Segmenting magnetic resonance images via hierarchical mixture modelling[☆]

Carey E. Priebe[*], Michael I. Miller, J. Tilak Ratnanather

Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

We present a statistically innovative as well as scientifically and practically relevant method for automatically segmenting magnetic resonance images using hierarchical mixture models. Our method is a general tool for automated cortical analysis which promises to contribute substantially to the science of neuropsychiatry. We demonstrate that our method has advantages over competing approaches on a magnetic resonance brain imagery segmentation task.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Automated cortical analysis; Magnetic resonance imagery; Mixture model; Classification; Segmentation

1. Introduction

Gaussian mixture modelling is ubiquitous throughout laboratories in the world for image analysis and segmentation, and has played a core role in medical image analysis, and in

particular Human Brain Mapping for the past several decades. A fundamental aspect of Human Brain Mapping research today is cortical reconstruction of brain volumes into their macroscopic components, including gray matter (G), white matter (W), or cerebrospinal fluid (C). Throughout the field of brain mapping, almost without exception the construction of cortical manifolds requires some form of segmentation via probabilistic model fitting. Reconstruction of volumetric submanifolds such as the thalamus, hippocampus, deep nuclei, and other structures are being performed via automatic segmentation methodologies by many investigators (Fischl et al., 2001; Grabowski et al., 2000; Holmes et al., 1998; Kapur et al., 1996; Joshi et al., 1999; Miller et al., 2000; Ratnanather et al., 2001; Shattuck et al., 2001; Robb, 1999; Teo et al., 1997; Wells-III et al., 1996; Xu et al., 1999). Such approaches boil down to some form of classical Bayesian segmentation and Neyman–Pearson likelihood-ratio testing on Gaussian models and Gaussian mixture models. Even segmentation methods which apply segmentation implicitly through mapping without direct Bayesian hypothesis testing, such as those approaches based on deformable template matching (Christensen et al., 1997) or active surface methods (Yezzi et al., 2002; Westin et al., 2000; Tagare, 1997; Shen et al., 2001; Sclaroff and Liu, 2001; Schultz and Conradsen, 1998; Montagnat et al., 1999, 2001; Montagnat and Delingette, 1997; Mignotte and Meunier, 2001; Kervrann and Heitz, 1999; Jain et al., 1996; Garrido and de la Blanca, 1998, 2000; Chen and Metaxas, 2000; Caunce and Taylor, 2001; Xu et al., 2000; Xu and Prince, 2000; Pham et al., 2000) are usually based on some form of Gaussian mixture modelling. Usually, the data models are implicitly conditionally Gaussian, with the resulting matching functions of the quadratic-norm type.

At the heart of such brain mapping efforts is mixture modelling in which each tissue class is modelled as a single Gaussian component. Sometimes additional "partial volume" components are included to account for the inadequacy of the single Gaussian model. Of course, selecting the dimension of the model, i.e. the number of partial volume fits, reduces to the model selection problem; as well there is no reason to believe that a single model order fits all tissue and or imaging modality types. The focus of this paper is to propose a method which allows for the flexibility of a Gaussian mixture model—with model complexity selected adaptively from the data—for each tissue class. Our procedure involves modelling each class as a semiparametric mixture of Gaussians. Thus, for each subject j, the subject-specific marginal f_j is taken to be a hierarchical mixture model—a mixture of Gaussian mixtures. For each of the three classes $c \in \{C, G, W\}$, the density f_j has one mixture component f_{jc}, and each of these components is itself a mixture of Gaussians. The major difficulty solved in this paper associated with employing such semi-parametric methods is to solve dynamically the model selection problem. The crucial step of determining class-conditional mixture complexities for (unlabeled) test data is hopeless in the unsupervised case—without labeled voxel observations known to belong to the three classes, the hierarchical mixture model is hopelessly unidentifiable. We accomplish this by matching models to a predefined data base of hand labelled experimental tissue samples. Thus, the use of a registry of expertly generated data becomes a fundamental part of our solution. This approach of course builds on the entire philosophy now emerging through the Biomedical Informatics Research Network (http://www.nbirn.net) initiatives in which data bases are federated and available for laboratory comparison and analysis. As we show, the performance improvement obtained is significant, implying that the emergence of hand

labelled information from laboratories across the country which are combined with automated methods for analysis promise tremendous improvements in future automated brain analysis methods.

Of course, the importance of such improvements is in that a fundamental aspect of human brain mapping today is the subtle examination of small changes in brain morphometry. For instance, to investigate whether schizophrenia or Alzheimer's disease alters the structure of gray matter in the cingulate gyrus of affected subjects, accurate segmentations into tissue types of the cingulate gyrus for both affected and control subjects are required. Such segmentations are currently obtained via labor- and cost-intensive hand-segmentation of all images in a sample. For the purpose of investigating subtle differences between the affected and control populations, larger sample sizes afford greater statistical power and are, of course, desirable. This leads to a conflict between the desire for a large sample size and the desire to keep the number of hand-segmentations small. This manuscript presents a method for taking a hand-segmented *training* subset of the images to be segmented and producing accurate automated segmentations for the remaining images in the sample. This will allow for a larger sample of segmented images, thereby increasing the power of subsequent statistical analyses.

Our method proceeds according to the following outline.

(1) For each subject/class pair in the available training data set, estimate the marginal subject-specific class-conditional probability densities. Notice that for this supervised step (since these are training images, individual voxel class labels are available) semi-parametric mixture complexity estimation is appropriate.
(2) Find the "closest" training model to the (unlabeled) test data.
(3) Fit a mixture to the test data, using the training model obtained in step (2) to determine the class-conditional mixture complexities and starting locations.
(4) Classify voxels from the test data using the plug-in Bayes rule, where the mixture component class labels are inherited from the selected training model (but the mixture itself is estimated from the test data in step (3)).

This manuscript is structured as follows. In Section 2 we describe a data set of MR images to be investigated. Section 3 gives the mixture modelling methodology in detail. Section 4 describes the experimental protocol employed, and Section 5 reports the results of our investigations. Finally, in Section 6, we conclude with a discussion of the implications of our results and our methodology.

2. Data

Dysfunction of specific subregions of the cerebral cortex has been implicated in the pathophysiology of several neuropsychiatric disorders, including schizophrenia. Although widespread thinning of the gray matter cortical mantle has been observed in subjects with schizophrenia (Zipursky et al., 1992), specific abnormalities of neuronal architecture may exist within portions of the cortical mantle and form the neuroanatomical basis for specific cognitive deficits or symptom groupings. Shown in Fig. 1 are examples of cingulate, occipital

Fig. 1. Panel 1 shows the cingulate gyrus, panel 2 is the occipital gyrus, and panel 3 is the medial prefrontal gyrus. (Adapted from Duvernoy, 1999).

and medial prefrontal in the human brain, which are studied here. Such brain areas consist of different mixtures of gray matter, white matter, and cerebrospinal fluid.

