MR IMAGE SEGMENTATION

Prepared by : Monil Shah

What is Segmentation ?

- Partitioning a region or regions of interest in images such that each region corresponds to one or more anatomic structures

Classification

- Classification means to assign to each point in the image a tissue class, where the classes are agreed in advance
- Note that the problems are inter-linked: a classifier implicitly segments an image, and a segmentation implies a classification

- An image can be represented as:

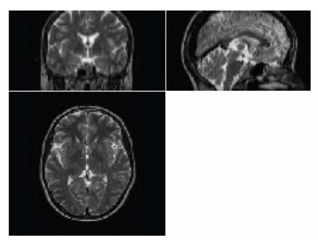
$$F_{P\times Q}=[f(x,y)]_{P\times Q}$$

- F is the set of all pixels and P() is a uniformity (homogeneity) predicate defined on groups of connected pixels, then segmentation is a partitioning of the set F into a set of connected subsets or regions (S1,S2,.....Sn) such that

$$\bigcup_{i=1}^{n} S_i = F \quad \text{with} \quad S_i \bigcap S_j = \emptyset, \quad i \neq j.$$

- The uniformity predicate P(Si) = true for all regions
- P (Si∪Sj) = false when Si is adjacent to Sj.

- Radiotherapy planning
- Surgical planning
- Clinical drug trials
- Multiple Sclerosis, MS Lesion quantification
- Determine Brain tumor volumetrices
- Artery Vein separation in MR Angiography
- Preprocessing for multimodality Image Registration



MR Image

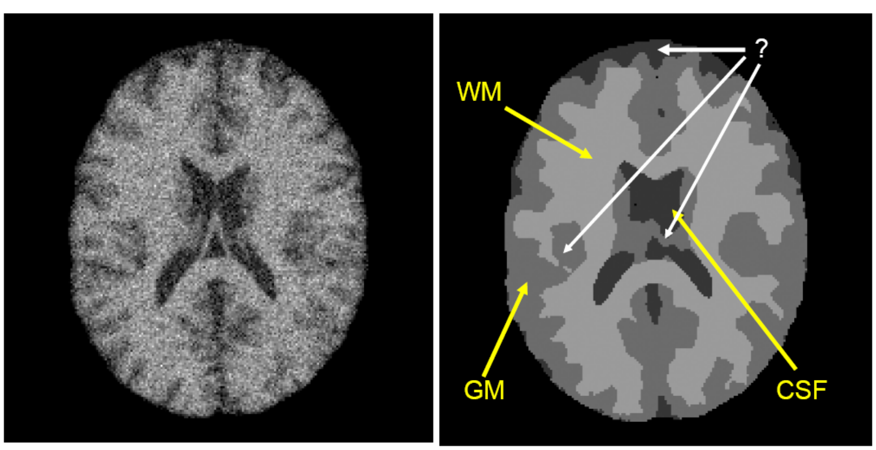
These two features facilitate segmentation

- Excellent contrast between soft tissues
- Brain images are approximately piecewise constant; with small number of classes

Practical Conditions

- There are image distortions (e.g. motion artifact, bias field,
- Partial volume effect (Many pixels contain more than one tissue type)
- – structures of interest (tumours, temporal lobes,
- ...) have complex shapes

- The noisy MRI image of the brain slice shown left is ideally piecewise constant, comprising grey matter, white matter, air, ventricles. The right image is a segmentation of the image at left.



Edge based Segmentation

- Mainly used for segmentation of large well defined structures such as brain Parenchyma but not to distinguish individual tissue types.
- Segmentation of White matter - Gray matter

- Edge: boundary between two regions with relatively distinct gray levels

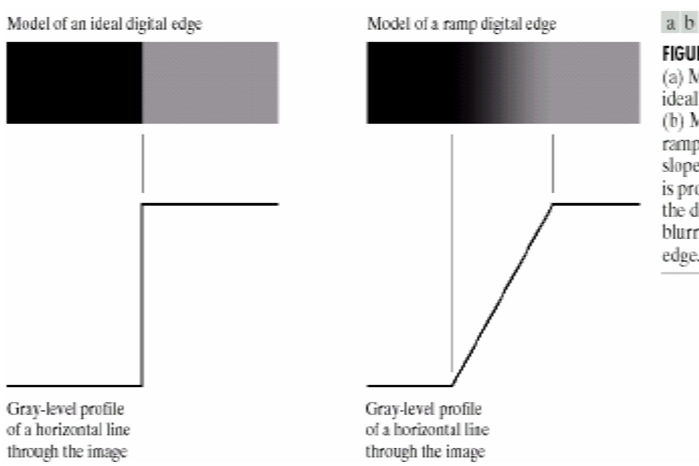
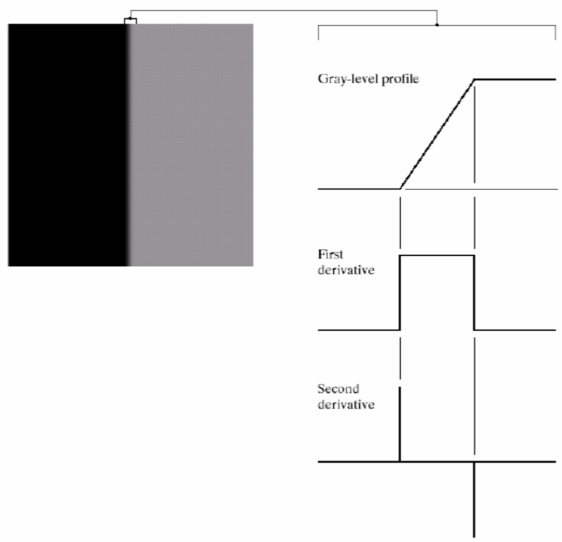


FIGURE 10.5 (a) Model of an ideal digital edge. (b) Model of a ramp edge. The slope of the ramp is proportional to the degree of blurring in the edge.

