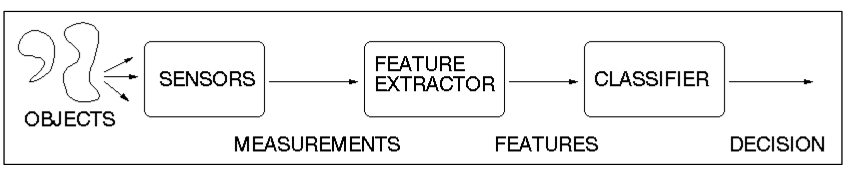
Medical Image Analysis
Lecture 3a

Guido Gerig
UNC Departments of Psychiatry and Computer Science
Course Autumn 1999/ Spring 2000

Repetition: Statistical Pattern Recognition

Pattern Classification System

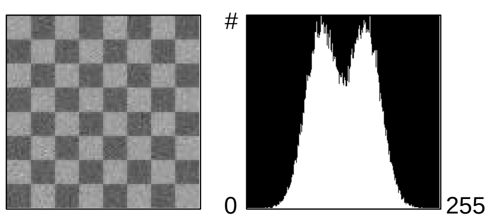


Real World: Example: Vending machines check coins (thickness, weight, color, etc.).
Digital Images: Measurements (patterns) are often pixel values (scalar in b/w images or <u>vector-valued in color images</u>).

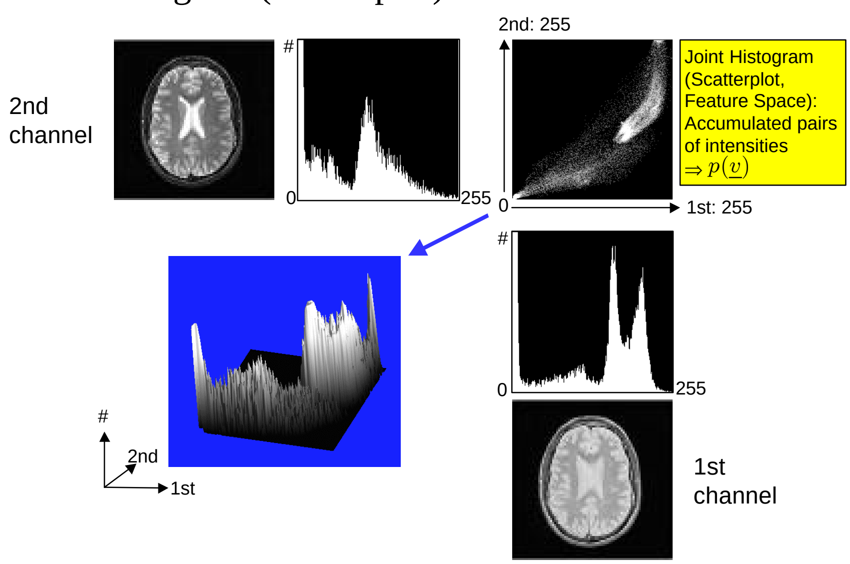
Segmentation by Pixel Classification

Concept:
Scalar or vectorial measurement at each pixel decides about category.
Scalar image data: $p(v)$ given by histogram
Vectorial image data: $p(\underline{v})$ given by joint histogram

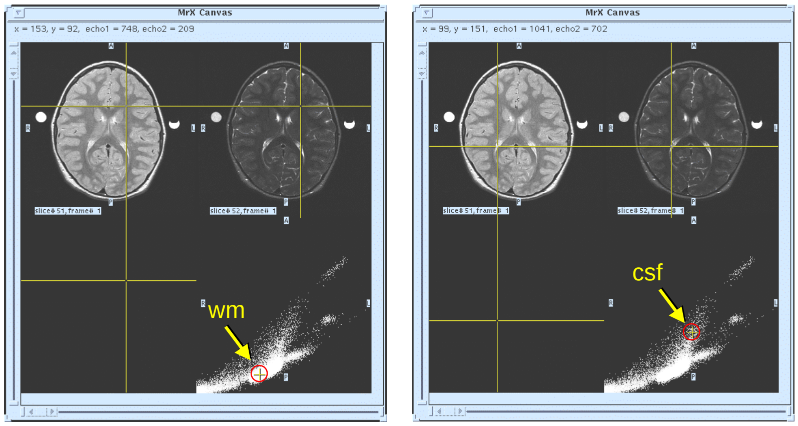
Scalar (black&white image):



