Anatomically Consistent Segmentation of the Human Cortex for Functional MRI Visualization

Patrick C. Teo
Dept. of Computer Science
Stanford University
Stanford, CA 94305
teo@white.stanford.edu

Guillermo Sapiro
Imaging Technology Dept.
Hewlett-Packard Labs
Palo Alto, CA 94304
guille@hplgui.hpl.hp.com

Brian A. Wandell
Dept. of Psychology
Stanford University
Stanford, CA 94305
brian@white.stanford.edu

Abstract

We describe a system that is being used to segment gray matter from volumetric representations of the human cortex obtained using magnetic resonance imaging. The segmentation algorithm identifies gray matter voxels and computes their connectivity. The method differs from existing schemes in that it exploits knowledge of the anatomy of human cortex and produces anatomically consistent segmentations. The method is based on a novel and computationally efficient technique of incorporating structural constraints into the segmentation algorithm. Because the gray matter segmentation is anatomically consistent, it can be used together with functional magnetic resonance imaging measurements to visualize the spatial pattern of cortical activity within the gray matter.

Keywords: Segmentation, human cortex, structural MRI, functional MRI, visualization.

1 Introduction

Magnetic resonance scanners can be used to measure various aspects of a source material. In one important application magnetic resonance imaging (MRI) is used as a non-invasive method of visualizing biological structures (sMRI). Recently, functional magnetic resonance imaging (fMRI) has provided a method of visualizing a correlate of neural activity in the brain.[1] The ability to measure cortical activity non-invasively is an important breakthrough, providing us with a new opportunity to study the activity of single human brains at relatively high spatial resolution.

Among the various parts of the brain, the cortex is the most prominent, and one of the most intensely studied. The cortex is divided into two hemispheres connected by a massive set of nerve fibers. The cortex is composed largely of two types of tissue: gray matter and white matter. Gray matter forms the outer layer of the cortex, encasing the inner white matter almost completely. In humans, gray matter is, on average, 3 mm thick. Gray matter tissue contains a high density of neurons (approximately $10^5/\text{mm}^3$ [31]) which form the computational units of the cortex. White matter is made up of nerve fibers that connect different parts of the cortex, as well as the cortex with other parts of the brain. Functional magnetic resonance imaging measures the neural activity in the gray matter.

Despite its complex outward appearance, the structure of each hemisphere is quite straightforward and consistent across human brains. The cortical gray matter surface is the same as that of a crumpled sheet; i.e., it does not have any holes or self-intersections. To visualize the cortex as a three-dimensional structure from sMRI data, most algorithms require some prior segmentation. This segmentation may be simply discriminating between white matter, gray matter, and cerebral spinal fluid (CSF, the fluid that fills the cranial cavity), or the segmentation may involve a more elaborate labeling of different anatomical structures. Based on this segmentation, the brain is rendered as a three dimensional surface and the observer sees mainly those portions outside the folds.

Because it is important to see neural activity buried deep within the three-dimensional folds of the brain, visualizing fMRI data requires novel visualization techniques. An increasingly popular way of visualizing such mappings is to superimpose fMRI measurements on flattened representations of the cortical surface [8, 9, 10, 32]. One method of obtaining a

[1] This correlate is the relative amount of oxygen in the surrounding blood-flow. Because (a) the relative amounts of oxygen around active areas of the cortex are different from those around inactive areas of the cortex, and (b) the paramagnetic properties of oxygenated and deoxygenated blood differ, MRI can be used as an indirect measure of neural activity.

