Biomedical Image Analysis: Challenges and Perspectives

Nikos Paragios

http://vision.mas.ecp.fr
Ecole Centrale de Paris/INRIA

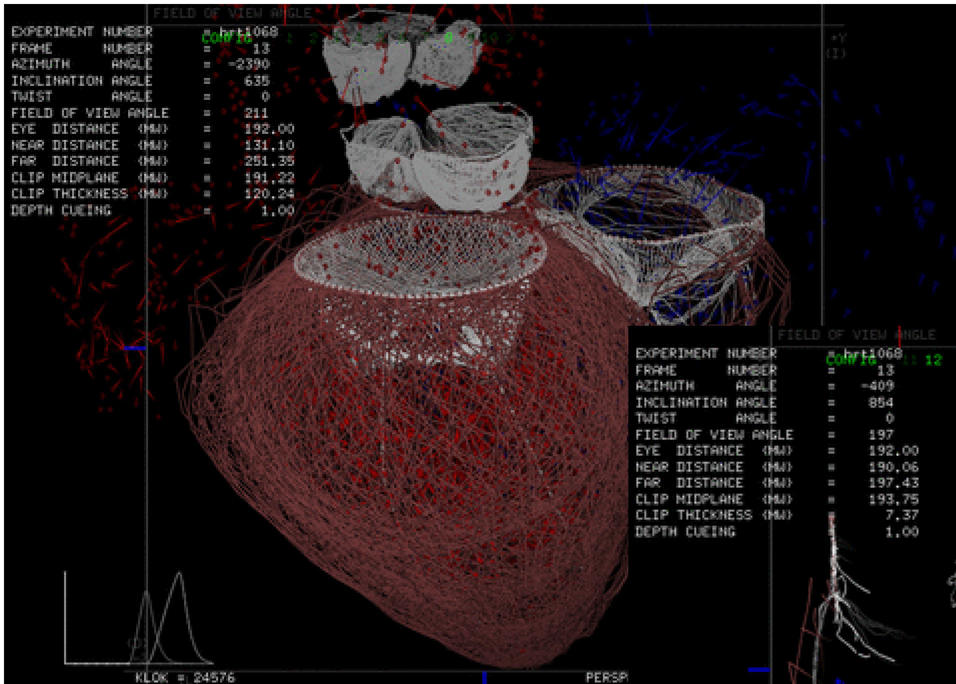
CENTRALE PARIS

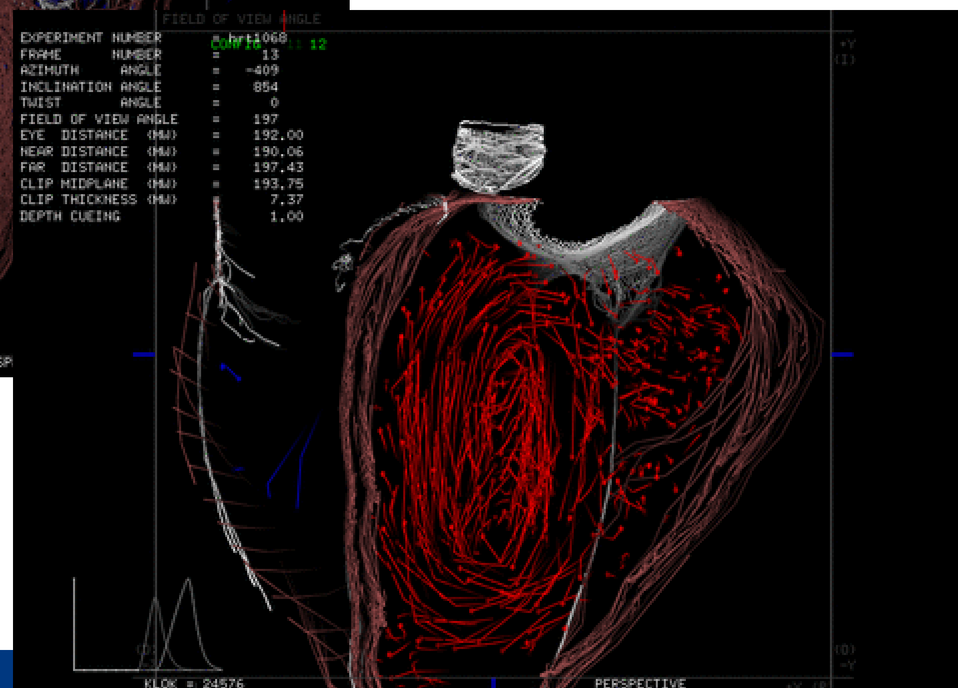
INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche SACLAY - ÎLE-DE-FRANCE

Heart Modeling





@ Pesking & Mqueen, NYU

Outline

Bio-engineering

On the need of Automatic Processing

Medical Images Modalities

Most Critical Methodologies in Medical Image Analysis

Examples, Applications where one can make the difference

Current State in Health-Care

Life expectancy is growing in the developing and developed countries

- Improve of the quality and conditions of life,
- Significantly advances in medicine in terms of prevention & early diagnosis
- Novel techniques for efficient treatment to a number of diseases that used to be uncured in the past

Advances in Prevention & Early Diagnosis

- Improve of the hardware of existing modalities (CT,MR,US) in terms of precision, definition and quality when acquiring medical images
- Possible the acquisition of novel non-invasive images/information spaces (f-MRI, DT-MRI, …, molecular imaging)
- Better visualization of human organs that provides to the physicians better understanding of diseases as well as better signs of detection
- Improve of the means of processing and visualization of such a rich information space

Current State in Health-Care

Advances in treatment

- Advances in computational biology, extraction of the DNA chain & new drug discoveries,
- More efficient ways to detect on non-healthy cells and focus the treatment on these cells
- More advanced hardware that allows the physicians better understanding of diseases,

All these are possible because :

- The cost of acquisition is dropping down
- Acquisition becomes more frequent, even for prevention
- Evidence comparison using information from different modalities can do a much better job than evidence from one modality

Current Limitations of Health Care

On the other hand

- In developed countries health case becomes an industrial routine with paramedics being in charge of the early stage of health exams
- Physicians spend lesser and lesser time with patients and tend to rapidly examine the outcome of medical exams, leading to often no-diagnosis at an early stage that becomes critical to efficient treatment
- The cost becomes a major issue, we are moving towards two-step health-care and therefore certain scans become less attainable unless hardly justified through preliminary results
- While hardware continues to evolve (Siemens recently introduced a CT 64 slices scanner, Philips their new 3D ultrasound scanner, etc) the consensus is that soon hardware will reach its limitations with marginal improvements in terms of performance
- What will make the difference in the future is the means of exploitation of data that are now standard packages of hardware

Need for Automatic Processing

Understanding and extracting content from images is a primary answer to a number of existing limitations of the health-care system.

Automatic Processing of Visual Content can Never Replace the Physician in the Health-Chain, it is intended to facilitate his tasks

Recent advances of computational devices (PCs) have made possible the development of complicated mathematical models that can provide answers to the visual understanding procedure

Why It will Work?

Computational Vision; that means understanding the environment from images is still on an embryonic stage;

Medical Imaging on the other hand is the most established branch of this domain with significant advances being made;

- Knowledge of the environment
- Control on the acquisition procedure
- Prior anatomical models that can account for the ill-posedness of the inference problems as well as for the lack of information
- Substantial amount of money invested in the domain from health-care providers and hardware manufacturers (Siemens, Phillips, General Electric)

Still a lot of progress is to be made, but prototypes exist for assisting diagnosis with automatic image extraction techniques

An example

Virtual colonoscopy (VC)

- x rays and computers to produce two- and three-dimensional images of the colon (large intestine) from the lowest part, the rectum. The procedure is used to diagnose colon and bowel diseases. VC can be performed with computed tomography (CT), sometimes called a CAT scan, or with magnetic resonance imaging (MRI).

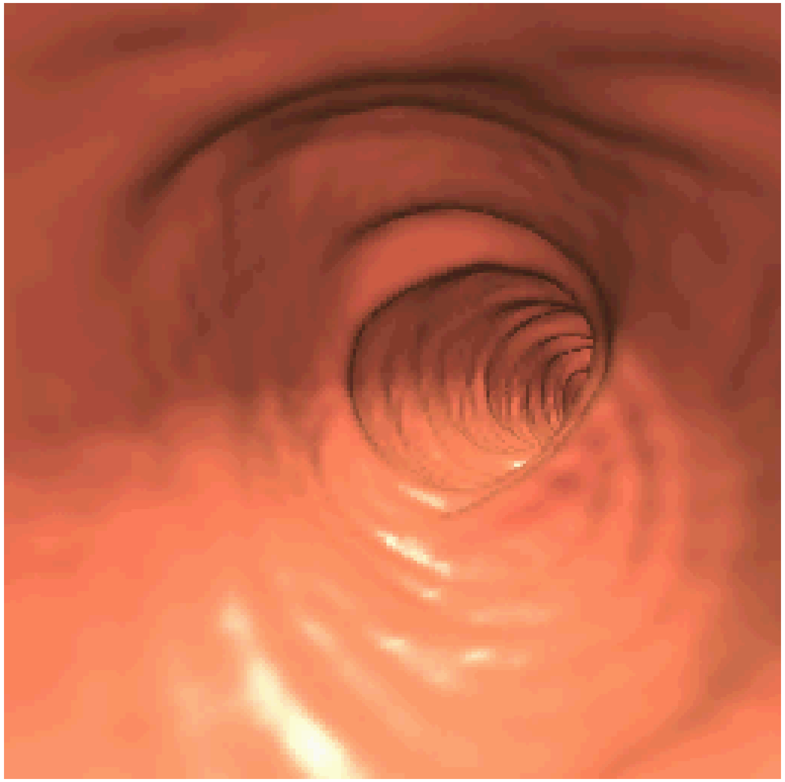
Versus:

Conventional colonoscopy (CC),

- the doctor inserts a colonoscope—a long, flexible, lighted tube—into the patient's rectum and slowly guides it up through the colon. Pain medication and a mild sedative help the patient stay relaxed and comfortable during the 30- to 60-minute procedure.

VC is more comfortable than CC colonoscopy

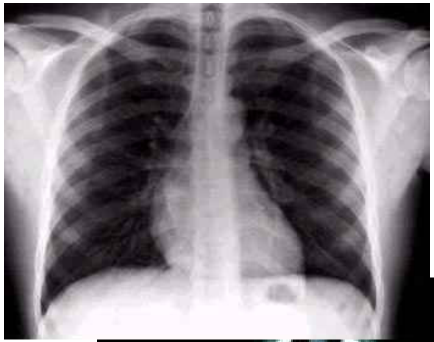
- it does not use a colonoscop, no sedation is needed, and you can return to your usual activities. VC provides clearer, more detailed images than a conventional x ray.



Some of the Medical Image Modalities

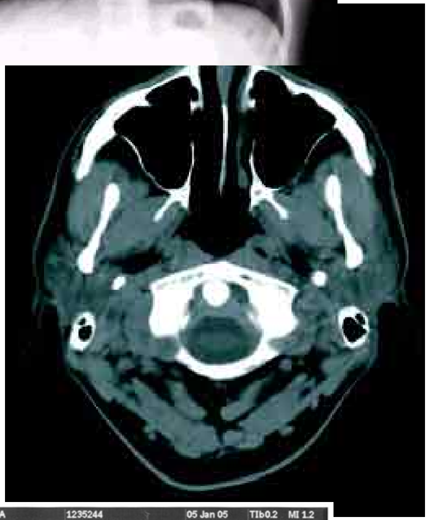
X-ray, Roentgenogram

- is the most basic tool in medical imaging; a two-dimensional shadow picture is used to examine soft and bony tissue.



CT, Computerized Tomograpy

- takes multiple cross sectional roentgenogram in a plane perpendicular to the patient. A computer uses mathematical operations to construct a 3-D image from the two-dimensional sections.



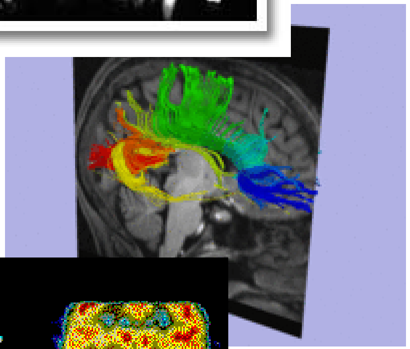
Ultrasound and Echocardiogram

- use sound waves to image organs and view organ function in real time. The sound waves are reflected back at differing intensities based on the density and penetration of the organ.

Medical Image Modalities

MRI, Magnetic Resonance Imaging

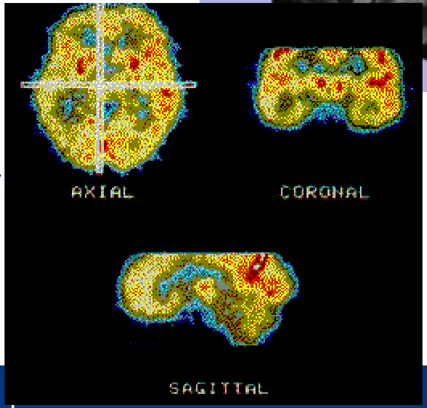
- creates a magnetic field which coordinates the spin of hydrogen ions. When the magnetic field is removed, the relaxation of spinning is measured. The different relaxation of the hydrogen ions, which is based on water content, is converted into an image.



Diffusion-tensor MRI, f-MRI

PET, Positron Emission Tomography

- uses radioactive labeled glucose, the primary energy source of all cells, to monitor organ metabolism.

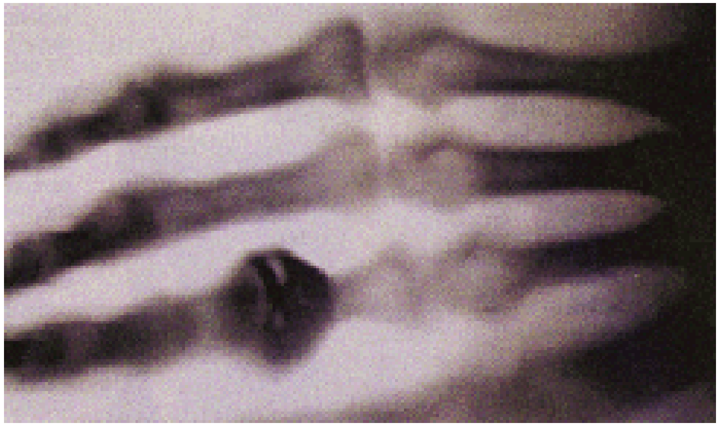


Courtecy T. Petters

X-rays

- Wilhelm Conrad Röntgen (1845-1923)

Nobel Prise in Physics, 1901



- “X” stands for “unknown”
- *X-ray imaging* is also known as
    - radiograph

Courtecy T. Petters

X-rays

Bertha Röntgen's Hand 8 Nov, 1895

A modern radiograph of a hand

- Calcium in bones absorbs X-rays the most
- Fat and other soft tissues absorb less, and look gray
- Air absorbs the least, so lungs look black on a radiograph

Courtecy T. Petters

Towards 3D imaging

X-ray imaging
1895

Mathematical results:
Radon transformation
1917

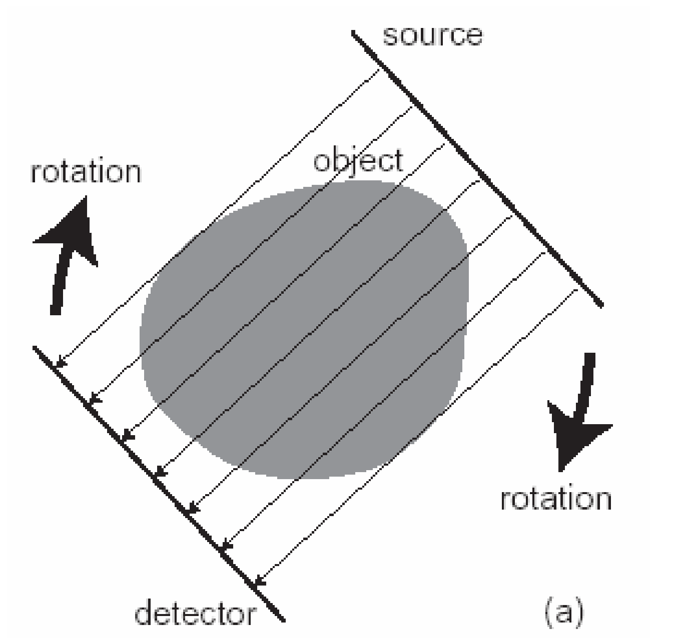
Computers can perform complex mathematics to reconstruct and process images
Late 1960's:

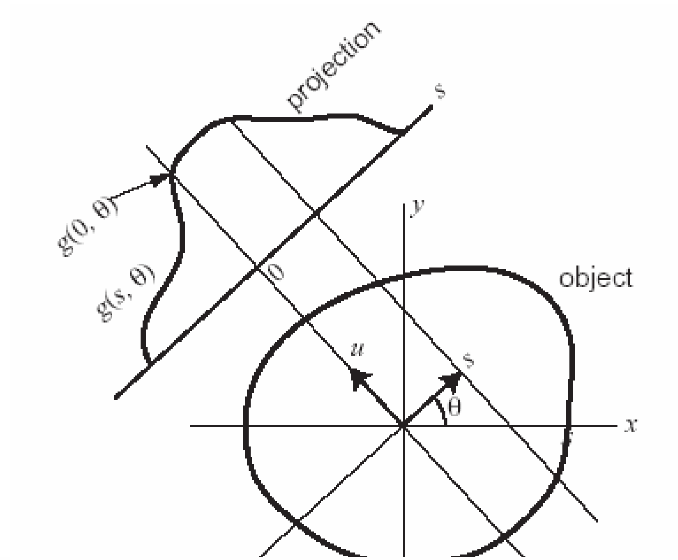
Development of CT
(computed tomography)
1972

- Image reconstruction from projection
- Also known as CAT (Computerized Axial Tomography)
- "tomos" means "slice" (Greek)

Courtecy T. Petters

Radon Transformation





$$g(s,\theta) = \iint_{-\infty}^{\infty} f(x,y)\,\delta(x\cos\theta + y\sin\theta - s)\,dx\,dy$$

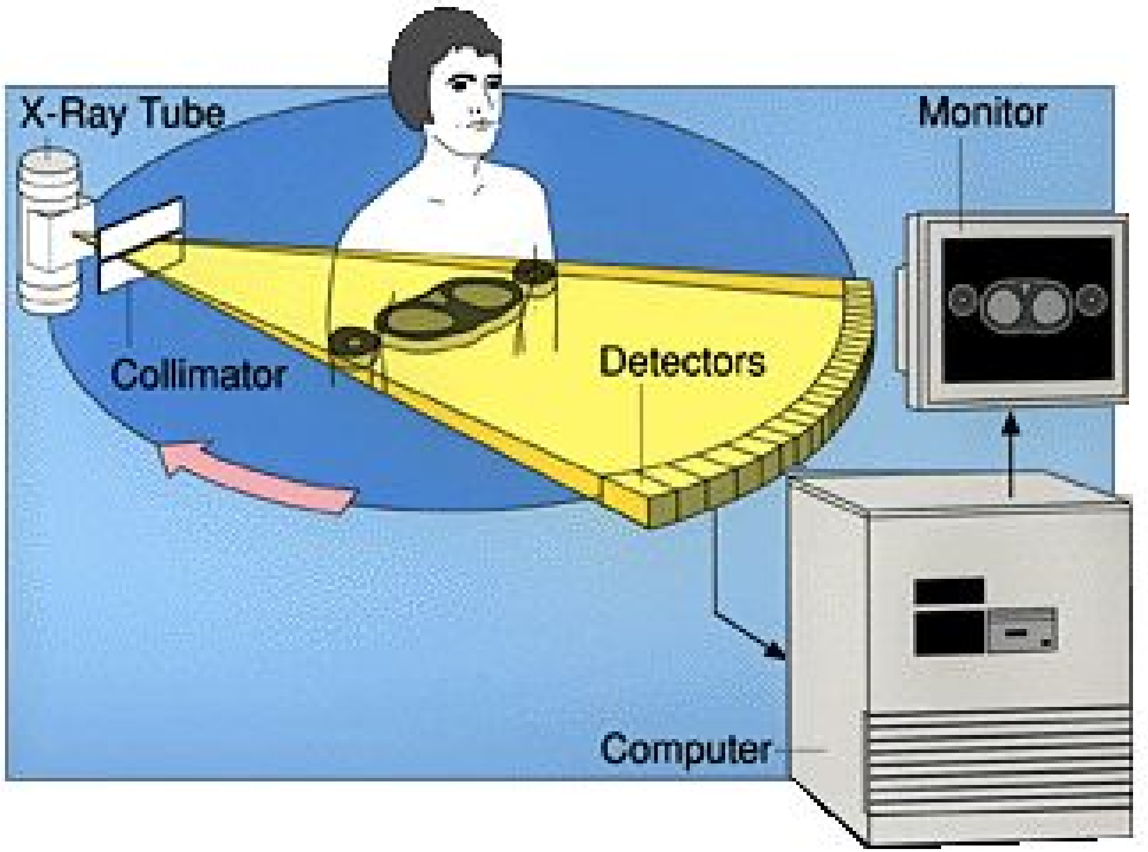
- Mathematical transformation (related to Fourier)
- Reconstruction of the shape of object (distribution *f(x,y)*) from the multitude of 2D projections

$g(s,\theta)$

Figure from www.imaginis.com/ct-scan/how_ct.asp

Courtecy T. Petters

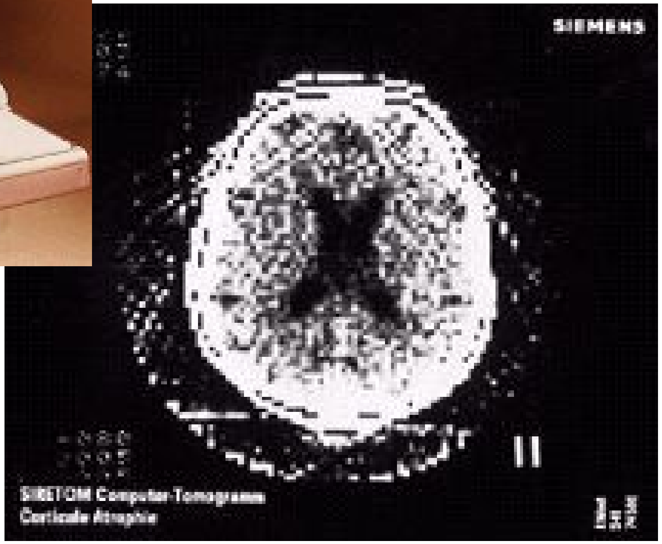
CT imaging



Courtecy T. Petters

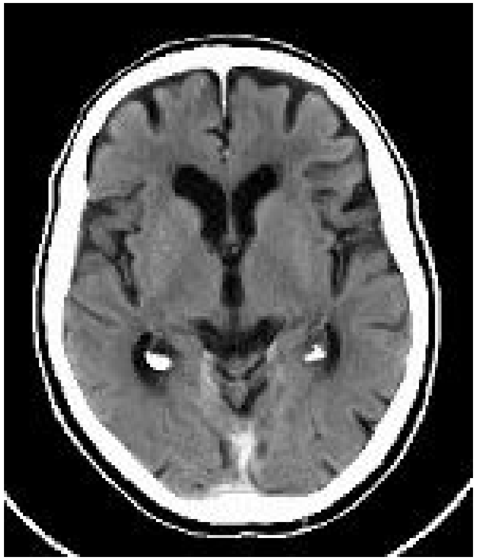
CT imaging, availability (since 1975)

1974



Original axial CT image from the dedicated Siretom CT scanner circa 1975. This image is a coarse 128 x 128 matrix; however, in 1975 physicians were fascinated by the ability to see the soft tissue structures of the brain, including the black ventricles for the first time (enlarged in this patient).

25 years later



Axial CT image of a normal brain using a state-of-the-art CT system and a 512 x 512 matrix image. Note the two black "pea-shaped" ventricles in the middle of the brain and the subtle delineation of gray and white matter (Courtesy: Siemens)

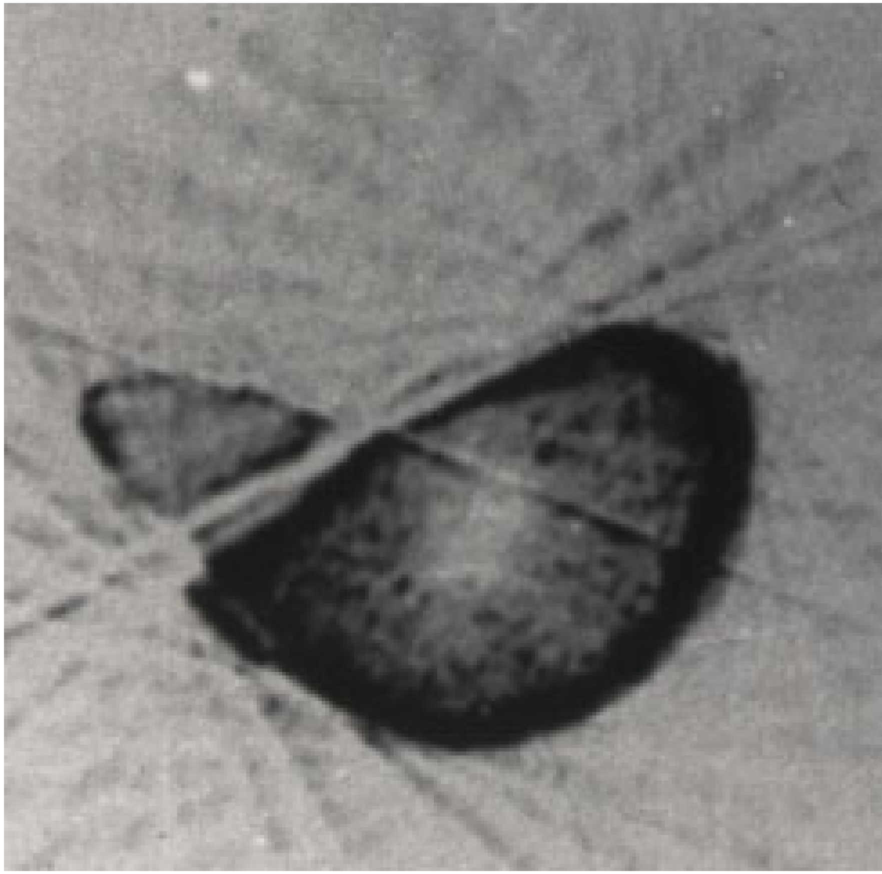
Courtecy T. Petters

Clinical Acceptance of CT!?

Dr James Ambrose 1972

- Radiologist, Atkinson - Morley's Hospital London
- Recognised potential of EMI-scanner

"Pretty pictures, but they will never replace radiographs" – Neuroradiologist 1972



Courtecy T. Petters

Then ……………..and Now

80 x 80 image

3 mm pixels

13 mm thick slices

Two simultaneous slices!!!

80 sec scan time per slice

80 sec recon time

512 x 512 image

<1mm slice thickness

<0.5mm pixels

0.5 sec rotation

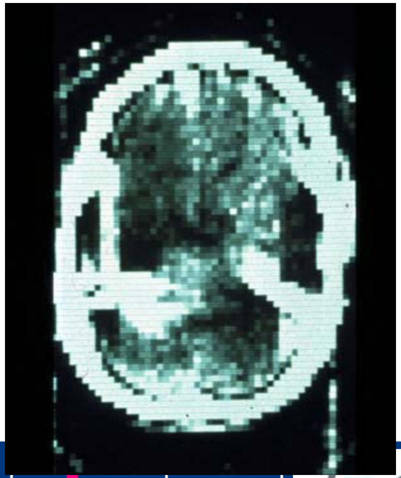
0.5 sec recon per slice

Isotropic resolution

Spiral scanning - up to 16 slices simultaneously

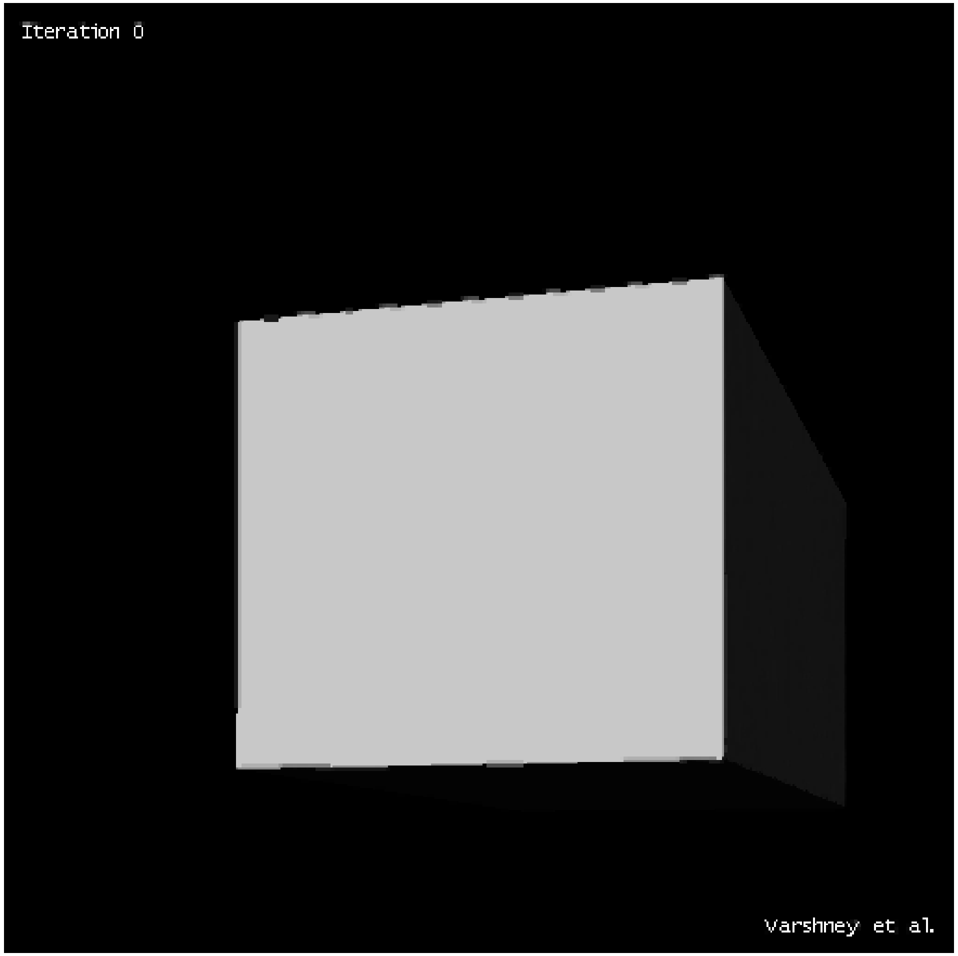
Courtecy T. Petters

30 Years of CT



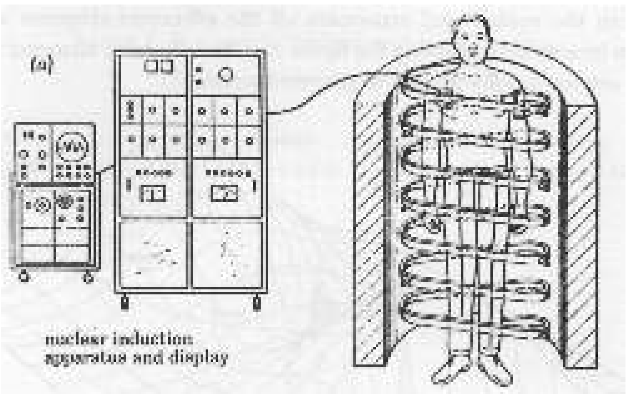
Some nice input

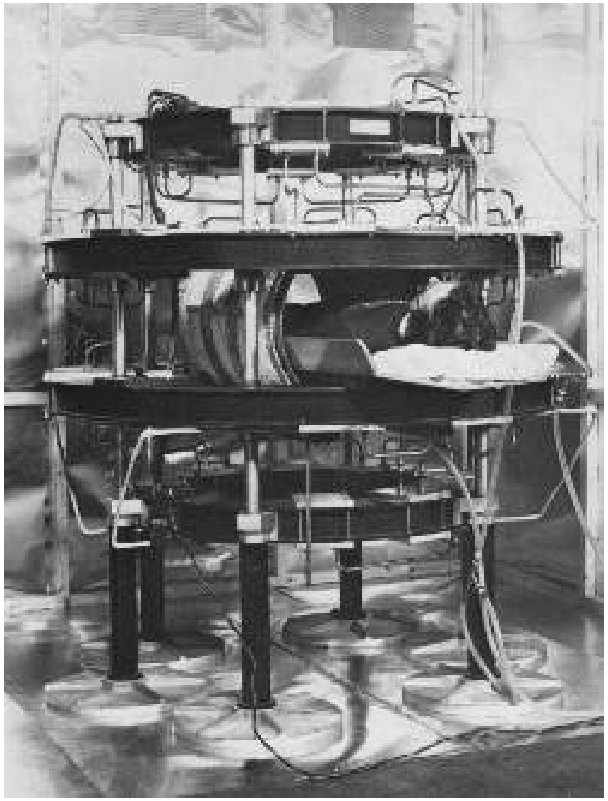
Some nice results



Courtecy T. Petters

Birth of MRI

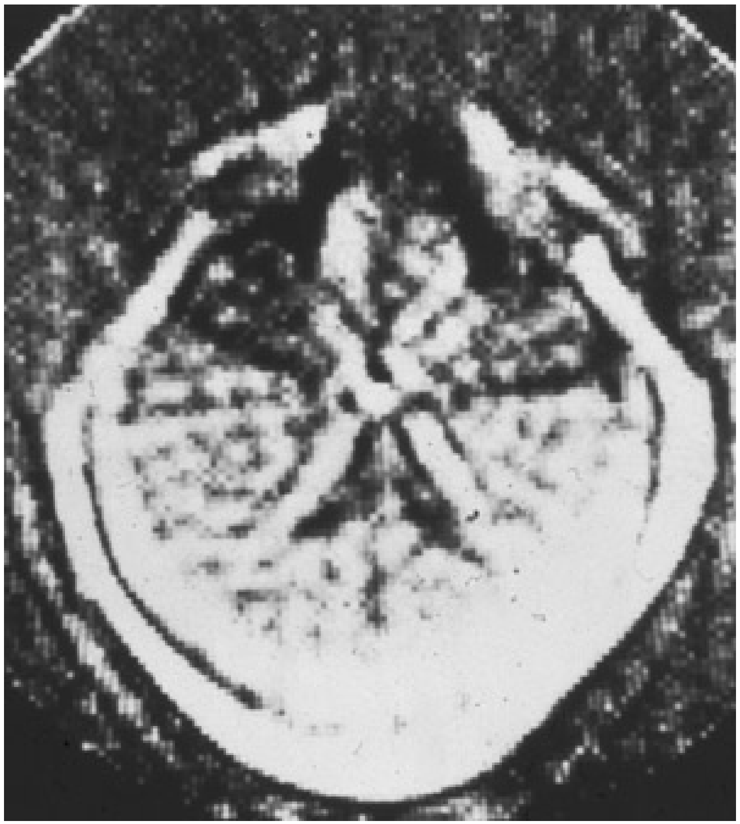




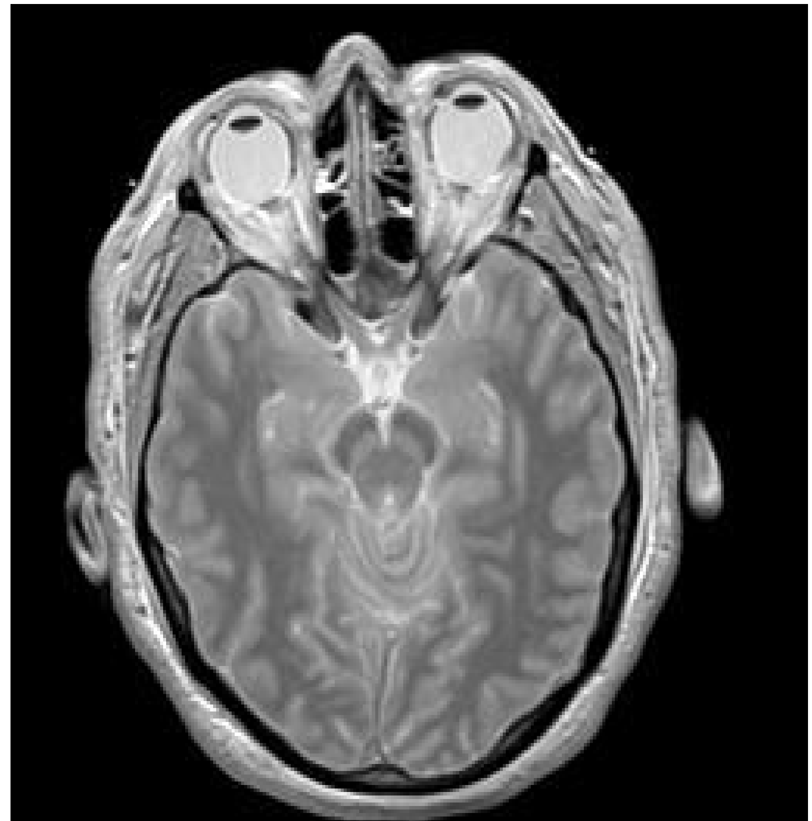
- Electro Marnetic signal emitted (in harmless radio frequensy) is acquired in the time domain
- image has to be reconstructed (Fourier transform)

Courtecy T. Petters

30 Years of MRI



First brain MR image



Typical T2-weighted MR image

Courtecy T. Petters

MR Imaging

"Interesting images, but will never be as useful as CT"

- (A different) neuroradiologist, 1982

Courtecy T. Petters

MR Imaging …more than T1 and T2

MRA - Magnetic resonance angiography

- images of vessels

MRS - Magnetic resonance spectroscopy

- images of chemistry of the brain and muscle metabolism

fMRI - functional magnetic resonance imaging

- image of brain function

PW MRI – Perfusion-weighted imaging

DW MRI – Diffusion-weighted MRI

- images of nerve pathways

Courtecy T. Petters

Magnetic Resonance Angiography

MR scanner tuned to measure only moving structures

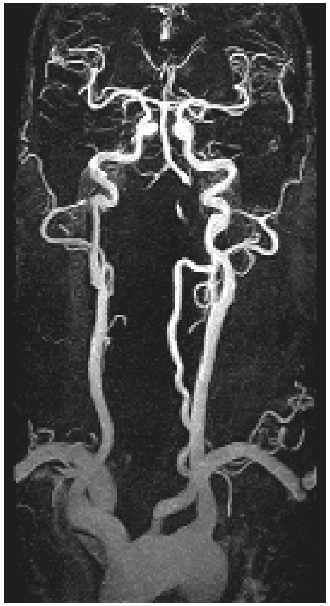
"Sees" only blood - no static structure

Generate 3-D image of vasculature system

May be enhanced with contrast agent (e.g. Gd-DTPA)

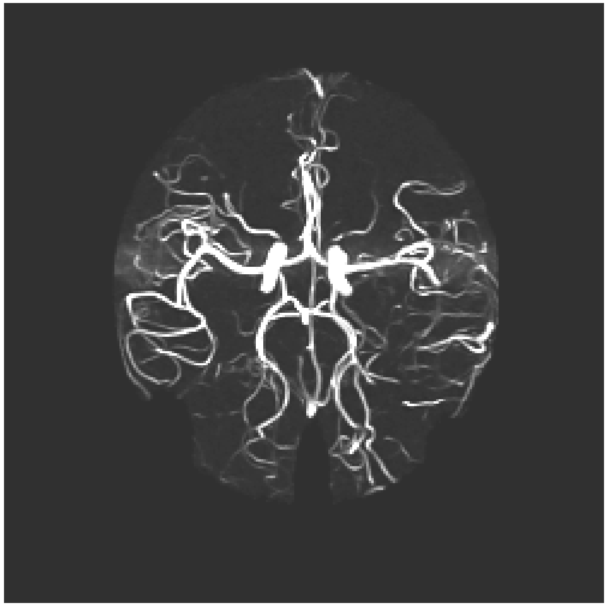
Courtecy T. Petters

MR Angiography



GD-enhanced

GD-enhanced



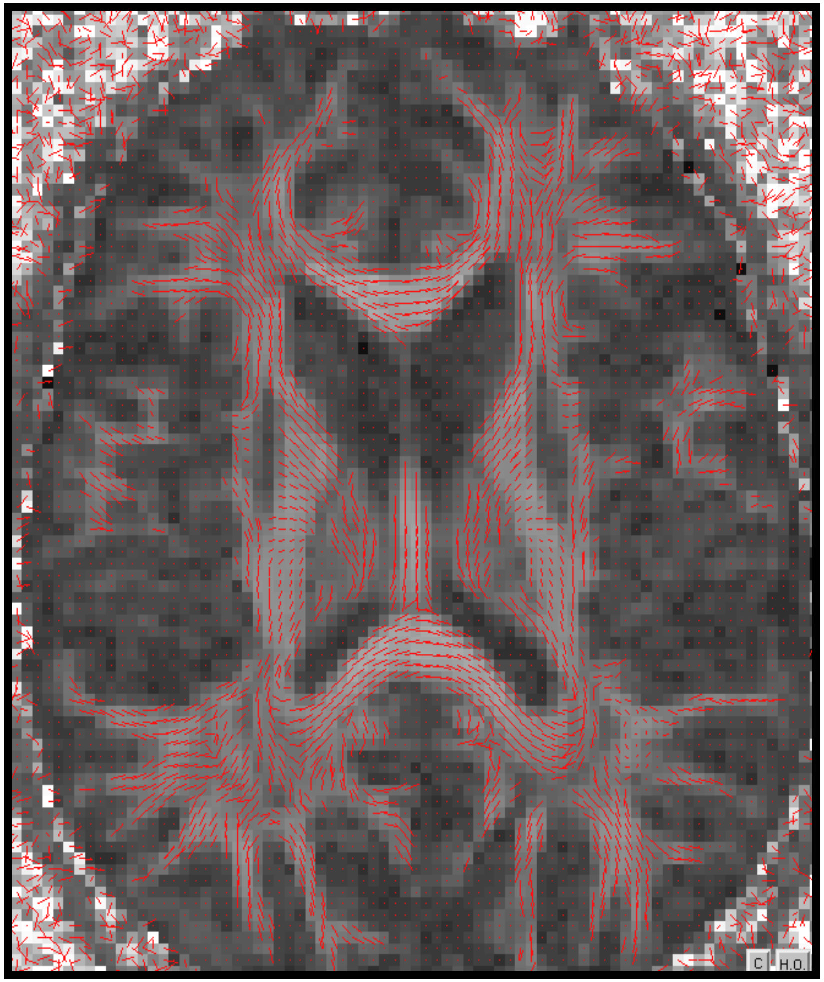
Phase-contrast

In-flow

Courtecy T. Petters

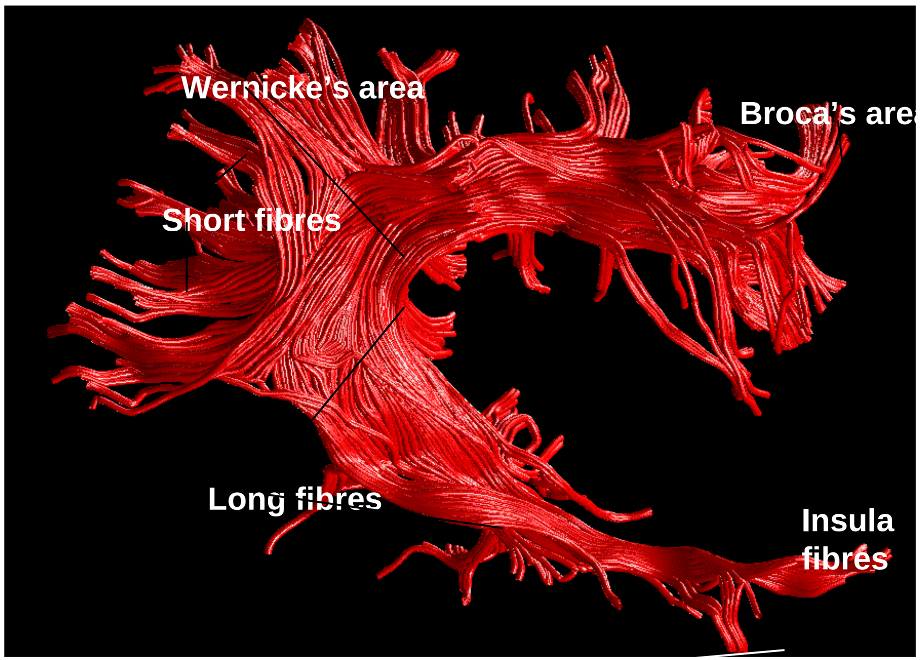
Diffusion-Weighted MRI

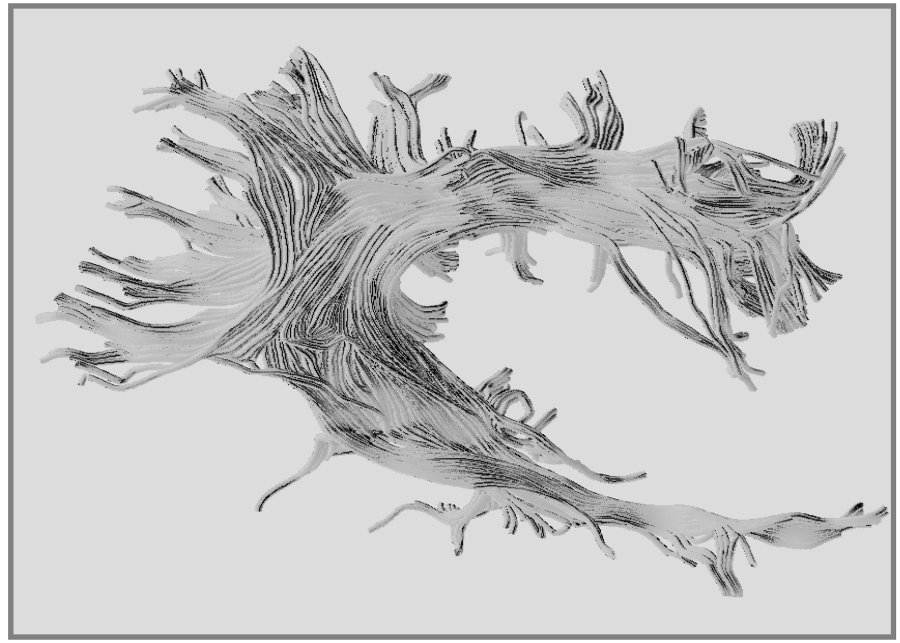
Image diffuse fluid motion in brain
Construct “Tensor image” – extent of
diffusion in each direction in each voxel
in image
Diffusion along nerve sheaths defines
nerve tracts.
Create images of nerve
connections/pathways



Courtecy T. Petters

Tractography

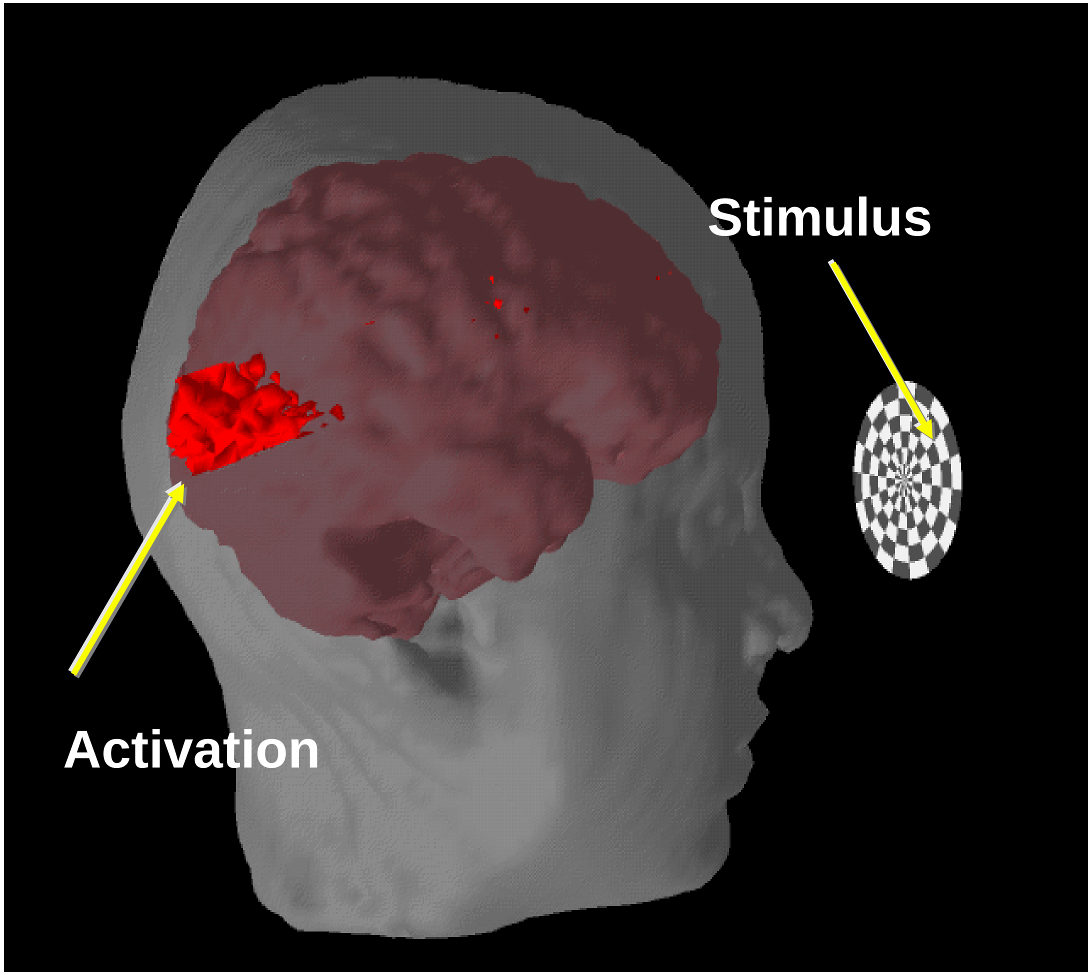




Courtecy T. Petters

fMRI

Subject looks at flashing disk while being scanned "Activated" sites detected and merged with 3-D MR image