The cingulate gyrus, a large C-shaped gyrus that courses around the corpus callosum and cerebral ventricles, has attracted considerable interest because of its role in attention and other cognitive functions that appear to be critical for understanding schizophrenia and other neuropsychiatric disorders (Csernansky and Bardgett, 1998). The medial prefrontal cortex has been implicated in a variety of recent functional and structural neuroimaging studies of both normal affect regulation and differences associated with affective disorders such as major depressive disorder or bipolar affective disorder (Swerdlow and Koob, 1987; Drevets et al., 1992; Drevets and Todd, 1997; Botteron and Figiel, 1997; Drevets, 1999). The occipital cortex has been the focus of several major functional neuroimaging studies of visual perception; see, e.g. (Yantis, 2001; Tootell et al., 1998).

High-resolution MR imaging now affords an unprecedented opportunity to acquire detailed images of the neuroanatomical configurations and tissue characteristics of the living human brain. Substantial advances in our understanding of neuropsychiatric disorders are anticipated as this technology is used more widely in populations of individuals with such illnesses. However, most neuropsychiatric researchers still apply manual methods to assess the boundaries and internal tissue segmentations of cortical subregions in MR scans. These manual methods are limited by their dependency on the constant availability of human experts and by between rater error (e.g., Bartzokis et al., 1993; Haller et al., 1997).

We investigate first a set of cingulate gyrus images provided by John Csernansky of the Washington University School of Medicine, St. Louis. High-resolution MR scans were acquired using a turbo-FLASH sequence (TR $= 20$ ms, TE $= 5.4$ ms, flip angle $= 30°$, ACQ-1, Matrix 256×256, Scanning Time $= 13.5$ min) with 1 mm $\times 1$ mm $\times 1$ mm isotropic resolution across the entire cranium. (See, for instance, Cho et al. (1993) for MR imaging background.) MR image data were reformatted via intensity scale adjustment to maximize tissue contrast, using the voxel intensities in the corpus callosum and the lateral ventricle as limiting values. MR scans were then interpolated into 0.5 mm $\times 0.5$ mm $\times 0.5$ mm isotropic resolution using trilinear interpolation.

We consider $J = 10$ subjects; five with schizophrenia (labeled "CG 1" through "CG 5") and five control subjects (labeled "CG 6" through "CG 10"). For each subject, the voxels in the cingulate gyrus of the right hemisphere, have been hand-segmented by expert neuroanatomists into one of three tissue classes; $\mathcal{C} := \{C, G, W\}$. (See Figs. 2 and 3.) Hand
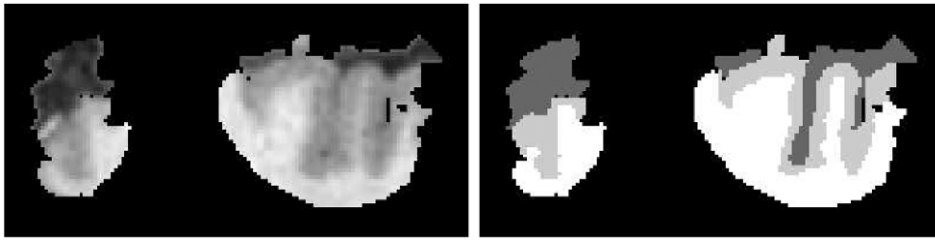
Fig. 2. Panel 1 shows the original MRI, and Panel 2 is the hand-segmentation of (one slice of) CG 9 into C (dark gray), G (light gray), W (white); black is unsegmented region.
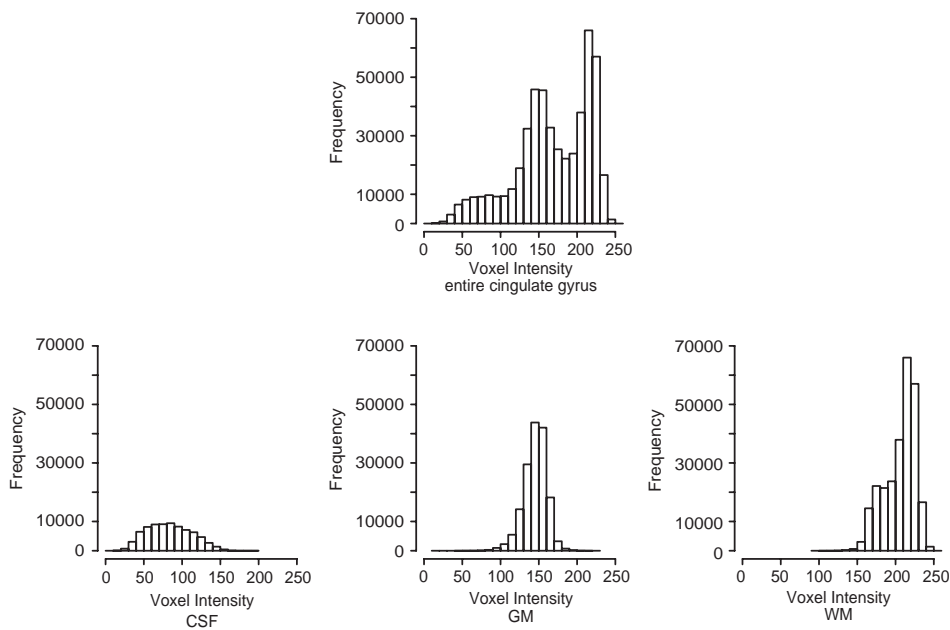

Fig. 3. Frequency histograms for CG 9: entire cingulate gyrus, right hemisphere (top), and C (bottom left), G (bottom center), and W (bottom right).

segmented files were created using Analyze (Robb, 1999) and a region of interest (ROI) defined by expert raters to isolate the cingulate gyrus. The ROI included white matter, gray matter and cerebrospinal fluid of one hemisphere that encompassed the cingulate gyrus and immediate surrounding region. A three-dimensional ROI encompassing the entire cingulate gyrus in the right hemisphere was outlined. The ROI was constructed in coronal sections. In each MR section, an enclosure which consisted of the gray matter of the right cingulate gyrus and its neighboring gray matter, white matter, and cerebrospinal fluid was drawn by hand. This was a fast procedure as it did not require manual segmentation of exact tissue types or precise anatomical structures. An example of such an ROI section is shown in

Table 1
Sample sizes for 10 cingulate gyri

	Number of voxels									
	CG 1	CG 2	CG 3	CG 4	CG 5	CG 6	CG 7	CG 8	CG 9	CG 10
C	150,128	90,035	91,452	85,063	121,138	45,709	108,617	84,402	76,886	125,409
G	210,636	186,529	136,856	124,227	130,717	154,993	181,538	169,583	161,138	131,230
W	333,341	359,746	263,401	243,651	274,943	193,633	320,293	276,141	264,607	339,410
Total	694,105	636,310	491,709	452,941	526,798	394,335	610,448	530,126	502,631	596,049

Fig. 2; note that some voxels were not labeled and thus not included in the ROI. The right cingulate gyrus was manually delineated, according to anatomical rules. For each scan, a set of cingulate manual segmentations by an expert rater was obtained in the ROI. The manual segmentations consisted of three tissue classes. Voxel classifications by the rater were based on visual inspection of morphological cues such as gyri curvature and connectivity, as well as visual intensity cues between three tissue classes.