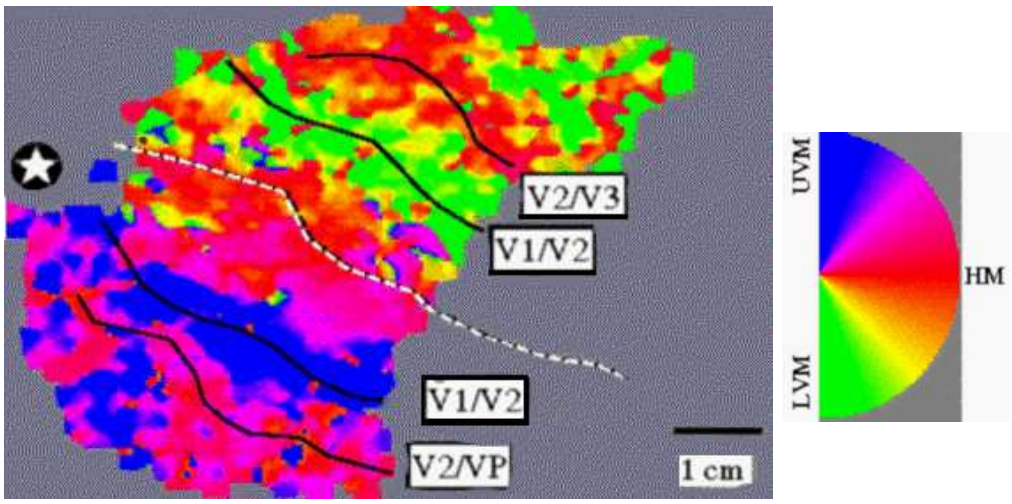


Figure 1: Functional MRI measurements in response to a visual stimulus overlaid on a flattened representation of the occipital lobe (source: Engel et al., 1997). The data represent the retinotopic, with respect to polar angle, of several visual areas in the occipital lobe. The color codes the stimulus angle at the moment each location within cortex was active. The angles are indicated by the key on the upper right: UVM is upper vertical meridian, HM is horizontal meridian, and LVM is lower vertical meridian. The left hemisphere represents only the right visual field. Reversals in the change of the polar angle representation identify the boundaries between several visual areas (V1,V2,V3,VP).

flattened representation is to compute the best planar representation of a region of gray matter such that distances on the plane are similar to the corresponding actual (geodesic) distances on the gray matter surface.

Figure 1 is an example of how fMRI measurements can be represented on a flattened region of the occipital lobe. Data from monkey and human studies show that neurons within area V1 are *retinotopically organized*: Neurons that are responsive to nearby regions of the visual field are located close to one another within the gray matter layer. Because of the retinotopic organization of visual areas, it is possible to create simple visual stimuli that generate continuous traveling waves of neural activity in visual cortex [31]. The figure illustrates how the spatial structure of these traveling waves, represented on the flattened cortical surface, can be used to determine the locations and boundaries of different areas of specialized processing.

In order to create flat maps, it is important not only to be able to identify regions in the MR data that correspond to gray matter but also to determine their connectivity. The most time-consuming aspect of creating flattened representations of visual cortex is to identify the layer of gray matter from the structural MRI of the brain. This task is often done manually, though as we discuss later, several algorithms have been proposed.

In this paper, we describe a semi-automatic system that is being used to segment gray

matter voxels in structural MR data. In addition to identifying gray matter voxels, the segmentation also computes the connectivity of gray matter voxels within the gray matter layer. Furthermore, the segmentation ensures that the topology of the connected gray matter voxels is anatomically correct; i.e., it cannot have holes or self-intersections. This is of particular importance if the gray matter segmentation will be used, together with fMR measurements, to visualize the spatial pattern of cortical activity within the gray matter layer. The presence of holes or self-intersections, caused by incorrect connections or erroneous segmentations, could have disastrous effects on these spatial patterns.

Gray matter segmentation is difficult for a variety of reasons. First, voxel intensities of gray matter tissue are spread over a large range. Second, given the current spatial resolution of MRI, regions of gray matter voxels can be as narrow as one or two voxels. This is thin compared to regions of white matter voxels which make up the bulk of the MR volume, so that a large percentage of gray matter voxels suffer from severe *partial volume* effects.[2] Partial volume reduces the effectiveness of intensity-based gray matter segmentation algorithms. Third, it is impossible to determine the connectivity of the gray matter voxels by only examining the segmented gray matter voxels. For example, the gray matter voxels on opposite sides of a sulcus (an infolding in the gray matter) could potentially be very close or even adjacent to one another on the sampling grid. Yet, the gray matter voxels on opposite sides of the sulcus should not be labeled as connected to each other. Because connectivity cannot be discerned from the intensity-based segmentation alone, ensuring that the topology of the segmented gray matter is anatomically correct is also impossible.

