MRI Tissue Classification Using Bayesian Hidden Markov Normal Mixture Models

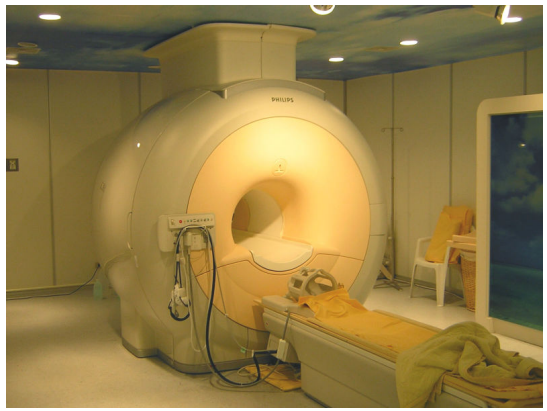
Dai Feng and Luke Tierney

Department of Statistics & Actuarial Science
University of Iowa

December 6, 2007

Magnetic Resonance Imaging (MRI)

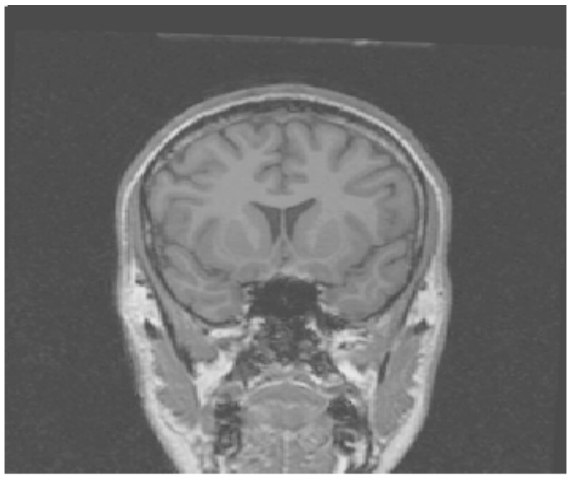
- MRI is a non-invasive method for imaging the inside of objects.
- Based on nuclear magnetic resonance.
- MRI has many medical applications.
- Different imaging modalities emphasise different tissue contrasts:
  - T1-weighted, used in structural brain imaging
  - T2, T2*-weighted, often used in functional MRI
  - Unweighted proton density (PD), may help for some anomalies
- Sometimes more than one image type is available.
- Each image is a 3D array of image intensities, one for each voxel (volume picture element).



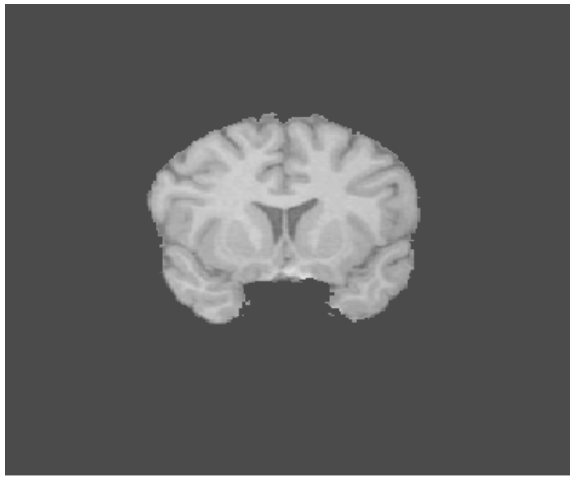
Clinical MRI Scanner

MR Imaging the Human Brain

- T1-weighted images are usually used for examining brain structure.
- 2D slices of a 3D image:



Full MR Image



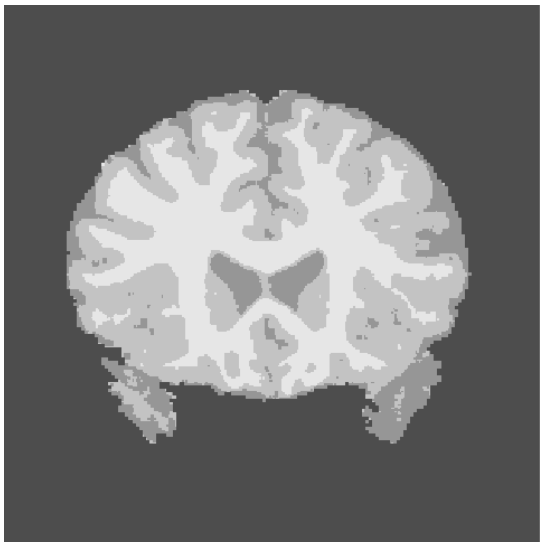
Brain-Only Image

- Data are often pre-processed:
  - alignment to eliminate motion
  - masking out non-brain tissue
  - warping to standard coordinates (multiple subjects)

Brain Tissue Classification

- Major brain tissue types:
  - White matter (WM)
  - Gray matter (GM)
  - Cerebrospinal fluid (CSF)

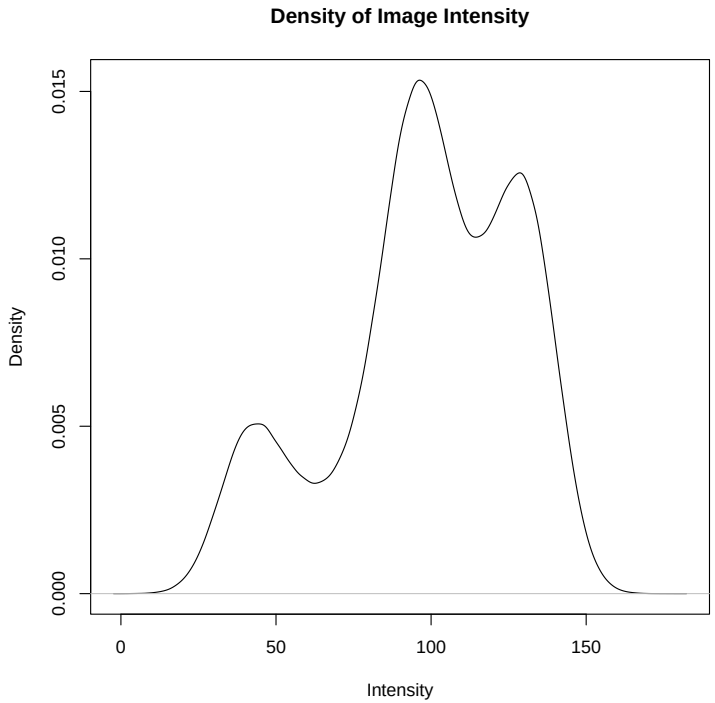
  There are others, but tissue classification usually focuses on these.
- Some applications:
  - Diagnosis of disease
  - Surgery preparation
  - Background for functional imaging
- Manual tissue classification is very labor intensive.
- Automated methods try to match quality of manual at lower cost.
- Usually based on intensities in a T1 MR image.



WM = light gray
GM = medium gray
CSF = dark gray

Basic Properties of the Data

- Data consist of *image intensities* $y_1, \ldots, y_N$ for N voxels in a 3D grid.
- A typical value of N is 256^3.
- Intensities are often scaled to $[0, 255]$ and rounded to an integer.
- Tissue types are denoted by $z \in \{1, \ldots, K\}$; we use $K = 3$.
- A density plot for a relatively low noise MR image:



A Simple Mixture Model

- A common model: given the tissue structure z, intensities are
  - independent
  - normally distributed,

$$y_i|z_i \sim \mathrm{N}(\mu(z_i), \sigma(z_i)^2)$$

- Mean and and SD depend on the tissue type.
- Assuming tissue types are independent leads to a simple mixture model

$$f(y_i) = \sum_{z=1}^{K} f(y_i|\mu(z), \sigma(z))p_z$$

- Parameters are easily estimated by the EM algorithm.
- Tissue types can be assigned using the Bayes classifier.

