Variational methods as a computational model for cortical visual maps

Pierre Kornprobst Thierry Viéville 2006-01-10

Odyssée

INRIA

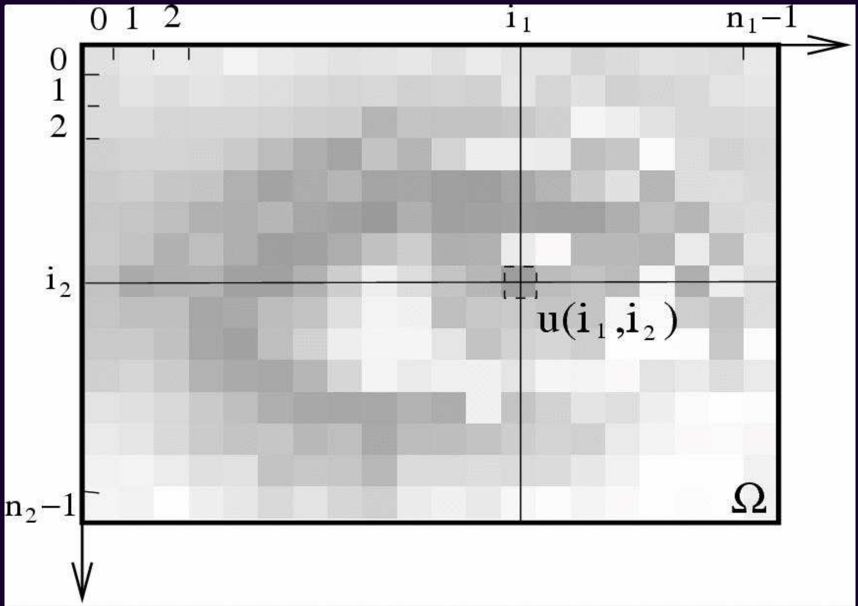
The 45mn talk step by step

- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

Introductory example: Isotropic Diffusion and Gaussian Filtering

- Retinotopic map: “images”
- Linear Gaussian Filtering : Image Smoothing
- The Heat Equation : Isotropic Diffusion
- A Variational Formulation : Image Regularization
- From this example to a general setting

Retinotopic map: “images”



- A digital image may be defined as a 2×2 array or as a discrete function, a “map”

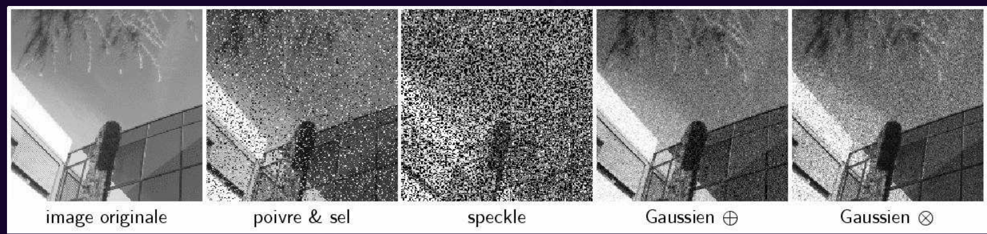
$$u : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^M$$

- From the analog and continuous world, it is obtained after both pixelization.

$$u(i_1, i_2) = \int_{\text{pixel}} u(x, y)$$

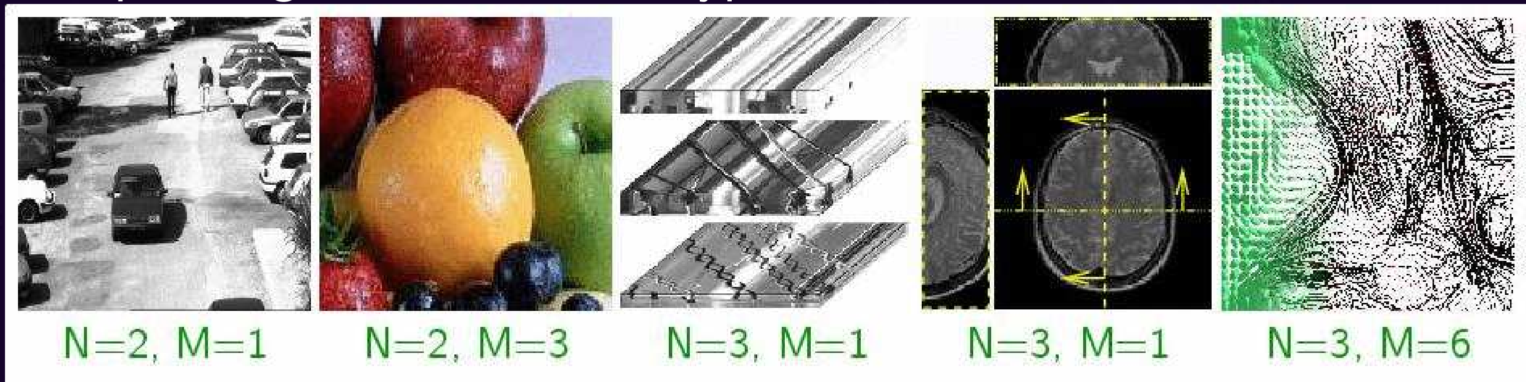
and quantification,

and with noise

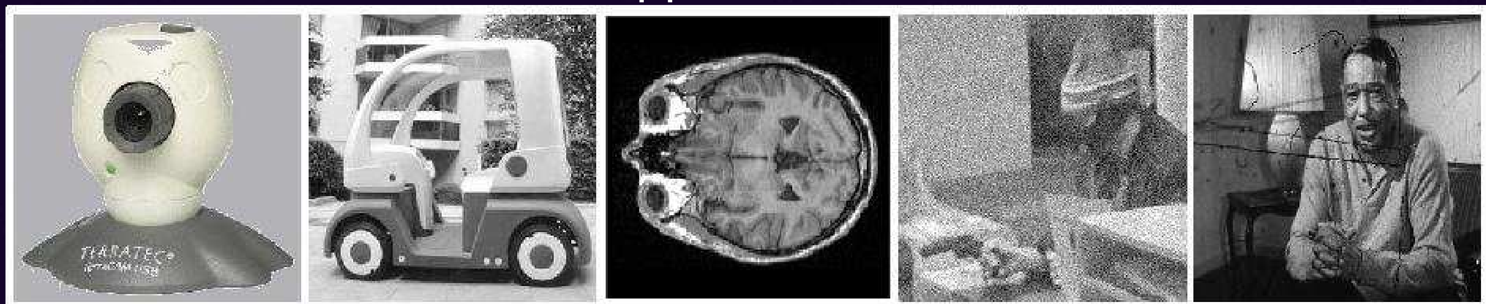


More general images (image sequences or bundle) . .

. . corresponding to various data type:



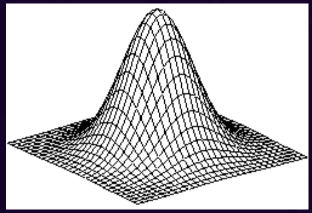
. . and in relation with various applications:

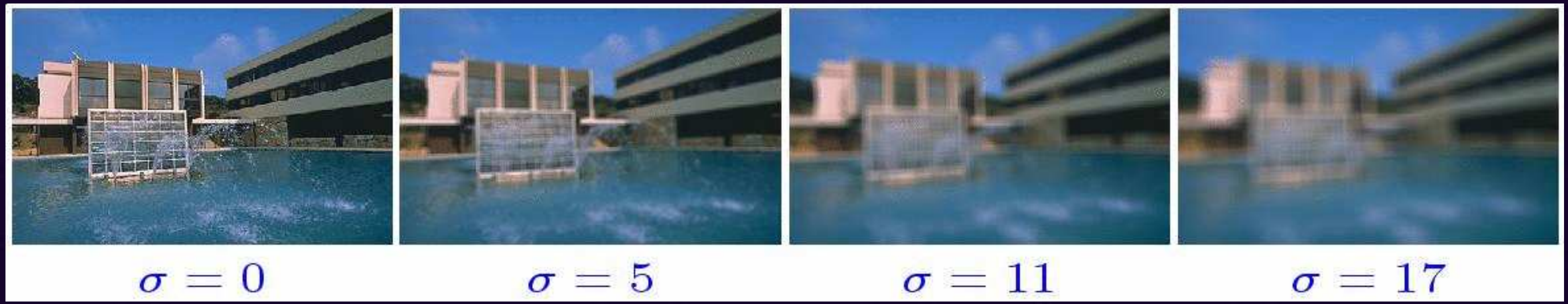


Linear Gaussian Filtering : Image Smoothing

- Let u_0 an image, the Gaussian Smoothing writes:

$$u_\sigma(x) = (G_\sigma * u_0)(x) \text{ with } G_\sigma(x) = \frac{1}{2\pi\,\sigma^2}\ \exp\left(-\frac{|x|^2}{2\,\sigma^2}\right).$$





- This is a standard front-end for multi-scale representation of an image.

The Heat Equation : Information Diffusion

- Let u_0 an image, the Isotropic Diffusion writes (Partial Differential Equation):

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \Delta u(t, x), & t \geq 0, \\ u(0, x) = u_0(x). \end{cases}$$

- The Laplacian Δ is an isotropic, elementary diffusion operator:

$$\Delta u = \sum_{i=1}^{2} \frac{\partial^2 u}{\partial x_i^2} \simeq \sum_{z \in V(x)} [u(z) - u(x)]$$

1	2	1
2	-12	2
1	2	1

(local balanced average)

- Main result: $u(t, x) = (G_{\sqrt{2t}} * u_0)(x)$

- Diffusion is an infinitesimal smoothing !

A Variational Formulation : Image Regularization

- Let u_0 an image, the Regularized Image writes:

$$\inf_u \; E(u), \qquad E(u) = \int_\Omega (1-\lambda)\,|u(x) - u_0(x)|^2 + \lambda\,|\nabla u(x)|^2 \, dx$$

- Main result (Euler-Lagrange equation):

$$\boxed{\frac{\partial u}{\partial t}(t,x) \equiv -\frac{1}{2}\nabla E = (1-\lambda)\,[u_0(x) - u(x)] + \lambda\,\Delta u(x) \text{ minimizes } E}$$

- When $\lambda \rightarrow 1$ the heat equation minimizes E.

- This gives : convergence + function specification !

what's to be done $\rightarrow$ *how to do it*

From this example to a general setting

- All main visual functions may be specified from a variational approach
- The partial differential equation is even more general
- Very robust and efficient implementations are derived
- Generalization to non-linear space (Beltrami flow)
- The link with biological neural networks has been built
- . . and it is not that complicated.

