

STATISTICAL ATLAS-BASED SUB-VOXEL SEGMENTATION OF 3D BRAIN MRI

Marcel Bosc^{1,2*}, Fabrice Heitz¹, Jean-Paul Armspach²

(1) LSIIT UMR-7005 CNRS / Strasbourg I University,
67400 Illkirch, France

(2) IPB UMR-7004 CNRS / Strasbourg I University,
67085 Strasbourg, France

ABSTRACT

We present a 3D brain MRI segmentation method in which a high resolution label image evolves under the influence of multiple constraints. The constraints are expressed in a versatile energy minimization framework which allows for evolutions only at label boundaries, effectively making it a surface evolution system. Constraints are defined using atlas-mapped parameters. The atlas, composed of a reference image and parameter values, is mapped onto the source image using a multi-resolution deformable image matching method. Variable scale image constraints are considered. The prior model currently includes: a relative distribution constraint, which gives the probability of observing a label at a given distance from another label, a thickness constraint and a surface regularization constraint. Issues related to partial volumes are addressed by the use of a high resolution label image and an accurate model of the acquisition process. High resolution segmentations are thus obtained from standard (eventually low resolution) MRIs.

1. INTRODUCTION

Automatically, and reliably segmenting an MR brain image is a difficult task [1, 2]. An expert, however, can easily distinguish one tissue type from another, using extensive prior knowledge on the shape, relative locations, and other characteristics of each tissue class. The goal in this automated segmentation system is to introduce strong prior knowledge, thus reducing the space of acceptable solutions, and increasing reliability. We seek to make prior knowledge explicit and express it in an intuitive manner. In many existing segmentation methods this knowledge is implicit and deeply hidden in the internal workings of the algorithm.

We believe that 3D image acquisition will evolve rapidly, changing the types of artifacts that hinder segmentation, eventually out-dating methods that are explicitly designed to address a specific artifact. Therefore we prefer putting the emphasis on developing prior models that describe the objects we want to segment. Certain organs have consistent shape characteristics that may be modeled, however, shapes like cortical folds are challenging, as they display very important anatomical variability and have a complex geometry.

The original approach presented here bears similarity with deformable surfaces methods in that surfaces evolve under the influence of constraints (figure 1). However, no attempt is made to use parametric or implicit level-set [2, 3] representations of the

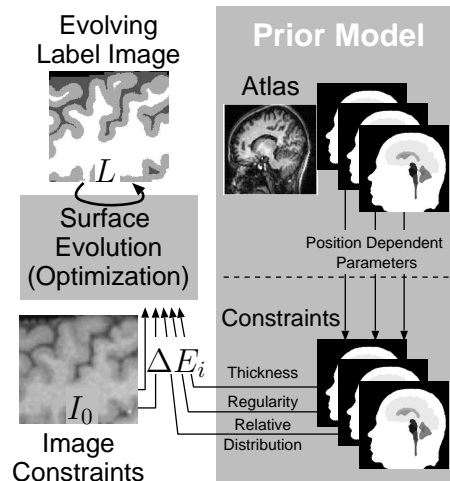


Fig. 1. Segmentation flowchart. A segmentation label image L evolves under the influence of multiple constraints coded by energy functions E_i . Constraints are defined using position dependent parameters mapped from a statistical atlas.

surface. Instead, the approach iteratively refines a segmentation label image L , at label boundaries. The label evolution framework (section 2) is simple and has the advantage of allowing topological changes. The label image is represented at a very high resolution, enabling the representation of fine details. Prior knowledge makes it possible to segment details that are smaller than the resolution of the source images. For example, in low resolution images ($2mm^3$) the cerebro-spinal fluid (CSF) regions within cortical folds are often too small to appear significantly. However we know that CSF regions are surrounded by gray matter of a relatively constant thickness. If gray matter regions are correctly positioned by the system, then CSF position will be implied (section 5.2.1). Note that the resolution of the label image is independent of the source image resolution. In particular, if the source image has non-isotropic sampling, this does not change the resolution of the label image (section 5.1.1).

