Active Mean Fields: Solving the Mean Field Approximation in the Level Set Framework

Kilian M. Pohl[1,2], Ron Kikinis[1], and William M. Wells[1,2]

[1]Surgical Planning Laboratory, http://www.spl.harvard.edu, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA
{pohl,kikinis,sw}@bwh.harvard.edu
[2] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA.

Abstract. We describe a new approach for estimating the posterior probability of tissue labels. Conventional likelihood models are combined with a curve length prior on boundaries, and an approximate posterior distribution on labels is sought via the Mean Field approach. Optimizing the resulting estimator by gradient descent leads to a level set style algorithm where the level set functions are the logarithm-of-odds encoding of the posterior label probabilities in an unconstrained linear vector space. Applications with more than two labels are easily accommodated. The label assignment is accomplished by the Maximum *A Posteriori* rule, so there are no problems of "overlap" or "vacuum". We test the method on synthetic images with additive noise. In addition, we segment a magnetic resonance scan into the major brain compartments and subcortical structures.

1 Introduction

Many clinical researchers rely on automatic segmentation techniques to analyze medical images [1]. Popular approaches for this task are curve evolution methods, which evolve the boundary of an object coupling image data with smoothness constraints of a zero-level set [2–9]. Some of these methods evolve multiple zero-level sets but they usually do not provide a simple interpretation for overlapping curves. We address this issue by using an alternative representation called LogOdds that views the entire level set function as a representation of posterior probabilities of label maps.

We derive the corresponding curve evolution framework, called Active Mean Fields (AMF), by revisiting the Mean Field approximation; a method frequently used in medical imaging for estimating the posterior probabilities of label maps [10, 11]. When estimating the solution to the Markov Random field model [12], simplifications result from approximating some random field variables by their mean value. Similar to other approximations of Markov Random field models [13–15] the methods by [10, 11] lack the notion of objects' boundaries – this often leads to fragmented label maps. We address this issue by incorporating a curve length prior into the Mean Field model. This results in the AMF algorithm, which approximates the solution via a level set framework in the LogOdds space.

The contributions of this paper are three-fold. First, we derive a new level set representation based on multinomial Logarithm-of-Odds (LogOdds). For the probability p of a binary variable, the LogOdds (also called logit) is the logarithm of the ratio between the probability p and its complement $1-p$. As a generalization of [16], here LogOdds defines a vector space structure that relates the evolution of multiple curves in the level set formulation to space conditioned probabilities. An advantage of this new representation is that it replaces the potentially ambiguous interpretation of overlapping zero-level sets with a simpler Maximum *A Posteriori* (MAP) Probability criteria.

Second, we compute the Mean Field solution of the posterior probabilities of label maps via a level set formulation. We do so by projecting the probabilities into the vector space of LogOdds maps and determining the solution via gradient descent. This, combined with our choice of prior model, results in a curve evolution algorithm coupling the curve shortening prior from the level set model with the posterior probabilities traditionally associated with the Mean Field approximation. The resulting curve evolution, called AMF, not only updates the zero-level set but also evolves the entire family of curves, as it is common in the level set community, which now correspond to levels of the posterior probabilities of labels.

Third, to the best of our knowledge, this is the first time for a level set framework to simultaneously segment 3D MR images into the three major brain compartments and subcortical structures. As we show in our example, if AMF is initialized by a noisy automatic segmentation [17] it can improve the 3D segmentations by removing outliers and islands that violate the smoothness constrains of the prior model.

This paper is organized as follows. In Section 2, we provide the mathematical definition of LogOdds, as well as their relationship with discrete probabilities. In Section 3, we derive the AMF which approximates the Mean Field solution via a level set framework. In Section 4, we apply AMF to synthetic and medical images.

2 Multinomial LogOdds

In this section, we generalize the binomial LogOdds representation discussed in [16] to discrete distributions, which we call multinomial LogOdds. We show that the LogOdds space has a one-to-one mapping to the space of discrete probabilities and defines a vector space. These two properties are very important for the derivations in Section 3 where we determine an approximation for the Mean Field solution via gradient descent.

LogOdds are an example of a class of functions that map the space of discrete distributions [18] to the Euclidean space. Let $\mathbb{P}_M$ be the open probability simplex for M labels $\mathbb{P}_M = \left\{ p \,|p = (p_1, \ldots, p_{M-1}, 1 - \sum_{i=1,\ldots,M-1} p_i) \in (0,1)^M \right\}$. Note that $\mathbb{P}_M$ is an M-1 dimensional space as the M^{th} entry is defined by the first M-1 entries. Furthermore, the space is open avoiding distributions that are certain about the assignment. For the specific case of $M = 2$, $\mathbb{P}_2 = \{(p, 1-p) | p \in (0,1)\}$ is the Bernoulli distribution [19]. Many binary classification problems use the Bernoulli distribution where p represents the probability that a voxel belongs to a particular anatomical structure and its complement $\bar{p} = 1 - p$ represents the probability of the voxel being in the background.