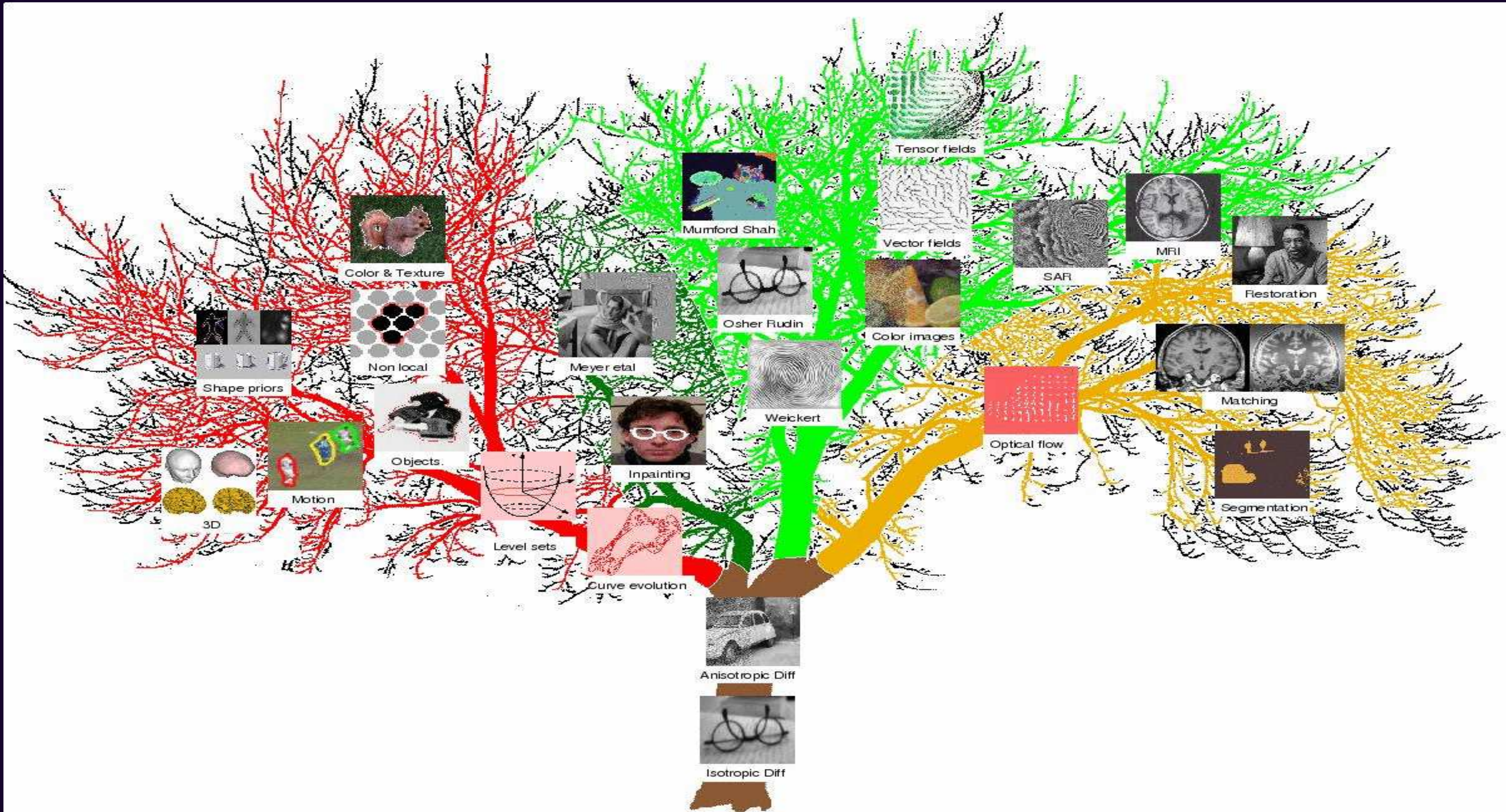
The 45mn talk step by step

- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

Specification of visual functions

- Image restoration (smoothing, etc..) including using a biological model.
- Image segmentation (object detection, ..)
- Image matching / registration (stereo, motion, ..)
- Others:
  - Focus of attention (winner take [almost] all)
  - Image completion (in-painting, ..)

. . and more !


Tensor fields
Mumford Shah
Vector fields
SAR
MRI
Restoration
Color & Texture
Osher Rudin
Color images
Non local
Meyer etal
Shape priors
Weickert
Matching
Optical flow
Objects.
Inpainting
Motion
Segmentation
3D
Level sets
Curve evolution
Anisotropic Diff
Isotropic Diff

Specification of visual functions: image restoration

- Basic model: find u observing u_0,

$$\underbrace{u_0}_{\textit{measured image}} = \underbrace{R}_{\text{image formation linear operator}} \underbrace{u}_{\textit{original image}} + \underbrace{\nu}_{\text{some additive Gaussian noise}}$$

- Basic specification: minimize,

$$\inf_u \; E(u) = \int_\Omega (u_0 - Ru)^2 \, dx + \lambda \int_\Omega \phi(|\nabla u|) \, dx$$

  - Data attach: least-square solution (statistically optimal . . but ill-posed)
  - Regularization: restrain the set of solutions
  - Meta-parameter : high-level control of the solution

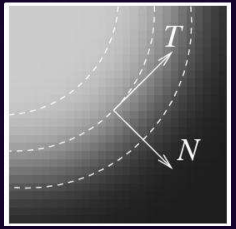
Specification of visual functions: image restoration

- Automatic derivation of the Euler-Lagrange equation:

$$(R^* R u - R^* u_0) - \frac{\lambda}{2} \operatorname{div}\Big(\underbrace{\boxed{\frac{\phi'(|\nabla u|)}{|\nabla u|}}}_{c(|\nabla u|)} \nabla u \Big) = 0$$

- with a geometrical interpretation of the non-linear diffusion:

$$\operatorname{div}\left(\frac{\phi'(|\nabla u|)}{|\nabla u|}\nabla u\right) = \underbrace{\frac{\phi'(|\nabla u|)}{|\nabla u|} u_{TT}}_{tangential} + \underbrace{\phi''(|\nabla u|)\, u_{NN}}_{\substack{normal \\ not\ across\ edges}}$$



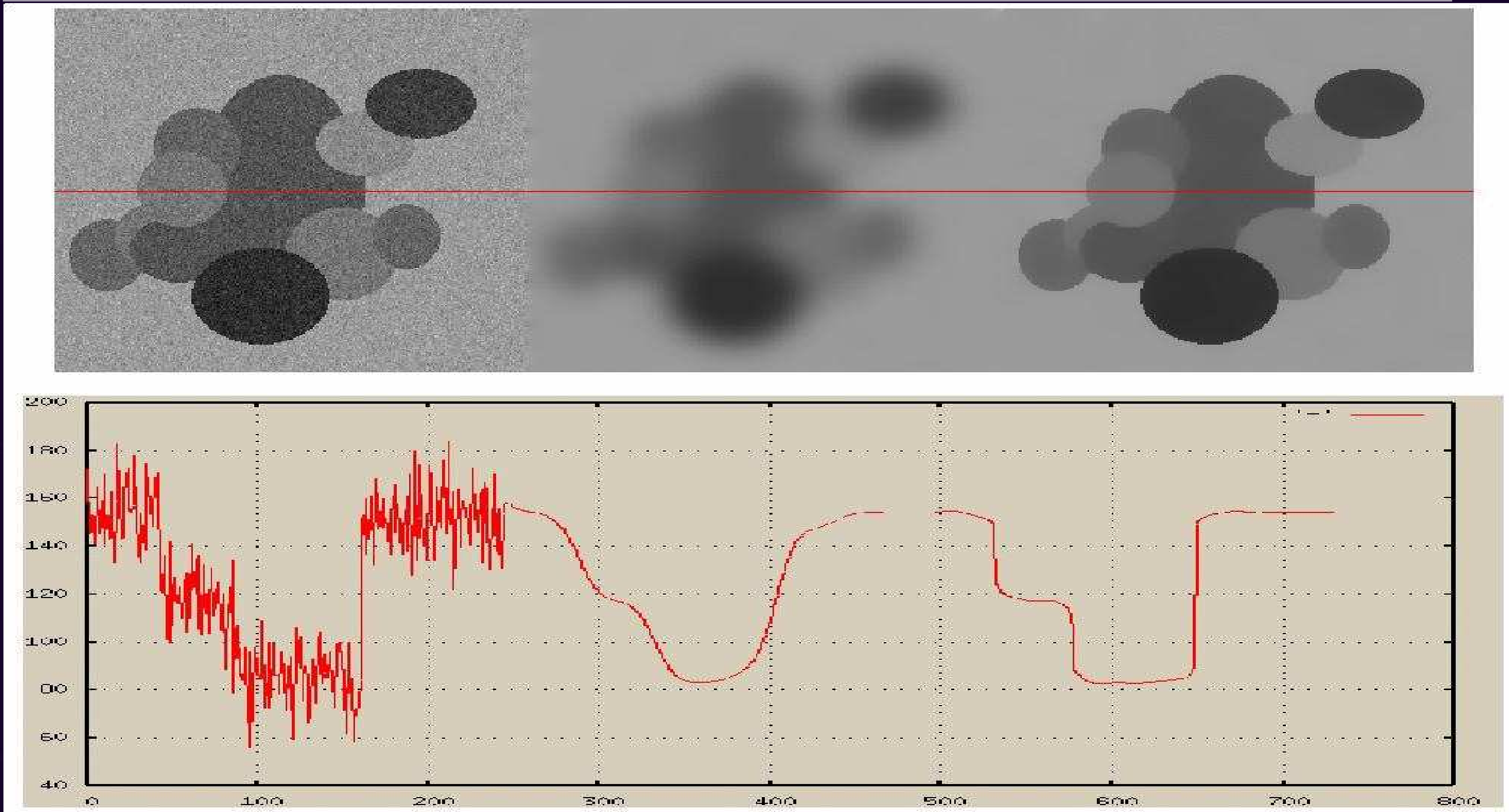
Specification of visual functions: image restoration

- A large choice of non-linear profile:

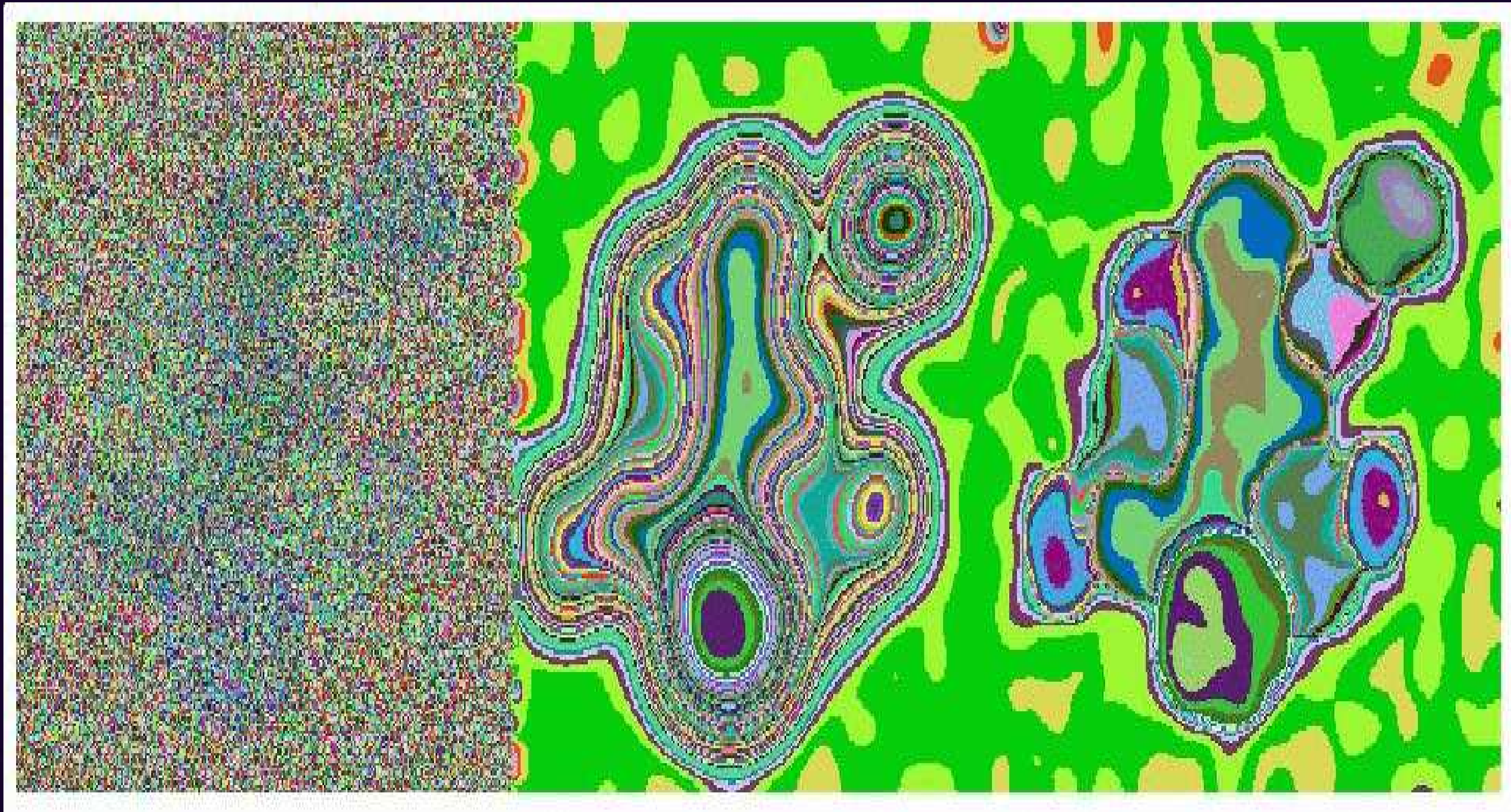
Author	$\phi(x)$		$\frac{\phi'(x)}{x}$
Malik & Perona	$\log(1+x^2)$		$\frac{2}{(1+x^2)}$
Tikhonov & Arsenin	x^2	convex	2
Geman & Reynolds	$\frac{x^2}{1+x^2}$		$\frac{2}{(1+x^2)^2}$
Green	$2\log[\cosh(x)]$	convex	$\begin{cases} 2 & x=0 \\ 2\tanh(x)/x & x\neq 0 \end{cases}$
Aubert & Vese	$2\sqrt{1+x^2}-2$	convex	$\frac{2}{\sqrt{(1+x^2)}}$

