Cortical Reconstruction Using Implicit Surface Evolution: A Landmark Validation Study

Duygu Tosun[1], Maryam E. Rettmann[2], Daniel Q. Naiman[3],
Susan M. Resnick[2], Michael A. Kraut[4], and Jerry L. Prince[1]

[1] Electrical and Computer Engineering Dept., Johns Hopkins University, MD, USA
[2] National Institute on Aging, National Institutes of Health, MD, USA
[3] Applied Mathematics and Statistics Dept., Johns Hopkins University, MD, USA
[4] Department of Radiology, Johns Hopkins Hospital, MD, USA

Abstract. A validation study was conducted to assess the accuracy of an algorithm developed for automatic reconstruction of the cerebral cortex from T1-weighted magnetic resonance (MR) brain images. Manually selected landmarks on different sulcal regions throughout the cortex were used to analyze the accuracy of three reconstructed nested surfaces – the inner, central, and pial surfaces. We conclude that the algorithm can find these surfaces with subvoxel accuracy, typically with an accuracy of one third of a voxel, although this varies by brain region and cortical geometry. Parameters were adjusted on the basis of this analysis in order to improve the algorithm's overall performance.[1]

1 Introduction

Many brain mapping procedures require automated methods to find and mathematically represent the cerebral cortex in volumetric MR images. Such reconstructions are used for characterization and analysis of the two-dimensional (2-D) geometry of the cortex – e.g., computation of curvatures, geodesic distance, segmenting sulci or gyri, surface flattening, and spherical mapping.

The cerebral cortex is a thin, folded sheet of gray matter (GM), bounded by the cerebrospinal fluid (CSF) on the outside, and by the white matter (WM) on the inside, as illustrated in Fig. 3. The boundary between GM and WM forms the inner surface, and the boundary between GM and CSF forms the pial surface. It is useful to define the central surface as well; it lies at the geometric center between the inner and pial surfaces, representing an overall 2-D approximation to the three-dimensional (3-D) cortical sheet. A 3-D reconstruction method, called Cortical Reconstruction Using Implicit Surface Evolution (CRUISE), has been developed for automatic reconstruction of these three nested cortical surfaces from T1-weighted SPGR volumetric axially acquired MR images.

The goal of the landmark validation study presented in this paper was to evaluate the performance of CRUISE, yielding quantitative measures of accuracy and suggesting optimal parameters. Sect. 2 briefly explains the cascaded

---
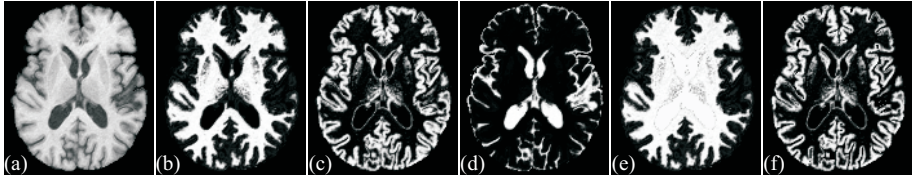
Fig. 1. Sample axial slices from (a) the skull-stripped MR image volume; (b) μ_{WM}; (c) μ_{GM}; (d) μ_{CSF}; (e) $\hat{\mu}_{\mathrm{WM}}$; (f) $\hat{\mu}_{\mathrm{GM}}$.

algorithms of CRUISE and the relation between its parameters and the location of the surfaces. Next, the paper describes the data and the validation study in Sect. 3. Based on the analysis reported in Sect. 3, a way to select optimal parameters is discussed in Sect. 4 and an analysis to validate the new parameters is presented. In Sect. 5, we summarize with some concluding remarks.

2   CRUISE: Cortical Reconstruction Using Implicit Surface Evolution

CRUISE is a data-driven method combining a robust fuzzy segmentation method, an efficient topology correction algorithm, and a geometric deformable surface model. Overall, the general approach we use in finding cortical surfaces from MR image data is described in [1], and several improvements have been incorporated, as described in [2,3,4]. The algorithm has been targeted toward and evaluated on the MR images acquired by the Baltimore Longitudinal Study of Aging [5] with the following parameters: TE = 5, TR = 35, FOV = 24, flip angle = $45°$, slice thickness = 1.5, gap = 0, matrix = 256×256, NEX = 1.