Joint Histogram (Scatterplot) of 2 Channels



Joint Histogram (Scatterplot) in *MrX



Upper row: First and second MR echo
Lower Right: Joint histogram (cursor points to gray matter (left) and csf (right))

*MrX: Medical image segmentation software developed by General Electric and Brigham and Womens's Harvard

Pixel-based Classification: Parametric

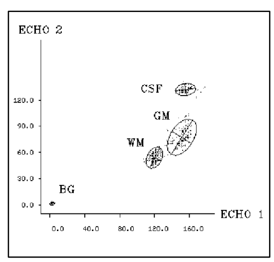
Estimate PDF's $p(\underline{v}|\omega_k)$ by:

Supervised: manual selection of training samples (seeds) for each category and estimation of parameters:

$$\underline{\mu} = E[\underline{v}] \quad \text{COV} = \sum(\underline{v} - \underline{\mu})(\underline{v} - \underline{\mu})^T$$

Assuming **parametric distribution function** for $p(\underline{v}|\omega_k)$ e.g. Gaussian:

$$p(\underline{v}|\omega_k) = G(\underline{\mu}, COV)$$



- **Calculate decisions**:

Posterior probability for class ω_i given measurement $\underline{v}$:

$$P(\omega_i|\underline{v}) = \frac{p(\underline{v}|\omega_i)P(\omega_i)}{p(\underline{v})}$$

Decision for each voxel with measurement $\underline{v}$: Assign class *i* with maximum posterior probability $P(\omega_i|\underline{v})$:

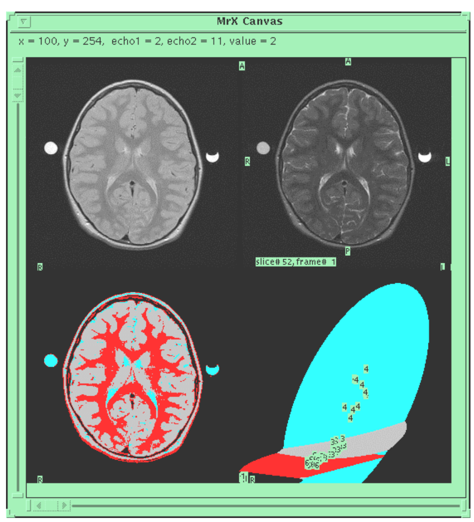
$$\forall \underline{v} : \text{ class } i \rightarrow argmax_{(i)} \; (P(\omega_i|\underline{v}))$$

$$\underline{v} \longrightarrow \begin{cases} P(\omega_1|\underline{v}) \\ P(\omega_2|\underline{v}) \\ \vdots \\ P(\omega_c|\underline{v}) \end{cases} \longrightarrow \text{Max: } P(\omega_i|\underline{v}) \longrightarrow \text{class } i$$

Statistical Pattern Recognition

Parametric Classifier
(Maximum Likelihood)

The feature space window (lower right) shows the training points for csf (4), for gray matter (3), for white matter (6) and the resulting parametric classification map (2D Gaussian distributions).

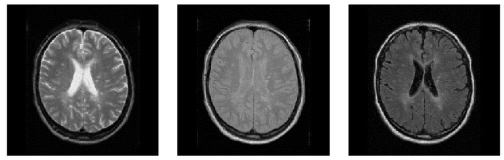


Parametric Classifier
(Maximum Likelihood)

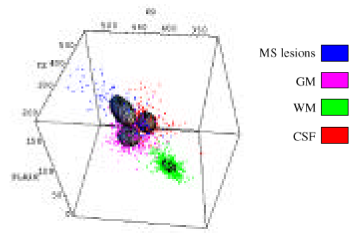
A classification can be done with more than 2 channels.