There are between 390,000 and 700,000 class-labeled voxels per subject. The smallest subject-specific class-conditional sample size is 45,709 (C for subject CG 6). See Table 1 for details.

3. Semiparametric mixture modelling

Model selection in Gaussian mixtures—the estimation of mixture complexity—is of fundamental importance in many applications of finite mixture models. An enormous body of literature exists regarding the application, computational issues, and theoretical aspects of mixture models when the number of components is known (see, e.g., McLachlan and Peel, 2000; Everitt and Hand, 1981; McLachlan and Basford, 1988; Titterington et al., 1990), but estimating the unknown number of components remains an area of intense research effort. See, for instance, Chen and Kalbfleisch (1996); Dacunha-Castelle and Gassiat (1997, 1999); Escobar and West (1995); Henna (1985); James et al. (2001); Keribin (2000); McLachlan (1987); Priebe and Marchette (2000); Roeder and Wasserman (1997) for recent progress in this area.

We model the class-conditional probability density functions via the "alternating kernel and mixture" (AKM) method (Priebe and Marchette, 2000; James et al., 2001). This methodology results in an automatic estimate of mixture complexity. We do not claim AKM is necessarily the best approach; but it is reasonable and illustrative, and can have advantages over other techniques. In Priebe and Marchette (2000) AKM is compared to the Bayesian methodology of Roeder and Wasserman (1997), while James et al. (2001) presents a comparison with numerous competing approaches.

3.1. Modelling

For each subject $j \in \mathscr{J} := \{1, \ldots, J\}$, consider magnetic resonance voxel observations $\mathscr{X}_j := \{X_{j1}, \ldots, X_{jn_j}\}$; the subject-specific sample sizes are denoted by n_j. Let the marginal

probability density function for subject j voxel observations be given by

$$f_j = \sum_{c \in \mathscr{C}} \pi_{jc} f_{jc}. \tag{1}$$

That is, $f_j = \pi_{jC} f_{jC} + \pi_{jG} f_{jG} + \pi_{jW} f_{jW}$. The subject-specific class-conditional marginals are denoted f_{jc}, and the subject-specific class-conditional mixing coefficients π_{jc} are non-negative and sum to unity. We assume that the X_{ji} are identically distributed according to f_j (but not independent).

Each subject-specific class-conditional marginal f_{jc} is itself modelled as a mixture of normals

$$f_{jc} = \sum_{t=1}^{k_{jc}} \pi_{jct} \, \varphi_{jct}, \tag{2}$$

where the subject-specific class-specific mixing coefficients satisfy $\pi_{jct} \geqslant 0$ for each j, c, t, and $\sum_t^{k_{jc}} \pi_{jct} = 1$ for each j, c, and $\varphi_{jct} = \varphi(\cdot; \mu_{jct}, \sigma^2_{jct})$ denotes the Gaussian probability density function with mean μ_{jct} and variance σ^2_{jct}. Thus the subject-specific marginals f_j are modelled as hierarchical mixtures—mixtures of Gaussian mixtures;

$$f_j = \sum_{c \in \mathscr{C}} \pi_{jc} \sum_{t=1}^{k_{jc}} \pi_{jct} \, \varphi_{jct}. \tag{3}$$

This model is precisely analogous to the model used in "mixture discriminant analysis" (Hastie and Tibshirani, 1996). The subject-specific class-conditional mixture complexities k_{jc} are to be estimated from the data via AKM.

3.2. Estimation

Given class-labelled training data, we have available subject-specific class-conditional sample sizes n_{jc} such that $n_j = \sum_c n_{jc}$. For the subject-specific mixing coefficients π_{jc} we use the empirical estimate

$$\widehat{\pi}_{jc} = n_{jc}/n_j, \tag{4}$$

the ratio of the subject-specific class-conditional sample size to the total subject-specific sample size. Let $\mathscr{X}_{jc}$ denote that subset of $\mathscr{X}_j$ for which the class label is c.

We estimate the subject-specific class-conditional mixture complexities k_{jc}, mixing coefficients π_{jct}, and mixture components φ_{jct}, via AKM. This estimation is semiparametric; the mixture complexities $\widehat{k}_{jc}$ are estimated from the data. AKM employs an iterative estimation scheme, comparing the k-component mixture estimate against the $k+1$-component mixture estimate. When the improvement obtained by adding a $k+1$st component is negligible (less than some penalty term) the iteration is terminated and the resulting k-component mixture is used as the estimate. This general version of model selection—looking for the "elbow" or "knee" in a complexity vs. penalized estimation accuracy curve—is quite common. A distinguishing feature of AKM is the process of using successive kernel estimates to guide the successive mixture estimates.

The filtered kernel estimator (Marchette et al., 1996) extends the basic kernel estimator by allowing multiple bandwidths driven by a pilot mixture model. Given a Gaussian mixture

$$f = \sum_{t=1}^{k} \pi_t \varphi_t \tag{5}$$

and a bandwidth h, define the filtered kernel estimator $\widetilde{f}(\cdot; \mathscr{X}, f, h)$ based on the mixture f and using the data $\mathscr{X} = \{X_1, \ldots, X_n\}$ to be

$$\widetilde{f}(x; \mathscr{X}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{k} \frac{\pi_t \varphi_t(X_i)}{f(X_i) h \sigma_t} \varphi_0 \left(\frac{x - X_i}{h \sigma_t} \right), \tag{6}$$

where σ_t^2 is the variance of the tth component of the mixture f and φ_0 is the standard zero mean, unit variance normal. This allows for different bandwidths, guided by the mixture model f.