The first processing step is to re-slice the image volume to axial cross-sections parallel to the line through the manually identified anterior and posterior commissures, followed by removing the cerebellum, extracranial tissue, and brain stem from the image using a semi-automatic algorithm. The remaining image volume is then re-sampled to obtain isotropic voxels each having size 0.9375 mm×0.9375 mm×0.9375 mm using cubic B-spline interpolation in order to make subsequent processing less sensitive to orientation.

The next step in processing this "skull-stripped" MR image volume is to apply a fuzzy segmentation algorithm [6], yielding three membership function image volumes representing the fractions of WM, GM, and CSF within each image voxel – i.e., μ_{WM}, μ_{GM}, and μ_{CSF}, while compensating for intensity inhomogeneity artifacts inherent in MR images, and smoothing noise. A sample axial slice from the skull-stripped MR image volume and its tissue segmentation results are shown in Figs. 1(a)-(d).

Fig. 2 illustrates the one-dimensional profiles of the membership functions. This profile starts in the WM, passes through the GM of the cortex, and ends in the CSF surrounding the cortex. An isosurface of μ_{WM} at an isolevel $\alpha = 0.5$

provides a good approximation to the GM/WM interface. It is apparent from Fig. 1(b), however, that such an isosurface will include non-cortical surfaces such as the subcortical interfaces near the brainstem and within the ventricles. To prevent undesirable parts of the isosurface from being generated, an automatic method [2] called *AutoFill* is used to modify μ_{WM} in order to fill these concavities with WM. The largest $\alpha = 0.5$ isosurface of the edited WM membership, $\hat{\mu}_{WM}$ (Fig. 1(e)), is a close approximation to the GM/WM interface within each hemisphere, connected across the corpus callosum at the top and through the brainstem at the bottom. A graph-based topology correction algorithm [3] followed by a topology-preserving geometric deformable surface model (TGDM) [4] is used to estimate a topologically correct and slightly smoothed "inner surface" on the GM/WM interface, as shown in Fig. 5(a).

The inner surface serves as an initial surface for finding both the central and pial surfaces. These surfaces are difficult to find due to the partial volume averaging effect, which makes adjacent GM banks within narrow sulci barely dis-
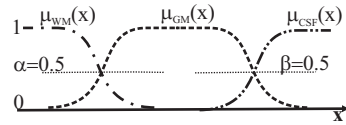

Fig. 2. One-dimensional (1-D) profiles of membership functions.

tinguishable because of the missing evidence of CSF. To compensate for this effect, CRUISE uses *anatomically consistent enhancement* (ACE) [2], which automatically modifies μ_{GM}, creating thin (artificial) CSF separations within tight sulci and yielding $\hat{\mu}_{GM}$ shown in Fig. 1(f). $\hat{\mu}_{GM}$ is used in two ways to find these surfaces. First, a gradient vector flow (GVF) external force [1] is computed directly from $\hat{\mu}_{GM}$, as if it was an edge map itself. A TGDM deformable surface is then initialized at the inner surface and is driven toward the central surface using the GVF forces, yielding a central surface as shown in Fig. 5(c). It is observed that the $\beta = 0.5$ isosurface of $\hat{\mu}_{WM} + \hat{\mu}_{GM}$ is a very good approximation to the pial surface. Accordingly, a region-based TGDM deformable surface model is used to drive the central surface toward the $\beta = 0.5$ isosurface of $\hat{\mu}_{WM} + \hat{\mu}_{GM}$, yielding an estimate of the pial surface as shown in Fig. 5(e).

When surfaces are computed using geometric deformable models such as TGDM, they contain no self-intersections. Also, an extra constraint is used to ensure the proper nesting of these three cortical surfaces with no mutual intersections. Figs. 5(b),(d), and (f) show the contours of these nested cortical surfaces superposed on a skull-stripped MR image (axial) cross-section.

3   Landmark Validation Study on Inner and Pial Surfaces

A validation study on the central surface using a set of 50 manually selected central surface landmarks – 5 on each hemisphere of 5 brains – was reported in our previous work [1]. The distance from each landmark to the central surface estimated by our algorithm served as a measure of accuracy. Overall, the mean

landmark error was 0.51 mm with a standard deviation of 0.41 mm illustrating subvoxel accuracy in our reconstruction of the central surface.