Slides from Terry Peters

fMRI in Neurosurgery Planning

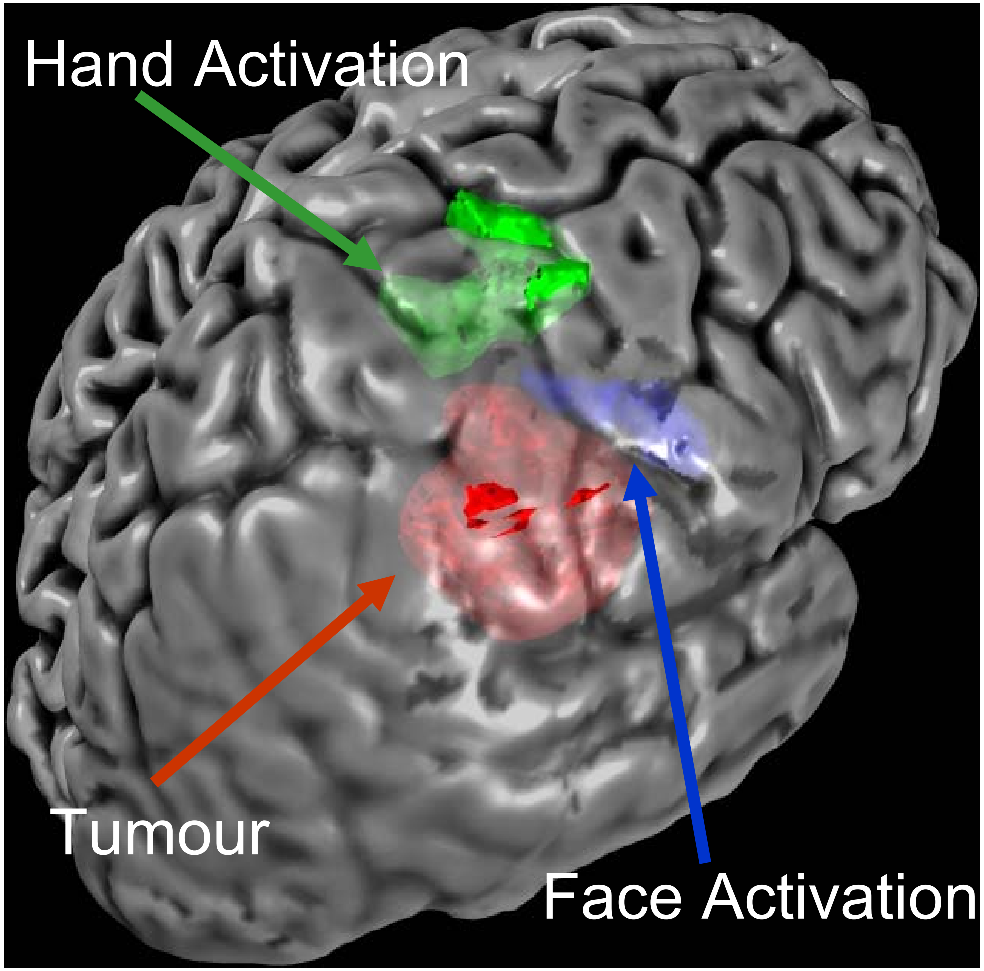
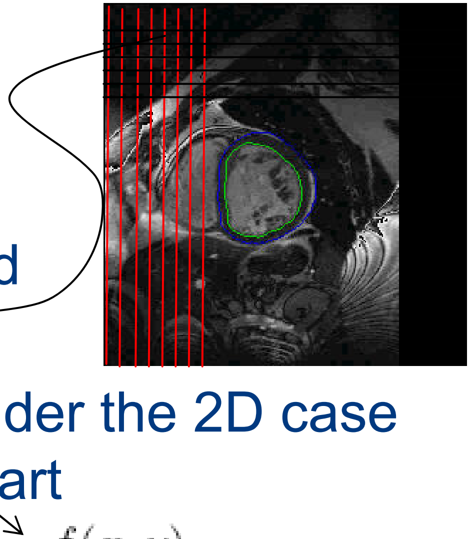


Image Formation; Representing Signals in the Form of Continuous Functions

Images correspond to a certain spatial and temporal resolution; where observations are completely independent

Such resolution is driven from the acquisition and differs from one modality to the other

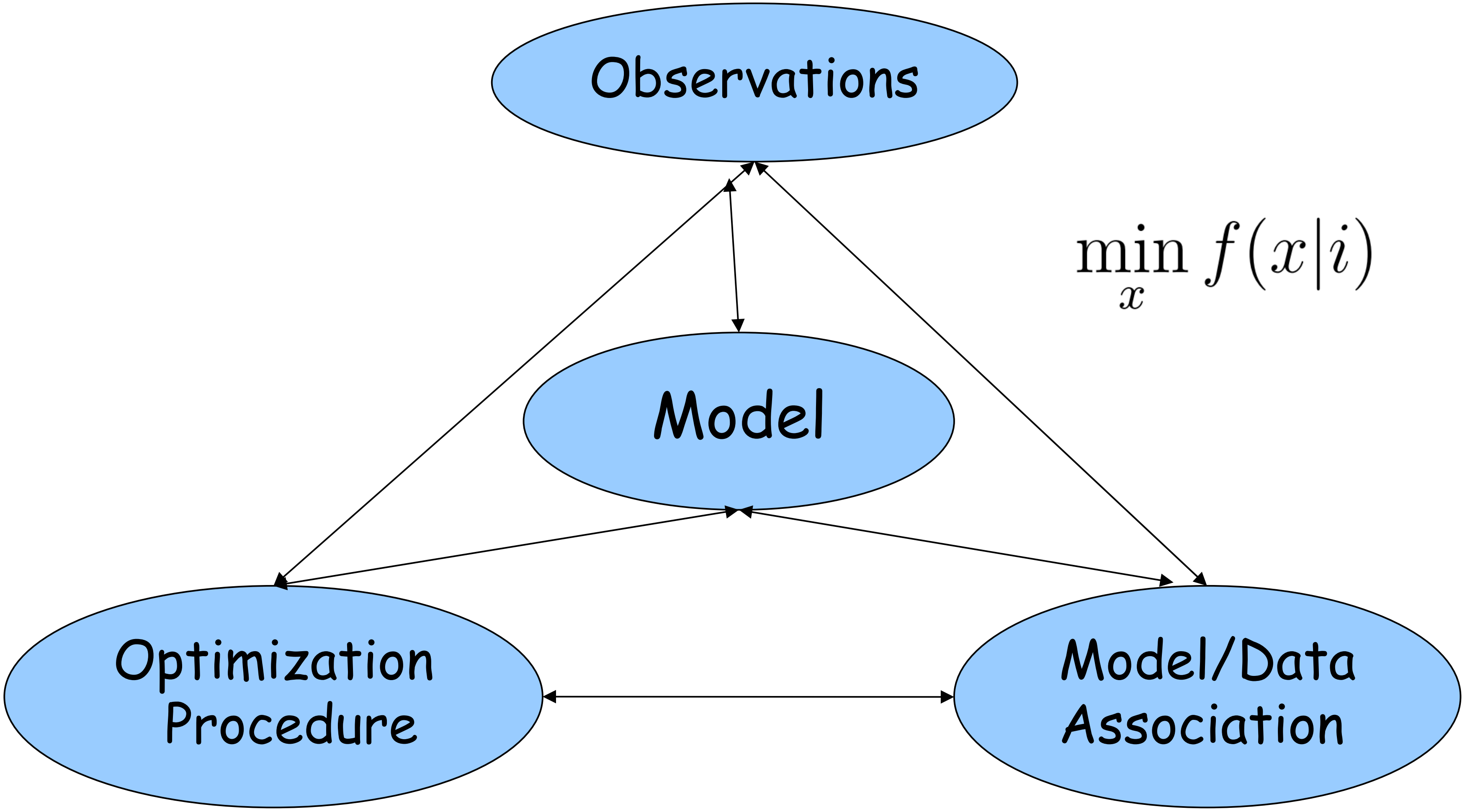


In order to simplify the notation, let us consider the 2D case as shown for the CT slice image of the heart

$f(x, y)$

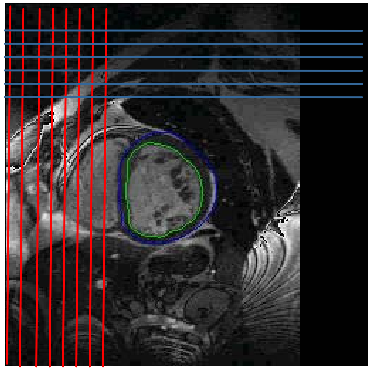
Medical imaging from a mathematical perspective consists of recovering inference from the data or recovering some qualitative/quantatita $Z = G(f(x,y))$ on from the image

Medical Imaging Paradigm



$$\min_x f(x|i)$$

Example of Medical Imaging Problem



Images correspond to very local measurements

being completely independent

Simple example: left ventricle extraction in CT images; the solution can be represented using a set of points

$$x = (x_1, \cdots, x_n)$$

Using this points, one can define the solution through some kind of interpolation strategy

$$\partial \mathcal{R} = \pi(x_1, \cdots, x_n)$$

And try to optimize the position of these points in the image plane, according to the lowest potential of an objective functions, like for example the strength of the edges over the curve

$$\min_{(x_1, \cdots, x_n)} \int_{\partial \mathcal{R}} f(\partial \mathcal{R}(p) | I) dp$$

Image Formation; Representing Signals in the Form of Continuous Functions

Often such a direct linear operator does not exist and therefore, feature extraction is reformulated either as a minimization problem;

$$argmin_Z\{Z - G(f(x,y))\}$$

That consists of finding the most optimal configuration that minimizes some distance between observations and the solution of the problem;

Or in a probabilistic fashion:

$$argmax_Z\{p(Z|G(f(x,y)))\}$$

Where certain assumptions on the space of plausible solutions are made and we consider the distribution to be known; that is possible in some cases but not for the most general medical imaging problems

Main Challenges

Curse of Dimensionality : find a compromise between the expression power of the model and its complexity

Curse of Non-linearity: the association of the model parameters and the observations are highly non-linear

Curse of Non-Convexity: the designed objective function leaves in a high-dimensional non-convex space

Curse of Non-Modularity: any solution is hardly portable to another application setting or another problem

How to Address these challenges

Curse of Dimensionality : Prior Knowledge either through anatomy of machine learning techniques towards dimensionality reduction

Curse of Non-linearity: Model Decomposition / Data association allows direct support estimation of parameter selection from the images

Curse of Non-Convexity: Regularization terms / dropping out of constraints can improve the optimality properties of the obtained solution

Curse of Non-Modularity: Model/Data Association/Inference Decomposition and use of gradient free methods

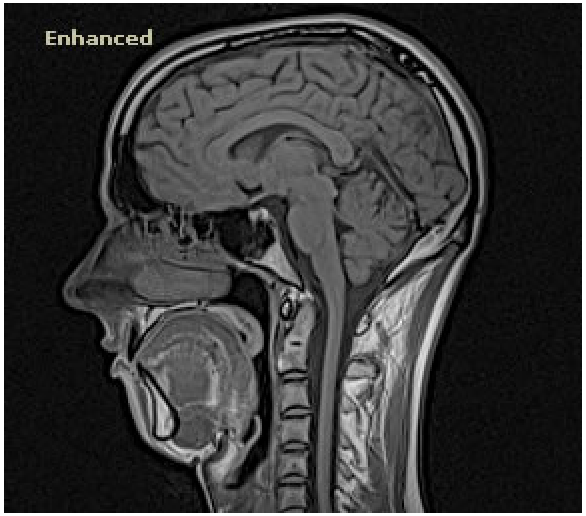
Medical Imaging: Challenges

In the most general case, given an observation f (one or multiple images), mathematical imaging consists of defining operators that once applied to these images returns to the user a vector of valuable information

- Images With Improved Quality (denoising)
- Images of smaller size where information has been preserved (compression)
- Image separated in uniform regions (segmentation), & Region of interest in the images that corresponds to a given anatomical structure (object extraction)
- Recovering the same structure of particular interest in a number of successive images
- Understanding Object Deformation from one image to the next
- Transformations that related images taken from different point of views of the same or different modality (registration)
- 3D visualization of the structure of interest (reconstruction from partial views)
- Measures that allow the differentiation of this image from other images in the training set, or from expected statistical patterns (diagnosis)
- Inference between existing mechanical models and images to made such models patient-specific (modeling organ behaviour)

Image Enhancement

Noise is a predominant factor of medical image acquisition in a number of modalities



Problem Formulation

Courtecy N. Azzabou

Additive Noise

$$I(\mathbf{x}) = U(\mathbf{x}) + n(\mathbf{x}) \quad \text{For} \quad \mathbf{x} \in \Omega$$

Gaussian distribution of the noise

$$n \sim \mathcal{G}(0, \sigma_n^2)$$

Patch definition

$$I_{\mathbf{N}_x} = \{I(\mathbf{y}) \text{ with } \mathbf{y} \in \mathbf{N}_{\mathbf{x}}\}$$

Courtecy N. Azzabou

State of the Art

Neighborhood filtering

- Sigma, Bilateral, NL-means , ….

PDE and Energy minimization based techniques

- Anisotropic diffusion and Total variation minimization

Transform domain and Sparse image modeling

- Wavelet, bandelet, contourlet, K-SVD …

Statistical Image models and Bayesian framework

- MRF models, non parametric models

Mathematical Definition of Image Enhancement

Let us consider an input signal *f* that has been corrupted by a white noise model *n*; white models refers to zero mean model

In such a case for a given image location (x,y) we observe g(x,y)

$$g(x, y) = f(x, y) + n(x, y)$$

In order to recover *f,* given that noise models are unknown we assume that the image in a small scale is local smooth, and that noise model affects observations with the same way;

$$min_{f(x,y)} \int \int_{\mathcal{N}(x,y)} (f(x,y) - g(u,v))^2 \, dudv$$

And in that case the optimal filtered value consists of the simple mean value of the observations

$$f(x,y) = \frac{1}{|\mathcal{N}(x,y)|} \sum \sum_{\mathcal{N}(x,y)} g(u,v) dudv$$

That encodes the zero-mean assumption of the noise model; what is known to be mean filtering…

Courtecy N. Azzabou

- A Data driven and a global image model that encodes image structure
- Linear image model

$$E(U) = \int_\Omega \left(\left[\frac{1}{Z(\mathbf{x})} \int_\Omega w_{\mathbf{xy}} U(\mathbf{y}) d\mathbf{y} \right] - U(\mathbf{x}) \right)^2 d\mathbf{x} + \lambda \int_\Omega (I(\mathbf{x}) - U(\mathbf{x}))^2 \, d\mathbf{x}$$

- The interaction between pixels is expressed using weights

$$w_{\mathbf{xy}} = \exp\left(-\frac{\|\Psi_I(\mathbf{x},\mathbf{y})\|^2}{2\sigma_{ph}^2} \right) \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma_s^2} \right)$$

Courtecy N. Azzabou

Examples of distance used to evaluate the photometric similarity

Distance between image patches $\Psi_I(\mathbf{x},\mathbf{y}) = \left\|I_{\mathbf{N}_x} - I_{\mathbf{N}_y}\right\|^2$

Distance between projection of patches on a subspace determined using PCA $\Psi_I(\mathbf{x},\mathbf{y}) = \|\mathbf{F}_\mathbf{x} - \mathbf{F}_\mathbf{y}\|^2$

Consider the statistical distribution of the distance L2 distance between patches

$$w_{\mathbf{xy}} = \frac{1}{2\sqrt{d\pi}} \exp - \left(\frac{\left[\frac{\Psi_I(\mathbf{x},\mathbf{y})}{2\sigma_n} - d \right]^2}{4d} \right) \exp \left(-\frac{\left\|\mathbf{x} - \mathbf{y}\right\|^2}{2\sigma_s^2} \right)$$

Courtecy N. Azzabou

Spatial Bandwidth selection : must be adapted to each pixel

$$w_{\mathbf{xy}} = \exp\left(-\frac{\|\Psi_I(\mathbf{x},\mathbf{y})\|^2}{2\sigma_{ph}^2}\right)\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma_s(\mathbf{x})^2}\right)$$

Why a variable bandwidth ??

Large spatial bandwidth values for smooth areas

Small bandwidth values for smooth areas

Courtecy N. Azzabou

$$E(U,\sigma_s) = \int_\Omega \left(\left[\frac{1}{Z(\mathbf{x})} \int_\Omega w_{\mathbf{xy}}(\sigma_s(\mathbf{x})) U(\mathbf{y}) d\mathbf{y} \right] - U(\mathbf{x}) \right)^2 d\mathbf{x}$$
$$+ \lambda \int_\Omega (I(\mathbf{x}) - U(\mathbf{x}))^2 d\mathbf{x} + \mu \int_\Omega \|\nabla \sigma_s(\mathbf{x})\|^2 d\mathbf{x}$$

Alternating minimization

Minimization with respect to *U* $U^{t+1} = U^t - dt\nabla E$

Minimization with respect to σ_s $\sigma_s^{t+1}(\mathbf{x}) = \sigma_s^t(\mathbf{x}) - dt\nabla E|_{\sigma_s(\mathbf{x})}$

Convex Functional Minimization

constrained problem

Courtecy N. Azzabou

Energy function

$$E(U) = \int_\Omega \left(\left[\frac{1}{Z(\mathbf{x})} \int_\Omega w_{\mathbf{xy}} U(\mathbf{y}) d\mathbf{y} \right] - U(\mathbf{x}) \right)^2 d\mathbf{x} + \lambda \int_\Omega (I(\mathbf{x}) - U(\mathbf{x}))^2 \, d\mathbf{x}$$

Constraint $\int_\Omega (I(\mathbf{x}) - U(\mathbf{x}))^2 d\mathbf{x} = \sigma_n^2$

Courtecy N. Azzabou

Problem formulation

Minimizing the following constrained problem

$$E_{reg}(U) = \int_{\Omega} \left\| \left[\frac{1}{Z(\mathbf{x})} \int_{\Omega} w_{\mathbf{xy}} U(\mathbf{y}) d\mathbf{y} \right] - U(\mathbf{x}) \right\|^2 d\mathbf{x}$$

Under the constraint

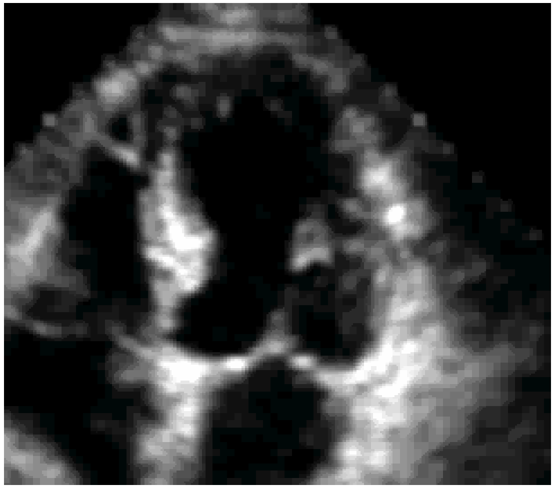
$$\int_{\Gamma_c(GL)} (I_c(\mathbf{x}) - U_c(\mathbf{x}))^2 d\mathbf{x} = \sigma_n^2(GL)$$

Euler Lagrange equation using the Lagrange multiplier updated as

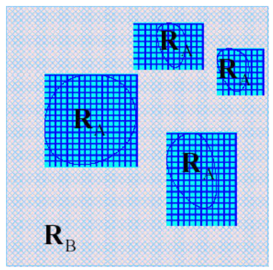
$$\lambda_{GL}^c \propto \frac{1}{2\sigma_n^2(GL)}$$

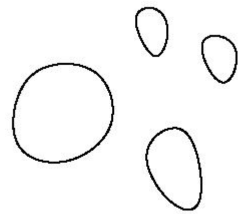
Courtecy N. Azzabou

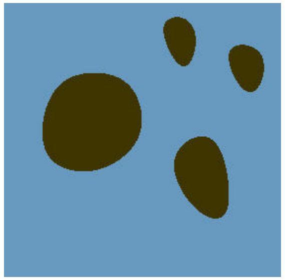
Some Results



Model-free Segmentation



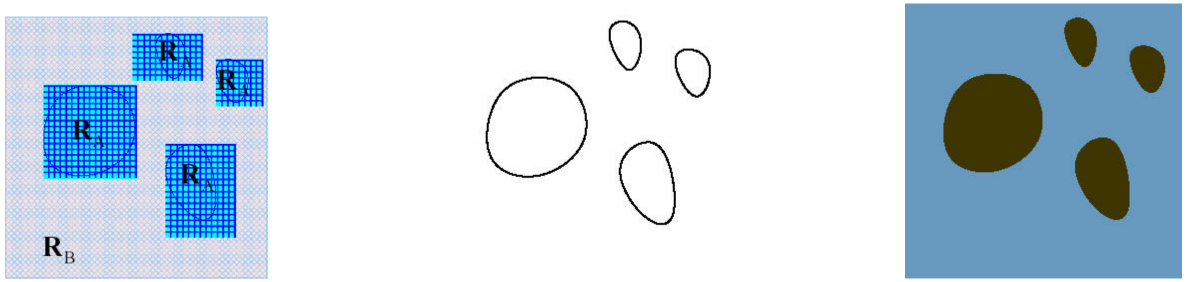




frame partition according to a given feature space is a core component of imaging, vision & graphics

- Model-free Grouping
  - Edge Detection, Geometric Flows, Snakes, Clustering, MRFs, Graph Theory
- Model-based Grouping
  - Deformable Templates, Active Shape/Appearance Models, etc,

Segmentation



Evolve an initial curve towards the lowest potential of a cost function that is a compromise between data-driven attraction terms and internal constraints

$$E[(C)(p)] = \alpha \int_0^1 E_{int}(C(p))dp + \beta \int_0^1 E_{img}(C(p))dp + \gamma \int_0^1 E_{con}(C(p))dp$$

- Internal term stands for curve smoothenss
- The image term guides the curve towards the desired image properties
- External term can stand for prior knowledge or user interaction
- One can minimize this cost function and recover a way of deforming the curve towards the lowest potential of the cost function

$$\begin{cases} \frac{\partial C}{\partial t}(p) = F(\mathcal{K})\,\mathcal{N} \\ C(p,0) = C_0(p) \end{cases}$$

On the Propagation of Curves

Snake Model (1987) [Kass-Witkin-Terzopoulos]

- Planar parameterized curve C:R-->RxR
- A cost function defined along that curve

$$E[(C)(p)] = \alpha \int_0^1 E_{int}(C(p))dp + \beta \int_0^1 E_{img}(C(p))dp + \gamma \int_0^1 E_{con}(C(p))dp$$

- The internal term stands for regularity/smoothness along the curve and has two components (resisting to stretching and bending)
- The image term guides the active contour towards the desired image properties (strong gradients)
- The external term can be used to account for user-defined constraints, or prior knowledge on the structure to be recovered
- The lowest potential of such a cost function refers to an equilibrium of these terms

Active Contour Components

The internal term…

$$E_{int}(C(p)) = w_{tension}(C(p)) \left| \frac{\partial C}{\partial p}(p) \right|^2 + w_{stiffness}(C(p)) \left| \frac{\partial^2 C}{\partial p^2}(p) \right|^2$$

- The first order derivative makes the snake behave as a membrane
- The second order derivative makes the snake act like a thin plate

The image term…

$$E_{img}(C(p)) = w_{line} E_{line}(C(p)) + w_{edge} E_{edge}(C(p)) + w_{term} E_{term}(C(p))$$

- Can guide the snake to
  - Iso-photes $E_{line}(C(p)) = I(C(p))$ edges $E_{edge}(C(p)) = |\nabla I(C(p))|^2$
  - and terminations

Numerous Provisions…: balloon models, region-snakes, etc…

Optimizing Active Contours

Taking the Euler-Lagrange equations:

$$\alpha\left(w_{tension}\frac{\partial^2 C}{\partial p^2}(p) - w_{stiffness}\frac{\partial^4 C}{\partial p^4}(p)\right) - \beta\nabla E_{img}(C(p) = 0$$

That are used to update the position of an initial curve towards the desired image properties

- Initial the curve, using a certain number of control points as well as a set of basic functions,
- Update the positions of the control points by solving the above equation
- Re-parameterize the evolving contour, and continue the process until convergence of the process…

Pros/Cons of such an approach

Pros

- Low complexity
- Easy to introduce prior knowledge
- Can account for open as well as closed structures
- A well established technique, numerous publications it works
- User Interactivity

Cons

- Selection on the parameter space and the sampling rule affects the final segmentation result
- Estimation of the internal geometric properties of the curve in particular higher order derivatives
- Quite sensitive to the initial conditions,
- Changes of topology (some efforts were done to address the problem)

Level Set: The basic Derivation

The Level Set Method

Osher-Sethian (1987)

- Earlier: Dervieux, Thomassett, (1979, 1980)

Introduced in the area of fluid dynamics

Vision and image segmentation

- Caselles-Catte-coll-Dibos (1992)
- Malladi-Sethian-Vermuri (1994)

Level Set Milestones

- Faugeras-keriven (1998) stereo reconstruction
- Paragios-Deriche (1998), active regions and grouping
- Chan-Vese (1999) mumford-shah variant
- Leventon-Grimson-Faugeras-etal (2000) shape priors
- Zhao-Fedkiew-Osher (2001) computer graphics

The Level Set Method

Let us consider in the most general case the following form of curve propagation:

$$C(p,t) = F(\mathcal{K})\mathcal{N}$$

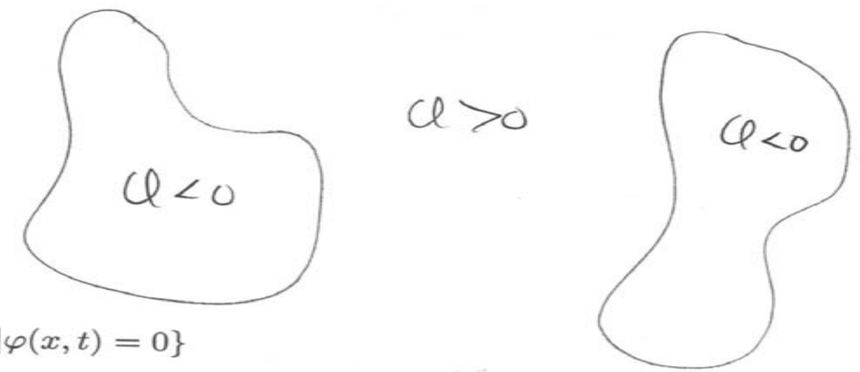
Addressing the problem in a higher dimension…

$$\varphi(x,y,t) : \mathcal{R}^2 \times [0,T) \rightarrow \mathcal{R}$$

The level set method represents the curve in the form of an implicit surface:

That is derived from the initial contour according to the following condition:

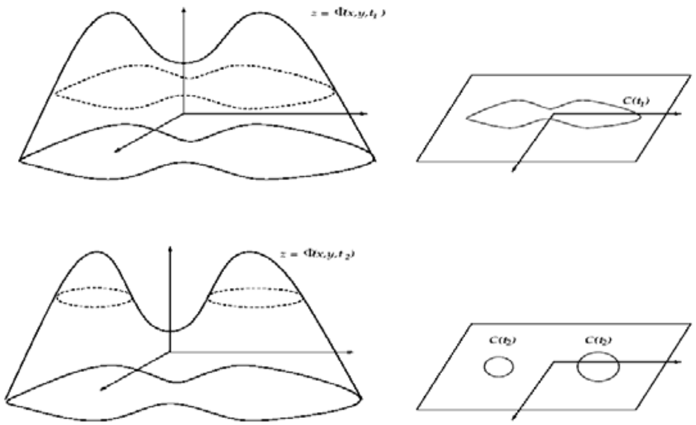
$$C(p,0) = \{(x,y) : \varphi(x,y,0) = 0\}$$



$\{x|\varphi(x,t) = 0\}$ defines $\Gamma(t)$.

The Level Set Method

Construction of the implicit function



$$C(p,0) = \{(x,y) : \varphi(x,y,0) = 0\}$$

$$C(p,t) = \{(x,y) : \varphi(x,y,t) = 0\}\ ,\ C(t) = \varphi^{-1}(0)$$

And taking the derivative with respect to time (using the chain rule)

$$\varphi(C(t),t) = 0 \quad \Rightarrow \quad \frac{\partial \varphi}{\partial C} \cdot \underbrace{\frac{\partial C}{\partial t}}_{F\mathcal{N}} + \underbrace{\frac{\partial \varphi}{\partial t}}_{\varphi_t} = 0 \qquad (1)$$

The Level Set Method

Let us consider the arc-length (c) parameterization of the curve, then taking the directional derivative of $\varphi(C(t), t)$, in that direction we will observe no change:

$$\varphi_s = 0 = \varphi_x x_s + \varphi_y y_s = \langle \nabla\varphi, C_s \rangle$$

leading to the conclusion that the $\nabla\varphi$ s ortho-normal to C where the following expression for the normal vector $\left[\mathcal{N} = -\frac{\nabla\varphi}{|\nabla\varphi|}\right]$

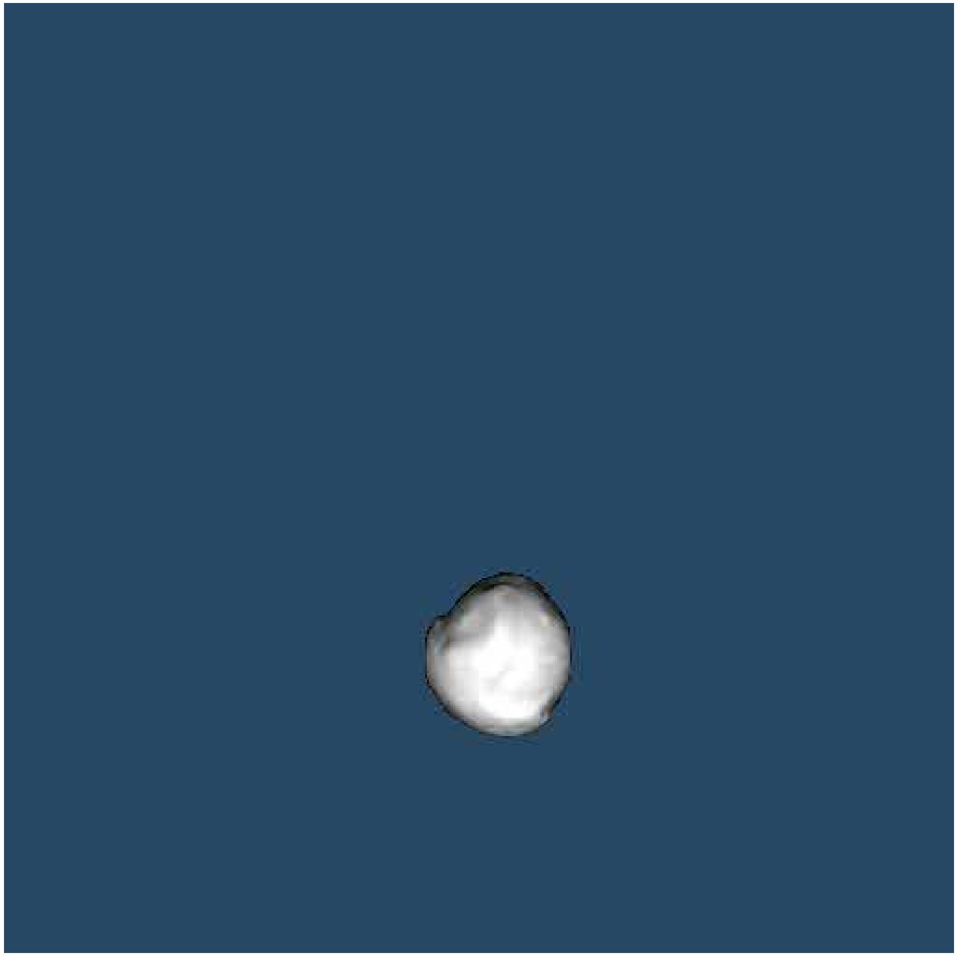
Embedding the expression of the normal vector to:

$$\varphi(C(t), t) = 0 \quad \Rightarrow \quad \frac{\partial\varphi}{\partial C} \cdot \underbrace{\frac{\partial C}{\partial t}}_{F\mathcal{N}} + \underbrace{\frac{\partial\varphi}{\partial t}}_{\varphi_t} = 0$$

the following flow for the implicit function is recovered:

$$\varphi_t = -\langle \nabla\varphi,\ F(\mathcal{K})\mathcal{N} \rangle = -F(\mathcal{K}) \left\langle \nabla\varphi, -\frac{\nabla\varphi}{|\nabla\varphi|} \right\rangle = F(\mathcal{K})\,|\nabla\varphi| \quad (2)$$

Some nice animations



Level Set Method (the basic derivation)

Where a connection between the curve propagation flow and the flow deforming the implicit function was established

Given an initial contour, an implicit function is defined and deformed at each pixel according to the equation (2) where the zero-level set corresponds to the actual position of the curve at a given frame

Euclidean distance transforms are used in most of the cases as embedding function

Geodesic Active Contours

[Caselles-Kimmel-Sapiro:95, Kichenassamy-Kumar-etal95]

Connection between level set methods and snake driven optimization

The geodesic active contour consists of a simplified snake model without second order smoothness

$$E[(C)(p)] = \alpha \int_0^1 \left| \frac{\partial C}{\partial p}(p) \right|^2 dp \; - \; \beta \int_0^1 |\nabla I(C(p))| \, dp$$

That can be written in a more general form as

$$E[(C)(p)] = \alpha \int_0^1 \left| \frac{\partial C}{\partial p}(p) \right|^2 dp \; + \; \beta \int_0^1 g\left(|\nabla I(C(p))|\right)^2 dp$$

Where the image metric has been replaced with a monotonically decreasing function:

$$g(;) = \frac{1}{1 + |\nabla G_s * I(;)|}$$

Geodesic Active Contours

[Caselles-Kimmel-Sapiro:95, Kichenassamy-Kumar-etal95]

Leading to the following more general framework…

$$E[(C)(p)] = \alpha \int_0^1 \left| \frac{\partial C}{\partial p}(p) \right|^2 dp + (1-\alpha) \int_0^1 g\left(|\nabla I(C(p))|\right)^2 dp$$
$$= \int_0^1 \left(E_{int}(C(p)) + E_{ext}(C(p))\right) dp$$

One can assume that smoothness as well as image terms are equally important and with some "basic math"

$$E[(C)(p)] = \underbrace{\int_0^L g\left(|\nabla I(C(s))|\right) ds}_{Geodesic\ Active\ Contour} = \int_0^1 \underbrace{g\left(|\nabla I(C(p))|\right)}_{attraction\ term} \underbrace{\left| \frac{\partial C}{\partial p}(p) \right|}_{regularity\ term} dp$$

That seeks a minimal length geodesic curve attracted by the desired image properties…

Geodesic Active Contours

That when minimized leads to the following geometric flow:

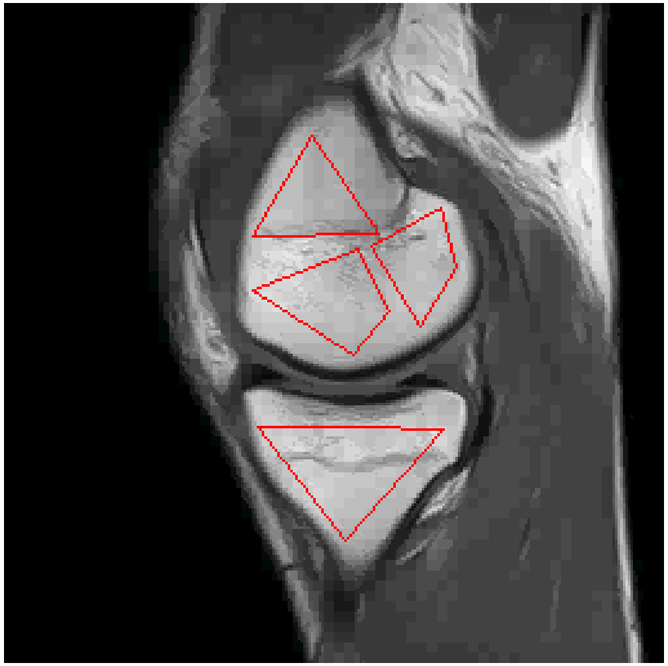
$$\frac{\partial C}{\partial t} = \underbrace{g\left(|\nabla I|\right) \mathcal{K} \mathcal{N}}_{boundary\ \ force} - \underbrace{\left(\nabla g\left(|\nabla I|\right) \cdot \mathcal{N}\right) \mathcal{N}}_{refinement\ \ force}$$

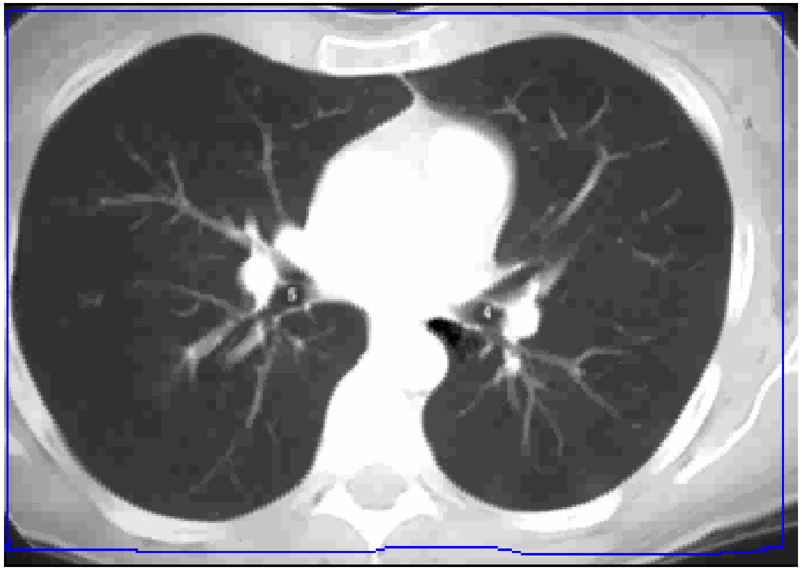
- Data-driven constrained by the curvature force
- Gradient driven term that adjusts the position of the contour when close to the real 0bject boundaries…

By embedding this flow to a level set framework and using a distance transform as implicit function,

$$\phi_t(p)) = g(p)\mathcal{K}(p)|\nabla \phi(p)| + \nabla g(p) \cdot \nabla \phi(p)$$

Some nice animations





Level Sets in imaging and vision…
the region-driven case

The Mumford-Shah framework

[chan-vese:99, yezzi-tsai-willsky-99]

The original Mumford-Shah framework aims at partitioning the image into (multiple) classes according to a minimal length

curve and reconstructing the noisy signal in each class

$$E(C,u) = \alpha \int\int_{\Omega} (I-u)^2 d\omega + \beta \int |C'| dc + \gamma \int\int_{\Omega - C} |\nabla u| d\omega$$

Let us consider - a simplified version - the binary case and the fact that the reconstructed signal is piece-wise constant

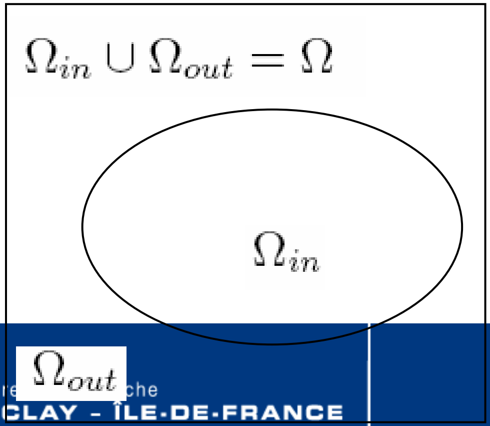
$$E(C,\mu_{in},\mu_{out}) = \alpha \int\int_{\Omega_{in}} (I-\mu_{in})^2 d\omega + \alpha \int\int_{\Omega_{out}} (I-\mu_{out})^2 d\omega + \beta \int |C'| dc$$

Where the objective is to reconstruct

the image, using the mean values for the

inner and the outer region

Tractable problem, numerous solutions…



The Mumford-Shah framework

[chan-vese:99, yezzi-tsai-willsky-99]

Taking the derivatives with respect to piece-wise constants, it straightforward to show that their optimal value corresponds to the means within each region:

$$\mu_{in} = \frac{\int \int_{\Omega_{in}} I(\omega) d\omega}{\int \int_{\Omega_{in}} d\omega}$$

While taking the derivatives with respect and using the stokes theorem, the following flow is recovered for the evolution of the curve:

$$C_t = \alpha \left((I - \mu_{in})^2 - (I - \mu_{out})^2 \right) \mathcal{N} + \beta \mathcal{K} \mathcal{N}$$

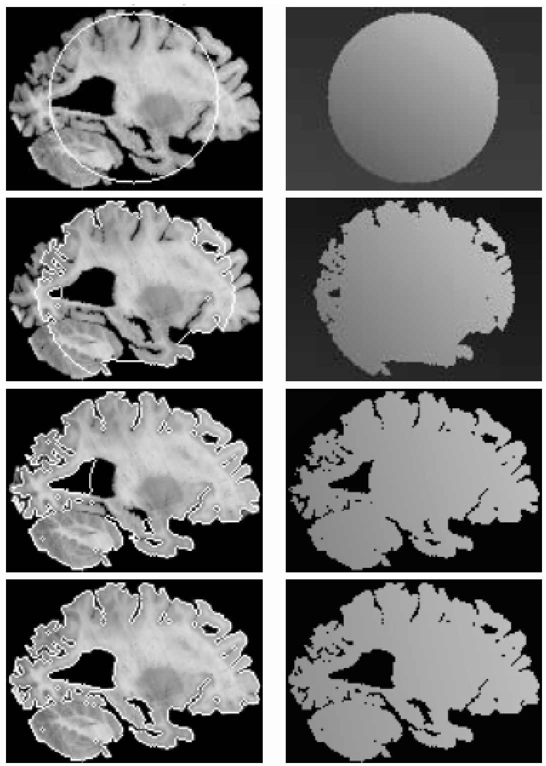
- An adaptive (directional/magnitude)-wise balloon force
- A smoothness force aims at minimizing the length of the partition

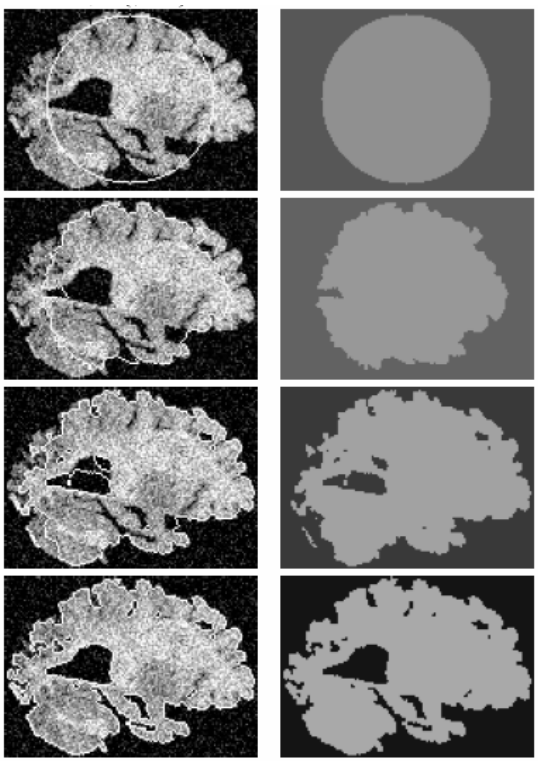
That can be implemented in a straightforward manner within the level set approach

$$\phi_t = \alpha \left((I - \mu_{in})^2 - (I - \mu_{out})^2 \right) |\nabla \phi| + \beta \mathcal{K} |\nabla \phi|$$

The Mumford-Shah framework – Criticism & Results

Account for multiple classes? Quite simplistic model, quite often the means are not a good indicator for the region statistics? Absence of use on the edges, boundary information





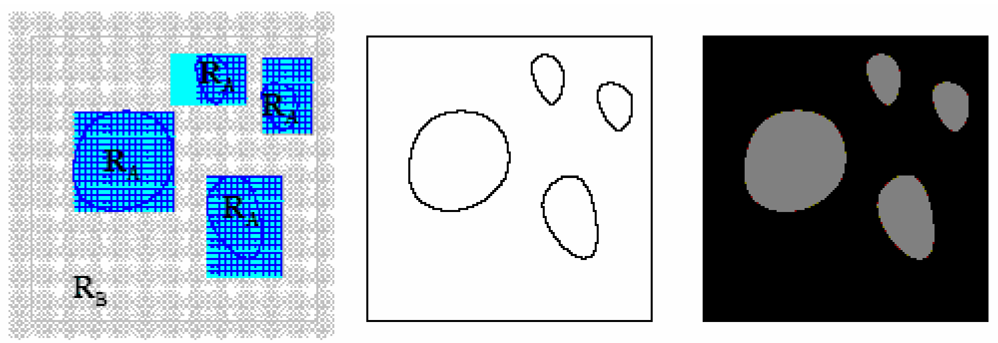
Geodesic Active Regions

[paragios-deriche:98]

Introduce a frame partition paradigm within the level set space that can account for boundary and global region-driven information

KEY ASSUMPTIONS

- Optimize the position and the geometric form of the curve by measuring information along that curve, and within the regions that compose the image partition defined by the curve:



(input image) (boundary) (region)

Geodesic Active Regions

We assume that prior knowledge on the positions of the objects to be recovered is available - $[p_C()]$ - as well as on the expected intensity properties of the object and the background $[\, p_A(), p_B() \,]$

$$E(\partial\mathcal{R}) = \underbrace{\alpha \int_0^1 \overbrace{g\left(\underbrace{p_C(I(\partial\mathcal{R}(c)))}_{boundary\ probability}\right)}^{boundary\ attraction} \overbrace{\left|\partial\dot{\mathcal{R}}(c)\right|}^{regularity} dc}_{Boundary\ Term}$$

$$\underbrace{-(1-\alpha) \overbrace{\iint_{\mathcal{R}_A} \log\left[\underbrace{p_A(I(x,y))}_{h_A\ probability}\right] dxdy}^{\mathcal{R}_A\ fitting\ measurement} -(1-\alpha) \overbrace{\iint_{\mathcal{R}_B} \log\left[\underbrace{p_B(I(x,y))}_{h_B\ probability}\right] dxdy}^{\mathcal{R}_B\ fitting\ measurement}}_{Region\ Term}$$

Geodesic Active Regions

Such a cost function consists of:

- The geodesic active contour
- A region-driven partition module that aims at separating the intensities properties of the two classes (see later analogy with the Mumford-Shah)

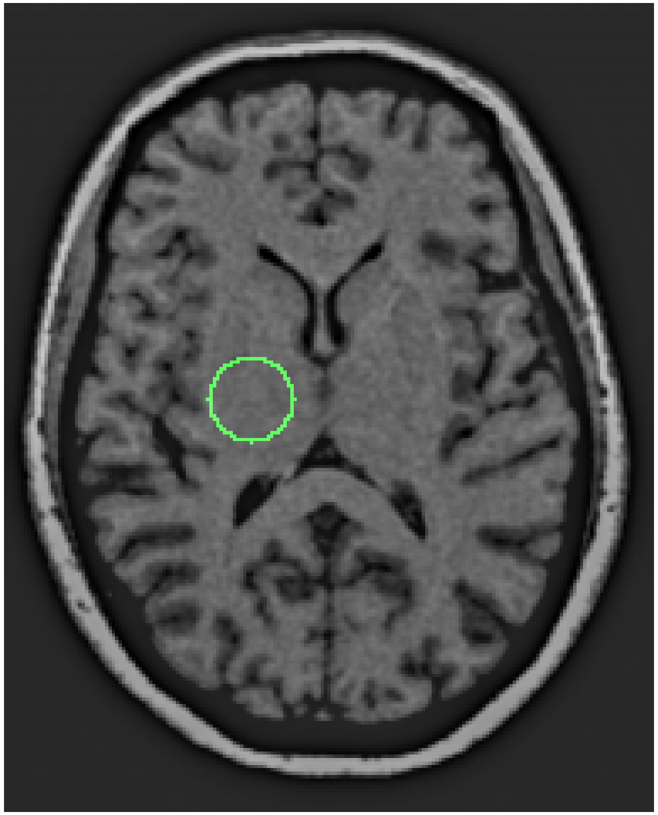
And can be minimized using a gradient descent method leading to:

$$\frac{\partial}{\partial t}u = \left[\underbrace{\alpha \mathbf{log} \left(\frac{\overbrace{p_B(I(u))}^{h_B\ probability}}{\underbrace{p_A(I(u))}_{h_A\ probability}} \right)}_{region-based\ force} + \underbrace{(1-\alpha)\left(g(p_C(I(u)))\mathcal{K}(u) - \nabla g(p_C(I(u))) \cdot \mathcal{N}(u)\right)}_{boundary-based\ force} \right] \mathcal{N}(u)$$

Which can be implemented using the level set method as follows…

$$\phi_t = \alpha \log \left(\frac{p_B(I)}{p_A(I)} \right) |\nabla \phi| + (1-\alpha)\left(g\mathcal{K}|\nabla\phi| + \nabla\phi\nabla g\right)$$

Some nice animations



…REMINDER…

Level Set & Geometric Flows

While evolving moving interfaces with the level set method is quite attracting, still it has the limitation of being a static approach

- The motion equations are derived somehow,
- The level set is used only as an implementation tool…

$$C(p,t) = F(\mathcal{K})\mathcal{N}$$

$$\varphi_t = -\langle \nabla\varphi,\ F(\mathcal{K})\mathcal{N} \rangle = -F(\mathcal{K}) \left\langle \nabla\varphi, -\frac{\nabla\varphi}{|\nabla\varphi|} \right\rangle = F(\mathcal{K})\,|\nabla\varphi|$$

- That is equivalent with saying that the problem has been somehow already solved…since there is not direct connection between the approach and the level set methodology

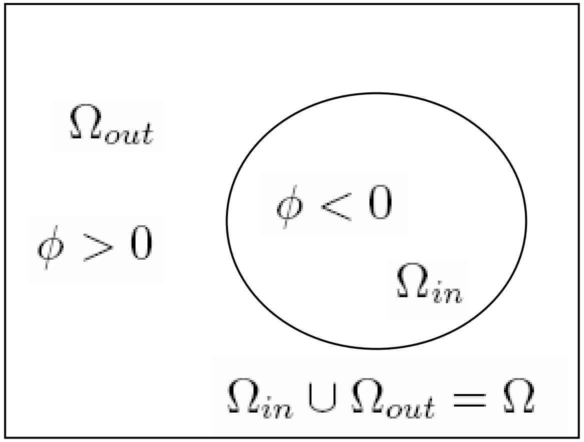
Level Set: Optimization space

Level Set Dictionary

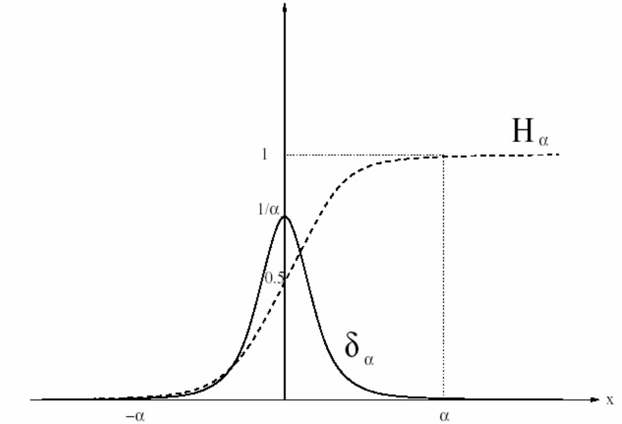
Let us consider distance transforms

as embedding function

$$\phi(s) = \begin{cases} -D(s,C), & s \in \Omega_{in} \\ 0), & s \in C \\ -D(s,C), & s \in \Omega_{out} \end{cases}$$



One can introduce the Dirac distribution



$$\delta_\alpha(\phi) = \begin{cases} 0 & , |\phi| > \alpha \\ \frac{1}{2\alpha}\left(1 + \cos\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}$$

$$H_\alpha(\phi) = \begin{cases} 1 & , \phi > \alpha \\ 0 & , \phi < -\alpha \\ \frac{1}{2}\left(1 + \frac{\phi}{\alpha} + \frac{1}{\pi}\sin\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}$$

Level Set Dictionary

Using the Dirac function and integrating within the image domain, one can estimate the length of the curve:

$$|C| = \int \int \delta_\alpha(\phi)|\nabla \phi| d\Omega$$

While integrating the Heaviside Distribution within the image domain

$$|\Omega_{in}| = \int \int H_\alpha(\phi) d\Omega$$

Such observations can be used to define regional partition modules as follows according to some descriptors

$$E_{regional}(\phi) = \underbrace{\int \int_\Omega H_\alpha(\phi) r_O(;)\ d\Omega}_{object} + \underbrace{\int \int_\Omega (1 - H_\alpha(\phi)) r_B(;)\ d\Omega}_{background}$$

That can be optimized with respect to the level set function (implicitly with respect to a curve position)

Level Set Optimization

$$\frac{\partial}{\partial \phi} E_{regional} = \frac{\partial}{\partial \phi} \int\int H_\alpha(-\phi) r_O d\Omega + \frac{\partial}{\partial \phi} \int\int (1 - H_\alpha(-\phi)) r_B d\Omega$$

$$= \frac{\partial H_\alpha(-\phi)}{\partial \phi} r_O + H_\alpha(-\phi) \frac{\partial r_O}{\partial \phi} + \frac{\partial (1 - H_\alpha(-\phi))}{\partial \phi} r_B + (1 - H_\alpha(-\phi)) \frac{\partial r_B}{\partial \phi}$$

And given that :

$$\frac{\partial H_\alpha(-\phi)}{\partial \phi} = -\delta(-\phi), \quad \frac{\partial r_B}{\partial \phi} = \frac{\partial r_O}{\partial \phi} = 0$$