Let $\widehat{f}^1$ be the single normal component with mean and variance determined by the sample moments of $\mathscr{X}_{jc}$—that is, $\widehat{f}^1$ is a trivial one-component mixture. Let $\widetilde{f}^1$ be the filtered kernel estimate based on $\widehat{f}^1$—that is, $\widetilde{f}^1$ is the standard kernel estimate using the normal reference rule to determine the bandwidth. For $k = 2, 3, \ldots$, define in turn first $\widehat{f}^k$ to be the k-component mixture best matched to the nonparametric estimate $\widetilde{f}^{k-1}$

$$\widehat{f}^k := \arg \min_{f \in \mathscr{F}^k} ||f - \widetilde{f}^{k-1}||_2^2, \tag{7}$$

where $||f - g||_2^2 := \int_{-\infty}^{\infty} (f(x) - g(x))^2 \, \mathrm{d}x$ is the integrated squared error and $\mathscr{F}^k$ is the class of k-component Gaussian mixtures. Subsequently define $\widetilde{f}^k$ to be the filtered kernel estimate based on the mixture $\widehat{f}^k$.

Let

$$\ell(f, \mathscr{X}) := \log \prod_{x \in \mathscr{X}} f(x) = \sum_{x \in \mathscr{X}} \log f(x). \tag{8}$$

(This would be the log-likelihood if the observations were independent; alas, this is not the case for the brain voxels. We address the issue of spatial dependence in the next subsection.) The estimate of mixture complexity used here (analogous to the estimates proposed in Priebe and Marchette (2000) and James et al. (2001)) is given by

$$\widehat{k}_{jc} = \arg \min\{k \in \{1, 2, \ldots\} : \ell(\widehat{f}^{k+1}, \mathscr{X}_{jc}) - \ell(\widehat{f}^k, \mathscr{X}_{jc}) < a(n_{jc}, k+1)\}. \tag{9}$$

The penalty term $a(n_{jc}, k+1)$ in the above equation—a function of sample size and model complexity—is *the* key practical issue in this version of model selection. A simple choice for $a(n_{jc}, k+1)$ is $3 \log(n_{jc})$. Asymptotic considerations require $a(n_{jc}, k+1)/n_{jc} \to 0$ as $n_{jc} \to \infty$ for fixed k, but not too fast. In practice, the choice of $a(n_{jc}, k+1)$ drives the parsimoniousness of the resultant model. Akaike (1974), Rissanen (1978), and many others since (see, notably, George and Foster, 2000) have weighed in on the choice of this penalty

term; nevertheless, for finite sample sizes the choice quite resembles an artistic balancing of parsimony and fit—the conventional bias-variance trade-off.

3.3. Spatial dependence

Because there is correlation amongst the voxel observations, and thus $\ell(\cdot, \cdot)$ defined above is not the log-likelihood, simple interpretation of the penalty term $a(n_{jc}, k+1)$ as a function of sample size calls for conditional covariance modelling of the three-dimensional spatial process. Covariograms (Cressie, 1993) for the subject-specific class-conditional random fields indicate that significant spatial correlation is evident at a distance of up to 10 voxels. (Stationarity and isotropy are indicated as well.) This suggests that the effective sample sizes are significantly smaller than the number of voxels n_{jc}.

One way to look at the effect of dependent observations on our estimate of mixture complexity is to note that, if the observations in the subject-specific class-conditional sample $\mathscr{X}_{jc}$ are positively correlated, then the terms $\ell(\widehat{f}^{k+1}, \mathscr{X}_{jc})$ and $\ell(\widehat{f}^{k}, \mathscr{X}_{jc})$ in Eq. (9) should be scaled accordingly. That is, if the "effective number of independent observations" is $n'_{jc} := n_{jc}/\delta$ for $1 \leqslant \delta \leqslant n_{jc}$, then substituting $\delta^{-1}\ell(\cdot, \cdot)$ into Eq. (9) will account for the correlation. The case $\delta = 1$ represents independence and, as (positive) spatial correlation increases, δ increases as well. This can be presented as

$$\widehat{k}_{jc} = \arg\min\{k \in \{1, 2, \ldots\} : \ell(\widehat{f}^{k+1}, \mathscr{X}_{jc}) - \ell(\widehat{f}^{k}, \mathscr{X}_{jc}) \\ < \delta a(n_{jc}/\delta, k+1)\}. \tag{10}$$

The idea of accounting for spatial correlation via an "effective number of independent observations" (Cressie, 1993) is analogous to the "resolution elements" found in Worsley et al. (1992), e.g. Solid mathematical justification for adopting this approach is lacking except in the most restricted cases (Leadbetter et al., 1983). Direct modelling of the dependency structure, via methods such as Markov random field models, is the subject of intense ongoing research. Our ad hoc approach is meant as a simple and straight-forward method enabling semiparametric hierarchical mixture modelling.

Estimating δ is a sticky wicket. Nevertheless, as will be seen in the experimental results presented below, a significant performance improvement can be obtained by employing even a crude covariogram-based estimate $\widehat{\delta}$ in the complexity selection methodology. Furthermore, Occam's Razor may suggest a preference to err on the side of parsimony. Our rule of thumb is to choose $\widehat{\delta}$ to be the number of voxels in a ball whose radius is given by the distance at which the omnidirectional empirical covariogram drops below some threshold, rounded to the nearest integer.

Using $a(n, k) = 3 \log(n)$, we arrive at the estimate of mixture complexity

$$\widehat{k}_{jc} = \arg\min\{k \in \{1, 2, \ldots\} : \ell(\widehat{f}^{k+1}, \mathscr{X}_{jc}) - \ell(\widehat{f}^{k}, \mathscr{X}_{jc}) < \widehat{\delta}3 \log(n_{jc}/\widehat{\delta})\}. \tag{11}$$

Given $\widehat{k}_{jc}$ selected in this manner, the mixture model estimate of the marginal probability density for subject j, tissue class c, is then given by

$$\widehat{f}_{jc} = \arg\min_{f \in \mathscr{F}^{\widehat{k}_{jc}}} ||f - \widetilde{f}^{\widehat{k}_{jc}-1}||_2^2, \tag{12}$$

the $\widehat{k}_{jc}$-component mixture identified in the AKM procedure. Results will be reported below for three choices of $\widehat{\delta}$: 1, 25, and 99. The case $\widehat{\delta} = 1$ corresponds to ignoring the dependence; the choices $\widehat{\delta} = 25$ and $\widehat{\delta} = 99$ are obtained with threshold values 2/3 and 1/3, respectively.

4. Segmentation method

We consider segmenting the cingulate gyrus for subject $\gamma \in \mathcal{J}$; in this case, we have $J - 1$ training subjects corresponding to indices $j \in \mathcal{J}_\gamma := \mathcal{J} \backslash \{\gamma\}$.

4.1. Training

For $j \in \mathcal{J}_\gamma$, we obtain

$$\widehat{f}_j = \sum_{c \in \mathcal{C}} \frac{n_{jc}}{n_j} \widehat{f}_{jc} = \sum_{c \in \mathcal{C}} \frac{n_{jc}}{n_j} \sum_{t=1}^{\widehat{k}_{jc}} \widehat{\pi}_{jct} \widehat{\varphi}_{jct} \tag{13}$$

via AKM. For γ we have the test data set $\mathcal{X}_\gamma$.