Example:

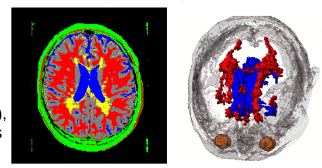
3channel classification using T2-weighted, PD-weighted and FLAIR MR for the segmentation of a multiple sclerosis dataset.



Training samples and probability density functions in 3D feature space



Segmentation Result: WM (red), GM (gray), CSF (blue), Lesions (yellow), skin/fat/muscles (green)



Pixel-based Classification: Non-Parametric

Assumption: Type of distributions unknown.

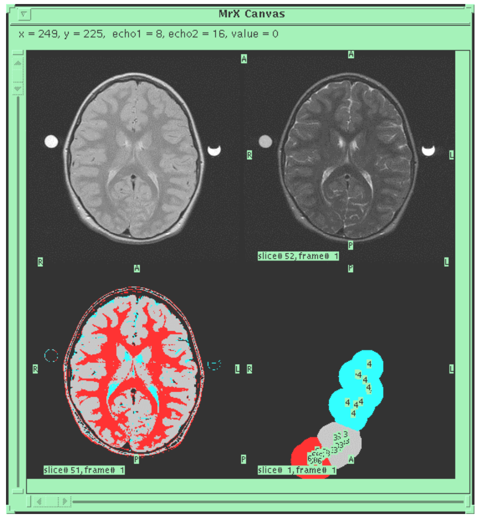
Classification Strategy: Estimate Probabilities using training samples for each class:

A) <u>Parzen Window Classifier</u>: Probability $P(\omega_i|\underline{v})$ is estimated in feature space (joint histogram) by:

- Define control volume around each point $\underline{v}$ in feature space (often circle with radius r if 2 channels)
- Within control volume, count class occurrence $N(\omega_i)$ and total #samples N_{tot}
- Probabilities $P(\omega_i|\underline{v})$ estimated as: $\frac{N(\omega_i)}{N_{tot}}$
- Decide about class *i* with largest probability $P(\omega_i|\underline{v})$

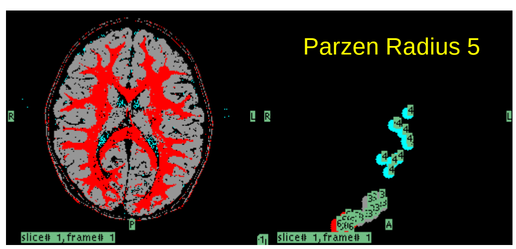
Parzen-Window Classifier with Radius 20

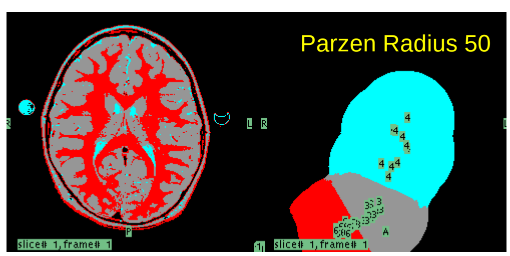
The feature space window (lower right) shows the training points for csf (4), for gray matter (3), for white matter (6) and the resulting classification map.
Locations with no training samples within volume are assigned to rejection class (black).



Parzen Window Classifier: Choice of control volume (radius in 2-channel classification):

If too large: segmentation of all voxels without rejection of outliers.
If too small: incomplete coverage of noisy distributions ⇒ noisy segmentation result.