- Here ϕ allows to control the regularity of the solution
- In fact ϕ allows to defined the underlying functional space of the solution

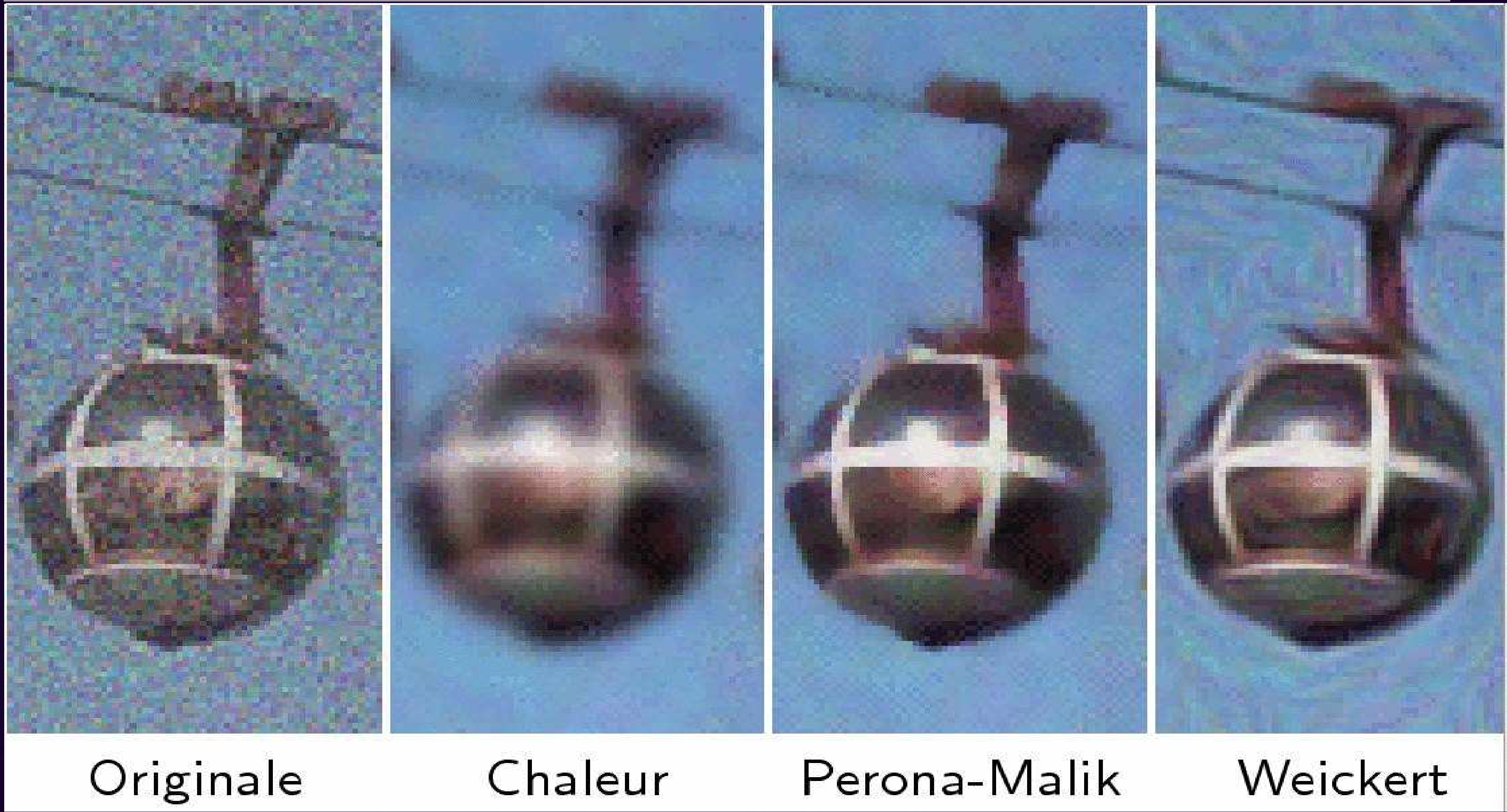
Specification of visual functions: Perona-Malik restoration



Specification of visual functions: Along isophotes diffusion



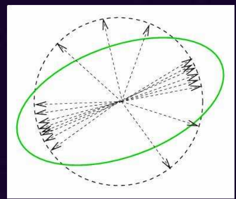
Specification of visual functions: A few examples



Specification of visual functions: a non variational approach

- Defining the structure tensor from the image gradient ∇u:

$$k_\rho * \nabla u_\sigma \nabla u_\sigma^t = k_\rho * \begin{pmatrix} u_{\sigma xx} & u_{\sigma xy} \\ u_{\sigma xy} & u_{\sigma yy} \end{pmatrix}$$



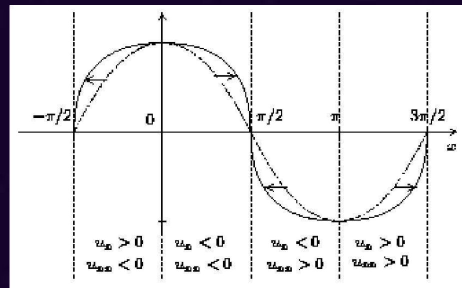
- Allows to propose the Weickert diffusion scheme:

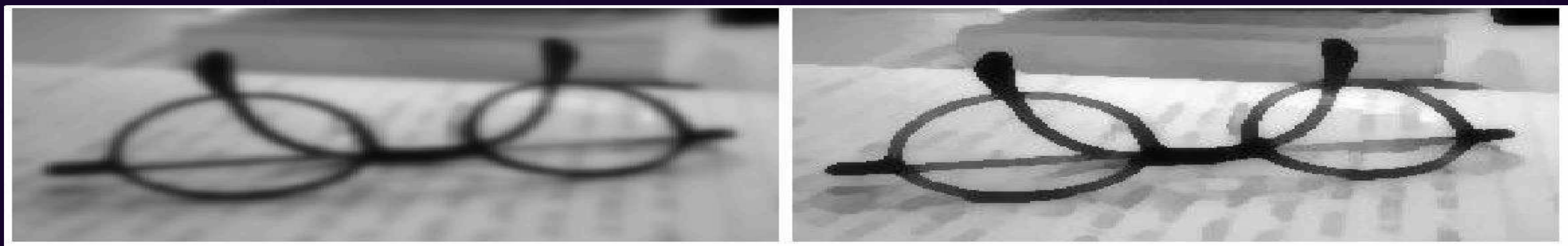
$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\underbrace{D(k_\rho * \nabla u_\sigma \nabla u_\sigma^t)}_{\text{matrix}} \nabla u \right)$$

Specification of visual functions: another non variational approach

- The Osher and Rudin shock-filter approach:

$$\begin{cases} u_t(t,x) = -\,|u_x(t,x)|\ \mathrm{sign}\,(u_{xx}(t,x)), \\ u(0,x) = u_0(x), \end{cases}$$





can not be derived from a variational approach (convergence not guarantied !)

Specification of visual functions: the Cottet-Ayyadi model

Cottet and Ayyadi consider the Hebbian adaptive diffusion processes:

$$\boxed{min \int ||\nabla u||^2_{L()}} \Rightarrow \dot{u} = -l(u)\, \Delta_{\mathbf{L}} u \text{ with } u(0) = u_0$$

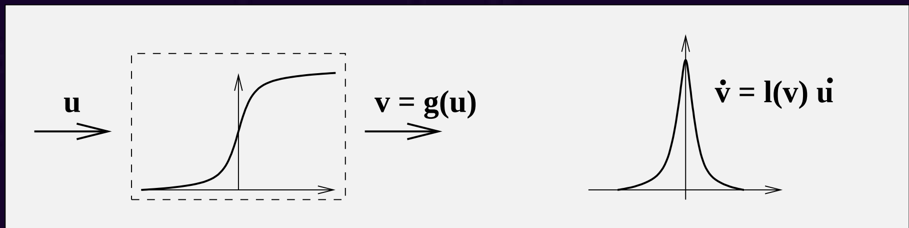
with contrast threshold s, adaptation time constant τ, spatial smoothing S:

$$\frac{\partial \mathbf{L}}{\partial t} + \frac{1}{\tau}\mathbf{L} = \frac{1}{\tau}\left[\rho^2\, \mathbf{P}_{\mathbf{g}^\perp} + \frac{3}{2}(1-\rho^2)\,\mathbf{I}\right] \text{ with } \begin{array}{l} \rho = \min\left(1, \frac{||\mathbf{g}||^2}{s^2}\right),\ \mathbf{g} = S * \nabla u \\ \mathbf{P}_{(g_1,g_2)^\perp} = \begin{pmatrix} g_2 g_2 & -g_1 g_2 \\ -g_1 g_2 & g_1 g_1 \end{pmatrix} \end{array}$$

Specification of visual functions: the Cottet-Ayyadi model

thus with:

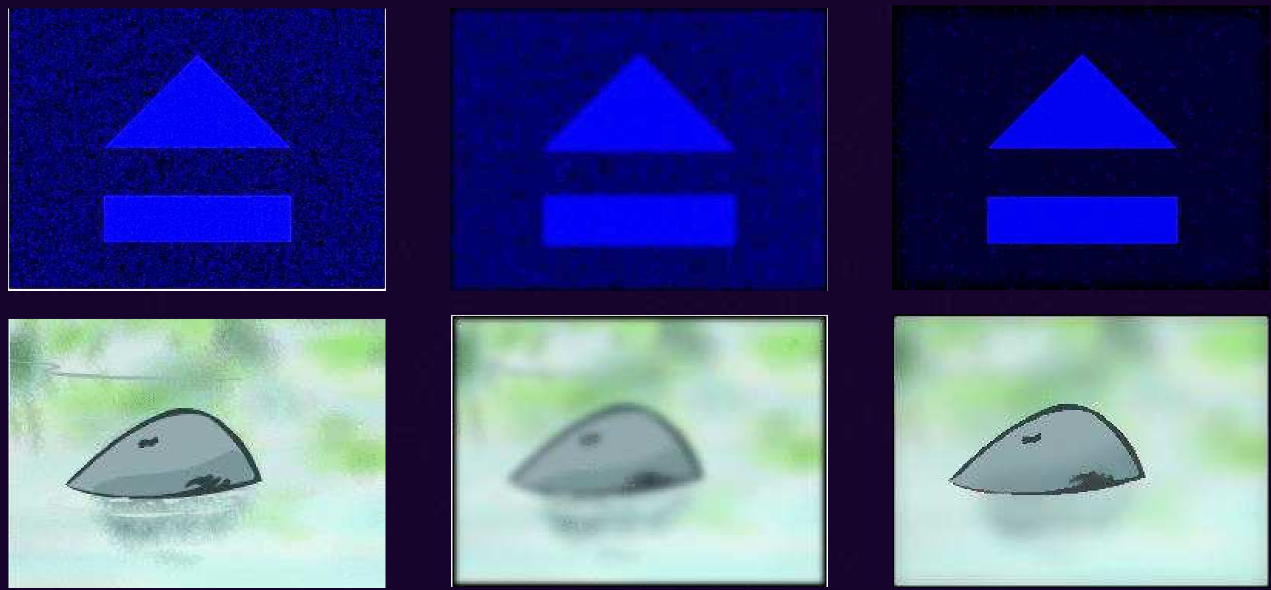
- anisotropic diffusion along edges but not across edges for high contrasted areas (i.e. $\mathbf{L} \equiv \mathbf{P}_{\mathbf{g}\perp}$ when ρ is close to 1 in the previous equation) but
- isotropic diffusion in almost uniform areas when low-contrast (i.e. $\mathbf{L} \equiv \mathbf{I}$ when ρ is close to 0 in the previous equation).



Including the non-linear relationship between:

- the neuronal state $\mathbf{u}$ (usually related to the membrane potential) and
- the neuronal output $\mathbf{v} \in [0, 1]^N$ (usually related to the average firing rate probability).

Specification of visual functions: the Cottet-Ayyadi model

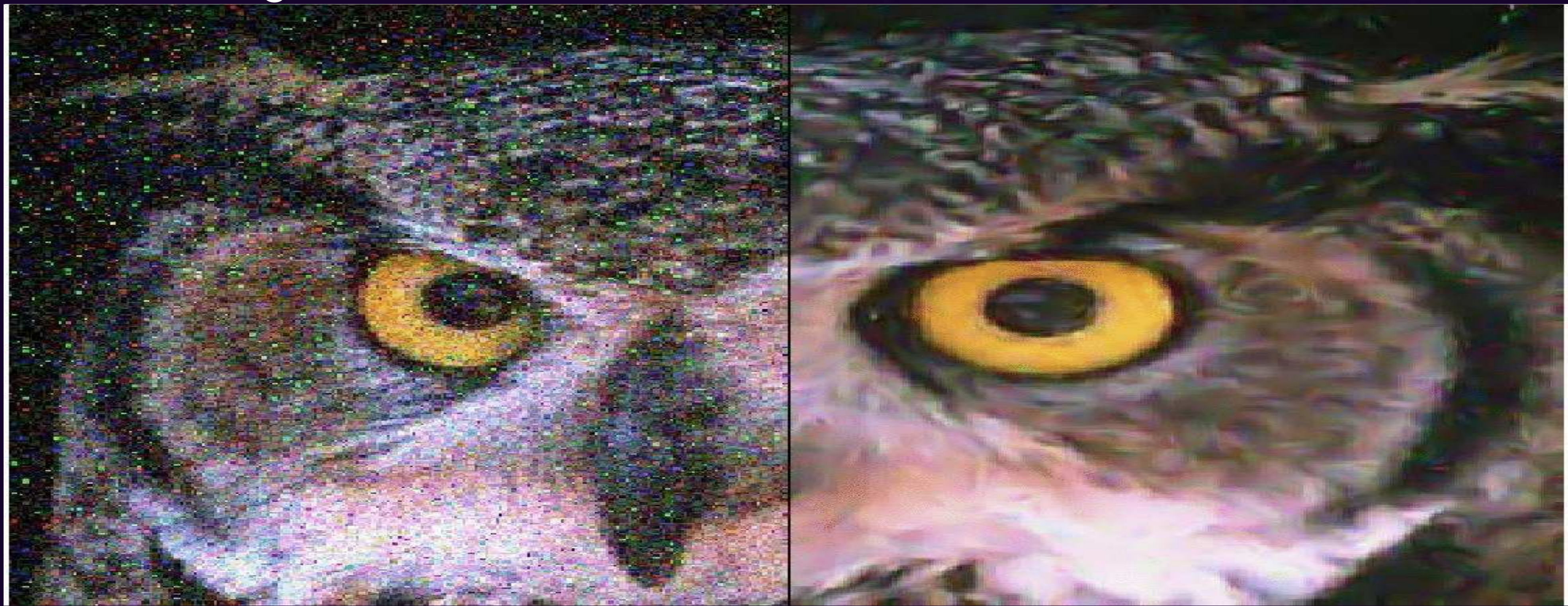


Raw *Isotropic* *Anisotropic*

The blue image contains a huge (80%) amount of noise. The complex image contains features at several scales. Edges are preserved, while an important smoothing has been introduced.

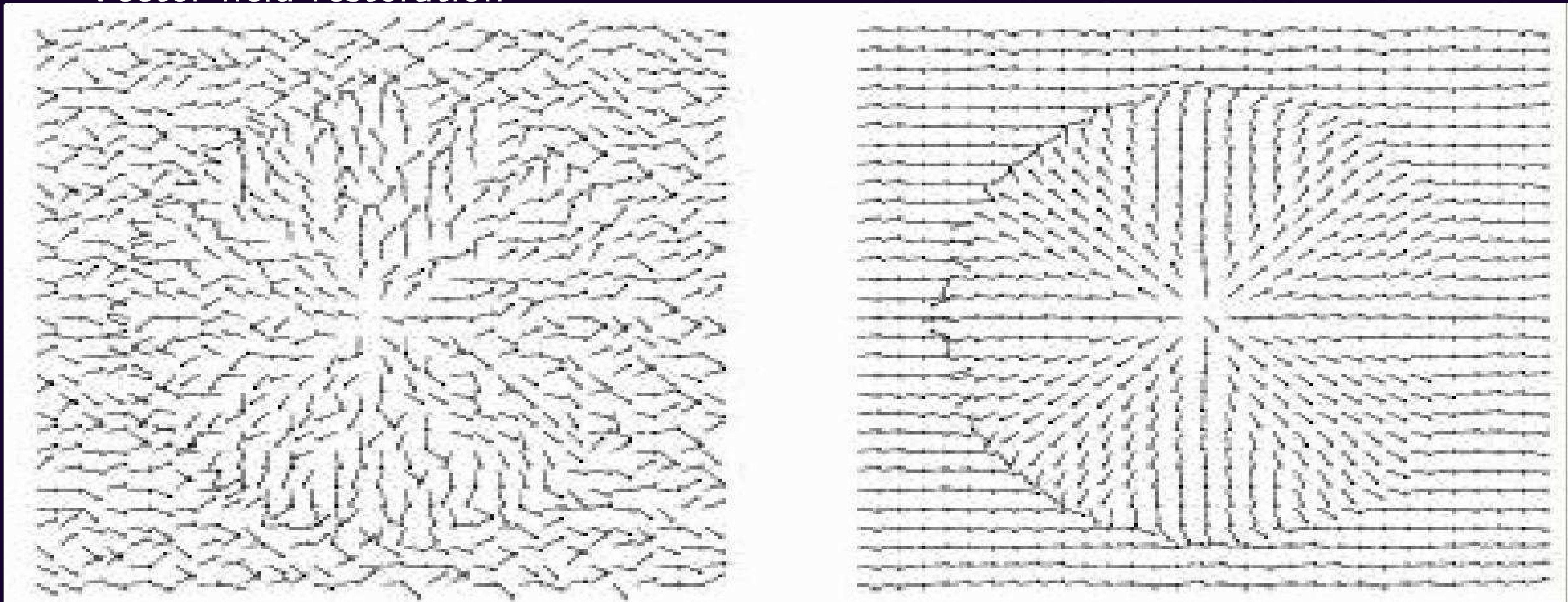
Specification of visual functions: restoration of complex images

- Color image restoration



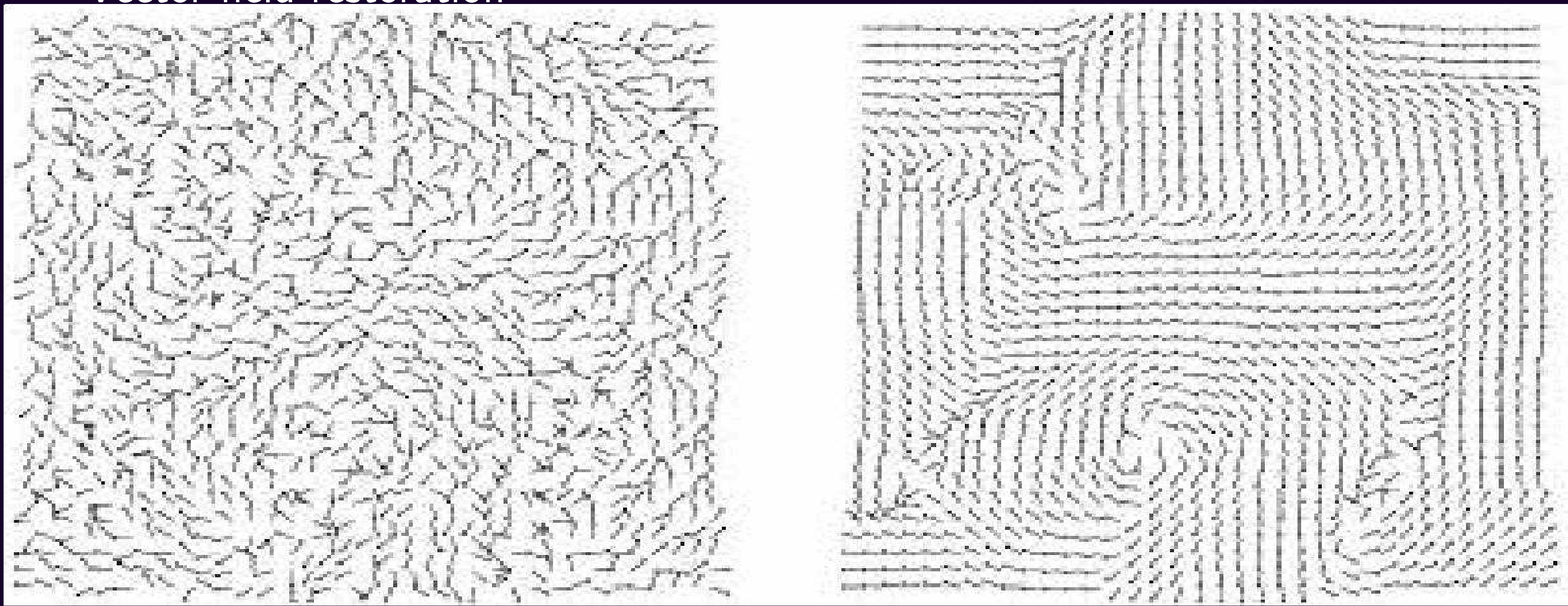
Specification of visual functions: restoration of complex images