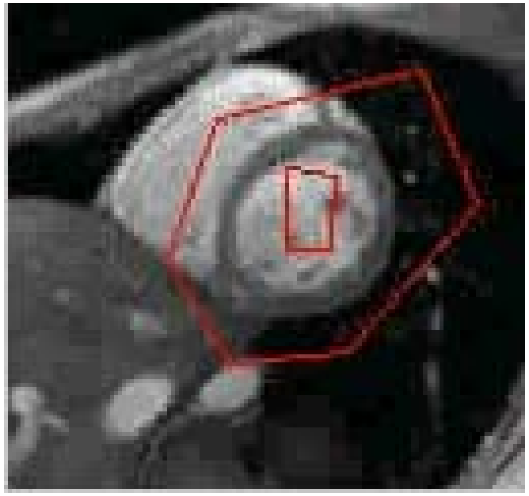
An adaptive (directional & magnitude wise) flow is recovered for the propagation of an initial surface towards a partition that is optimal according to the regional descriptors…

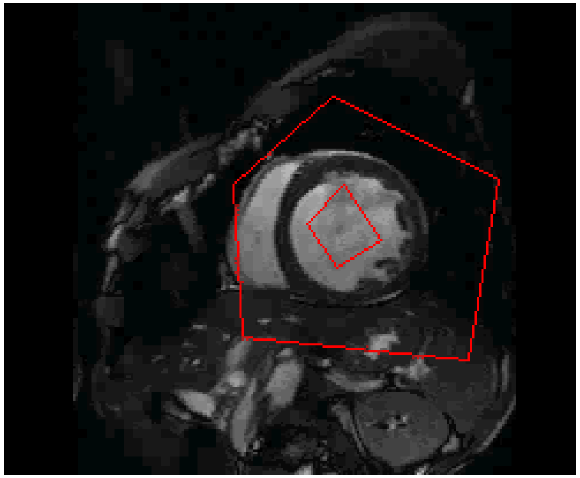
$$\frac{\partial}{\partial \tau} \phi = \delta_\alpha(\phi)(r_B(;) - r_O(;))$$

The same idea can be used to introduce contour-driven terms…

$$E_{geodesic}(\phi) = \int\int_\Omega \delta_\alpha(\phi) b(;) |\nabla \phi| \, d\Omega$$

Some nice animations





Level Set Optimization

and optimize them directly on the level set space

Curve-driven terms:

$$\frac{\partial}{\partial \tau}\phi = \delta_\alpha(\phi)\mathrm{div}\left(b(;)\frac{\nabla\phi}{|\nabla\phi|}\right)$$

$$\frac{\partial}{\partial \tau}\phi = \delta_\alpha(\phi)(r_B(;) - r_O(;))$$

Global region-driven terms:

According to some image metrics…defined along the curve and within the regions obtained through the image partition according to the position of the curve, that can be multi-component but is representing only one class

Multi-Phase Motion

PROS

- Taking the level set method to another level
- Dealing with multiple (multi-component) objects, and multiple tasks
- Introducing interactions between shape structures that evolve in parallel

CONS

- Computationally expensive
- Difficult to guarantee convergence
- Numerically unstable & hard to implement
- Prior knowledge required on the number of classes and in some cases on their properties…

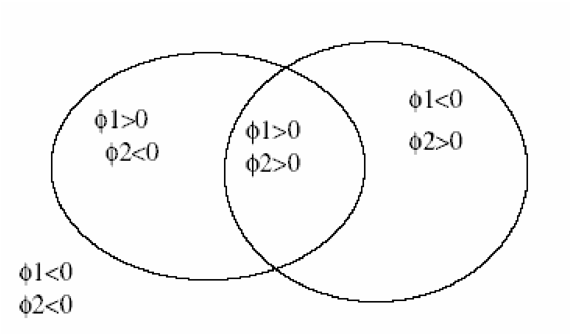
PARTIAL SOLUTION: The multi-phase Chan-Vese model

Multi-Phase Motion [vese-chan:02]

Introduce classification according to a combination of all level sets at a given pixel

LEVEL SET DICTIONARY

- Class 1: $H_\alpha(\phi_1)H_\alpha(\phi_2) = 1$
- Class 2: $H_\alpha(\phi_1)(1 - H_\alpha(\phi_2)) = 1$
- Class 3: $(1 - H_\alpha(\phi_1))H_\alpha(\phi_2) = 1$
- Class 4: $(1 - H_\alpha(\phi_1))(1 - H_\alpha(\phi_2)) = 1$



And therefore by taking these products one can define a modified version of the mumford-shah approach to account with four classes while using two level set functions…

$$E(\phi_1, \phi_2, \mu) = \int\int H_\alpha(\phi_1)H_\alpha(\phi_2)(I-\mu_{11})^2 + \int\int H_\alpha(\phi_1)(1-H_\alpha(\phi_2))(I-\mu_{12})^2$$

$$\int\int(1-H_\alpha(\phi_1))H_\alpha(\phi_2)(I-\mu_{21})^2 + \int\int(1-H_\alpha(\phi_1))(1-H_\alpha(\phi_2))(I-\mu_{22})^2 + w\int\int \delta_\alpha(\phi_1)$$

Multi-Phase Motion

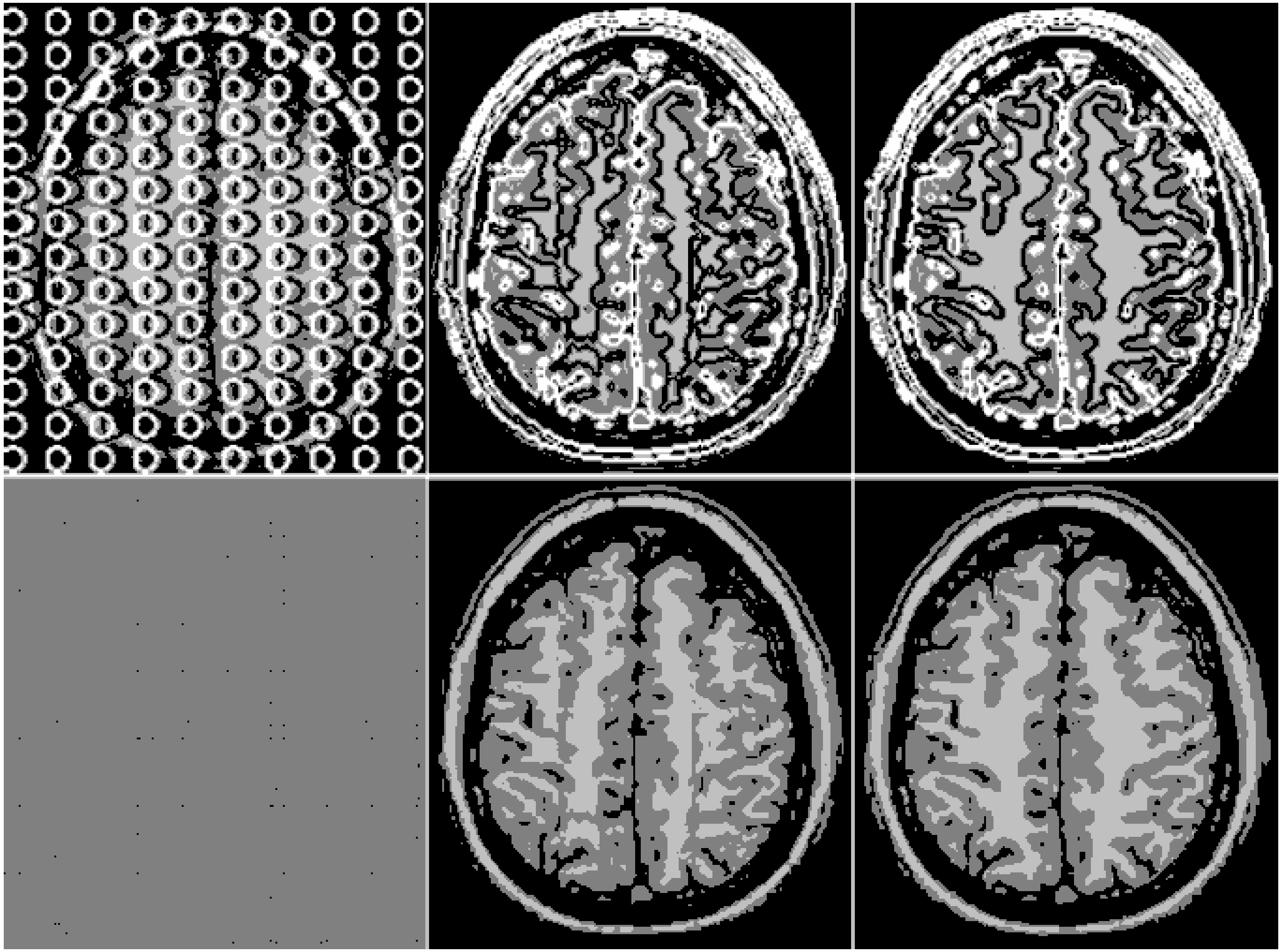
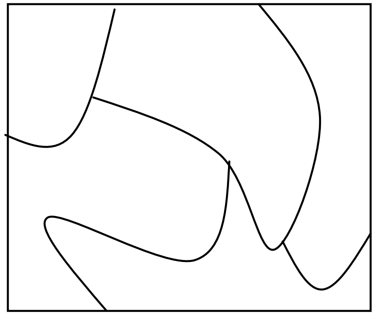


Image Segmentation in the form of a labeling problem



- Let us consider an image composed of m classes

$$\{\omega_1, \omega_2, ..., \omega_m\}$$

- Segmentation consists of finding a partition

$$\{\Omega_1, \Omega_2, ..., \Omega_m\},\ \mathcal{R} = \cup\Omega_i; \Omega_i \cup \Omega_j = \emptyset$$

- That separates this classes…if we assume that each class has a characteristic function; or some measure that can account if a given observation belong to this class;

$$p(f|\omega_k)$$

$$argmax \sum \Pi_{\Omega_i} p(f|\omega_i)$$

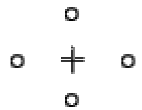
- That consists of associating to each pixel the class that is best supported by the observation;

Markov Random Fields

See segmentation in a probabilistic manner

Let Ω be the image domain, $\mathcal{L}$ be a discrete set of labels (random variables) that are associated with this domain and $\mathcal{I}$ the information space associated with that can either be the image itself or high-dimensional data recovered from the image after applying certain filter operators…

Let us also consider a discrete neighborhood system $\mathcal{V}$ living on Ω where each pixel is associated with its for direct neighbors:



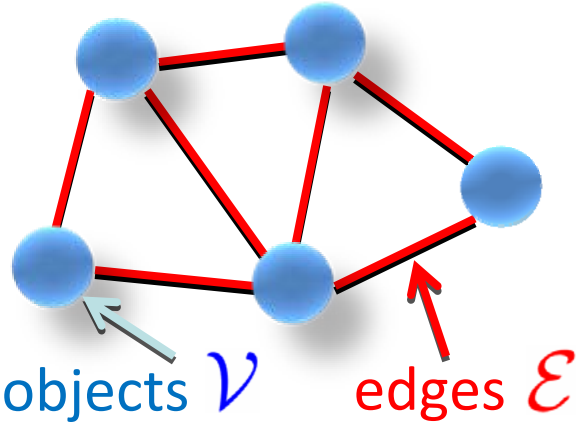
Then image segmentation into a certain number of classes is equivalent with associating each element of the domain Ω to a label from the set $\mathcal{L}$

Such a decision should be supported by the data $\mathcal{I}$ and has to be locally consistent (neighborhood pixels should be assigned mostly the same label)

Discrete MRF optimization

Given:

- Objects $\mathcal{V}$ from a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- Discrete label set $\mathcal{L}$



- Assign labels (to objects) that minimize MRF energy:

$$\min_{\{x_p\}} \sum_{p \in \mathcal{V}} \underbrace{\bar{g}_p(x_p)}_{\text{unary potential}} + \sum_{pq \in \mathcal{E}} \underbrace{\bar{f}_{pq}(x_p, x_q)}_{\text{pairwise potential}}$$

- MRF optimization ubiquitous in vision (and beyond)

MRFs and Optimization

- Deterministic Methods:
  Iterated Conditional Modes/Highest Confidence First
- Non-Deterministic Methods:
  Mean-field and Simulating Annealing, etc
- Graph-cut based techniques such as *α*-expansion:
  Min cut/max flow, etc
- Message-passing techniques:
  Belief Propagation Networks generalized by TRW methods

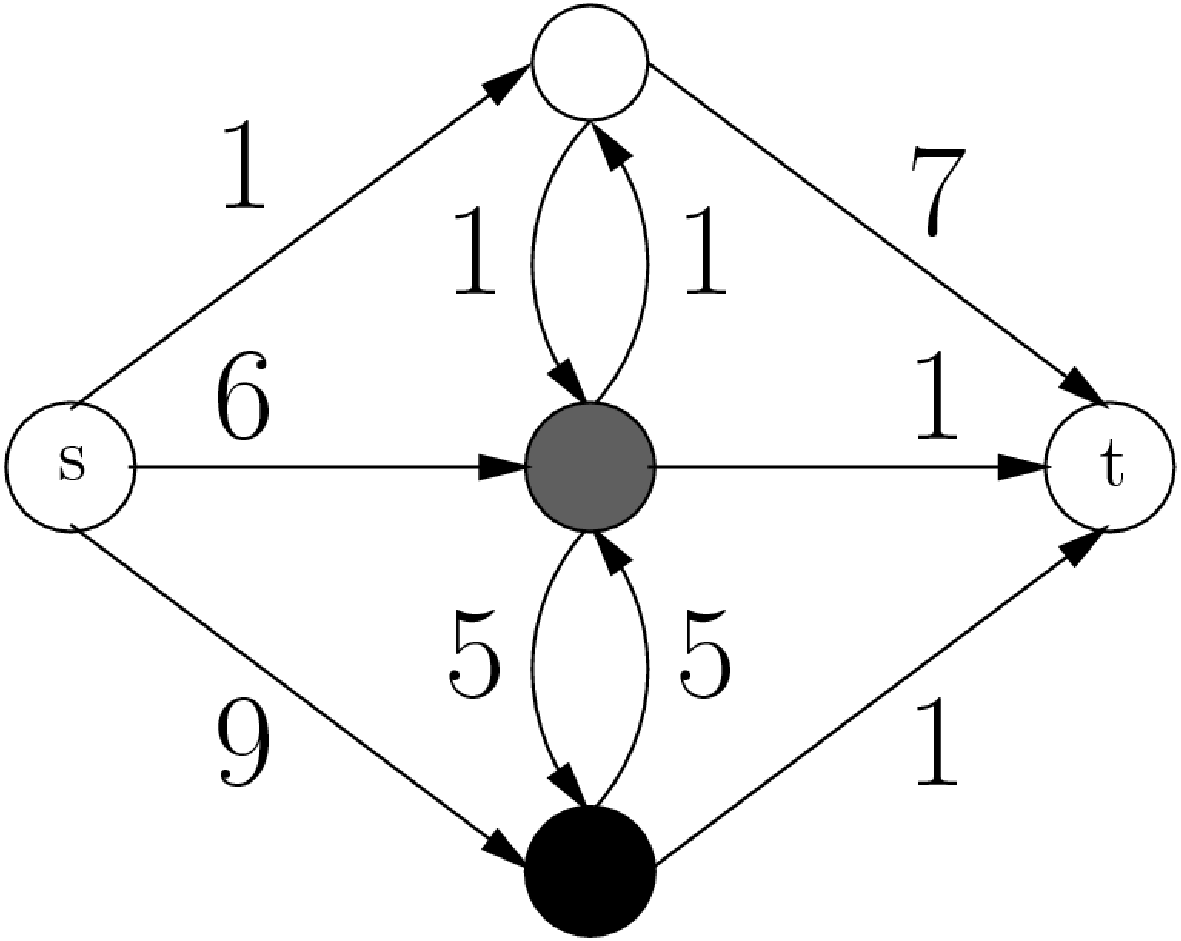
- The above statement is more or less true for almost all state-of-the-art MRF techniques

Max-Flow/Min-Cut Theorem

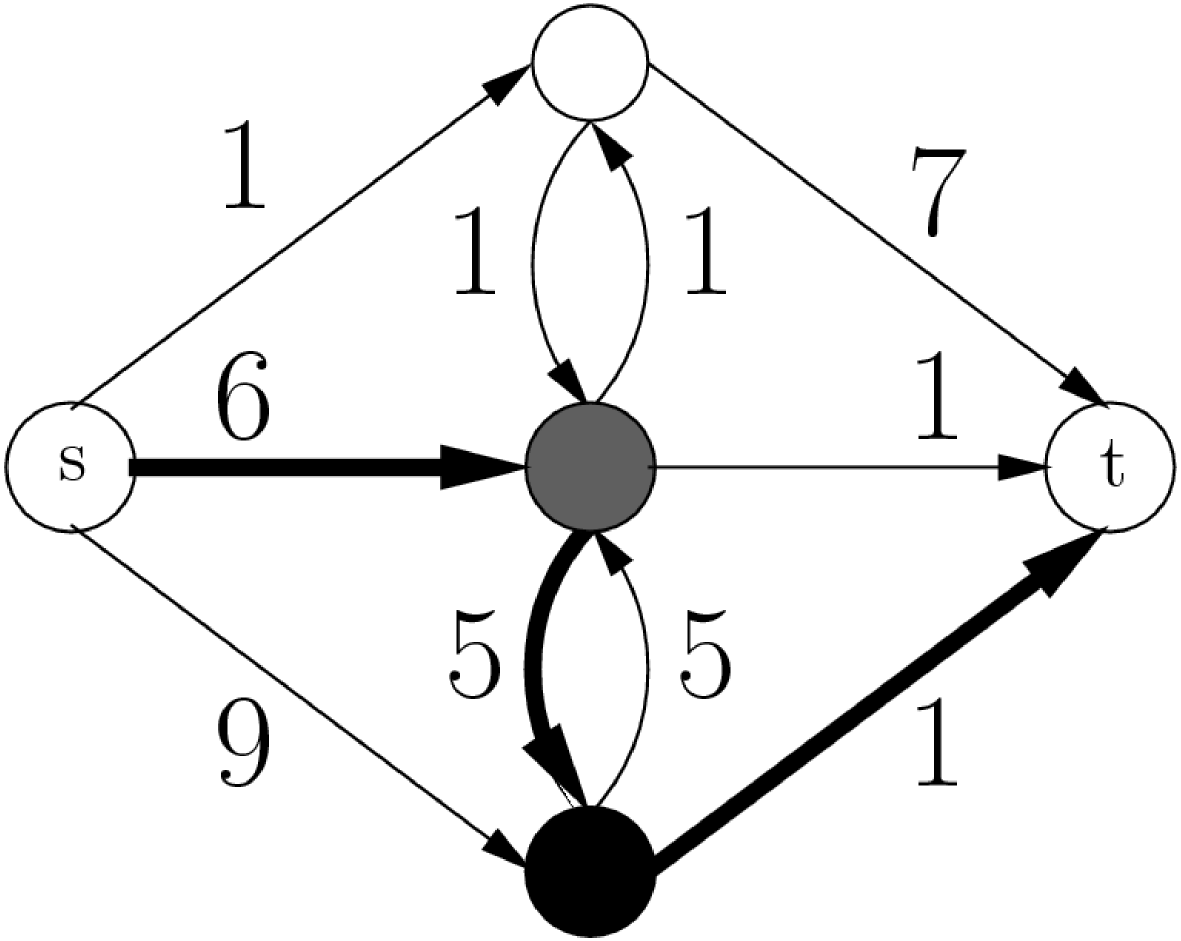
For any network having a single origin and a single destination node, the maximum possible flow from origin to destination equals the minimum cut value for all the cuts in the network.

If we want to find the minimum cut, we can compute the maximum flow and look for the cut(s) that separate origin and destination by cutting through bottlenecks of the network.

