\alpha = 0.01$ and $S = 1$ in Fig. 2(b). The differences among MAMS, SAMS or SASS at their optimum, $\alpha = 1$, $S = 1$, are not statistically significant. On the other hand, their optimal performance is statistically significantly better than the benchmark, with all p-values less than 10^{-4}.

Because dice computed over the entire surface can be deceiving by suppressing small structures, we show in Fig. 3a the dice for each individual structure averaged over 39 subjects for MAMS ($S = 1$). The structures with the worst dice are the frontal pole and corpus callosum. Fig. 3a also shows the difference in dice between MAMS and the benchmark for each structure. For each structure, we perform the one-sided paired-sampled t-test between MAMS and the benchmark, where each subject is considered a sample. Results are shown in Fig. 3(b). Structures with p-values less than 0.05 are shown in dark blue. MAMS achieves statistically significant improvement over the benchmark for 16 structures. For the remaining 19 structures, the differences between the two segmentations are not statistically significant.
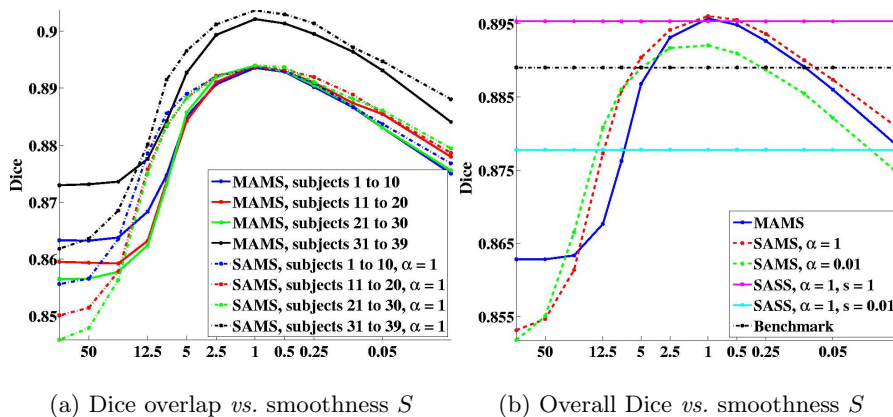
(a) Dice overlap *vs.* smoothness S (b) Overall Dice *vs.* smoothness S

Fig. 2. Summary of registration-segmentation results. S is plotted on a log scale.

The optimal SAMS and SASS perform similarly to optimal MAMS, which is the unfortunate result of the local nature of gradient ascent. This is especially problematic on cortical surfaces where two adjacent sulci might appear quite similar locally. Incorporating multiscale features with multiple warp scales should therefore be more effective. Ultimately, however, a major hindrance is the accuracy of the manual segmentation. In well-defined regions such as pre- and post-central gyri, accuracy is already as high as 95% and within the range of inter-rater variability. In ambiguous regions, such as the frontal pole, inconsistent manual segmentation leads to poor atlas and less accurate segmentation validation.

Further work involves the application of our framework to other data sets and experiment with other models of data fidelity and regularization. We expect the optimal smoothness would change but it would be interesting to verify if the optimal smoothness for a given experimental condition stays almost constant for all subjects.

To conclude, we proposed a joint registration-segmentation framework that incorporates consistent atlas construction, multiple atlases and MRF priors on both registration warps and segmentation labels. We showed that atlas "sharpness" and warp regularization are important factors in segmentation. With the proper choice of atlas "sharpness" and warp regularization, even with a less complex MRF model, the joint registration-segmentation framework has better segmentation accuracy than the state-of-the-art benchmark algorithm.

(a) Dice Measures   (b) Spatial Map of Segmentation Improvement
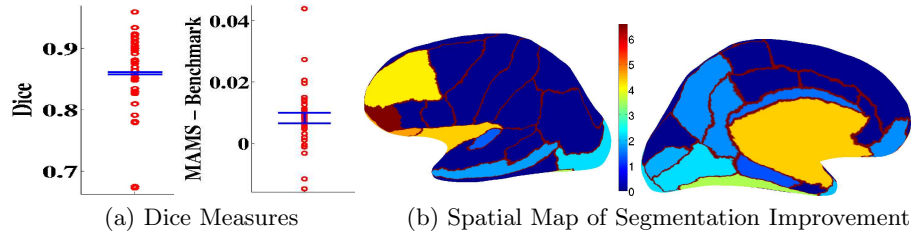
Fig. 3. (a) Dice of MAMS ($S = 1$) for individual structures (left) and improvement over the benchmark (right). (b) Spatial distribution of $-\log_{10}(p)$, where p is the p-value of the one-sided paired-sampled t-test between MAMS and the benchmark algorithm for each structure. Structures with p-values below 0.05 are colored dark blue.

References

1. Ashburner and Friston. Unified Segmentation. *NeuroImage*, 26:839-851, 2005.
2. Twining *et al.* A unified information theoretic approach to groupwise non-rigid registration and model building. *IPMI*, 9:1–14, 2005.
3. Desikan *et al.* An auto. labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31:968-980, 2006.
4. B. Fischl *et al.* Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*, 9(2):195–207, 1999.
5. B. Fischl *et al.* Automatically Parcellating the Human cerebral Cortex. *Cerebral Cortex*, 14:11-22, 2004.
6. T. Kapur *et al.* Enhanced Spatial Priors for Segmentation of Magnetic Resonance Imagery. *MICCAI*, 457-468, 1998.
7. K. Van Leemput. Probabilistic Brain Atlas Encoding Using Bayesian Inference. *MICCAI*, 704–711, 2006.
8. A. Klein and J. Hirsch. Mindboggle: a scatterbrained approach to automate brain labeling. *NeuroImage*, (24)261-280, 2005.
9. G. Lohman and D.Y. von Cramon. Automatic labelling of the human cortical surface using sulcal basins. *Medical Image Analysis*, 4:179-188, 2000.
10. S. Makrogiannis *et al.* A Joint Transformation and Residual Image Descriptor for Morphometric Image Analy. using an Equiv. Class Formulation. *MMBIA*, 2006.
11. Mangin *et al.* Auto. construction of an attributed relational graph representing the cortex topography using homotopic trans.. *Proc. SPIE*, 2299:110-121, 1994.
12. K.M. Pohl *et al.* A Bayesian model for joint segmentation and registration. *NeuroImage*, 31:228-239, 2006.
13. M.E. Rettmann *et al.* Automated Sulcal Segmentation Using Watersheds on the Cortical Surface. *NeuroImage*, 15:329-344, 2002.
14. D. Riviere *et al.* Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical Image Analysis*, 6:77-92, 2002.
15. X. Tao *et al.* Using a Statistical Shape Model to Extract Sulcal Curves on the Outer Cortex of the Human Brain. *Trans. Medical Imaging*, 21:513-524, 2002.
16. J. Zhang. The Mean Field Theory in EM Procedures for Markov Random Fields. *Second Edition, IEEE Transactions on signal Processing*, 40(10):2570-2583, 1992.
17. S. Zheng *et al.* A Learning Based Algorithm for Automatic Extraction of the Cortical Sulci. *MICCAI*, 695-703, 2006.
18. P. Wyatt and J. Noble. MAP MRF Joint Segmentation and Registration. *MICCAI*, 580–587, 2002.
19. A. Yezzi *et al.* A variational Framework for Joint Segmentation and Registration. *IEEE MMBIA*, 2001.