Incorporating Spatial Information

- Adjacent voxels are likely to contain the same tissue type.
- A more realistic model accounts for this spatial homogeneity in z.
- The Potts model family provides simple models for spatial homogeneity:

$$p(z) \propto \exp\left\{\sum_i \alpha_i(z_i) + \sum_{i\sim j} w_{ij}(z_i, z_j)\right\}$$

- The specific model is determined by
  - a neighborhood structure $i \sim j$
  - a weight function $w_{ij}(\cdot,\cdot)$
  - an external field $\alpha_i(\cdot)$
- This is an example of a Markov random field model.

Incorporating Spatial Information

The simple Potts model

- The simple Potts model spatial prior distribution has
  - $\alpha_i(z) \equiv 0$
  - $w_{ij}(z_i, z_j) = \beta 1_{\{z_i = z_j\}}$ with $\beta \geq 0$
  - six neighbors: right, left, front, back, above, below

  The value of β is often determined in a pilot study.
- The spatial model can be fit by maximizing the likelihood

$$\prod_i f(y_i|\mu(z_i), \sigma(z_i))p(z)$$

  by alternately
  - maximizing over μ, σ for fixed z
  - maximizing over each z_i for fixed values of μ, σ and the other z values.
- This is the Iterated Conditional Modes (ICM) algorithm.

Incorporating Spatial Information
A Bayesian formulation

- Alternatively, we can
  - specify a prior distributions $p(\mu, \sigma)$ on μ, σ
  - use MCMC to compute characteristics of the posterior distribution

$$p(\mu, \sigma, z|y) \propto \prod_i f(y_i|\mu(z_i), \sigma(z_i))p(z)p(\mu, \sigma)$$

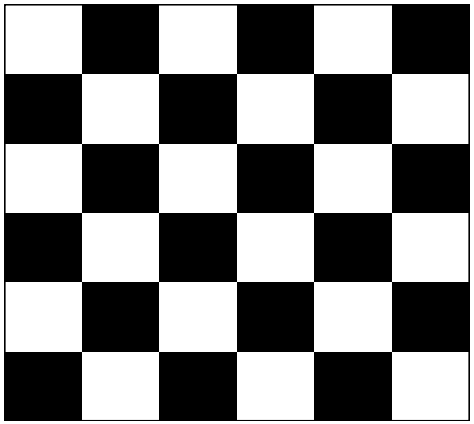
- Assume μ, σ, z are independent and
  - μ has a normal distribution
  - σ has an inverse Gamma distribution
- Then the full conditionals satisfy
  - $\mu \sim$ normal
  - $\sigma \sim$ inverse Gamma
  - $z \sim$ simple Potts model with external field

$$\alpha_i(z_i) = \log f(y_i|\mu(z_i), \sigma(z_i))$$

Incorporating Spatial Information

Some computational issues

- The Potts model can be sampled by single site updating.
- If the voxels are organized in a checkerboard pattern,



  then black voxels are conditionally independent given white ones.
- Black and white voxels can each be updated as a group.
- This can be used for vectorized computation in R or Matlab.
- This can also be used for parallel computation.

Incorporating Spatial Information
Computing a tissue classification using MCMC

- For each voxel we keep track of the number of times it is assigned to each tissue type.
- The most frequent type is assigned as the voxel's tissue type.
- This is the marginal posterior mode (MPM) estimate.
- Modest run length of 100–1000 seem to be adequate in many cases.

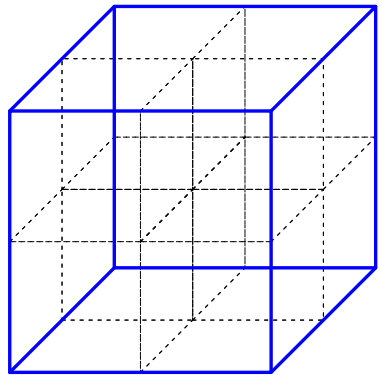
Partial Volume Effect

- Typical images have voxels of size $1mm^3$.
- Some voxels contain more than one tissue type.
- This is known as the *partial volume effect*.
- One approach is to introduce intermediate classes.
- Usually this is done by introducing two classes:
  - CSF/GM
  - GM/WM
- Voxels containing WM and CSF are very rare and are ignored.
- This helps reduce confounding in estimation.
- A number of studies have used this approach with
  - mixture models fit by EM
  - hidden Markov mixture models fit by ICM and variations
- Modeling the distribution of y_i for intermediate classes is challenging.