- Basic idea: computation of a local derivative operator.

a b
FIGURE 10.6
(a) Two regions separated by a vertical edge.
(b) Detail near the edge, showing a gray-level profile, and the first and second derivatives of the profile.



- The first order derivative can be used to detect the presence of edge at a point in an image

- The second order derivative is used to determine whether an edge pixel lies on the dark or light side of an edge

- An imaginary line joining the extreme positive and negative values would cross zero near the midpoint of the edge.

Gradient Operator

Gradient

$$\nabla \mathbf{f} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

Gradient magnitude

$$\nabla f = mag(\nabla \mathbf{f}) = \sqrt{\left(G_x^{\ 2} + G_y^{\ 2}\right)} \qquad \nabla f \approx |G_x| + |G_y|$$

Gradient direction

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

Laplacian

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\nabla^2 f = 8z_5 - \sum_{i=1}^{8} z_i$$

$$\nabla^2 f = 4z_5 - (z_2 + z_4 + z_6 + z_8)$$

FIGURE 10.13 Laplacian masks used to implement Eqs. (10.1-14) and (10.1-15), respectively.

0	−1	0
−1	4	−1
0	−1	0

−1	−1	−1
−1	8	−1
−1	−1	−1

- In this method, the image is first smoothed to reduce noise. Then, the second derivative (Laplacian) of the smoothed images is computed. The obtained image will have positive and negative regions. The boundary between the positive and negative regions is taken as the image edges.
- This process is the same as convolving the image with the Laplacian of a 2-D Gaussian and finding its zero-crossings.

Convolution: $$g(x,y) = f(x,y) * \nabla^2 h(x,y)$$

Gaussian function: $$h(x,y) = \exp(-\frac{x^2+y^2}{2\sigma^2})$$

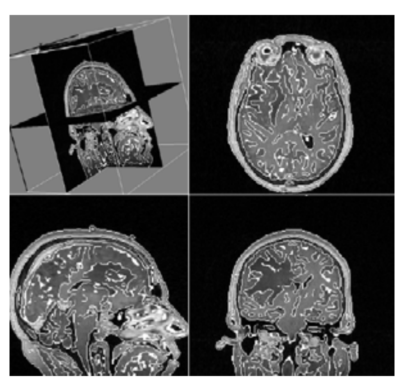
$$r^2 = x^2 + y^2$$

Second derivative: $$\nabla^2 h(r) = (\frac{r^2-\sigma^2}{\sigma^4})\exp(-\frac{r^2}{2\sigma^2})$$

- Zero crossings are approximated by thresholding the LoG image by setting all positive values to white and all negative values to black, as zero crossing occurs between positive and negative values of the Laplacian.

Advantages of Zero Crossing

- Edges are thinner
- Noise is reduced
- Rugged Performance



Region Based Segmentation

- Let R represent the entire image. Segmentation is a process that partitions R into n sub regions R1,R2,...,Rn such that:

$$\text{a) } \bigcup_{i=1}^{n} R_i = R$$

$$\text{b) } \mathrm{R}_{\mathrm{i}} \text{ is a connected region.}$$

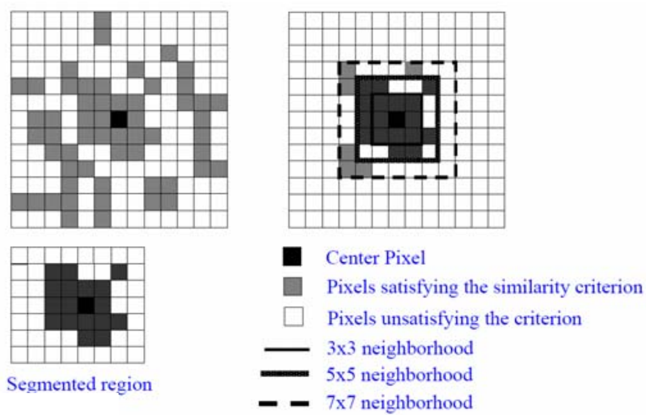
$$\text{c) } \mathrm{R}_{\mathrm{i}} \cap R_j = \phi$$

$$\text{d) } \mathrm{P}(\mathrm{R}_{\mathrm{i}}) = True$$

$$\text{e) } \mathrm{P}(\mathrm{R}_{\mathrm{i}} \cup R_j) = False$$

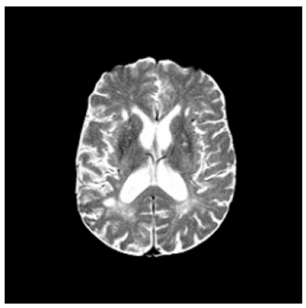
Region Growing

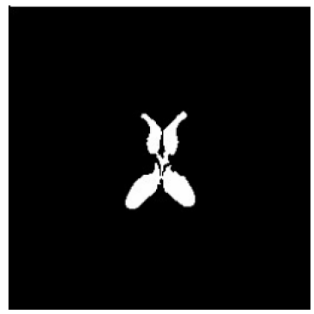
- Pixel aggregation: starts with a set of seed point and from these grows regions by appending to each seed point those neighboring pixels that have similar properties (e.g., gray-level, texture, color).