2. LABEL EVOLUTION FRAMEWORK

At each iteration, labels in the segmentation map L ($L : \Omega \rightarrow \{1 \dots n\}$, where $\Omega \subset \mathbb{Z}^3$, figure 2) are changed so as to minimize

*This work was supported by the *Ligue Française Contre la Sclérose en Plaques* and Région Alsace through a student grant

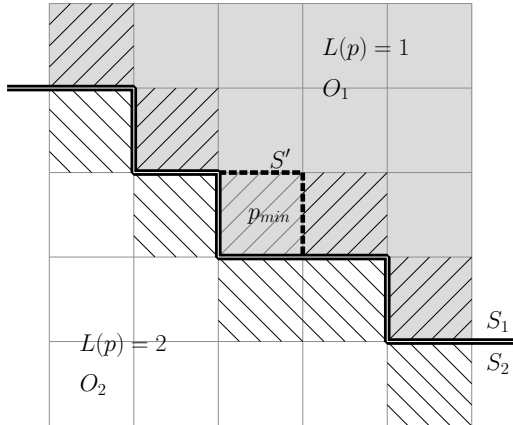


Fig. 2. Label evolution framework: close-up view of label image $L(p)$. Only labels at boundary voxels (hashed) may be changed. S_1 and S_2 are the boundary surfaces of objects O_1 and O_2 . Label image L is changed at point p_{min} where energy decrease is maximal. S' is the new boundary surface after the label has changed.

a global energy criterion

$$E(L) \triangleq \sum \alpha_i E_i(L) + E_b(L), \quad (1)$$

where $\alpha_i E_i$ are energy terms related to each constraint (see sections 3 and 5). The E_b term imposes that changes take place only at boundaries between labels, thus effectively making this a surface evolution system: $E_b = \infty$ if there are any isolated values in L , $E_b = 0$ otherwise.

At each iteration, for each point p on a boundary between two regions, we compute the energy change $\Delta E(p, l)$, if $L(p)$ were to change to a different label value l . This gives an energy variation image $\Delta E(p, l)$ for each different value of l . Note that this image is sparse, containing non zero values only at boundary voxels. We then change L at the point p_{min} that minimizes the energy (that is, $\Delta E(p_{min}, l_{min}) < \Delta E(p, l)$ for all p, l).

In theory ΔE should be recomputed on all points for each iteration, which would be computationally prohibitive. However, the impact of a label change has a limited range r on energy variation, meaning that changing a label at a position p has no significant impact on the value of the energy contribution at point p' if $\|p' - p\| > r$ (Markovian-like property). Therefore, at each iteration, possible label changes are sorted in increasing order of $\Delta E(p, l)$, and label changes are considered only if they occur at a position that is more distant than r to any previously changed label. This ensures that the total energy $E(L)$ decreases at each iteration. The energy interaction range r may be adjusted, depending on the constraints that are being considered.

The iterations stop when there are no more points where changing a label would decrease the total energy ($\Delta E(p, l) > 0$ for all p, l). This represents a local minimum of the energy function $E(L)$.

3. STATISTICAL ATLAS

Each constraint depends on parameters, which convey information on the shapes and sizes of the objects being segmented. The relative weight of each constraint in the final energy is defined by an

influence coefficient α_i . Any parameter or influence coefficient may be chosen to be position dependent. This allows for certain constraints being more important than others in certain regions of the image. For example, the relative distribution constraint (section 5.2.1) is essential for segmenting cortical folds, but has much less importance in ventricular regions. Or, in another example, the parameters expressing the expected width of gray matter may change for different areas of the brain.

The statistical atlas is made up of a reference MR image, a (low resolution) segmentation map and a mapping of the different parameters and influence coefficients. Statistical shape characteristics such as thickness, curvature are coded using probability distribution functions estimated from a training data-base. The statistical atlas is mapped onto the source image using a multi-resolution deformable image matching method developed by our group [4]. Probability distributions of shape characteristics have been estimated from the segmentation of 22 patients. The segmentations were computed semi-automatically using a watershed-based method [5]. Each patient image was registered on the atlas reference image using deformable registration [4]. This gives a series of possible values for a given atlas position, which are then modeled by parameterized probability distribution functions.