Write

$$\mathcal{L}(\gamma, f) := \prod_{x \in \mathcal{X}_\gamma} f(x). \tag{14}$$

(This would be the likelihood if the subject-specific voxel observations were independent.) We proceed by choosing the index γ^* specifying the training subject "closest" to the test data set. However, even though we proceed under the assumption that the subject-specific class-conditional tissue mixture models $\widehat{f}_{jc}$ for one training subject can be usefully transferred to the test sample, there is good reason to treat the subject-specific class-conditional mixing coefficients as nuisance parameters; while the tissue class probability distributions may be transferrable, we wish to assume that the percentage of gray matter, say, may vary from training subject to test subject. Toward this end we define

$$\mathcal{F}_j := \left\{ \sum_{c \in \mathcal{C}} p_{jc} \widehat{f}_{jc} : p_{jc} \in (0, 1) \ \forall c \text{ and } \sum_{c \in \mathcal{C}} p_{jc} = 1 \right\}, \tag{15}$$

the collection of densities which agree with $\widehat{f}_j$ up to subject-specific class-conditional mixing coefficients. Then

$$\gamma^* := \arg \max_{j \in \mathcal{J}_\gamma} \max_{f \in \mathcal{F}_j} \mathcal{L}(\gamma, f). \tag{16}$$

The selection of training subject γ^* provides the test subject γ with class-conditional complexity estimates $\widehat{k}_{\gamma^* c}$, as well as initial conditions obtained from the $\widehat{f}_{\gamma^* c}$ and $n_{\gamma^* c}/n_{\gamma^*}$, for

use in modelling $\mathscr{X}_\gamma$. (There are, of course, choices other than the likelihood-based option (Eq. (14)) for determining the "closest" training subject γ^*.)

4.2. Testing

4.2.1. Mixture modelling
Given γ^* selected as above, we estimate f_γ by estimating the parameters

$$
\begin{aligned}
\theta = [&\pi_{\gamma C}, \pi_{\gamma G}, \\
&\pi_{\gamma C1}, \ldots, \pi_{\gamma C(\widehat{k}_{\gamma C}-1)}, \mu_{\gamma C1}, \sigma^2_{\gamma C1}, \ldots, \mu_{\gamma C\widehat{k}_{\gamma C}}, \sigma^2_{\gamma C\widehat{k}_{\gamma C}}, \\
&\pi_{\gamma G1}, \ldots, \pi_{\gamma G(\widehat{k}_{\gamma G}-1)}, \mu_{\gamma G1}, \sigma^2_{\gamma G1}, \ldots, \mu_{\gamma G\widehat{k}_{\gamma G}}, \sigma^2_{\gamma G\widehat{k}_{\gamma G}}, \\
&\pi_{\gamma W1}, \ldots, \pi_{\gamma W(\widehat{k}_{\gamma W}-1)}, \mu_{\gamma W1}, \sigma^2_{\gamma W1}, \ldots, \mu_{\gamma W\widehat{k}_{\gamma W}}, \sigma^2_{\gamma W\widehat{k}_{\gamma W}}]'
\end{aligned}
\tag{17}
$$

via

$$
\widehat{f_\gamma} := \arg\max_{\theta \in \Theta} \prod_{x \in \mathscr{X}_\gamma} \sum_{c \in \mathscr{C}} \pi_{\gamma c} \sum_{t=1}^{\widehat{k}_{\gamma c}} \pi_{\gamma ct} \varphi_{\gamma ct}(x).
\tag{18}
$$

Notice that this involves estimation of $3\sum_{c \in \mathscr{C}} \widehat{k}_{\gamma^* c} - 1$ parameters. Algorithmically, this M-estimate is obtained using the EM algorithm (see, for instance, McLachlan and Krishnan, 1997) with the training model $\widehat{f_{\gamma^*}}$ as the starting point;

$$
\widehat{f_\gamma} := EM(\mathscr{X}_\gamma; \widehat{f_{\gamma^*}}).
\tag{19}
$$

The estimate obtained in this manner can be written as

$$
\widehat{f_\gamma} = \sum_{c \in \mathscr{C}} \widehat{\pi}_{\gamma c} \widehat{f_{\gamma c}} = \sum_{c \in \mathscr{C}} \widehat{\pi}_{\gamma c} \sum_{t=1}^{\widehat{k}_{\gamma c}} \widehat{\pi}_{\gamma ct} \widehat{\varphi}_{\gamma ct}.
\tag{20}
$$

Class complexities and component labels are inherited from $\widehat{f_{\gamma^*}}$, so that (for $c = G$, for instance) the mixing coefficient $\widehat{\pi}_{\gamma G}$ indicates the amount (proportion) of gray matter in the test image while $\widehat{f_{\gamma G}} = \sum_{t=1}^{\widehat{k}_{\gamma G}} \widehat{\pi}_{\gamma Gt} \widehat{\varphi}_{\gamma Gt}$ provides a model for that gray matter.

This unsupervised modelling of the test subject mixture allows estimation of the proportion and character of the different tissue types as they appear in the test subject, with a goal toward improving the segmentation for this subject.

4.2.2. Classification
We next consider the Bayes plug-in classifier

$$
g(x) = \arg\max_{c \in \mathscr{C}} \widehat{\pi}_{\gamma c} \widehat{f_{\gamma c}}(x).
\tag{21}
$$

The voxel x is to be labelled as belonging to the class which maximizes posterior probability of class membership. See Devroye et al. (1996) for a thorough discussion of the Bayes plug-in classifier.

This is (operationally) equivalent to the likelihood ratio test procedure (our estimates are indeed marginal density estimates) given by considering

$$LRT_{C/G}(x) = \frac{\widehat{\pi}_{\gamma C} \widehat{f}_{\gamma C}(x)}{\widehat{\pi}_{\gamma G} \widehat{f}_{\gamma G}(x)} =: r_1(x) \tag{22}$$

and

$$LRT_{G/W}(x) = \frac{\widehat{\pi}_{\gamma G} \widehat{f}_{\gamma G}(x)}{\widehat{\pi}_{\gamma W} \widehat{f}_{\gamma W}(x)} =: r_2(x). \tag{23}$$

Our automatic segmentation is then given by the following rules:

- $r_1(x) > 1$ implies voxel x is to be labelled as tissue class C.
- $r_2(x) < 1$ implies voxel x is to be labelled as tissue class W.
- $r_1(x) < 1$ & $r_2(x) > 1$ implies voxel x is to be labelled as tissue class G.
- $r_1(x) > 1$ & $r_2(x) < 1$ should not occur.
    - (Fig. 3 suggests stochastic ordering, $C <^{st} G <^{st} W$; this event is labelled "unknown" in the event that it does occur.)
- (Ties are to be broken arbitrarily.)