The multinomial LogOdds function $\text{logit}(\cdot) : \mathbb{P}_M \rightarrow \mathbb{R}^{M-1}$ of a discrete distribution $p \in \mathbb{P}_M$ is defined as the logarithm of the ratio between the i^{th} and last entry of p:

$$[\text{logit}(p\,)]_i \triangleq \log\left(\frac{p_i}{p_M}\right),$$

with $i \in \{1,\ldots,M-1\}$. The inverse of the log odds function $\text{logit}(\cdot)$ is the generalized logistic function

$$[\sigma(t\,)]_i \triangleq \begin{cases} \frac{e^{t_i}}{Z} & \text{, for } i \in \{1,\ldots,M-1\} \\ \frac{1}{Z} & \text{, if } i = M \end{cases}, \tag{1}$$

where $Z \triangleq 1 + \sum_{j=1,\ldots,M-1} e^{t_j}$ is the normalization factor. Having defined $\text{logit}(\cdot)$ and $\sigma(\cdot)$, we now induced from $\mathbb{P}_M$ the M-1 dimensional space of LogOdds $\mathbb{L}_{M-1} \triangleq \{\text{logit}(p) | p \in \mathbb{P}_M\}$. Note that $\mathbb{L}_{M-1}$ is equivalent to (M-1) dimensional real vector space. In Appendix A, we make use of this vector space structure to induce a vector space on $\mathbb{P}_M$.

3 Approximating the Mean Field Solution via Curve Evolution

We now combine the Mean Field approximation with the level set framework by using the LogOdds parametrization. We do so by embedding the Mean Field parameters into the LogOdds space. We then determine the optimal parameters via gradient descent which we is realized in the level set formulation. This results in the AMF algorithm which computes space conditioned probabilities while incorporating regional as well as boundary properties of objects.

3.1 Using Gradient Descent

We now derive a model for segmenting medical images via the Mean Field approximation. The segmentation problem can be described as assigning each voxel of the image I to an anatomical compartment, which results in the label map $\mathcal{T}$. Without priors, the relationship between the label map $\mathcal{T}$ and the image I is generally unclear as the image might not visualize some anatomical boundaries or is corrupted by noise and other image artifacts. Some of these difficulties can be addressed by the use of prior models. This results in estimating posterior probabilities which can, in some cases, be accomplished via the Mean Field approximation [11, 15].

The Mean Field approach makes the problem of estimating the posterior probabilities $P(\mathcal{T}|I)$ feasible by approximating $P(\mathcal{T}|I)$ as a factorized distribution $Q(\mathcal{T};\theta) = \prod_x Q_x(\mathcal{T}_x;\theta)$, where θ are the parameters defining $Q(\mathcal{T};\theta)$. The approach now computes the parameter setting $\widehat{\theta}$ that minimize Kullback-Leibler (KL) divergence between the true posterior probability $P(\mathcal{T}|I)$ and the approximation $Q(\mathcal{T};\theta)$

$$D(Q(\mathcal{T};\theta)||P(\mathcal{T}|I)) = E_Q\left(\log\frac{Q(\mathcal{T};\theta)}{P(\mathcal{T}|I)}\right) = \sum_{\mathcal{T}\in\mathbb{T}} Q(\mathcal{T};\theta)\log\frac{Q(\mathcal{T};\theta)}{P(\mathcal{T}|I)},$$

where $\mathbb{T}$ is the space of all label maps $\mathcal{T}$ and $E_Q(\cdot)$ is the expected value. In a nutshell, the Mean Field approximation determines the solution to

$$\widehat{\theta} \triangleq \min_\theta D(Q(\mathcal{T};\theta)||P(\mathcal{T}|I)). \tag{2}$$

Frequently, the multinomial distribution $Q(\mathcal{T};\theta)$ is parametrized by the component probabilities, in which case Equation (2) is a constrained minimization problem. We obtain an unconstrained problem by using the LogOdds parametrization.
$\theta \triangleq (\theta_1,\ldots,\theta_n) \in \mathbb{L}_M^n$ parametrizes the multinomial distribution Q_x as
$Q_x(\mathcal{T}_x = j;\theta_x) = [\sigma(\theta_x)]_j$. Equation (2) is now an unconstrained problem whose solution can be approximated via the following gradient descent:

$$\theta^{(k+1)} = \theta^{(k)} - \lambda \cdot \frac{\partial}{\partial\theta} D(Q(\mathcal{T};\theta)||P(\mathcal{T}|I))_{|\theta=\theta^{(k)}},$$

where λ is the step size parameter.

In the remainder of this section we derive the update term, which can be rewritten using the expected value $E_Q[\log(P(\mathcal{T}))]$ of the log prior of the label map $\log(P(\mathcal{T}))$, and the KL divergence $D(Q(\mathcal{T};\theta)||P(I|\mathcal{T}))$ of the estimated probability distribution $Q(\mathcal{T};\theta)$ and label likelihood $P(I|\mathcal{T})$. For notational convenience, we will continue to use the KL divergence even when its second argument is a not a probability distribution over $\mathcal{T}$. (We also note that the likelihood could be re-normalized without affecting the solution.)

$$\begin{aligned}\frac{\partial}{\partial\theta} D(Q(\mathcal{T};\theta)||P(\mathcal{T}|I)) &= \frac{\partial}{\partial\theta} E_Q[\log(Q(\mathcal{T};\theta)) - \log(P(\mathcal{T}|I))] \\ &= \frac{\partial}{\partial\theta} E_Q[\log(Q(\mathcal{T};\theta)) - \log(P(I|\mathcal{T}))] - \frac{\partial}{\partial\theta} E_Q[\log(P(\mathcal{T}))] \quad (3) \\ &= \frac{\partial}{\partial\theta} D(Q(\mathcal{T};\theta)||P(I|\mathcal{T})) - \frac{\partial}{\partial\theta} E_Q[\log(P(\mathcal{T}))].\end{aligned}$$

The first term drives the estimate $Q(\mathcal{T};\theta)$ towards the normalized label likelihood $P(I|\mathcal{T})$. The prior $P(\mathcal{T})$ is defined in Section 3.3 in such a way that the second term $\frac{\partial}{\partial\theta} E_Q[\log(P(\mathcal{T}))]$ encourages smoothness along the boundary of the object.