4. PREPROCESSING AND INITIALIZATION

The high resolution segmentation is computed from the unmodified source image, in its original position. This avoids applying geometrical transforms which involve re-sampling, particularly for non-isotropic images.

A non-linear intensity transfer function is fitted on the joint-histogram of the source image and the registered atlas reference image. This allows to determine the transfer function V that gives the mean image value corresponding to each label value (section 5.1.1).

The initial segmentation image $L(p)$ is obtained by standard segmentation techniques (thresholding) for regions that have strong anatomical variability, and by label mapping from the atlas for regions with low anatomical variability.

5. CONSTRAINTS

For each constraint we will give the expression of the corresponding energy function E_i , and then give the exact, or approximate expression of the energy variation $\Delta E_i(p, l) \triangleq E'_i - E_i$ due to the the label image L changing to l at position p . There are two types of constraints: image constraints and prior model constraints. Note that the constraints defined below are actually constraint classes, and that several instances of a constraint class may be used. For example, several relative distribution constraints (section 5.2.1) may be combined, each one modeling a distribution around a different label.

5.1. Image Constraint

To guarantee sub-voxel precision, image constraints need an accurate model of the acquisition process.

5.1.1. Image Acquisition Model

The physical acquisition system is modeled as a two step process. First a mean intensity is associated to a tissue by an intensity map-

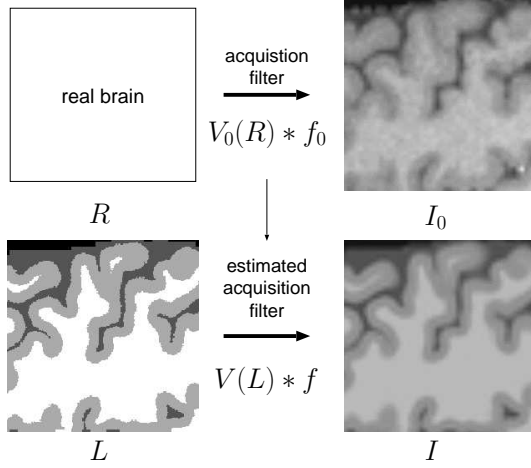


Fig. 3. Acquisition model: L is the current segmentation. I_0 is the acquired image that we are segmenting. I is the image that is predicted from L by the acquisition model. Image constraints attempt to minimize the difference between I and I_0 .

ping V_0 , then a low-pass filtering f_0 is performed. In our system we will mimic this acquisition process, applying an intensity mapping V followed by a low-pass filter f (Figure 3).

Many image segmentation methods, implicitly, or explicitly compare a segmented image $V(L)$ (which is piece-wise constant) to the acquired source image I_0 . This is not accurate, as the acquired image has undergone an acquisition process that involves low-pass filtering (filter f_0). This is the so-called partial volume effect, producing unclassifiable voxels.

The filter f is a Butterworth low-pass filter whose parameters are estimated from I_0 . The intensity mapping function V is estimated as explained in section 4.

Images with non-isotropic sampling are also well modeled in this framework.

5.1.2. Variable Scale Image Constraint

The energy associated with this constraint is a quadratic energy term, corresponding to a Gaussian observation model:

$$E \triangleq \|(I - I_0) * h\|^2 \quad (2)$$

h is a smoothing filter that determines the scale of this constraint. If h is a Dirac δ -function then we obtain a sub-voxel constraint that allows for precise positioning of the segmentation, but has a very short range. Its energy profile is very rough and in certain cases it might converge to a local minimum, far from the global optimum. Changing the spread of the smoothing filter h allows to smooth the energy profile and allows for long range image influence. Several variable image constraints may be used simultaneously.