5. Results

An example of the subject-specific, class-conditional marginal density estimates obtained via the AKM procedure presented above is presented in Fig. 4. Table 2 presents the complexity estimation results $\widehat{k}_{jc}$ for the 10 cingulate gyri under consideration for three choices of $\widehat{\delta}$. From Fig. 4 we see that these (representative) fits are quite accurate. From Table 2 we see that $\widehat{\delta} = 1$ leads to a very high complexity, while $\widehat{\delta} = 25$ and $\widehat{\delta} = 99$ provide a reasonable framework to investigate the parsimony-fit trade-off.

There is no "gold standard" available for this investigation; that is, we do not know the *true* voxel class labels. The expert's hand-segmentation is the "lead standard" against which
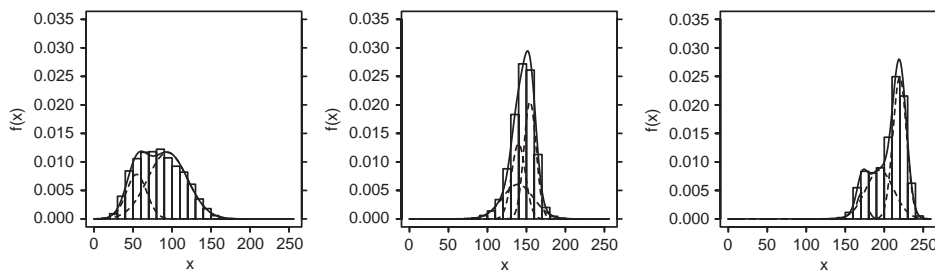

Fig. 4. Depicted are mixtures ($\widehat{\delta} = 25$) and histograms for *CG 9*. From left to right: C, G, W. Estimated model complexities are $\widehat{k}_{CG9C} = 2$, $\widehat{k}_{CG9G} = 3$, $\widehat{k}_{CG9W} = 3$.

Table 2
Mixture complexity estimation results for 10 cingulate gyri

	Estimated number of components									
	CG 1	CG 2	CG 3	CG 4	CG 5	CG 6	CG 7	CG 8	CG 9	CG 10
$\widehat{\delta} = 1$										
C	3	1	4	3	3	2	26	3	2	2
G	2	16	3	1	3	1	1	1	3	3
W	3	13	5	16	17	13	6	16	12	11
$\widehat{\delta} = 25$										
C	3	1	2	3	3	1	2	2	2	2
G	1	2	3	1	2	1	1	1	3	3
W	3	2	3	4	3	2	3	4	3	5
$\widehat{\delta} = 99$										
C	1	1	1	1	1	1	2	1	1	1
G	1	1	2	1	2	1	1	1	1	2
W	1	2	3	3	3	2	2	2	3	3

Table 3
Segmentation results for 10 cingulate gyri

	Probability of misclassification									
	CG 1	CG 2	CG 3	CG 4	CG 5	CG 6	CG 7	CG 8	CG 9	CG 10
PV1	0.15	0.09	0.23	0.20	0.28	0.15	0.15	0.16	0.20	0.24
PV2	0.09	0.06	0.13	0.11	0.18	0.09	0.09	0.11	0.12	0.15
AKM ($\widehat{\delta} = 25$)	0.13	0.07	0.12	0.09	0.11	0.10	0.07	0.10	0.09	0.11
AKM ($\widehat{\delta} = 99$)	0.10	0.07	0.13	0.10	0.15	0.09	0.07	0.10	0.08	0.10

we compare. Competing segmentation methodologies include:

- An overly-simplistic model in which each tissue class is modelled with but a single normal.
- Partial volume approaches (PV1 and PV2 in Table 3).
  These two partial volume methodologies are sketched here.

The partial volume approaches PV1 (Ratnanather et al., 2001) and PV2 (Ratnanather et al., 2004) involve a five-component Gaussian mixture fit via the EM algorithm; initial conditions are hand-selected based on the frequency histogram for the entire data set (e.g., the top panel in Fig. 3). One mixture component is initialized to account for each of the three tissue classes, and a "partial volume" mixture component is initialized between C & G and another between G & W. The difference between the two approaches is in the manner in which these partial volumes are allocated to the various tissue classes. These partial volume approaches are competitive with the most successful semi-automated MR

brain image segmentation methodologies available in the literature, and are in use in current brain segmentation projects (Ratnanather et al., 2001, 2004).

In method PV1, $\tau_{C/G} = 1/2$ of the *C/G* partial volume component is allocated to C and the other $1 - \tau_{C/G}$ to G. That is, a voxel observation *x* for which the *C/G* partial volume component maximizes (over the five mixture components) posterior probability of class membership is assigned to tissue class C (resp. *G*) if it takes a value smaller (resp. larger) than the mean of the *C/G* partial volume component. The *G/W* partial volume component is allocated analogously.

Method PV2, on the other hand, optimizes probability of misclassification over $\tau_{C/G} \in [0, 1]$ and $\tau_{G/W} \in [0, 1]$. For the cingulate gyrus data set under consideration herein, this optimization results in $\tau_{C/G}^* = 0.85$ and $\tau_{G/W}^* = 0.92$.

Table 3 presents the probability of misclassification results for the 10 cingulate gyri under consideration for the AKM hierarchical mixture model with two choices of $\hat{\delta}$, as well as analogous results for the two competing partial volume approaches described above. Despite the small sample size of just 10 subjects, Table 3 suggests that the hierarchical mixture model methodology may provide superior automatic segmentation results to the competing approaches. For example, a paired one-sided *t*-test of AKM ($\hat{\delta} = 99$) vs. PV2 yields a *p*-value of $p = 0.03$.

6. Discussion and conclusions

Voxel segmentation is one step of a sequence of procedures used in cortical analysis. Knowledge of volumes of gray matter, the coordinates of gray matter voxels and vertices of the gray/white surface is essential in a careful scientific analysis of cortical regions in normal and diseased subjects.

A real constraint on the ability to detect differences in two populations of MR brain imagery, and hence to investigate important neuropsychiatric hypotheses, is the cost of obtaining hand-segmentated brain volumes. This cost restricts the available sample sizes. Automated segmentation promises to allow for significantly larger sample sizes, and thus to advance the state of the science.

A hierarchical mixture modelling approach is demonstrated to provide automatic segmentation of magnetic resonance images superior to that of competing partial volume approaches. Our modelling methodology provides a key aspect missing in the methods currently in use: automatic subject-specific, class-conditional model complexity selection. The statistically innovative aspect of our work is the use of classified training data to obtain a model complexity for the unclassified test data, and hence allowing for richer test data models (and, subsequently, superior segmentation performance).

