Image segmentation, a historical and mathematical perspective

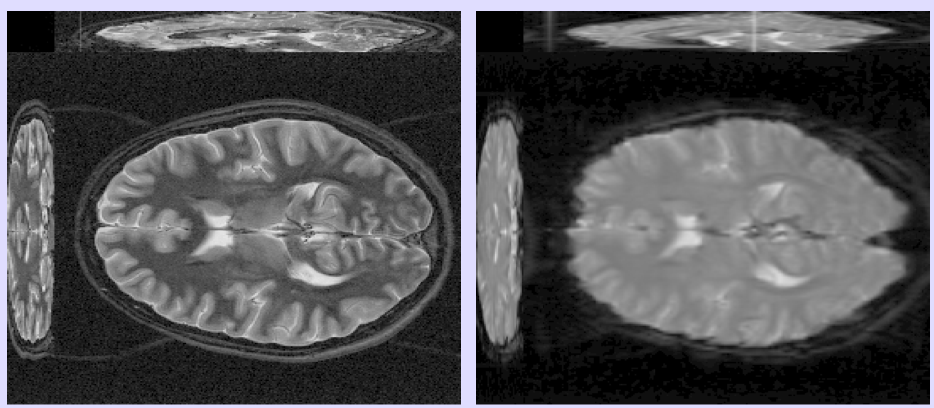
Olivier Faugeras

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

- Image segmentation has a (very) long history: Brice and Fenema (1970), Pavlidis (1972), Rosenfeld and Kak (1976).

- Segmentation is ill-defined:



- A division of the pixels (voxels) of an image into distinct groups (“objects”, “organs”).

- The set of group boundaries.

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

- Initialization: take all pixels as regions.
- For every pair of regions (Ω_i, Ω_j) such that $var(\Omega_i \cup \Omega_j) < \lambda$, merge Ω_i and Ω_j.
- How do we choose the threshold λ?
- No control on the smoothness of the boundaries.
- Solves the constrained optimization problem (Morel-Solimini 1995):

$$\min_{var(\Omega_i) < \lambda} Card(\{\Omega_i\})$$

- λ is a scale parameter.

- Start with the previous algorithm ($\lambda = 0$).

- Given two adjacent regions Ω_i and Ω_j, compute the length of the "weak part" of their common boundary $\partial(\Omega_i, \Omega_j)$ (jump of the intensity across the boundary is less than some threshold).

- Merge Ω_i and Ω_j if the ratio of the length of the weak part of $\partial(\Omega_i, \Omega_j)$ and the length of $\partial(\Omega_i, \Omega_j)$ is larger than a second threshold.

- Solves the optimization problem (Morel-Solimini 1995):

$$E(\partial\Omega) = \mu\, length(\partial\Omega) - \int_{\partial\Omega} \left| \frac{\partial I}{\partial n} \right| d\sigma$$

- Primitive version of the "snakes" technique.

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

The formalization of region growing (1985, 1989)

- A segmentation of an image I_0 is a pair $(\partial\Omega, I)$, where I is some approximation of I_0. I_0 is defined in Ω.

- The energy associated with a segmentation $(\partial\Omega, I)$ is the sum of three terms:

$$E(\partial\Omega, I) = \alpha \int_{\Omega \setminus \partial\Omega} |\nabla I|^2 \, dx + \beta length(\partial\Omega) + \int_{\Omega \setminus \partial\Omega} (I - I_0)^2 \, dx$$

- If I is imposed to be constant within each region

$$E(\partial\Omega, I) = \alpha length(\partial\Omega) + \int_{\Omega \setminus \partial\Omega} (I - I_0)^2 \, dx$$

- There exist minimal segmentations made of a finite set of C^1 curves.
- What is known (Morel-Solimini 1995, Aubert-Kornprobst 2000):
  - There exist minimal segmentations (non-uniqueness).
  - The set of segmentations is small (compact).
  - The boundaries are rectifiable (finite length).
  - The boundaries can be enclosed in a single rectifiable curve.

- **Initialization.** Set $I_0 = g$, piecewise constant on the pixels. $\partial\Omega_0$ is the union of the boundaries of all pixels.

- **Recursive merging.** Merge recursively all pairs of regions whose merging decreases the energy E.

- The scale parameter α can be adjusted.

- The full Mumford-Shah functional can be minimized using the ideas of Γ-convergence (De Giorgi).

- Practically nothing is known for 3D or 3D+t images (see the recent book by Guy David)

Plan

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

- Automatically detect contours of objects.

- A contour pixel x: $\|\nabla I(x)\|$ is high.

- Contrast inversion: function g.

- Energy function:

$$E(c) = \underbrace{\int_0^1 \|c'(q)\|^2 \, dq + \beta \int_0^1 \|c''(q)\|^2 \, dq}_{\text{Internal energy}} + \lambda \underbrace{\int_0^1 g^2(\|\nabla I(c(q))\|) \, dq}_{\text{External energy}}$$

- This energy is minimized using the associated Euler-Lagrange equations.

- $E(c)$ is not intrinsic.

- Impossible to detect more than one (changes in topology) convex object.

- Numerical problems occur when solving

$$\begin{cases} \dfrac{\partial E}{\partial t}(t,q) = -\nabla E(t,q) \\ c(0,q) = c_0(q) \end{cases}$$

- Define the energy (Riemaniann metric)

$$E_2(c) = \int_0^1 g(\|\nabla I(c(q))\|)\|c'(q)\|\, dq$$

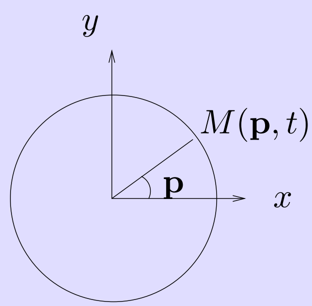
- Intrinsic criterion.

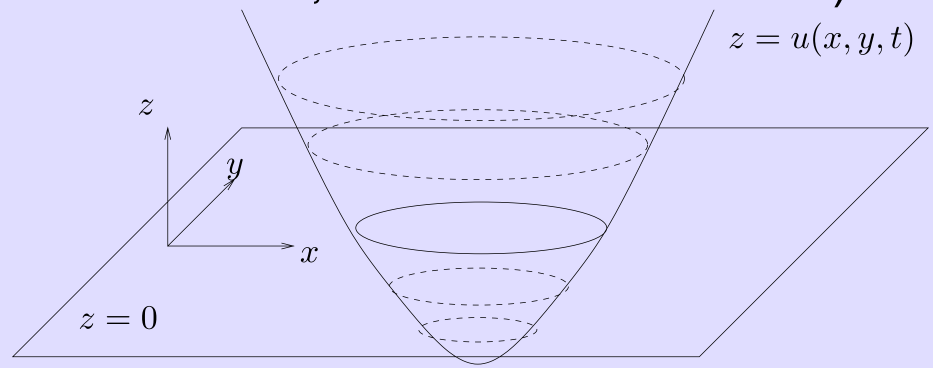
- Aubert and Blanc-Ferraud (1998) showed that E is equivalent to E_2.

- Euler-Lagrange and gradient descent:

$$\frac{\partial E}{\partial t} = (\kappa g - \nabla g \cdot n)n$$

- Basic idea (Dervieux-Thomasset 1979-80, Osher-Sethian 1988):





$$u(M(\mathbf{p},t),t) = 0$$

- Partial Differential Equation:

$$\frac{\partial u}{\partial t} = g(\|\nabla I\|)\text{div}\left(\frac{\nabla u}{\|\nabla u\|}\right)\|\nabla u\| + \nabla g \cdot \nabla u + \text{boundary conditions}$$

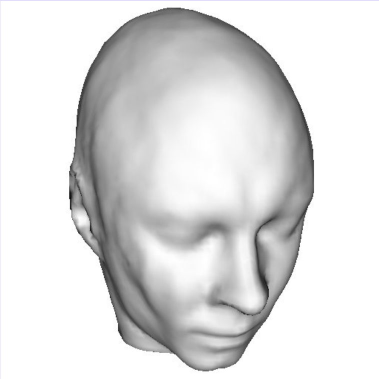
- There is a unique viscosity solution (Crandall-Lions 1982) to the previous equation (Caselles-Catte-Coll-Dibos 1993).

- $u(t,x)$ asymptotically fits the desired contour.

- Proceed in four steps

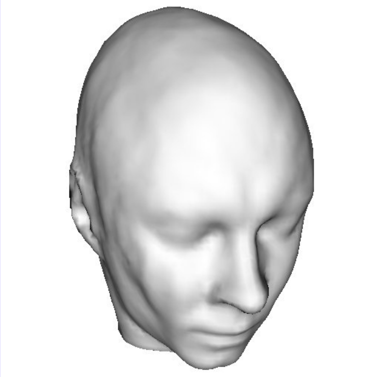
1. Segmentation of the skin surface

The geodesic snake shrinks until it reaches in the volume image high intensities corresponding to the skin.

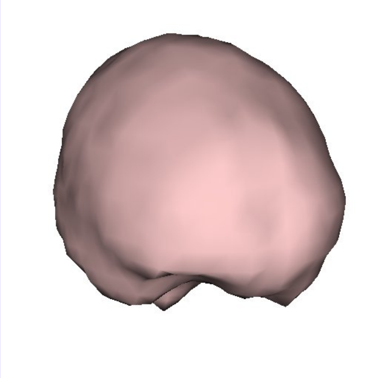


- Proceed in four steps

  1. **Segmentation of the skin surface**
     The geodesic snake shrinks until it reaches in the volume image high intensities corresponding to the skin.



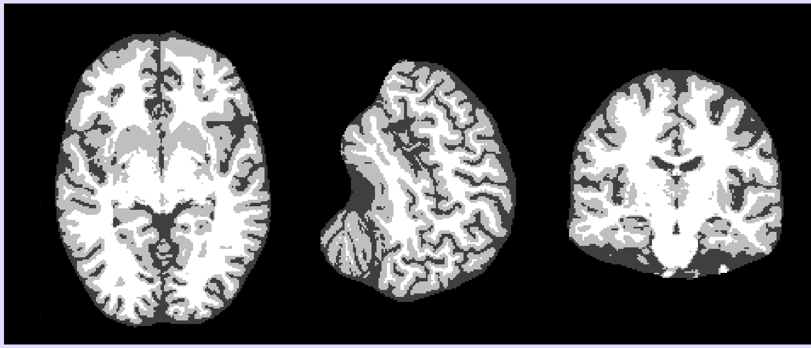
  2. **Segmentation of the brain outlines**
     A geodesic snake at the center of the brain inflates until it reaches in the volume image low intensities corresponding to the CSF or to the skull.