- Vector field restoration



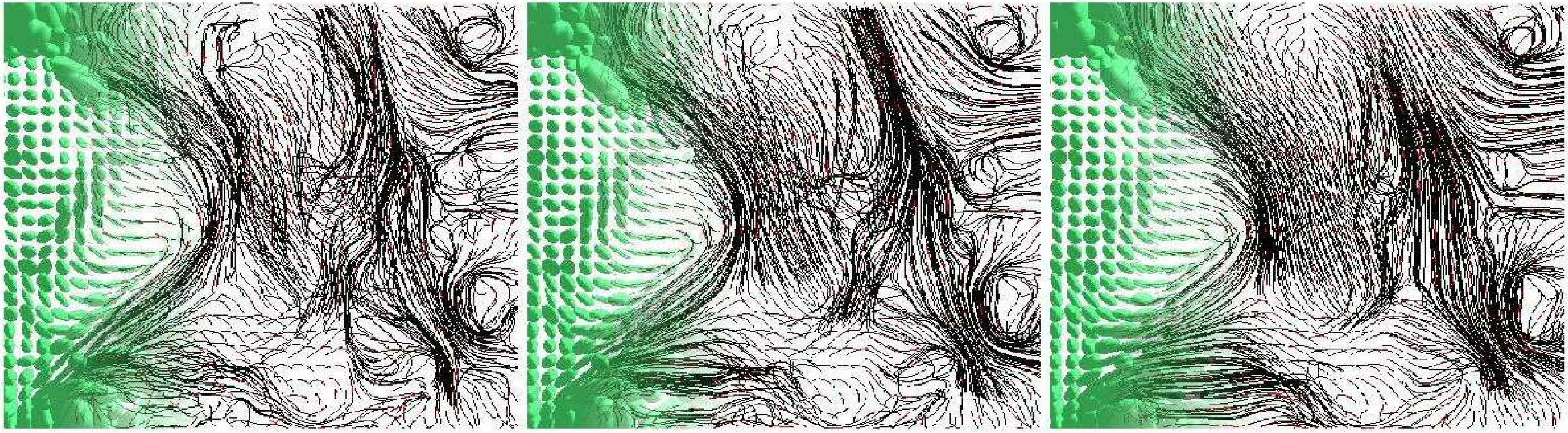
Specification of visual functions: restoration of complex images

- Vector field restoration



Specification of visual functions: restoration of a tensor field

- Saturation of a tensor field $T = R^T \, D \, R$ with
  (i) diffusion on D and (ii) regularization of R with orthonormal preservation



Original DT-MRI tensors/fibers   Regularization after 20 it.   Regularization after 50 it.

Specification of visual functions: image segmentation

Specification:

$$min_{u,K} \quad \underbrace{\int_W (u-u_0)^2}_{\text{approximation quality "data attach"}} \quad + \quad \underbrace{\alpha^2 \int_{W-K} ||\nabla u||^2}_{\text{homogeneity of each component "regularization"}} \quad + \quad \underbrace{\beta \int_K 1}_{\text{parsimony of the segmentation "edge length"}}$$

$$min_{u,z} \quad \int_W (u-u_0)^2 \quad + \quad \alpha^2 \int_W z^2 \, ||\nabla u||^2 \quad + \quad \beta \int_W \lambda_\epsilon(z)$$

where:

- $\beta > 0$ controls the fine/coarse grained segmentation and
- $\alpha > 0$ controls the scale,

while "resistance to noise" $(\equiv \beta/\alpha^4)$

and "sensibility to contrast/threshold" $(\equiv (\beta/\alpha^6)^{1/4})$

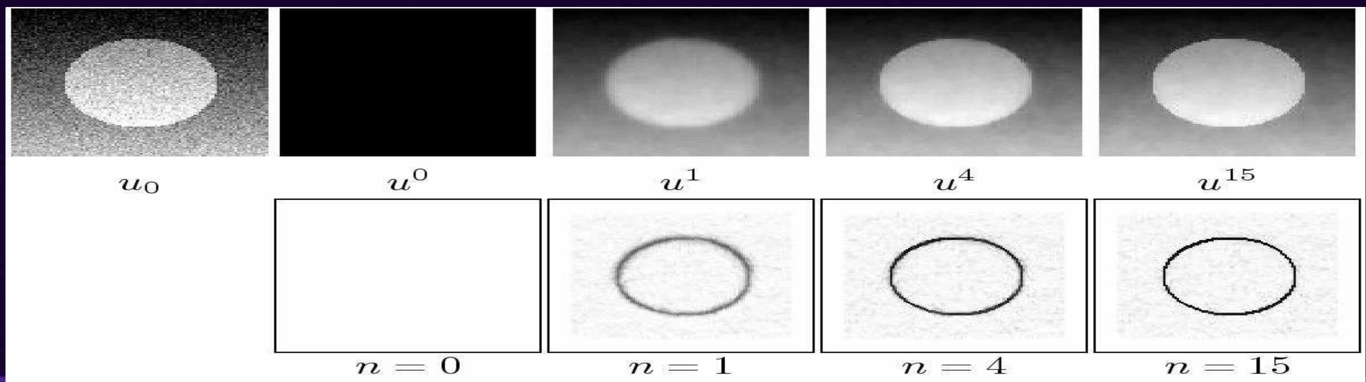
Specification of visual functions: image segmentation

Here $\int_K 1$ is the length of K in the Hausdorff sense (i.e. using the limit of the diameters of a covering)

The border K may be represented by an auxiliary function

$$z : W \longrightarrow [0,1] \text{ with } z/K \simeq 0 \text{ and } z/(W-K) \simeq 1$$

writing $\lambda_\epsilon(z) = \epsilon\,||\nabla z||^2 + \frac{(z-1)^2}{4\,\epsilon}$.



Specification of visual functions: image segmentation

Up to ϵ the Blake & Zisserman equations:

$$\left\{ \begin{array}{lcl} \dot{v} & \equiv & -(v-w) + \alpha^2 \left(z^2 \Delta v + 2\, z \nabla z^T \nabla v\right) \\ \dot{z} & \equiv & -\alpha^2\, z\, ||\nabla v||^2 + \beta \left(\epsilon\, \Delta z - \frac{z-1}{4\epsilon}\right) \end{array} \right.$$

solve the Mumford-Shah problem.

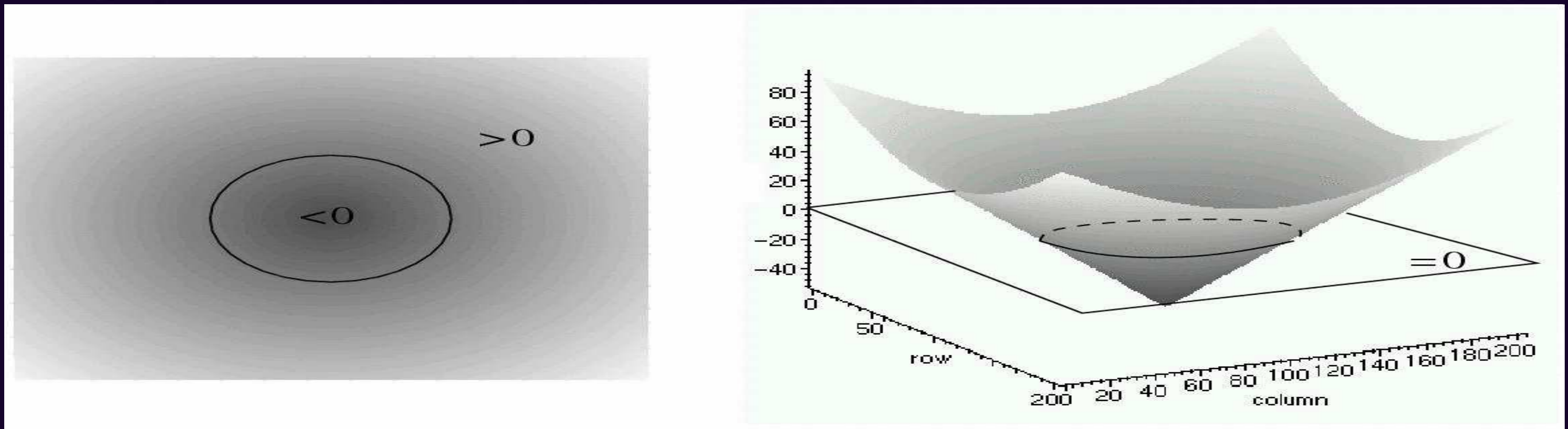
u_0 u u (random colormap) K

Specification of visual functions: image segmentation

- More generally, it involves two unknowns
  - u is a function defined on an N-dimensional space
  - K is an $(N-1)$-dimensional set.
- $E \to \mathcal{H}^{N-1}(\partial E)$ is not lower semi-continuous w.r.t. any compact topology.
- Solutions:
  - identifying the set of edges as the jump set of a BV function (see below)
  - approximation by elliptic functional (as done previously)
  - Chambolle discrete approximation by a suitable finite-difference scheme
  - etc..

Specification of visual functions: object segmentation

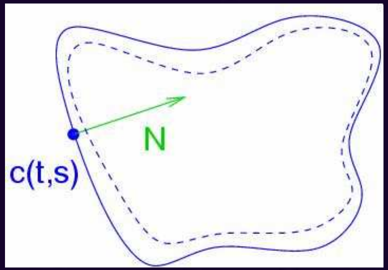
- Considering the figure/background segmentation
- The segmentation curve is defined a function level-set (Osher & Sethian)



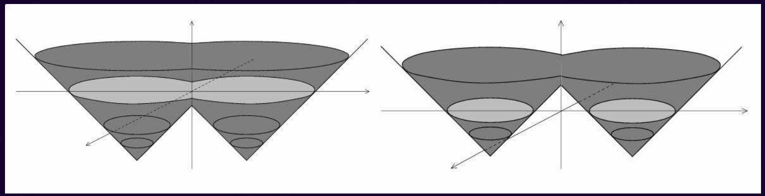
Specification of visual functions: object segmentation

- The level-set evolution induces the curve evolution

$$\begin{cases} \frac{\partial c}{\partial t} = v\, N, \\ c(0,q) = c_0(q). \end{cases} \Longrightarrow \begin{cases} \frac{\partial u}{\partial t} = v\, |\nabla u| \\ u(0,x) = u_0(x). \end{cases}$$