3.2 The Derivative of the KL Divergence of $Q(\mathcal{T};\theta)$ and $P(I|\mathcal{T})$

To simplify the computation of the derivative of the KL divergence $D(Q(\mathcal{T};\theta)||P(I|\mathcal{T}))$ we assume that the likelihood of the label map $P(I|\mathcal{T}) = \prod_{x\in\mathbb{I}} P(I_x|\mathcal{T}_x)$ is factorized over the image domain $\mathbb{I}$, which is typically a valid assumption. In this case,
$D(Q(\mathcal{T};\theta)||P(I|\mathcal{T}))$ is the sum of KL divergences over $\mathbb{I}$:

$$D(Q(\mathcal{T};\theta)||P(I|\mathcal{T})) = \sum\nolimits_{x\in\mathbb{I}} D(Q_x(\mathcal{T}_x;\theta_x)||P(I_x|\mathcal{T}_x))$$

For temporary convenience, we omit the voxel index x. If we now denote the probability of label i according to the parameter $\theta \triangleq (\theta_1,\ldots,\theta_M)$ with $q_i \triangleq Q(\mathcal{T}=i;\theta) = [\sigma(\theta)]_i$ and the normalized likelihood of label i as $p_i \triangleq P(I|\mathcal{T}=i)$ then the derivative of the KL divergence with respect to θ_i is

$$\frac{d}{d\theta_i} D(Q(\mathcal{T};\theta)||P(I|\mathcal{T})) = \frac{d}{d\theta_i} q_i \log\frac{q_i}{p_i} + \frac{d}{d\theta_i} q_m \log\frac{q_m}{p_m} + \sum\nolimits_{j\neq\{i,m\}} \frac{d}{d\theta_i} q_j \log\frac{q_j}{p_j}. \quad (4)$$

The derivative of q_j is $\frac{d}{d\theta_i} q_j = \frac{d}{d\theta_i} \frac{e^{\theta_j}}{1+e^{\theta_i}+\sum_{j\neq i} e^{\theta_j}} = \begin{cases} -q_j q_i & , j \neq i \\ q_i(1-q_i) & , i = j \end{cases}$

and $\frac{d}{d\theta_i} q_j \log \frac{q_j}{p_j} = \begin{cases} -q_j q_i (1 + log \frac{q_j}{p_j}) & , j \neq i \\ q_i(1-q_i)(1 + log \frac{q_i}{p_i}), & i = j \end{cases}$ so that Equation (4) can be rewritten as

$$
\begin{aligned}
& q_i(1-q_i)(1+\log\frac{q_i}{p_i}) - q_i(1-q_i-\textstyle\sum_{j\neq\{i,m\}} q_j)(1+\log\frac{q_m}{p_m}) - q_i \textstyle\sum_{j\neq\{i,m\}} q_j (1+\log\frac{q_j}{p_j}) \\
=& q_i \cdot (1-q_i) \cdot (\log\frac{q_i}{p_i} - \log\frac{q_m}{p_m}) - q_i \textstyle\sum_{j\neq\{i,m\}} q_j \cdot (\log\frac{q_j}{p_j} - \log\frac{q_m}{p_m}) \\
=& [\sigma(\theta)]_i (1-[\sigma(\theta)]_i)(\theta_i - [\text{logit}(p)]_i) \quad - \textstyle\sum_{j\neq\{i,m\}} [\sigma(\theta)]_i [\sigma(\theta)]_j (\theta_j - [\text{logit}(p)]_j) \quad (5)
\end{aligned}
$$

When used within gradient descent, the derivative of KL divergence combines a driving force towards the LogOdds of the label likelihood $[\text{logit}(p_x)]_i$ with a second term, which we call the *coupling* term. In areas with high uncertainty ($q_{x_i} \approx 0.5$) the equation weighs heavily to move towards the LogOdds function $[\text{logit}(p_x)]_i$. However, if θ_{x_i} has high certainty about the label ($q_{x_i} \approx 0$ or $q_{x_i} \approx 1$) then the likelihood term is less important.

Unlike with binary representations of curves, our method allows zero-contours to overlap as the curves now represent level set of the posterior probabilities $Q(\mathcal{T};\theta)$, where $Q(\mathcal{T};\theta)$ is a multinomial distribution in $\mathbb{P}^n_m$. In $\mathbb{P}^n_m$, the probability maps indicate a label map via the MAP criteria.