- Proceed in four steps

3. Classification of brain tissues into three classes

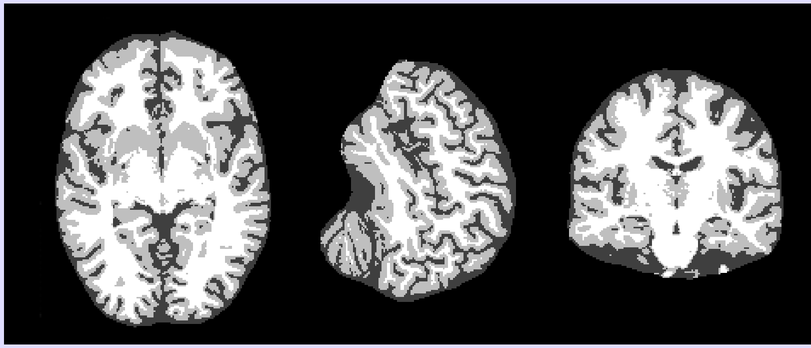
Separation of the grey matter, the white matter and the CSF + correction of the nonunifomities in the MR image.



- Proceed in four steps

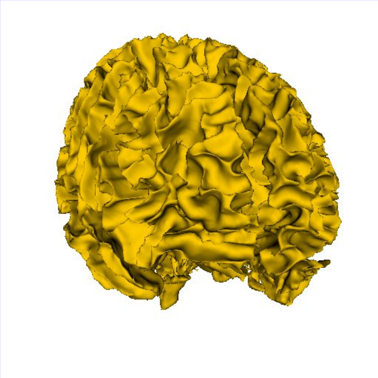
3. Classification of brain tissues into three classes

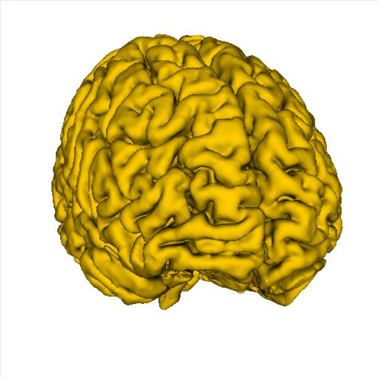
Separation of the grey matter, the white matter and the CSF + correction of the nonunifomities in the MR image.



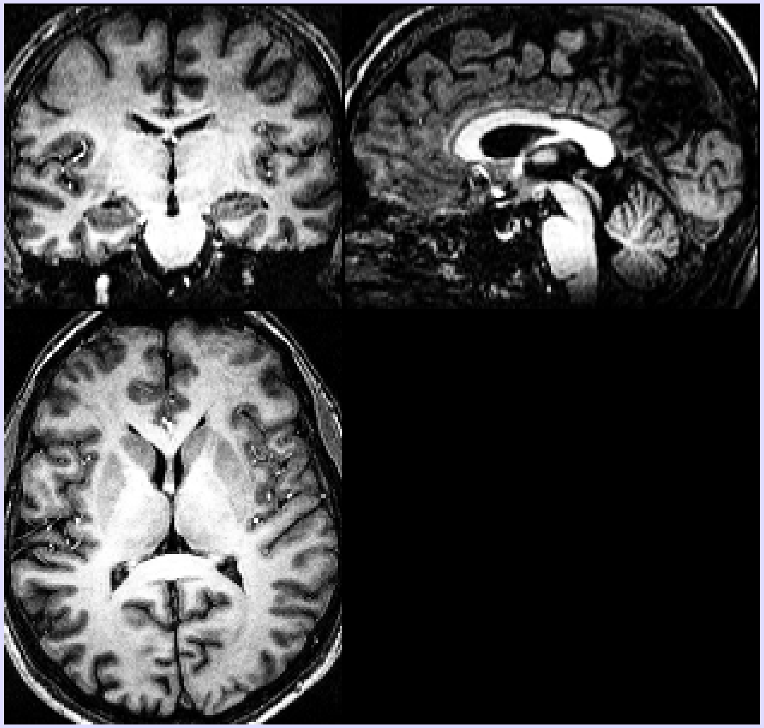
4. Extraction of the internal and external surfaces of the cortex

Two surfaces approximate the results of the classification while guaranteeing a correct geometry (J.Prince et al. 2003)

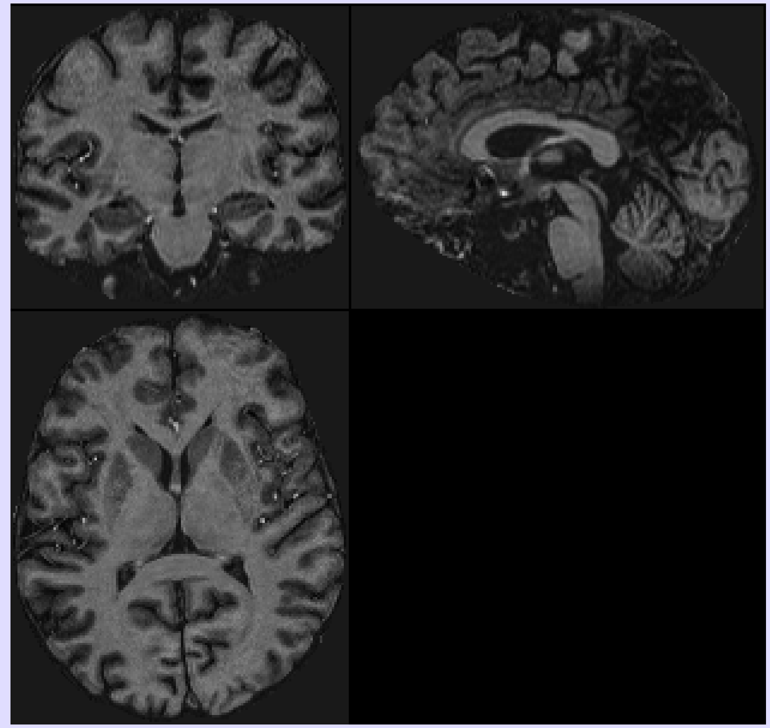




Results: correction of the inhomogeneities

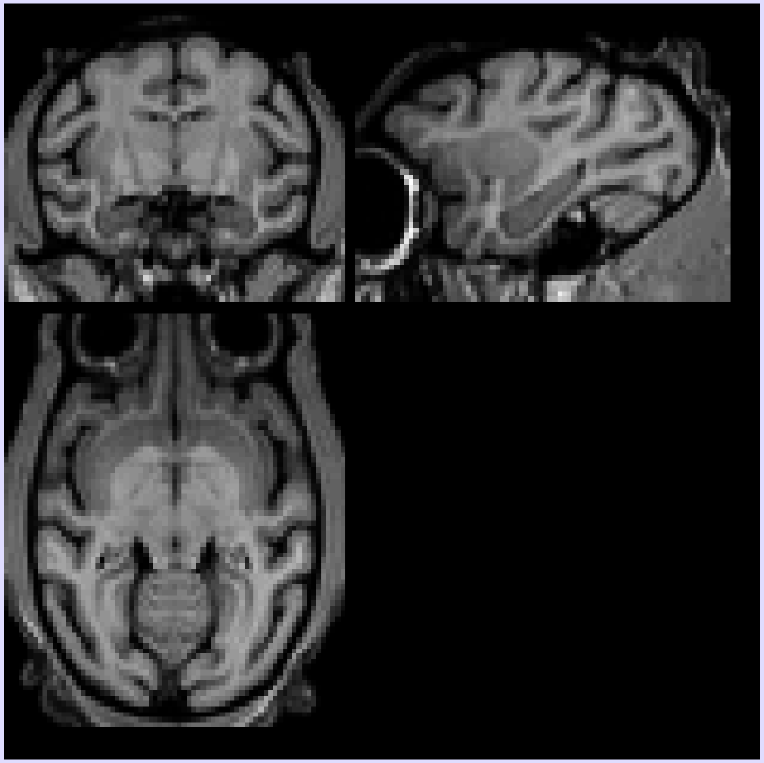


Initial image

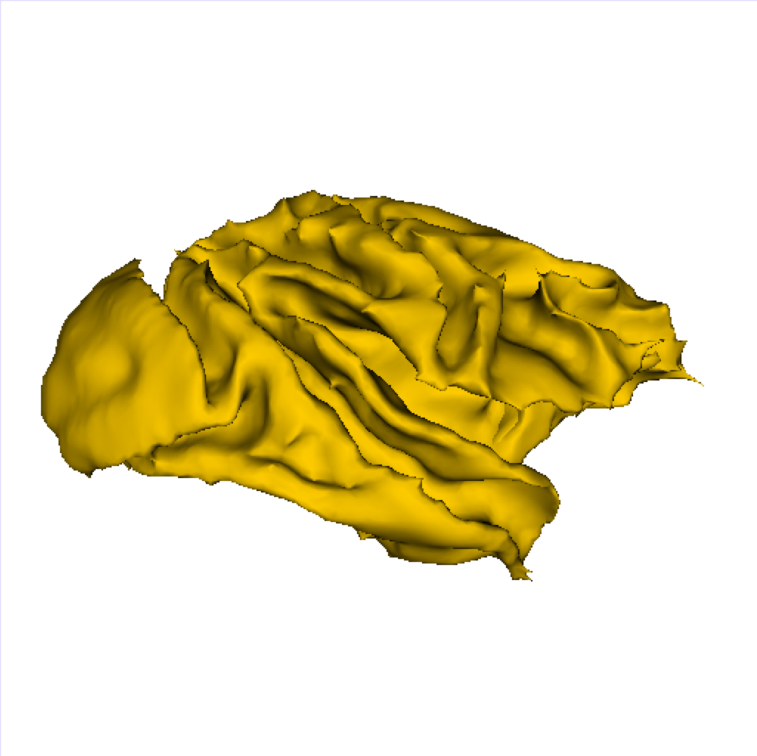


Corrected image

The same techniques can be applied **to monkey data**, thereby allowing to verify their pertinence (e.g., Guy Orban's lab. in Leuven)