The associated energy variation is:

$$\Delta E(p, l) = (V(l) - V(L(p)))^2 \|f * h\|^2 - 2(V(l) - V(L(p))) [(I_0 - I) * h * f](p) \quad (3)$$

where V is the function that maps label values to mean image values. Note that the convolution term $[(I_0 - I) * h * f]$ may be computed in low resolution. The only high resolution computation involved is the downsampling of $V(L)$.

5.2. Prior Model Constraints

Prior model constraints are independent of the source image. They describe the shapes and positions of the objects that are being segmented. We will call $O_k \triangleq \{p \in \Omega \mid L(p) = k\}$ the object (3D region) defined by all points of label k .

5.2.1. Relative Label Distribution Constraint

This constraint plays an important role in the segmentation of cortical folds. It is similar in its goal to the coupled surface approach described in [2]. It models the knowledge that white and gray matter are distributed in a specific manner around CSF in cortical folds. This is modeled as the probability of observing a label at a given distance of an object O_k :

$$E \triangleq \sum_{p \in \Omega} -\log(P(L(p) | D_k(p))) \quad (4)$$

where D_k is the (chamfer) distance map from object O_k . The probability P is a parameterized sigmoid function, whose parameters are position dependent. This constraint is also useful to model the relative positions of other tissues. The approximate associated energy variation is:

$$\Delta E(p, l) = -\log\left(\frac{P(l | D_k(p))}{P(L(p) | D_k(p))}\right) \quad (5)$$

5.2.2. Thickness Constraint

This constraint describes the thickness of an object O_k . Its associated energy is:

$$E \triangleq \sum_{p \in \bar{O}_k} -\log(P(t_k(p))) \quad (6)$$

where $t_k(p)$ is the thickness of label k at boundary voxel p . \bar{O}_k is the set of all boundary voxels inside O_k . $P(t)$ is the probability of observing a given thickness t .

Thickness: The thickness of an object O_k at a point x_0 could be defined by using the skeleton of the object. However, this definition would not be reliable, due to the inherent instability of the skeleton. We therefore propose a modified thickness definition.

Let d be the distance map inside O_k . Starting from point x_0 we extend a path $C(x_0) = (x_0 \cdots x_n)$ where x_{i+1} is the neighbor of x_i that has the largest value of $\frac{d(x_{i+1}) - d(x_i)}{\|x_{i+1} - x_i\|}$. The path is stopped when $\sum_{i=0}^{n-1} \|x_{i+1} - x_i\| > d(x_n) + \lambda$, where λ is a smoothing factor that represents the trade-off between precision and stability. We use $\lambda = 2$. The thickness is then defined as $t_k(x_0) = d(x_n)$. The approximate energy variation is:

$$\Delta E(p, l) = \mp \log\left(\frac{P(t_k(p) \pm 1)}{P(t_k(p))}\right) \quad (7)$$

depending on whether the thickness of the object k is increasing ($l = k$), or decreasing ($L(p) = k$).

5.2.3. Surface Regularization Constraint

This constraint describes the curvature of the surface S_k of an object O_k , thus imposing smoothness and orientation. S_k is the set of all boundary surface elements of O_k . A surface element is the

square that separates neighboring voxels with different values of $L(p)$. The energy associated to this constraint is:

$$E \triangleq \sum_{s \in S_k} -\log(P(K(s))) \quad (8)$$

where $K : S_k \rightarrow \mathbb{R}^3$ is a robust curvature operator that will be presented, and $P(K(s))$ is the probability of observing a given curvature at surface element s .

Most surface evolution methods include a regularization term that favors smooth surfaces. Parametric surface representations (splines) or implicit (level set [3]) representation can compute differential values of the surfaces. However differential values are local, by definition. What is desired in a surface evolution system, is to control the smoothness of surfaces at a given scale. Imposing strong regularization constraints at a very fine scale leads to stiffening the surface, and may block any further surface evolution. Moreover, computing differential properties is inherently unstable. This led us to introduce a different definition of curvature that is robust, is well defined for any continuous surface, and whose scale can be tuned.