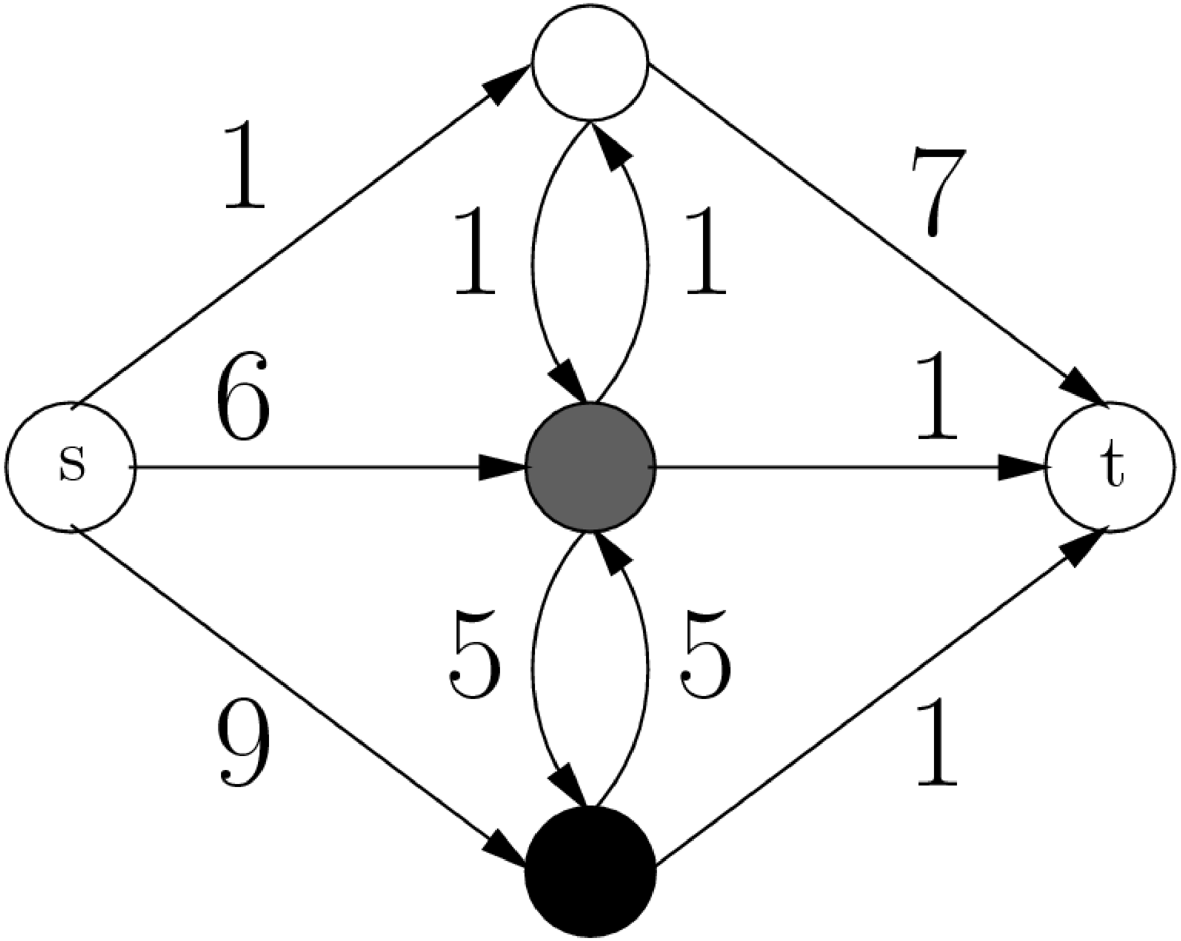
Example: Ford & Fulkerson



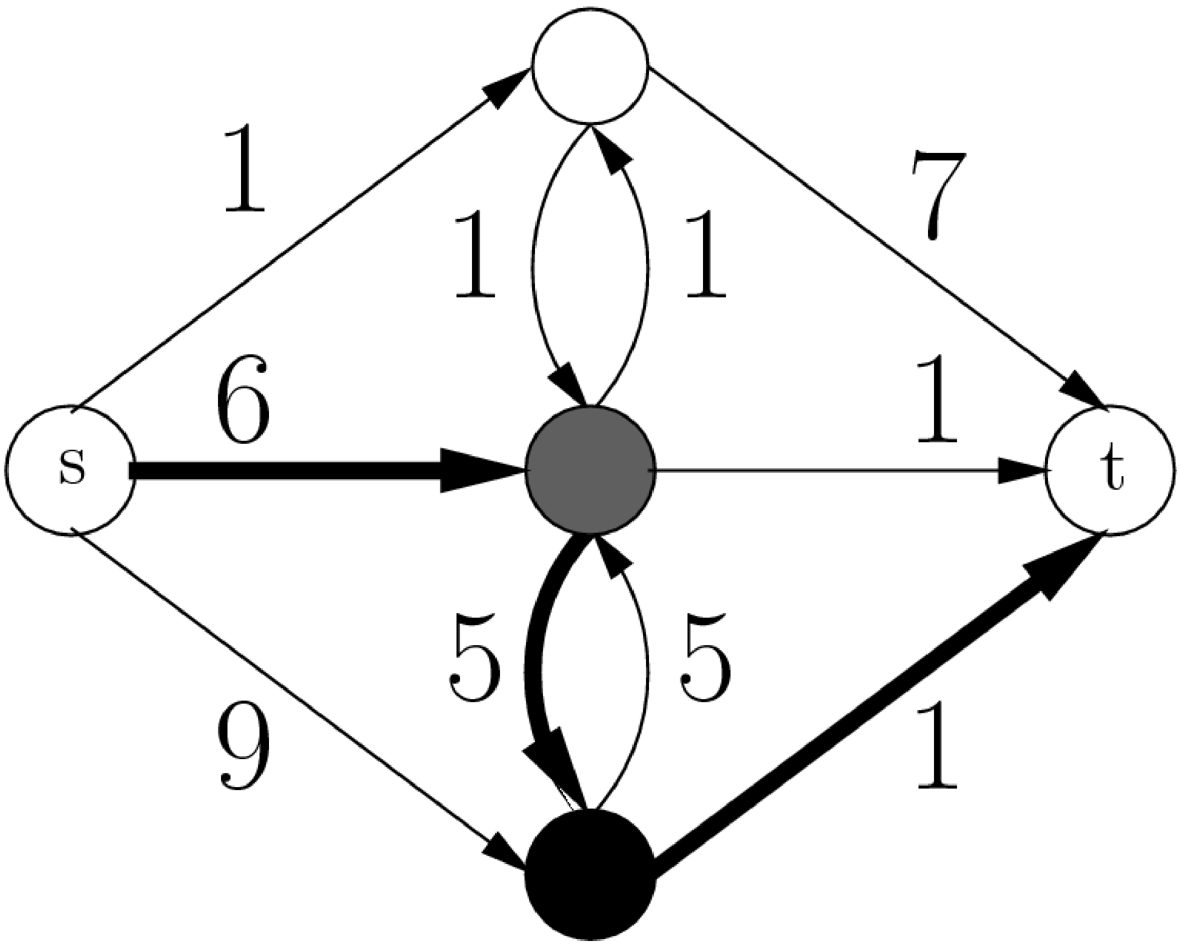
Example: Ford & Fulkerson



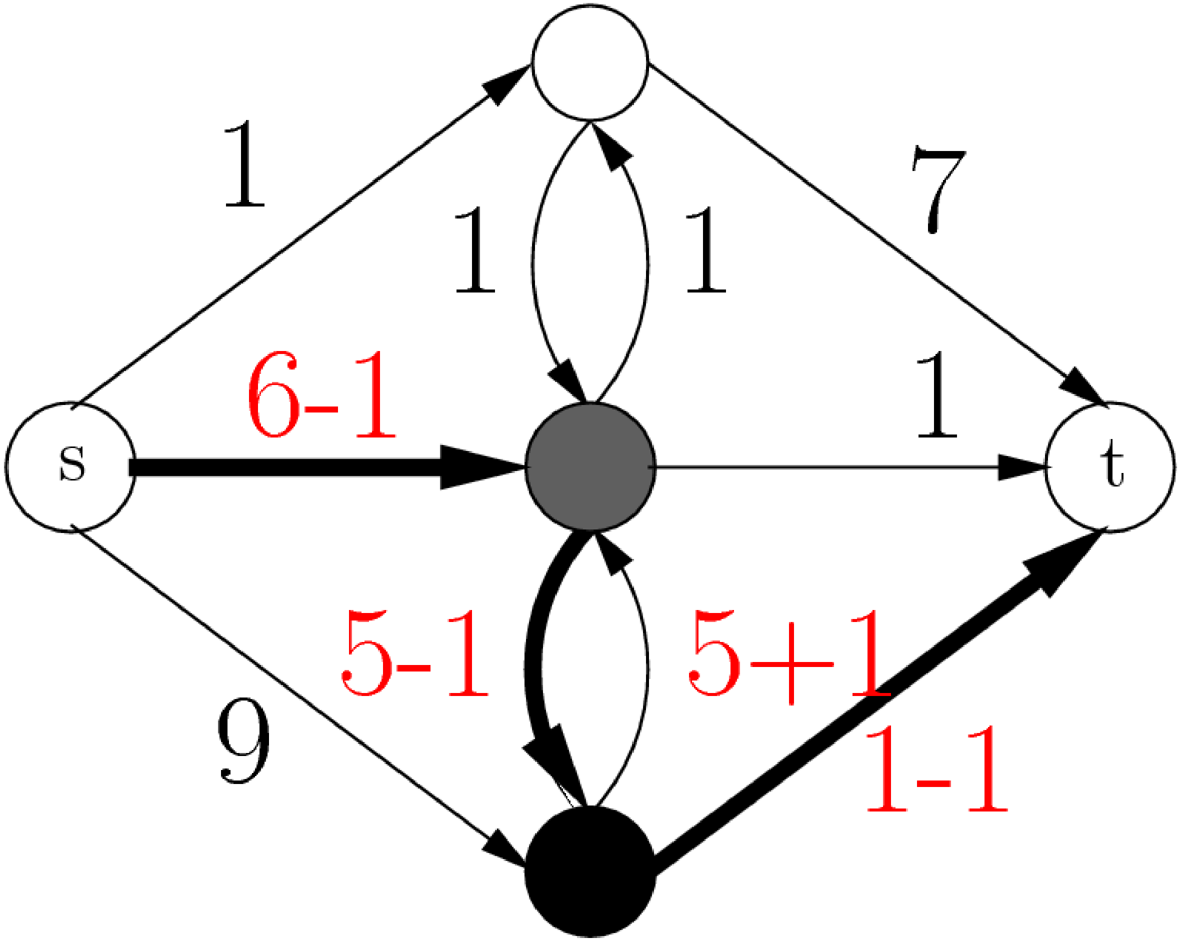
Example: Ford & Fulkerson



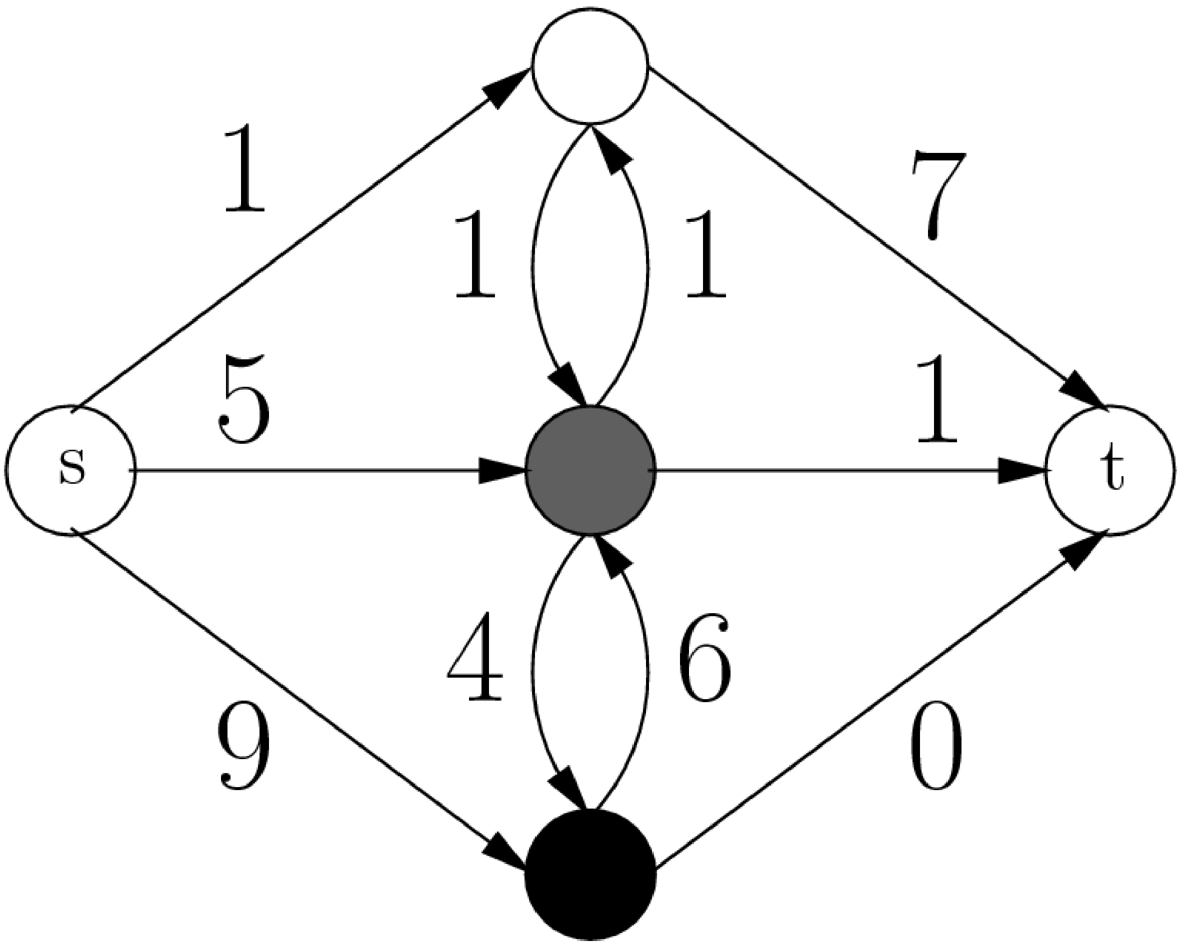
Example: Ford & Fulkerson



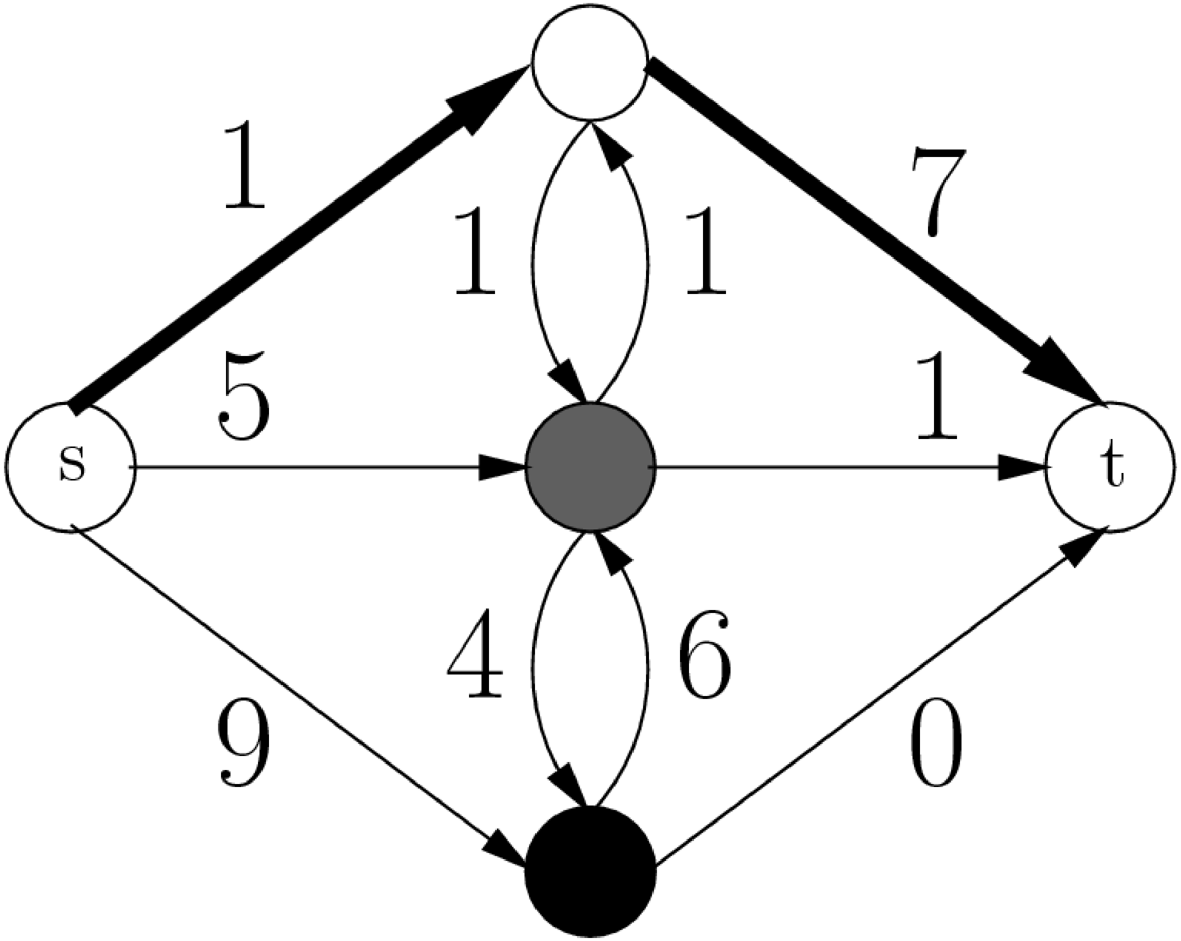
Example: Ford & Fulkerson



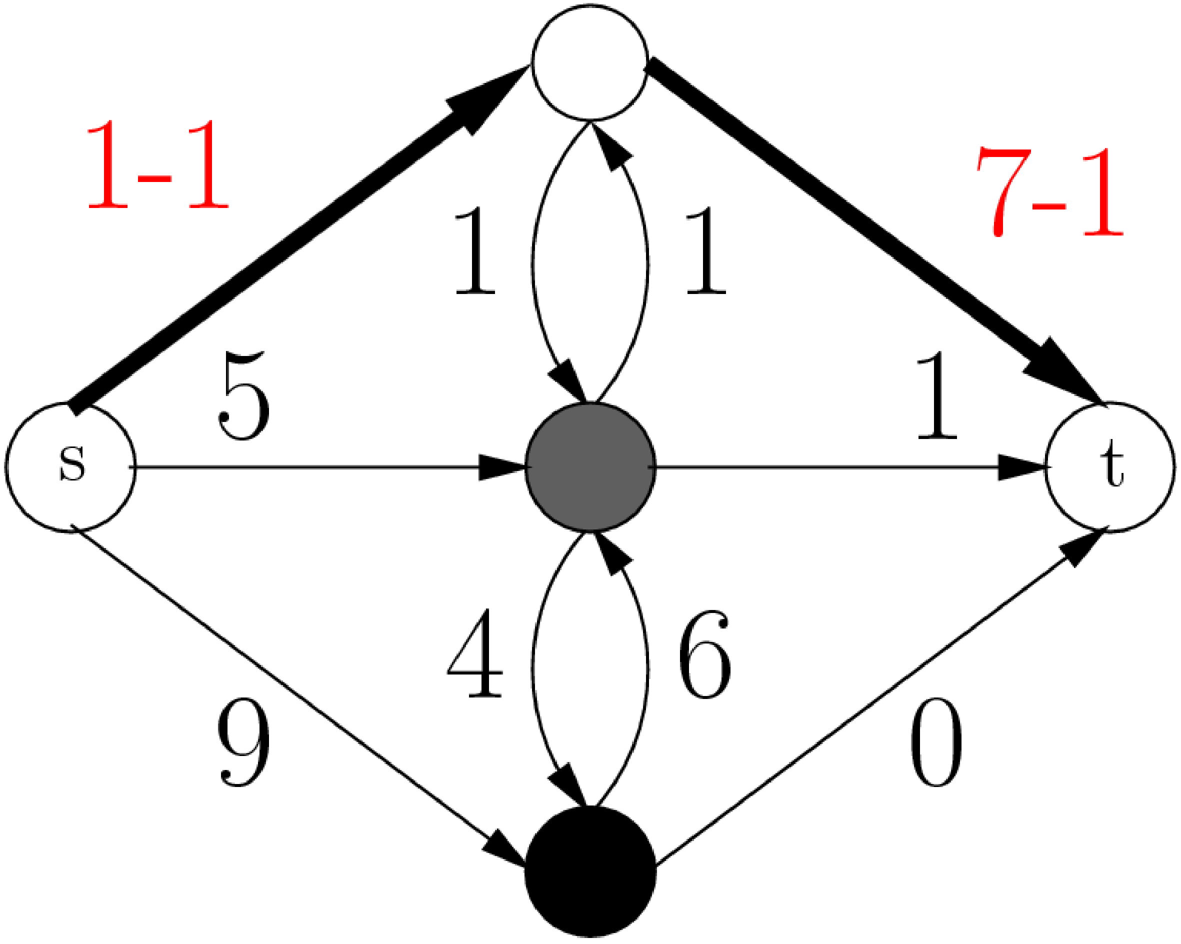
Example: Ford & Fulkerson



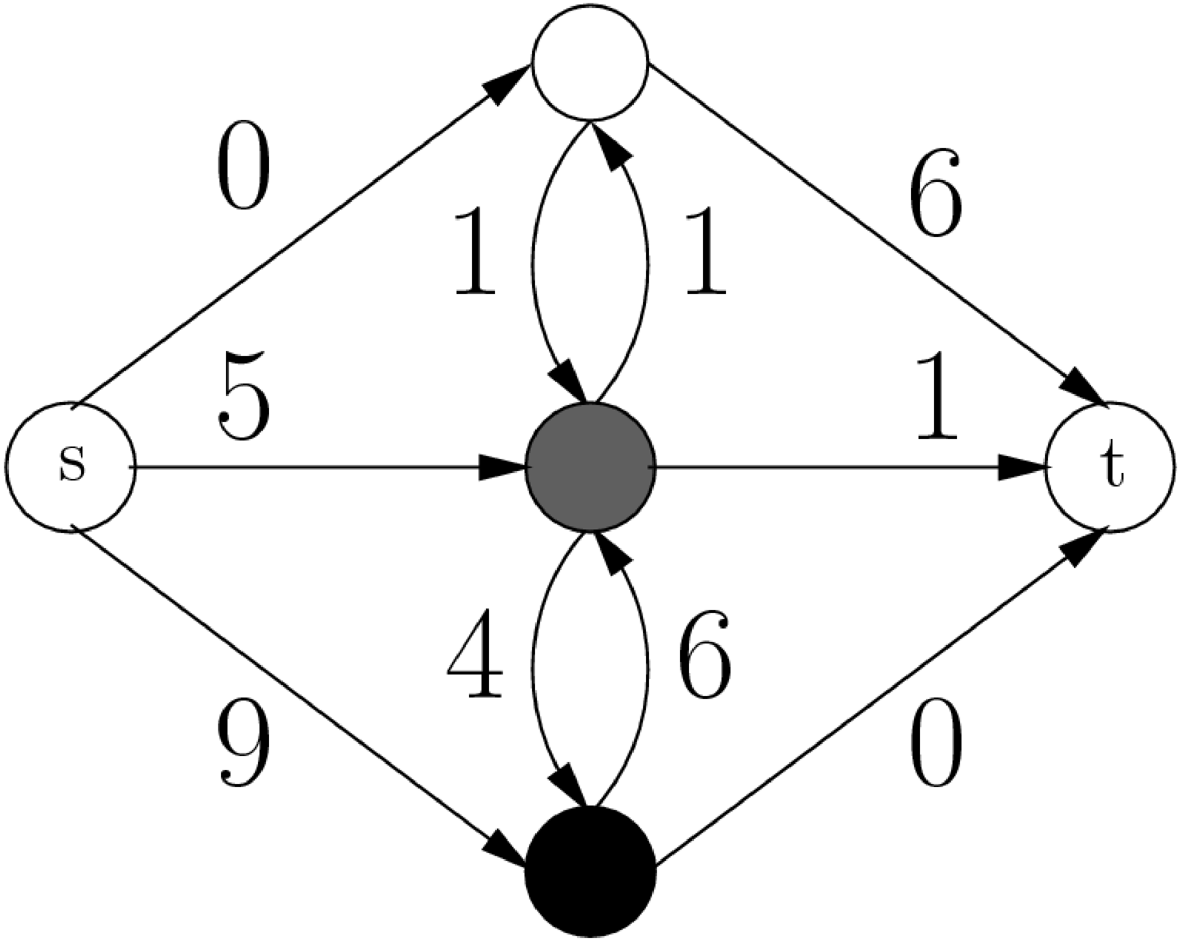
Example: Ford & Fulkerson



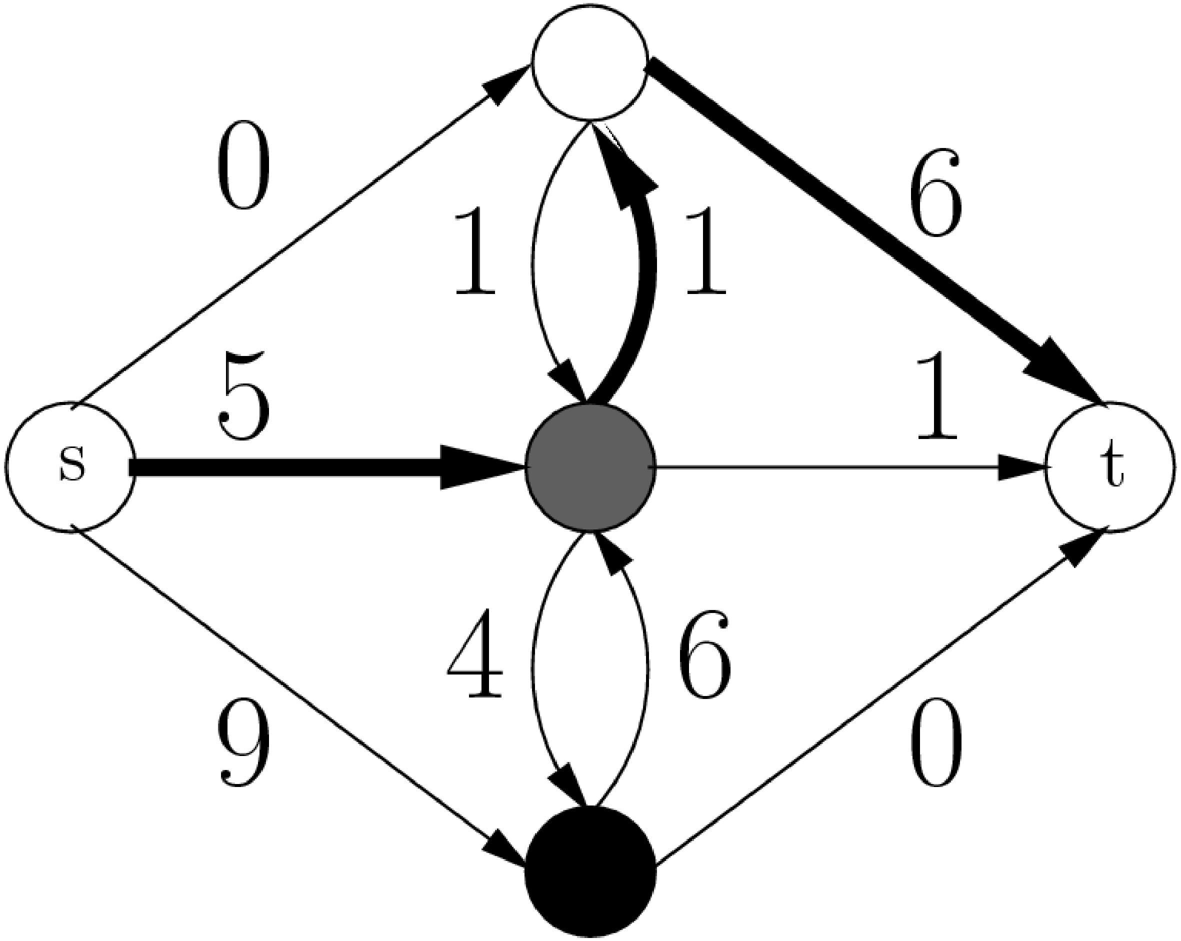
Example: Ford & Fulkerson



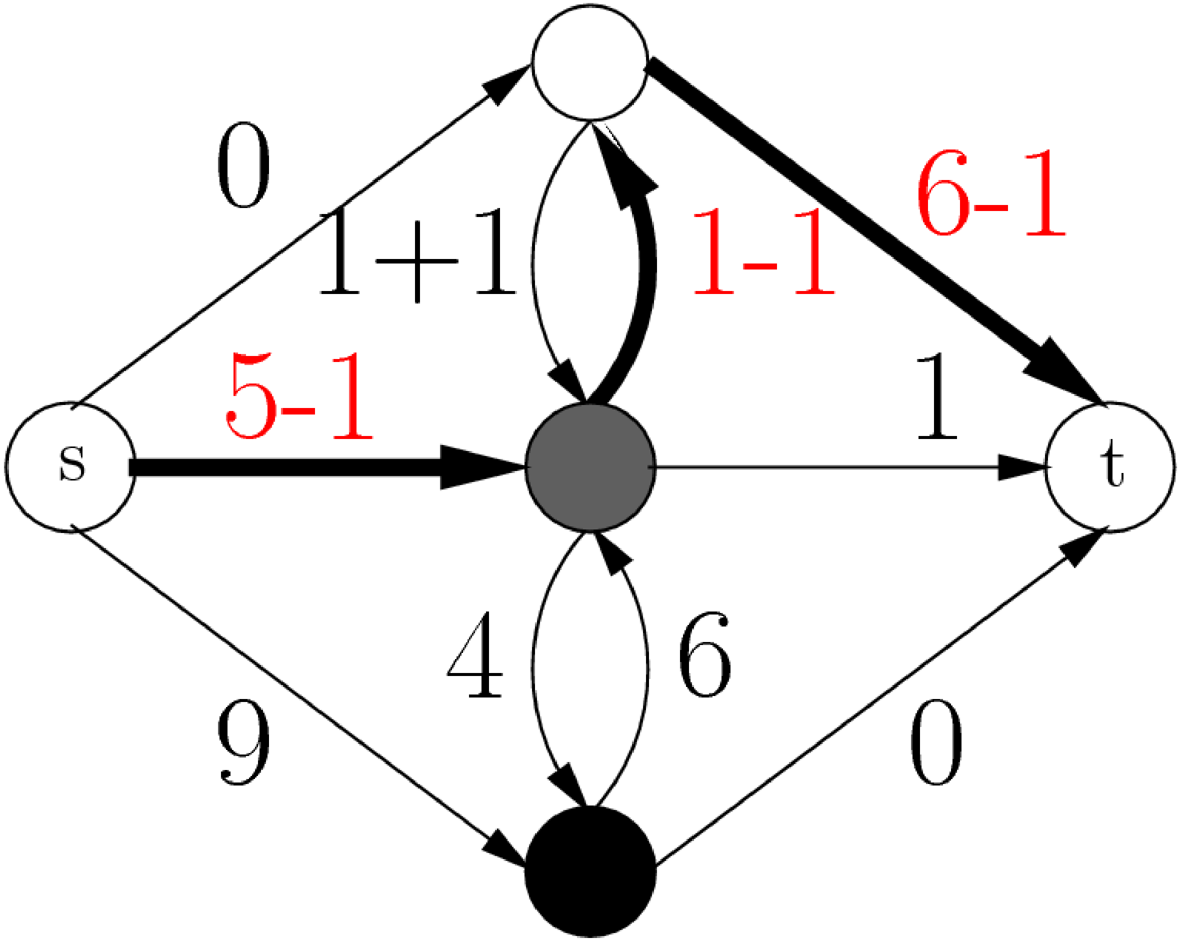
Example: Ford & Fulkerson



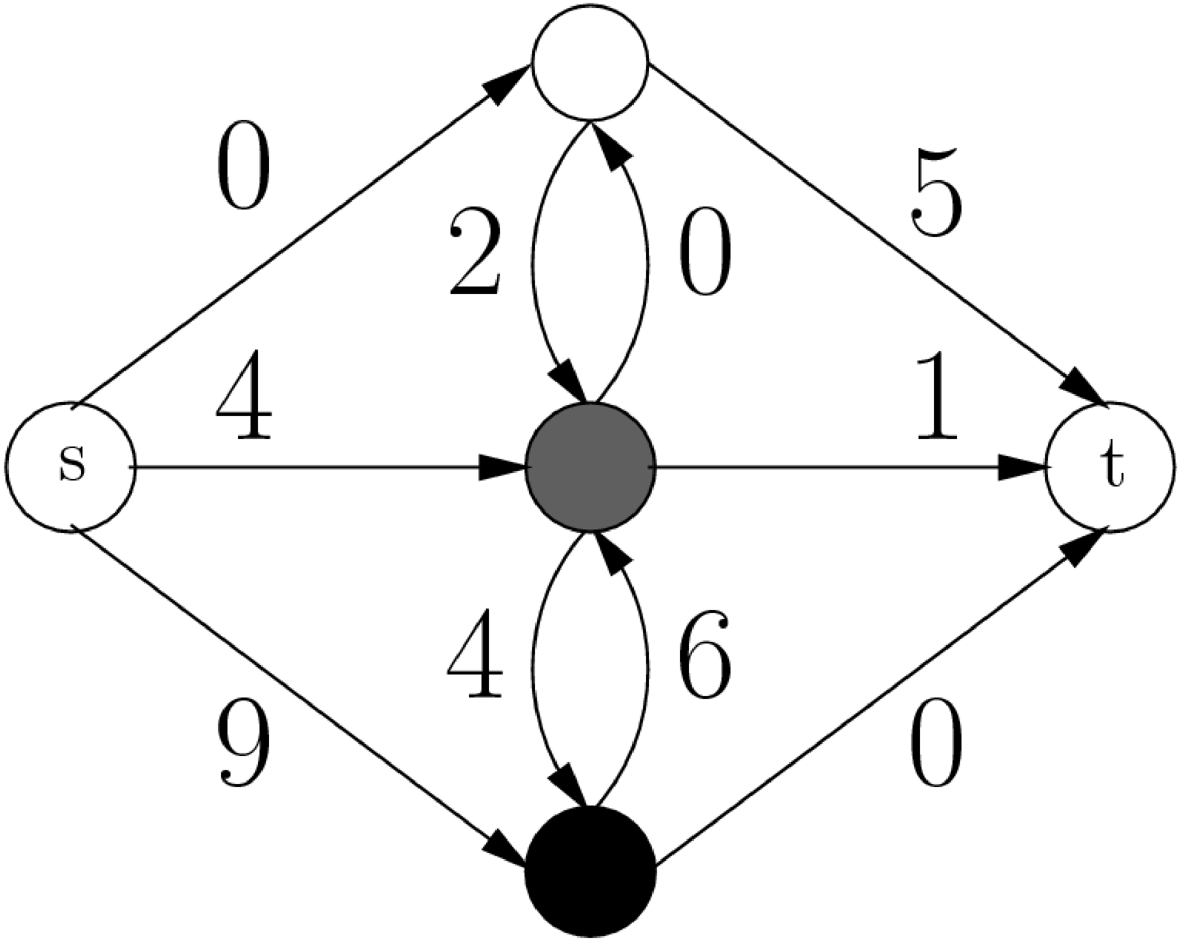
Example: Ford & Fulkerson



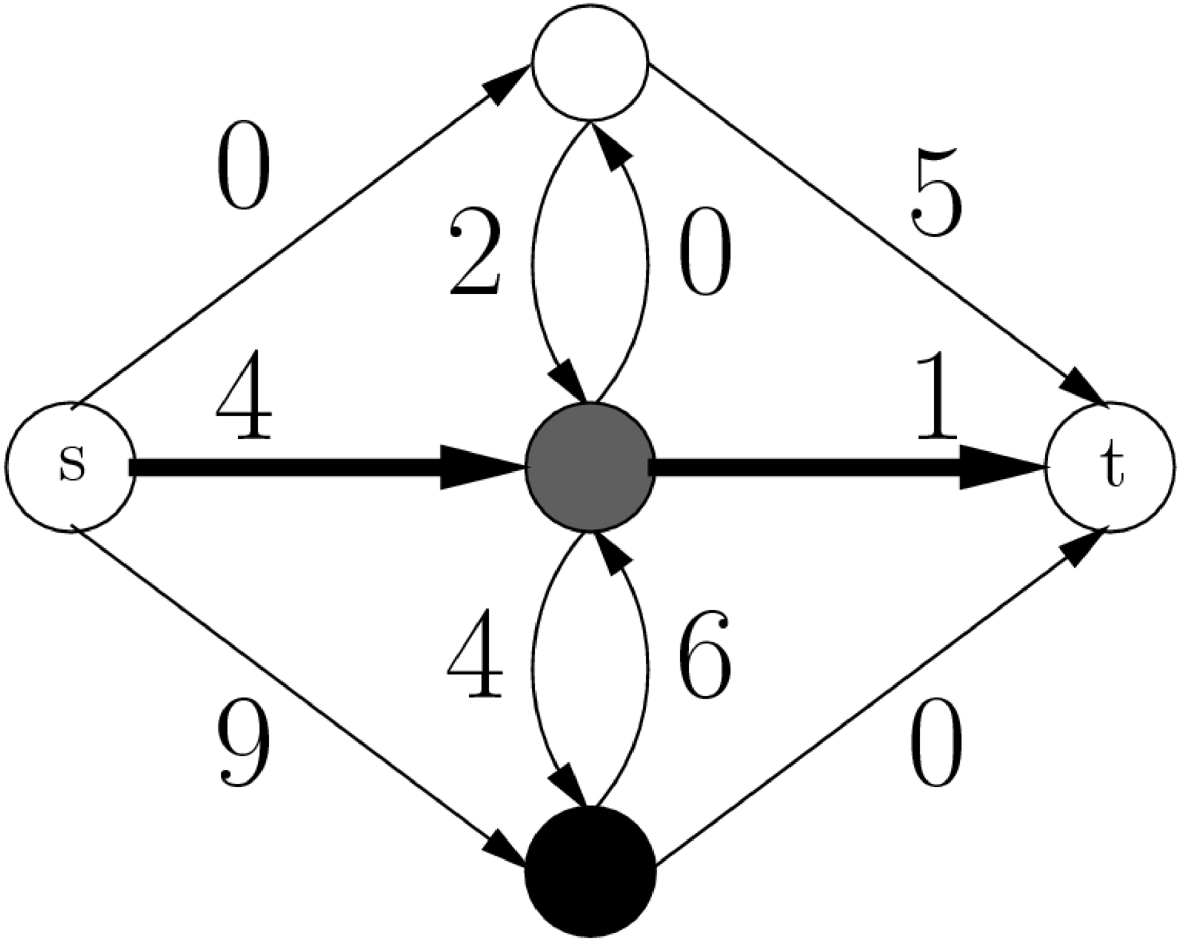
Example: Ford & Fulkerson



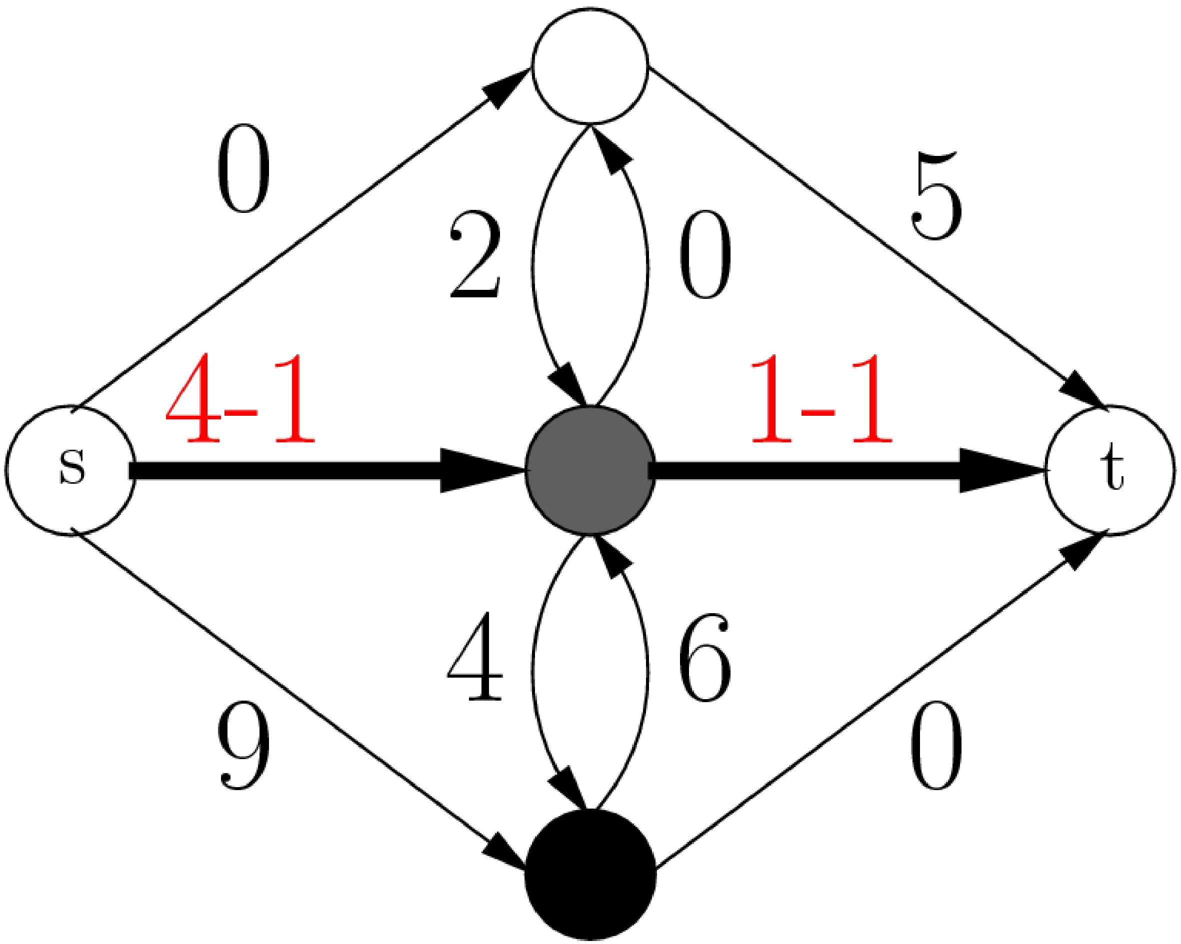
Example: Ford & Fulkerson



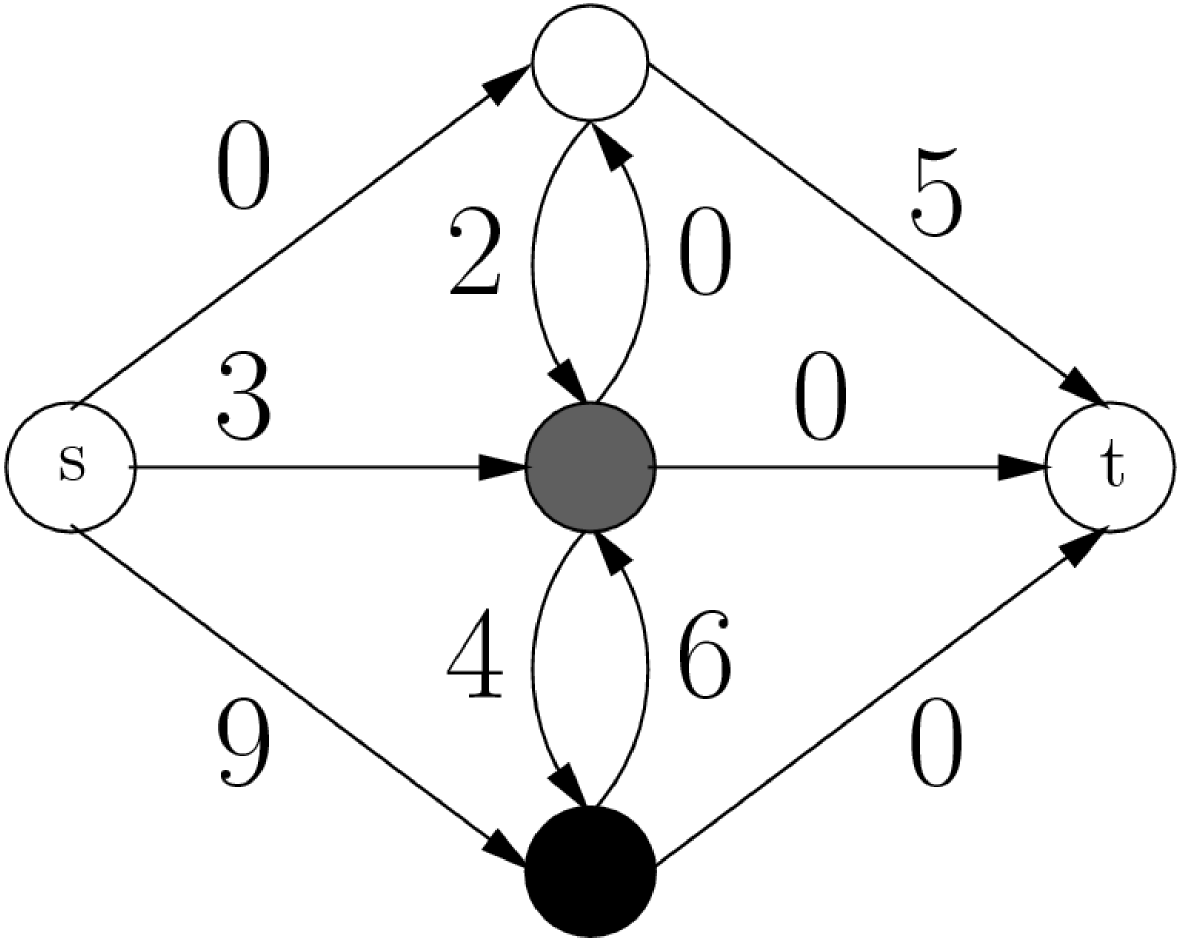
Example: Ford & Fulkerson



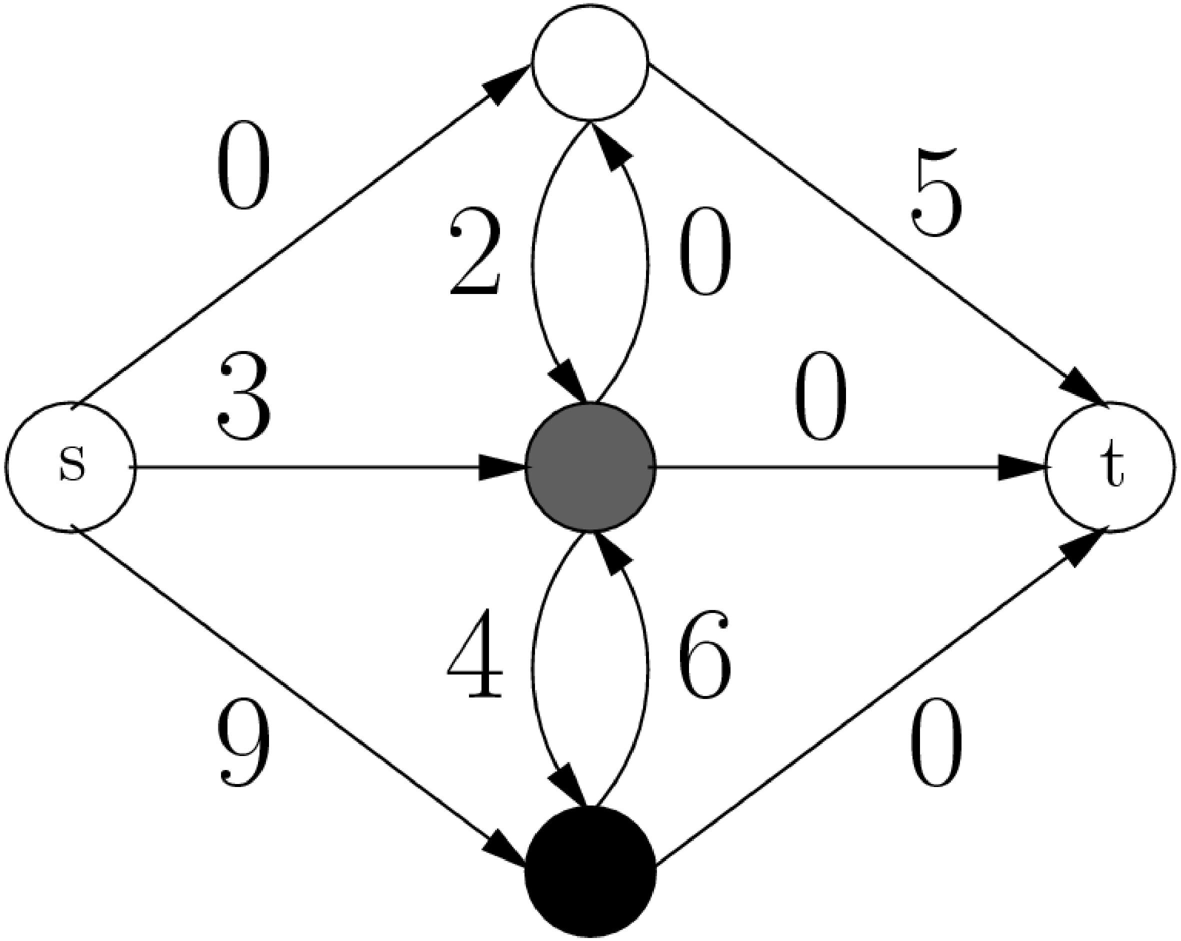
Example: Ford & Fulkerson



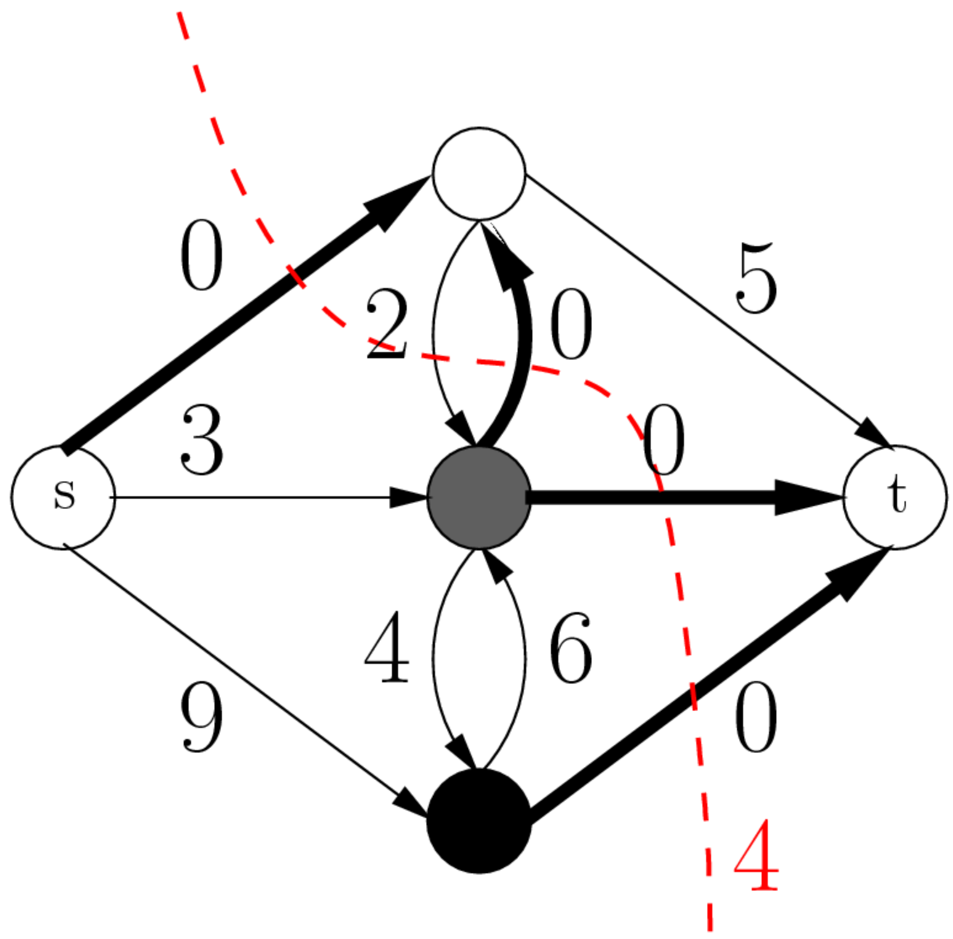
Example: Ford & Fulkerson



Example: Ford & Fulkerson



Example: Ford & Fulkerson



Markov Random Fields & Segmentation

Therefore, segmentation is equivalent with recovering a labeling $\mathcal{L}(\Omega)$ where the posterior density given the data is maximized while smoothness characterized the decision process

$$E(\mathcal{L}(\Omega) = E_{data}(p(\mathcal{I}|\mathcal{L}(\Omega)) + \alpha E_{smooth}(p(\mathcal{L}(\Omega))$$

that under conditional independence on the labeling process leads to

$$E(\mathcal{L}(\Omega) = \sum_{\omega \in \Omega} -\log\left(p(\mathcal{I}(\omega)|\mathcal{L}(\omega))\right) + \alpha \sum_{\omega \in \Omega} \sum_{\phi \in \mathcal{N}(\Omega)} \mathcal{V}(\omega, \phi)$$

Where the first term measures the fitness of the data associated at ω with the label and the second term penalizes discontinuities on the labeling process at the local neighborhood scale

- This is done through direct comparison between the label of each pixel and the labels of the neighborhood pixels

Markov Random Fields

In the simplistic case where each hypothesis is represented using a Gaussian distribution and the only information is the data itself one can simplify the model in the following way

$$E(\mathcal{L}(\Omega) = \sum_{\omega \in \Omega} \left(\log \sigma_{\mathcal{L}(\omega)} + (\mathcal{I}(\omega) - \mu_{\mathcal{L}(\omega)})^2 \right) + \alpha \sum_{\omega \in \Omega} \sum_{\phi \in \mathcal{N}(\Omega)} \mathcal{V}(\omega, \phi)$$

Where $\mathcal{V}(\omega, \phi)$ at the simplest case is a binary function equal to 1 when the labels are different (thus increasing the cost of such a labeling) and equal to 0 when they are the same

Minimizing the cost function with respect to the labeling vector $\mathcal{L}(\Omega)$ equivalent with solving the segmentation problem,

Such an approach does not take into account natural discontinuities of the image like transitions between two different classes and therefore will fail on the classes boundaries

Markov Random Fields

One can modify the cost function to account for the image discontinuities

$$E(\mathcal{L}(\Omega) = ... + \alpha \sum_{\omega \in \Omega} \sum_{\phi \in \mathcal{N}(\Omega)} g(|\nabla \mathcal{I}(\omega)|)\mathcal{V}(\omega, \phi)$$

Where $g(|\nabla \mathcal{I}(\omega)|)$ is a boundary indicator often defined in the following manner;

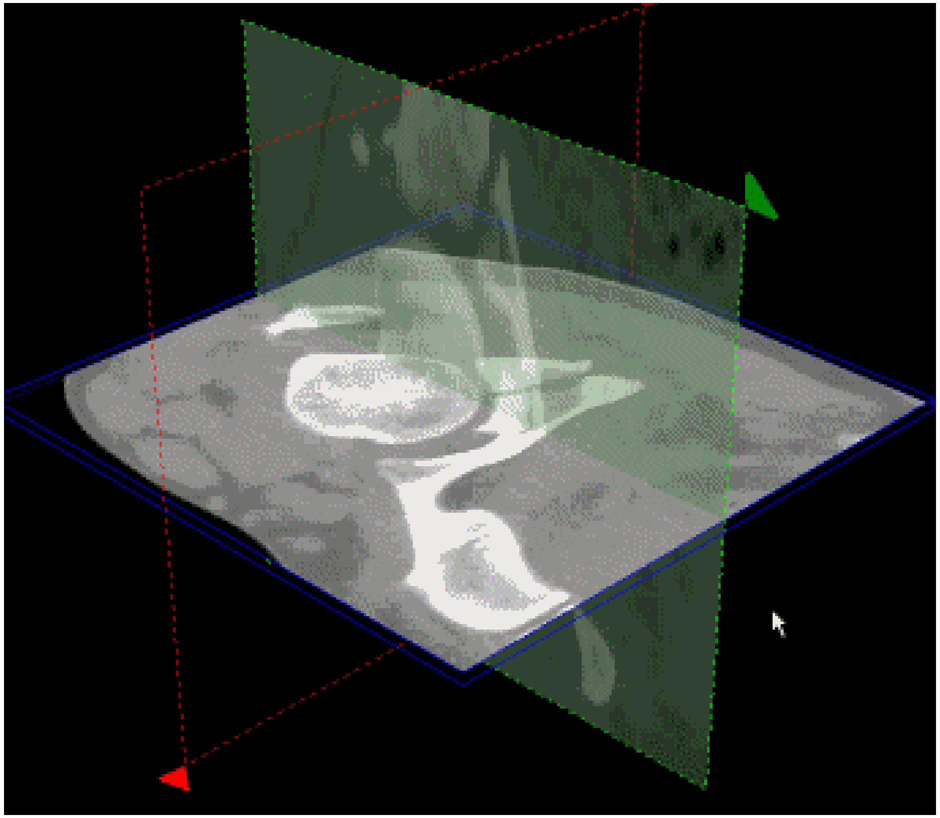
$$g(x) = \frac{1}{1 + |x|^\beta}$$

Such a modification will downsize the effect of introducing a discontinuity in areas where the gradient is strong that eventually means boundaries between two different classes

One can minimize this cost function towards global/local minimum

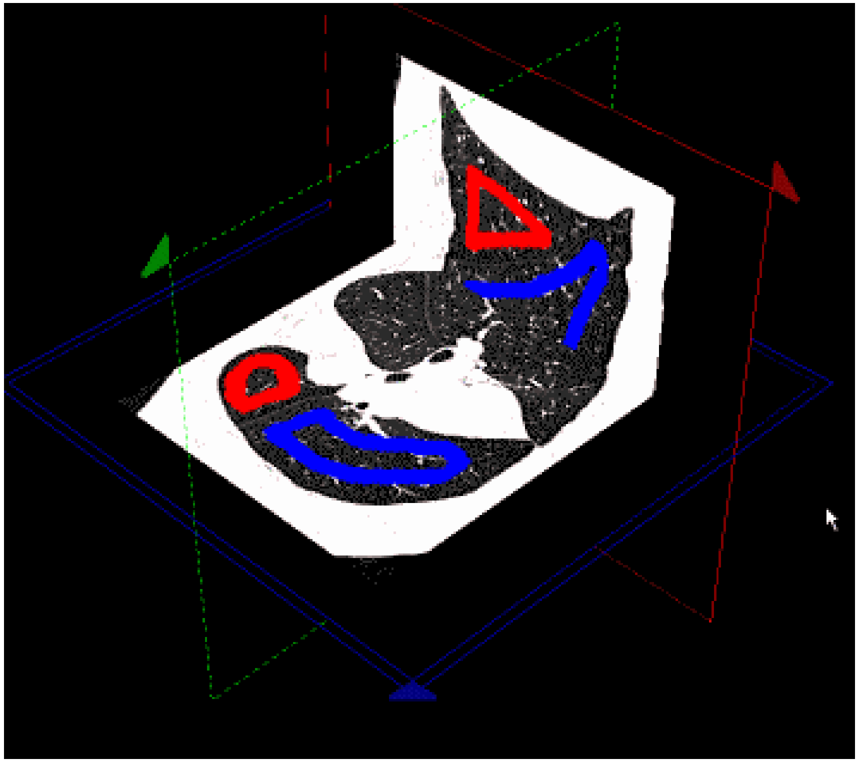
- Simulated annealing
- Combinatorial Optimization (Graph cuts)
- Mean-field annealing (Iterated conditional Modes, Highest Confidence First)

Bone Segmentation



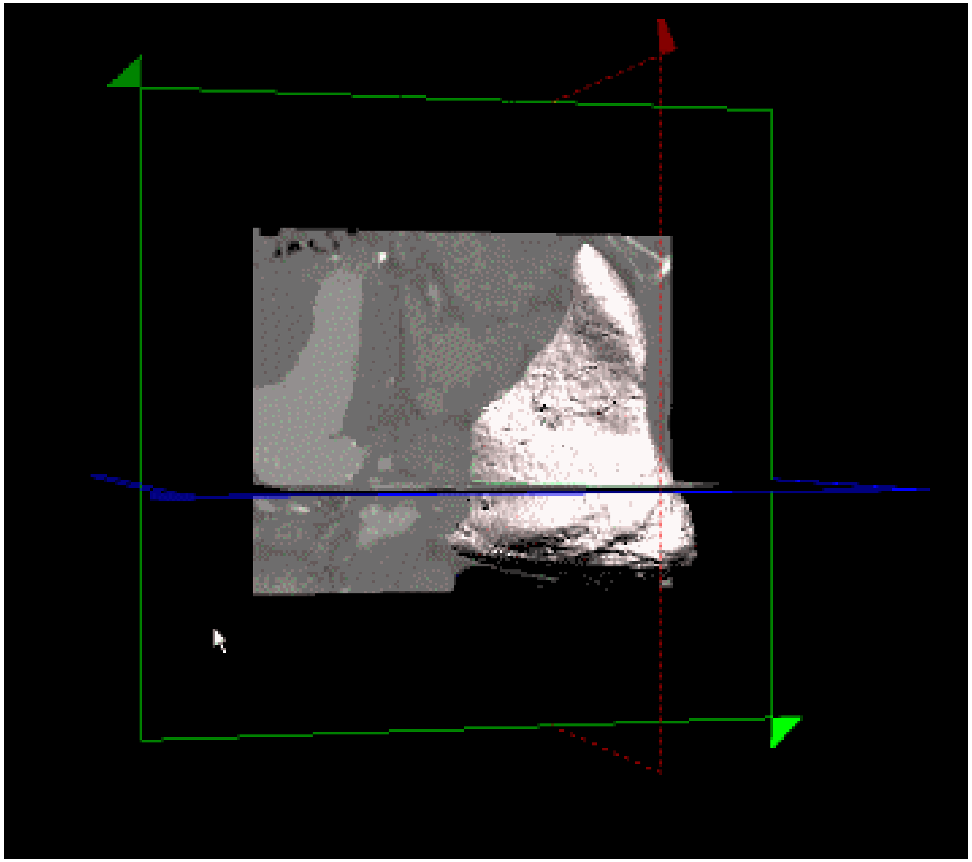
@ Y. Boykov, UWO

kidney Segmentation



@ Y. Boykov, UWO

Liver Segmentation



@ Y. Boykov, UWO

Knowledge-based Segmentation

Knowledge-based Object Extraction

Objective:

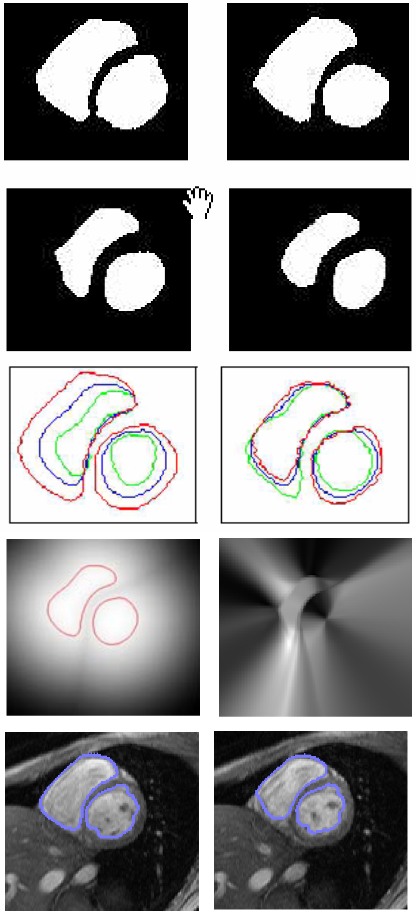
recover from the image a structure
of a particular – known to some extend
– geometric form

Methodology

- Consider a set of training examples
- Register these examples to a common pose
- Construct a compact model that expresses the variability of the training set
- Given a new image, recover the area where the underlying object looks like that one learnt

Advantages of doing that on the LS space:

- Preserve the implicit geometry
- Account with multi-component objects…
- … all wonderful staff you can do with the LS



Active Shape Models

Learning: recover a geometric form of an object from a training set

- Register a set of training examples to a common pose
- Recover a probabilistic notion of a model that involves a small number of parameters from the training set

Segmentation: find in an image a region that refers to the same object according to some intensity properties while respecting the prior knowledge

- Find a global transformation between an average model and the image that positions the model to the most prominent desired image
- Deform the model locally within the bounds of the probability density function – the one has learnt – to better account for local deformations of the object
- Do that in an incremental manner, iterate between extraction of interesting features and model fitting

Some Definitions (Modeling aspect)

Training Set: $\mathcal{S} = \{s_1, s_2, ..., s_n\}$ that refer to n shapes in a discrete form using a finite number of control points

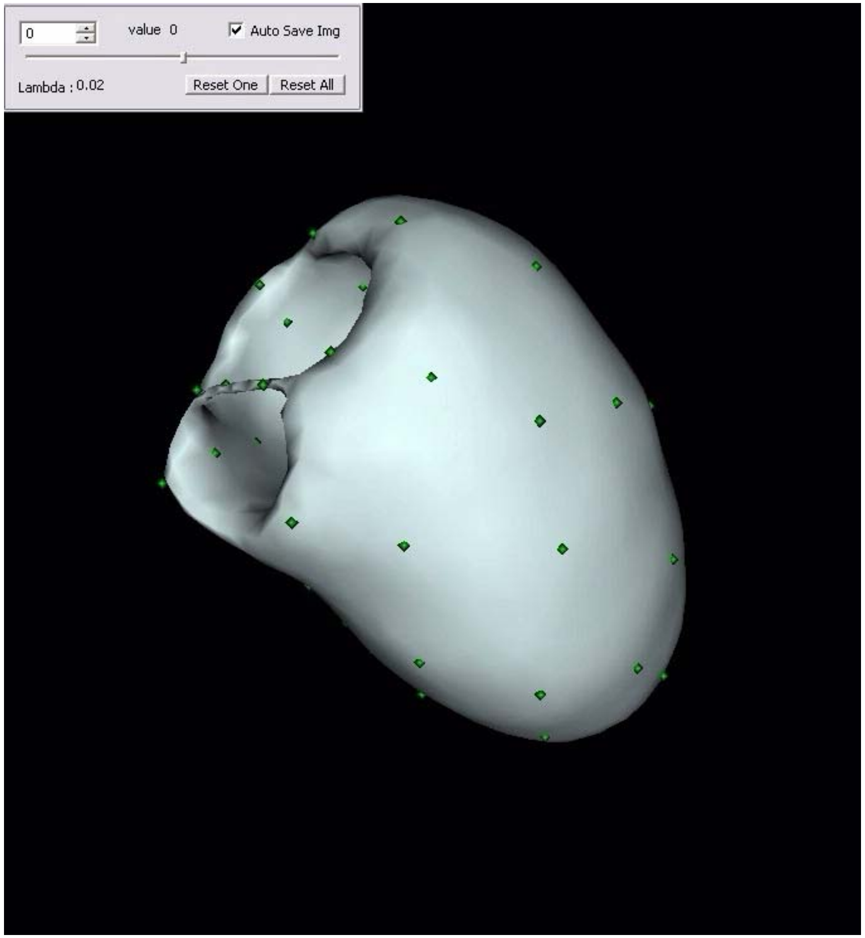
Registered Training Set $\overline{\mathcal{S}} = \{\overline{s}_1, \overline{s}_2, ..., \overline{s}_n\}$ where an one-to-one correspondence has been recovered between all shapes for each control point

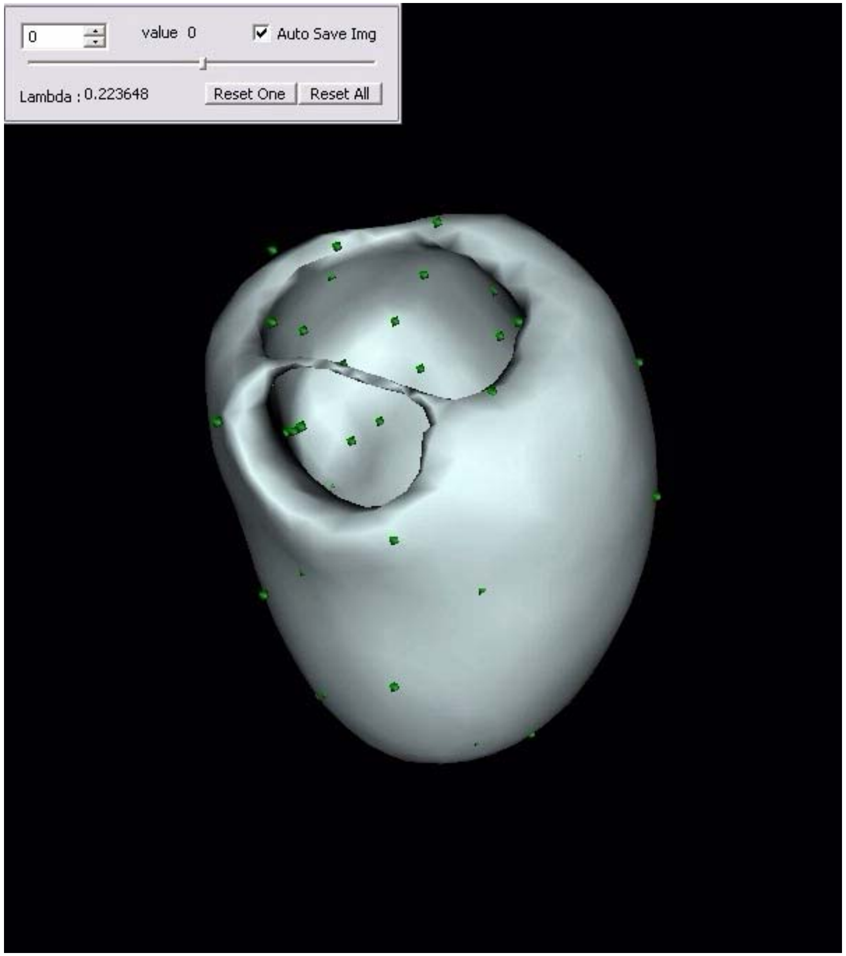
Average Model: $$\overline{s} = \frac{1}{n}\sum_{i=1}^{n} \overline{s}_i$$

An orthogonal basis of modes of variation $\mathcal{U}_k$ such that one can reproduce the training set through a global transformation $\mathcal{T}$ of the average model and a linear combination (a set of parameters λ_k) of the principal modes of variations?

$$s = \overline{s}(\mathcal{T}) + \sum_{1}^{k} \lambda_k \mathcal{U}_k$$

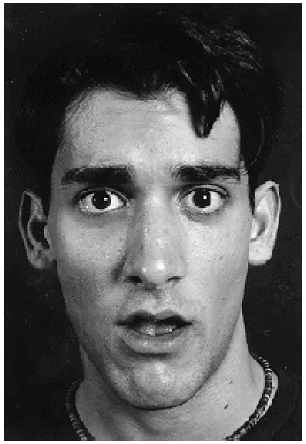
Some nice animations





Definition & Registration of the Training Set

Ground truth provided by the user,



Manual approach

- Parameterization of the contour: user-defined criteria, compromise low complexity and fair approximation of the shape of the target
- Registration: done in an implicit manner through the user by point clicking

Automatic approach

- Parameterization of the contour: according to the number of samples in the training set
- Registration: done separately using various point-matching registration techniques or distance transforms

Model Building: Singular Value Decomposition of the Registered training set

Building the Model

Once training samples have been registered one can estimate the average model:

$$\overline{s} = \frac{1}{n}\sum_{i=1}^{n} \overline{s}_i$$

that explicitly assumes one-to-one correspondences between the control points of each sample

And use it to recover the relative variation of the training set from the mean shape through a simple subtraction process:

$$\widehat{s}_i = \overline{s}_i - \overline{s}$$

If we assume that the training set follows a Gaussian distribution, then one can use the subtracted results to define the covariance matrix of this distribution

Then by concatenating $\widehat{s}_i$ and performing singular value decomposition on this matrix one can recover the most prominent directions of variation

Complete Model

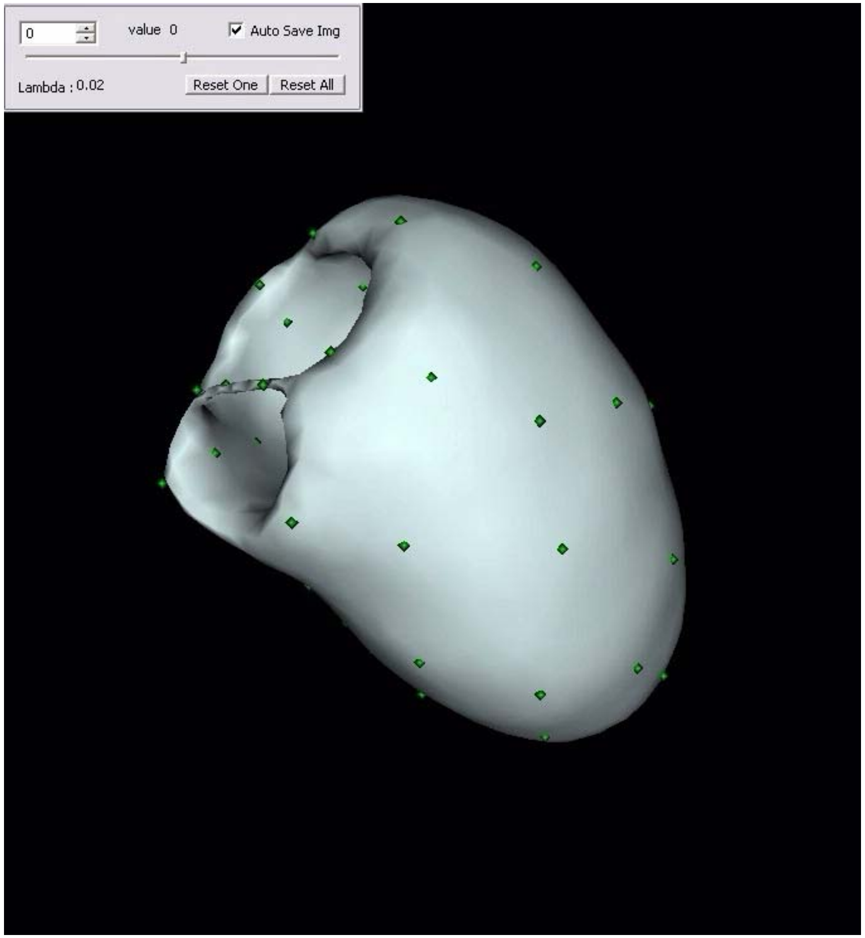
Consists of the average shape $\overline{s}$ and the principal modes of variations $\mathcal{U}_k$ that can be used to represent any element of the training set through a linear combination of the average model and the principal modes of variations:

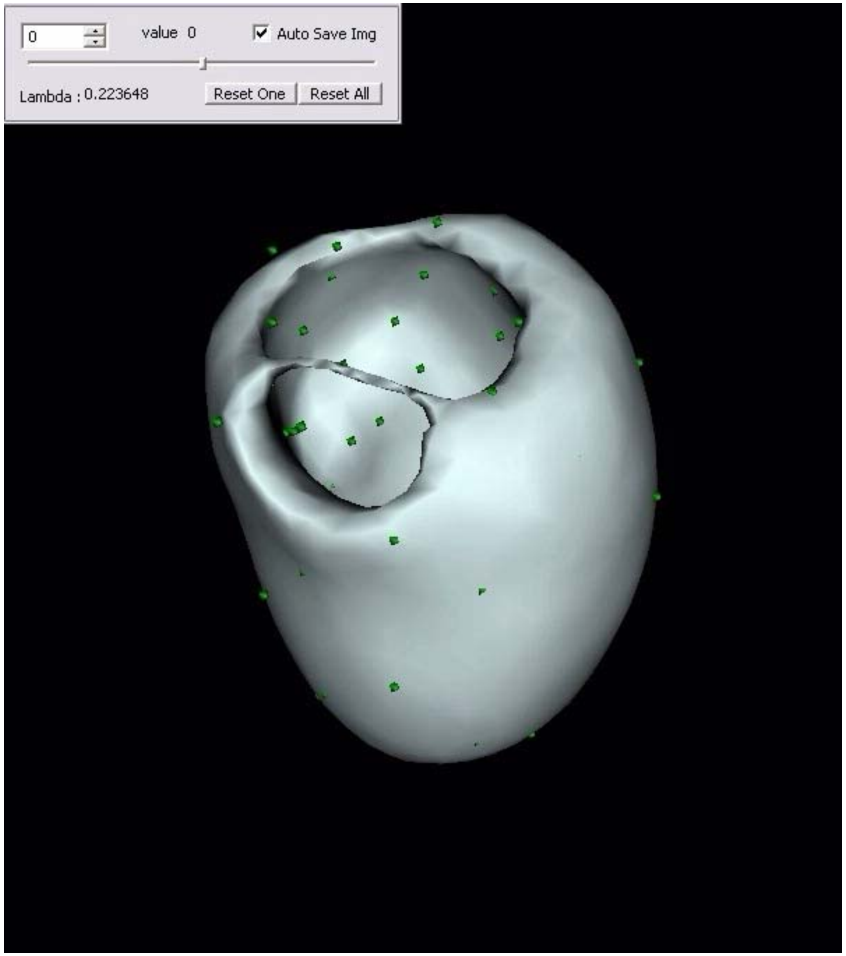
$$s = \overline{s} + \sum_{k=1}^{m} \lambda_k \mathcal{U}_k$$

Such a task is equivalent with finding the set of parameters that best approximate any given shape

The number of retained components is determined according to the magnitude of the eigen values

Some nice animations

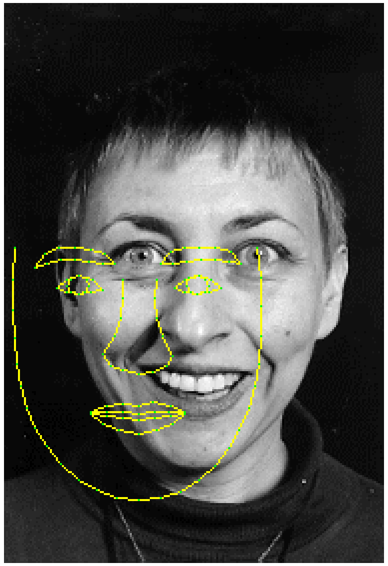




Segmentation Concept

In the case of segmentation, one can assume that the object observed in the image belongs to the family of shapes generated by the training set, where its exact position is unknown as well as its exact form

- Changes on the object pose require finding a global transformation between the mean model and the image
- exact localization of the object requires defining appropriate image features that can guide both the global transformation as well as the local deformations towards the object position…



$$s = \overline{s}(\mathcal{T}) + \sum_{k=1}^{m} \lambda_k \mathcal{U}_k$$

That consists of estimating

$$(\mathcal{T}, \lambda_k; k \in [1, m])$$

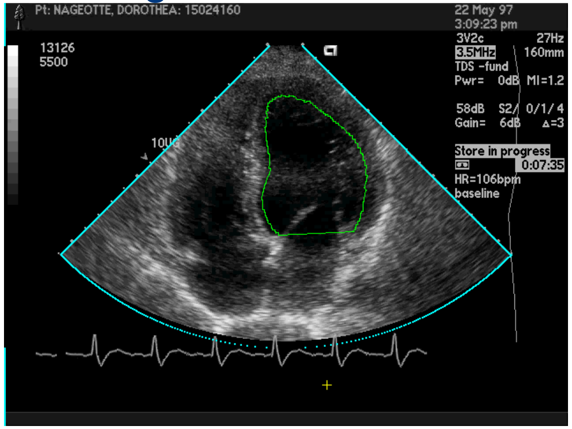
Segmentation

If we assume that the true object position is known s_{img}, then recovering these parameters is trivial since one has to minimize the distance between the model and the actual position

$$\min_{\lambda_k, \mathcal{T}} \quad \left| s_{img} - \overline{s}(\mathcal{T}) - \sum_{k=1}^{m} \lambda_k \mathcal{U}_k \right|^2$$

That is equivalent with solving a linear system with more constraints than the number of unknown variables, and can be done in a straightforward manner

Such a system is recovered through a least square optimization and therefore can be very sensitive to outliers…



$$\min_{\lambda_k, \mathcal{T}} \quad \rho\left(\left| s_{img} - \overline{s}(\mathcal{T}) - \sum_{k=1}^{m} \lambda_k \mathcal{U}_k \right| \right)$$

One can address such a limitation through a robust minimization

the Solution

Calculus of variations with respect to the unknown parameters

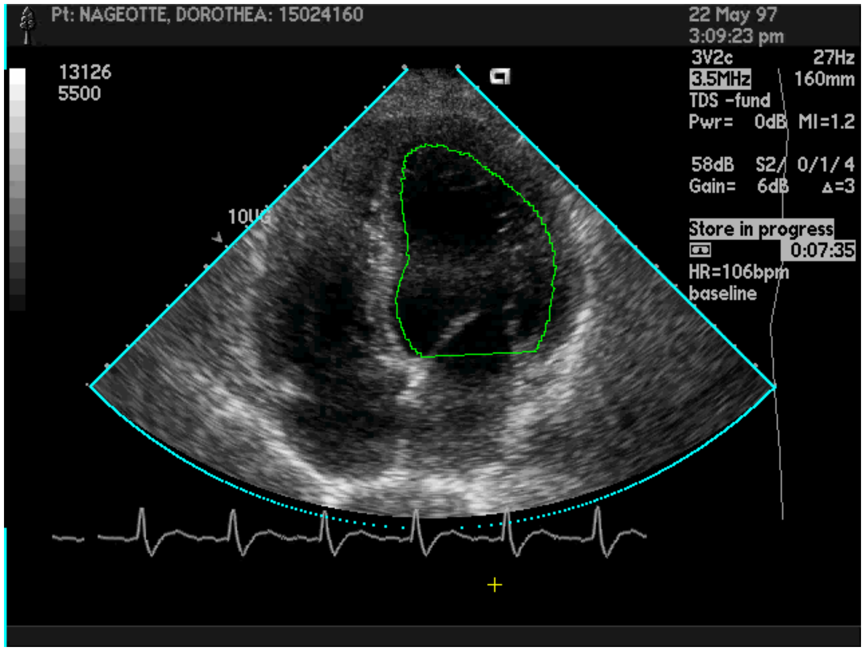
$$\frac{\partial}{\partial \lambda_k}\rho\left(\left|s_{img} - \overline{s}(\mathcal{T}) - \sum_{k=1}^{m}\lambda_k \mathcal{U}_k\right|\right) = 0$$

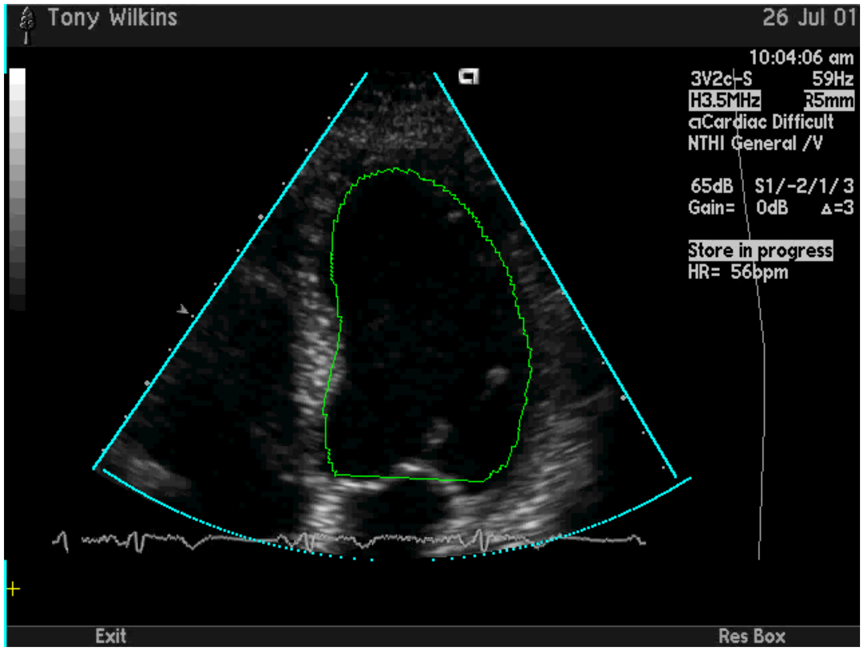
$$\frac{\partial}{\partial \tau}\rho\left(\left|s_{img} - \overline{s}(\mathcal{T}) - \sum_{k=1}^{m}\lambda_k \mathcal{U}_k\right|\right) = 0$$