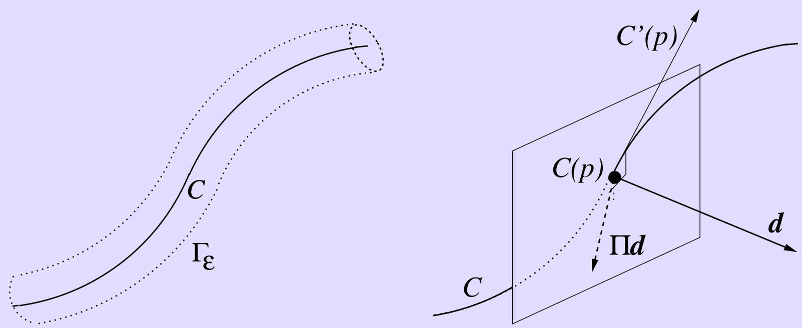


MR Image



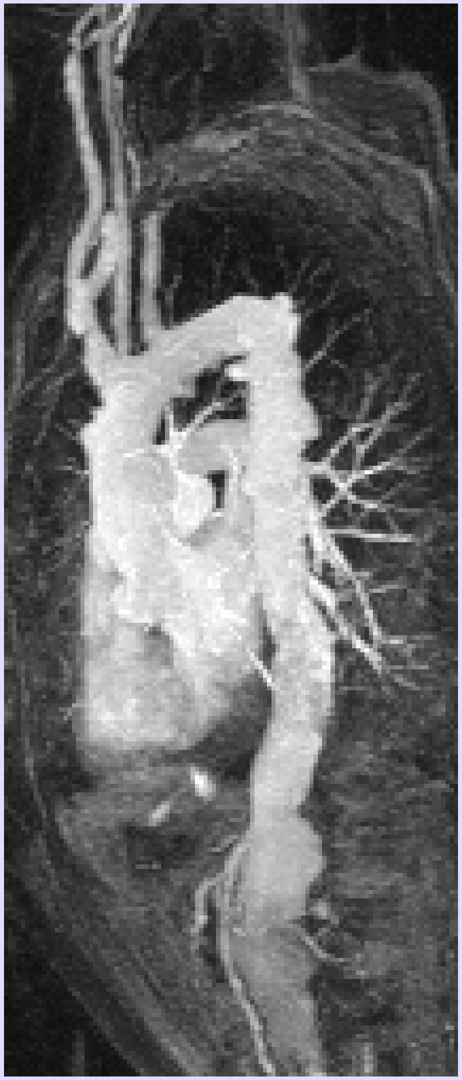
Left hemisphere of the cortex

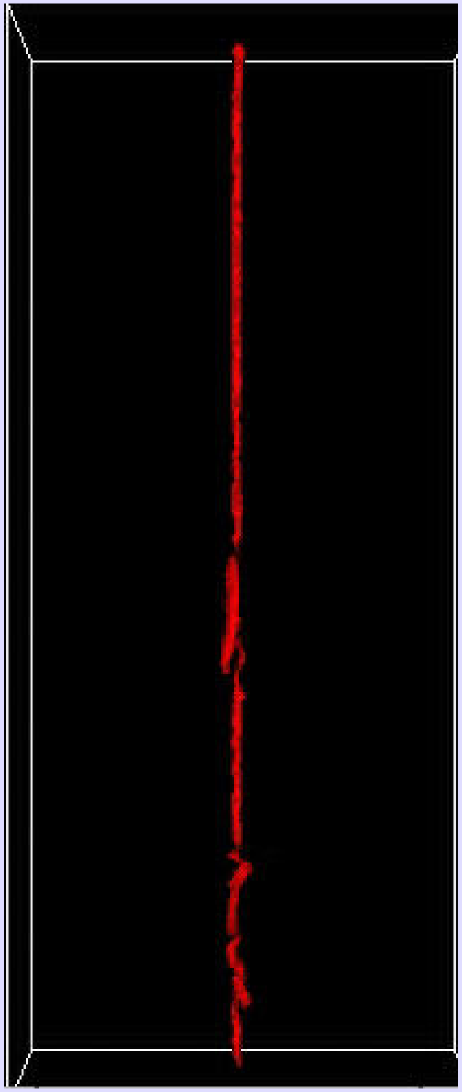
- Geodesic snakes are co-dimension 1.
- **Goal**: detect and characterize the shape and size of blood vessels in MRA images.
- **Methodology**: generalization of the previous approach to curves in 3D space through the idea of ε-level sets.

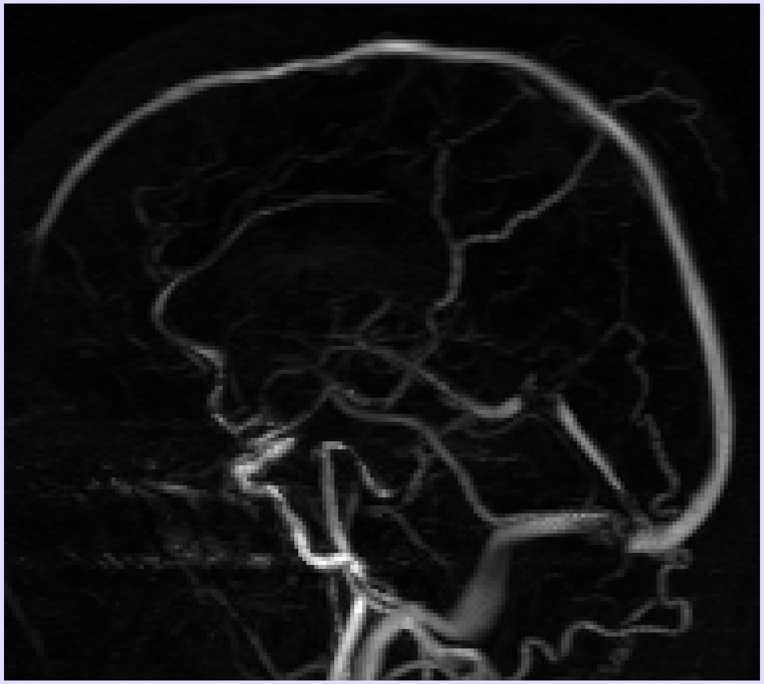


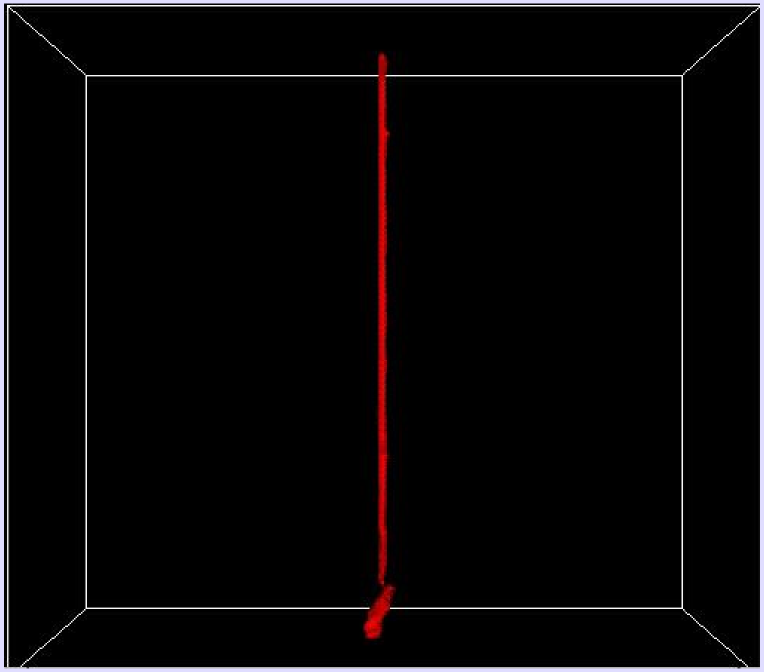
- It is equivalent to smoothing with the **smallest** principal curvature rather than with the **mean** curvature.

Geodesic snakes *Generalization to 3D curves: aorta data (courtesy Siemens)*









Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

INRIA
Odyssee Lab

Active regions

- The contour approach is limited to the contours!

- Let Ω be a region, define:

$$J(\Omega) = \int_{\Omega} f(x, \Omega)\, dx$$

- Examples of functions f:
  1. $f(x, \Omega) = (I(x) - \mu_{\Omega})^2$ $\quad \mu_{\Omega}$ mean intensity in Ω.
  2. $f(x, \Omega) = \rho(\sigma_{\Omega})$ $\quad \sigma^2_{\Omega}$ intensity variance in Ω.
  3. $f(x, \Omega) = -\log h_{\Omega}(I(x))$ $\quad h_{\Omega}$ intensity histogram in Ω.

$$E(R) = J(\Omega) + J(\Omega^c) + \lambda\, length(\partial\Omega)$$

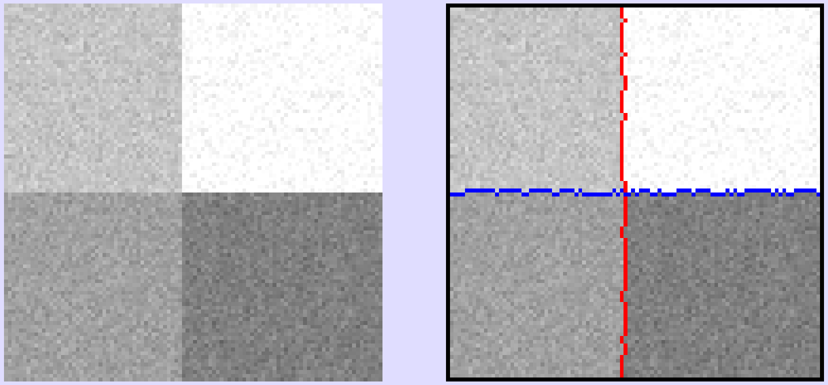
- **Problem**: How do we compute the derivative of E with respect to the boundaries shape.
- **Answer**: Use the tools of shape derivatives invented by, e.g. Jacques Solomon Hadamard.
- More recent work by Delfour and Zolesio 2001
- See also the field of Shape Optimization.