Region Growing

A T-2 weighted MR brain image (left) and the segmented ventricles (right) using the region-growing method.



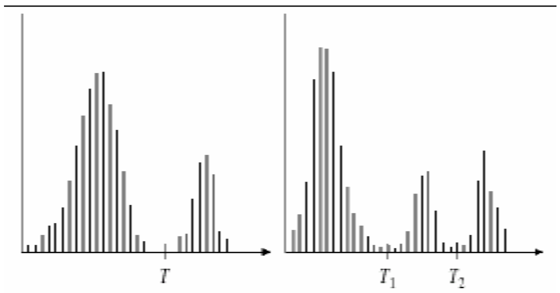


Problems:

1. Selection of initial seeds that properly represent regions of interest
2. Selection of suitable properties
3. Formulation of a stopping rule.
4. Only well defined regions can be robustly identified

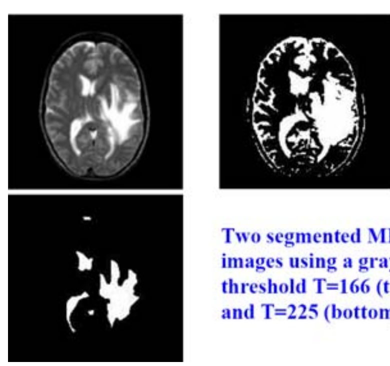
Intensity Thresholding

•Gray level histogram of an image f(x,y) composed of a light object on a dark background.
• To extract the object: select a threshold T that separates the gray levels of the background and the object.



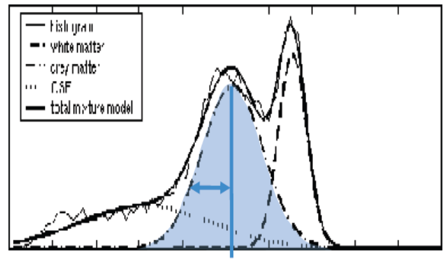
- Threshold in general can be calculated as:
- T= T(x,y,p(x,y),f(x,y))
- T depends only on f(x,y): global threshold
- T depends on f(x,y) and p(x,y): local threshold

The method is limited:
Clinical Applications are hindered
Variability of anatomy in brain MRI



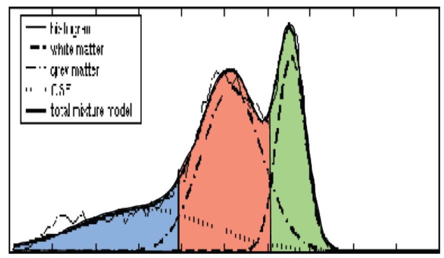
Two segmented MR brain images using a gray value threshold T=166 (top right) and T=225 (bottom)

Intensity distributions of white matter, gray matter and CSF are modeled as Gaussian distributions
mean = average intensity
variance = variation around the average intensity



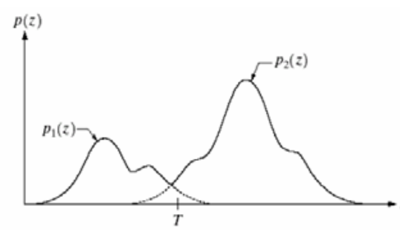
The threshold value is computed by:
–Detecting the valley between the modes in the histogram of the image.

Once the mean and variance of each tissue type is known, voxels can be classified based on their intensity



Optimal Global Thresholding

FIGURE 10.32 Gray-level probability density functions of two regions in an image.



$$p(z) = P_1 \cdot p_1(z) + P_2 \cdot p_2(z)$$

$$P_1 + P_2 = 1$$

Problem: how to optimally determine T to minimize the segmentation error?

Optimal Global Thresholding

Error 1: $E_1(T) = \int_{-\infty}^{T} p_2(z)dz$

Error 2: $E_2(T) = \int_{T}^{\infty} p_1(z)dz$

Total error: $E(T) = P_1 E_2(T) + P_2 E_1(T)$

Goal: what's the value of T to minimize E(T)?