3.3 Determining the Smoothing Term

We now compute $\frac{\partial}{\partial\theta} E_Q[\log(P(\mathcal{T}))]$, the second term of Equation (3). First, we define the probabilistic model for the label map prior $P(\mathcal{T})$ as a distribution preferring smooth boundaries in $\mathcal{T}$. We do so by making the prior $P(\mathcal{T})$ a function of the arc length of the binary maps that is defined by $\mathcal{T}$ [20]. For this purpose, we define $\mathcal{T}$ as a vector of indicator random variables $\mathcal{T}_x \in \{e_1, \ldots, e_m\}$ with the indicator $[e_j]_j = 1$ and zero otherwise. We can then extract from $\mathcal{T}$ a binary map $[\mathcal{T}]_i \triangleq (\mathcal{T}_{1_i}, \ldots, \mathcal{T}_{n_i})$ for each label i. The arc length $L([\mathcal{T}]_i)$ of the binary map $[\mathcal{T}]_i$ is defined as the length of the boundaries in $[\mathcal{T}]_i$. Based on the arc length for each label we can then specify the prior as $P(\mathcal{T}) \triangleq \frac{1}{Z} e^{-\sum_i L([\mathcal{T}]_i)} = \frac{1}{Z} \prod_i e^{-L([\mathcal{T}]_i)} \sim \prod_i P([\mathcal{T}]_i)$ and rewrite the derivative of the expected value as

$$
\begin{aligned}
\frac{\partial}{\partial\theta} E_{Q(\mathcal{T};\theta)}[\log(P(\mathcal{T}))] &= -\textstyle\sum_{i=1,\ldots,m} \frac{\partial}{\partial\theta} E_{Q(\mathcal{T};\theta)}[L([\mathcal{T}]_i)] \\
&= -\textstyle\sum_{i=1,\ldots,m} \frac{\partial}{\partial\theta} E_{Q([\mathcal{T}]_i;\theta)}[L([\mathcal{T}]_i)] .
\end{aligned}
$$

It is intractable to compute $E_{Q([\mathcal{T}]_i;\theta)}$ as we have to sum over all possible label maps $[\mathcal{T}]_i$ with $i = 1, \ldots, m$. According to Lemma 2 of Appendix B, however, an approximation for the expected value is

$$
\begin{aligned}
E_{Q([\mathcal{T}]_i;\theta)}[L([\mathcal{T}]_i)] &\approx -\int_{-\infty}^{\infty} \sigma(\alpha)(1-\sigma(\alpha)) L(\mathcal{H}([\theta]_i - \alpha)) d\alpha \\
&= -\int_{-\infty}^{\infty} \sigma(\alpha)(1-\sigma(\alpha)) \iint_{\mathbb{I}} \delta(\theta_{x_i} - \alpha) dx\, d\alpha,
\end{aligned}
$$

where $\mathcal{H}(y) \triangleq \{1 \text{ for } y > 0, 0 \text{ otherwise}\}$ is the Heaviside Function, $\delta(\cdot)$ the Dirac Delta function, and $[\theta]_i \triangleq (\theta_{1_i}, \ldots, \theta_{n_i})$ are the parameters of label i. The above approximation would be accurate if $L([\mathcal{T}]_i)$ were a sum of functions (or $P(\mathcal{T})$ were independent in space) as outlined in Lemma 2.

We compute the derivative of the above approximation by applying the Euclidean curve shortening flow [21], which states that $\frac{d}{d\theta}\int\int_{\mathbb{I}} \delta(\theta-\alpha)dx = \kappa_\alpha(\theta)|\nabla_\alpha(\theta)|\delta(\theta-\alpha)$, where $\nabla_\alpha(\theta)$ is the derivative of the LogOdds map θ with respect to the α-contour in θ and $\kappa_\alpha(\theta) = div(\frac{\nabla_\alpha(\theta)}{|\nabla_\alpha(\theta)|})$ is the corresponding curvature. Thus, we approximate the derivative of the expected value as the weighted integral over the curve shortening flow of all contours in the LogOdds map θ,

$$\frac{\partial}{\partial\theta} E_Q\left[\log(P(\mathcal{T}))\right] \approx -\int_{-\infty}^{\infty} \sigma(\alpha)(1-\sigma(\alpha))\kappa_\alpha(\theta)|\nabla_\alpha(\theta)|\delta(\theta-\alpha)d\alpha.$$

The derivative for each voxel location x and label i is defined as

$$\frac{\partial}{\partial\theta_{x_i}} E_Q\left[\log(P(\mathcal{T}))\right] \approx -\sigma(\theta_{x_i})\cdot(1-\sigma(\theta_{x_i}))\kappa_{\theta_{x_i}}(\theta_{x_i})|\nabla_{\theta_{x_i}}(\theta_{x_i})|. \tag{6}$$

Combing the results of this section, we compute the solution to the Mean Field approximation as defined in Equation (2) through the following curve evolution

$$\begin{aligned}\theta_{x_i}^{(k+1)} = \theta_{x_i}^{(k)} - \lambda\cdot\Big(&[\sigma(\theta_x^{(k)})]_i\cdot(1-[\sigma(\theta_x^{(k)})]_i)\cdot(\theta_{x_i}^{(k)} - [\text{logit}(p_x)]_i)\\ &-\sum\nolimits_{j\neq\{i,m\}}[\sigma(\theta_x^{(k)})]_i\cdot[\sigma(\theta_x^{(k)})]_j\cdot\left(\theta_{x_j} - [\text{logit}(p_x)]_j\right)\\ &+\cdot\sigma(\theta_{x_i}^{(k)})\cdot(1-\sigma(\theta_{x_i}^{(k)}))\cdot\kappa_{\theta_{x_i}^{(k)}}(\theta_{x_i}^{(k)})|\nabla_{\theta_{x_i}^{(k)}}(\theta_{x_i}^{(k)})|\Big).\end{aligned} \tag{7}$$