The focus of the validation study presented in this paper is to quantify the accuracy of the inner and pial surfaces estimated by CRUISE. In addition, we assess how the accuracy varies both across the cortical surface as well as within different cortical geometries illustrated in Fig. 3 – sulcal "fundi" (the bottom of a cortical fold), "banks" (the sides of the fold), and "gyri" (the top of the fold). Twelve volunteers (three neuroanatomists) participated in this validation study. Each participant identified a series of landmarks at the GM/WM and GM/CSF interfaces on the skull-stripped MR image volumes. The landmarks effectively yielded a "user implied surface" at the corresponding cortical layer. Throughout this study, we refer to these surfaces as the *implied surfaces* and the estimated surfaces by CRUISE as the *reference surfaces*. To quantify the agreement between the reference and implied surfaces, we define a "landmark offset" as the minimum distance between the given landmark and the corresponding reference surface, with the sign negative inside and positive outside (Fig. 4). These measurements will be used to quantify the accuracy of the estimated surfaces and to infer any systematic bias of CRUISE in the inward or outward directions.

The reference surfaces are defined primarily by the 0.5 isolevels of the $\hat{\mu}_{\mathrm{WM}}$ and $\hat{\mu}_{\mathrm{GM}} + \hat{\mu}_{\mathrm{WM}}$ as described in Sect. 2. Thus, the value of $\hat{\mu}_{\mathrm{WM}}$ at the GM/WM interface landmarks, and the value of $\hat{\mu}_{\mathrm{GM}} + \hat{\mu}_{\mathrm{WM}}$ at the GM/CSF interface landmarks could provide insight into a systematic bias of CRUISE. This is made clear by referring back to Fig. 2. Consider the isosurfaces of $\hat{\mu}_{\mathrm{WM}}$ and $\hat{\mu}_{\mathrm{GM}} + \hat{\mu}_{\mathrm{WM}}$ generated at higher

Fig. 3. Illustration of the three cortical geometries.

isolevel values – $\alpha > 0.5$ and $\beta > 0.5$. The isosurfaces estimated at these isolevels would give an inner and pial surface inside the surfaces estimated at $\alpha = 0.5$ and $\beta = 0.5$ indicating an outward bias of CRUISE. We are particularly interested in these measures because α and β are easily changed, to improve the performance of CRUISE.

First, we will conduct an analysis of the landmark offsets and the membership function values at the landmarks to quantify the accuracy of CRUISE as well as any systematic bias. Second, we will select the optimal α and β threshold that best fit the users' data. Finally, we will repeat the landmark validation study with the new thresholds.
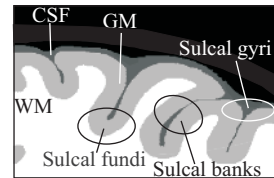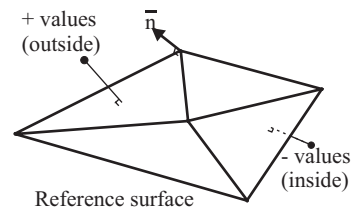

Fig. 4. Illustration of the "landmark offset".

3.1   Interactive Program for Landmark Picking

Our initial landmark validation study on the central surface [1] included only one user who picked a single landmark in each region at integer number coordinates. We have designed a new landmark validation study for the inner and pial surfaces with an extended number of landmarks and users, using two MR image volumes of different individuals. Landmarks are located on the GM/WM and GM/CSF interfaces in eleven cortical regions comprising the Sylvian, superior frontal, superior temporal, cingulate, and parieto-occipital sulci of each hemisphere, and the central sulcus of the left hemisphere. Within each region, landmarks are defined on the three cortical geometries – sulcal "fundi", "banks", and "gyri".

A visualization program was written using Open Data Explorer [7] to provide a standard way of picking the landmarks on the pre-selected axial slices of the skull-stripped MR brain volume so a statistical analysis approach could be used to compare the data of different users. The visualization program has two primary displays. The first display, shown in Fig. 6(a), provides the tissue interface and geometry information required to pick the landmarks on that axial slice. In addition, a counter is incremented after each landmark selection indicating the number of picks remaining. In the second primary display, shown in Fig. 6(b), an enlarged view around a 10×10 voxel region (outlined by the blue box in the first primary display) in which the user is required to pick the landmarks at least 0.50 mm apart from each other is shown. The interface allows the user to adjust several parameters including the center and size of the enlarged view in the second display as well as the colormap scaling – linear or logarithmic (provides more contrast at low intensities) – to improve the contrast between neighboring tissues. Each landmark is selected in this second primary display and the selected point is marked in red in all displays. In order to get a sense of the location of the point in 3-D, the two orthogonal slices through this point are also displayed, as shown in Figs. 6(c) and (d). The landmark is automatically recorded as the physical position of the selected point with floating number coordinates. The user also has the flexibility of removing any of the previously recorded landmarks.