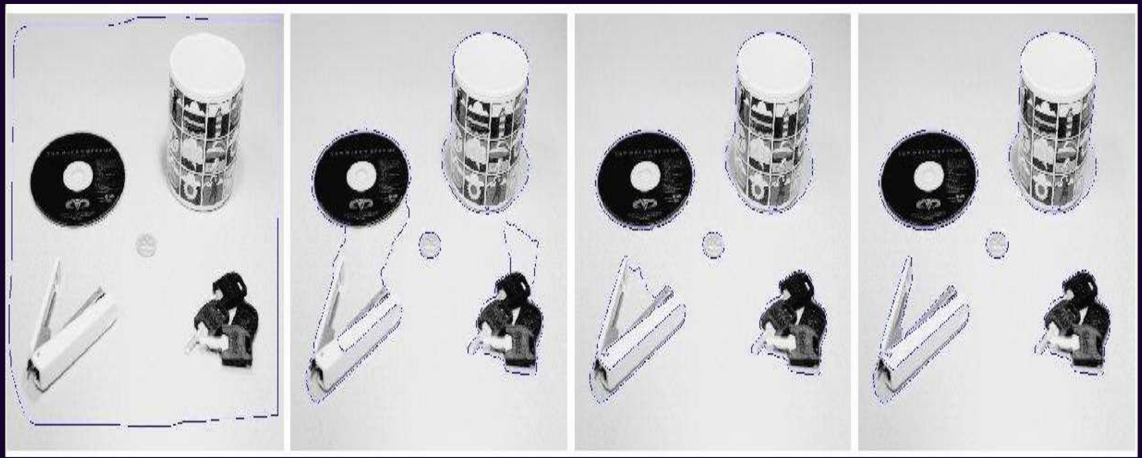


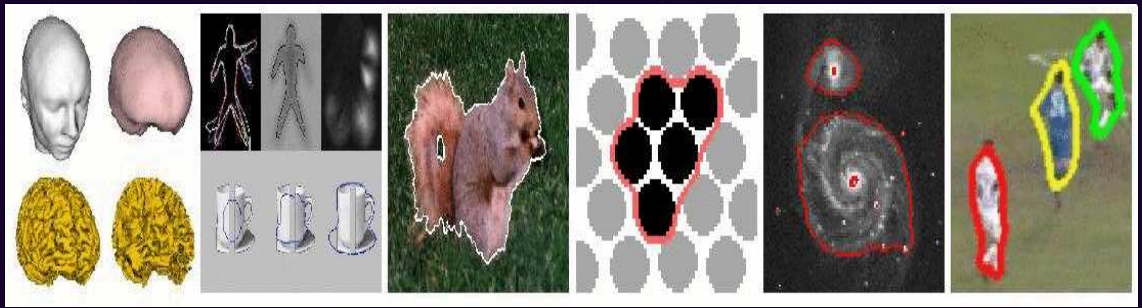
- Including with topological changes

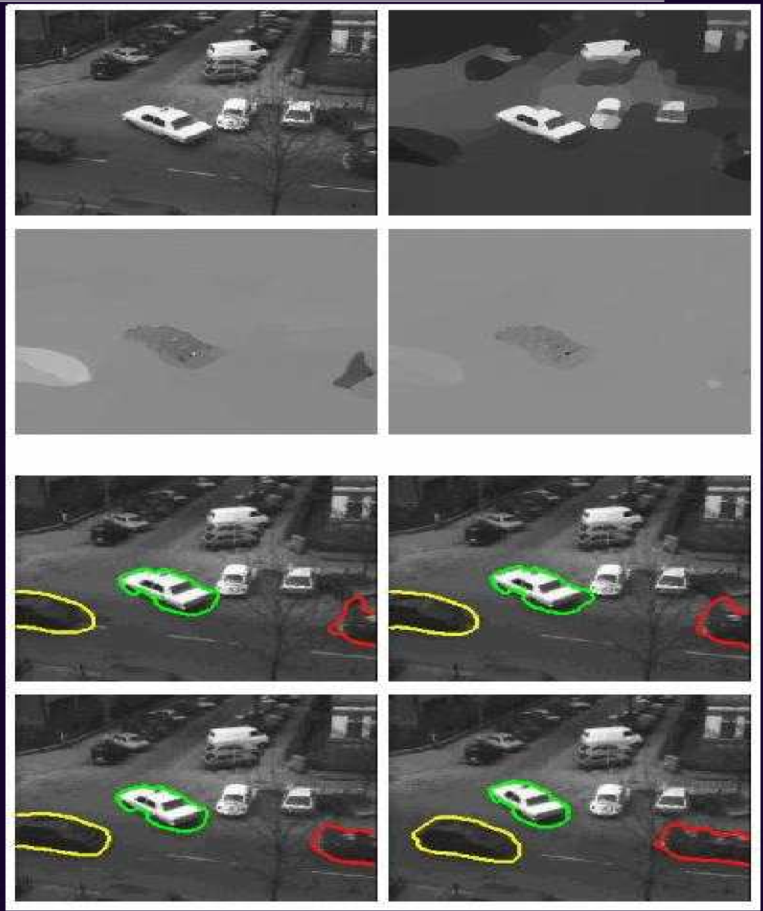


- Including in higher dimensions

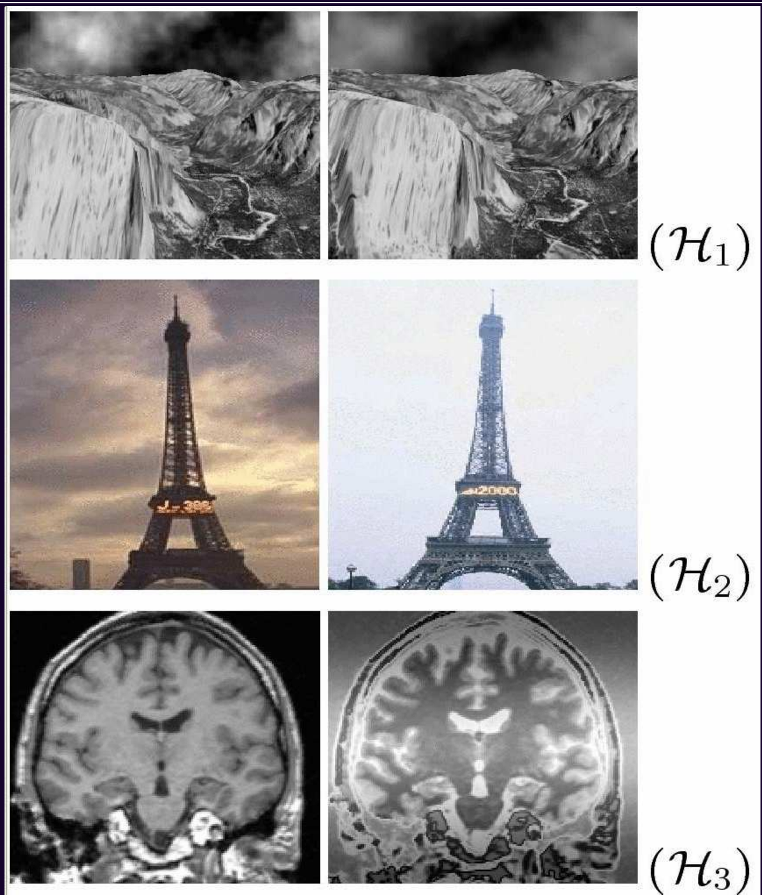
Specification of visual functions: object segmentation







Specification of visual functions: image matching



A large variety of problems / conditions:

($\mathcal{H}_1$) Intensity conservation

($\mathcal{H}_2$) Global intensity variation

($\mathcal{H}_3$) Local intensity variation

but a synthetic approach.

Specification of visual functions: image matching

- $(\mathcal{H}_1)$ Assuming intensity conservation
$$u(t+\delta t, x+\delta x) \simeq u(t,x)$$
defines the optical-flow constraint:
$$v = \frac{dx}{dt}, \quad v \cdot \nabla u(t,x) + \frac{\partial u}{\partial t}(t,x) = \varepsilon \simeq 0$$

- Approximate equation: true only for Lambertian surfaces in translation

- The approximation is better on edges (where $|\nabla u(t,x)| >> |\varepsilon|$)

- Aperture problem: only 1 equation, for a 2D problem

Specification of visual functions: image matching

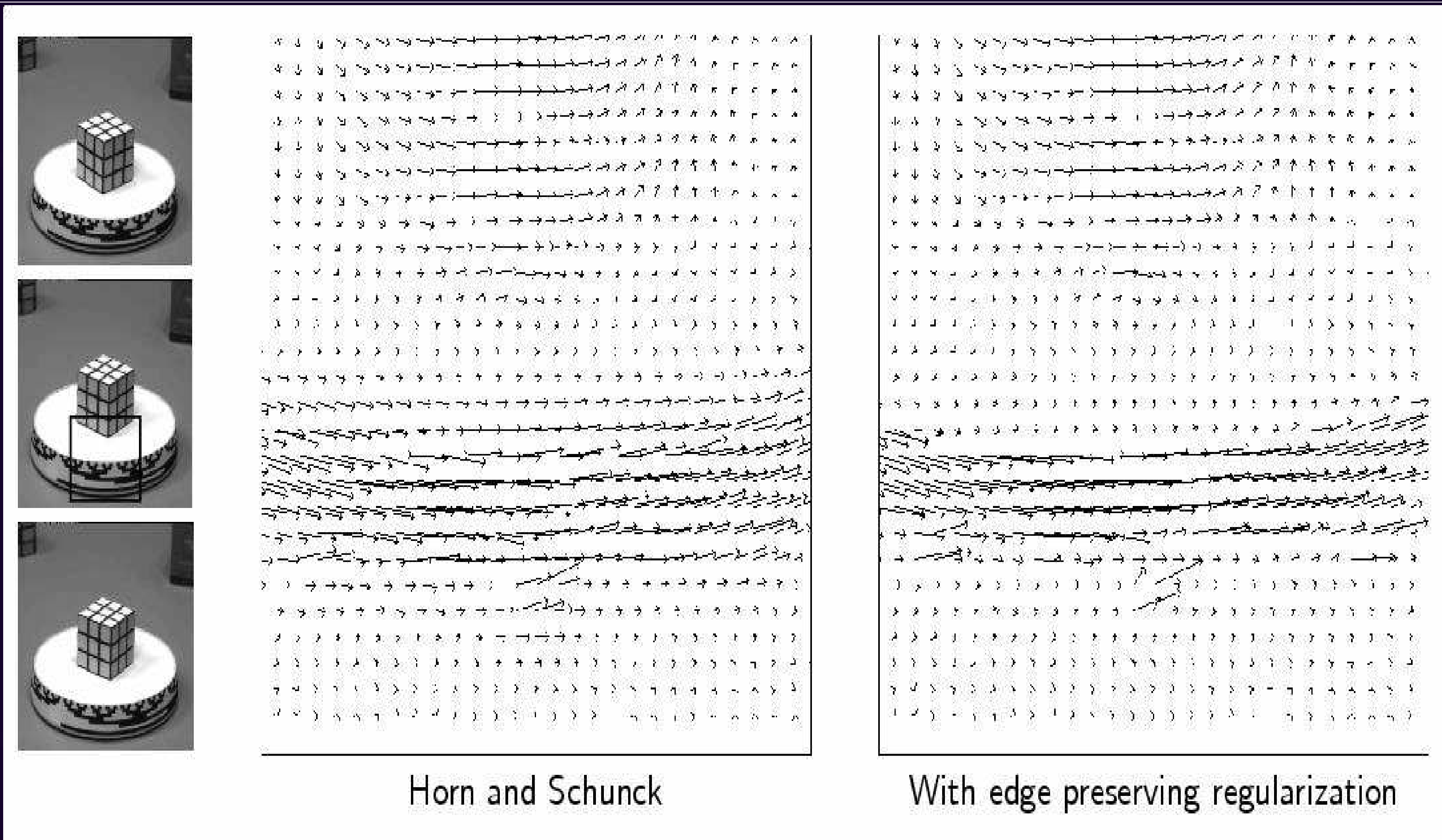
Specification of the solution:

$$\inf_u \int_\Omega A(v) + S(v)$$

$$\begin{aligned}
A(v) &= [v \cdot \nabla u + u_t]^2 \\
S(v) &= \sum_{j=1}^{2} \int_\Omega |\nabla v_j|^2 dx && \text{(Horn \& Schunck)} \\
&= \sum_{j=1}^{2} \int_\Omega \phi(|\nabla v_j|) dx && \text{(Preservation of discontinuities)} \\
&= \int_\Omega \varphi(\mathrm{div}(v), \mathrm{rot}(v)) dx && \text{(Differential properties)} \\
&= \int_\Omega \frac{\mathrm{trace}\left((\nabla v)^T \mathtt{D}(\nabla u)(\nabla v)\right) dx}{|\nabla u|^2 + 2\lambda^2} && \text{(Image properties)} \\
&= etc..
\end{aligned}$$