Result: $P_1 p_1(T) = P_2 p_2(T)$

Optimal Global Thresholding: Gaussian PDFs

$$p(z) = \frac{P_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right) + \frac{P_2}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z-\mu_2)^2}{2\sigma_2^2}\right)$$

Solution: $AT^2 + BT + C = 0$

$$A = \sigma_1^2 - \sigma_2^2$$

$$B = 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)$$

$$C = \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 + 2\sigma_1^2 2\sigma_2^2 \ln(\sigma_2 P_1 / \sigma_1 P_2)$$

Drawbacks of previous Methods

- Edge detection algorithms are vulnerable to artifacts and noise
- Region growing algorithms depend on seed selection and whether the regions are well defined and therfore is not robust
- Thresholding is only good for well separated intensities

Classification of Images

- The “labels” we wish to assign to objects are typically few and known in advance
  – e.g. WM, GM, CSF and air for brain MRI
- Objects of interest usually form coherent continuous shapes
  – If a pixel has label c, then its neighbors are also likely to have label c
  – Boundaries between regions labeled c, d are continuous
- Image noise means that the label to be assigned initially at any pixel is probabilistic, not certain
- One way to accommodate these considerations is Hidden Markov Random Fields

Segmentation by classification of voxels

- Every pixel is classified according to its probability of being a member of a particular tissue class.
- The Maximum Likelihood (ML) estimator assigns to each voxel **x** that class which maximizes

$$\Pr(\mathbf{x} \in C), \text{ where } C \in \{WM, GM, CSF, \ldots\}$$

- The Maximum a Posteriori (MAP) estimator

$$P(\mathbf{x} \mid \mathbf{y}) = P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x}) / P(\mathbf{y})$$

Normalizes intensities $P_{MAP}(\mathbf{x} \mid \mathbf{y}) = P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x})$

In practice do not work well on brain

Each class is defined as a probability density function

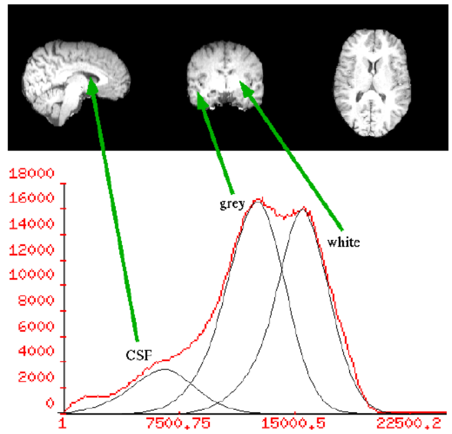
- Each class, say the one with label l (= GM say,) has an associated PDF, with parameters θ_l ,Often it is a Gaussian $\theta_l = (\mu_l, \sigma_l)$

$$p(y_i \mid l) = f(y_i; \theta_l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{(y_i - \mu_l)^2}{2\sigma_l^2} \right)$$

Tissue Probability Density Function

Note the huge overlap between the Gaussian pdf For GM and that for WM This means that there are many misclassifications of GM pixels as WM and vice versa.

Even small amounts of noise can change the ML classification.



HMRF Model

- Hidden Markov random field model (HMRF) is a stochastic process generated by a Markov random field whose state sequence is observed through a field of observations.
- Markov random field theory (MRF) encodes the spatial information of an image through contextual constraints of neighboring pixels. Neighboring pixels have same class labels or similar intensities.
- A statistical model is incomplete without its parameters being determined
- This is solved by using the Expectation-maximization (EM) algorithm.

MRF Model

A lattice of sites = the pixels of an image $S = \{1,\ldots,N\}$

A family of random variables, whose values define a configuration
$R = \{r_i, i \in S\}$
$(R_1 = r_1, \ldots, R_N = r_N) \equiv R = r$

The set of possible intensity values $D = \{1,\ldots,d\}$

An image viewed as a random variable $Y = \{\mathbf{y} = (y_1,\ldots,y_N) \mid y_i \in D, i \in S\}$

The set of class labels for each pixel (eg GM, WM, CSF) $L = \{1,\ldots,l\}$

A classification (segmentation) of the pixels in an image $X = \{\mathbf{x} = (x_1,\ldots,x_N) \mid x_i \in L, i \in S\}$

Pair wise Independence

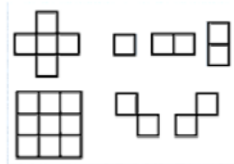
Classification is independent of image *and* neighbouring voxels are independent of each other
$$P(\mathbf{y},\mathbf{x}) = \prod_{i \in S} P(y_i, x_i)$$

Assume a parametric (eg Gaussian) form for distribution of intensity given label l
$$p(y_i \mid l) = f(y_i; \theta_l)$$

Markovian assumptions: probability of class label *i* depends only on the local neighbourhood N_i
$$P(\mathbf{x}) > 0, \forall \mathbf{x} \in X$$
$$P(x_i \mid x_{S-\{i\}}) = P(x_i \mid x_{N_i})$$

MRF equivalent to Gibbs distribution

- Deep theorem of Hammersley & Clifford
  - MRFs are equivalent to systems with a local potential function
- Define *cliques* made of neighbouring pixels



- Each clique has a local "potential energy" *U* which enables the probability of the pixel label to be calculated efficiently

$$\log P(\mathbf{x} \mid \mathbf{y}) = -U(\mathbf{x} \mid \mathbf{y})$$

$$\text{Equivalently}: P(\mathbf{x} \mid \mathbf{y}) = \exp(-U(\mathbf{x} \mid \mathbf{y}))$$

According to the local characteristics of MRFs, the joint probability of any pair of (*Xi*, *Yi*), given *Xi*'s neighborhood configuration is:

$$\mathrm{P}(y_i, x_i \mid x_{\mathcal{N}_i}) = P(y_i \mid x_i) P(x_i \mid x_{\mathcal{N}_i})$$

Thus, we can compute the marginal probability distribution of *Yi* dependent on the parameter set θ (random variable) and , $x_{\mathcal{N}_i}$