3.2   Data

First six participants picked landmarks on two different MR image volumes and the remaining six participants picked landmarks on only one image volume. Ten landmarks were picked on each of 33 selected axial cross-sections yielding a total of 330 landmarks equally distributed across the cortical regions and geometries for each tissue interface. Two measures were computed in this validation study – the signed distance (SD), or landmark offset, and the surface defining membership function value (SDMFV) defined as the value of $\hat{\mu}_{WM}$ and $\hat{\mu}_{GM} + \hat{\mu}_{WM}$ at the landmark for the inner and pial surface, respectively.

3.3   Statistical Analysis

We first test the effects of variable intensity inhomogeneity and colormap scaling by comparing the algorithm versus user across the two different brains, the three different geometries (sulcal "fundi", "banks", and "gyri"'), and eleven different sulcal regions. Since it is hard to quantify the intensity inhomogeneity and colormap scaling, in the following statistical analysis, differences in SD and SDMFV across "brain", "geometry", and "sulci" provide information on the effects of variation in intensity inhomogeneity and colormap scaling. The GM/WM and GM/CSF interface data are analyzed separately.

Multivariate analysis of variance (MANOVA) [8] is a statistical technique for analyzing correlated measurements (i.e., SD and SDMFV) depending on several kinds of factors (i.e., brain, geometry, sulci) operating simultaneously. The measurements of the first six users were tested separately in a balanced additive MANOVA design with "sulci", "geometry", and "brain" as the factors. Another balanced additive design with "sulci" and "geometry" as the factors was used to test the measurements of the users 7–12 separately because they had completed landmarks for a single brain. MANOVA revealed a significant effect of "geometry" and "sulci" for all twelve users, but "brain" failed to reach significance (5% level) for both GM/WM and GM/CSF interface data of the users 1, 2, 5 and 6. Significant differences in performance of the algorithm relative to user across different aspects of the cortical geometry and across different sulcal regions may reflect variability in noise, intensity inhomogeneity, abnormalities in the original MR brain volume, and colormap scaling function for the different brain features. On the other hand, the absence of an effect for "brain" reflects the intra-rater consistency of picking landmarks in different MR images.

3.4   Landmark Offset on the Inner Surface

The average landmark offsets for the inner surface for different geometries are shown in Table 1. The overall mean landmark offset is -0.35 mm with a standard deviation of 0.65 mm. Only 16% of the landmarks are farther than 1.00 mm from the estimated inner cortical surfaces, indicating that gross errors are not common.

3.5   Landmark Offset on the Pial Surface

The average landmark offsets on the pial surface are also shown in Table 1. The overall mean landmark offset is -0.32 mm with a standard deviation of 0.50 mm, and only 8% of the landmarks are farther than 1.00 mm from the reference surfaces. Smaller standard deviations of the pial surface landmark offsets compared with inner surface landmark offsets indicate a higher stability for the pial surface. The higher stability on the pial surface could be due to the ACE-processing in CRUISE. In separate experiments, we have found that in ACE-processed regions, ACE is more dominant than the membership isolevel criterion in defining the surface location. Smaller landmark offset statistics (-0.16 ± 0.43 mm) are
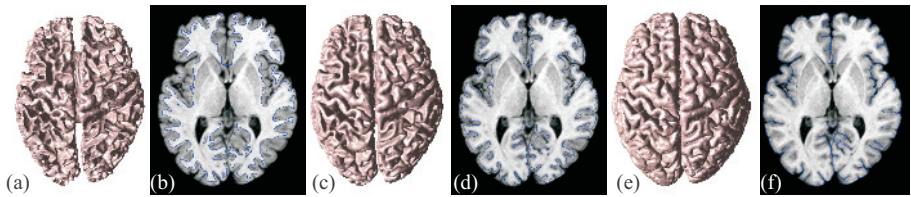
Fig. 5. Estimated surface of a sample brain and the surface superposed on an axial cross-section of the skull-stripped MR image: (a-b) inner, (c-d) central, (e-f) pial.
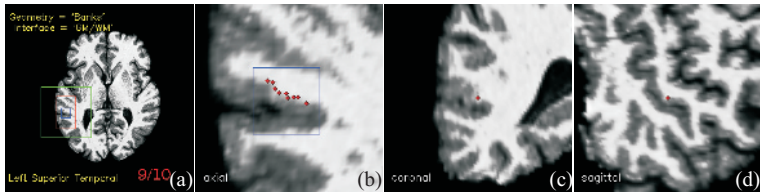