A Higher Resolution Spatial Model

We have adopted a different approach:

- Each voxel is divided in the x, y, z directions, producing 8 subvoxels.



- Each subvoxel is viewed as beckoning to only one tissue type.
- The observed voxel intensity y_i is

$$y_i = v_{i1} + \cdots + v_{i8}$$

where $v_{i1}, \ldots, v_{i8}$ are the *unobserved* subvoxel intensities.

A Higher Resolution Spatial Model

The subvoxel-level model

- Conditional on the tissue types the v_{ij} are independent normals.
- A spatial model is used at the subvoxel level.
- To capture the fact that CSF and WM rarely coexist in a voxel we use

$$w_{ij}(z_i, z_j) = \begin{cases} \beta_1 & \text{if } z_i = z_j \\ -\beta_2 & \text{if } \{z_i, z_j\} = \{\text{CSF,WM}\} \\ 0 & \text{otherwise} \end{cases}$$

We call β_2 the *repulsion parameter*.

A Higher Resolution Spatial Model

The posterior distribution

- The complete data posterior distribution is given by

$$p(\mu, \sigma, z|v) \propto \prod_i \prod_j f(v_{ij}|\mu(z_{ij}), \sigma(z_{ij}))p(z)p(\mu)p(\sigma)$$

- The observed data adds the constraints

$$y_i = v_{i1} + \cdots + v_{i8}$$

- Full conditional distribution of the unobserved data:

$$(v_{i1}, \ldots, v_{i8})|y, \mu, \sigma, z \sim \text{singular multivariate normal}$$

  independently for each voxel.
- Vectorized or parallel sampling of the subvoxel intensities is possible.

A Higher Resolution Spatial Model
MCMC sampling

- MCMC sampling alternates the two steps
  - generate new subvoxel data v given current μ, σ, and z values
  - generate new μ, σ, and z values given current subvoxel data v
- For each subvoxel we record how often it is assigned to each tissue.
- After a suitable run we have several options for tissue classification:
  - Assign each subvoxel to the tissue with the highest marginal posterior probability to get a high resolution image.
  - Use posterior probabilities as estimates of tissue proportions, at the voxel or subvoxel level.
  - Create a voxel-level classification based on proportion estimates or subvoxel classes.

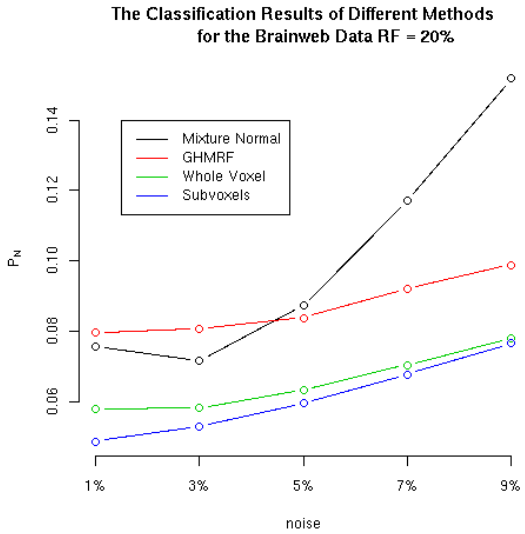
Empirical Results
BrainWeb Simulated Brain Database

- Challenge in evaluating classification methods: knowing ground truth.
- The BrainWeb project provides
  - a brain "phantom" with partial volume information
  - simulated MRI images with various noise levels

  Project URL: `http://www.bic.mni.mcgill.ca/brainweb/`
- β values were chosen by a pilot study based on minimizing mis-classification rates for whole-voxel classification.
  - For the whole voxel model this produces $\beta = 0.7$.
  - For the subvoxel model we obtained $\beta_1 = \beta_2 = 0.6$.
  - Results are not very sensitive to changes of ± 0.1 in β values.