For a point O on a surface S we define its curvature as:

$$K(O, Z) \triangleq \frac{4\pi}{Z^2} \int_Z \overrightarrow{OM} ds \quad (9)$$

where $Z \subset S$ is a surface patch containing O , Z^2 is the square of the area of patch Z , and M is a point on Z . We chose $Z = B_S(r)$ to be a geodesic disk of radius r : $B_S(r) \triangleq \{s \in S | d_S(O, s) < r\}$ (here, d_S is the geodesic distance on S). Note that $\int_Z \overrightarrow{OM} ds = \overrightarrow{OG}$ where G is the center of gravity of Z . It can be shown that $K(O, r) = K(O, B_S(r))$ correctly estimates surface curvature for small r . In the test case where S is a sphere of radius R , then $K(O, r) = \frac{1}{R}$. Note that this does not depend on the size r of Z . We have also shown that $K(O, Z)$ is fairly independent of the shape of Z . For correct results O should be close to the geodesic center of gravity of Z . This is only an issue at surface boundaries. In our system this may only occur at image edges, and is correctly dealt with.

The approximate value of the energy variation associated with this constraint is:

$$\Delta E(p, l) = - \left[\frac{\nabla P(K_p)}{P(K_p)} \right]^T dC \quad (10)$$

where dC is the displacement of the center of gravity of all surface elements surrounding p when the label at point p changes to l , and K_p is the average curvature on surface elements surrounding p .

6. EXPERIMENTAL RESULTS

The segmentation system has been successfully applied to both $1mm^3$ and $2mm^3$ images from different MRI machines. The images were segmented at label image L resolution of $.25mm^3$, which was found to be sufficient to represent thin structures in CSF regions. Figure 4 shows segmentation results after 100 iterations for a portion of brain containing cortical folds. Convergence was accelerated by setting $r = 0$ (section 2) for the first 75 iterations. With $r = 0$ boundary surfaces advance rapidly, and then oscillate. After 75 iterations, r is set to its correct value (1.5 times the grey matter width) and boundary surfaces evolve smoothly (without oscillations) to their final positions. Note that most image features

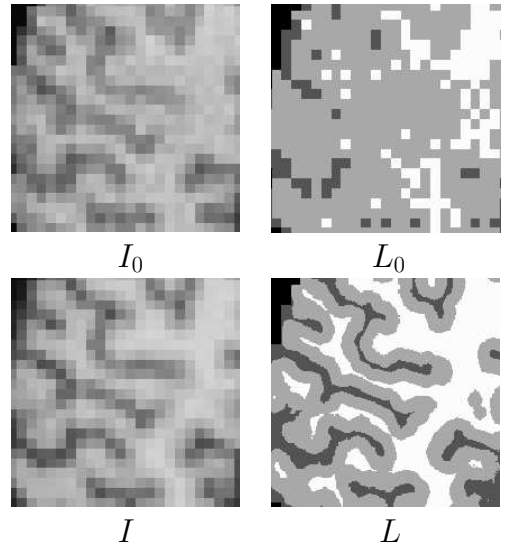


Fig. 4. Segmentation results for a low resolution ($2mm^3$) image. The source image I_0 and the predicted image I are zoomed without interpolation to emphasize their resolution. The initial segmentation L_0 was obtained through rough thresholding, giving a very bad initialization. The segmentation algorithm converges to label map L , demonstrating the robustness of the approach.

have correctly converged after only 20 iterations, the remaining features are ones where the initial segmentation was very far from the correct solution.

At the time of submission of this paper, a systematic evaluation of the performance of our segmentation approach is under way, using a data-base of anatomical structures manually labeled by experts. To this end we generate low resolution MRIs from manually labeled high resolution MR images and compare the automatic segmentation of the low resolution images with the manual segmentation provided by the experts. Receiver Operating Characteristics (ROC) analysis [5] may then be used to assess the accuracy of the segmentation procedure.