This update function defines the AMF algorithm. In a level set framework, the first term of the update formulation corresponds to the image coupling term. This coupling term is defined by the LogOdds of the corresponding normalized likelihoods, which are normally determined beforehand (e.g. with Gaussian classification techniques as in Section 4.3). The second part of our method defines the curve shortening flow, which controls the smoothness of the boundary. Both terms are weighted by the product $[\sigma(\theta_x^{(k)})]_i\cdot(1-[\sigma(\theta_x^{(k)})]_i)$ or $[\sigma(\theta_x^{(k)})]_i\cdot[\sigma(\theta_x^{(k)})]_j$ so that it may be possible to use the "narrow-band" style frequently discussed in the level set community.

The above derivations are greatly simplified by embedding θ in the vector space of LogOdds. The more usual parametrization requires each entry of θ to be confined to the interval $[0,1]$ and each vector θ_x needs to sum up to one. The corresponding gradient descent would therefore need to map each update to the manifold of discrete probabilities. Another advantage of the LogOdds representation is that our algorithm can simultaneously evolve multiple curves. The curves are level sets of LogOdds maps, which define posterior probabilities in our case. Applying MAP rule, each voxel is clearly assigned to a label. AMF is therefore free of complications with overlap or vacuum, which is a common problem in other multi-label level set formulations.

This completes our derivation of the AMF method. The resulting algorithm combines local constraints at each voxel location with global smoothness constraints of the boundaries.

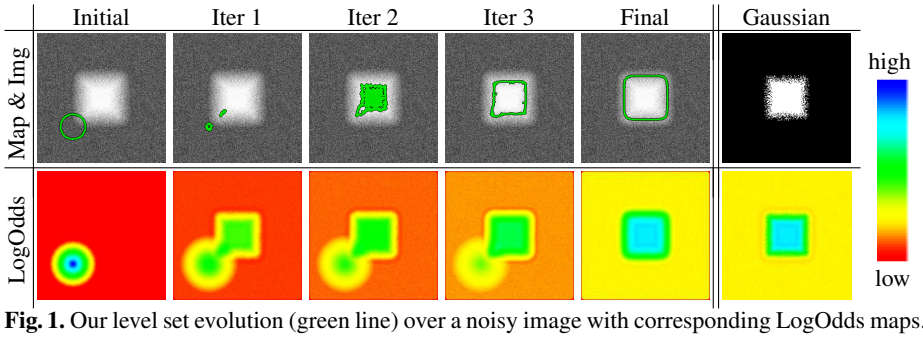

Fig. 1. Our level set evolution (green line) over a noisy image with corresponding LogOdds maps. The fragmented maps (last column) are obtained from a Gaussian likelihood model. In comparison, our results are smooth and connected even though the initial curve did not overlap with the square.

4 Experiments

We now apply the AMF to two examples. We first discuss the curve evolution of our algorithm on a noisy image that was segmented by a Gaussian classifier into a fragmented label map. The corresponding probability maps are the inputs to our algorithm, which robustly identifies the boundary of the structure. The second experiment includes real MRI images, in which AMF automatically segments the major brain compartments as well as subcortical structures. Due to the LogOdds parametrization, our method naturally evolves families of curves.

4.1 Segmenting Noisy Images

We now apply the AMF algorithm of Section 3 to a noisy image of a square (see top row of Figure 1). Before doing so, we compute the likelihood through a Gaussian intensity model, which results in a noisy LogOdds map (bottom, right) and, when thresholded, in a fragmented segmentation (top, right). The robustness of the classifier is greatly impacted by the noise as the approach ignores dependencies between neighboring voxels.

We initialize our curve evolution with the distance map of a small circle (see green circle in top, left image and distance map below) and the input is the noisy LogOdds map of the normalized likelihood (bottom, right). The initial curve is disconnected from the square forcing our method to split the zero-level set into two separate curves by Iteration 1. The circle connected to the square is expanding while the other curve is shrinking. Our curve evolution further evolves both curves until the connected curve converges to the shape of the square and the disconnected curve vanishes.

The evolution produces the LogOdds maps shown in the bottom row of Figure 1. Initially, the dark blue region shrinks, i.e. the number of voxels with high certainty about the presence of the square is decreasing. The shrinking is due to the discrepancy between the initial LogOdds map and the input label likelihoods. As the method progresses, the blue region assimilates towards the predefined LogOdds map. Unlike the segmentation produced through thresholding the initial likelihoods, our level set method filters out the noise. The final LogOdds map is smooth and the binary map shows the square as one connected region.

4.2 Segmenting Magnetic Resonance Images

In this experiment, we apply the AMF algorithm to a real 3D Magnetic Resonance (MR) scan (T1-weighted, matrix=$256\times256\times124$, dimension=$0.9375\times0.9375\times1.5$mm) to automatically segment the scan into the major brain compartments (gray matter = dark yellow, white matter = white, cerebrospinal fluid = blue) as well as the ventricles (right = yellow, left = pink), the thalamus (right = red, left = orange), and the caudate (right = turquoise, left = green). Figure 2 shows example slices of the segmentations, which were produced in 1.8 hours on a PC (dual processor Xoen, 3.0 GHz, 2 gig ram). We also segmented the scan using the approach of [17]. We determined the accuracy of each subcortical segmentation by computing its Dice score with respect to the manually generated label maps, which we view as ground truth.