An example: log likelihood energy (Schnörr 04)

- Histogram estimation by Parzen windowing: non parametric case
- Shape derivative:

$$\frac{1}{\mid \Omega \mid}\int_{\Omega}\frac{g_{\sigma}(\boldsymbol{I}(\boldsymbol{x})-\boldsymbol{I}(\boldsymbol{y}))}{p(\boldsymbol{I}(\boldsymbol{x}),\Omega)}\,d\boldsymbol{x}-\frac{1}{\mid \Omega^c \mid}\int_{\Omega^c}\frac{g_{\sigma}(\boldsymbol{I}(\boldsymbol{x})-\boldsymbol{I}(\boldsymbol{y}))}{p(\boldsymbol{I}(\boldsymbol{x}),\Omega^c)}\,d\boldsymbol{x}-\log\left(\frac{p(\boldsymbol{I}(\boldsymbol{y}),\Omega)}{p(\boldsymbol{I}(\boldsymbol{y}),\Omega^c)}\right)$$

- Implementation by level-sets (Vese and Chan 2001, Rousson and Deriche 2002): N level sets can find up to 2^N regions:



4 regions segmentation

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

- Color is multidimensional: use parametric representations.

- Idea based on the classical structure tensor::

$$J_\sigma = G_\sigma * (\nabla I \nabla I^\top) = \begin{pmatrix} G_\sigma * I_x^2 & G_\sigma * I_x I_y \\ G_\sigma * I_x I_y & G_\sigma * I_y^2 \end{pmatrix}$$

  where G_σ is a Gaussian kernel with standard deviation σ.

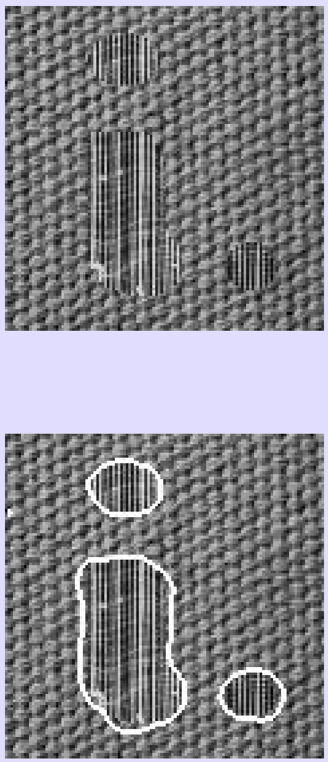
- Properties:
  - only 3 feature channels at a fixed scale (reduced number compared to a set of Gabor filters),
  - include orientation information,

- For color images is: $J_\sigma = G_\sigma * \left(\sum_{i=1}^{3} \nabla I_i \nabla I_i^\top \right)$.

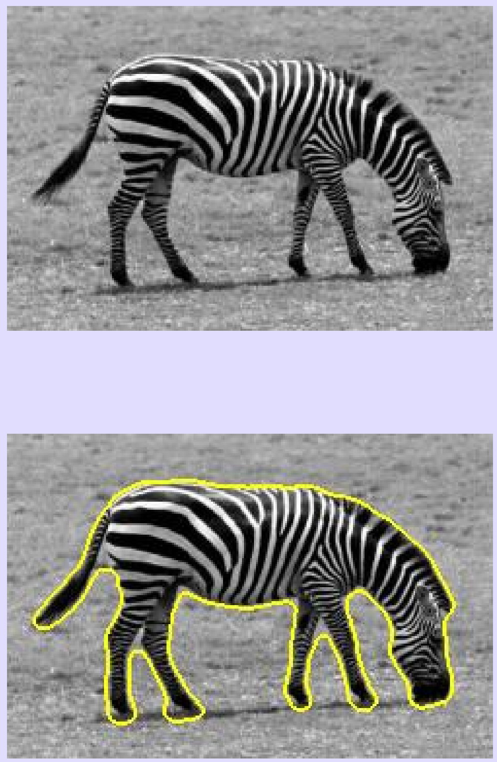
Color and texture *Example: Intensity and Texture (Rousson, Deriche et al. 2002-today)*

- Results on gray images:

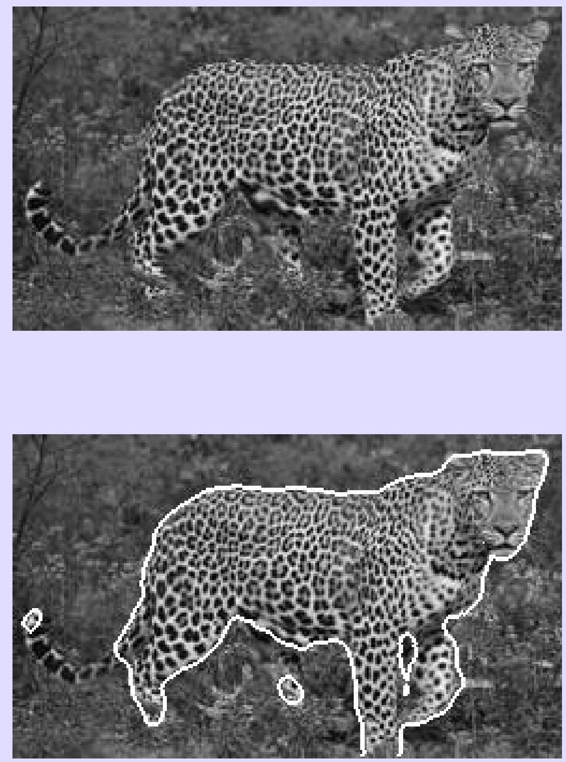
$$u(t=0) = (I, |I_x|, |I_y|, \pm\sqrt{\pm I_x I_y})$$



init 1 init 2



init 1 init 2

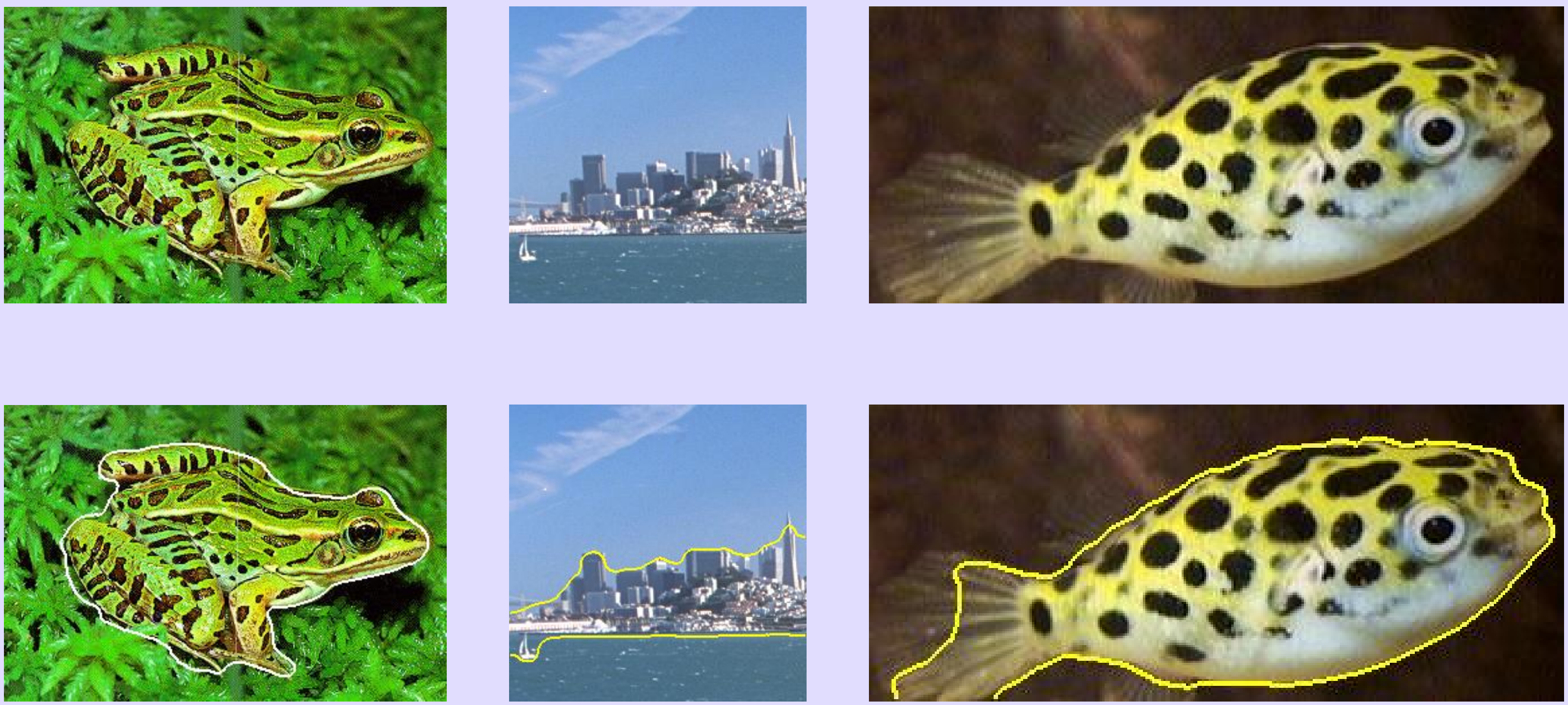


init 1 init 2

Example: Color and Texture (Rousson, Deriche et al. 2002-today)

- Results on color images:

$$u(t=0) = (I_l, I_a, I_b, \sqrt{J_\sigma^{(1,1)}}, \sqrt{J_\sigma^{(2,2)}}, \pm 2\sqrt{\pm J_\sigma^{(1,2)}})$$



- Optic flow constraint: $I_x u + I_y v + I_z = 0$

- Lucas and Kanade: $E(u,v) = \frac{1}{2}\int_{B_\sigma(x_0,y_0)} (I_x u + I_y v + I_z)^2 dxdy$

- A minimum (u,v) of E satisfies $\partial_u E = 0$ and $\partial_v E = 0$, leading to the linear system:

$$\begin{pmatrix} G_\sigma * I_x^2 & G_\sigma * I_x I_y \\ G_\sigma * I_x I_y & G_\sigma * I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -G_\sigma * I_x I_z \\ -G_\sigma * I_y I_z \end{pmatrix}.$$

  Instead of the sharp window B_σ, we use a convolution with a Gaussian kernel G_σ.