Fig. 6. Interactive program for landmark picking: (a) An example axial cross-section, (b) enlarged view, and (c)-(d) orthogonal cross-sections through the selected point.
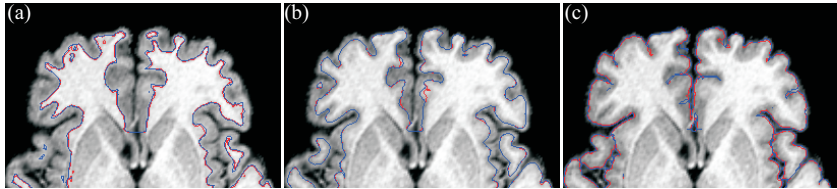

Fig. 7. Estimated surfaces superposed on an axial cross-section of the skull-stripped MR image: (a) inner, (b) central, (c) pial (blue: $\alpha = 0.5$ and $\beta = 0.5$; red: $\alpha = 0.64$ and $\beta = 0.72$).

observed in the ACE-processed regions as compared with the landmark offset statistics $(-0.47 \pm 0.51$ mm) in the regions not processed by ACE.

Consistent negative mean landmark offsets (more pronounced on the sulcal fundi regions) and the mean SDMFVs larger than 0.5 (reported in Table 1) may be interpreted as an outward bias of CRUISE. To address this observation, a simple threshold selection study is described in Sect. 4.

4   Threshold Optimization and Evaluation

In Sect. 3, we reported the landmark offsets on the inner and pial surfaces inferring an outward bias of CRUISE. Based on the observed SDMFV at the landmarks, here we want to estimate the α and β thresholds that best fit the landmark data and repeat the validation analysis with the surfaces estimated with the new thresholds. For these purposes, we divide the landmark data into two groups; the first group (training) is used for thresholds estimation, and the

Table 1. Statistics of Landmark Errors.

	Inner Surface						Pial Surface					
	SD		AD		$\hat{\mu}_{\text{WM}}$		SD		AD		$\hat{\mu}_{\text{GM}}+\hat{\mu}_{\text{WM}}$	
Geometry	mean	stdev	>1 mm	>2 mm	mean	stdev	mean	stdev	>1 mm	>2 mm	mean	stdev
Fundus	-0.60	0.60	22%	2.1%	0.73	0.21	-0.48	0.50	14%	0.4%	0.73	0.21
Bank	-0.37	0.55	11%	1.0%	0.68	0.22	-0.30	0.45	6.4 %	0.1 %	0.65	0.22
Gyrus	-0.09	0.70	16%	1.0%	0.61	0.25	-0.16	0.48	5.2 %	0.0 %	0.63	0.23
All	**-0.35**	**0.65**	**16%**	**1.4%**	**0.67**	**0.23**	**-0.32**	**0.50**	**8.4%**	**0.2%**	**0.67**	**0.23**

SD: signed distance in mm, AD: absolute distance in mm

Table 2. Membership function values at the training data landmarks

	Inner Surface		Pial Surface	
	$\hat{\mu}_{\text{WM}}$		$\hat{\mu}_{\text{GM}} + \hat{\mu}_{\text{WM}}$	
Geometry	mean	stdev	mean	stdev
Fundus	0.71	0.21	0.78	0.21
Bank	0.64	0.22	0.69	0.23
Gyrus	0.58	0.25	0.69	0.25
All	**0.64**	**0.23**	**0.72**	**0.23**

second group (test) is used to repeat the analysis. The grouping is based on the intra-user consistency on picking landmarks reported in Sect. 3.3. In particular, the data of the users 1, 2, 5, and 6 form the training data.

SDMFV statistics for the training data are reported in Table 2. Although we observed that the α and β thresholds should be functions of the cortical geometry – i.e., the ideal thresholds are different for different parts of the brain –, in this study, we choose a simpler approach and set α and β to the observed mean SDMFV, and repeated our previous analysis with these thresholds.