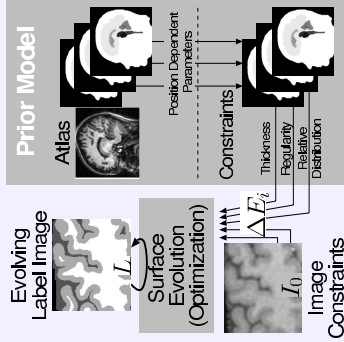
7. REFERENCES

- [1] D. Pham and J. Prince, "Adaptive fuzzy segmentation of Magnetic Resonance Images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 737–752, 1999.
- [2] X. Zeng, L. H. Staib, R. T. Schultz, and J. S. Duncan, "Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 100–111, October 1999.
- [3] J.A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, 1999.
- [4] O. Musse, F. Heitz, and J.-P. Armspach, "Topology preserving deformable image matching using constrained hierarchical parametric models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1081–1093, 2001.
- [5] G. Bueno, O. Musse, F. Heitz, and J.-P. Armspach, "3D segmentations of anatomical structures in MR images on large data bases," *Magnetic Resonance Imaging*, vol. 19/1, pp. 73–88, 2001.

Marcel Bosc,^{1,2} Fabrice Heitz,¹ Jean-Paul Armspach²

(1) LSIIT UMR-7005 CNRS / Strasbourg I University, 67400 Illkirch, France
 (2) IPB UMR-7004 CNRS / Strasbourg I University, 67085 Strasbourg, France

General Framework



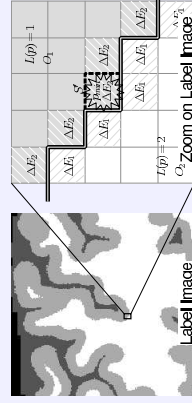
Design Goal:

Guide segmentation using complex prior knowledge

Segmentation method:

- High-resolution label image evolves at label boundaries.
- Evolution minimizes energy (cost) function.
- Energy function is a sum of terms called constraints.

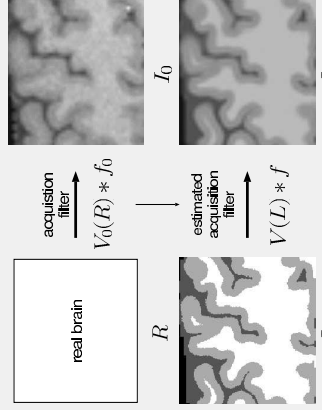
At each iteration, points where labels will change are chosen so as to guarantee energy decrease. Experiments prove that segmentation converges regardless of initialization.



Constraints are defined using atlas-mapped parameters

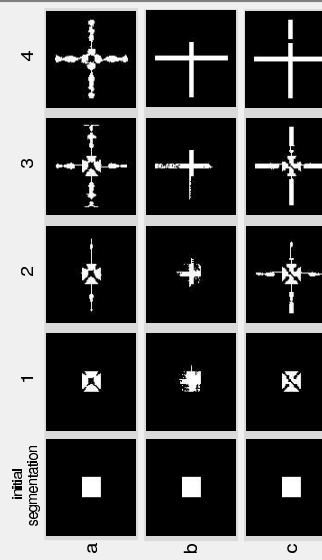
The atlas, composed of a reference image and parameter values, is mapped onto the source image using a multi-resolution deformable image matching method. Position-dependent parameters (such as thickness of a label) may be learned through statistical methods.

Variable Scale Image Constraint



Accurate model for acquisition process

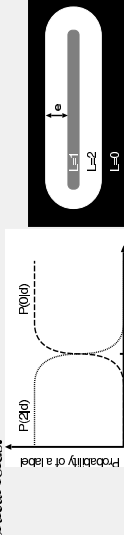
Model for so-called partial-volume effects. Acquisition model makes it possible to compare a high resolution segmentation L with a low-resolution source image I_0 . A predicted image I is computed from label image L .