- Any other method can be used for OF extraction.

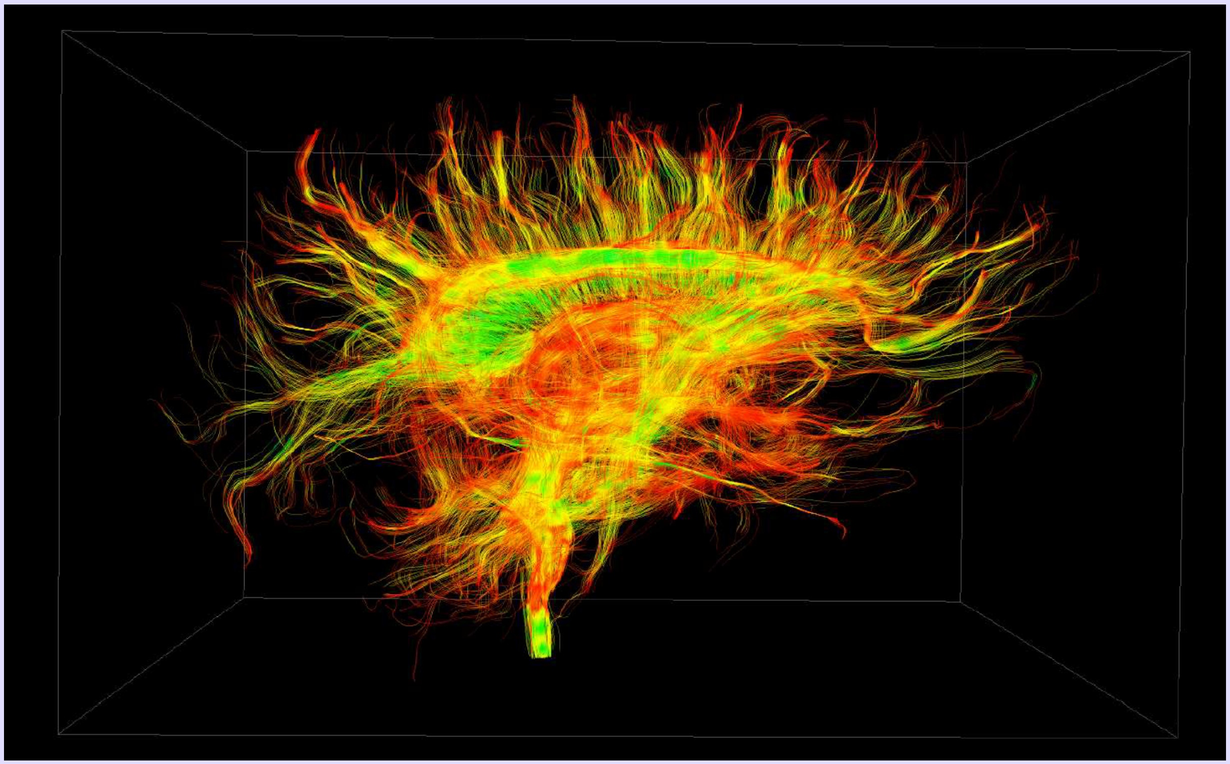
Motion *Example: Color, Motion and Texture (Paragios, Rousson, Deriche et al. 2002-today)*

Tracking of 3 players in the *soccer* sequence $(180 \times 130 \times 40)$.

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

Diffusion Tensor Imagery: Understanding the structure of neural fibers.

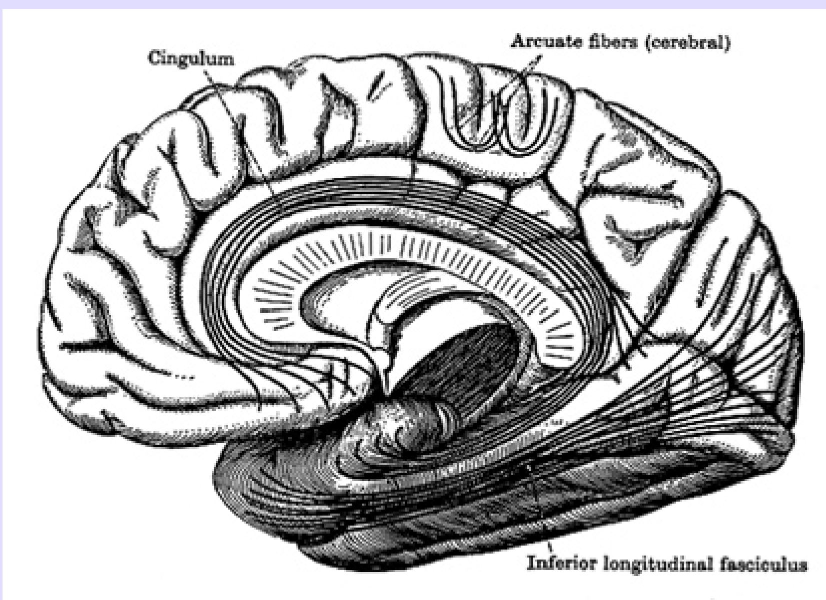


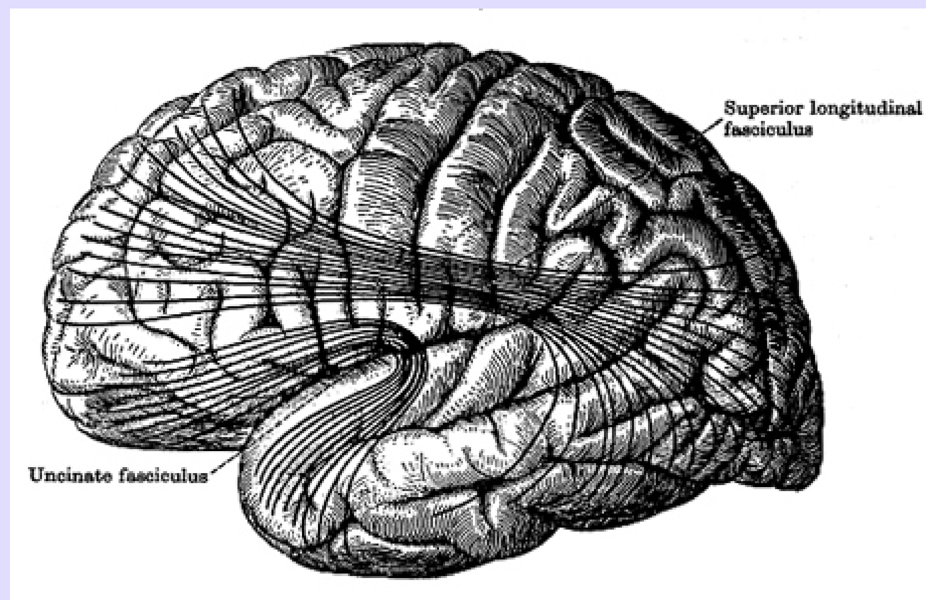
How to segment these fibers ?

- **Diffusion tensor imagery** : a MR modality that measures the motion of water molecules in tissues.

⇒ The water molecules move more easily along the fibers.

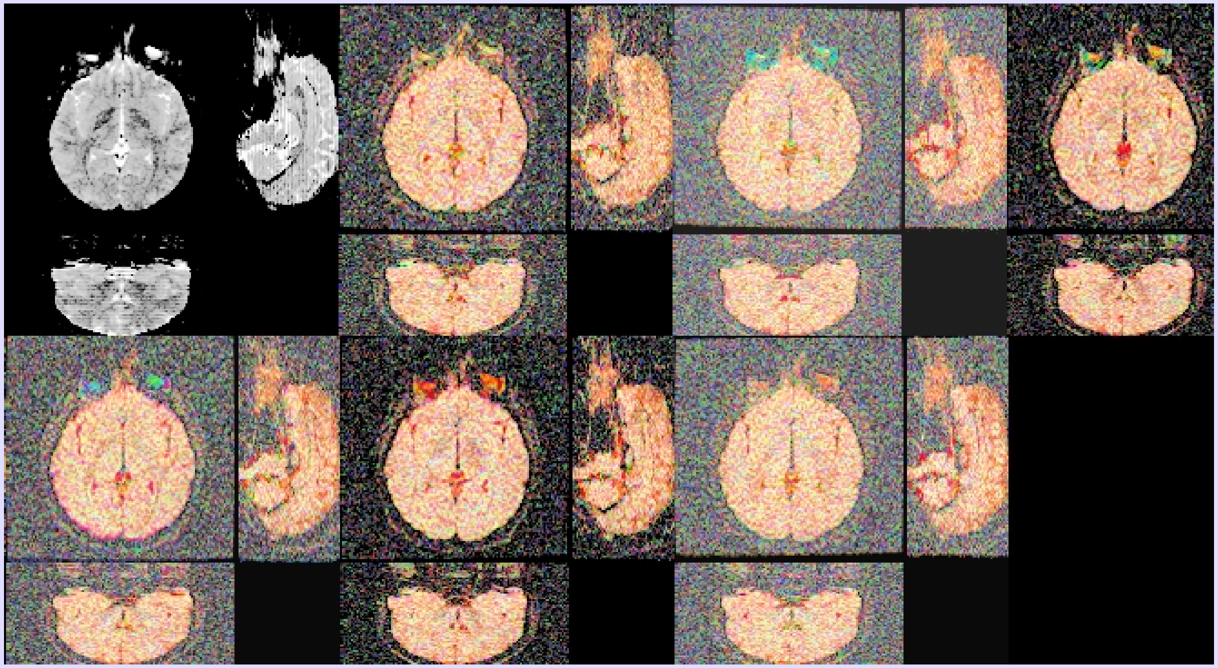
⇒ dtMRI allows us to measure the spatial structure of these fiber bundles





- MRI allows, under some circumstances, to measure **the amount of diffusion of water molecules** inside the tissues.