INRIA

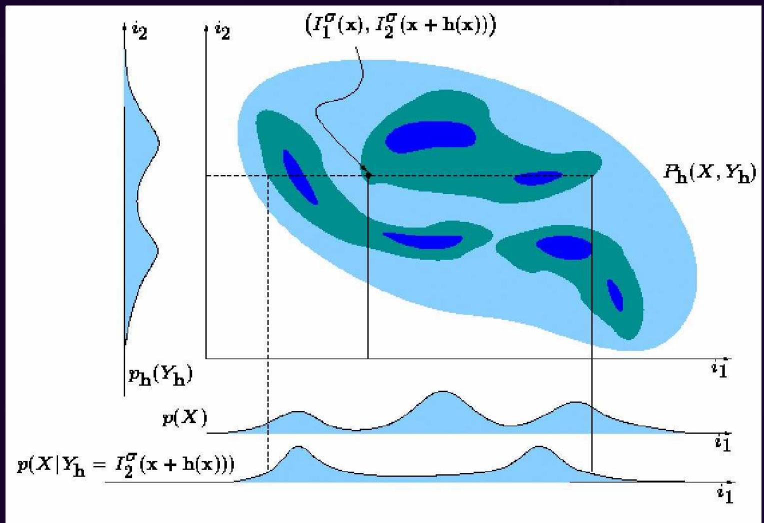
Specification of visual functions: image matching



Specification of visual functions: image matching

- $(\mathcal{H}_2)$ Assuming global intensity variation between, say

$$I_1 = u(t, x) \text{ and } I_2 = u(t + \delta t, x + \delta x) \text{ viewed as random variables}$$



A(v) is now computed on the joint histogram:

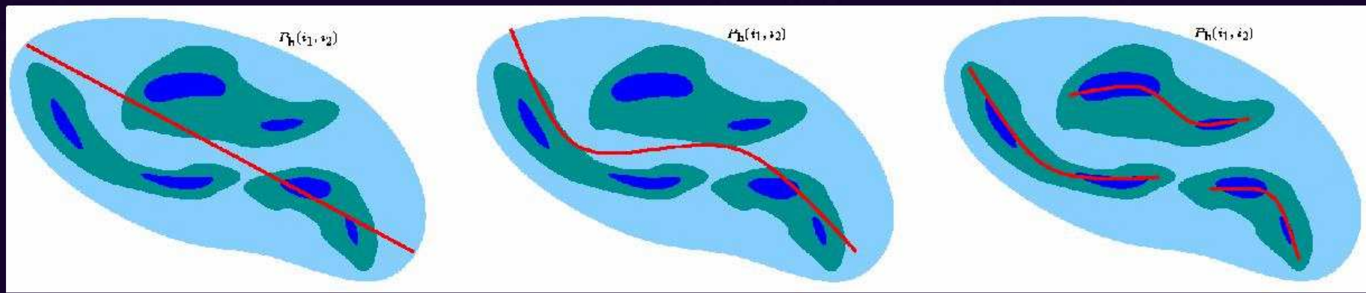
– Using Parzen density estimation

– i.e. Gaussian smoothing of the histogram

Specification of visual functions: image matching

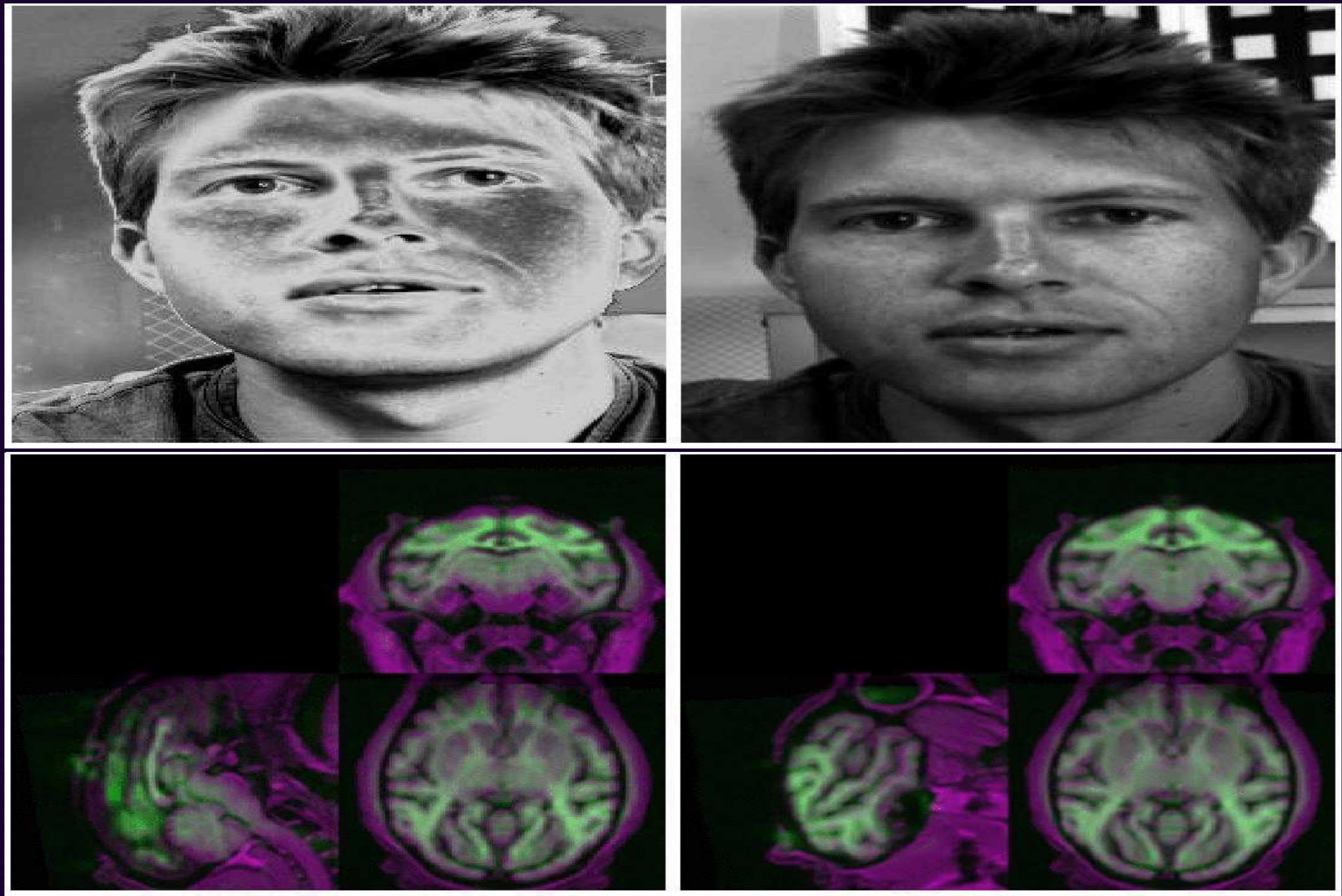
- The chosen criterion depends on the relation between the two images:

Cross correlation	Correlation ratio	Mutual information



Affine relation	Functional relation	Statistical relation

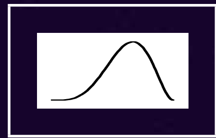
Specification of visual functions: image matching



Specification of visual functions: focus of attention

Combining diffusion and binarization:

$$min_v \quad \underbrace{||\nabla v||^2}_{\text{smoothness}} + \underbrace{\psi(v)}_{\text{binarization}} \qquad \psi : [0,1] \rightarrow \mathcal{R}$$

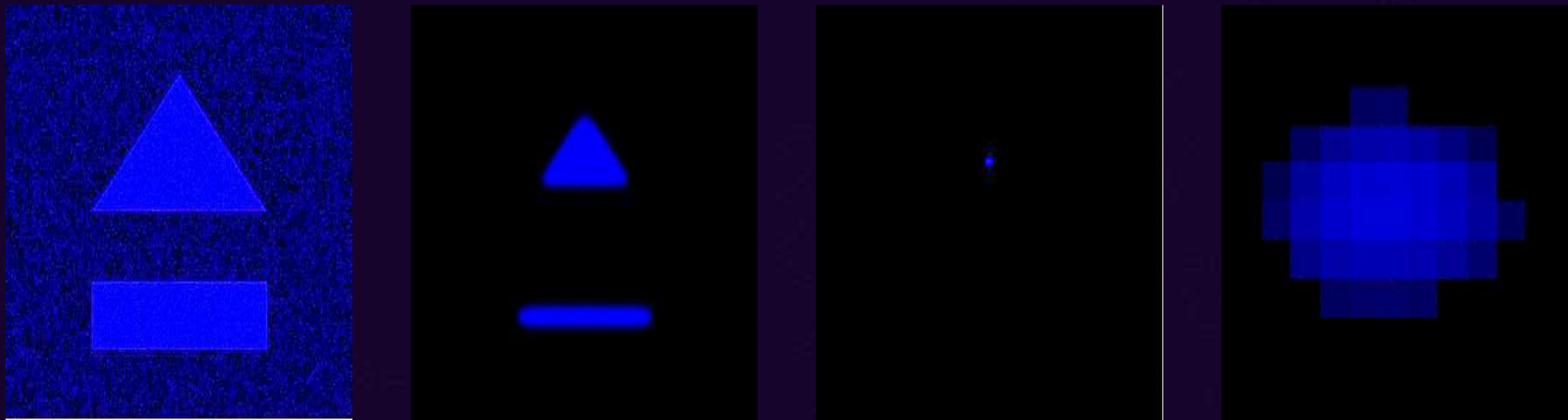


for some skew-symmetric bi-modal function $\psi()$ defining a threshold
- initialized to the distribution mean and
- incremented/decremented during the process

to maintain a small binarization with respect to diffusion

- the iteration is stopped when the output has a predefined small size.

Specification of visual functions: focus of attention



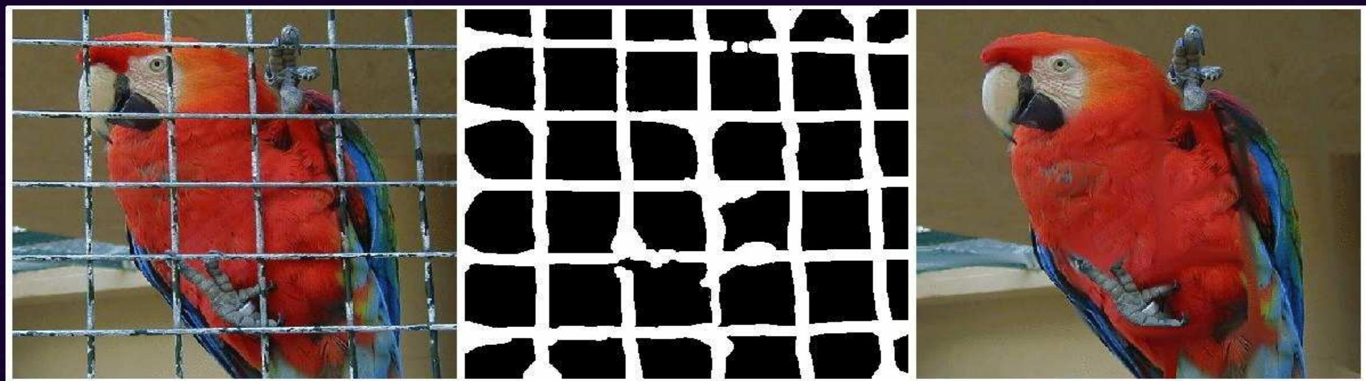
Input *Intermediate* *Output* *Output (zoom)*

An example of result for the winner-take-all mechanism implemented using the proposed method. The very noisy (more than 80%) original image is on the left; the intermediate result shows how diffusion is combined with erosion yielding the final result, shown also with a zoom.