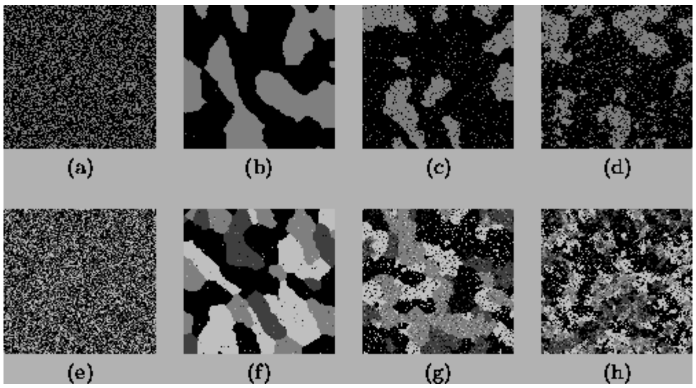
$$p(y_i \mid x_{\mathcal{N}_i}, \theta) = \sum_{\ell \in \mathcal{L}} p(y_i, \ell \mid x_{\mathcal{N}_i}, \theta) = \sum_{\ell \in \mathcal{L}} f(y_i; \theta_\ell) p(\ell \mid x_{\mathcal{N}_i}),$$

where $\theta = \{\theta_\ell, \ell \in \mathcal{L}\}$,

We call this the *hidden Markov random field* (HMRF) model. Note, the concept of an HMRF is different from that of an MRF in the sense that the former is defined with respect to a pair of random variable families (*X*,*Y*) while the latter is only defined with respect to *X*. More precisely, an HMRF model can be described by the following:

Model Simulation

Image simulation by the FGM model and the GHMRF model.
The first row shows the 3-class case. (a) FGM model;
(b)-(d) GHMRF mode with standard deviation 0.23, 0.4, 0.5, respectively.
The second row shows the 5-class case. (e) FGM model;
(f)-(h) GHMRF model with standard deviation 0.3, 0.47, 0.55, respectively.



MRF-MAP estimation

We seek a labelling of the image according to the MAP criterion:

$$\hat{x} = \arg\max_{\mathbf{x} \in X} \{ P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x}) \}$$

Recall that

$$P(x) = \frac{1}{Z} \exp(-U(x))$$

And, if we have Gaussian distributions of pixel values for each class, and $x_i = l$

$$p(y_i \mid x_i) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{(y_i - \mu_l)^2}{2\sigma_l^2} \right)$$

Pixel wise independence & ICM

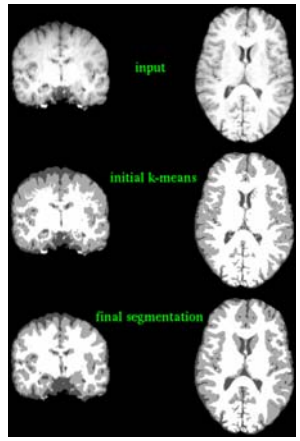
$$P(\mathbf{y} \mid \mathbf{x}) = \prod_{i \in S} p(y_i \mid x_i)$$

So that

$$U(\mathbf{y} \mid \mathbf{x}) = \sum_{i \in S} \left[\frac{(y_i - \mu_l)^2}{\sigma_l^2} + \log \sigma_l \right]$$

The final step is to use Besag's Iterated Conditional Modes algorithm:

Update the class label x_i^k at iteration k by minimising $U(x_i \mid \mathbf{y}, x_{N_i - \{i\}})$



Expectation-maximization

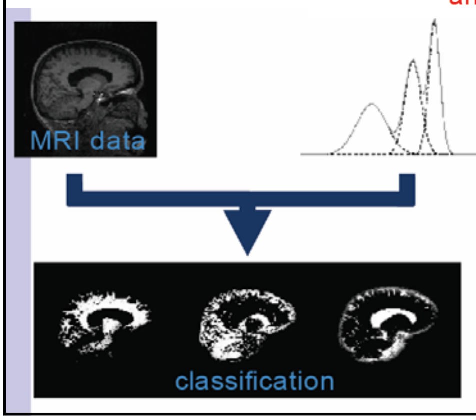
- statistical model is complete only if both its functional form and its parameters are determined. The procedure for estimating the unknown parameters is known as *model fitting*.
- We have assumed that we know the model parameters $\theta_l = (\mu_l, \sigma_l)$ ahead of time. Unfortunately, this is rarely the case. The next idea is to estimate the parameters and do the segmentation cooperatively using the EM algorithm

Expectation Step

Start: make initial estimates $\theta_i, i = 1 \ldots$

$$Q(\theta \mid \theta^k) = \sum_{x \in X} \underbrace{p(\mathbf{x} \mid \mathbf{y}, \theta^k) \log(p(x, y \mid \theta))}$$