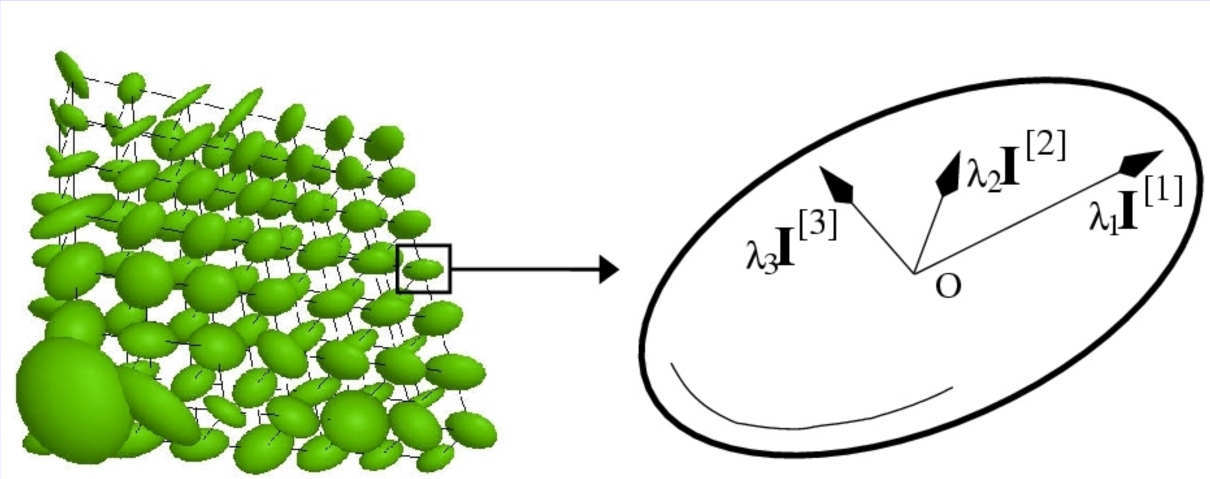
- We acquire **a large number of volume images** of the brain using different orientations and intensities of the magnetic field.



(An example with 7 images)

MR images of the diffusion tensor : Principle (2)

- From these “raw” images, a volume of **Diffusion Tensors** can be estimated.
- These tensors characterize the amount of **diffusion of the water molecules** in the tissues.
- We can represent them with **ellipsoids** :



- **Key observation**: the set of positive definite matrixes can be endowed with a structure of **Riemannian** space derived from the **Fisher information matrix**

- The *information* geodesic distance $\mathcal{D}$ was shown to be (S.T. Jensen 1976 cited in Atkinson and Mitchell 1981):

$$\mathcal{D}(\Sigma_1, \Sigma_2) = \frac{1}{2}\mathrm{tr}(\log^2(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}))$$

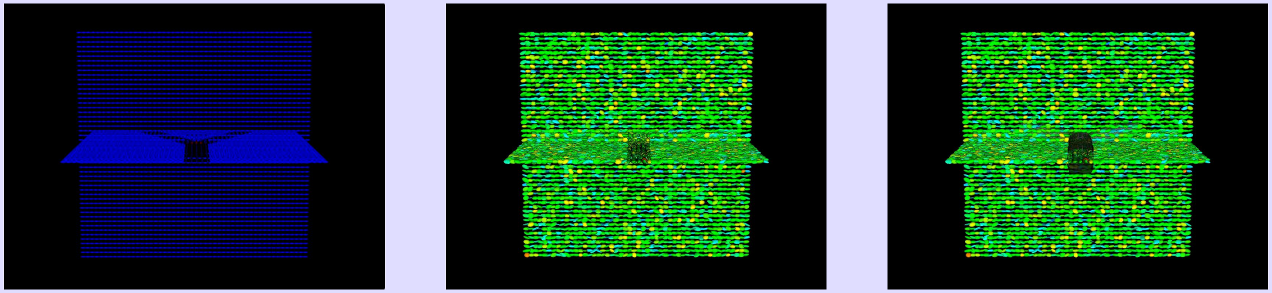
- Expressions can be derived for the geodesics, distance, mean, covariance matrix, Riemann-Christoffel and Ricci tensors, Scalar curvature.

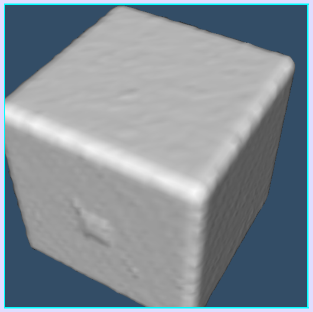
- These ideas are actively explored in (Lenglet, Rousson et al. 2004, Pennec et al. 2004, Joshi et al. 2004).

- Region-based segmentation of DTI may help in analyzing white matter structures.
- The active region formalism can be used in the framework of the Riemannian space of positive definite matrixes.

Experiments on synthetic data (Lenglet, Rousson et al. 2004)

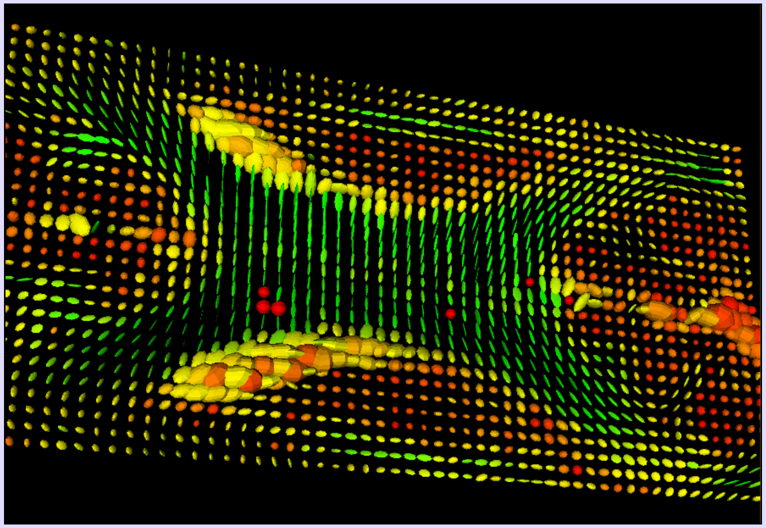
- Fibers bundle junction:

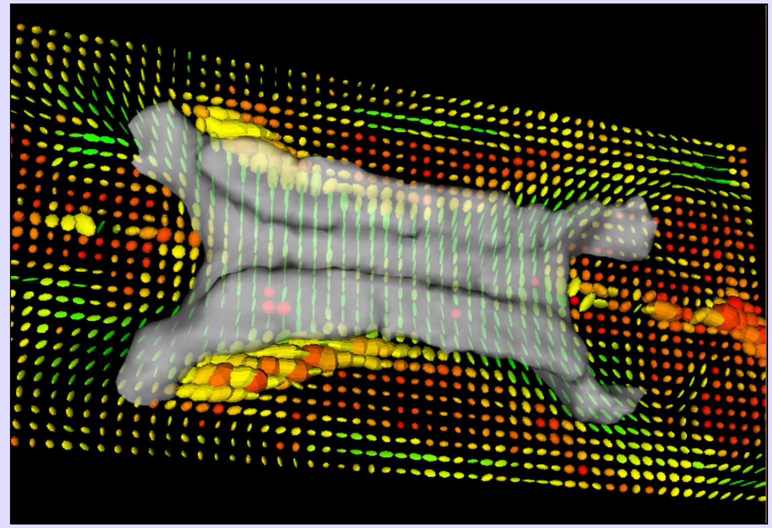


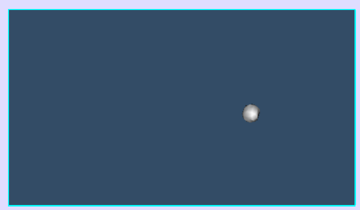


Real data (Lenglet, Rousson et al. 2004)

- Corpus callosum:



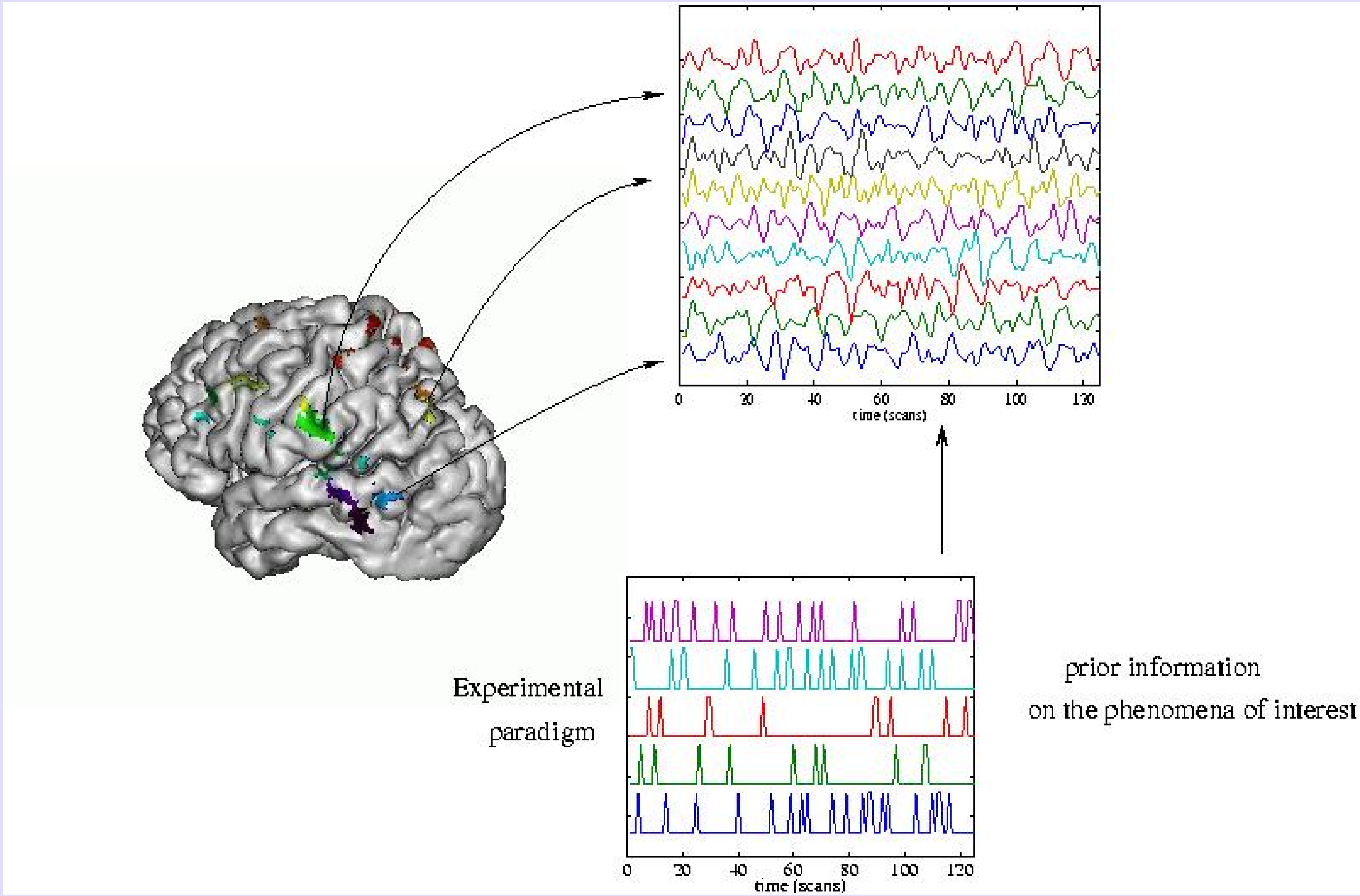




Outline

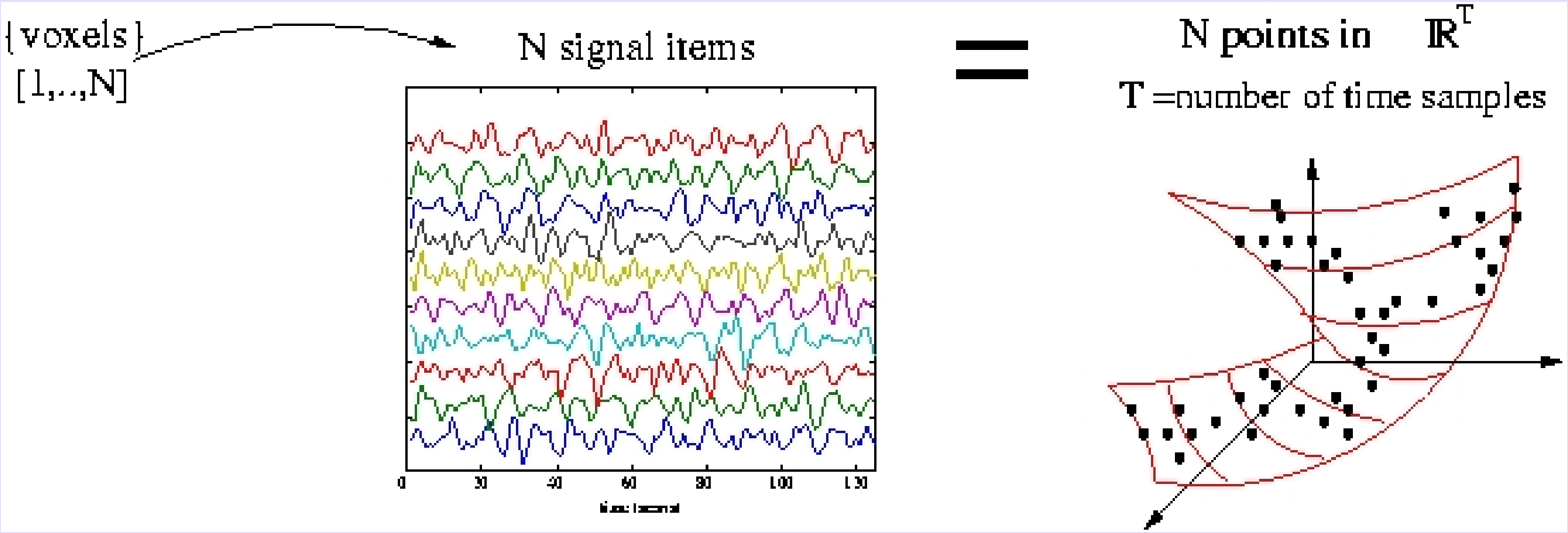
- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

Modeling fMRI datasets



fMRI time courses reflect

task-related activity + physiological confounds + measurement errors + spontaneous activity ...

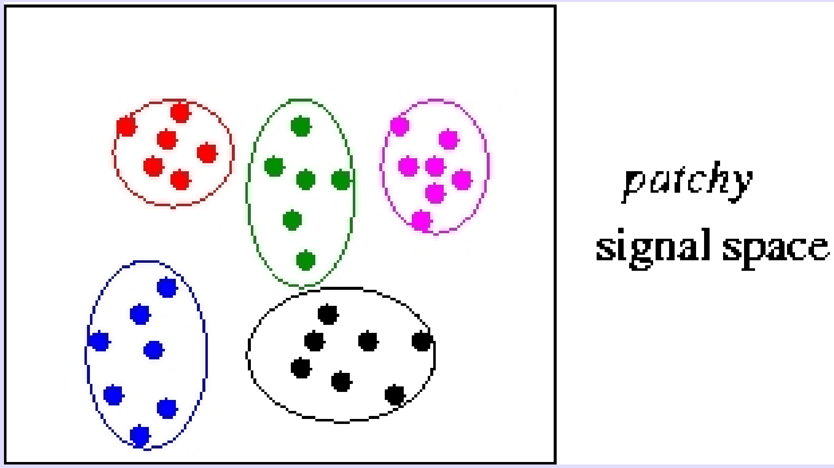


Find **reduced representations** of the data that retain its *essential features*. i.e. account for (dis-)similarities of the temporal patterns across the dataset.

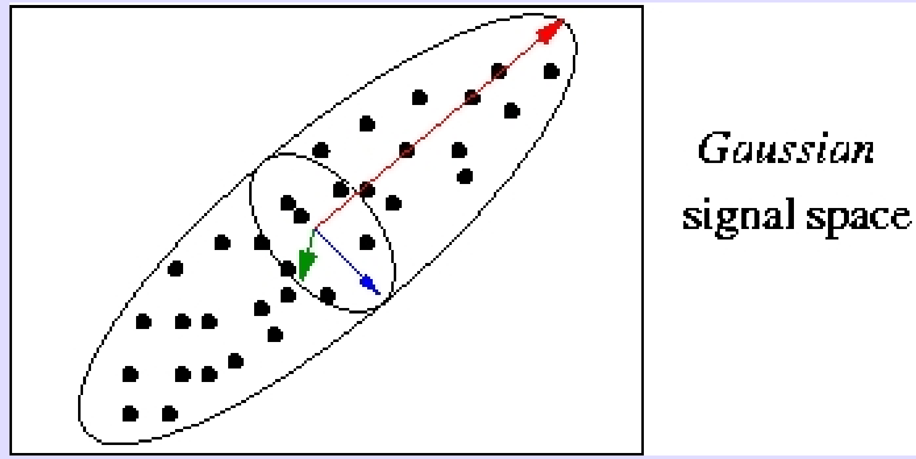
Question : How to model the signal space globally?

Some approaches

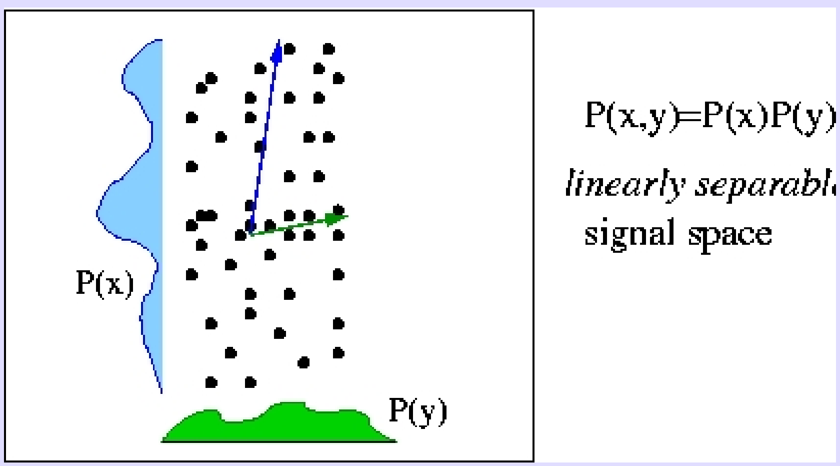
Clustering



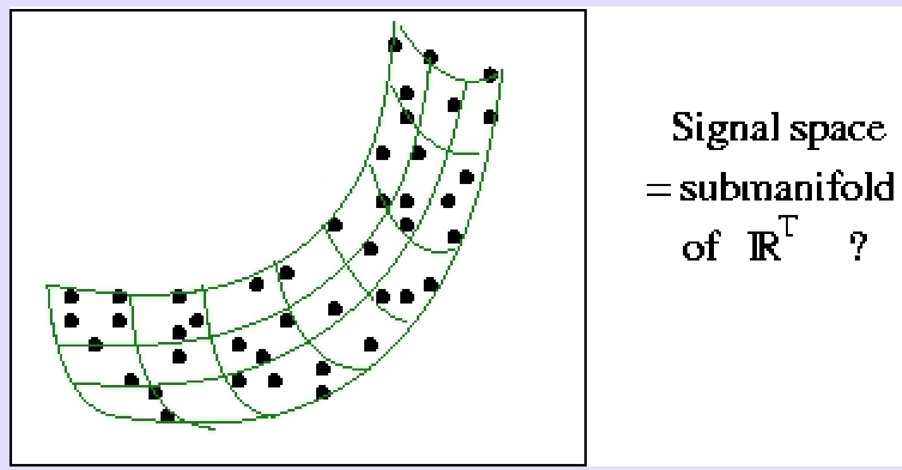
PCA



ICA



LE



Our hypotheses:

- The signal lives in a d-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^T$
  d is not known a priori.
- The different dimensions of $\mathcal{M}$ may be interpreted as the main effects (physiology, acquisition, activation, connectivity).

The Laplacian embedding technique (Belkin and Niyogi 2003) yields an estimate of d and a parameterization of $\mathcal{M}$, i.e. a data-driven characterization of the signal space.