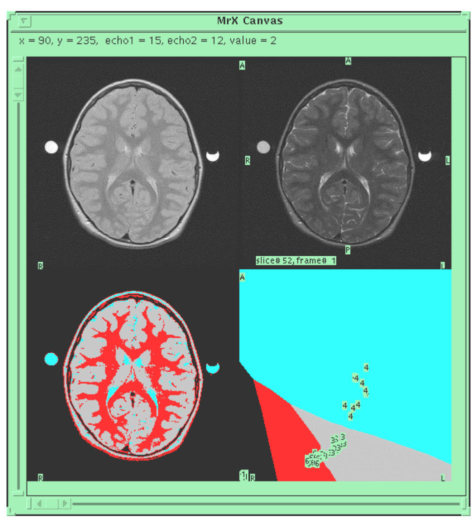
B) <u>KNN: k-nearest-neighbor classifier</u>

Probability $P(\omega_i|\underline{v})$ is estimated in feature space (joint histogram) by a procedure similar to Parzen window:

- expand volume around each point $\underline{v}$ in feature space until it contains *k* training objects
- decide about most probable class among these k training objects (class with most points)
- assign most probable class to feature vector $\underline{v}$

Knn Classifier with 3 neighbors.

The feature space window (lower right) shows the training points for csf (4), for gray matter (3), and for white matter (6) and the resulting classification map.
(here: no rejection class)

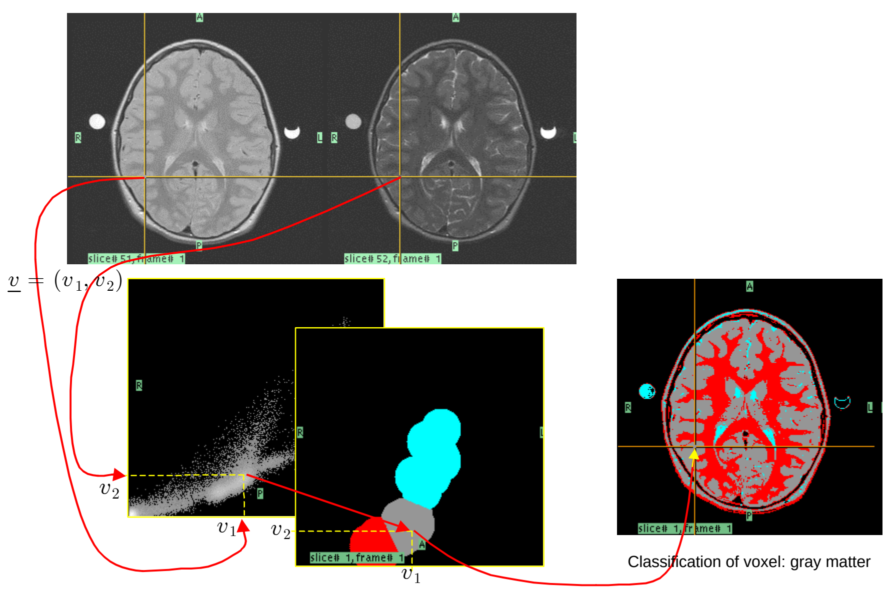


Final Classification of 3D image volume

A 2-D feature space with 2 channels offers an efficient segmentation strategy:

- Classify each point in the 2D feature space (all possible combinations of 2 measurements).
- Take each voxel within the original dataset; the two measurements $\underline{v} = (v_1, v_2)$ define the index to the 2D classification map.
- Read the classification category at this index and assign it to the voxel: $\text{class}\,(\underline{v}) = map(v_1, v_2)$.

Statistical Pattern Recognition



Classification of voxel: gray matter

Critique of pixel-based Statistical Classification

- sensitive to noise (overlap of clusters in feature space)
- each voxel classified independently of spatial neighborhood
- <u>choice of training samples decisive</u>: only limited representation of true probability densities, variability
- intensity gradient (bias field) and partial voluming: erroneous classifications
- textured regions cannot be classified
- no discrimination of classes with similar spectral properties
- classified voxels ≠ object/region segmentation: need for postprocessing