While specific details of the model fitting employed herein may be beneficially altered for particular applications, the general approach of (1) semiparametric estimation of subject-specific class-conditional marginal densities for a set of training volumes, (2) nearest neighbor matching of the test data to the training models providing for automated class-conditional mixture complexities, (3) parameter fitting of the selected training model to the test data, and (4) plug-in Bayes classification of unlabelled voxels, provides an advance in the state-of-the-art for automated MR brain image segmentation.

It is instructive to know which voxels are misclassified. In fact, for AKM the misclassified voxels are likely to be near class boundaries. This knowledge should be exploited to improve performance. However, such improvements may need to take into account that the "lead-standard" against which we test is based on expert hand-segmentation, and must not be considered a perfect "gold-standard."

A useful generalization of our methodology involves "combining classifiers"; that is, rather than considering just the closest training model, one can perform the classification as above for each training model and then combine the results by weighting the classification obtained via the jth training model inversely proportional to $\max_{f \in \mathscr{F}_j} \mathscr{L}(\gamma, f)$. For our cingulate gyrus investigation this alteration improves the performance for seven of the 10 subjects. Note, however, that computational requirements increase proportional to the number of models used in the combination.

One aspect of our methodology for which further investigation is particularly warranted is our ad hoc accounting for spatial dependence (our $\widehat{\delta}$). It will be beneficial to develop and use a more sophisticated spatially varying model to replace this crude covariogram estimate of the "effective number of independent observations" used herein.

The application of our methodology is not limited to just the cingulate gyrus. Experiments have been performed on medial prefrontal gyrus and occipital gyrus data sets; results indicate, again, the potential advantage of AKM. Indeed, MR brain image segmentation via AKM is a general tool for automated cortical analysis.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Contr. 19, 716–723.

Bartzokis, G., Mintz, J., Marx, P., Osborn, D., Gutkind, D., Chiang, F., Phelan, C.K., Marder, S.R., 1993. Reliability of in vivo volume measures of hippocampus and other brain structures using MRI. Magnet. Resonance Imaging 11, 993–1006.

Botteron, K.N., Figiel, G.S., 1997. The neuromorphometry of affective disorders. In: Krishnan, K.R.R. (Ed.), Brain Imaging in Clinical Psychiatry. Dekker, New York, NY, pp. 145–184.

Caunce, A., Taylor, C.J., 2001. Building 3d sulcal models using local geometry. Med. Image Anal. 5, 69–80.

Chen, J., Kalbfleisch, J.D., 1996. Penalized minimum distance estimates in finite mixture models. Can. J. Statist. 24, 167–175.

Chen, T., Metaxas, D., 2000. Image segmentation based on the integration of markov random fields and deformable models. In: Medical Image Computing and Computer-Assisted Intervention—Miccai 2000, Vol. 1935. Lecture Notes in Computer Science. Springer, Berlin, pp. 256–265.

Cho, Z.H., Jones, J.P., Singh, M., 1993. Foundations of Medical Imaging, Wiley, New York.

Christensen, G.E., Joshi, S.C., Miller, M.I., 1997. Volumetric transformation of brain anatomy. IEEE Trans. Med. Imaging 16, 864–877.

Cressie, N.A.C., 1993. Statistics for Spatial Data, Wiley, New York.

Csernansky, J.G., Bardgett, M.E., 1998. Limbic-cortical neuronal damage and the pathophysiology of schizophrenia. Schizophrenia Bull. 24, 231–248.

Dacunha-Castelle, D., Gassiat, E., 1997. The estimation of the order of a mixture model. Bernoulli 3, 279–299.

Dacunha-Castelle, D., Gassiat, E., 1999. Testing the order of a model using locally conic parameterization: population mixtures and stationary ARMA processes. Ann. Statist. 27, 1178–1209.

Devroye, L., Gyorfi, L., Lugosi, G., 1996. A Probabilistic Theory of Pattern Recognition, Springer, Berlin.

Drevets, W.C., 1999. Prefrontal cortical-amygdalar metabolism in major depression. Ann. NY Acad. Sci. 877, 614–637.

Drevets, W.C., Todd, R.D., 1997. Depression, mania and related disorders. In: Guze, S.B. (Ed.), Adult Psychiatry. Mosby, St Louis, MO, pp. 99–142.

Drevets, W.C., Videon, T.O., Price, J.L., Preskorn, S.H., Carmichael, S.T., Raichle, M.E., 1992. A functional anatomical study of unipolar depression. J. Neurosci. 12, 3628–3641.

Duvernoy, H.M., 1999. The Human Brain: Surface, Three-Dimensional Sectional Anatomy With MRI, and Blood Supply, Springer, New York.

Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. J. Amer. Statist. Assoc. 90, 577–588.

Everitt, B.S., Hand, D.J., 1981. Finite Mixture Distributions, Chapman and Hall, London.

Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: Constructing Geometrically Accurate and Topologically Correct Models of the Human Cerebral Cortex. IEEE Trans. Med. Imaging 20, 70–80.

Garrido, A., De la Blanca, N.P., 1998. Physically-based active shape models: initialization and optimization. Patt. Recogn. 31, 1003–1017.

Garrido, A., de la Blanca, N.P., 2000. Applying deformable templates for cell image segmentation. Patt. Recogn. 33, 821–832.

George, E.I., Foster, D.P., 2000. Calibration and empirical bayes variable selection. Biometrika 87, 731–747.

Grabowski, T.J., Frank, R.J., Szumski, N.R., Brown, C.K., Damasio, H., 2000. Validation of partial tissue segmentation of single-channel magnetic resonance images of the brain. Neuroimage 12, 640–656.

Haller, J.W., Banerjee, A., Christensen, G.E., Gado, M., Joshi, S., Miller, M.I., Sheline, Y., Vannier, M.W., Csernansky, J.G., 1997. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology 202, 504–510.

Hastie, T., Tibshirani, R., 1996. Discriminant analysis by Gaussian mixtures. IEEE Trans. Patt. Anal. Machine Intell. 18, 607–616.

Henna, J., 1985. On estimating of the number of constituents of a finite mixture of continuous distributions. Ann. Inst. Statist. Math. 37, 235–240.

Holmes, C.J., Hoge, R., Woods, R.P., Evans, A.C., Toga, A.W., 1998. Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22, 324–333.

Jain, A.K., Zhong, Y., Lakshmanan, S., 1996. Object matching using deformable templates. IEEE Trans. Patt. Anal. Mach. Intell. 18, 267–278.

James, L.F., Priebe, C.E., Marchette, D.J., 2001. Consistent estimation of mixture complexity. Ann. Statist. 29, 1281–1296.

Joshi, M., Cui, J., Doolittle, K., Joshi, S., Van Essen, D., Wang, L., Miller, M.I., 1999. Brain segmentation and the generation of cortical surfaces. Neuroimage 9, 461–476.