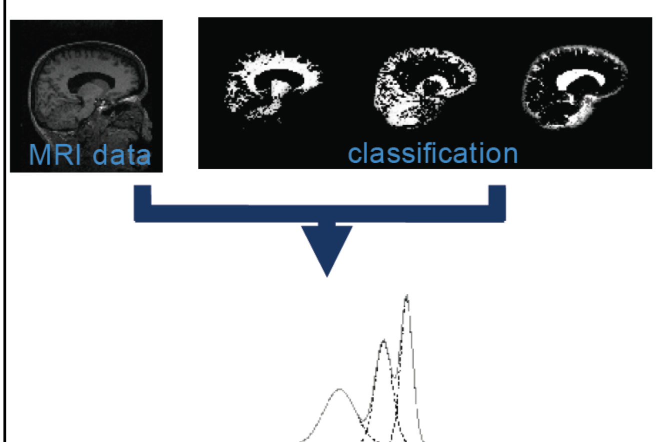
Expectation over the missing tissue labels based on the current parameter estimation and the observed data



Statistical classification of the image voxels based on the current parameter estimation

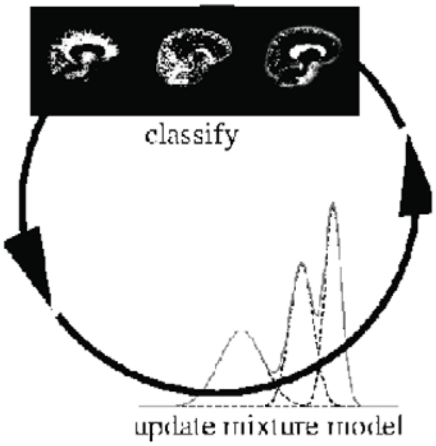
Maximization step

$$\theta^{k+1} = \arg\max_{\theta} Q(\theta \mid \theta^k)$$



Parameter re-estimation based on the current classification

Summarize EM



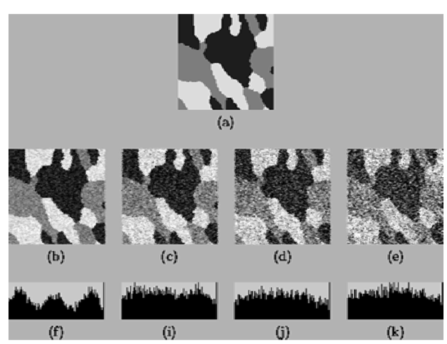
Simultaneous classification and parameter estimation

Likelihood is guaranteed to increase at each iteration

Experiment

- Figure 2(a) shows a simulated 3-class image sampled from an MRF model using the Gibbs sampler.
- Figure 2(b)-(e) show the same images with added Gaussian noise with standard deviation of 28, 47, 66, and 95.
- Because image contrast is what we are most interested in for examining qualities of an image, a measurement of the noise is more meaningful with image contrast being taken into account.

$$NCR = \frac{\text{standard deviation of the noise}}{\text{mean inter-class contrast}}.$$



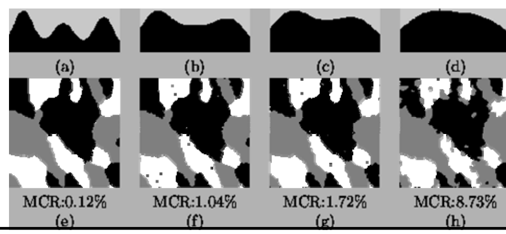
class	Class 1			Class 2			Class 3		
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3
$NCR = 0.3$	30	28	0.372	125	28	0.299	220	28	0.329
$NCR = 0.5$	30	47	0.372	125	47	0.299	220	47	0.329
$NCR = 0.7$	30	66	0.372	125	66	0.299	220	66	0.329
$NCR = 1.0$	30	95	0.372	125	95	0.299	220	95	0.329

- The standard HMRF-EM algorithm are then applied to the four test images until there is no significant change in the value of the *Q*-function. To measure the segmentation accuracy, we also define the misclassification ratio (MCR), which is

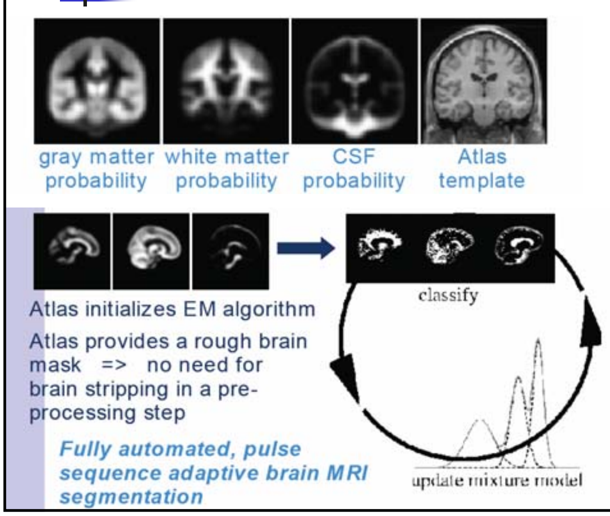
$$MCR = \frac{\text{number of mis-classified pixels}}{\text{total number of pixels}},$$

- the HMRF-EM algorithm gives more accurate estimates for parameters;
- the HMRF-EM algorithm provides automatic segmentation with much lower MCR.

class	Class 1			Class 2			Class 3			
parameter	μ_1	σ_1	ω_1	μ_2	σ_2	ω_2	μ_3	σ_3	ω_3	K
$NCR = 0.3$	32.0	24.6	0.378	124.9	27.8	0.300	219.2	24.8	0.322	7
$NCR = 0.5$	36.1	35.6	0.377	124.6	46.1	0.304	213.5	38.1	0.320	9
$NCR = 0.7$	42.2	46.3	0.372	123.0	62.7	0.307	207.9	49.3	0.321	10
$NCR = 1.0$	52.1	61.7	0.355	127.7	80.5	0.363	203.6	62.1	0.281	31