Various gray-matter segmentation techniques have been proposed and many of them use generic image segmentation techniques that do not fully take advantage of knowledge of the anatomy of the cortex [33, 29, 34, 30, 6, 3]. In our application, however, the gray matter segmentation is ultimately used to reveal spatial patterns of cortical activity within the gray matter layer so that the gray matter segmentation and its connectivity information needs to be anatomically consistent. There are several techniques that do benefit from some amount of anatomical knowledge, but the manner in which such knowledge is employed tends to be local and statistical [16, 28]. A notable exception is the method proposed by Joliot and Mazoyer [17] which will be discussed later in the paper.

Unlike generic image segmentation techniques, existing methods using deformable membranes (so called snakes or balloons) often do produce anatomically consistent segmenta-

[2] Partial volume effects occur when a voxel contains more than one tissue type. For example, the intensity of a voxel straddling the gray/white matter boundary or gray matter/CSF boundary would have a mean intensity value different from a voxel containing gray matter exclusively.

tions [5, 8, 18, 21, 25, 27]. If these membranes are initialized to be topologically equivalent to a sheet, they are automatically consistent with the topology of the gray matter layer. Unfortunately, the deformation of these membranes are quite prone to being captured by local minima. These minima occur when the membrane is required to deform into sulci that are deep but have narrow entrances. The region of human cortex near the occipital lobe, for example, is highly convoluted and has many such sulci. As a result, good initialization of such techniques is often required. On the other hand, these methods have the advantage that they incorporate smoothness as part of their segmentation criterion and are capable of producing sub-pixel classification (of the boundary between white and gray matter, for example). The segmentation method proposed in this paper could, quite straightforwardly, be used to initialize some of these algorithms.

2 Method

Our method consists of four stages. First, the white matter and CSF regions in the MR volume are segmented. Second, the desired cortical white matter component is selected by the user. Third, the consistency of the white matter topology with the actual anatomy is verified. Finally, gray matter is segmented by growing out from the white matter boundary while connectivity of the segmented gray matter is computed simultaneously.

2.1 Segmentation of White Matter and CSF

In the first stage, voxels containing white matter tissue and cerebral spinal fluid (CSF) are segmented. White matter is segmented before gray matter because (a) the variability of white matter voxels is lesser than the variability of gray matter voxels, and (b) white matter regions are also much larger in size than gray matter regions. Although only white matter and CSF are being classified, three classes are considered. The third class roughly corresponds to gray matter but, as shall be seen later, the segmentation of this class is poor, and will consequently be rejected in favor of the results of a secondary gray matter segmentation algorithm. The intensity of a voxel belonging to each class is modeled as an independent random variable with a normal distribution. Thus, the likelihood of a particular voxel belonging to a certain class is:

$$\Pr(V_i = v | C_i = c) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{1}{2}\frac{(v-\mu_c)^2}{\sigma_c^2}\right) \quad (1)$$

where i is a spatial index ranging over all voxels in the MR volume and c is one of {white, gray, CSF}. V_i and C_i correspond to the intensity and classification of voxel i respectively. Currently, the user interactively adjusts the parameters μ_c and σ_c until a satisfactory segmentation is produced. Typically, these parameters remain the same over all MR measurements that were obtained by a given scanning protocol.