Where $\mathcal{T} = (\tau_1, \tau_2, ...\tau_o)$ are the parameters of the transformation (3 for rigid, 4 for similarity, 6 for affine)

Such a system has low dimensionality and a unique, closed form solution where the number of constraints is greater to the number of unknowns to be recovered…(each control point creates refers to two equations, one at the horizontal and one at the vertical direction)

Some nice animations

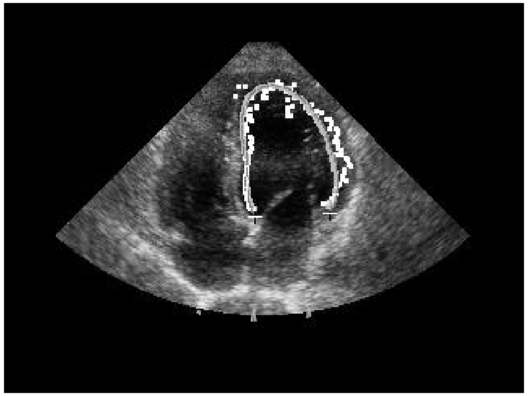
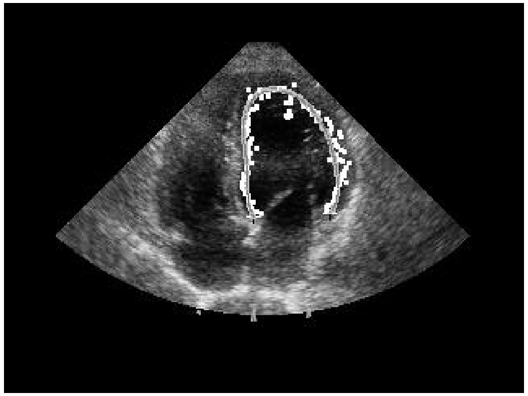




Finding Correspondences…(the simple case)

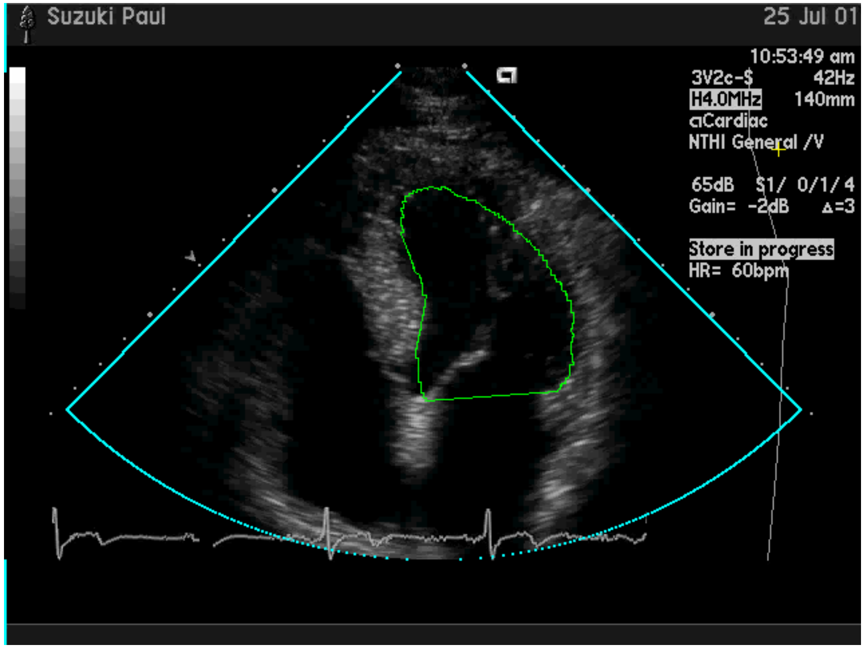
However the true position of the object is unknown and the one to be recovered and the notion of exact correspondences for the model control points is ambiguous

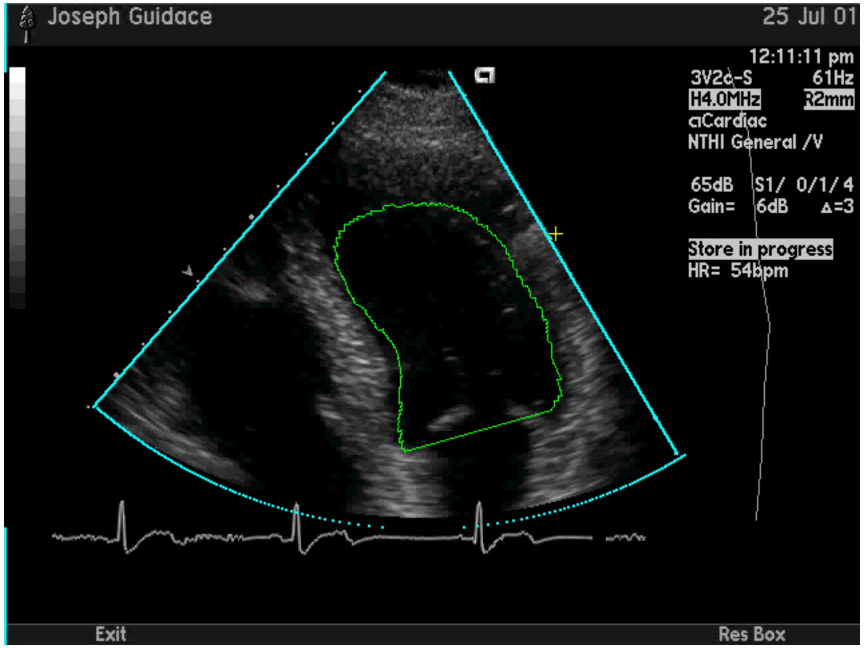
One can assume that object boundaries refer to discontinuities on the image plane, therefore given an initial position of the active shape, one can seek for the closest edge at every control point of the model:

Simple edge detectors can be used, the only issue to be addressed is the search area…

More nice animations



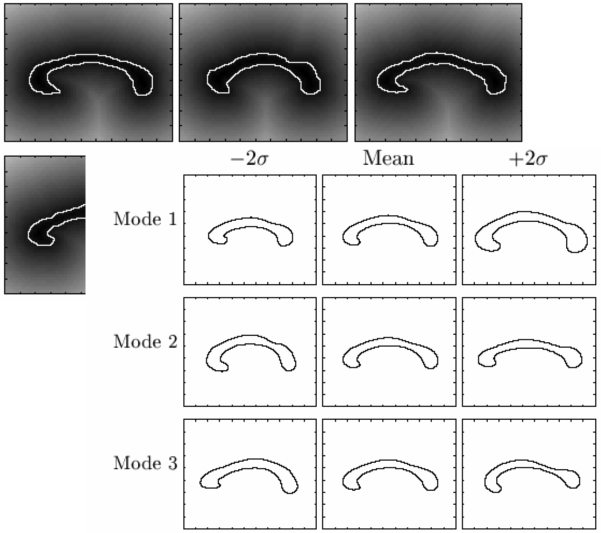


Knowledge-based Segmentation

[leventon-faugeras-grimson-etal:00]

Concept: Alternate between segmentation

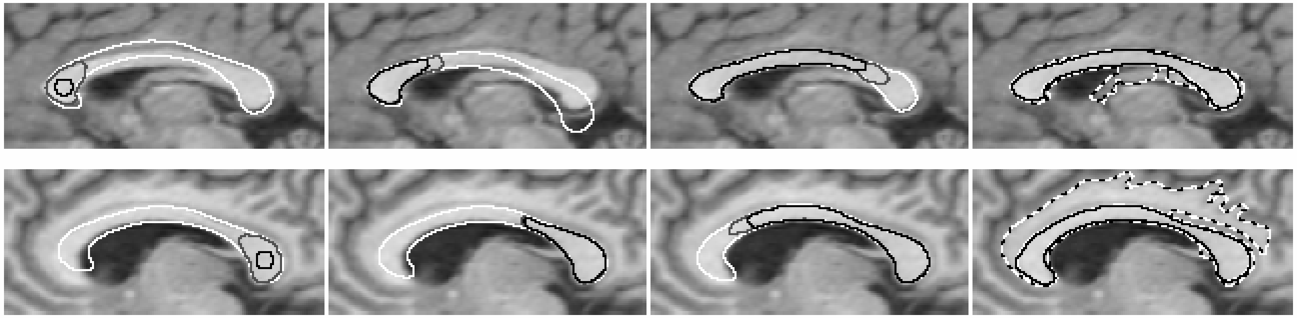
& imposing prior knowledge



- Learn a Gaussian distribution of the shape to be recovered from a training set directly at the space of implicit functions
  - The elements of the training set are registered
  - A principal component analysis is use to recover the covariance matrix of probability density function of this set
- ALTERNATE
  - Evolve a let set function according to the geodesic active contour
  - Given its current form, deform it locally using a MAP criterion so it fits better with the prior distribution
  - Until convergence…

Knowledge-based Segmentation

[leventon-faugeras-grimson-etal:00]



Limitations:

- Data driven & prior term are decoupled
- Building density functions on high dimensional spaces is an ill posed problem,
- Dealing with scale and pose variations (they are not explicitely addressed)

Knowledge-based Segmentation

[chen-etal:01]

Concept level:

- Use an average model as prior in its implicit function
- For a given curve find the transformation that projects it closer to the zero-level set of the implicit representation of the prior
- For a given transformation evolve the curve locally towards better fitting with the prior…
- Couple prior with the image driven term in a direct form…

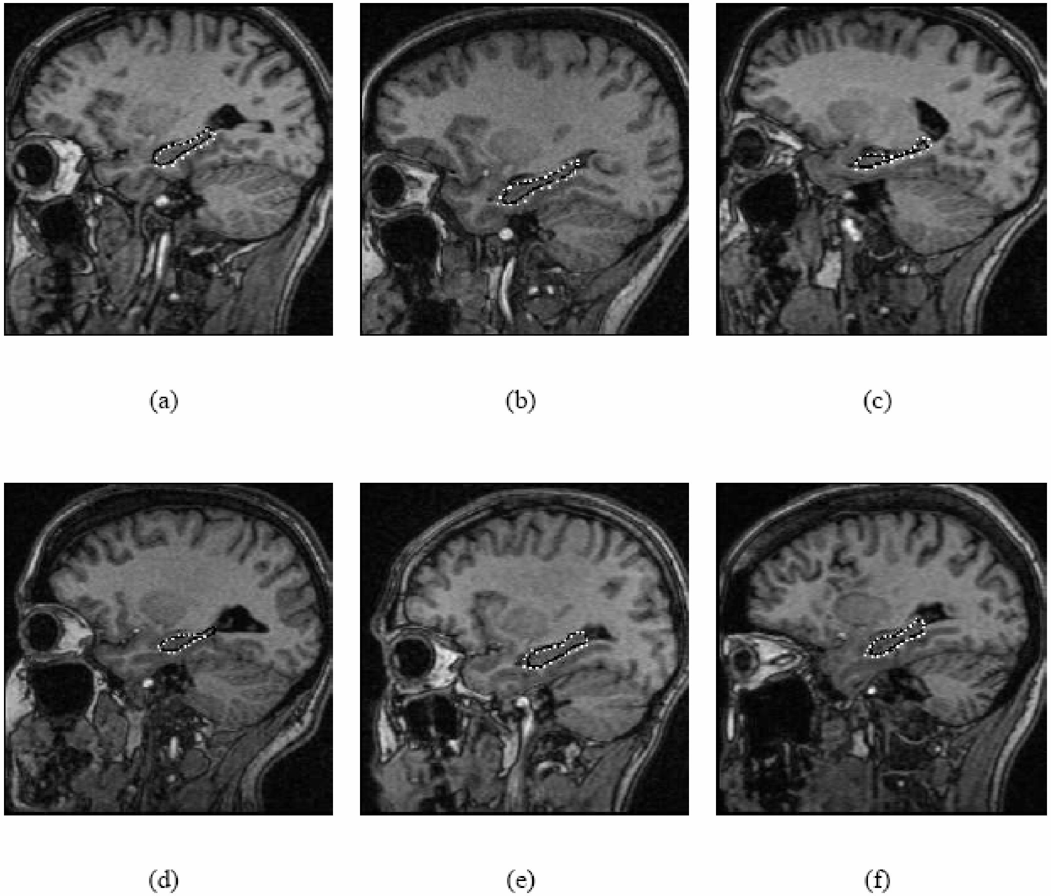
$$\min_{u,\mu,R,T} \int_{\Omega} \delta(u)\{g(|\nabla I|) + \frac{\lambda}{2} d^2(\mu R x + T)\}|\nabla u|$$

Issues to be addressed:

- Model is very simplistic (average shape) – opposite to the leventon's case where it was too much complicated…
- Estimation of the projection between the curve and the model space is tricky…not enough support…data term can be improved…

Knowledge-based Segmentation

[chen-etal:01]



Knowledge-based Segmentation

[tsai-yezzi-etal:01]

At a concept level, prior knowledge is modeled through a Gaussian distribution on the space of distance functions by performing a singular value decomposition on the set of registered training set,

The mumford-shah framework determined at space of the model is used to segment objects according to various data-driven terms

The parameters of the projection are recovered at the same time with the segentation result…

- A more convenient approach than the one of Leventon-etal
- Which suffers from not comparing directly the structure that is recovered with the model…

Knowledge-based Segmentation

[paragios-rousson:02]

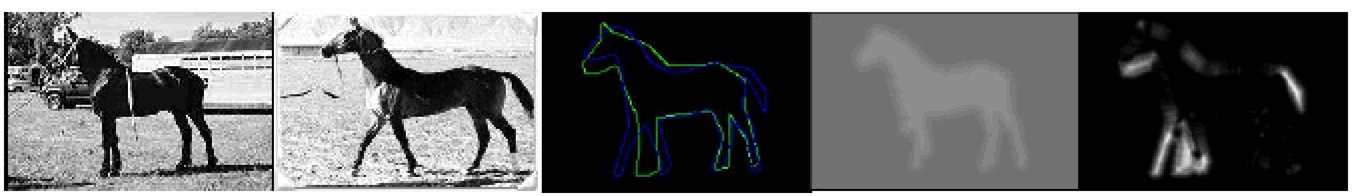
Prior is imposed by direct comparison between the model and evolving contour modulo a similarity transformation…

The model consists of a stochastic level set with two components,

- A distance map that refers to the average model
- And a confidence map that dictates the accuracy of the model

$$p_s(\phi) = \frac{1}{\sqrt{2\pi}\sigma_m(s)} e^{-\frac{(\phi - \phi_m(s))^2}{2\sigma_m(s)^2}}$$

Objective: Recover a level set that pixel-wise looks like the prior modulo some transformation



Model Construction

$$p_s(\phi) = \frac{1}{\sqrt{2\pi}\sigma_m(s)} e^{-\frac{(\phi - \phi_m(s))^2}{2\sigma_m(s)^2}}$$

From a training set recover the most representative model;

If we assume N samples on the training set, then the distribution that expresses at a given point most of these samples is the one recovered through MAP

$$E(\phi_m(s), \sigma_m(s)) = -\log \sum_{i=1}^{N} p_s(\phi_i(s)) = \sum_{i=1}^{N} \left[\log(\sigma_m(s)) + \frac{(\phi_i - \phi_m(s))^2}{2\sigma_m(s)^2} \right]$$

Where at a given pixel, we recover the mean and the variance that best describes the training set composed of implicit functions at this point, where the mean corresponds to the average value

Constraints on the variance to be locally smooth is a natural assumption

$$E(\phi_m, \sigma_m) = \alpha \sum_{i=1}^{N} \int\int_{\Omega} \left[\log(\sigma_m) + \frac{(\phi_i - \phi_m)^2}{2\sigma_m^2} \right] d\Omega + \int\int_{\Omega} \psi\left(\nabla \sigma_m\right) d\Omega$$

Imposing the (Static) Prior

Define/recover a morphing function "A" that creates correspondence between the model and the prior

$$\phi(;\tau) = \phi_m(\mathcal{A}(;\tau))$$

In the absence of scale variations, and in the case of global morphing functions one can compare the evolving contour with the model according to

$$E(\phi, \mathcal{A}) = \int\int_\Omega \delta_\epsilon(\phi)(\phi - \phi_m(\mathcal{A}))^2 d\Omega$$

Static Prior (continued)

$$E(\phi, \mathcal{A}) = \int\int_{\Omega} \delta_{\epsilon}(\phi)(\mathcal{S}\phi - \phi_m(\mathcal{A}))^2 d\Omega$$

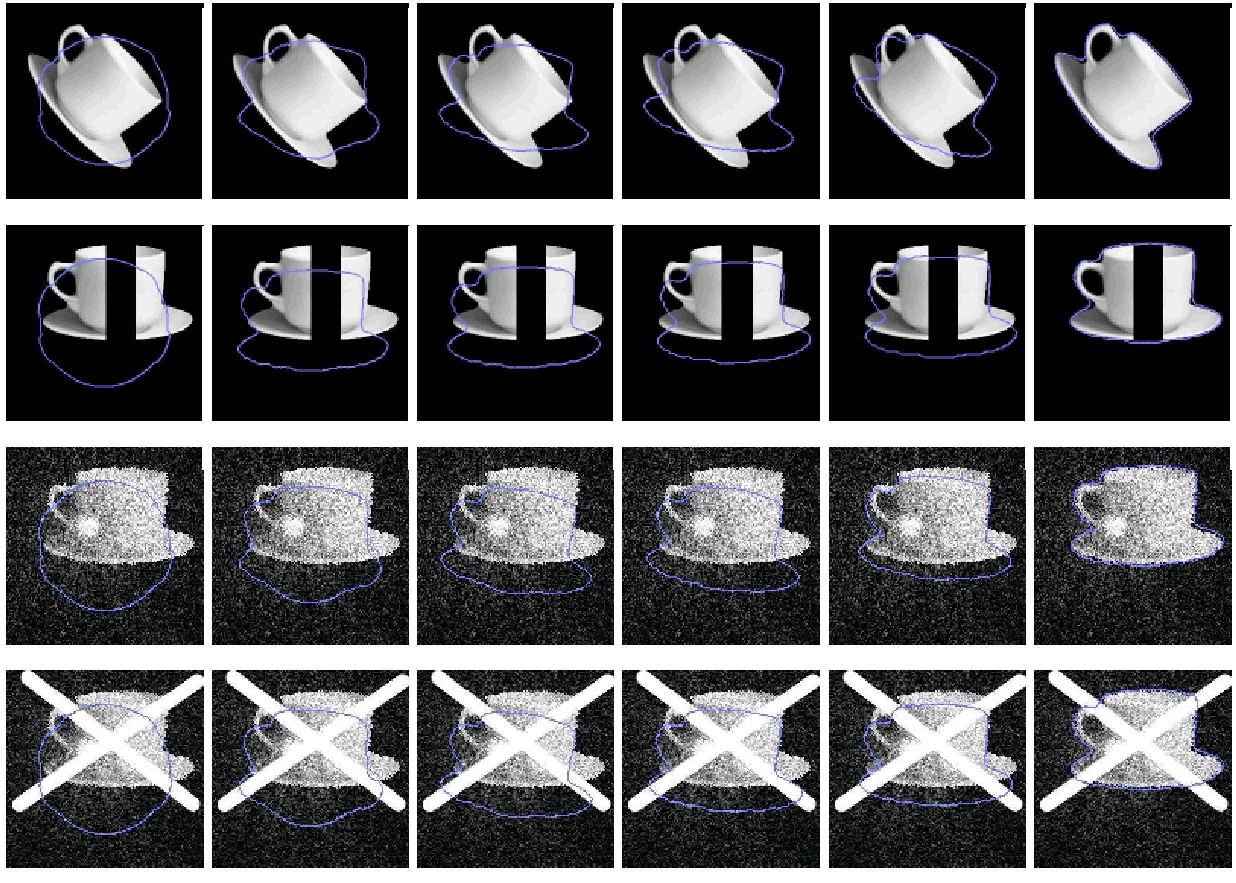
Where the unknowns are the morphing function and the position of the level set

Calculus of variations with respect to the position of the interface are straightforward:

$$\frac{d}{d\tau}\phi = -2\underbrace{\delta_{\epsilon}(\phi)\mathcal{S}(\mathcal{S}\phi - \phi_m(\mathcal{A}))}_{shape\ consistency\ force} - \underbrace{\left[\frac{\partial}{\partial \phi}\delta_{\epsilon}(\phi)\right](\mathcal{S}\phi - \phi_m(\mathcal{A}))^2}_{area\ force}$$

The second term is a constant inflation term aims at minimizing the area of the contour and eventually the cost function and can be ignored…since it has no physical meaning.

Static Prior, Concept Demonstration



Static Prior (continued)

One can also optimize the cost function with respect to the unknown parameters of the morphing function

$$\frac{d}{d\tau}\Theta = -2\int\int_{\Omega}\delta_{\epsilon}(\phi)(\mathcal{S}\phi - \phi_m(\mathcal{A}))(-\nabla\phi_m(\mathcal{A})\cdot\frac{\partial}{\partial\Theta}\mathcal{A})\; d\Omega$$

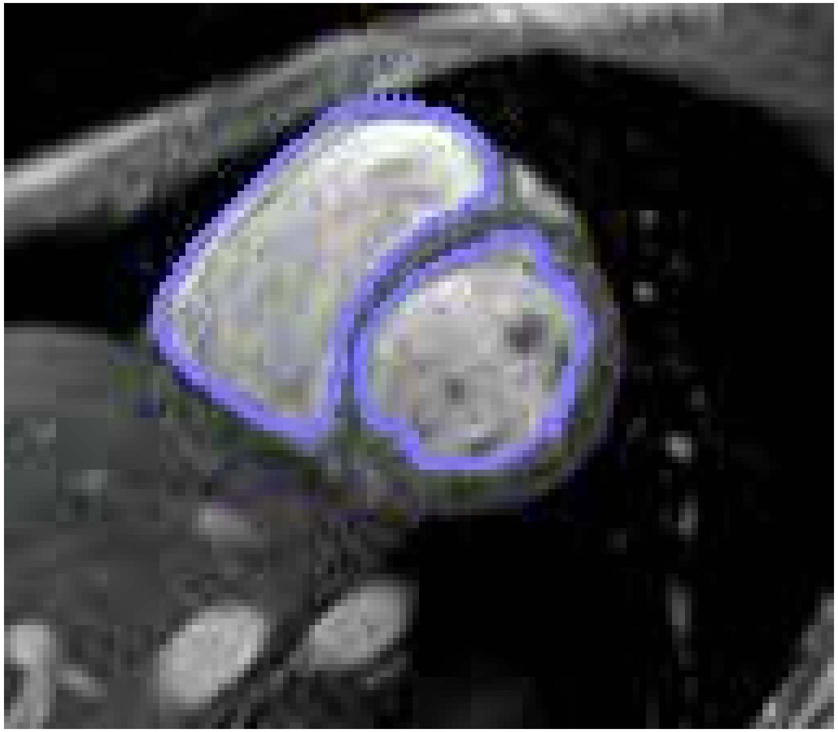
$$\frac{d}{d\tau}\mathcal{S} = -2\int\int_{\Omega}\delta_{\epsilon}(\phi)(\mathcal{S}\phi - \phi_m(\mathcal{A}))(-\phi - \nabla\phi_m(\mathcal{A})\cdot\frac{\partial}{\partial\mathcal{S}}\mathcal{A})\; d\Omega$$

$$\frac{d}{d\tau}\begin{bmatrix}\mathcal{T}x\\ \mathcal{T}y\end{bmatrix} = -2\int\int_{\Omega}\delta_{\epsilon}(\phi)(\mathcal{S}\phi - \phi_m(\mathcal{A}))(-\nabla\phi_m(\mathcal{A})\cdot\frac{\partial}{\partial\begin{bmatrix}\mathcal{T}x\\ \mathcal{T}y\end{bmatrix}}\mathcal{A})\; d\Omega$$

Leading to a nice “self-sufficient” system of motion equations that update the global registration parameters between the evolving curve and the model

However, the variability of the model was not considered up to this point and areas with high uncertainties will have the same impact on the process

Some Results



Taking Into Account the Model Uncertainties

Maximizing the joint posterior (segmentation/morphing) is a quite attractive criterion in "inferencing"

$$p(\mathcal{A}, \phi | \phi_m) = \frac{p(\phi_m | \mathcal{A}, \phi)}{p(\phi_m)} p(\phi, \mathcal{A}) = \frac{p(\phi_m(\mathcal{A}) | \phi)}{p(\phi_m)} p(\phi, \mathcal{A})$$

Where the Bayes rule was considered and given that the probability for a given prior model is fixed and we can assume that all (segmentation/morphing) solutions are equally probable, we get

$$p(\phi_m(\mathcal{A}) | \phi) = \prod_{\omega \in \Omega} p(\phi_m(\mathcal{A}(\omega)) | \mathcal{S}\phi(\omega))$$

Under the assumption of independence...within pixels…and then finding the optimal implicit function and its morphing transformations is equivalent with

$$E(\phi, \mathcal{A}) = -\log \left[\prod_{\omega \in \Omega} p(\phi_m(\mathcal{A}(\omega)) | \mathcal{S}\phi(\omega)) \right] = -\int \int_{\Omega} \log(p_{\omega}(\mathcal{S}\phi(\omega))) d\Omega$$

Taking Into Account the Model Uncertainties

That can be further developed using the Gaussian nature of the model distribution at each image pixel

$$E(\phi, \mathcal{A}) = \int\int_{\Omega} \left(\log(\sigma_m(\mathcal{A})) + \frac{(\mathcal{S}\phi - \phi_m(\mathcal{A}))^2}{2\sigma_m(\mathcal{A})^2} \right) d\Omega$$

A term that aims at recovering a transformation and a level set that when projected to the model, it is projected to areas with low variance (high confidence)

A term that aims at minimizing the actual distance between the level set function and the model and is scaled according to the model confidence…

- would prefer have a better match between the model and level set in areas where the variability is low,
- while in areas with important deviation of the training set, this term will be less important

Taking the derivatives…

Calculus of variations regarding the level set and the morphing function:

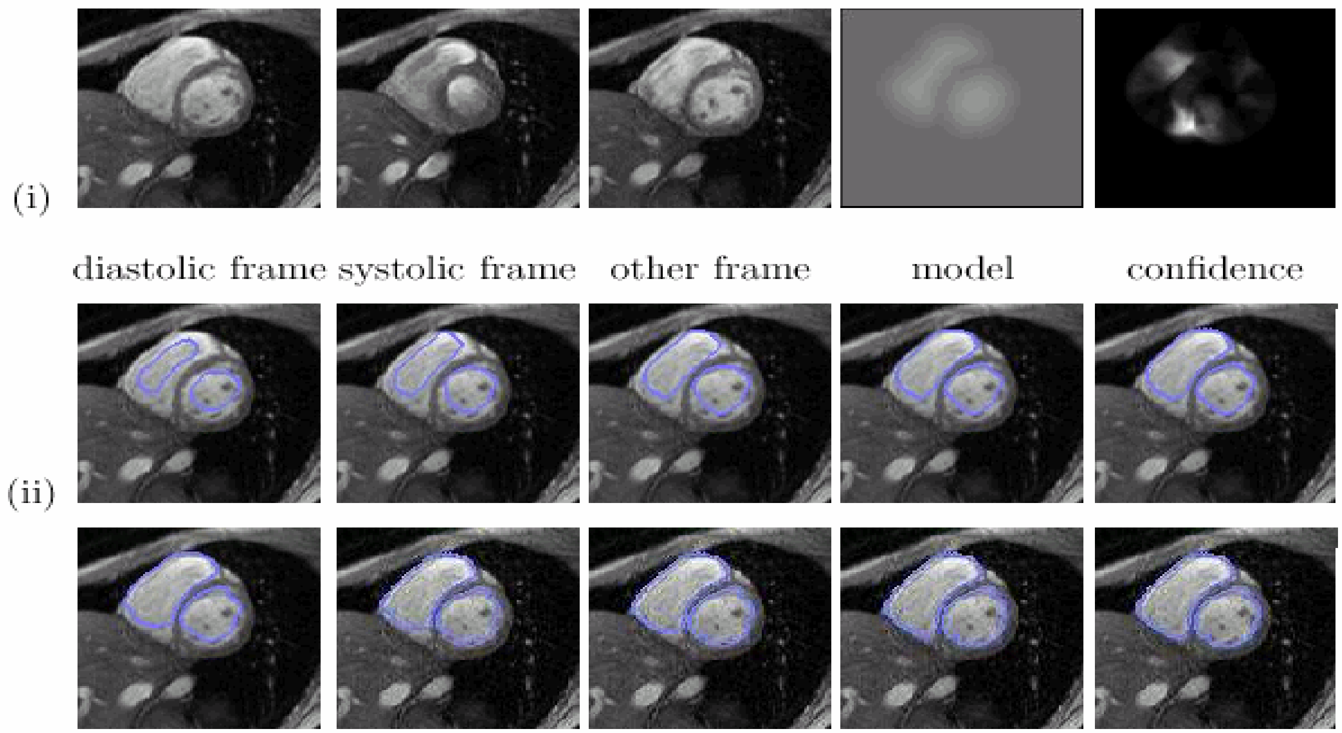
$$\frac{d}{d\tau}\phi = -\underbrace{\left[\frac{\partial}{\partial\phi}\delta_\epsilon(\phi)\right]\left(\log(\sigma_m(\mathcal{A})) + \frac{(\mathcal{S}\phi - \phi_m(\mathcal{A}))^2}{\sigma_m(\mathcal{A})^2}\right)}_{area\ \ force} - 2\underbrace{\delta_\epsilon(\phi)\mathcal{S}\frac{(\mathcal{S}\phi - \phi_m(\mathcal{A}))}{\sigma_m(\mathcal{A})^2}}_{shape\ \ consistency\ \ force}$$

The level set deformation flow consists of two terms:

- that is a constant deflation force (when the level set function collapses, eventually the cost function reaches the lowest potential)

- An adaptive balloon (directional/magnitude-wise) force that inflates/deflates the level set so it fits better with the prior after its projection to the model space…In areas with high variance this term become less significant and data-terms guide the level set to the real object boundaries...

Some medical results



Implicit Active Shapes

[rousson-paragios:03]

The Active Shape Model of Cootes et al. is quite popular to object extraction. Such modeling consists of the following steps:

Let us consider a training set $\mathcal{S}_i$ of N **registered surfaces** (implicit representations can also be used for registration [4]). Distance maps are computed for each surface:

$$\mathcal{S}_i \rightarrow \phi_i, \quad i = 1..N$$

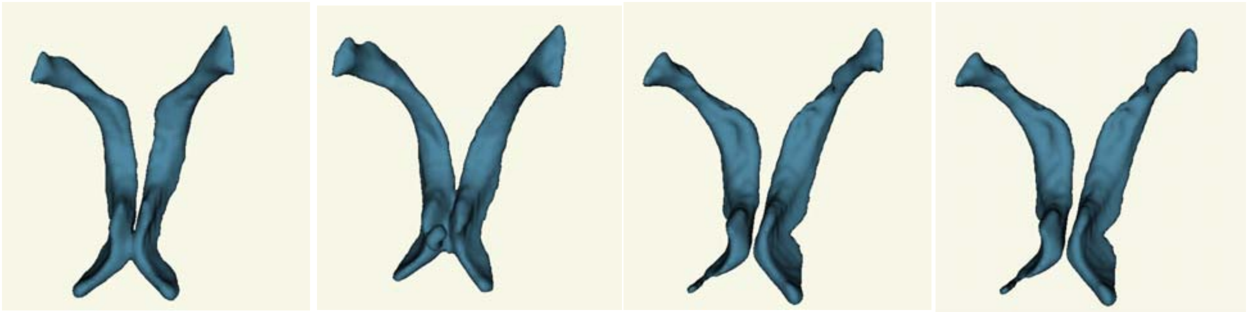
The samples N are centered with respect to the average representation :

$$\psi_i = \phi_i - \phi_{\mathcal{M}}$$

Implicit Active Shapes

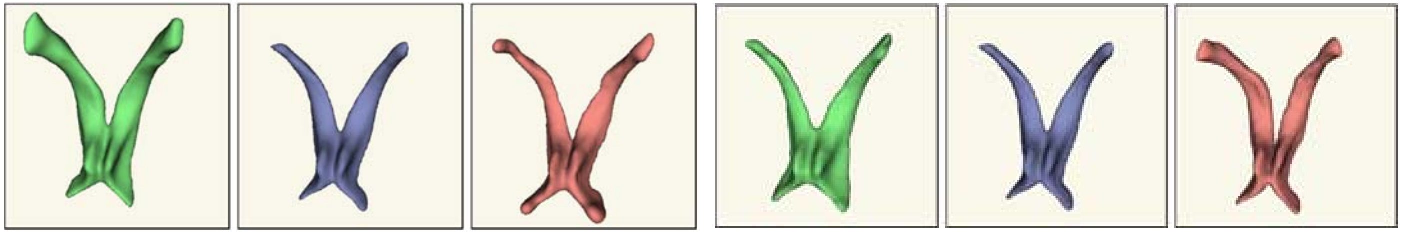
[rousson-paragios:03]

Training set:

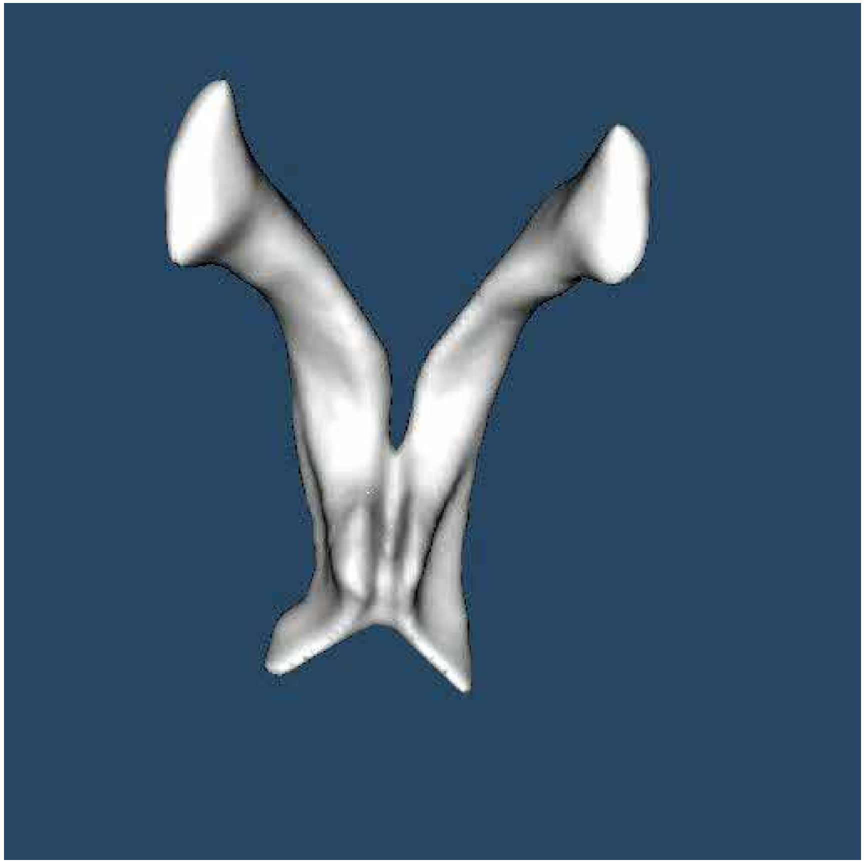


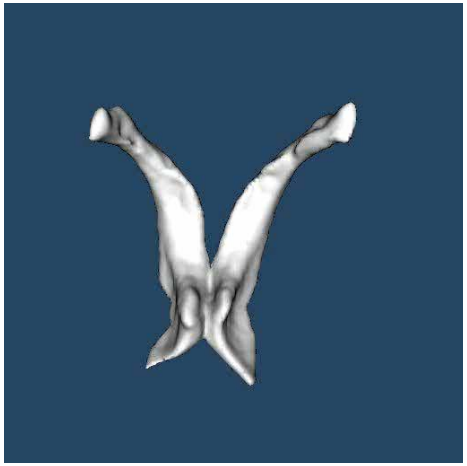
The **principal modes of variation** U_j are recovered through Principal Component Analysis (PCA). A new shape ϕ can be generated from the (m) retained modes:

$$\phi = \phi_{\mathcal{M}} + \sum_{j=1}^{m} \lambda_j U_j$$



The model…





The prior

A level set function that has minimal distance from a linear from the model space…

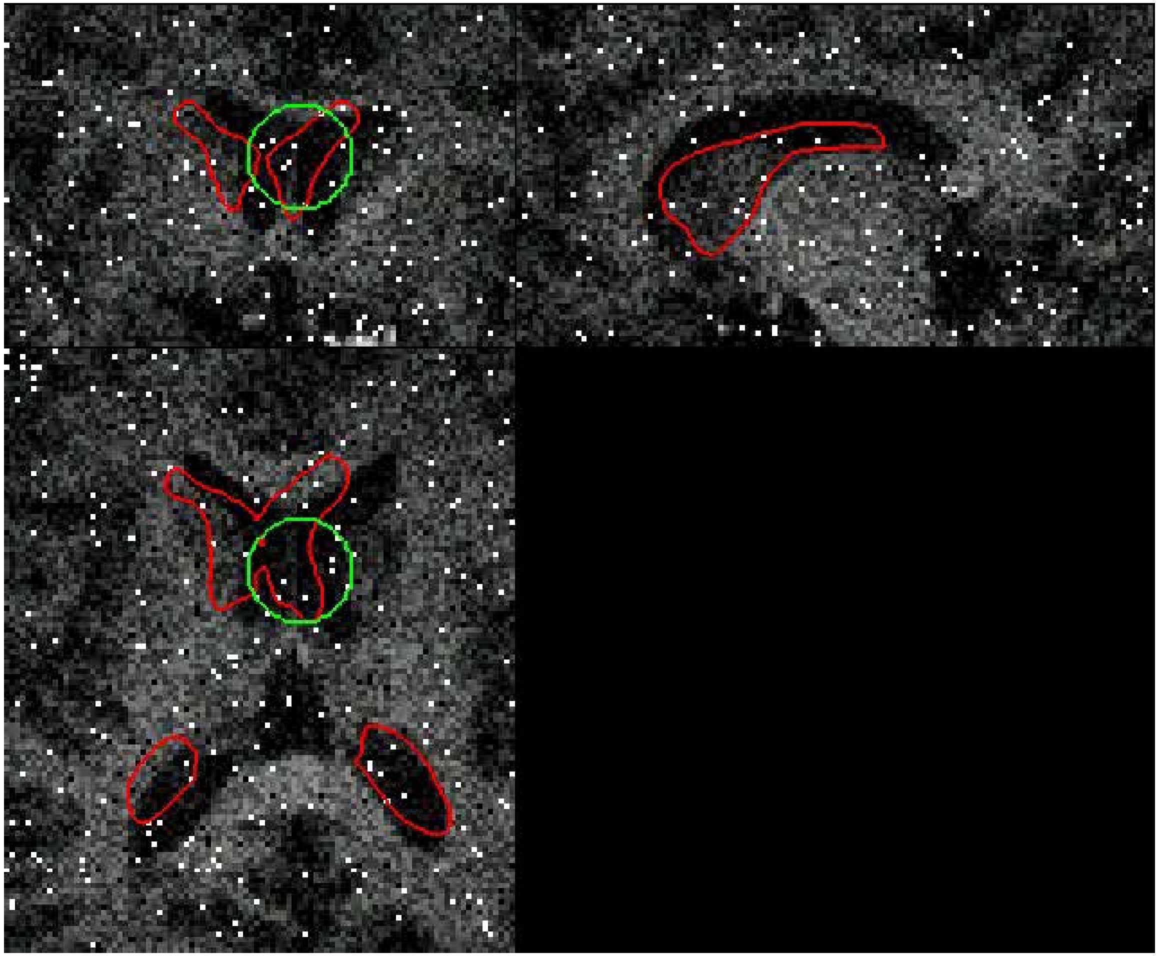
$$E(\phi, \mathcal{A}, \alpha) = \iint_{\Omega} \delta_{\epsilon}(\phi) \left(\mathcal{S}\phi - \left(\phi_{\mathcal{M}}(\mathcal{A}) + \sum_{j=1}^{m} \lambda_j \, U_j(\mathcal{A}) \right) \right)^2 d\Omega$$

The unknown consist of:

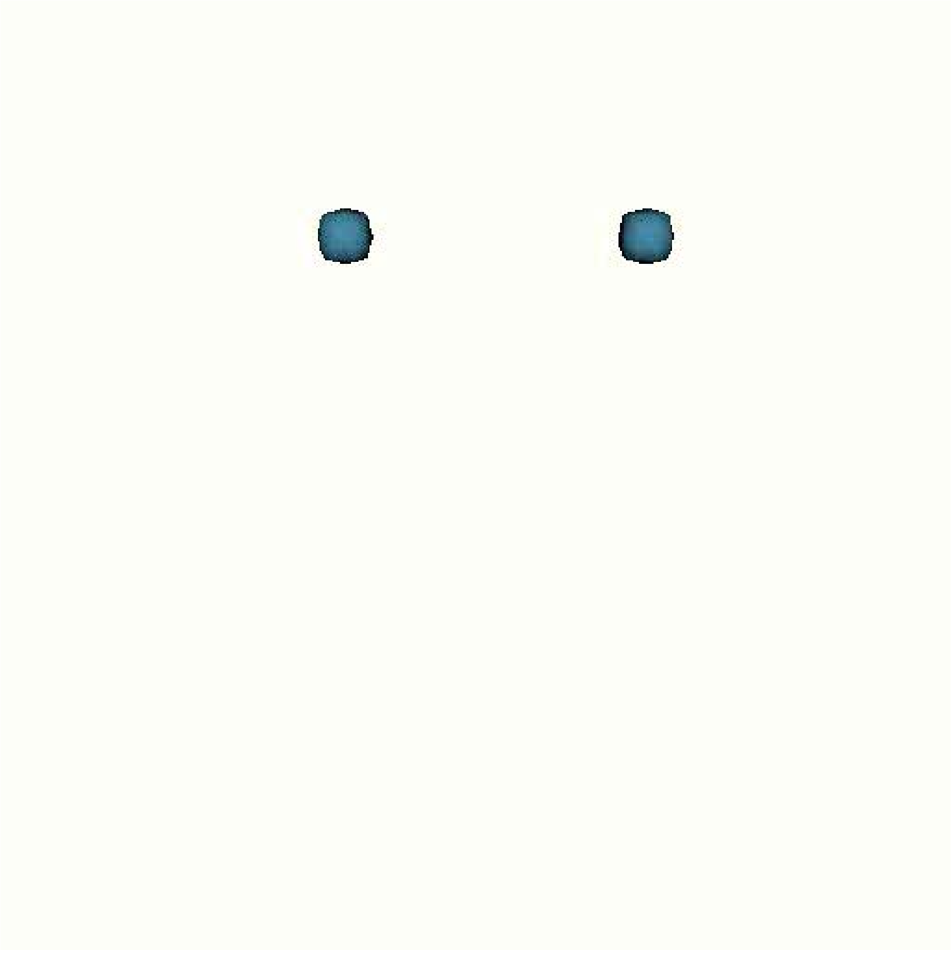
- The form of the implicit function
- The global transformation between the average mode and the image,
- The set of linear coefficients that when applied to the set of basis functions provides the optimal match of the current contour with the model space

And are recovered in a straightforward manner using a gradient descent method…

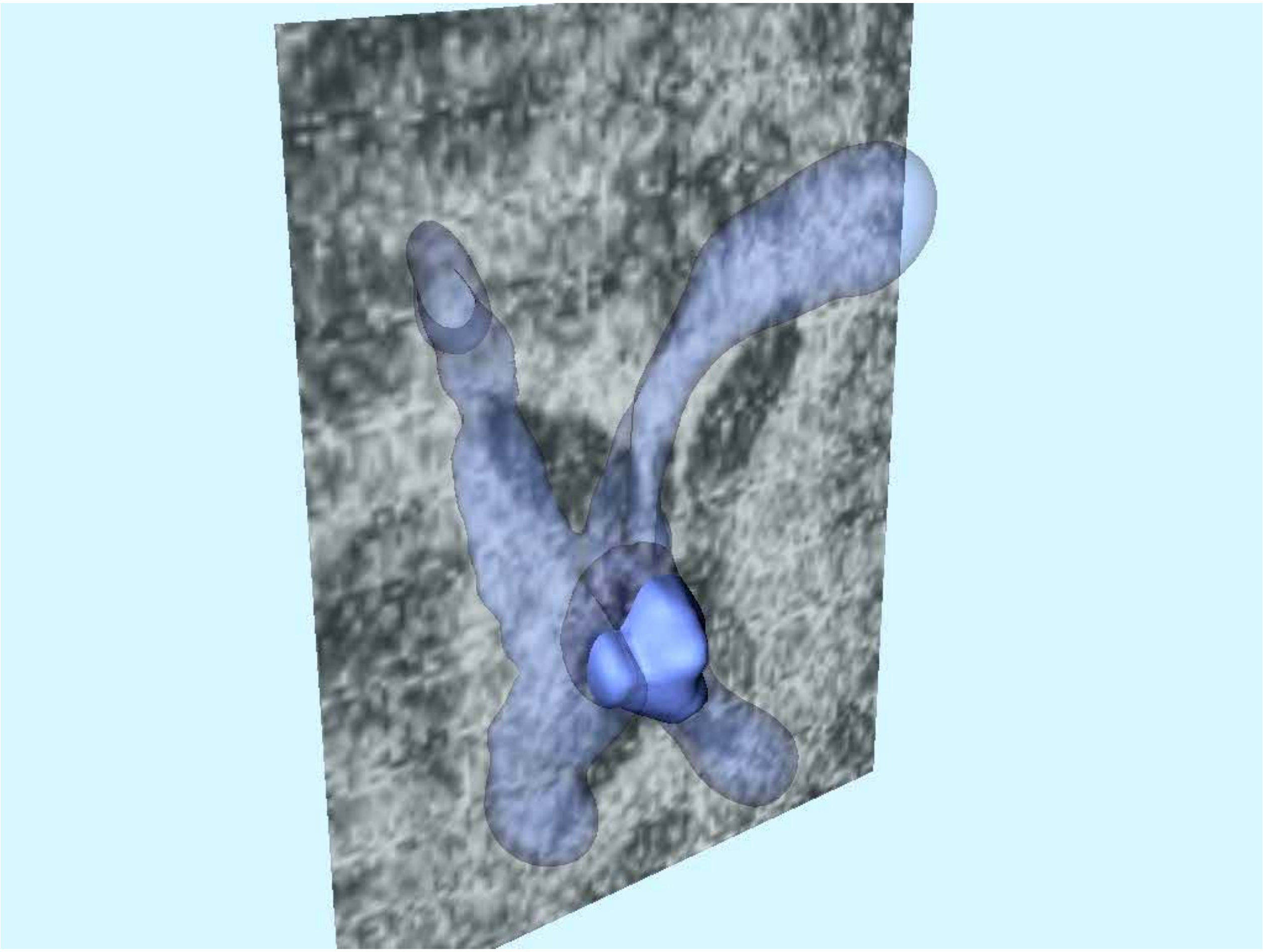
Some nice results…



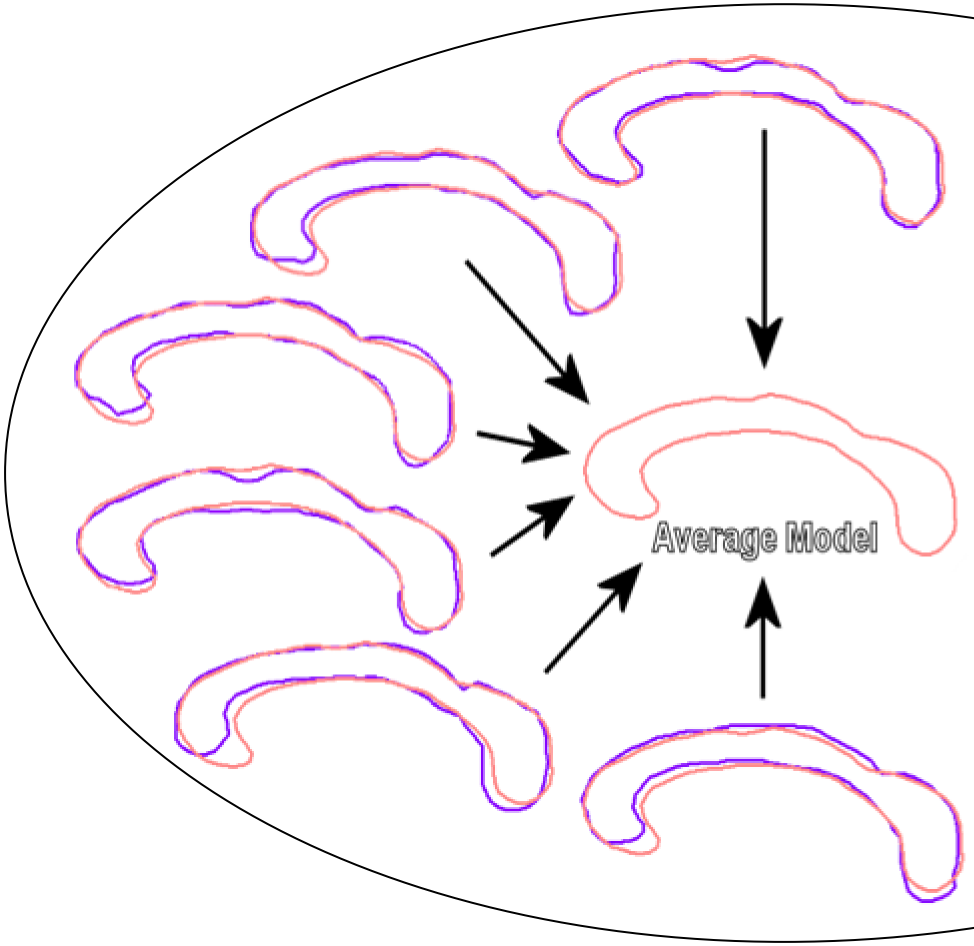
Some nice results…



Some nice results…



Going Beyond PCA



In the registration process each element was associated with uncertanties measures,

Such measures indicate the quality of the samples at a local scale and should be used when building models

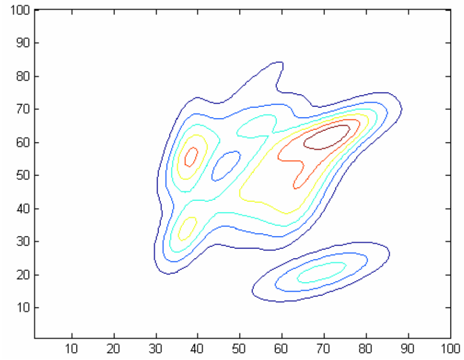
Non-parametric density approximations

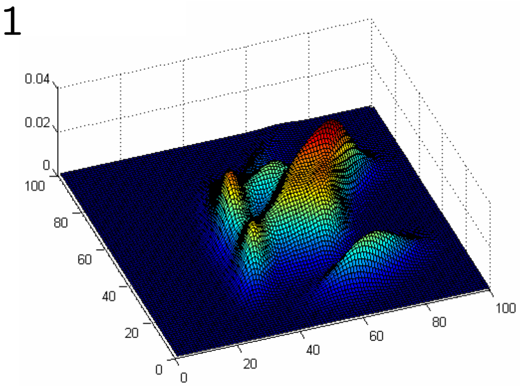
Opposite to Gaussian assumptions, kernels can do a better job in the approximation of non-linear functions.

$$N_i(\Theta) = \frac{1}{\|V_i(\hat{\Theta}_i)\|^{1/2}} \, \mathsf{K}\left(V_i(\hat{\Theta}_i)^{-1/2}(\Theta - \hat{\Theta}_i)\right)$$

The idea is to express the density as a combination of these kernels where the importance of the kernel is dictated from the uncertainties measures of the registration process.

$$\hat{f}_S(\boldsymbol{\Theta}) = \frac{1}{M}\sum_{i=1}^{M} N_i(\Theta)$$





Non-parametric Prior models & segmentation with Uncertainties

Integration between image and prior shape terms :

$$E(\Theta) = E_{im}(\Theta) + E_{shape}(\Theta) = E_{im}(\Theta) - log(\hat{f}_S(\Theta)))$$

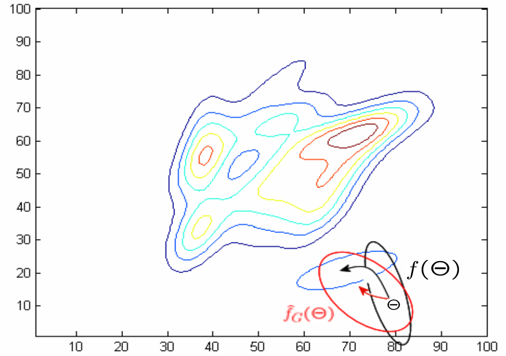
The prior model refers to a hybrid estimator :

$$N_i'(\Theta) = \frac{1}{\|V_i(\hat{\Theta}_i) + V(\hat{\Theta})\|^{1/2}} \mathrm{K}\left([V_i(\hat{\Theta}_i) + V(\hat{\Theta})]^{-1/2}(\Theta - \hat{\Theta}_i)\right)$$

The solution as well as the uncertainties estimates are jointly propagated

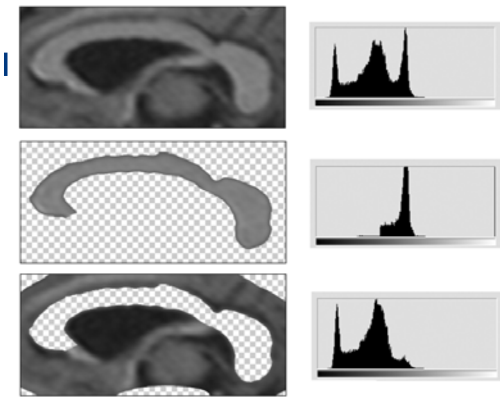
The prior term becomes far less important in areas

and in the direction of weak uncertainties

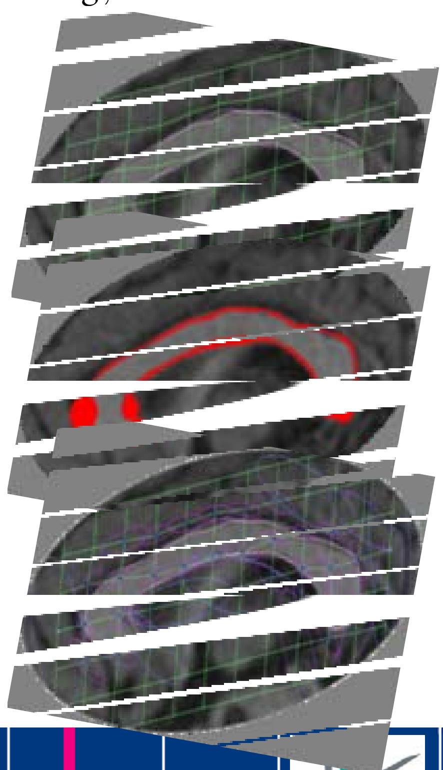


Segmentation with non-parametric prior

The image term aims at separating the histograms betwee background:



$$E_{im}(\Theta) = \int_{\Omega_{\mathcal{S},in}} p_{in}[I(T(\mathbf{x},\Theta))]\, d\mathbf{x} + \int_{\Omega_{\mathcal{S},out}} p_{out}[I(T(\mathbf{x},\Theta))]\, d\mathbf{x}$$



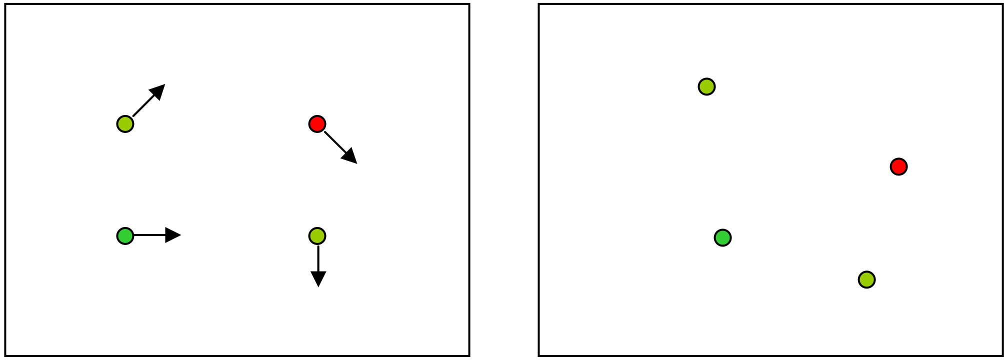
We can minimize this cost using a gradient descent method while the Hessian matrix of the cost function can provide information on

$$V(\Theta) = (\sum_i \mathcal{X}^T_{\tilde{\mathbf{x}}_i}.\Sigma_i^{-}.\mathcal{X}_{\tilde{\mathbf{x}}_i})^{-}$$

the quality of the solution that integrate both shape and image information

Segmentation with non-parametric prior & uncertainties estimation

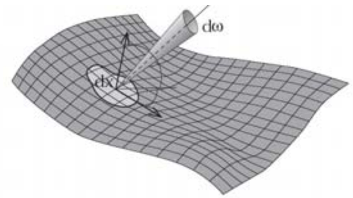
Optical Flow Estimation / Deformable Registration



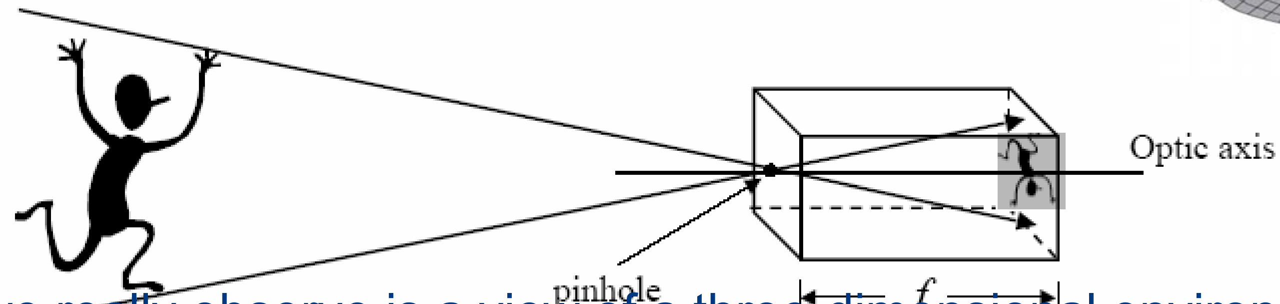
f g

- Let us consider two images (f,g) that are the projections of a real 3-D scene observed by a camera or acquisitions of the same organ part
- Let us consider (for simplicity we assume that the camera is static) some moving objects that are part of this scene and move in an independent fashion,
- Optical flow estimation **consists of recovering a motion field (3D or 2D)** that explains the motion of these objects in time

Image Formation



The Pinhole Model



What we really observe is a view of a three-dimensional environment, each pixel can be thought of as representing the “appearance” of a particular portion of this environment.