Kapur, T., Grimson, W.E.L., Wells-III, W.M., Kikinis, R., 1996. Segmentation of brain tissue from magnetic resonance images. Med. Image Anal. 1, 109–127.

Keribin, C., 2000. Consistent estimation of the order of mixture models. Sankhya 62, 49–62.

Kervrann, C., Heitz, F., 1999. Statistical deformable model-based segmentation of image motion. IEEE Trans. Image Process. 8, 583–588.

Leadbetter, M.R., Lindgren, G., Rootzen, H., 1983. Extremes and Related Properties of Random Sequences and Processes, Springer, New York.

Marchette, D.J., Priebe, C.E., Rogers, G.W., Solka, J.L., 1996. The Filtered Kernel Estimator. Comput. Statist. 11, 95–112.

McLachlan, G.J., 1987. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. Appl. Statist. 36, 318–324.

McLachlan, G.J., Basford, K.E., 1988. Mixture Models: Inference and Applications to Clustering, Marcel Dekker, New York.

McLachlan, G.J., Krishnan, T., 1997. The EM Algorithm and Extensions, Wiley, New York.

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models, Wiley, New York.

Mignotte, M., Meunier, J., 2001. A multiscale optimization approach for the dynamic contour-based boundary detection issue. Comput. Med. Imaging and Graph. 25, 265–275.

Miller, M.I., Massie, A.B., Ratnanather, J.T., Botteron, K.N., Csernansky, J.G., 2000. Bayesian construction of geometrically based cortical thickness metrics. Neuroimage 12, 676–687.

Montagnat, J., Delingette, H., 1997. Volumetric medical images segmentation using shape constrained deformable models. In: Cvrmed-Mrcas'97, Vol. 1205. Lecture Notes in Computer Science. Springer, Berlin, pp. 13–22.

Montagnat, J., Delingette, H., Malandain, G., 1999. Cylindrical echocardiographic image segmentation based on 3d deformable models. In: Medical Image Computing and Computer-Assisted Intervention, Miccai'99, Proceedings, Vol. 1679. Lecture Notes in Computer Science. Springer, Berlin, pp. 168–175.

Montagnat, J., Delingette, H., Ayache, N., 2001. A review of deformable surfaces: topology geometry and deformation. Image and Vision Comput. 19, 1023–1040.

Pham, D., Xu, C., Prince, J., 2000. Current methods in medical image segmentation. Ann. Rev. Biomed. Eng. 2, 315–337.

Priebe, C.E., Marchette, D.J., 2000. Alternating kernel and mixture density estimates. Comp. Statist. Data Anal. 35, 43–65.

Ratnanather, J.T., Botteron, K.N., Nishino, T., Massie, A.B., Lal, R.M., Patel, S.G., Peddi, S., Todd, R.D., Miller, M.I., 2001. Validating cortical surface analysis of medial prefrontal cortex. Neuroimage 14, 1058–1069.

Ratnanather, J.T., Wang, L., Nebel, M.B., Hosakere, M., Han, X., Csernansky, J.G., Miller, M.I., 2004. Validation of semiautomated methods for quantifying cingulate cortical metrics in schizophrenia. Psychiat. Res. Neuroimaging, in press.

Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.

Robb, R.A., 1999. Biomedical Imaging, Visualization and Analysis, Wiley, New York.

Roeder, K., Wasserman, L., 1997. Practical Bayesian density estimation using mixtures of normals. J. Amer. Statist. Assoc. 92, 894–902.

Schultz, N., Conradsen, K., 1998. 2d vector-cycle deformable templates. Signal Process. 71, 141–153.

Sclaroff, S., Liu, L.F., 2001. Deformable shape detection and description via model-based region grouping. IEEE Trans. Patt. Anal. Mach. Intell. 23, 475–489.

Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13, 856–876.

Shen, D.G., Herskovits, E.H., Davatzikos, C., 2001. An adaptive-focus statistical shape model for segmentation and shape modeling of 3-d brain structures. IEEE Trans. Med. Imaging 20, 257–270.

Swerdlow, N.R., Koob, G.F., 1987. Dopamine, schizophrenia, mania and depression: toward a unified hypothesis of cortico-striato-pallido-thalamic function. Behav. Brain Sci. 10, 197–245.

Tagare, H.D., 1997. Deformable 2-d template matching using orthogonal curves. IEEE Trans. Med. Imaging 16, 108–117.

Teo, P.C., Sapiro, G., Wandell, B.A., 1997. Creating connected representations of cortical gray matter for functional MRI visualization. IEEE Med. Trans. 16, 852–863.

Titterington, D.M., Smith, A.F.M., Makov, U.E., 1990. Statistical Analysis of Finite Mixture Distributions, Wiley, New York.

Tootell, R., Hadjikhani, N., Mendola, J., Marett, S., Dale, A., 1998. From retinotopy to recognition: fMRI in human visual cortex. Trends Cog. Sci. 2, 174–183.

Wells-III, W.M., Grimson, W.E.L., Kikinis, R.L., Jolesz, F.A., 1996. Adaptive segmentation of MRI Data. IEEE Trans. Med. Imaging 15, 429–442.

Westin, C.-F., Lorigo, L.M., Faugeras, O., Grimson, W.E.L., Dawson, S., Norbash, A., Kikinis, R., 2000. Segmentation by adaptive geodesic active contours. In: Medical Image Computing and Computer-Assisted Intervention—Miccai 2000, Vol. 1935. Lecture Notes in Computer Science. Springer, Berlin, pp. 266–275.

Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metabol. 12, 900–918.

Xu, C., Prince, J.L., 2000. Gradient vector flow deformable models. In: Handbook of Medical Imaging. Academic Press, San Diego, CA.

Xu, C., Pham, D.L., Rettmann, M.E., Yu, D.N., Prince, J.L., 1999. Reconstruction of the human cerebral cortex from magnetic resonance images. IEEE Trans. Med. Imaging 18, 467–480.

Xu, C., Pham, D.L., Prince, J.L., 2000. Medical image segmentation using deformable models. In: SPIE Handbook on Medical Imaging—Vol. III. Medical Image Analysis, SPIE, pp. 129–174.

Yantis, S., 2001. Visual Perception, Psychology Press, Philadelphia, PA.

Yezzi, A., Tsai, A., Willsky, A., 2002. A fully global approach to image segmentation via coupled curve evolution equations. J. Visual Commun. Image Representation 13, 195–216.

Zipursky, R.B., Lim, K.O., Sullivan, E.V., Brown, B.W., Pfefferbaum, A., 1992. Widespread cerebral gray volume deficits in schizophrenia. Arch. Gen. Psychiatry 49, 195–205.