The segmentation of [17] received a Dice Score of 0.778 for the left caudate, 0.770 for the right caudate, 0.895 for the left thalamus, and for the right thalamus 0.896. The label map is fragmented and has many misclassified regions. For example, in the temporal region a part of the skull is identified as gray matter. The segmentation AMF improves this segmentation using the corresponding space conditioned probabilities of [17] for the definition of the label likelihoods p_x in Equation (7). This improvement is also reflected in the Dice score, which is higher for each structure (left caudate: 0.789, right caudate: 0.774, left left thalamus: 0.897, right thalamus: 0.906).

The second and forth column of Figure 2 show example slices of the label map generated by our method. Our curve evolution model is not only robust enough to simultaneously segment the 3D Volume into 10 compartments, but also produces a much smoother label map with fewer islands than [17]. Unlike in the results of [17], the 3D model and slice of Example 1 show a subcortical region composed of oval-shaped structures, which closely match the expected anatomy in that region. In addition, the skull is properly separated from the brain (see temporal region in Example 2). Furthermore, the label map in the supra-sella region does not seem to be influenced by the noise in the image.

In the final experiment, we test the robustness of AMF by again segmenting the 3D MR image scan of Figure 2. This time the approach is initialized with a set of LogOdds maps representing nine circles as shown in Figure 3. Figure 3 shows the segmentation corresponding to the MR image of Example 1 of Figure 2. The method converges again to a solution that is very similar to the previously discussed results. Based on the these results, joining the Mean Field approach with the smoothness constraints of the level set formulation seems to be a robust framework for removing outliers and islands.

5 Conclusion

We described a new approach for estimating the posterior probabilities of tissue labels. We combined conventional likelihood models with a curve length prior on boundaries, and obtained posterior distributions by way of the Mean Field method. We used the LogOdds parametrization to facilitate optimization of the estimator by gradient descent, and with our choice of prior model, the influence of the prior is defined by the curve shortening flow. As demonstrated by our experiments, the approach can robustly segment multiple 3D objects in MR scans.

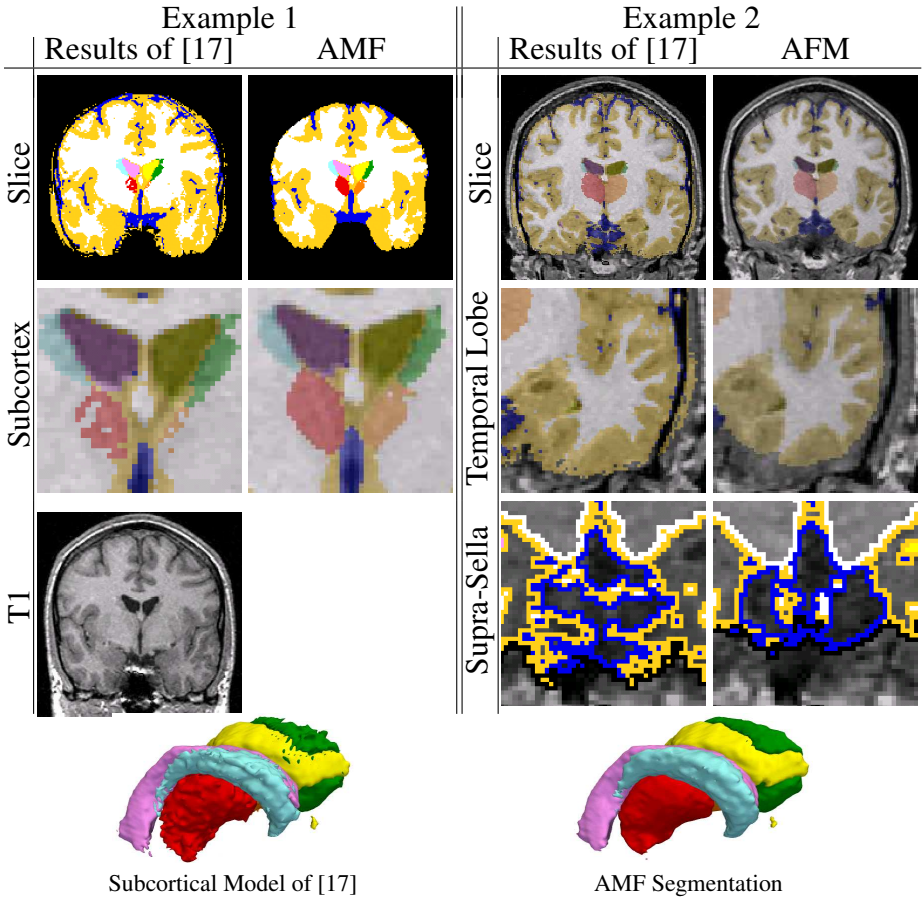

Fig. 2. The subcortical 3D models and two samples slices of the 3D data set segmented by [17] and AMF. The maps of [17] are fragmented and show many falsely identified regions. AMF produces smoother segmentations where most of the outliers are removed.

We tested the accuracy of our model by automatically identifying a single square in a noise synthetic image. In the final experiment, our approach segmented a 3D MR scan into the major brain compartments and subcortical structures; to our knowledge, this is the first time for a level set approach achieved this. The AMF algorithm accurately identified the structures and generated a smooth segmentation.