Such “appearance” is measured through the amount of incident light from the corresponding direction.

In practice the measurement depends on the camera’s radiometric and colorimetric response as well as the surface properties of the objects in the scene,

Optical Flow Definition

Let us a 3D patch (*s*) at time t and its projection to the image $\vec{\mathcal{P}}(s)$ according to some camera model

Let us consider the <u>same</u> 3D patch (*s*) at time t+1 and its projection to $\vec{\mathcal{P}'}(s)$ a new image according to the some camera model where either the camera has moved or the object has moved or both therefore

$$\vec{\mathcal{P}}(s) \# \vec{\mathcal{P}'}(s)$$

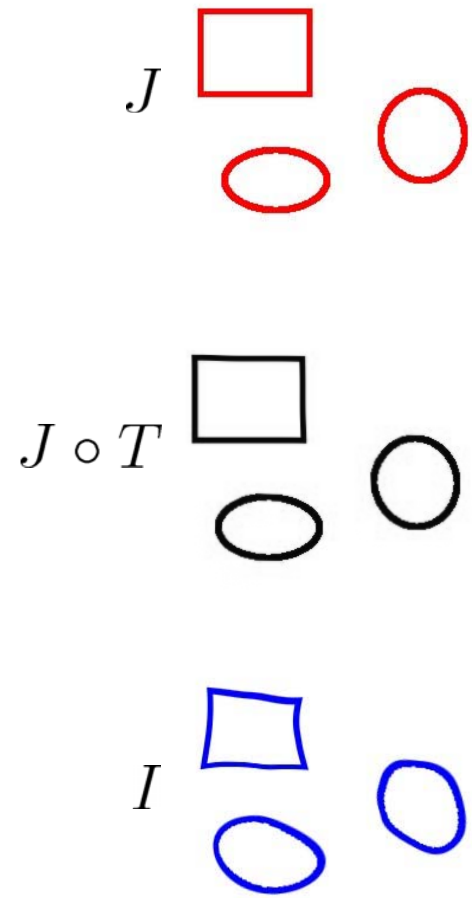
Optical Flow consists of recovering a motion vector (u, v) that explains the 2D motion of the 3D patch (*s*) in time:

$$(u, v) = (\vec{\mathcal{P}}(s) - \vec{\mathcal{P}'}(s))$$
$$(u, v) + \vec{\mathcal{P}}(s) = \vec{\mathcal{P}'}(s)$$

Basic Idea of Intensity-based Registration

Image registration as an optimization problem

$$T^* = \arg\min_T \phi\left(I, J \circ T\right)$$



Target and source Image:

$$I, J : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}$$

Transformation:

$$T(\mathbf{x}) = \mathbf{x} + D(\mathbf{x})$$

Image metric:

$$\phi : (I, J) \mapsto \mathbb{R}$$

The Aim of Registration…

…is to recover the transformation which involves

- the definition of a **transformation type**
- the definition of a **image metric**
- the definition of an **optimization strategy**

Energy Formulation

Data term based on image metric

(e.g. SSD)

$$E_{\text{data}}(D) = \int_{\Omega} \left(I(\mathbf{x}) - J(\mathbf{x} + D(\mathbf{x}))\right)^2 d\mathbf{x}$$

Smoothness constraints due to the ill-posedness

(e.g. penalizing gradients of the displacement field)

$$E_{\text{smooth}}(D) = \int_{\Omega} \|\nabla D(\mathbf{x})\| \, d\mathbf{x}$$

Optical Flow Estimation

$$f(\vec{\mathcal{P}}(s)) = g(\mathcal{P}(s) \vec{+} (u,v))$$

$$E(u,v) = min \int \int |f(\vec{\mathcal{P}}(\omega)) - g(\vec{(}\mathcal{P}(\omega) + (u,v))| d\omega$$

One can perform estimation through the optimization of a cost function

Optical Flow Constraint

Where $(u, v) \in [\Omega \rightarrow R, \Omega \rightarrow R]$

One can consider a Taylor expansion of the constraint

$$g(\mathcal{P}(s) \vec{+} (u, v)) = u\ g_x(\vec{\mathcal{P}}(s)) + v\ g_y(\vec{\mathcal{P}}(s)) + g(\vec{\mathcal{P}}(s))$$

That will lead to the follow constraint on the estimation of the optical flow, the famous Horn-Schunck equation [1981]

PROBLEM:

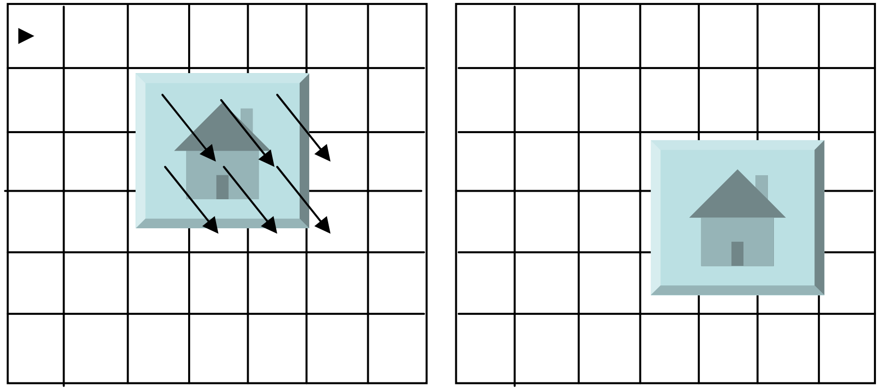
- **number of constraints is lower than the number of unknown variables**

SOLUTION

- **Consider the problem in subspaces with more than one constraints**
- **Consider regularization constraints when recovering dense motion fields**

The Lucas-Kanade approach

Let us consider the image divided into a certain number equal size of patches (blocks)



One can consider that

each block has constant

velocity in the image

plane and therefore

the same constraint

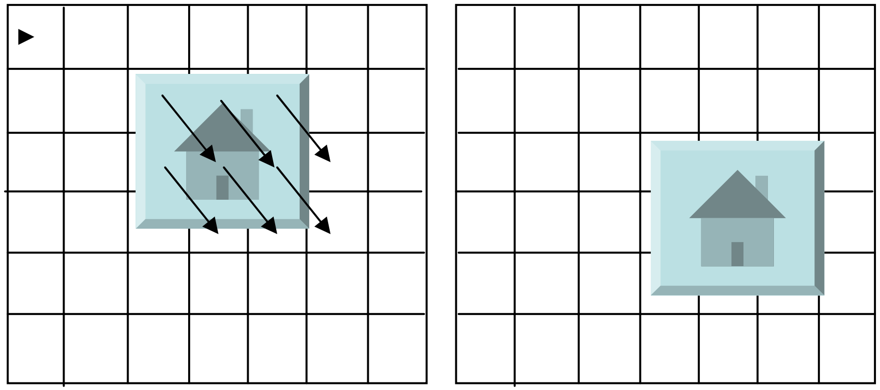
should be applicable to all pixels within the image blob

$$E(u,v) = min \int_{-w}^{w}\int_{-w}^{w}((u,v)\nabla g(\omega)+g(\omega)-f(\omega))^2 d\omega$$

The lucas-kanade method consists of estimating optical flow for each block by solving a linear system of through the calculus of variations

The Lucas-Kanade approach

$$\begin{cases} \dfrac{\partial E}{\partial u}(u, v) = 0 \\ \dfrac{\partial E}{\partial v}(u, v) = 0 \end{cases}$$



That can lead to a $\mathbf{A} \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{B}$

linear inference problem

$$\mathbf{A} = \begin{bmatrix} \sum g_x g_x & \sum g_x g_y \\ \sum g_x g_y & \sum g_y g_y \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} -\sum g_x (f - g) \\ -\sum g_y (f - g) \end{bmatrix}$$

Introducing Regularization Constraints

In the most general $$(u,v) \in [\Omega \rightarrow R, \Omega \rightarrow R]$$

One can define the optical flow constraint in the entire image domain:

$$E(u,v) = min \int\int |(u,v)\nabla g(\omega) + g(\omega) - f(\omega)| d\omega$$

Solving ill-posed problems is a frequent problem in computer vision

Since quite often the scenes that we do observe are smooth, one can assume that the optical flow estimates (the displacement of these scenes) are smooth

$$(u,v) : (u_x, u_y, v_x, v_y) \rightarrow \mathbf{0}$$

Or

$$u_x^2 + u_y^2 + v_x^2 + v_y^2 = \epsilon \rightarrow 0$$

Introducing Regularization Constraints

Such a smoothness constraint can be integrated with the optical flow constraint leading to the following objective function…

$$E(u,v) = min \int\int |(u,v)\nabla g(\omega) + g(\omega) - f(\omega)| d\omega$$
$$+ w \int\int (u_x^2 + u_y^2 + v_x^2 + v_y^2) d\omega$$

One can optimize this cost function:

- Partial Differential Equations & Gradient Descent Methods
- Mean-field/simulated Annealing and Dynamic Programming, Graph-based Optimization techniques…

PROBLEMS:

- the cost function is not convex
- the initial conditions are quite important
- Sensitive to the initial conditions, non-robust behaviour

Introducing Robustness to the Process

Standard least square estimators suffer from outliers…

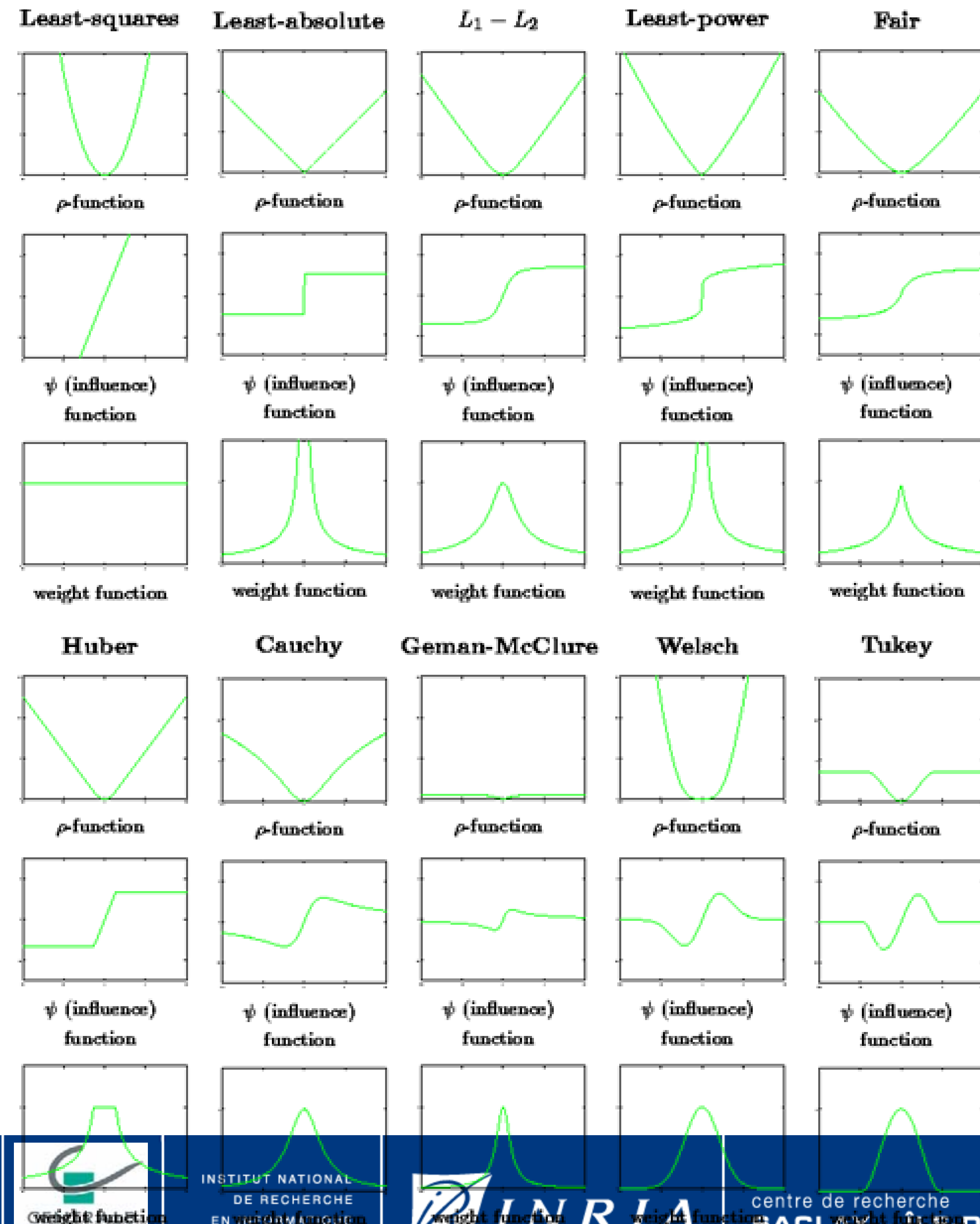
$$E(u,v) = min \int \int |(u,v)\nabla g(\omega) + g(\omega) - f(\omega)| d\omega$$

One can replace the $\rho(;)$-two norm with more appropriate error norms like

$$E(u,v) = min \int \int \rho\left((u,v)\nabla g(\omega) + g(\omega) - f(\omega)\right) d\omega$$

Where is a bounded error function and the optimization of the new objective function turns out to be equivalent with a re-weighted least square approach

Such an approach is far more appropriate for motion estimation where the number of constraints is marginal to the number of variables to be recovered.


Least-squares
Least-absolute
$L_1 - L_2$
Least-power
Fair
ρ-function
ψ (influence) function
weight function
Huber
Cauchy
Geman-McClure
Welsch
Tukey
ρ-function
ψ (influence) function
weight function

Parametric Motion Models

Let us consider the case where motion is observed because of the change of the position of the camera…

In that case the changes we observe in the images can be somehow modeled using a global transformation between the two views, In other words one can assume a motion model

$$\mathcal{T}(\Theta;\omega)$$

That consists of a limited number of parameters Θ

If we consider planar patches of constant depth, one can say that their motion can be approximated using such a simplistic model, or

$$(u(\omega),v(\omega))=\mathcal{T}(\omega)$$

Examples of Parametric Models

Similarity

$$\mathcal{T}(x,y) = \underbrace{s\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}}_{\begin{bmatrix} a & b \\ -b & a \end{bmatrix}} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix}$$

Consists of a scale (zoom in/out), translation and rotation

Affine

$$\mathcal{T}(x,y) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

That is an extension of the similarity model

Parametric/Global Registration

One now can reformulate the problem of motion estimation using these parametric models

$$E(\mathcal{T}) = min \int\int \rho(\mathcal{T}(\omega)\nabla g(\omega) + g(\omega) - f(\omega))d\omega$$

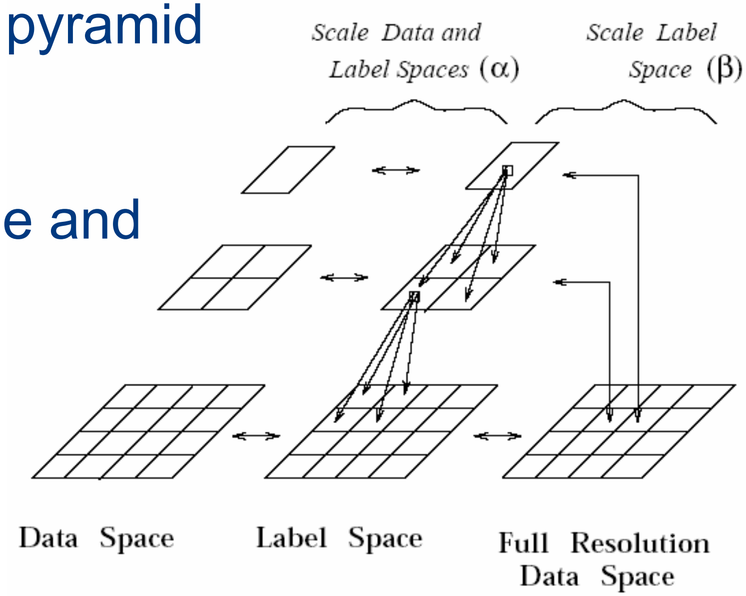
Such objective function refers to 4-6 unknown parameters (the parameters of the motion model) while one constrain can be recovered for every pixel of the image

In the case of similarity model, recovering these parameters is equivalent with solving the following linear system

$$\begin{cases} \frac{\partial E}{\partial a}(\mathcal{T}) = 0, & \frac{\partial E}{\partial b}(\mathcal{T}) = 0 \\ \frac{\partial E}{\partial c}(\mathcal{T}) = 0, & \frac{\partial E}{\partial d}(\mathcal{T}) = 0 \end{cases}$$

Multi-Scale Approaches

Smooth the input images and create a pyramid



Consider the images at the lowest scale and estimate motion in this scale

Use this estimation as an initial guess for the next scale, and re-estimate the flow

Repeat the process until convergence

Motivation for the Use of MRFs

- Most common approach is gradient descent

BUT:

- Optimization of non-convex objective functions is difficult (many local minima)
- Introduction of arbirtry metric is not possible (sensitive to the derivative of the cost function)
- Computational complexity is an issue

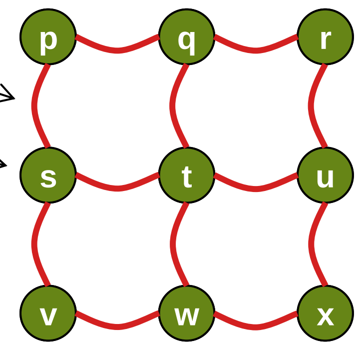
Discrete Labeling Problem

Markov Random Field formulation with pairwise interactions

$$E_{\mathrm{mrf}}(\mathbf{l}) = \sum_{p \in G} V_p(l_p) + \sum_{(p,q) \in N} V_{pq}(l_p, l_q)$$

Edges

Nodes



Unary potentials (matching):

$$V_p(l_p) = \underbrace{\int_{\Omega} \hat{\eta}(\mathbf{x}) \left(I(\mathbf{x}) - J(\mathbf{x} + \mathbf{d}^{l_p}) \right)^2 d\mathbf{x}}_{\text{or any other local image metric}}$$

Pairwise potentials (smoothness):

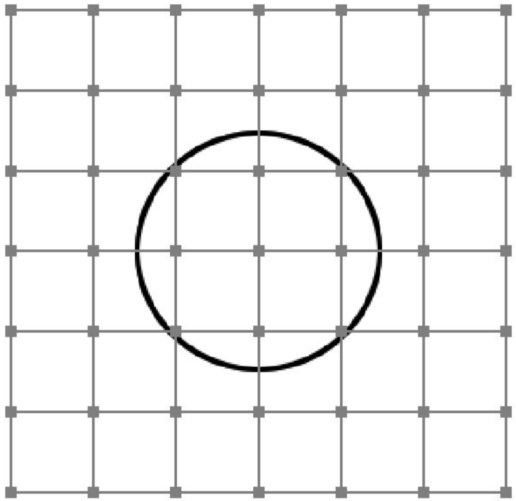
$$V_{pq}(l_p, l_q) = \lambda \left\| \mathbf{d}^{l_p} - \mathbf{d}^{l_q} \right\|$$

Dimensionality Reduction

- Linear combination of control points

$$D(\mathbf{x}) = \sum_{i}^{M} \eta(\mathbf{x})\, \mathbf{d}_i$$

$$\eta : \text{basis functions} \quad \mathbf{d} : \text{displacements}$$



- e.g. Free-Form Deformations (Sederberg et al. 1986; Rueckert et al. 1999), or **THIN PLATE SPLINES, etc...**

(Weighted) Block Matching

- Redefinition of data term w.r.t. control lattice

$$E_{\text{data}}(D) = \sum_{i}^{M} \int_{\Omega} \hat{\eta}(\mathbf{x}) \ (I(\mathbf{x}) - J(\mathbf{x} + D(\mathbf{x})))^2 \, d\mathbf{x}$$

$$\text{with} \quad \hat{\eta}(\mathbf{x}) = \frac{\eta(\mathbf{x})}{\int_{\Omega} \eta(\mathbf{y}) \, d\mathbf{y}}$$

- Pixel-wise image metrics weighted by normalized basis functions
  - image points closer to a control point gain more influence on its matching energy
- Statistical image metrics (e.g. mutual information, cross correlation)
  - evaluation of image metric in local patches centered at the control points
  - block size depends on control lattice resolution

From a Continuous to a Discrete Model

Let us consider a predefined ordered set of displacements

$$\Theta = \{\mathbf{u}_1, ..., \mathbf{u}_i\}$$

Associated with a set of labels

$$\{\omega_1, ..., \omega_i\}$$

Then, the problem of image registration can be expressed as follows

$$E(\omega) = \sum_{\mathbf{m}\in\mathcal{G}} \iint_{\Omega} \eta(|\mathbf{x} - \mathbf{m}|)\rho_h(g(\mathbf{x}), f(\mathbf{x} + \mathbf{u}_{\omega(\mathbf{x})}))d\mathbf{x}$$

That is associate to each node of the deformation, a displacement that once used along with interpolation method, optimizes the similarity metric between the target and the transformed source.

Where each node of the transformation domain is associated with a particular cost, and when the set of labels yields to infinity, we have a formulation that is almost equivalent to the continuous one

Registration via Block Matching

Simple iterative algorithm:

1. Find control point displacements that minimize the matching energy
2. Interpolate dense displacement field
3. Warp source image

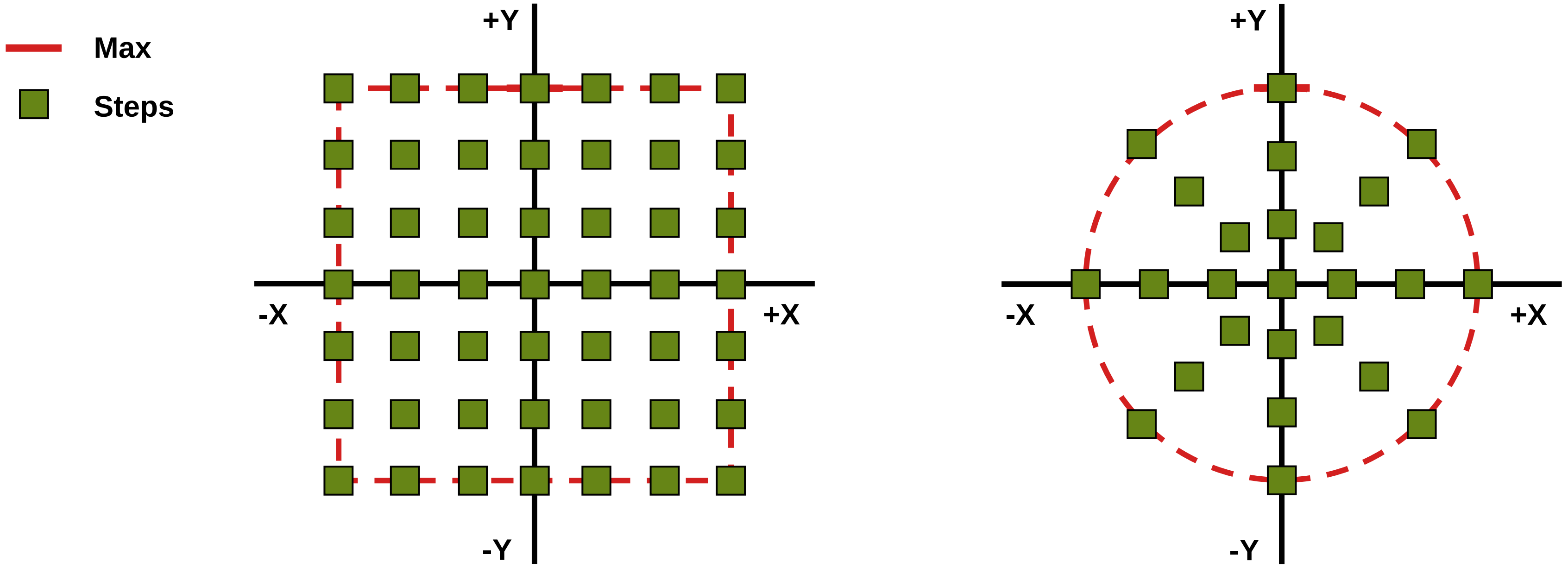
Our solution:

- Cast the problem as a discrete labeling of a Markov Random Field
- Use efficient global optimization techniques

Search Space Discretization

Consider a discrete set of labels and a quantized version of the displacement space

$$L = \{l_1, ..., l_m\} \quad \Theta = \{\mathbf{d}_1, ..., \mathbf{d}_m\}$$

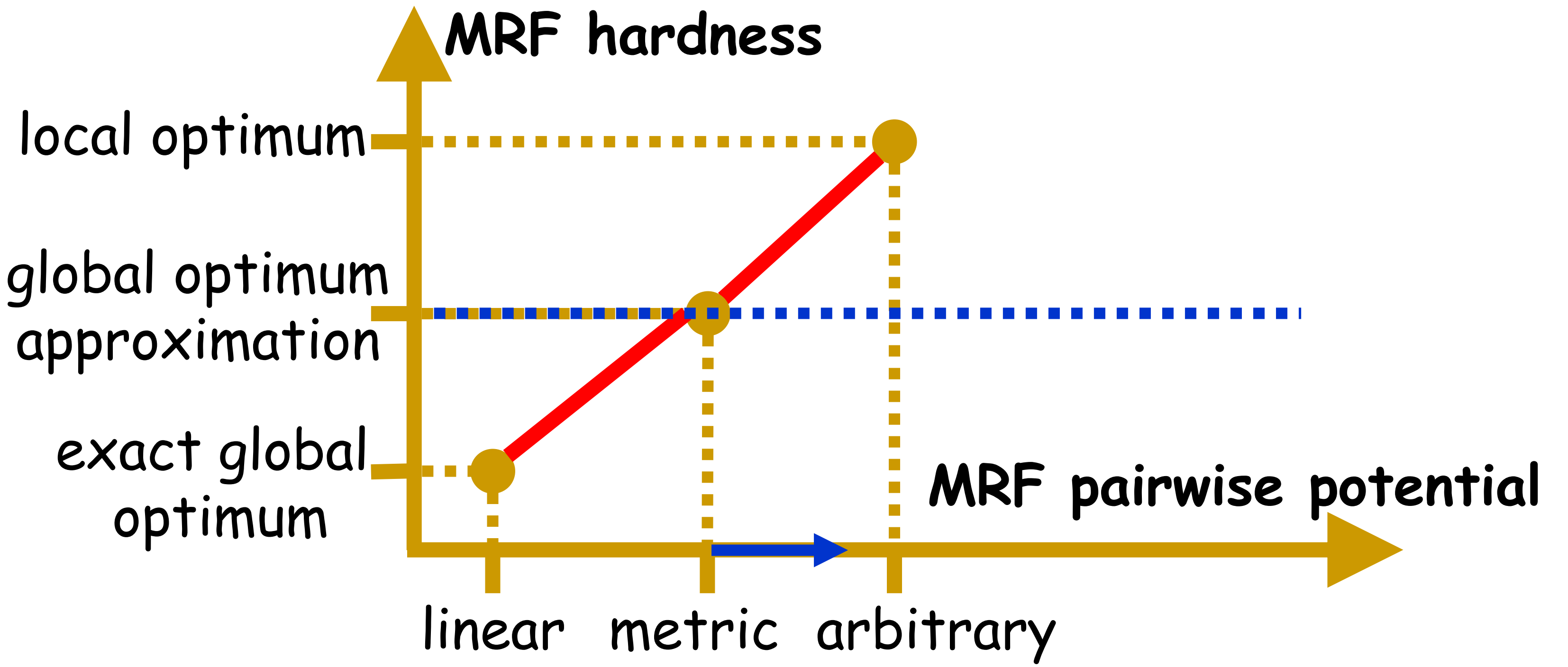


Dense sampling

Sparse sampling

MRF hardness

Courtecy N. Komodakis



Move left in the horizontal axis,

and remain low in the vertical axis
(i.e., still be able to provide approximately optimal solutions)

But we want to be able to do that efficiently, i.e. fast

How to handle MRF optimization?

- Unfortunately, discrete MRF optimization is extremely hard (a.k.a. NP-hard)
  - E.g., highly non-convex energies

- So what do we do?
  - Is there a principled way of dealing with this problem?

- Well, first of all, we don't need to panic. Instead, we have to stay calm and **RELAX!**

- Actually, this idea of relaxing may not be such a bad idea after all…

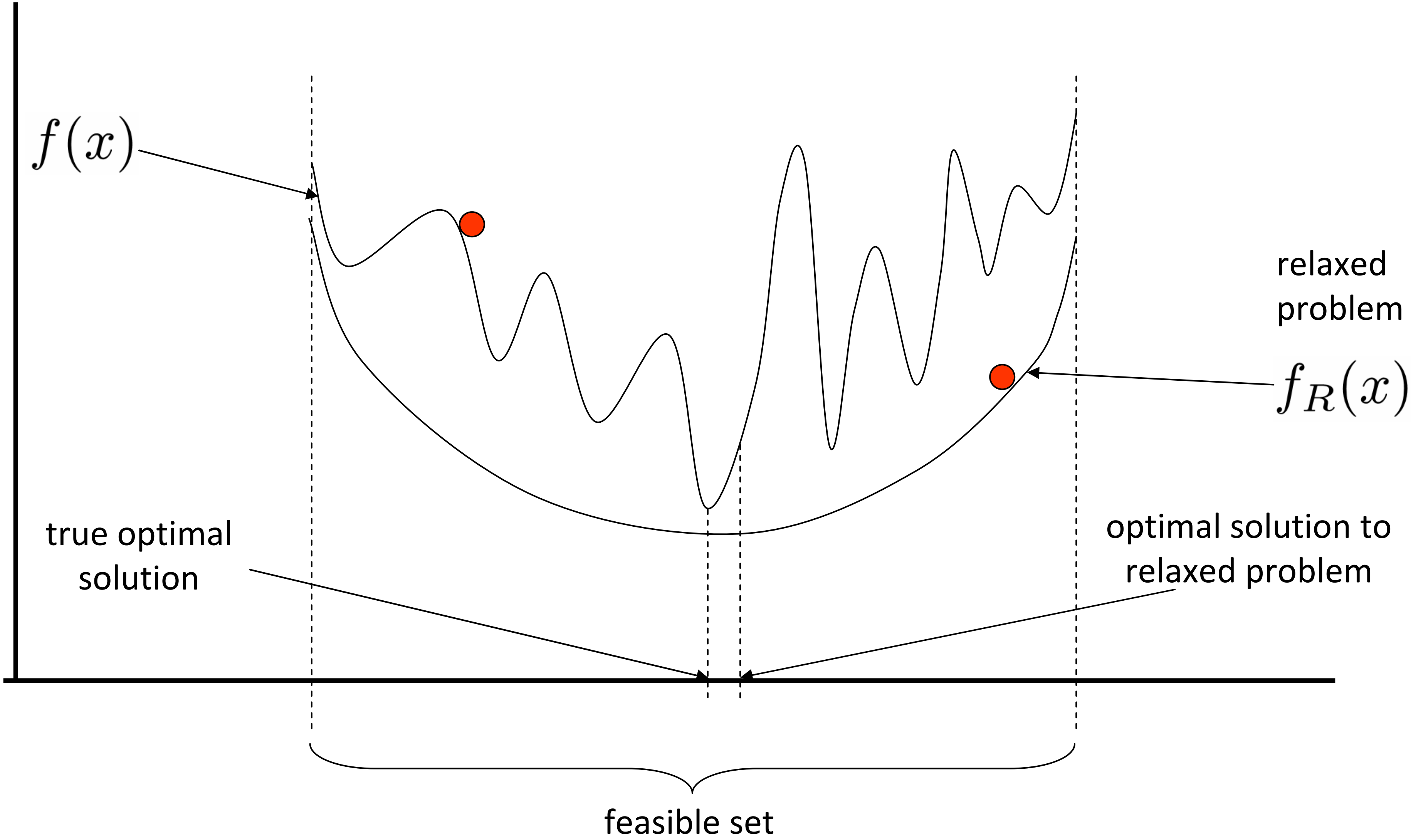
The relaxation technique

Very successful technique for dealing with difficult optimization problems

- It is based on the following simple idea:
  - try to approximate your original difficult problem with another one (the so called **relaxed problem**) which is easier to solve
- Practical assumptions:
  - Relaxed problem must always be easier to solve
  - Relaxed problem must be related to the original one

The relaxation technique

Courtecy N. Komodakis



How do we get a convex relaxation?

- By dropping some constraints
  (so that the enlarged feasible set is convex)

- By modifying the objective function
  (so that the new function is convex)

- By combining both of the above

Courtecy N. Komodakis

Linear programming (LP) relaxations

Optimize linear function subject to linear constraints, i.e.:

$$\min \mathbf{c}^T \mathbf{x}$$
$$\text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Very common form of a convex relaxation, because:
  - Typically leads to very efficient algorithms (important due to large scale nature of problems in computer vision)
  - Also often leads to combinatorial algorithms

Example 1: Primal-dual schema

The primal-dual schema

Courtecy N. Komodakis

- Say we seek an optimal solution x^* to the following integer program (this is our **primal** problem):

$$\min \mathbf{c}^T \mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{N}$$

(NP-hard problem)

- To find an approximate solution, we first relax the integrality constraints to get a primal & a dual linear program:

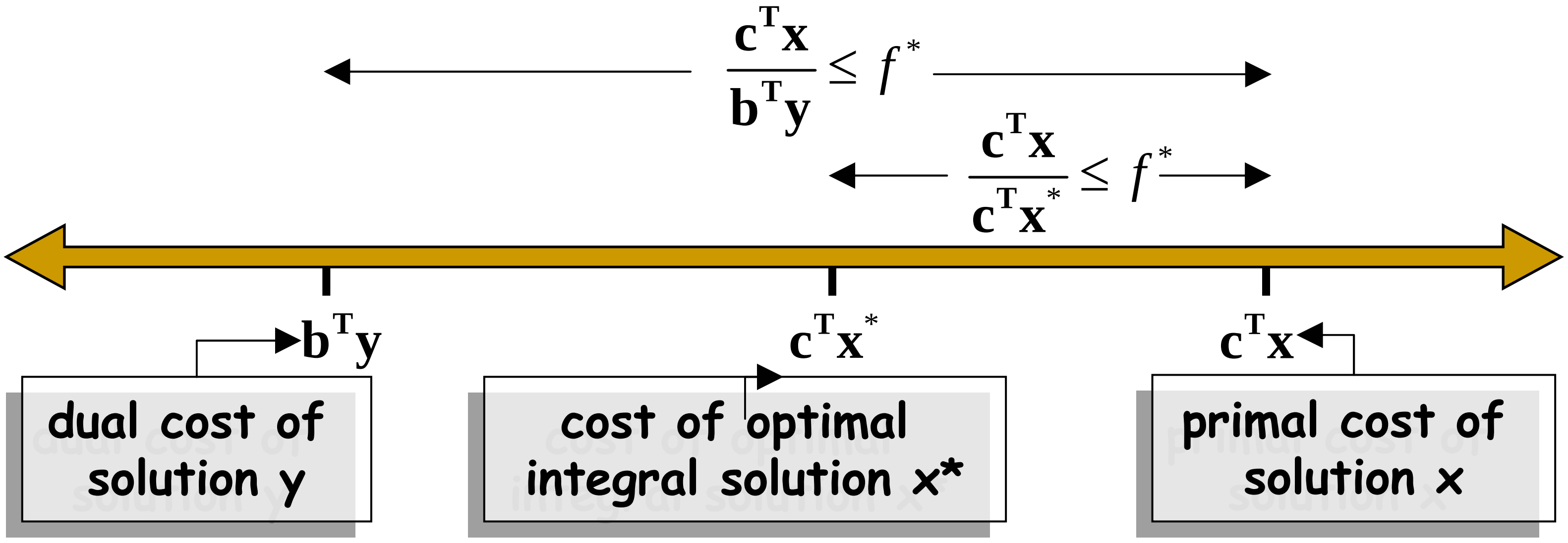
primal LP: $$\min \mathbf{c}^T \mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0$$

dual LP: $$\max \mathbf{b}^T \mathbf{y} \quad \text{s.t.} \quad \mathbf{A}^T \mathbf{y} \leq \mathbf{c}$$

The primal-dual schema

Courtecy N. Komodakis

Goal: find integral-primal solution x, feasible dual solution y such that their primal-dual costs are “close enough”, e.g.,

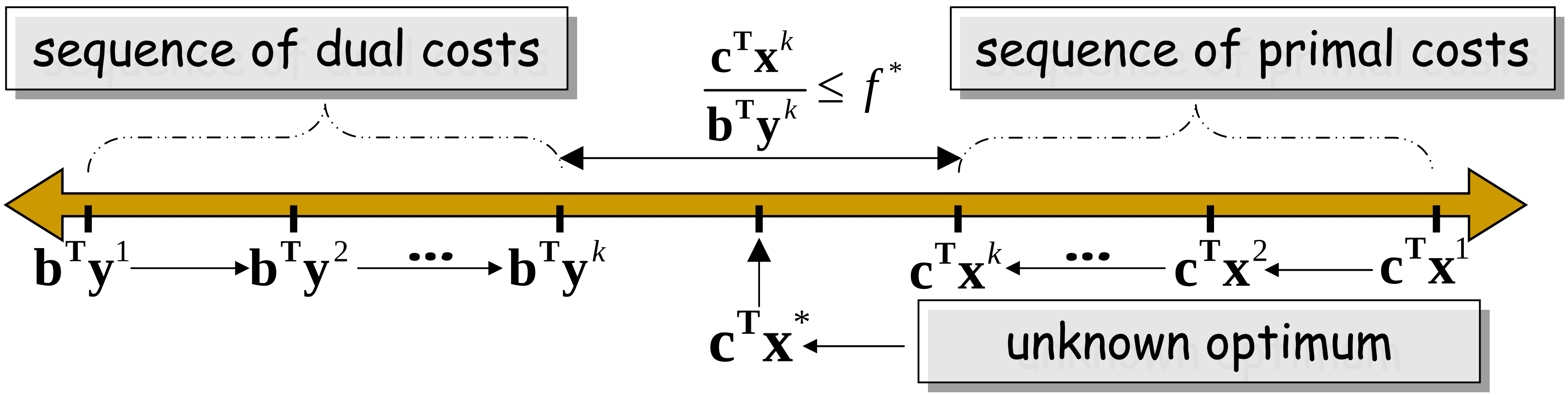


Then x is an f^*-approximation to optimal solution x^*

The primal-dual schema

Courtecy N. Komodakis

The primal-dual schema works iteratively



- Global effects, through local improvements!
- Instead of working directly with costs (usually not easy), use RELAXED complementary slackness conditions (easier)
- Different relaxations of complementary slackness → **Different approximation algorithms!!!**

Standard LP-relaxation for MRFs

$$\min_{\mathbf{x}} \sum_{p \in \mathcal{V}} \sum_{l \in \mathcal{L}} \bar{g}_p(l) x_p(l) + \sum_{pq \in \mathcal{E}} \sum_{l,l' \in \mathcal{L}} \bar{f}_{pq}(l,l') x_{pq}(l,l')$$

$$\text{s.t.} \sum_{l \in \mathcal{L}} x_p(l) = 1,$$

← only one label per object

$$\sum_{l' \in \mathcal{L}} x_{pp'}(l,l') = x_p(l),$$

$$\sum_{l \in \mathcal{L}} x_{pp'}(l,l') = x_{p'}(l'),$$

← consistency between $x_{pp'}(l,l')$ and $x_p(l)$, $x_{p'}(l')$

$$x_p(l) \geq 0, \quad x_{pp'}(l,l') \geq 0.$$

$x_p(l) = 1 \Leftrightarrow$ label l assigned to object p

$x_{pp'}(l,l') = 1 \Leftrightarrow$ labels l,l' are assigned to objects p, p'

The primal-dual schema for MRFs

During the PD schema for MRFs, it turns out that:

each update of primal and dual variables ⇄ solving max-flow in appropriately constructed graph

- Resulting flows tell us how to update both:
  - the dual variables, as well as
  - the primal variables

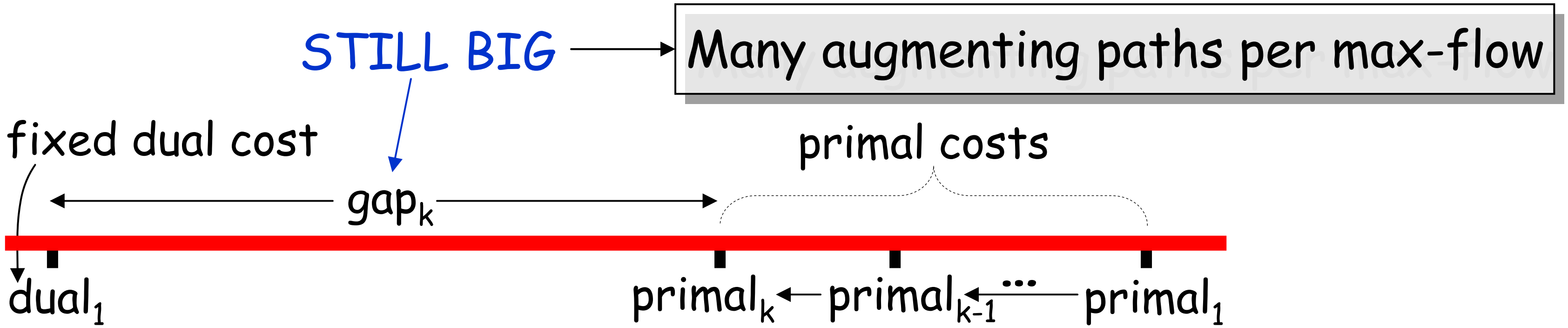
  } ← for each iteration of primal-dual schema

- Max-flow graph defined from current primal-dual pair (x^k,y^k)
  - (x^k,y^k) defines **connectivity** of max-flow graph
  - (x^k,y^k) defines **capacities** of max-flow graph

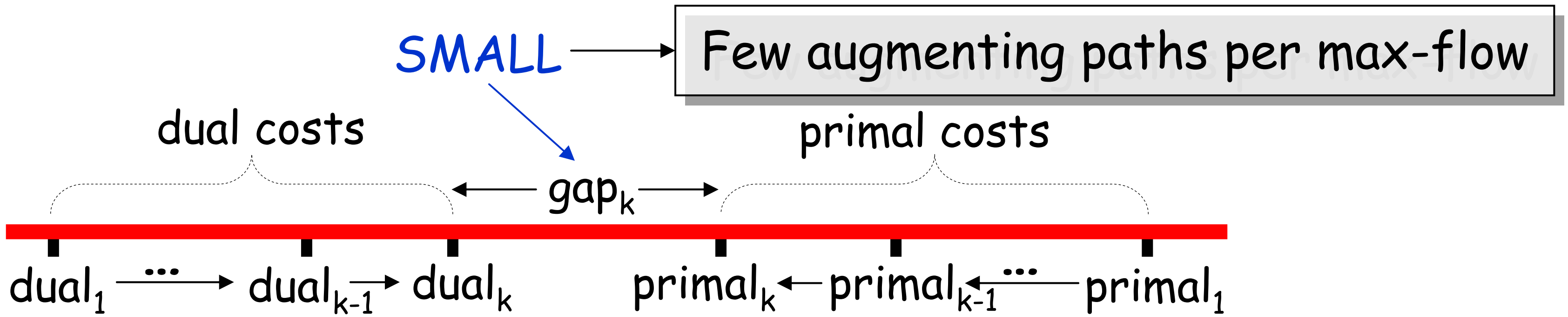
- Max-flow graph is thus continuously updated

Computational efficiency

MRF algorithm only in the primal domain (e.g., a-expansion)



- MRF algorithm in the primal-dual domain (Fast-PD)



Theorem: primal-dual gap = upper-bound on #augmenting paths (i.e., primal-dual gap indicative of time per max-flow)

Courtecy N. Komodakis

Beyond Pairwise Energies: Efficient Optimization for Higher-order MRFs

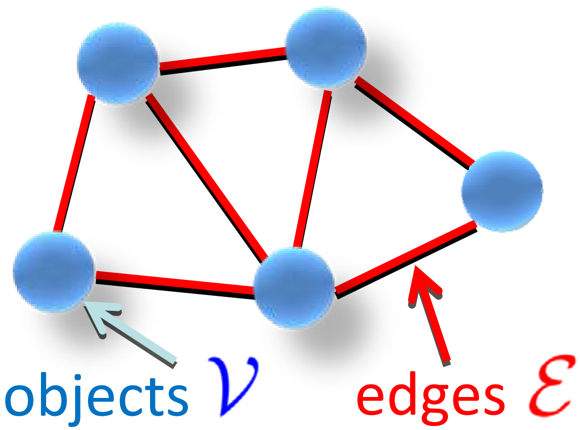
Introduction

Discrete MRF optimization

Courtecy N. Komodakis

Given:

- Objects $\mathcal{V}$ from a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- Discrete label set $\mathcal{L}$



- Assign labels (to objects) that minimize MRF energy:

$$\min_{\{x_p\}} \underbrace{\sum_{p \in \mathcal{V}} \bar{g}_p(x_p)}_{\text{unary potential}} + \underbrace{\sum_{pq \in \mathcal{E}} \bar{f}_{pq}(x_p, x_q)}_{\text{pairwise potential}}$$

unary potential — pairwise potential

- MRF optimization ubiquitous in vision (and beyond)

Introduction

Courtecy N. Komodakis

MRFs extremely popular in vision and beyond

- Stereo matching, optical flow, segmentation, object recognition, image completion, …
- Computer graphics, pattern recognition, …

- MRF optimization
  - Large amount of work over the last years
  - A success story
  - But mostly for **pairwise** MRFs

Introduction

Courtecy N. Komodakis

Why do we need higher-order MRFs?