The cortical surfaces were estimated using the $\alpha = 0.64$ and $\beta = 0.72$ thresholds, and the landmark validation study was repeated on both inner and pial surfaces. Table 3 gives the landmark offsets of the test data on the surfaces estimated with the original and the new α and β thresholds. Overall, there is a 79% improvement on the inner surface, and a 64% improvement on the pial surface. Different percentile improvements on the different geometries support the idea of defining the thresholds as functions of the cortical geometry. A sample axial cross-section with the surfaces estimated with the original and the new α and β thresholds are shown in Fig. 7. We repeated the central surface landmark validation study reported in [1] with the new central surfaces. Slight differences with no substantial improvement or change were observed on the reported values. These results show the robustness of the central surface reconstruction.

5   Discussion and Future Work

The purpose of this work was to evaluate the accuracy of the CRUISE algorithms developed for the automatic reconstruction of the three nested surfaces of the cerebral cortex from MR brain images. This was accomplished by conducting a validation study using landmarks. The surfaces can be found with subvoxel accuracy, typically with an accuracy of one third of a voxel. Currently, we utilize "user implied surfaces", derived from user selected landmarks, to quantify the

Table 3. Landmark errors for the test data

Geometry	α	Inner Surface SD mean	stdev	AD >1 mm	>2 mm	β	Pial Surface SD mean	stdev	AD >1 mm	>2 mm
Fundus	0.50	-0.64	0.62	24%	2.7%	0.50	-0.52	0.56	18%	1.7%
Bank	0.50	-0.44	0.54	13%	1.0%	0.50	-0.41	0.56	14%	1.3%
Gyrus	0.50	-0.10	0.74	17%	1.1%	0.50	-0.24	0.57	11%	1.3%
All	**0.50**	**-0.39**	**0.68**	**18%**	**1.6%**	**0.50**	**-0.39**	**0.57**	**14%**	**1.4%**
Fundus	0.64	0.38	0.61	12%	1.6%	0.72	-0.39	0.75	18%	4.4%
Bank	0.64	-0.16	0.62	7%	1.9%	0.72	-0.14	0.59	9%	0.1%
Gyrus	0.64	0.29	0.92	20%	6.6%	0.72	0.12	0.59	10%	0.1%
All	**0.64**	**-0.08**	**0.78**	**13%**	**3.4%**	**0.72**	**-0.14**	**0.68**	**12%**	**1.8%**

SD: signed distance in mm, AD: absolute distance in mm

accuracy of CRUISE. In future work, we plan to create a nested surface truth model from the visible human cyrosection and MR image data [9], and validate our methods against this data.

A simple experiment to improve CRUISE by selecting new threshold values which were more in accordance with the user implied surfaces was presented in Sect. 4. Based on the statistics reported in this work, a more extensive study is currently underway to formulate CRUISE thresholds as functions of cortical geometry. We expect a variable threshold scheme – based on local cortical geometry – will provide even higher accuracy in the CRUISE reconstruction algorithms.

References

1. C. Xu, D. L. Pham, M. E. Rettmann, D. N. Yu, and J. L. Prince. Reconstruction of the human cerebral cortex from magnetic resonance images. *IEEE Trans. Med. Imag.*, 18(6):467–480, June 1999.
2. X. Han, C. Xu, M. E. Rettmann, and J. L. Prince. Automatic segmentation editing for cortical surface reconstruction. In *Proc. of SPIE Medical Imaging*, Feb 2001.
3. X. Han, C. Xu, U. Braga-Neto, and J. L. Prince. Topology correction in brain cortex segmentation using a multiscale, graph-based algorithm. *IEEE Trans. Med. Imag.*, 21:109–121, 2002.
4. X. Han, C. Xu, and J. L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Trans. on Pattern Anal. Machine Intell.*, 25:755–768, 2003.
5. S. M. Resnick, A. F. Goldszal, C. Davatzikos, S. Golski, M. A. Kraut, E. J. Metter, R. N. Bryan, and A. B. Zonderman. One-year age changes in MRI brain volumes in older adults. *Cerebral Cortex*, 10(5):464–72, May 2000.
6. D. L. Pham. Robust fuzzy segmentation of magnetic resonance images. In *IEEE Symposium on Computer-Based Medical Systems*, pages 127–131, 2001.
7. OpenDX Developers. http://www.opendx.org/, 1997.
8. H. Scheffe. *The analysis of variance*. Wiley, New York, 1959.
9. V. Spitzer, M. J. Ackerman, A. L. Scherzinger, and D. J. Whitlock. The visible male: A technical report. *J. Am. Med. Informat. Assoc.*, 3:118–130, 1996.