Next, the posterior probabilities of each voxel belonging to each class is computed using Bayes' Rule and anisotropic smoothing. The posterior probability is computed for each voxel independently using Bayes' Rule together with a homogeneous prior:

$$\Pr(C_i = c|V_i = v) = \frac{1}{K} \Pr(V_i = v|C_i = c) \Pr(C_i = c) \tag{2}$$

where K is a normalizing constant independent of c. Adopting a homogeneous prior implies that $\Pr(C_i = c)$ is the same over all spatial indices i. The prior probability typically reflects the relative frequency of each class. For example, since white matter voxels occur more frequently than gray matter voxels, the prior probability of white matter is larger than that of gray matter.

In the second step, the posterior volumes are smoothed anisotropically in three dimensions while preserving discontinuities. Figure 2 shows an example of a posterior derived from a homogeneous prior along with its smoothed counterpart. The anisotropic smoothing technique applied is a 3D extension of the original 2D version proposed by Perona and Malik [22]. This step involves simulating a discretization of the following partial differential equation for a small number of iterations:

$$\frac{\partial P_c}{\partial t} = \mathrm{div}(g(||\nabla P_c||)\nabla P_c) \tag{3}$$

where $P_c = \Pr(C = c|V)$ represents the volume of posterior probabilities for class c. $g(||\nabla P_c||) = \exp(-(||\nabla P_c||/\eta_c)^2)$ and η_c represents the rate of diffusion for class c. The function $g(\cdot)$ controls the local amount of diffusion such that diffusion across discontinuities in the volume is suppressed. These parameters typically remain unchanged across different MR data sets. The intuition for applying anisotropic smoothing on the posterior probabilities is deferred to the discussion section.

Finally, segmentation is carried out by labeling each voxel with the class that yields the maximum posterior probability estimate. That is,

$$C_i^* = \underset{c \in \{\text{white, gray, CSF}\}}{\arg\max} \Pr^*(C_i = c|V_i = v) \tag{4}$$