Automated Segmentation



- The prior probability for tissue type k is provided by the statistical brain atlas
- No unknown spatial model parameters to be estimated

Bias field Correction

- Since it is a complete approach to segmenting piecewise-constant images, the HMRF-EM framework can be applied to brain MR images.
- However, MR images are often corrupted by a low-frequency spatially varying artifact known as the *bias field*.
- This is mainly caused by the variation of the interaction between the human body and the magnetic field. The bias field can also be due to the differential attenuation of signals and nonlinearities of the sensitivities of the receiver coils, etc.
- Although such an artifact has little impact on visual diagnosis, the performance of most automatic image analysis techniques, especially intensity-based segmentation, can degrade dramatically. Therefore, a robust, automatic, and inexpensive way of correcting for this artifact is required.

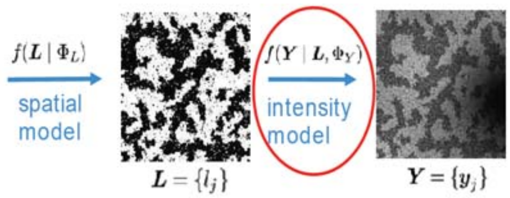
Bias field affects Histogram

Original image without bias field and its histogram

Bias field corrupted image and its histogram

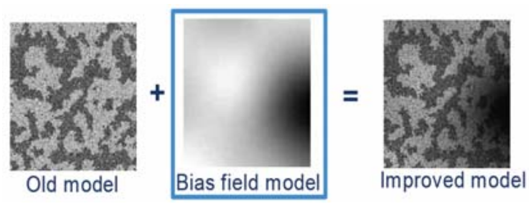
Estimating the bias field amounts to estimating the pdf, given prior assumptions about the bias field and expected pdf

- Include an explicit model for the bias field in the intensity model
- Bias field is usually assumed to be multiplicative
- After logarithmic transformation => bias field becomes additive



Gain field

$$I_i^{measured} = I_i^{original} \cdot \gamma_i + N_i$$



$$y_i^{measured} = y_i^{original} + \beta_i$$

Applying EM to bias field correction

developed by W. M. Wells

- Given the class labels x, it is further assumed that the ideal intensity value at pixel i follows a Gaussian distribution with parameter $\theta(x_i) = (\mu_{x_i} \sigma_{x_i})$

$$p(y_i^* | x_i) = g(y_i^*; \theta(x_i)).$$

- With the bias field bi taken into account, the above distribution can be written in terms of the observed intensity yi as

$$p(y_i | x_i, B) = g(y_i - b_i; \theta(x_i)).$$

- Thus, the intensity distribution is modelled as a Gaussian mixture, given the bias field. It follows that

$$p(y_i | B) = \sum_{j \in \mathcal{L}} \{g(y_i - b_i; \theta(j)) P(j)\}.$$

Applying EM to bias field correction

developed by W. M. Wells

The MAP principle is then employed to obtain the optimal estimate of the bias field, given the observed intensity values:

$$\hat{B} = \arg\max_{B} p(\mathbf{y}|B)p(B),$$

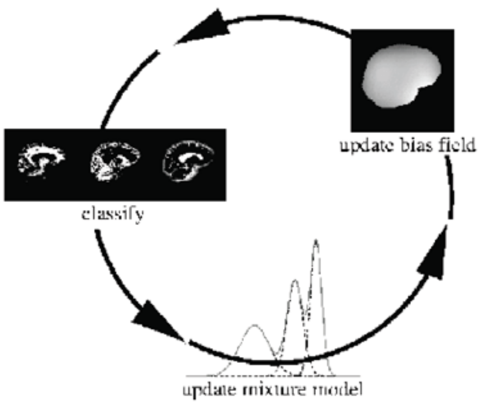
A zero-gradient condition is then used to assess this maximum,

$$Wij = \frac{p(y_i|x_i, \beta)p(x_i)}{p(y_i|\beta)}$$

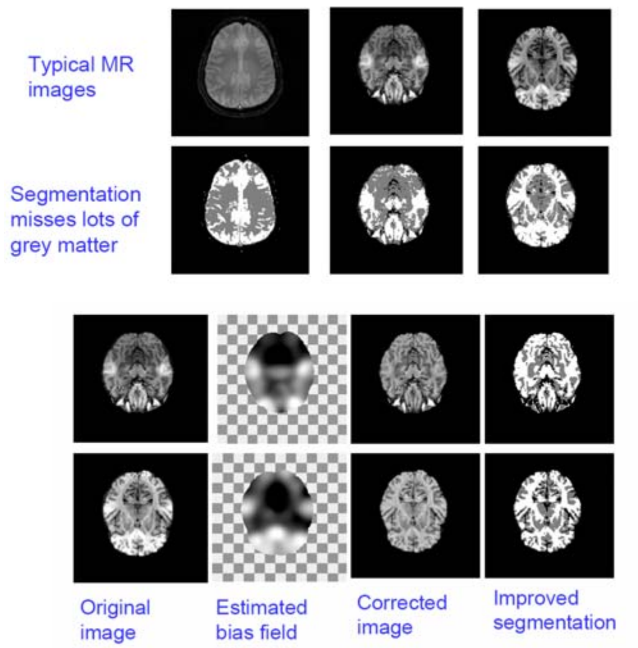
$$bi = \frac{[FR]_i}{[F\psi^{-1}\mathbf{1}]_i}, \text{ with } \mathbf{1} = (1, 1, \cdots, 1)^T,$$

Wij is the posterior probability that pixel i belongs to class j given the bias field estimate.