Acknowledgments: We thank Karl Krissen [22] for his level set implementation in the 3D Slicer (www.slicer.org), which we used as a starting point for our AMF implementation. We also appreciate the valuable comments by Sylvain Bouix, Torsten Rohlfing, Mert Sabuncu, and Kinh Tieu. This research was supported by the NIH (NIBIB NAMIC U54-EB005149, NCRR NAC P41-RR13218, NINDS R01-NS051826, NCRR mBIRN U24-RR021382, U41-RR019703), the NSF (JHU ERC CISST), US Army (SBIR W81XWH-04-C0031), and the Brain Science Foundation.

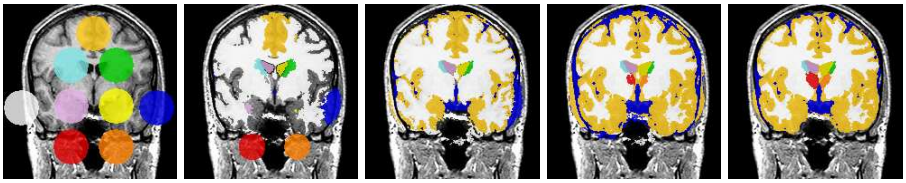

Fig. 3. Revisiting Example 1 of Figure 2, where the level set framework is initialized by nine circles. The resulting level set evolution is very similar to the segmentation of Figure 2.

Appendix A Define a Vector Space for Discrete Probabilities

The function $\text{logit}(\cdot)$ and its inverse comprise a homeomorphism between $\mathbb{P}_M$ and $\mathbb{L}_{M-1}$ so that we can borrow the vector space structure on $\mathbb{L}_{M-1}$ to induce one on $\mathbb{P}_M$.

A.1 Addition in $\mathbb{P}_M$

The *probabilistic addition* $p_a \oplus p_b$ in $\mathbb{P}_M$ is constructed by mapping p_a and p_b into LogOdds space, performing the addition between $\text{logit}(p_a)$ and $\text{logit}(p_b)$, and then mapping the result back into $\mathbb{P}_M$ via the logistic function. This operation is equivalent to a normalized multiplication of two discrete probabilities within $\mathbb{P}_M$:

$$p_a \oplus p_b \triangleq \sigma(\text{logit}(p_a) + \text{logit}(p_b)) = \frac{1}{\sum_{i=1,\ldots,M} p_{a_i} \cdot p_{b_i}} (p_{a_1} \cdot p_{b_1}, \ldots, p_{a_M} \cdot p_{b_M}). \quad (8)$$

$(\mathbb{P}_M, \oplus)$ with the zero element being the uniform distribution $\left(\frac{1}{M}, \ldots, \frac{1}{M}\right)$ forms an Abelian group as the *probabilistic addition* $\oplus$ is closed in $\mathbb{P}_M$. The additive inverse of a discrete probability $p \in \mathbb{P}_M$ is its complement $\bar{p}$, defined as $\bar{p}_i \triangleq \frac{1}{1+\sum_{j \neq i} \frac{p_i}{p_j}}$, for all $i \in \{1, .., M\}$. Similar to [16], it can be shown that for certain probabilistic models of p_a and p_b the *probabilistic addition* carries out the arithmetic of Bayes' rule.

A.2 Scalar Multiplication in $\mathbb{P}_M$

For $\mathbb{P}_M$ to be a vector space we also need to define a scalar multiplication operator. As with the probabilistic addition, the *probabilistic scalar multiplication* $\alpha \circledast p$ in $\mathbb{P}_M$ is defined as the logistic function of the product between α and the LogOdds $\text{logit}(p)$:

$$\alpha \circledast p \triangleq \sigma(\alpha * \text{logit}(p)) = \frac{1}{\sum_{i=1,\ldots,M} p_i^\alpha} (p_1^\alpha, \ldots, p_M^\alpha).$$

This operation is equivalent to exponentiating the discrete distribution with α and normalizing it. The technique of exponentiating and normalizing probabilities is frequently used in areas such as Markov Random fields [12] for controlling the "sharpness" of discrete distributions. As shown for the Binomial case in [16], α can also represent the certainty in the boundary location within an image space.

This completes our discussion of vector space $(\mathbb{P}_M, \oplus, \circledast)$ with 1 as the identity element of the scalar multiplication and $\bar{p} = -1 \circledast p$ the complement of p. By construction, this vector space is equivalent to $(\mathbb{L}_{M-1}, +, *)$ and its addition and scalar multiplication can be used to perform statistical computations in $\mathbb{P}_M^n$.

Appendix B Define the Expected Value via the Heaviside Function

Lemma 1: If Y is a random variable with Bernoulli distribution $P(Y=1) = p \in [0,1]$ then the expected value of $F(\cdot) : \{0,1\} \to \mathbb{R}$ with respect to Y can be defined by the integral over the LogOdds space and the Heaviside function $\mathcal{H}(y) \triangleq \{1 \text{ for } y > 0, 0 \text{ otherwise}\}$: $E_{P(Y)}(F(Y)) = \int_{-\infty}^{\infty} \sigma(\alpha) \cdot (1-\sigma(\alpha)) \cdot F(\mathcal{H}(\text{logit}(p) - \alpha)\, d\alpha$.