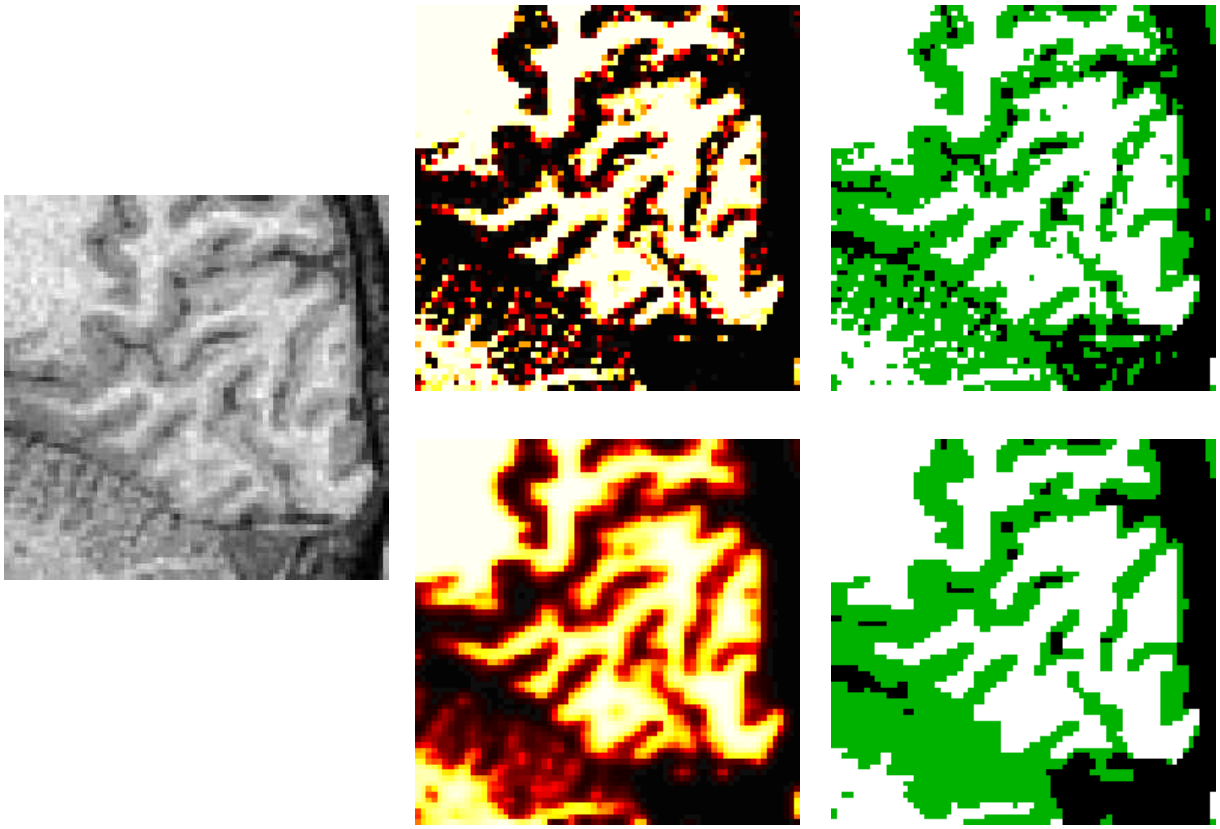


Figure 2: (Top row) Left: Intensity image of MR data. Middle: Image of posterior probabilities corresponding to white matter class. Right: Image of corresponding MAP classification. Brighter regions in the posterior image correspond to areas with higher probability. White regions in the classification image correspond to areas classified as white matter; black regions correspond to areas classified as CSF. (Bottom row) Left: Image of white matter posterior probabilities after being anisotropically smoothed. Right: Image of MAP classification computed with smoothed posteriors.

where $\Pr^*(C_i = c|V_i = v)$ corresponds to the anisotropically smoothed posterior. Figure 2 shows the MAP segmentation results obtained first using a posterior with a homogeneous prior, and subsequently with its smoothed counterpart.[3] As can be seen from the figure, the segmentation results for white matter and CSF are quite good after anisotropic smoothing is applied to the posterior volumes; the gray matter segmentation, on the other hand, remains fairly poor. As a result, only the white matter and CSF segmentations are retained.

[3]Since we were using a head coil and working on small areas, we didn't find any need to use the technique described in [33], which can be easily incorporated in our system if necessary.

2.2 Selection of Cortical White Matter

As gray matter surrounds white matter almost completely, accurate white matter segmentation can be used to obtain accurate gray matter segmentation. Thus, attention is next focused on the white matter segmentation that was computed in the previous stage. Since the white matter segmentation is represented on a grid, its digital topology needs to be defined. White matter connectivity is defined using 26-neighbor adjacency; that is, two distinct white matter voxels are adjacent to each other if their spatial coordinates differ by no more than one. Two white matter voxels are connected to each other if there is a path of white matter voxels connecting the two such that all neighboring pairs of white matter voxels along the path are 26-neighbor adjacent. Gray matter connectivity is also defined using 26-neighbor adjacency. CSF connectivity, on the other hand, is defined using 6-neighbor adjacency; that is, two distinct voxels classified as CSF are adjacent to each other if exactly one of their spatial coordinates differ by one. The reason for defining the connectivity of CSF differently is to prevent intersections between regions of CSF and white matter (or gray matter) [19].

At this stage, the user selects a voxel in the cortical white matter component, via a graphical user interface, and a flood-filling algorithm employing 26-neighbor adjacency is used to extract the corresponding white matter connected component [13]. The flood-filling algorithm begins by marking the user's selection and then proceeds iteratively, marking all unmarked voxels adjacent to existing marked voxels until there are no more unmarked voxels adjacent to marked ones. The purpose of this stage is, primarily, to remove extra-cortical components like the cerebellum. If the MR volume has been cropped to an appropriate region of interest within the cortex, the cortical white matter component typically corresponds to the largest white matter component.