- Pairwise MRFs often unable to provide a faithful modeling (e.g., compare **global optimum** vs **ground truth**)

- Higher-order potentials
  - Can capture multiple interactions
  - Allow for far more expressive priors
  - Much more accurate modeling

Introduction

Courtecy N. Komodakis

Unfortunately not much work on higher-order MRFs

- Mostly over the last few years (e.g., [Lan et al. 06], [Potetz 07], [Kohli et al. 07, 08], [Werner 08])

- Reasons?
  - Hardness
    (much more difficult problems to optimize)
  - Computational cost
    (typically prohibitive, e.g., exponential with clique size)

Contributions of this work

Courtecy N. Komodakis

- Powerful optimization framework for high-order MRFs
  - Relies on dual decomposition
  - Extremely general
  - Extremely flexible
    (easily adaptable to the problem's structure)
  - Leads to efficient algorithms that provide high-quality solutions

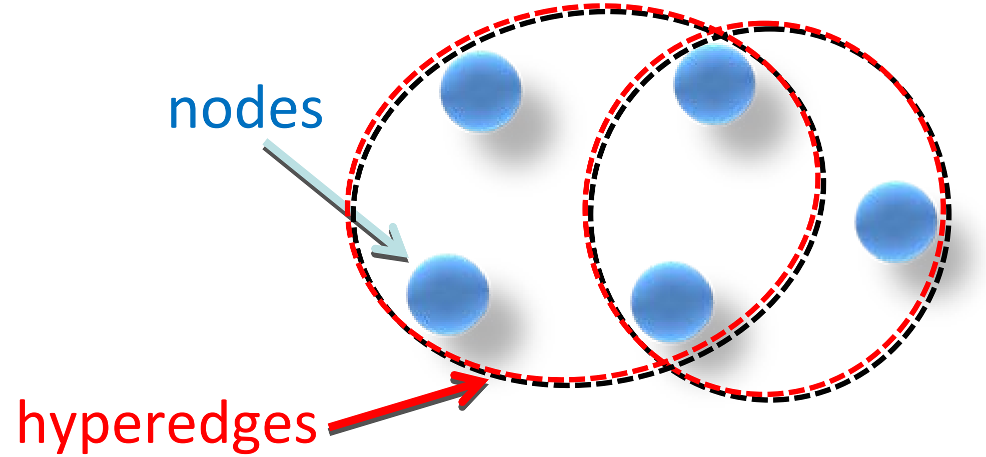
Optimization framework

Courtecy N. Komodakis

Optimization of high-order MRFs

Hypergraph $G = (\mathcal{V}, \mathcal{C})$

- Nodes $\mathcal{V}$
- Hyperedges/cliques $\mathcal{C}$



- High-order MRF energy minimization problem

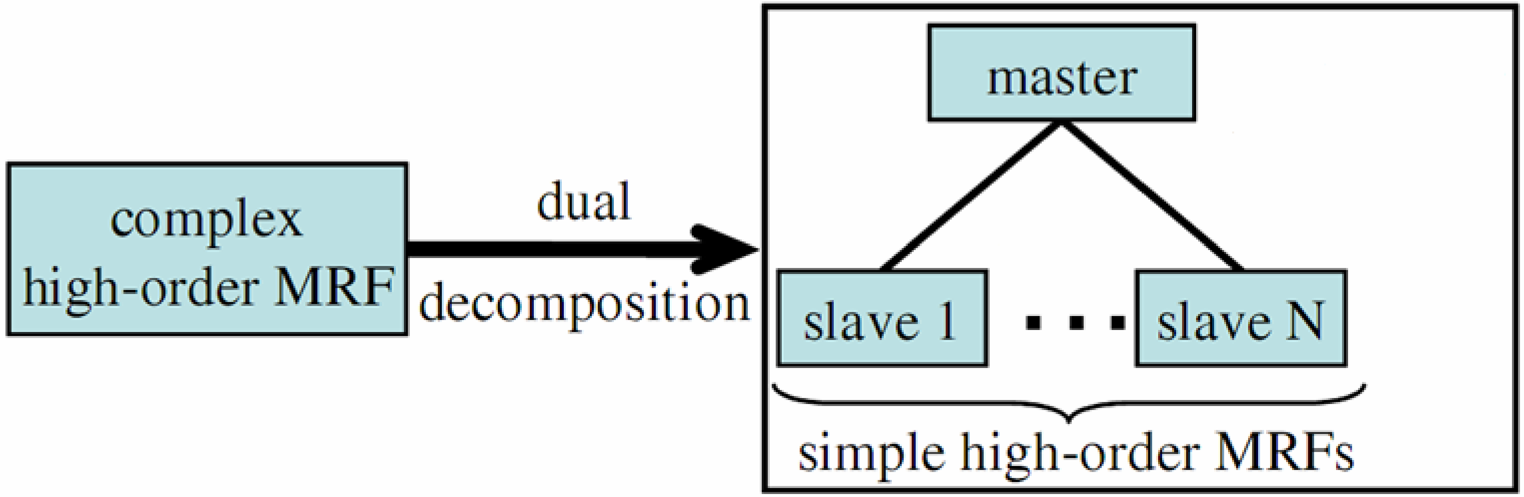
$$\mathrm{MRF}_G(\mathbf{U}, \mathbf{H}) \equiv \min_{\mathbf{x}} \sum_{q \in \mathcal{V}} \underbrace{U_q(x_q)}_{\text{unary potential (one per node)}} + \sum_{c \in \mathcal{C}} \underbrace{H_c(\mathbf{x}_c)}_{\text{high-order potential (one per clique)}}$$

unary potential (one per node) high-order potential (one per clique)

Courtecy N. Komodakis

Optimization of high-order MRFs

To handle the problem in full generality, we will rely on **dual decomposition** [Komodakis et al. 07]

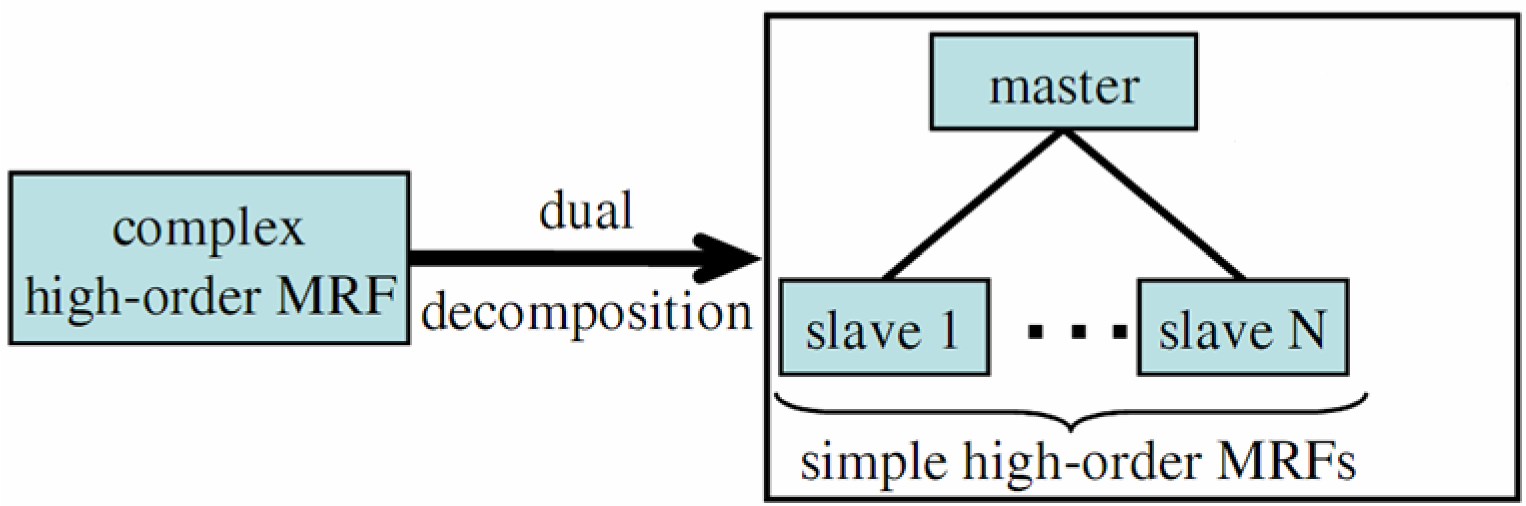


- Master = coordinator (has global view)
  Slaves = subproblems (have only local view)

Courtecy N. Komodakis

Optimization of high-order MRFs

To handle the problem in full generality, we will rely on **dual decomposition** [Komodakis et al. 07]

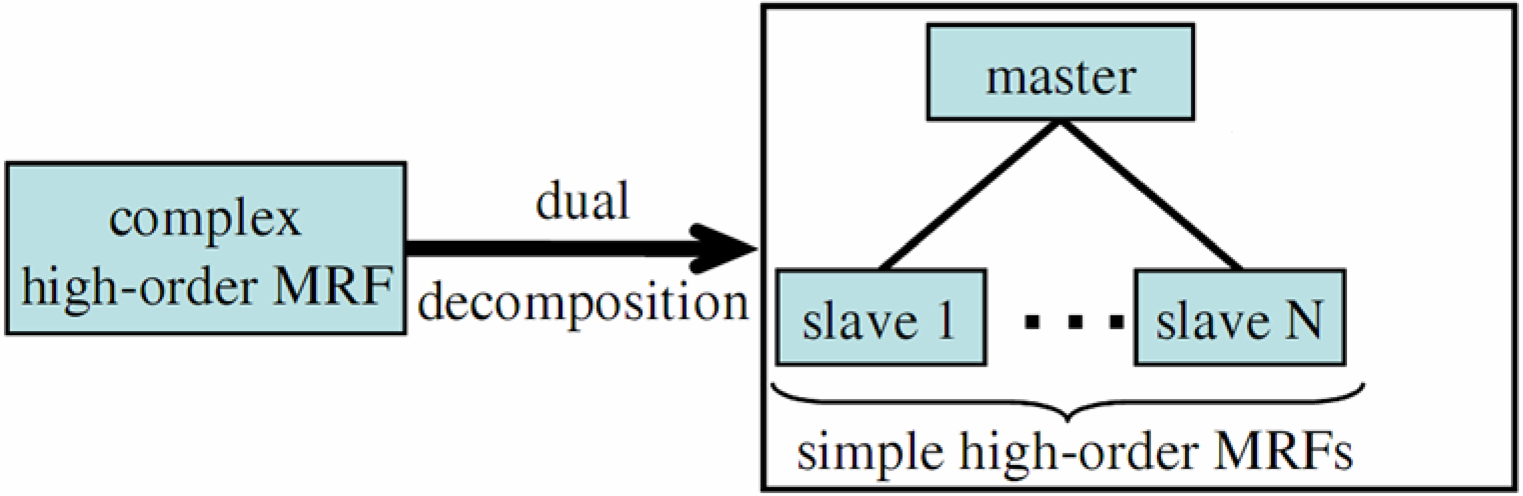


- Master = $\mathrm{MRF}_G(\mathbf{U}, \mathbf{H})$ (i.e., MRF on hypergraph *G*)

Courtecy N. Komodakis

Optimization of high-order MRFs

To handle the problem in full generality, we will rely on **dual decomposition** [Komodakis et al. 07]

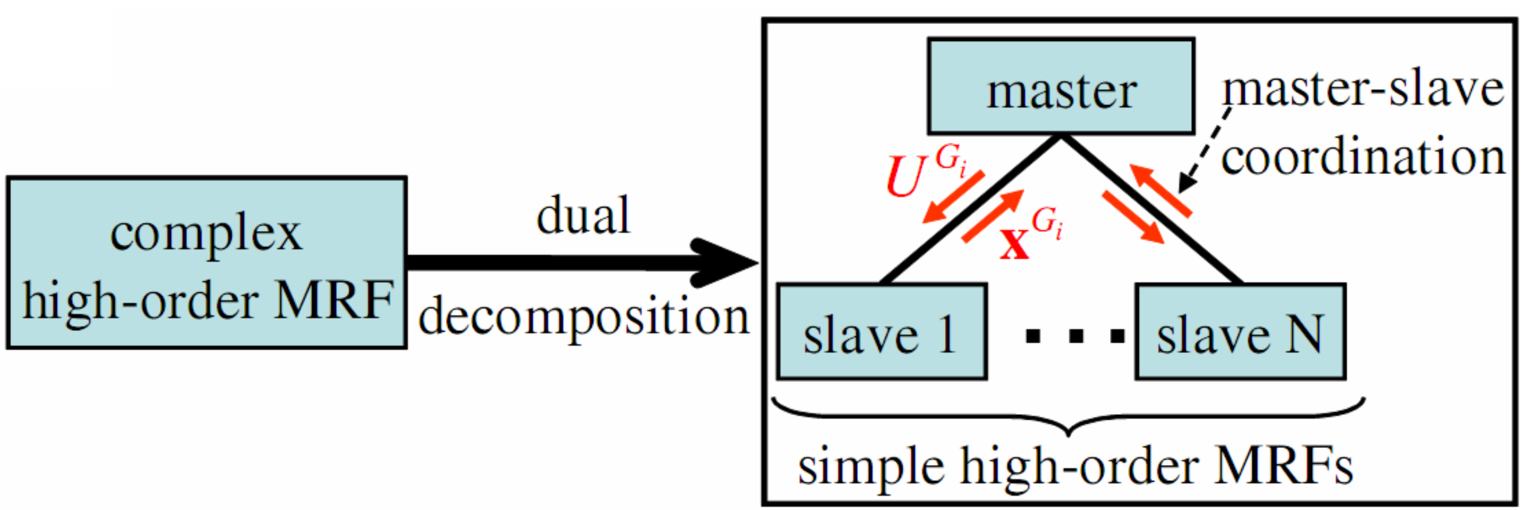


- Set of slaves = $\{\mathrm{MRF}_{G_i}(\mathbf{U}^{G_i}, \mathbf{H}^{G_i})\}$
  (MRFs on sub-hypergraphs G_i whose union covers G)

Courtecy N. Komodakis

Optimization of high-order MRFs

To handle the problem in full generality, we will rely on **dual decomposition** [Komodakis et al. 07]



- Optimization proceeds in an iterative fashion via **master-slave coordination**

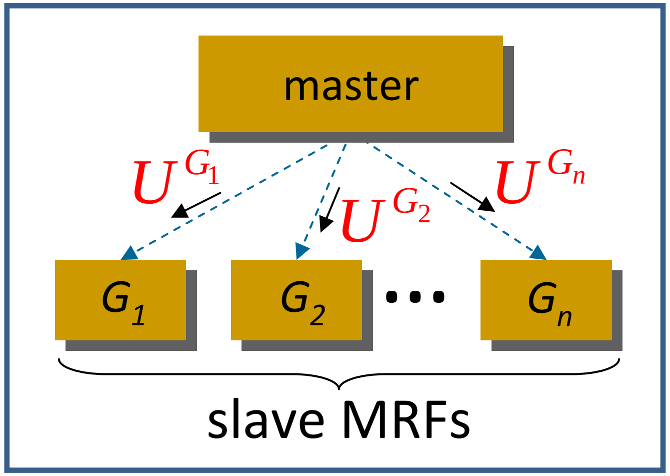
Courtecy N. Komodakis

"Hey, Slaves, …"

Master sends to slaves current unary potentials and requests them to optimize their problems

$$\left\{ U^{G_i} \right\}$$

master talks to slaves



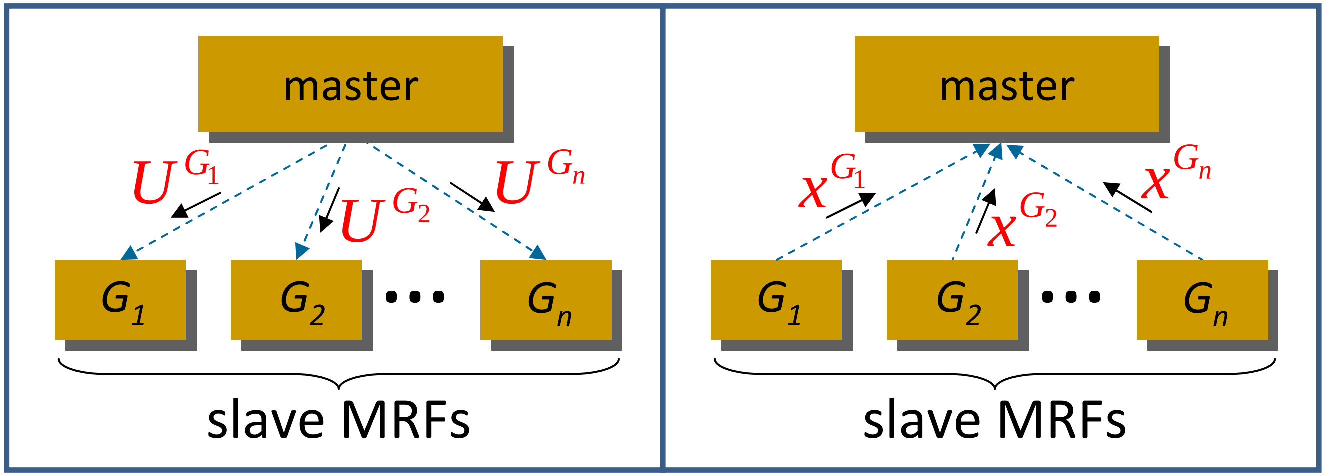
Courtecy N. Komodakis

"What is it that you want master?"

Slaves obey to the master by solving $\{\mathrm{MRF}_{G_i}(\mathbf{U}^{G_i}, \mathbf{H}^{G_i})\}$

and sending back to him the resulting minimizers $\{x^{G_i}\}$

master talks to slaves

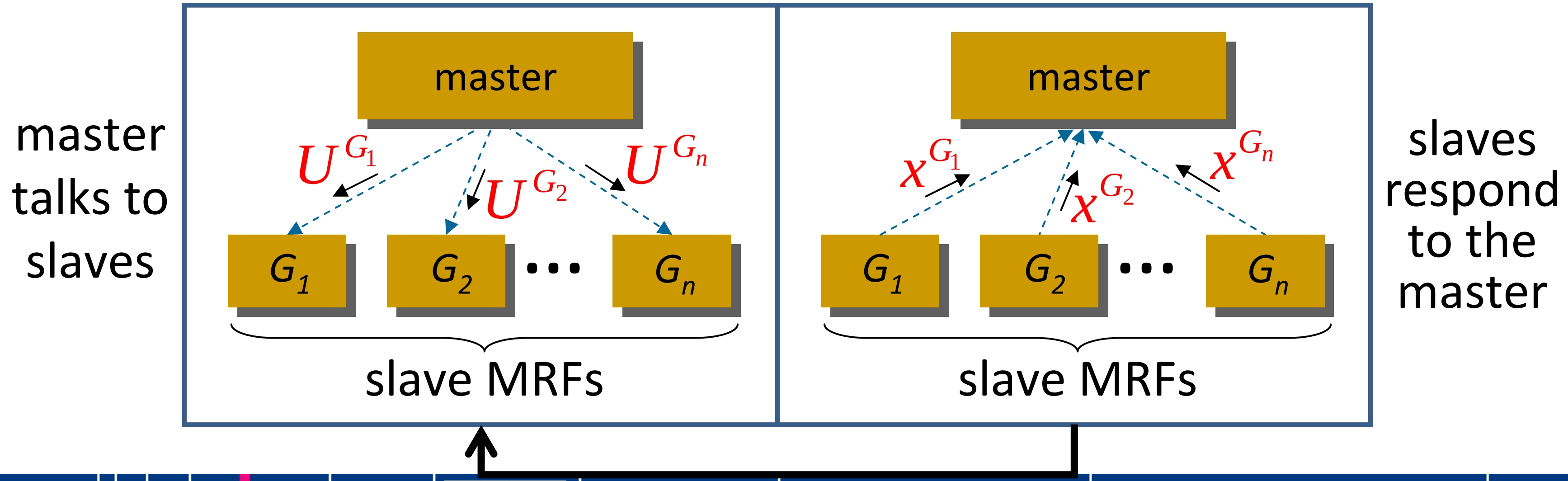


slaves respond to the master

Courtecy N. Komodakis

Optimization of high-order MRFs

- Master collects minimizers and reupdates $\{U^{G_i}\}$

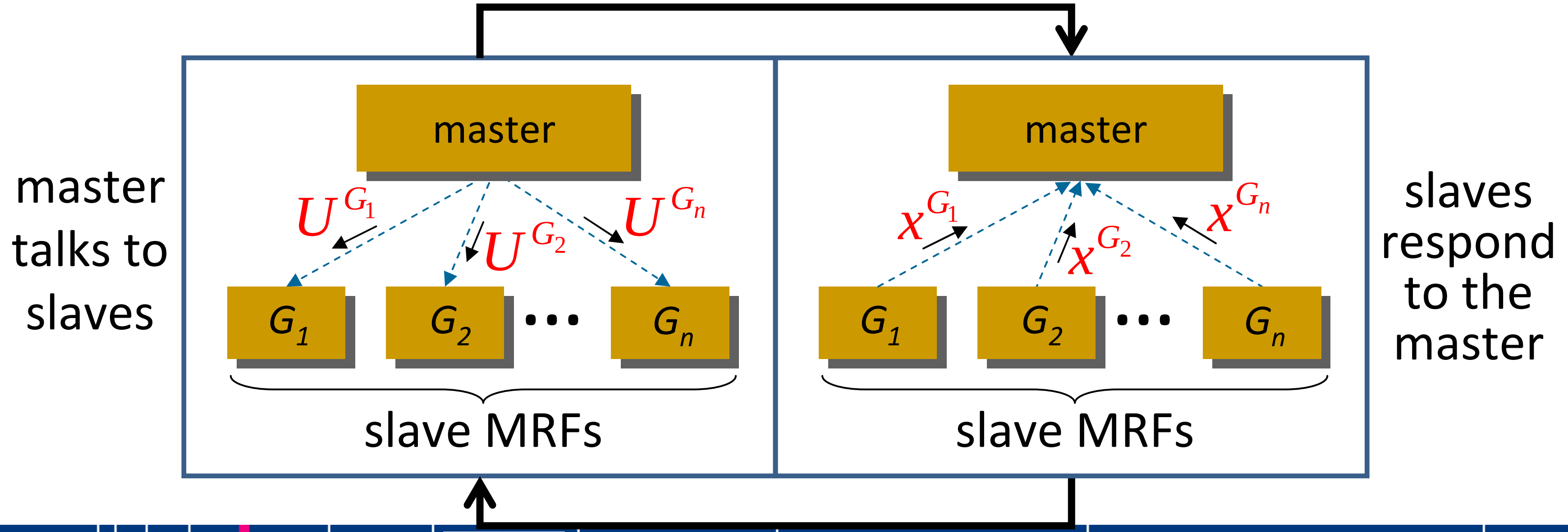


Courtecy N. Komodakis

Optimization of high-order MRFs

Process repeats until convergence
(many variations of the above basic scheme possible)

At the end, a solution can be extracted by, e.g., copying the slave minimizers



Courtecy N. Komodakis

Optimization of high-order MRFs

- User only needs to focus on how to solve the slaves (rest are taken care by the framework)

Slaves, i.e., MRF_{G_i} can be freely chosen by the user as long as (provides great flexibility)

$$G = \cup_i G_i$$

- For each choice of slaves, master solves a (possibly different) dual relaxation
  - Sum of slave energies = lower bound on MRF optimum
  - Dual relaxation = maximum such bound

- Choosing more difficult slaves $\Rightarrow$ tighter lower bounds $\Rightarrow$ tighter dual relaxations

Courtecy N. Komodakis

Generic optimizer for high-order MRFs

Uses the following simple choice of slaves:

- One slave per clique $c \in \mathcal{C}$
- Corresponding sub-hypergraph $G_c = (\mathcal{V}_c, \mathcal{C}_c)$:

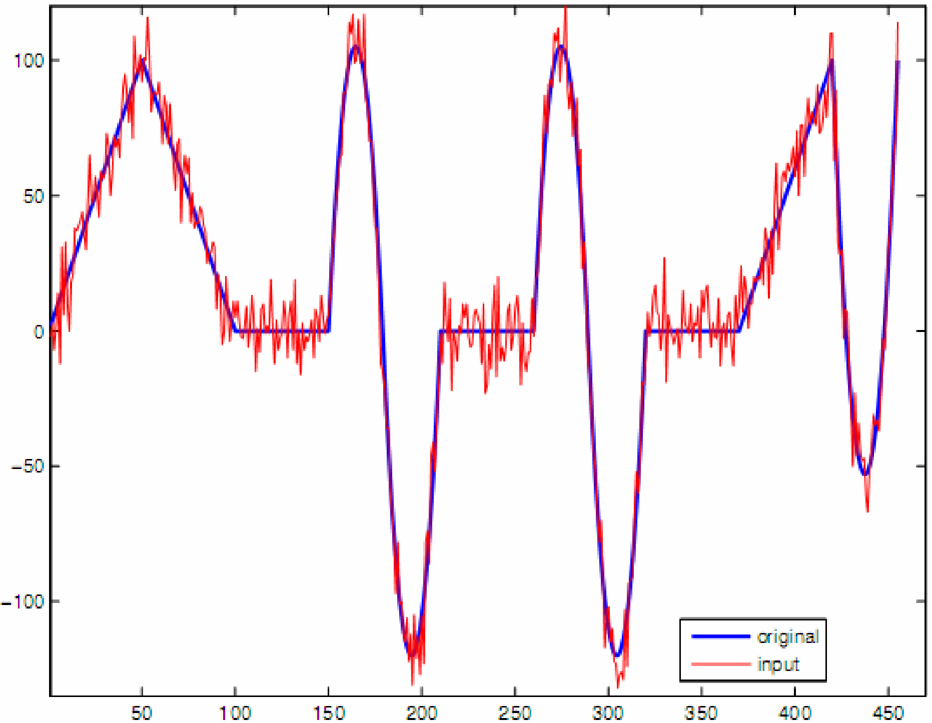
$$\mathcal{V}_c = \{q | q \in c\}, \mathcal{C}_c = \{c\}$$

- Resulting slaves often easy (or even trivial) to solve even if global problem is complex and NP-hard
  - widely applicable algorithm
- Corresponding dual relaxation is an LP
  - Generalizes well known LP relaxation for pairwise MRFs (at the core of most state-of-the-art methods)

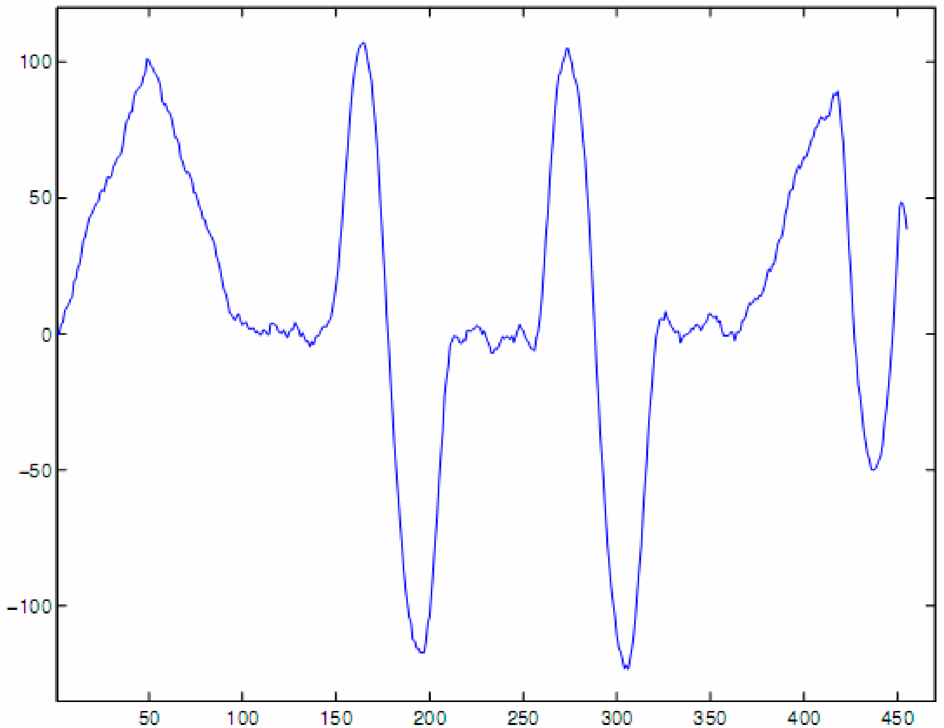
Experimental results

Courtecy N. Komodakis

Signal reconstruction



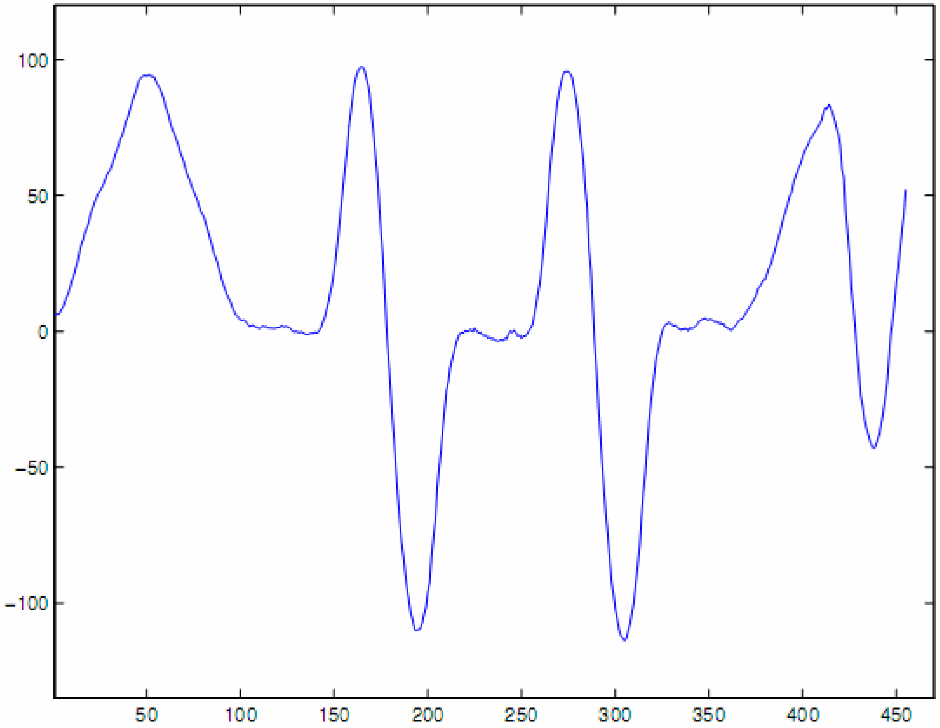
corrupted signal (red)
original signal (blue)



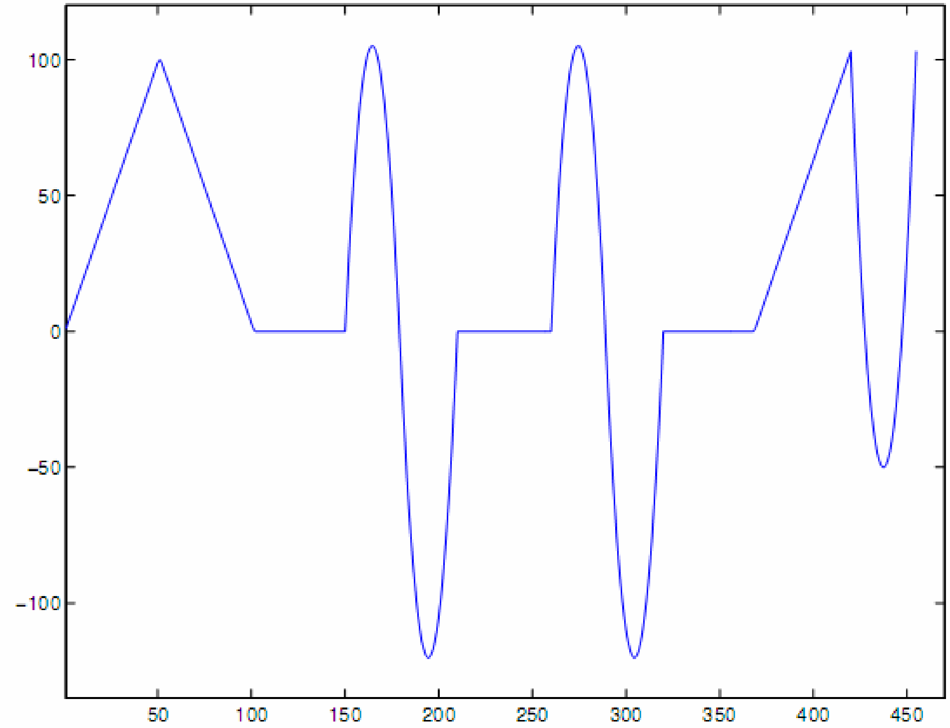
Gaussian filter result

Signal reconstruction

Courtecy N. Komodakis



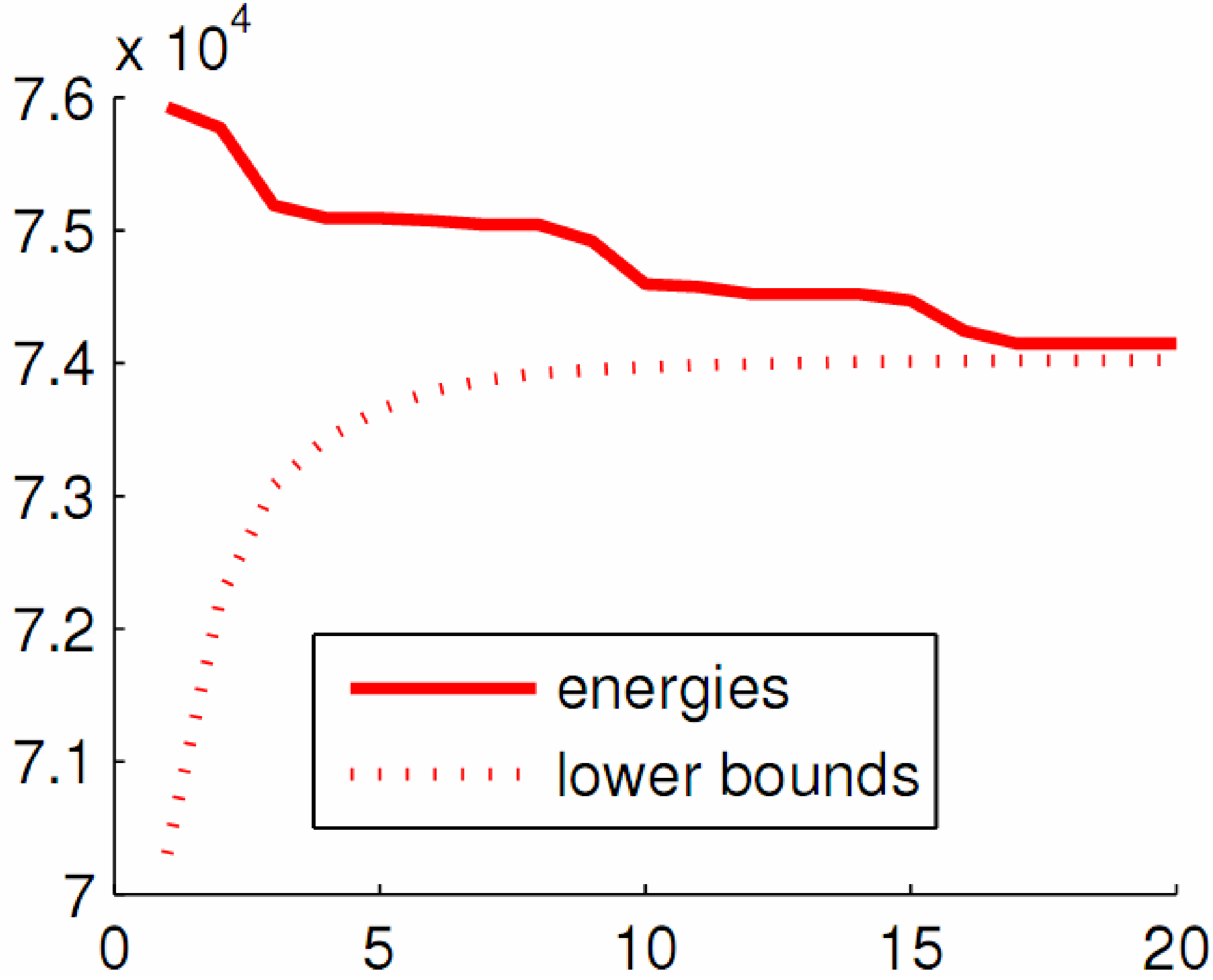
bilateral filter result



our result with 4th-order truncated potential

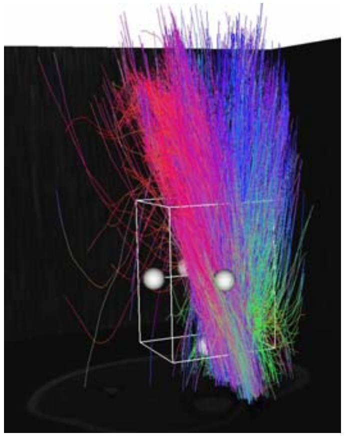
Courtecy N. Komodakis

Random $\mathcal{P}^n$ Potts model



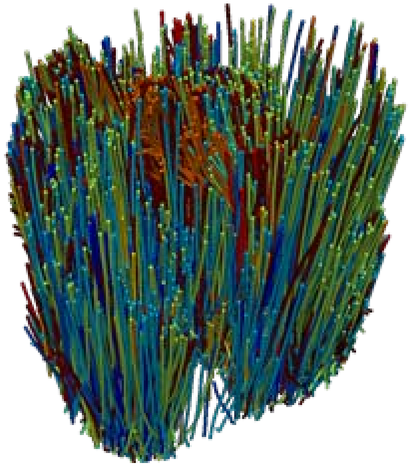
Random $\mathcal{P}^{3\times3}$ Potts model on 50x50 grid with 10 labels

Diffusion Tensor Imaging in Human Skeletal Muscle



Nikos Paragios

http://vision.mas.ecp.fr



Radhouene Neji
Salma Essafi
Ecole Centrale de Paris

Jean-Francois Deux
Guillaume Bassez
Alain Rahmuni
CHU-Henry Mondor Hospital

Brownian Motion in the human body

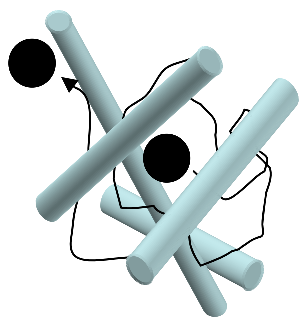
Isotropic Diffusion in fluids

Anisotropic Diffusion in the human body

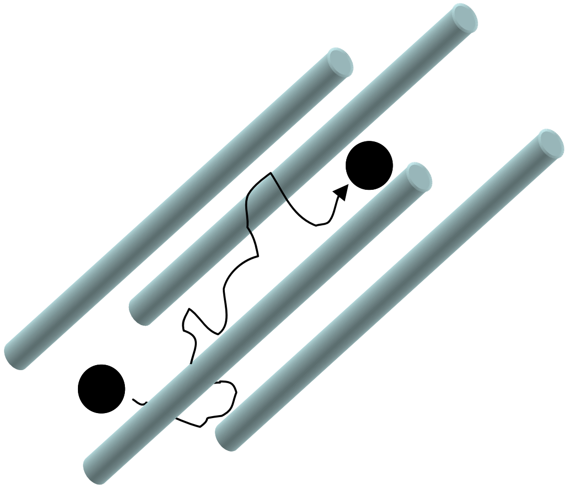
BUT after averaging over the voxel: isotropic

Only in very structured tissues anisotropic diffusion is measured

Isotropic



Anisotropic

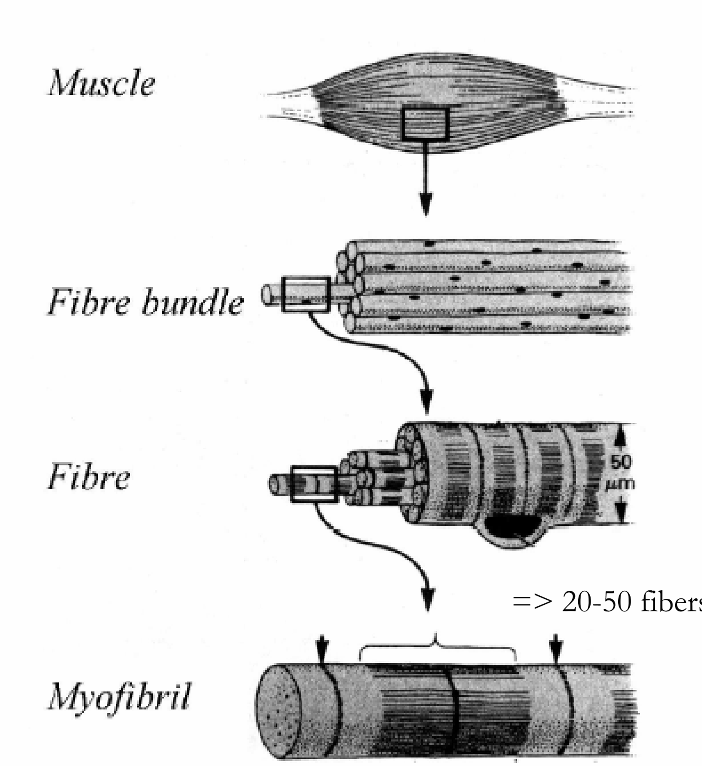


Images from S. Mori

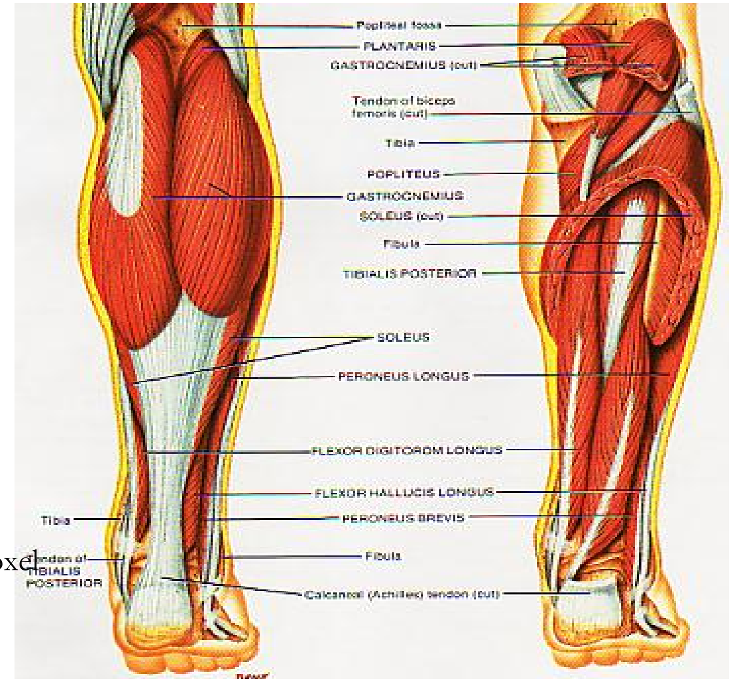
Anisotropic Diffusion in Muscle

Very structured tissue in muscle

=> Using DTI to reconstruct muscle structure



=> 20-50 fibers per voxel



Motivation for the project (I)

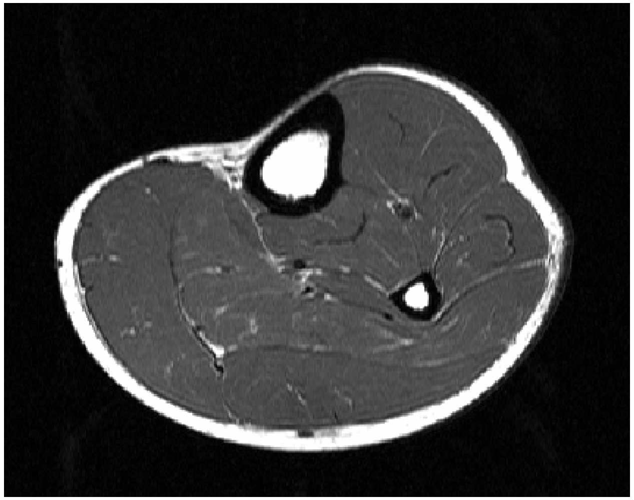
Why is the knowledge about structure important?

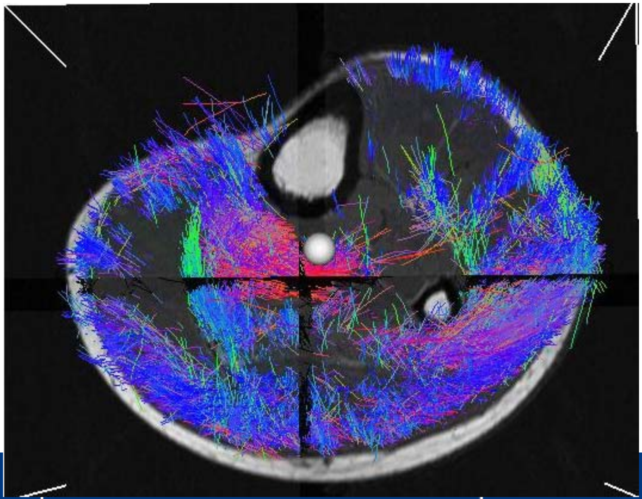
- Mechanical behavior depends strongly on muscle architecture

Why using DTI?

- "Standard" imaging modalities do not give information about the architecture
- Biopsys have several disadvantages:
  - Local
  - No measurements over time possible
  - No in vivo measurements

DTI is the only approach available to non-invasively study the three-dimensional architecture of muscle





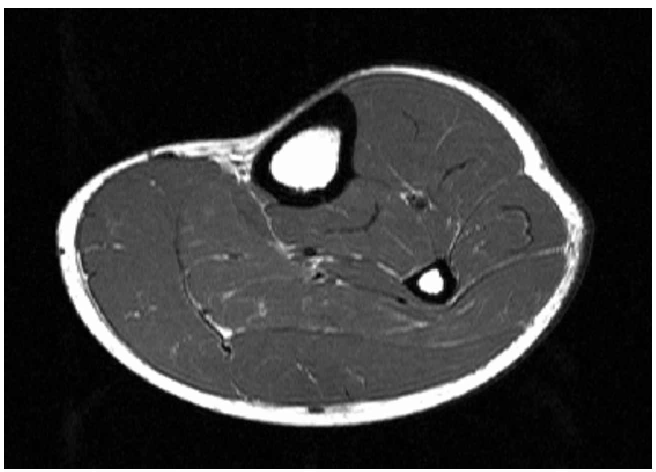
Motivation for the project (II)

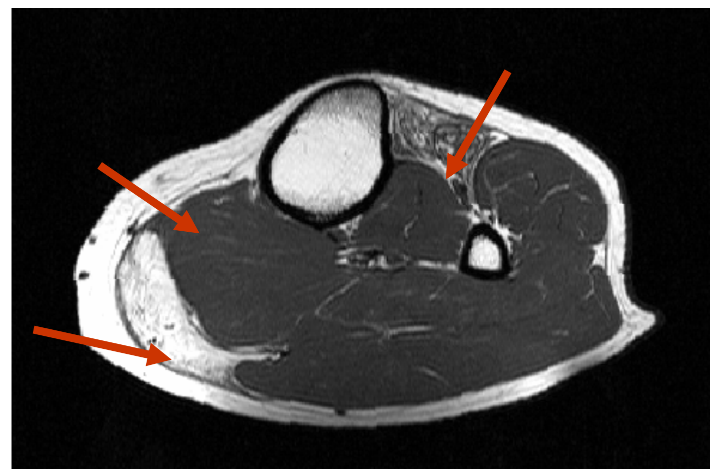
Muscular Diseases (Myopathy):

- Muscle cells are replaced by fat cells
- Lower performance of the muscle

How is the evolution of the disease?

How much muscle performance rests?