Empirical Results
BrainWeb Simulated Brain Database

- Comparison of whole-voxel mis-classification rates:



- Mean squared errors in estimating tissue proportions:

Simple Mixture	Whole Voxel	Whole Voxel Post. prob	Subvoxel	Subvoxel Post. prob
0.1847	0.1135	0.0766	0.0615	0.0499

Empirical Results

Iowa Mental Health Research Center Data

- Data on 4 patients available.
- Only one of the 4 images has been analyzed so far.
- Noise level in images is relatively high.
- Two slices of each have been classified manually.
- Mis-classification rates of automatic classification procedures:

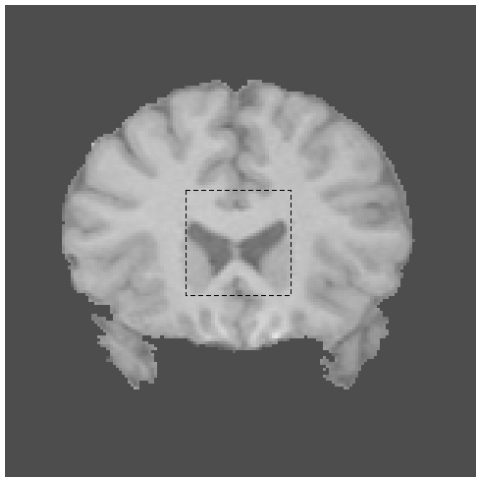
Simple Mixture	Standard HM Mixture	High-Resolution HM Mixture
20.9	21.3	13.6

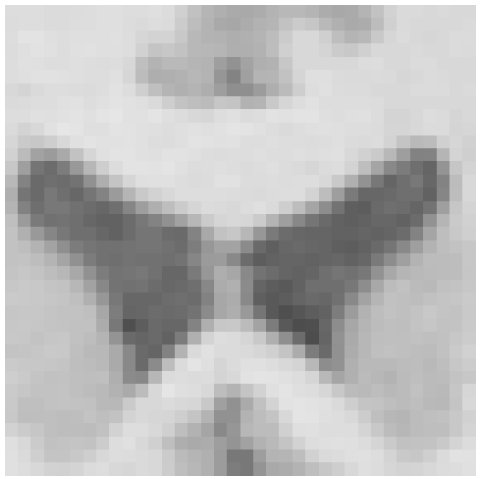
- Current software for the high resolution model takes 20 minutes for 100 iterations.
- Results are consistent with results for 1000 iterations.

Empirical Results
Iowa Mental Health Research Center Data

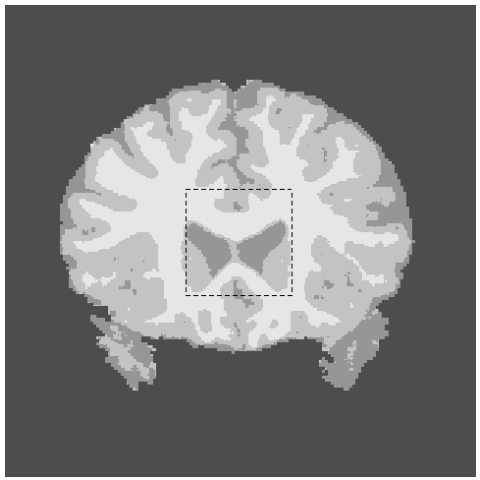
High resolution classification:

Original MR Image





Classification Results



Future Work

- Explore other neighborhood structures and weightings.
- Using data T2 and/or PD modalities in addition to T1.
- Joint estimation of the Potts model parameters.
- Alternate sampling algorithms and implementations to
  - improve mixing
  - improve performance
- Extensions to support identifying pathologies, such as
  - tumors
  - lesions