2.3 Verification of White Matter Topology

After the cortical white matter component has been selected, its consistency with cortical anatomy is verified. Since the gray matter is a sheet with no holes or self-intersections, and it encases white matter almost completely, it is sufficient to ensure that the selected white matter component is free of cavities and handles. Cavities are non-white matter regions in the classification that are completely surrounded by white matter while handles are also holes but are not completely surrounded by white matter. An everyday example of a cavity is the inside of a tennis ball; likewise, an example of a handle is the hole of a doughnut. Figure 3 shows an example of a white matter cavity and a white matter handle. Handles

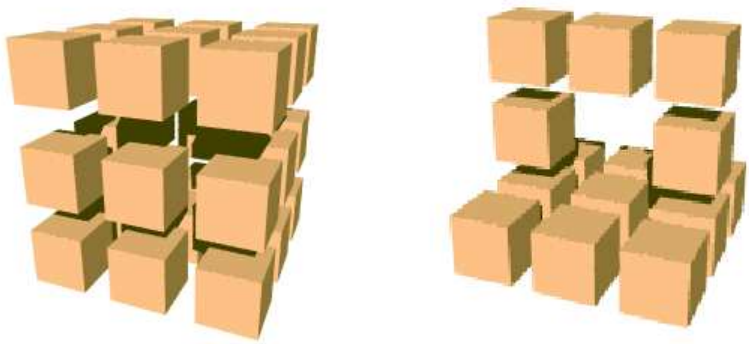
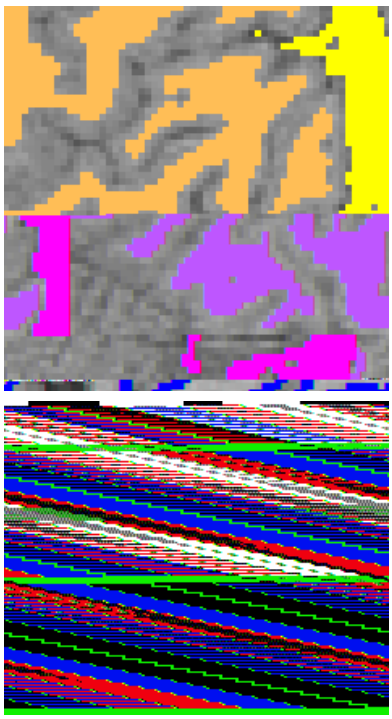


Figure 3: Left: example of a white matter cavity. Note that the middle cube is missing. Right: example of a white matter handle.

and cavities are detected at this stage and brought to the attention of the user who can then edit the segmentation results to remove them.

Cavities are first identified using a flood-filling algorithm. The algorithm identifies non-white matter connected components that are connected to the boundary of the bounding volume containing the cortical white matter component. This is accomplished by repeatedly initiating the flood-filling algorithm from non-white matter voxels on the boundary. After that, all non-white matter connected components that have *not* yet been filled must be encased entirely by white matter; i.e., they must be cavities. The user can then choose to fill these cavities or to edit the segmentation so as to remove the cavities. In practice, these cavities are typically small, and are usually filled.

Identifying handles is more complicated; in the current implementation, the number of handles is computed instead. The number of handles is computed from the Euler characteristic χ which is equal to the sum of the number of connected components and cavities, minus the number of handles. Since the number of connected components is one (as a result of second stage), and the number of cavities is zero (as a result of the previous step), the number of handles is simply one minus the Euler characteristic. The Euler characteristic is, in turn, computed as the sum of the local Euler characteristic over all $2 \times 2 \times 2$ voxel neighborhoods:

$$\chi^{\text{local}} = \sum_i \frac{v_i}{8} - \frac{e_i}{4} + \frac{f_i}{2} - o_i \tag{5}$$

where i ranges over all $2 \times 2 \times 2$ voxel neighborhoods and v_i, e_i, f_i, o_i represent the number of vertices, edges, faces and octants in the i-th neighborhood respectively [20]. Since there are only 256 possible $2 \times 2 \times 2$ neighborhood configurations, the local Euler characteristic of each possible configuration is precomputed and stored in a table. The Euler characteristic,