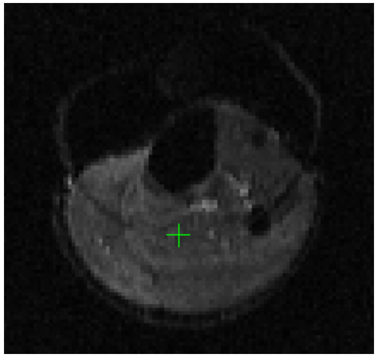
healthy

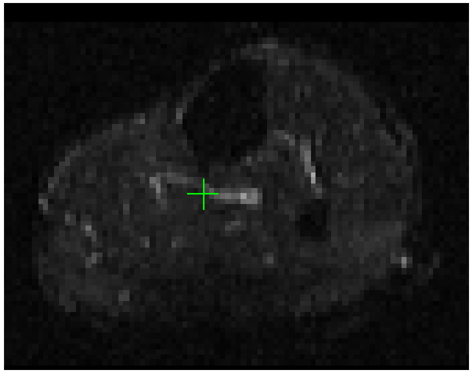
Non-healthy

Image Acquisition

Diffusion Weighted Echo-Planar Imaging with:

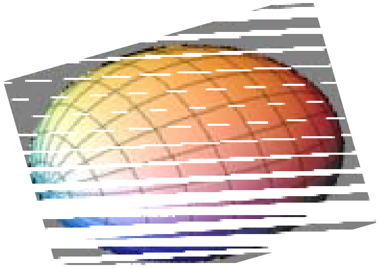
- Siemens Symphony 1,5 T
- 12 gradient directions
- b-value: 0 s/mm^2 and 450 s/mm^2
- Effective voxel size: 1.8 x 1.8 x **7.8** mm^3 later 1.8 x 1.8 x **5.6** mm^3
- Volume: 23 x 23 x **15.6** cm^3 later 23 x 23 x **11.2** cm^3
- Acquisition matrix: 128 x 128 x 20





Images taken at CHU-Henry Mondor Hospital (Créteil, Paris)

Mathematical Model



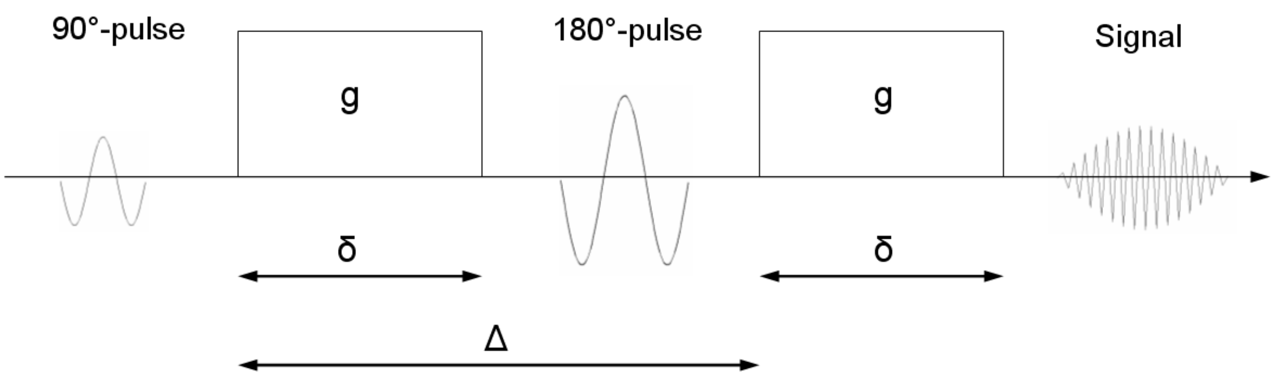
Basser proposed to use an **ellipsoid** to model anisotropic Gaussian Diffusion

Ellipsoids are described by a 3x3 symmetric positive-definite matrix (6 DOF)

=> Each voxel has to measured from at least 6 non-collinear directions

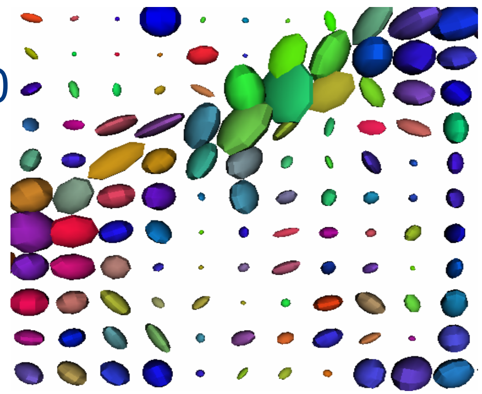
Stejskal-Tanner imaging sequence to measure diffusion

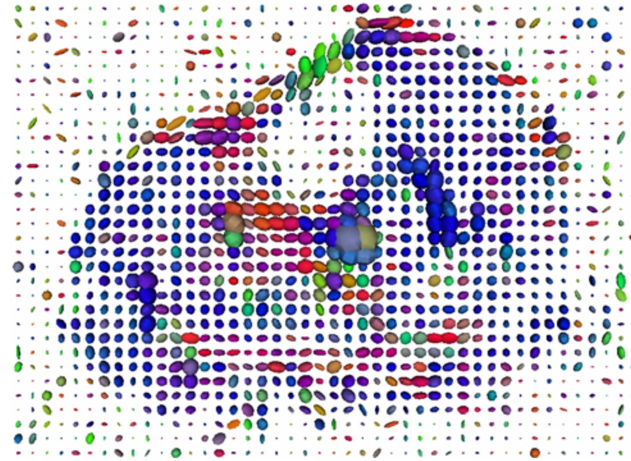
Spins that have completed a location change due to the Brownian motion during the time period Δ



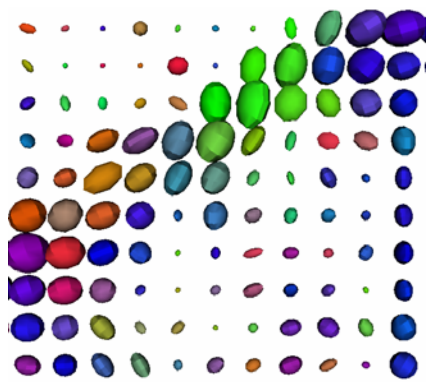
Effect of regularization

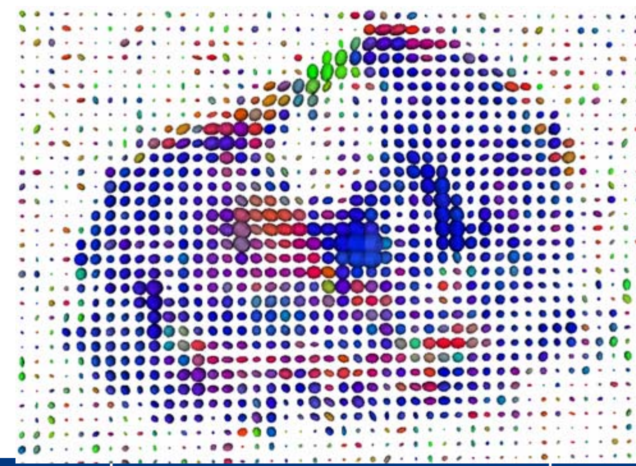
With $\alpha = 0$





With $\alpha = 0.5$



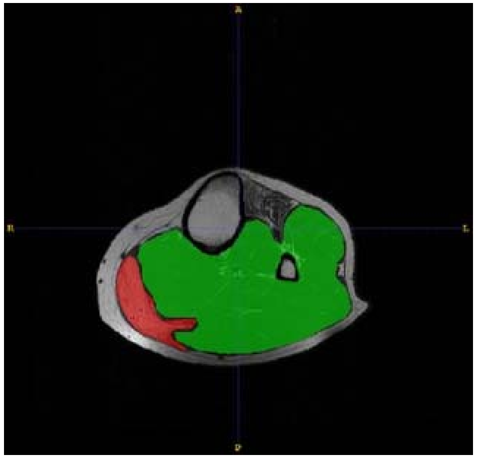


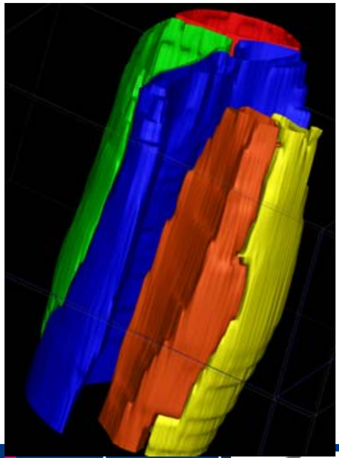
Segmentation

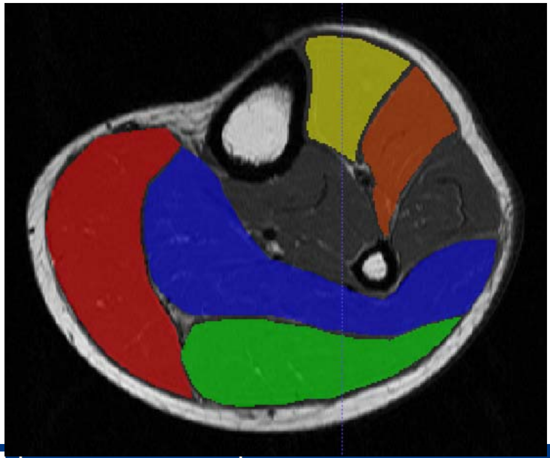
Segmentation of **muscle groups** and **ill/heatlhy**

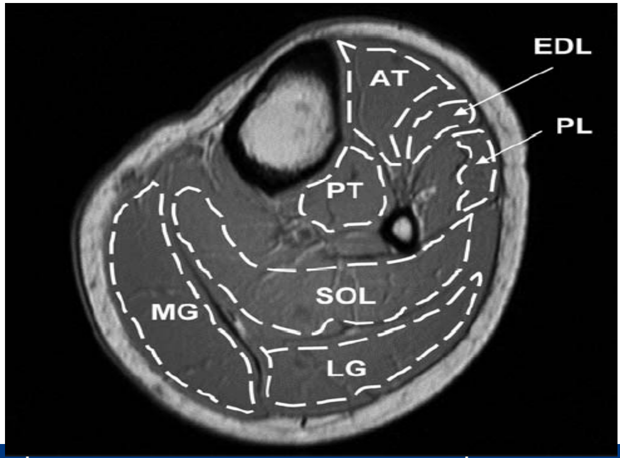
Segmentation approaches

- on T1 / T2 slides: manually (time consuming), serving as ground truth
- in tensor space: automatically, → next slides
- in fiber space: automatically grouping similar fibers together → fiber clustering









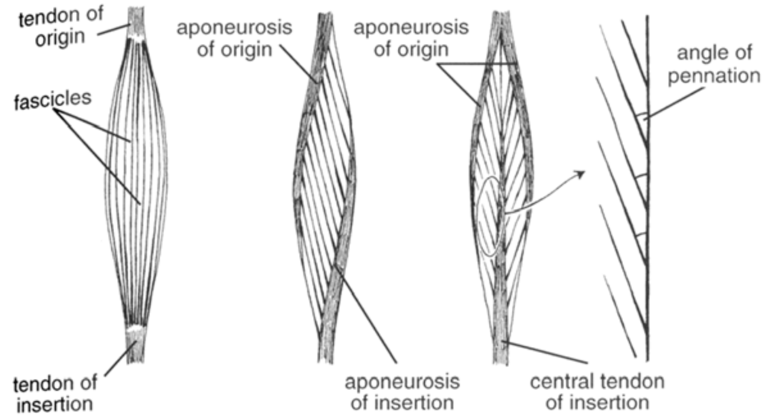
Analysis in Fiber Space – Muscle Groups

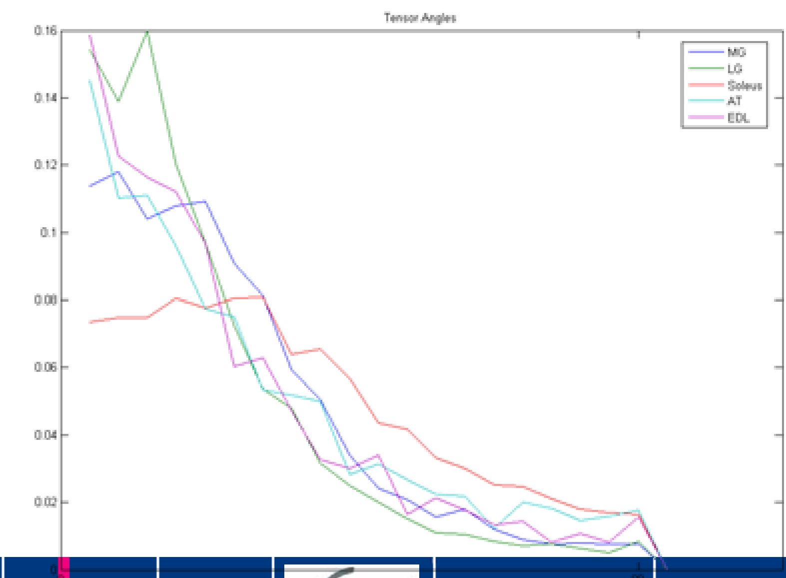
Pennation angle in different muscles

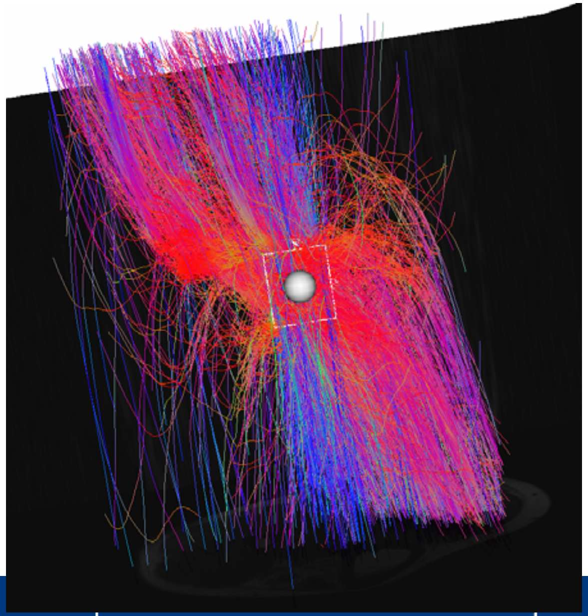
- Nonpennate, Unipennate, Bipennate

Calculation not possible because tendon plates were not identified

Calculation of tensor angle show a equal distribution for muscle groups







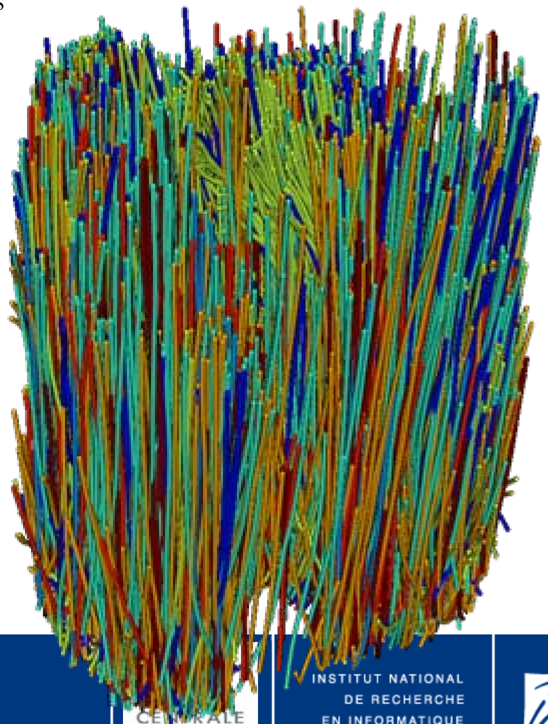
Cluster Analysis in Fiber Space – M.G.

Automatic segmentation into muscle groups

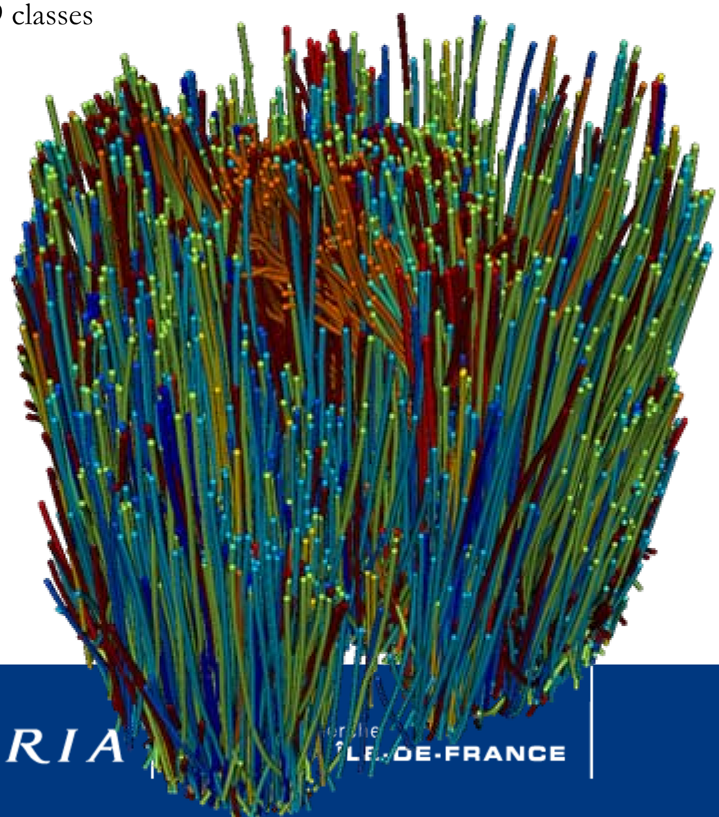
Difficult because of low quality of fiber field and complexity of task

- Linear regression with alignment in curve space
- Alignment in time and curve space failed

7 classes



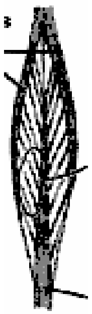
9 classes

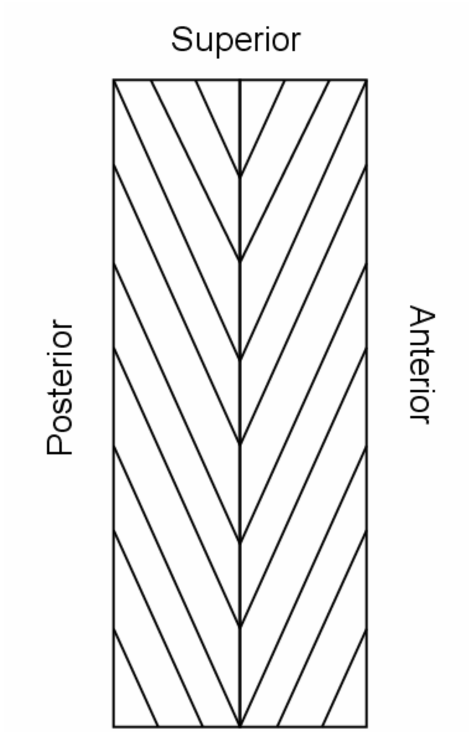


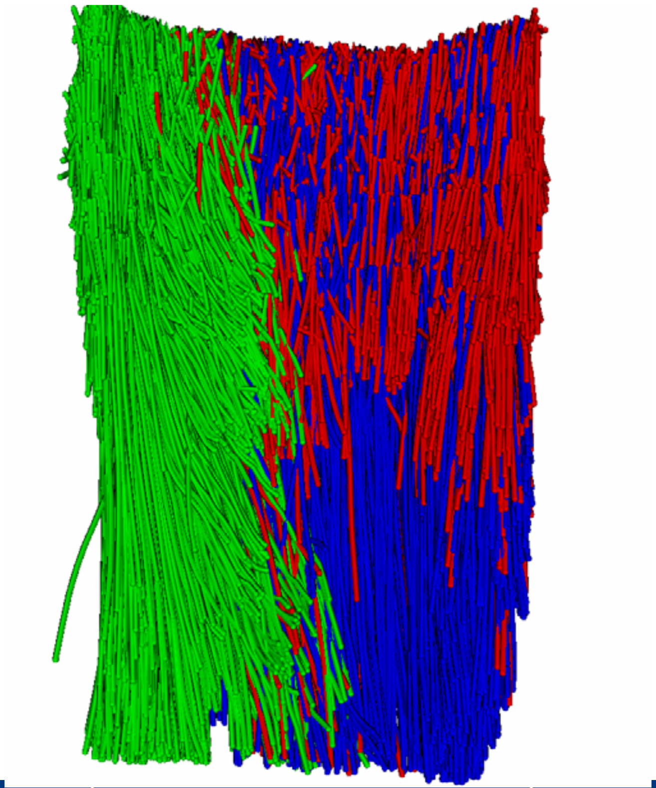
Cluster Analysis in Fiber Space – M.G.

Finding tendon plate in bipennate muscle (Soleus)

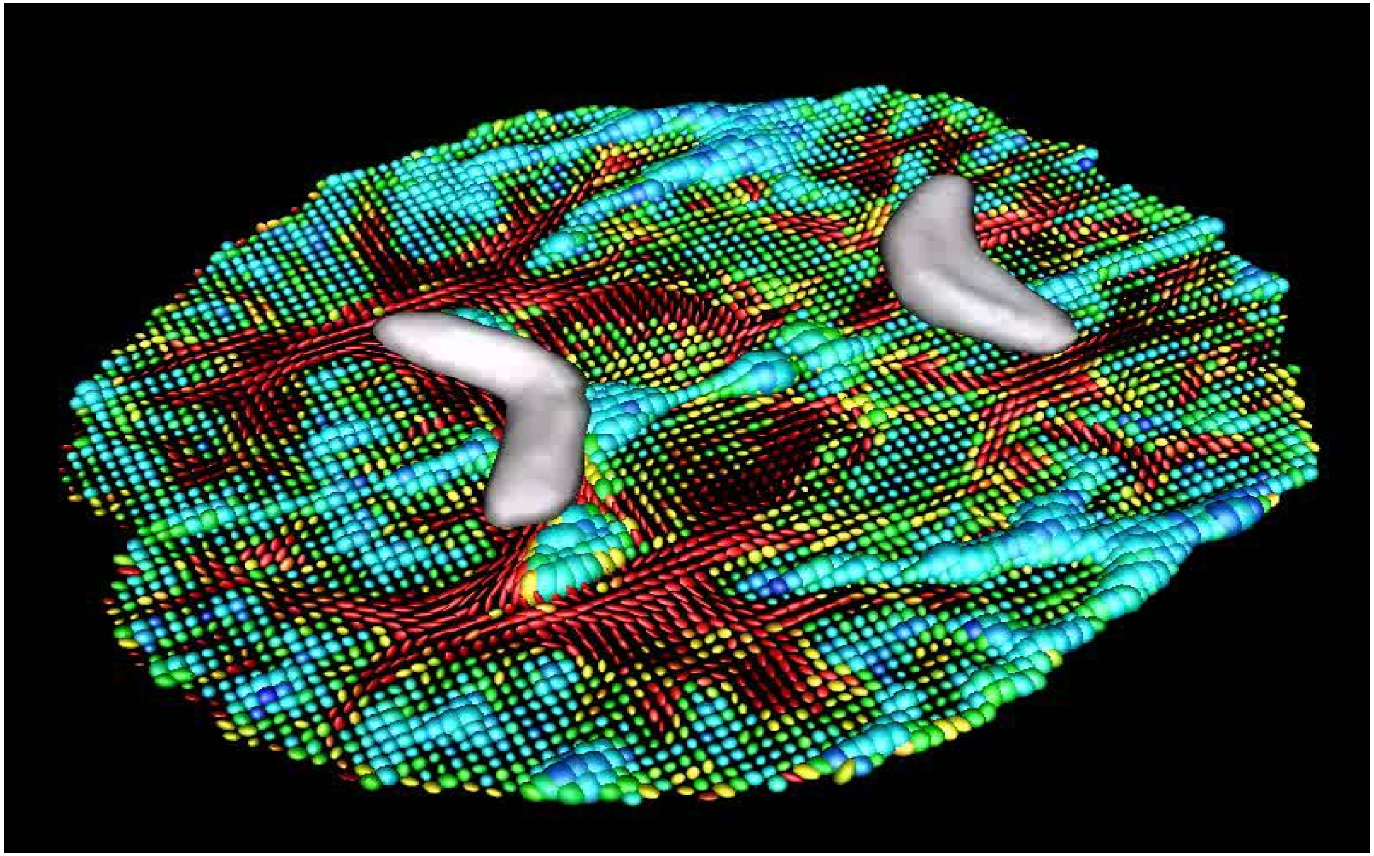
Using JCA with a linear regression model and 3 classes







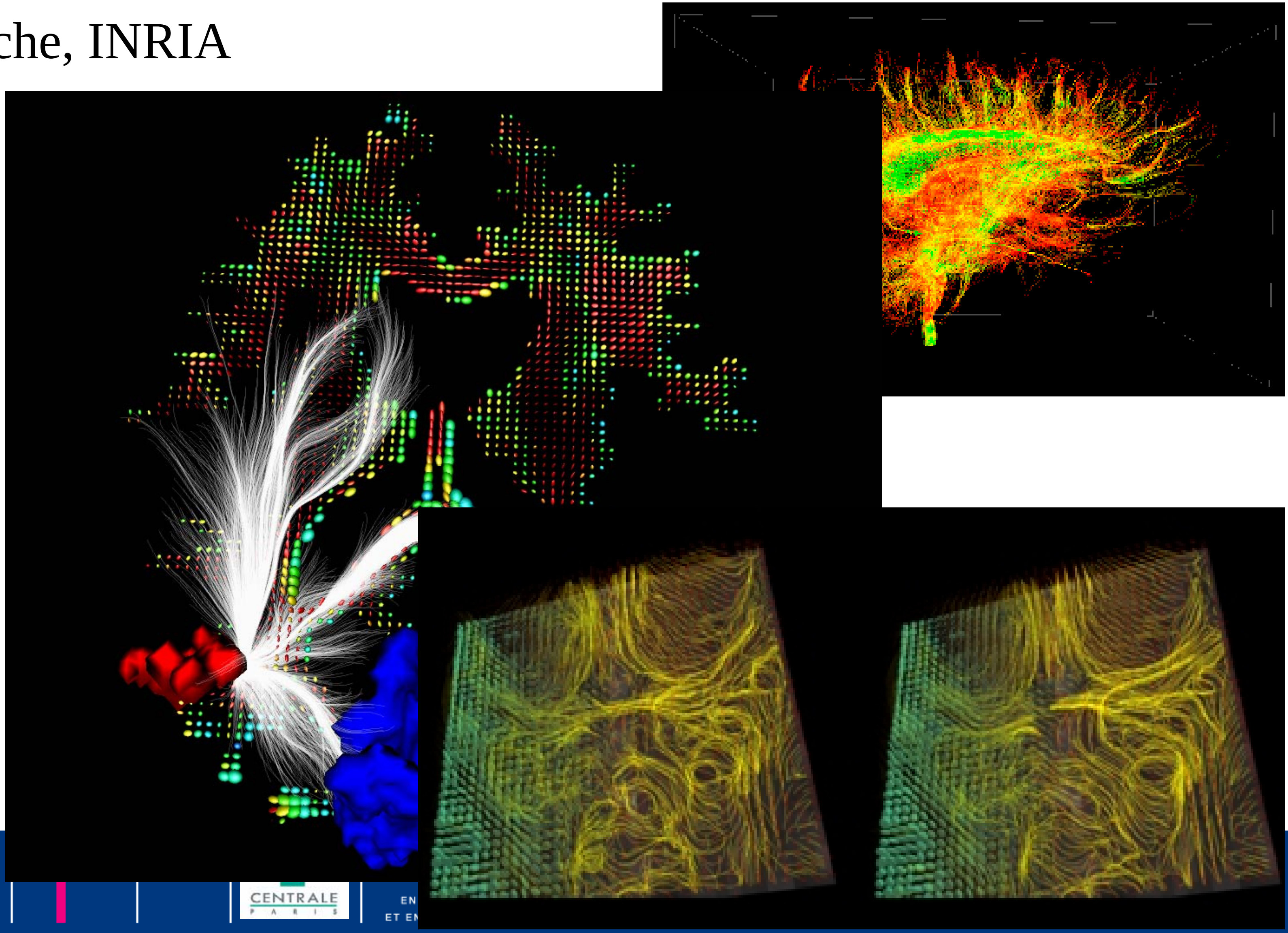
Brain Registration Segmentation



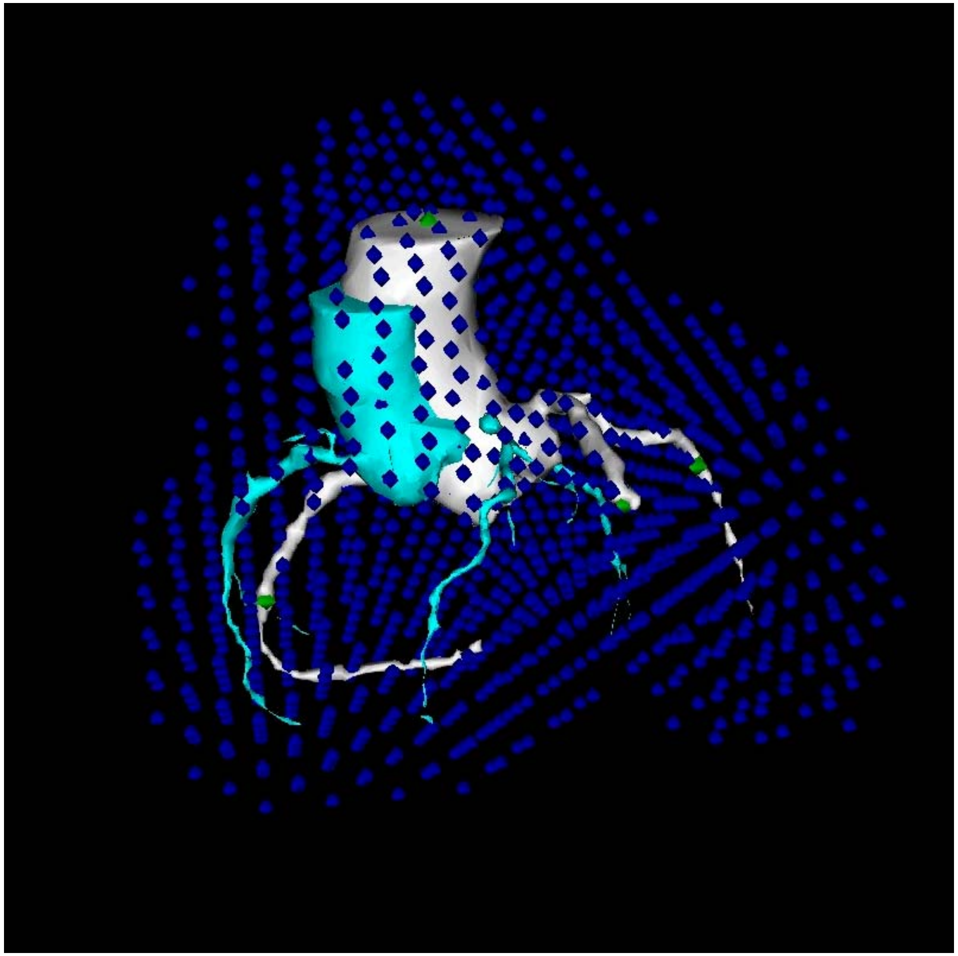
@R. Deriche, INRIA

Diffusion-Tensor Imaging

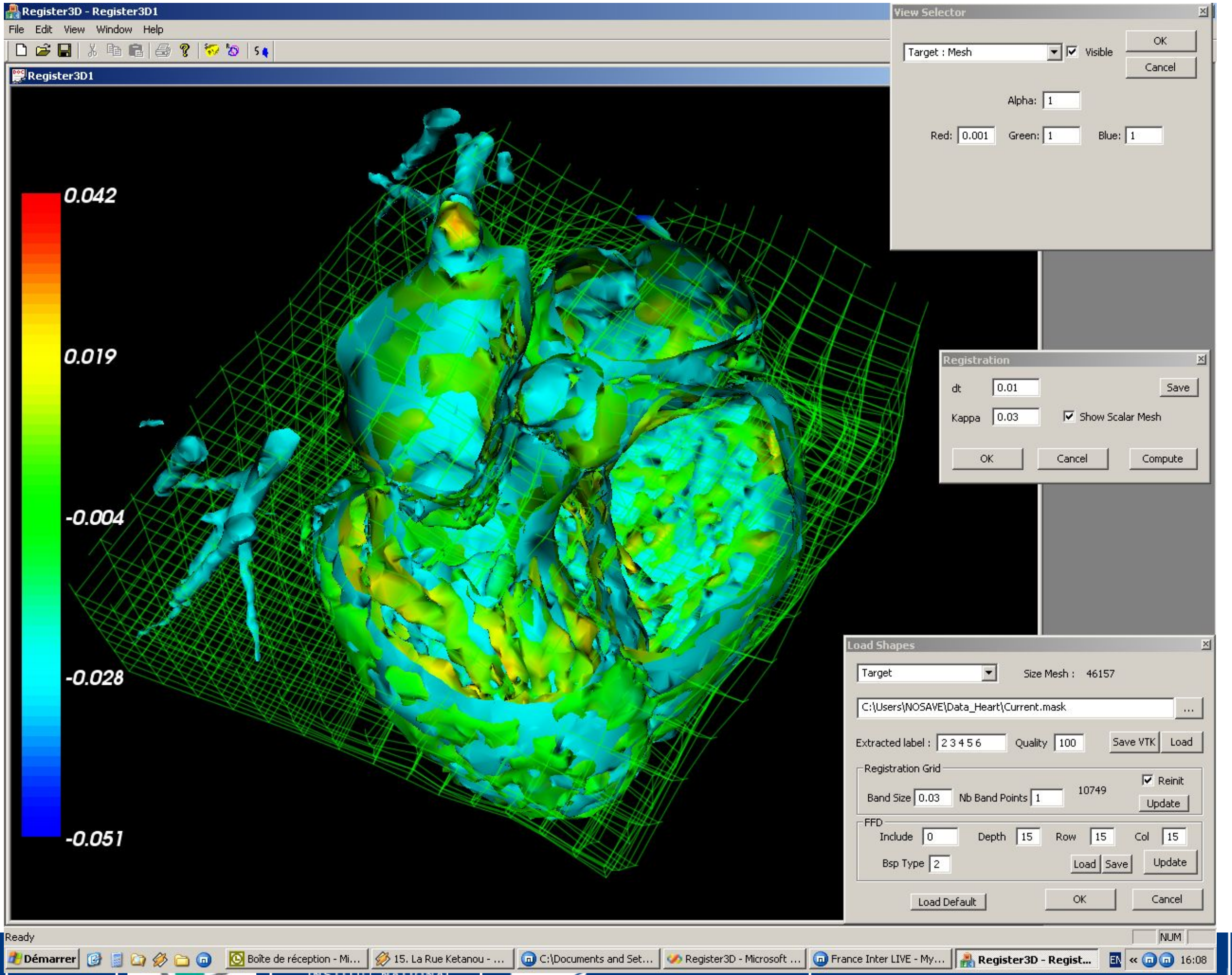
@R. Deriche, INRIA



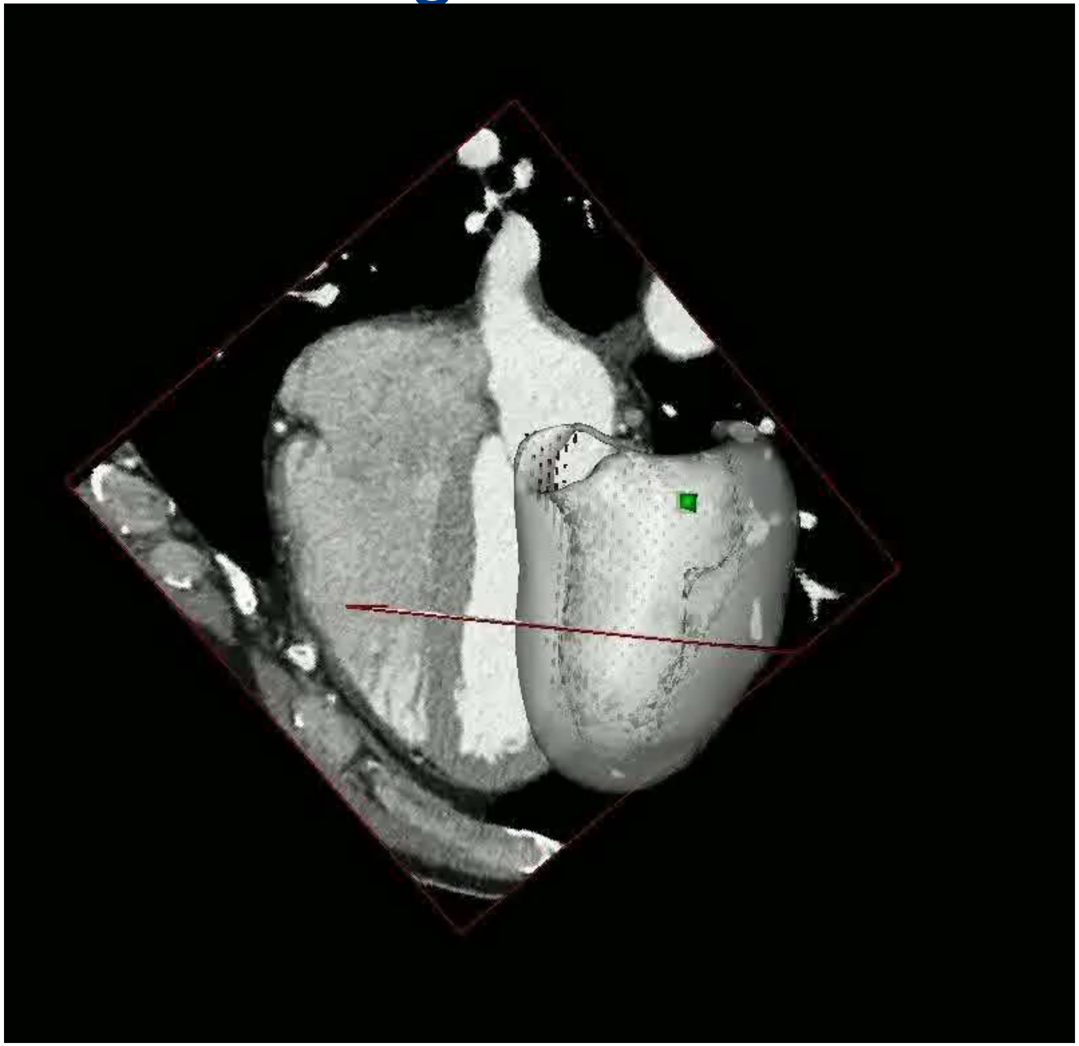
Shape Registration with Uncertainties on Implicit Spaces



Live Demonstration (part 2)

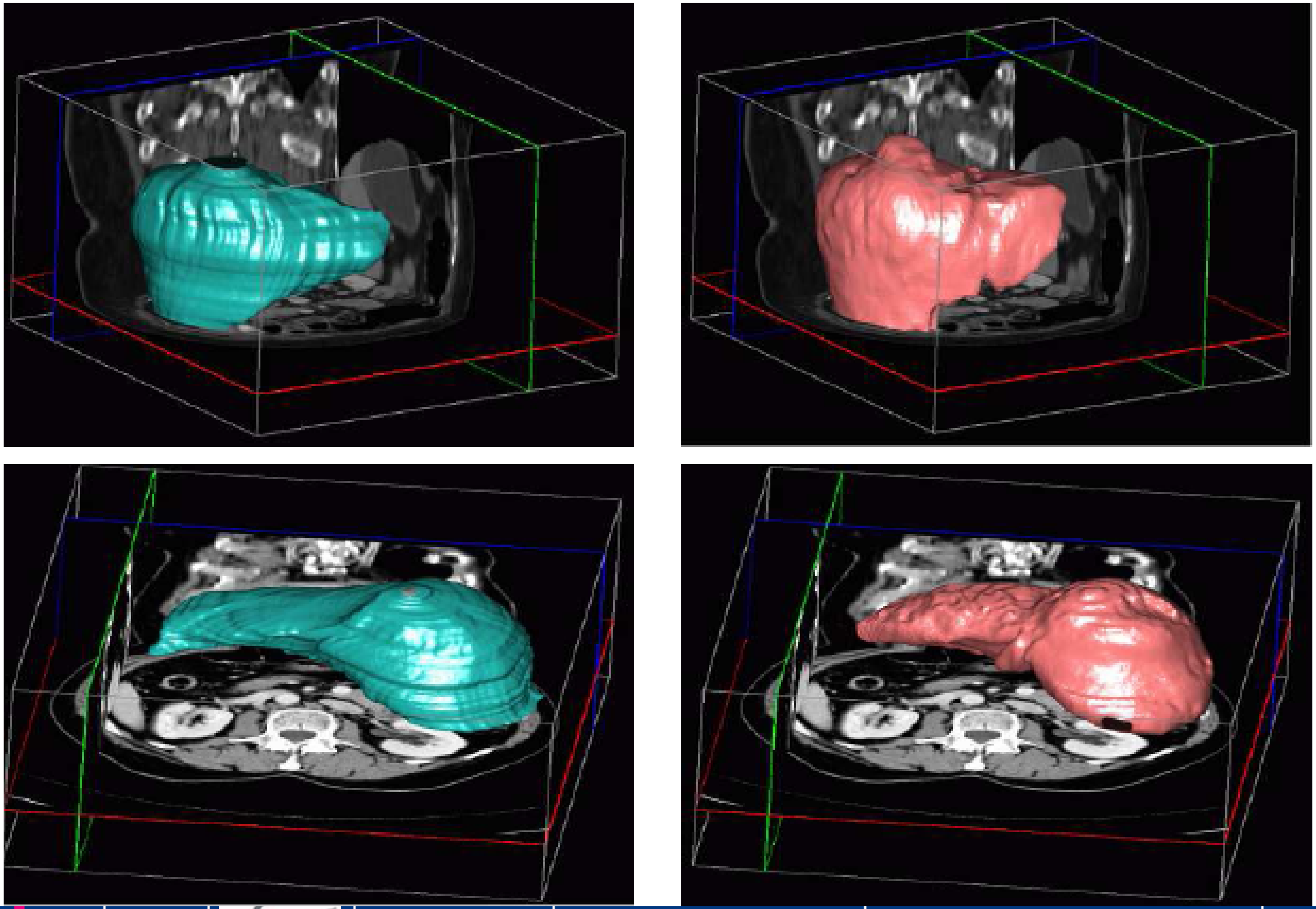


Application to 3D Segmentation - ICA

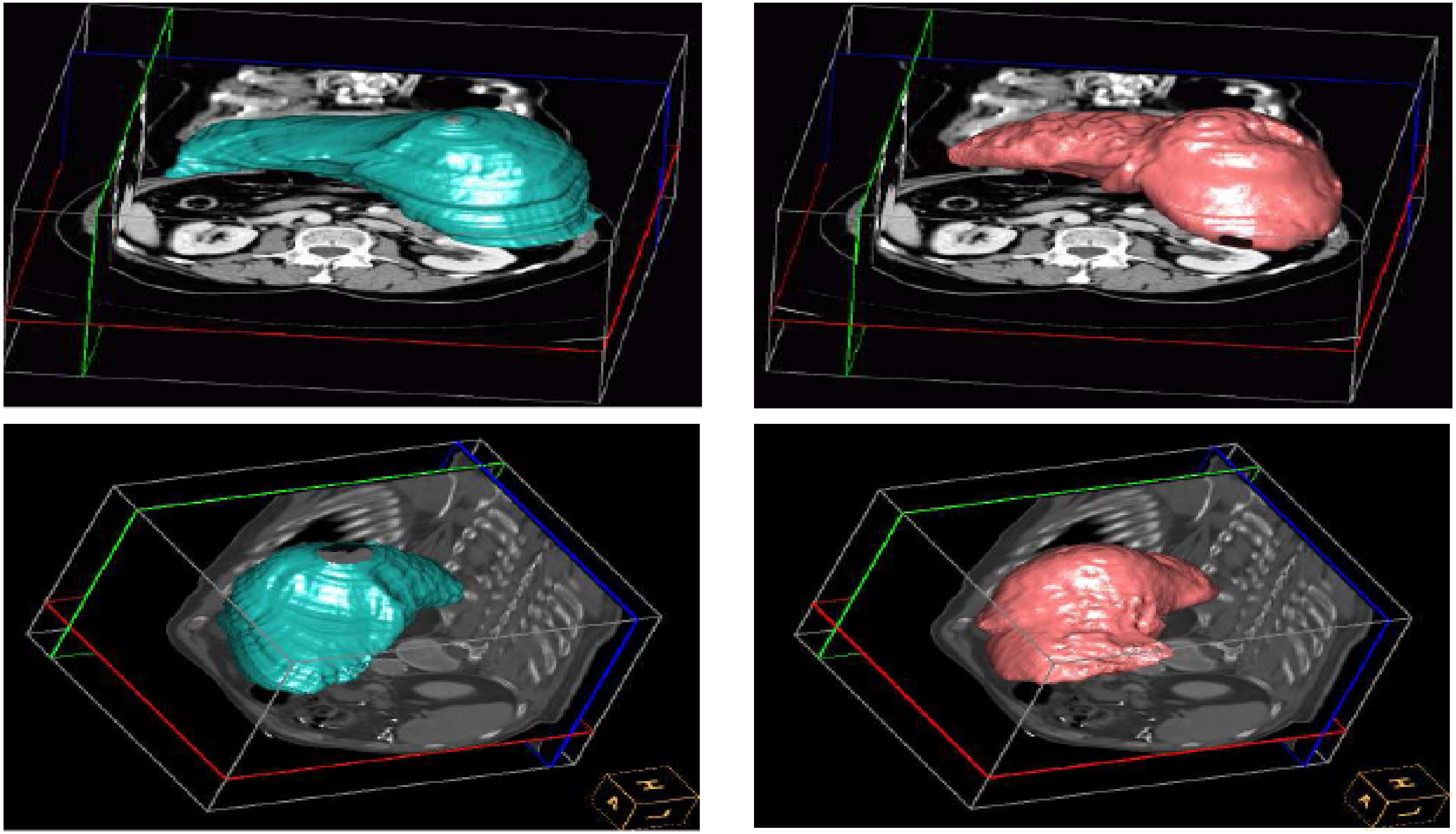


Liver Segmentation Using Sparse 3D Models with Optimal Data Support

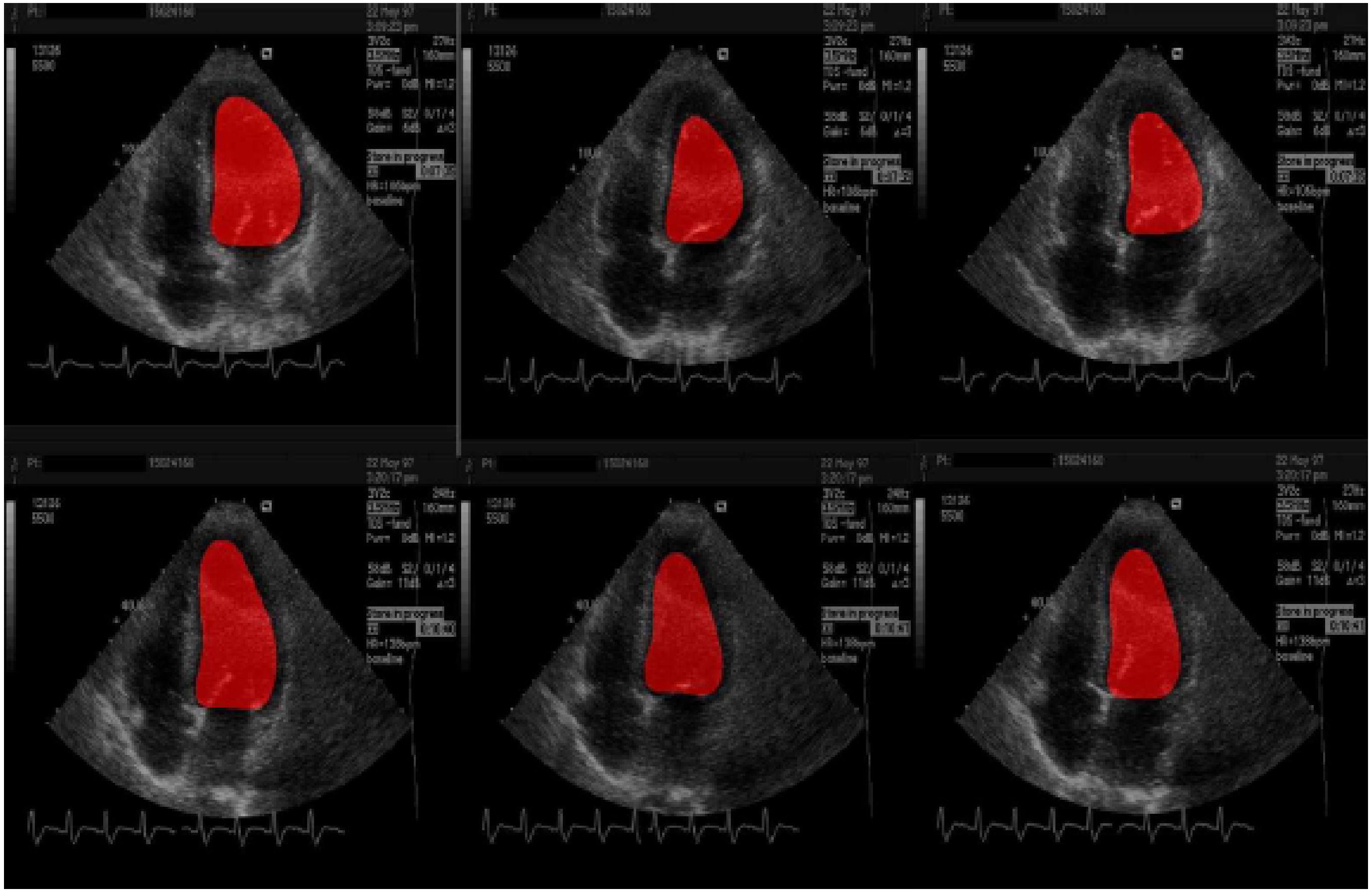
Some Visual Results



Some Visual Results



Other applications of this model



Future Work

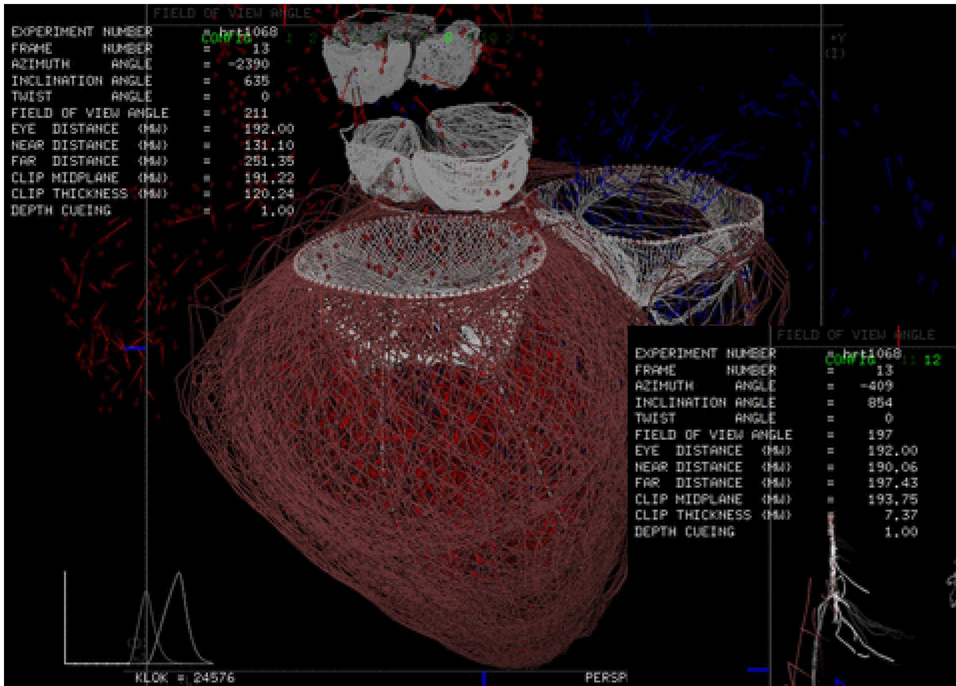
Introducing statistical behavior to the model,

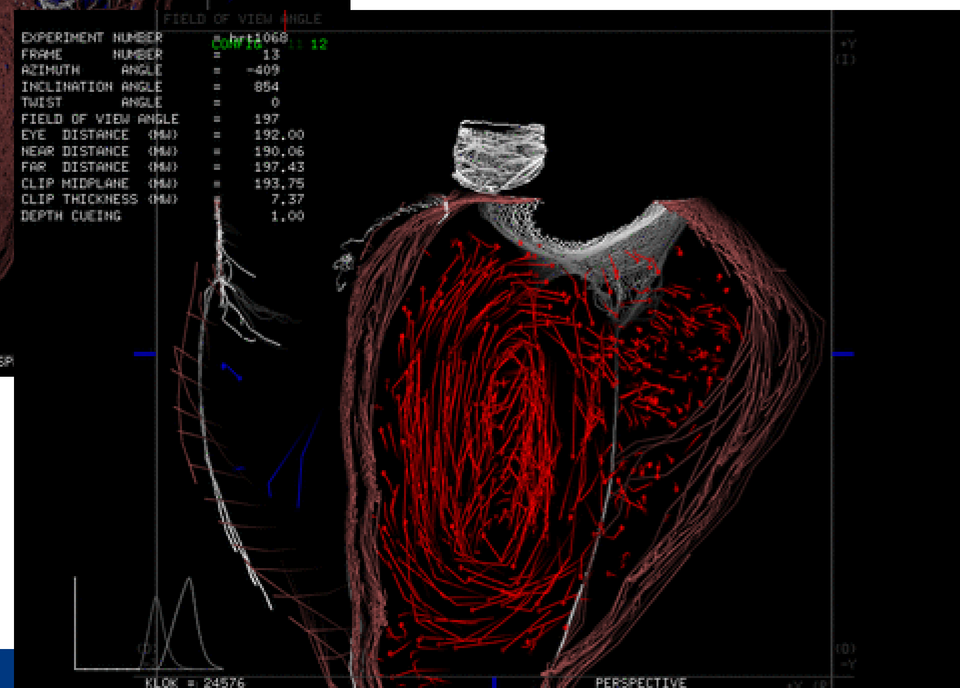
- either through modeling the distribution of the interpolation matrix,
- or through the distribution of the key contours among training examples such that using the same interpolation matrix can fully describe the training set,

Joint optimization on the model complexity (rank of the subspace) and the corresponding parameters through discrete optimization,

Investigation of generalized/non-linear interpolation methods to increase the capacity of the model toward capturing local variation,

Heart Modeling





@C. Perskin, D. Mqueen, NYU

Future

More and more data to be treated…

More and more computational power

Better and better image quality

Need for better and better mathematical models

Future : Introducing anatomical properties to the existing mathematical models, or developing better anatomical models;

Helping the physicians, improve performance of diagnosis and being able to provide solutions to new emerging problems