Proof: $E_{P(Y)}(F(Y)) \triangleq pF(1) + (1-p))F(0) = \int_0^p F(1)d\beta + \int_p^1 F(0)d\beta$

$$= \int_0^1 F(\mathcal{H}(p-\beta))d\beta = \int_{-\infty}^{\infty} F(\mathcal{H}(\text{logit}(p) - \alpha)) \left(\frac{d}{d\alpha}\sigma(\alpha)\right) d\alpha$$

$$= \int_{-\infty}^{\infty} \sigma(\alpha) \cdot (1-\sigma(\alpha)) \cdot F(\mathcal{H}(\text{logit}(p) - \alpha))\, d\alpha \qquad \blacksquare$$

Lemma 2: If $Y = (y_1, \ldots, y_n)$ is a vector of independent random variable with Bernoulli distribution $P = (p_1, \ldots, p_n)$ where $P(y_x = 1) = p_x \in [0,1]$ for $x \in \mathbb{I} = \{1, \ldots, n\}$ and the function $F(\cdot)$ is defined as $F(Y) = \sum_{x \in \mathbb{I}} f_x(y_x)$ then the expected value of F with respect to Y is defined within the LogOdds space as

$$E_{P(Y)}(F(Y)) \triangleq \int_{-\infty}^{\infty} \sigma(\alpha)(1-\sigma(\alpha))F(\mathcal{H}(\text{logit}(P) - \alpha))\, d\alpha$$

Proof: $E_{P(Y)}(F(Y)) = \sum_{x \in \mathbb{I}} E_{P(Y_x)}(F(Y_x))$. Then according to Lemma 1

$$E_{P(Y)}(F(Y)) = \sum_{x \in \mathbb{I}} \int_{-\infty}^{\infty} \sigma(\alpha) \cdot (1-\sigma(\alpha)) \cdot f_x(\mathcal{H}(\text{logit}(p_x) - \alpha))\, d\alpha$$

$$= \int_{-\infty}^{\infty} \sigma(\alpha) \cdot (1-\sigma(\alpha)) \sum_{x \in \mathbb{I}} f_x(\mathcal{H}(\text{logit}(p_x) - \alpha))\, d\alpha$$

$$= \int_{-\infty}^{\infty} \sigma(\alpha) \cdot (1-\sigma(\alpha))F(\mathcal{H}(\text{logit}(P) - \alpha)\, d\alpha \qquad \blacksquare$$

References

1. B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Sgonne, D. Salat, E. Busa, L. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A. Dale, "Automatically parcellating the human cerebral cortex," *Cerebral Cortex*, vol. 14, pp. 11–22, 2004.
2. A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 199–209, 1997.
3. K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset, "Model based detection of tubular structures in 3d images," *Computer Vision and Image Understanding*, vol. 80, no. 2, pp. 130–171, 2000.

4. M. Leventon, W. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1316 – 1323, 2000.
5. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137 – 154, 2003.
6. M. Rousson, N. Paragios, and R. Deriche, "Active shape models from a level set perspective," Tech. Rep. 4984, Institut National de Recherche en Informatique et en Automatique, 2003.
7. J. Yang, L. H. Staib, and J. S. Duncan, "Neighbor-constrained segmentation with level set based 3D deformable models," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 940–948, 2004.
8. M. Xu, P. Thompson, and A. Toga, "An adaptive level set segmentation on a triangulated mesh," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 191 –201, 2004.
9. P. Yushkevich, J. Piven, H. Hazlett, R. Smith, S. Ho, J. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 1, pp. 1116–28, 2006.
10. T. Kapur, *Model based three dimensional Medical Imaging Segmentation*. PhD thesis, Massachusetts Institute of Technology, 1999.
11. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897 – 908, 1999.
12. J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Society. Series B.*, vol. 48, no. 3, pp. 259–302, 1986.
13. B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. Dale, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, 2002.
14. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
15. J. Marroquin, E. Santana, and S. Botello, "Hidden markov measure field models for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1380–1387, 2003.
16. K. Pohl, J. Fisher, M. Shenton, R. W. McCarley, W. Grimson, R. Kikinis, and W. Wells, "Logarithm odds maps for shape representation," in *Medical Image Computing and Computer-Assisted Intervention*, no. 4191 in Lecture Notes in Computer Science, Springer-Verlag, 2006.
17. K. M. Pohl, J. Fisher, W. Grimson, R. Kikinis, and W. Wells, "A Bayesian model for joint segmentation and registration," *NeuroImage*, vol. 31, no. 1, pp. 228–239, 2006.
18. M. G. Kendall and W. R. Buckland, *A Dictionary of Statistical Terms*. Longman Group, 1976.
19. M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, ch. 4: Bernoulli Distribution, pp. 31 –33. Wiley, 3rd ed., 2000.
20. M. Taron, N. Paragios, and M.-P. Jolly, "Modelling shapes with uncertainties : Higher order polynomials, variable bandwidth kernels and non-parametric density estimation," in *IEEE International Conference on Computer Vision*, 2005.
21. M. Grayson, "The heat equation shrinks embedded plane curves to round points," *Journal of Differential Geometry*, vol. 26, no. 2, pp. pp. 285–314, 1987.
22. K. Krissian and C. Westin, "Fast sub-voxel re-initialization of the distance map for level set methods," *Pattern Recognition Letters*, vol. 26, no. 10, pp. 1532–1542, 2005.