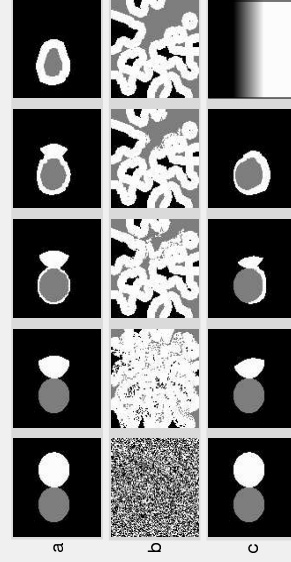
Three 2D evolutions illustrating the multi-resolution-like properties of variable scale image constraints. (a) Large scale constraints allow for global, though inaccurate convergence. (b) Small scale constraints allow for accurate convergence on a local scale, that may miss large features. (c) Combining both gives desired results (c-4).

Relative Distribution Constraint

Describes the probability of observing a label at a given distance from another label. This is used to model the distribution of gray matter and white-matter around CSF in cortical folds.



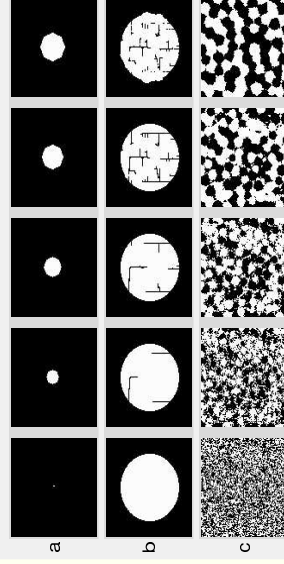
Example of probability distributions that are used to model cortical folds.



Three (a,b,c) 2D evolutions (0-1-2-3-4) illustrating the relative distribution constraint, using the probability distributions described above. In the first two evolutions (a,b) the ϵ parameter is constant. Evolution (c:1-2-3) demonstrates an atlas mapped position parameter ϵ , whose values are shown at (c-4): white label is thicker towards the bottom of the image.

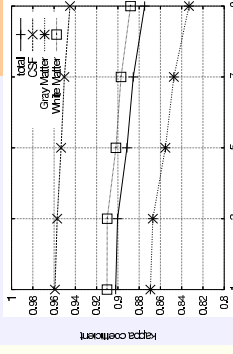
Thickness Constraint

Describes the probability of observing an object (label) of a given thickness. This is used to limit the thickness of CSF in different parts of the brain. Thickness is defined for each point on an object's surface.

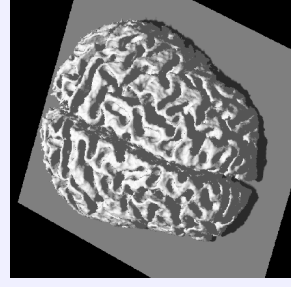
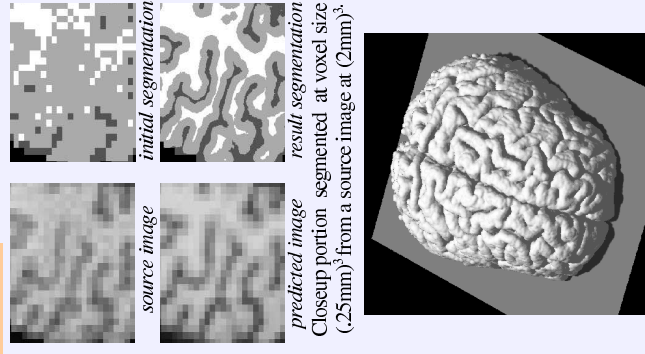


Three (a,b,c) 2D evolutions (0-1-2-3-4) illustrating the thickness constraint. The first evolution (a) increases thickness form a single point, demonstrating the consistency of thickness definition. Evolution (b) decreases the thickness of a disk. This is most effectively done by splitting the object. Evolution (c) demonstrates two competing thickness constraints (one on the white label, the other on the black). Initial random image effectively evolves into black and white regions of desired thickness.

Results



Evaluation results on BrainWeb images using the Kappa chance-corrected agreement coefficient.



3D rendering of white matter (left) and gray matter (right) segmented at voxel size $(2.5\text{mm})^3$ from a source image with voxel size $(1\text{mm})^3$.