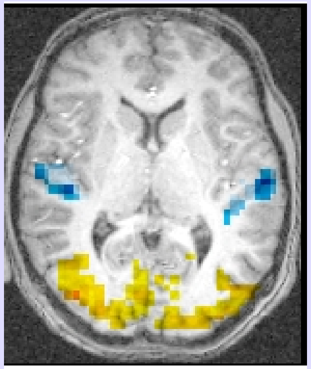
- It is mathematically equivalent to the **graph-cuts** technique (Kolmogorov and Zabih 2002, Shi and Malik).
- Its implementation is closely related to solving the heat equation on the unknown manifold (Laplace-Beltrami operator).

- 0ne-session event-related experiment
- Localizes the main brain functions: primary visual areas, primary auditory areas, reading, computation, motor (left and right hand clicks).
- Standard preprocessing: slice timing, band-pass filtering, spatial normalization.

Exploratory analysis with Laplacian embedding approach.

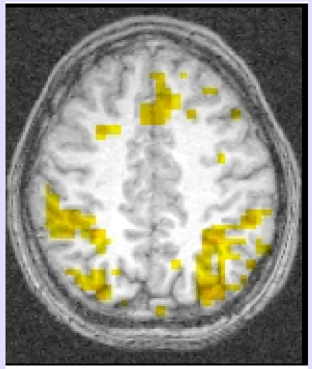
Localizer experiment

LE 1
z=8mm



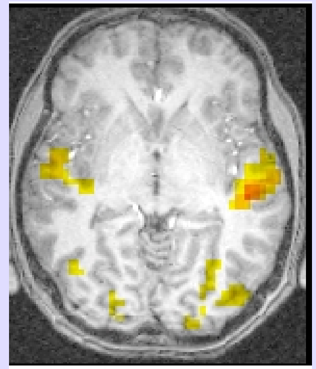
visu-auditory

LE 2
z=44mm



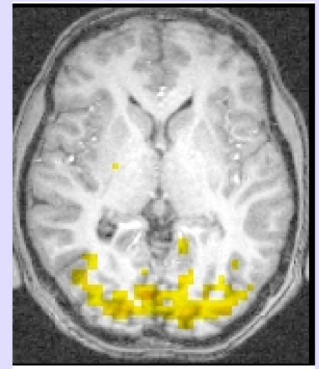
computation

LE 3
z=0mm



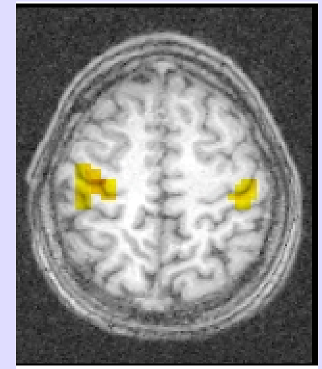
understanding

LE 4
z=4mm



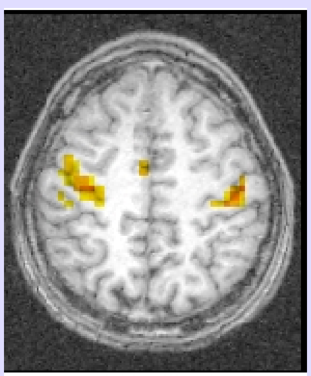
prim. visu

LE 5
z=56mm



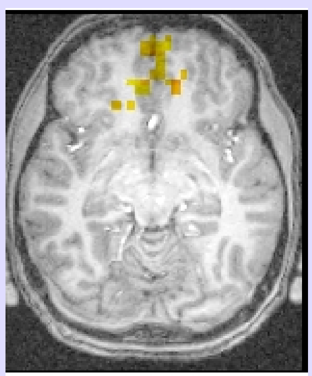
motor

LE 6
z=52mm



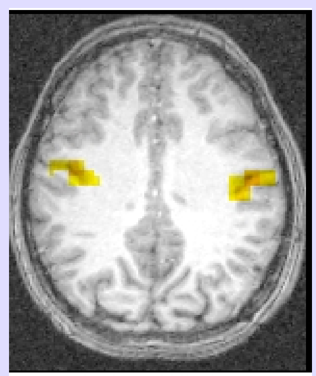
motor

LE 7
z=-8mm



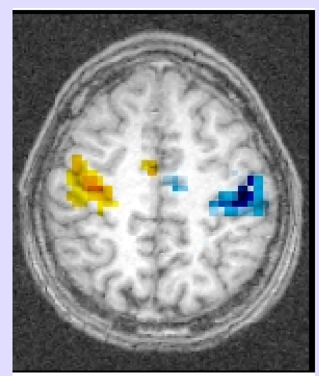
?

LE 8
z=36mm



?

LE 9
z=52mm

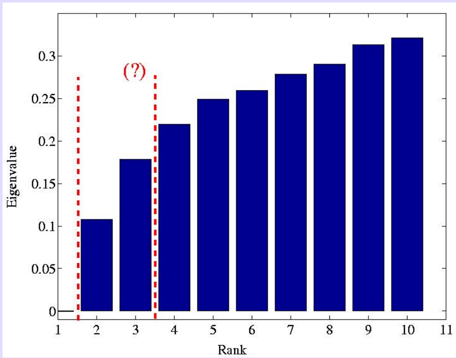


L-R motor

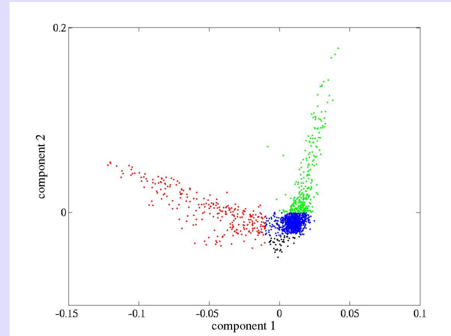
Example2: Supervised

- 1 session of real data [Vanduffel-Fize-etal:01]
- Study of monkey vision: passive observation of static/moving textures
- $N = 12320$ voxels, $T = 120$ scans
- After estimation of voxel-based hemodynamic responses from multi-session data, classification of the resulting *hrf's*.

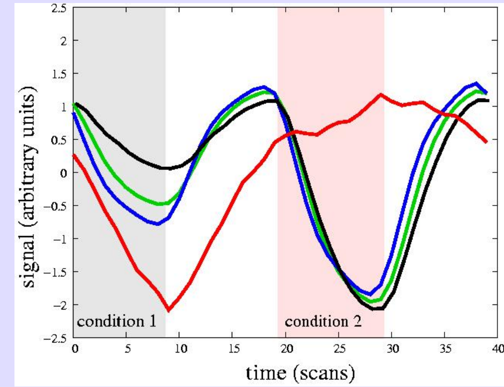
Supervised analysis: Classification of hemodynamic responses (Bertrand Thirion 2004)



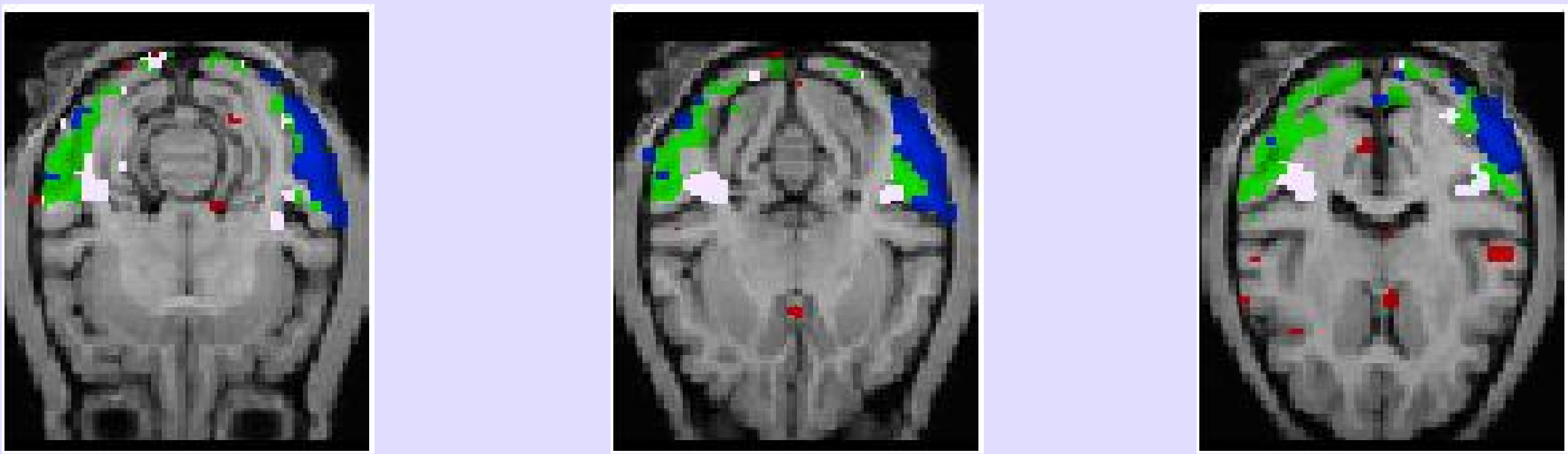
Laplacian eigenvalues



Laplacian 2D representation



Time courses



Spatial distribution of the clusters

Outline

- Introduction
- Pre-history: the non-variational approach
- Mumford-Shah
- Snakes and variations thereof
- Active regions
- More features
- More structure
- More dimensions
- Conclusion

Conclusion and Perspectives: Mathematics

- Clear increase in the mathematical sophistication of image segmenters.
- We are going away from 19th century mathematics and beginning to use 20th century maths!
- What are the challenges:

1. Well-posedness.
2. Numerical schemes.
3. Geometry, in particular random geometry.

Conclusion and Perspectives: Segmentation

- We are clearly driven by the technology. . . but
  - we use very few geometric and physical image formation models.
- We would very much like to use prior knowledge, to acquire knowledge automatically . . . but
  - we use very few of the effective models of data distribution and classification procedures developed (statistical learning theory and theoretical computer science).
- We see very little of our segmentation, matching, warping programs in the hospital . . . but
  - we build very few systems.

The final word

- Mathematics are necessary but not sufficient ...