R is the residual and F is a low pass filter



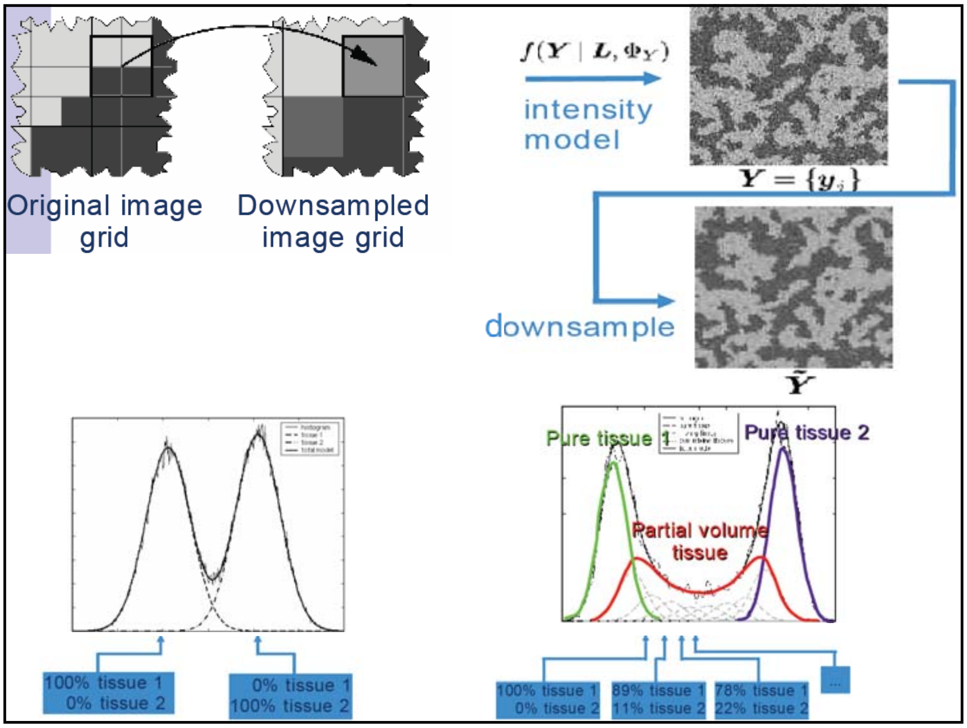
- The E step assumes that the bias field is known and calculates the posterior tissue class probability Wij.
- In the M step, the bias field B is estimated given the estimated Wij in the E step.
- Once the bias field is obtained, the original intensity I* is restored by dividing I by the inverse log of B. Initially, the bias field is assumed to be zero everywhere.



Partial Volume Segmentation

- Assumed so far that each voxel belongs to one single tissue type
- In reality, many voxels in brain MR images are a mixture of several tissue types at the same time
  - -Complex shape of the tissue interfaces in the brain
  - -Limited spatial resolution of MRI

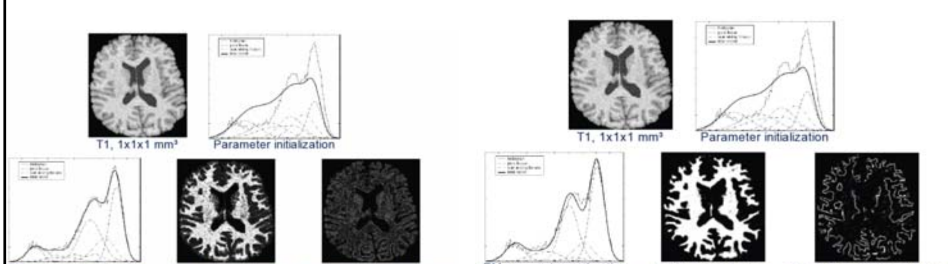
- Consistently misplacing the tissue borders in a 1 mm isotropic brain MRI with a single pixel in each slice results in large volume errors
- *[Niessen et al., 1999]*:
  ~ 30% for white matter
  ~ 40% for gray matter
  ~ 60% for CSF
- Partial volume voxels make lesion segmentation more difficult



- *Expectation step:* find the function

Involves a partial volume image classification: Not only probability for pure tissuesBut also probability for e.g. 22% of tissue1 and 78% of tissue 2

- Mixing combination in a voxel is independent of the mixing combinations in other voxels All non-pure mixing combinations are equally probable
- Markov random field model
- Clustered regions of the same tissue type before downsampling
- Homogeneous regions of pure tissues bordered by partial volume voxels after downsampling



Drawbacks of HMRF-EM

- MRF model tends to minimize the boundary length between tissues.
- This discourages classifications from accurately following the complex shape of the tissue interfaces
- MRF over-smooths the classifications in cases where the intensity information doesn't prevent it