Specification of visual functions: image completion

Same kind of criterion as for restoration with a distance to the image statistic

Before Mask After



The 45mn talk step by step

- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

AWYDWKAM: The foundation of the PDE approach

- We define a multi-scale analysis (or equivalently the scale-space) as a family of operators $\{T_t\}_{t\geq 0}$,
  which applied to the original image $u_0(x)$
  yield a sequence of images $u(t,x) = (T_t\, u_0)(x)$.

- We are going to list below a series of axioms to be satisfied by $\{T_t\}_{t\geq 0}$.
  (X denotes the space $C_b^\infty(\Re^2)$ and $u_0 \in C_b(\Re^2)$)
  These formal properties are very natural from an image analysis point of view.

AWYDWKAM: Axioms and Properties

(A1) *Recursivity:*

$T_0(u) = u, \;\; T_s \circ T_t(u) = T_{s+t}(u)$ for all s, $t \geq 0$ and all $u \in X$.

(A2) *Regularity:*

$|\, T_t(u + h\, v) - (T_t(u) + h\, v) \,|_{L^\infty} \leq c\, h\, t$ for all h and t in $[0, 1]$ and all u, $v \in X$.

(A3) *Locality:*

$(T_t(u) - T_t(v))(x) = o(t), \;\; t \to 0^+$ for all u and $v \in X$ such that $\nabla^\alpha u(x) = \nabla^\alpha v(x)$ for all $|\, \alpha \,| \geq 0$ and all x ($\nabla^\alpha u$ stands for the derivative of order α).

(A4) *Comparison principle:*

$T_t(u) \;\leq\; T_t(v)$ on $\Re^2$, for all $t \geq 0$ and $u, v \in X$ such that $u \leq v$ on $\Re^2$.

(I1) *Gray-level shift invariance:*

$T_t(0) = 0, \;\; T_t(u + c) = T_t(u) + c$ for all u in X and all constant c.

(I2) *Translation invariance:*

$T_t(\tau_h.u) = \tau_h.(T_t u)$ for all h in $\Re^2$, $t \geq 0$, where $(\tau_h.u)(x) = u(x + h)$.

AWYDWKAM: The main result

Alvarez et al. theorem: Under assumptions A1, A2, A3, A4, I1, and I2:

(i) There exists a continuous function $F : \Re^2 \times S^2 \rightarrow \Re$

satisfying $F(p, A) \geq F(p, B)$ for all $p \in \Re^2$, A and B in S^2 with $A \geq B$ such that

$$\delta_t(u) = \frac{T_t(u) - u}{t} \rightarrow F(\nabla u, \nabla^2 u), \quad t \rightarrow 0^+$$

uniformly for $x \in \Re^2$, uniformly for $u \in X$.

(ii) Then $u(t, x) = (T_t u_0)(x)$ is the unique viscosity (say "weak") solution of

$$\begin{cases} \dfrac{\partial u}{\partial t} = F(\nabla u, \nabla^2 u), \\ u(0, x) = u_0(x), \end{cases}$$

and $u(t, x)$ is bounded, uniformly continuous on $\Re^2$.

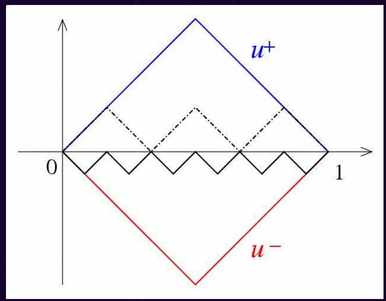
AWYDWKAM: What the hell is a "weak" solution ?

- A way to deal with non-linear degenerated equations:
$$\frac{\partial u}{\partial t}(t,x) + H(t,x,\nabla u(x), \nabla^2 u(x)) = 0, \quad t \geq 0, x \in \Omega$$
Here $H \; : \;]\,0,T\,] \times \Omega \times R \times R^N \times S^N \; \rightarrow \Im$ is continuous, elliptic and degenerated

  Here $u \in C(]\,0,T\,] \times \Omega)$ but not differentiable everywhere

- using test functions $\phi \in C^2(]\,0,T\,] \times \Omega)$ allowing to bound the solution

E.g. the eikonal equation: $\left\{ \begin{array}{l} |u'(x)| = 1 \quad \text{in} \quad [0,1], \\ u(0) = u(1) = 0, \end{array} \right.$



YAUHF: In which functional space do we work ?

We consider *functions of bounded variation*

. (= distributions which derivatives are measurable)

$$BV(\Omega) = \{u \in L^1(\Omega) / Du \in \mathcal{M}(\Omega)\}$$

with mainly an hyper-surface as singular set S_u (where upper/lower limits u^+/u^- differ) and which total variation is of the form (n_u is the normal to S_u) :

$$Du = \nabla u \cdot \mathcal{L}_N + (u^+ - u^-)\, n_u \cdot \mathcal{H}^{N-1}_{|S_u} + \underbrace{C_v}_{cantor\ \ part}$$

$\mathcal{H}$ is the Hausdorff measure (i.e. length, surface, etc.. of a curved space); while we consider $C_u = 0$ in practice.

In fact not optimal for textures, small structures:

an oscillatory component is also considered $v = div(g), g \in L^\infty$

YAUHF: In which functional space do we work ?

An example of BV + OSC decomposition:

YAUHF: Which properties to define the minimization ?

$$u_\bullet = \text{Argmin}_{u \in V} E(u)$$

- Inferior semi-continuity $\text{liminf}_{u_n \to u_\bullet} F(u_n) \geq F(u_\bullet)$
- Coercivity $\lim_{|u| \to +\infty} E(u) = +\infty$
- Convexity (for unicity)

allows to define a minimizing series of the energy (notion of Γ-convergence).

YAUHF: What the hell is Γ-convergence ?

$$\Gamma\text{-}\lim_{k\to\infty} E_k = E$$
$$\Leftrightarrow$$
$$\inf_{u_k\to u} \liminf_{k\to\infty} E_k(u_k) = \sup_{u_k\to u} \limsup_{k\to\infty} E_k(u_k)$$
$$\Leftrightarrow$$
$$\forall u_k \to u, E(u) \leq \liminf_{k\to\infty} E_k(u_k) \ \ \& \ \ \exists u_k \to u, \limsup_{k\to\infty} E_k(u_k) \leq E(u)$$

Main result:

If u_k is a minimizer of E_k and $u_k \to u$ then u is a minimizer of E

thus allowing to approximate a “singular” energy by a series of regular energy.

The 45mn talk step by step

- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

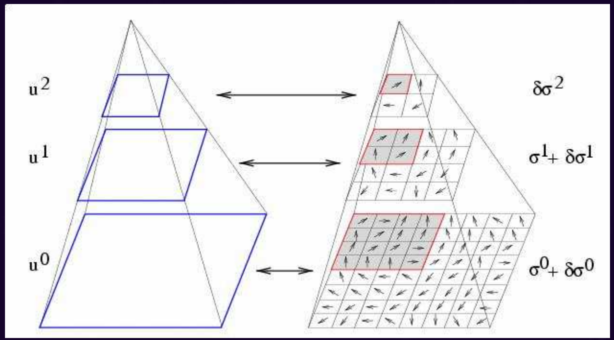
Implementing variational approaches: standard schemes

- Finite difference methods:

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) \quad \rightarrow \quad \frac{u_{i,j}^{n+1} - u_{i,j}^{n}}{\Delta t} = \Delta u_{i,j}^{n}$$

$$\rightarrow \quad u_{i,j}^{n+1} = u_{i,j}^{n} + \frac{\Delta t}{h^2}\left[u_{i+1,j}^{n} + u_{i-1,j}^{n} + u_{i,j+1}^{n} + u_{i,j-1}^{n} - 4u_{i,j}^{n}\right]$$

- Including multi-resolution framework



- Including semi-implicit schemes (solving a linear equation at each step)

Implementing variational approaches: standard schemes

- Linearization methods

$$0 = \frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) \quad \longrightarrow \quad 0 = \frac{\Delta t}{h^2}\left[u^n_{i+1,j} + u^n_{i-1,j} + u^n_{i,j+1} + u^n_{i,j-1} - 4u^{n+1}_{i,j}\right]$$

$$\longrightarrow \quad u^{n+1}_{i,j} = \frac{1}{4}\left[u^n_{i+1,j} + u^n_{i-1,j} + u^n_{i,j+1} + u^n_{i,j-1}\right]$$

- More generally

$$\min_u E(u) \quad \longrightarrow \quad u^{n+1}_{i,j} = (1-\alpha)\, u^n_{i,j} - \alpha\, \mathbf{M}\, \nabla E(u^n_{i,j})$$

  – the matrix $\mathbf{M}$ allowing to solve the linear part of ∇E,
  – the $\alpha \in]0,1]$ parameter controls the convergence.

Implementing variational approaches: Chambolle et al. scheme

- The continuous criterion is 1st approximated on a grid:

$$min_{u,K} \quad \int_W (u-u_0)^2 \quad + \quad \alpha \int_{W-K} ||\nabla u||^2 \quad + \quad \beta \int_K 1$$

$$min_u \quad \int_W (u-u_0)^2 \quad + \quad h^{n-1} \sum_{\mathbf{p}} \sum_{\mathbf{q}} \phi(\mathbf{q}) f_{\alpha,\beta} \left(\frac{(u(\mathbf{p}) - u(\mathbf{p} + h\,\mathbf{q}))^2}{h} \right)$$

  where:

$\phi(t)$	is a positive even, finite and small support profile with $\phi(0) = 0$ and $\int t^2 \phi(t) < +\infty$
$f_{\alpha,\beta}(t) = \beta\, f\left(\frac{\alpha}{\beta}\, t\right)$	is a suitable non-decreasing function $f(t) \leq \min(t, 1)$ (e.g. arctan)

- The Γ-convergence when $h \leftarrow 0$ is verified, and numerical approximations valid.

- The length $\int_K 1$ minimization is obtained thanks to the non-linear function $f()$.

Implementing variational approaches: Software architecture

- The software architecture is straightforward:
  - Map loaded with default values
  - Until convergence (on the criterion or the inter-iteration distance)
    - For each cortical map pixel (in sequence, randomly or in parallel)
      - Apply a local operator of the form

$$u_{i,j}^{n+1} = F(\{\cdots u_{i+u,j+v}^{n} \cdots\}, u \in \{-w..w\}, v \in \{-h..h\})$$

- Existing middle-ware defines image iterators and take into account the application of the operator on the map boundary must use performant full compiled code (see e.g. CImg open-source)

Implementing variational approaches: Convergence/complexity

- Complexity in $O(S)$ for an image of size $S = N^d$
- . . with "exponential fast" convergence (contraction) $\epsilon(t) < K\epsilon(t-1) < K^t\epsilon(0)$
- Parallel implementation is straight-forward
- Convergence to a local-minimum is garanty by construction
- . . and "convexification" allows to control which minimum
  $\rightarrow$ default/a-priori value closest solution

Implementing variational approaches: Hebbian schemes

- Consider the problem $\min_{\mathbf{u}} |\mathbf{u}|^2$ with $\mathbf{C}\,\mathbf{u} = \mathbf{u}_0$

  Any sequence $\mathbf{u}^{n+1} = \mathbf{u}^{n} - \gamma$ with $\left\{ \begin{array}{rcl} \gamma^T \mathbf{g} & > & 0 \\ |\gamma| & < & \varepsilon \end{array} \right.$

  writing $\mathbf{g} = (\mathbf{C}\,\mathbf{C}^T)^{-1}\,\mathbf{u}^n - \mathbf{C}^T\,\mathbf{u}_0$

  converges towards the minimum.

- Here γ is related to $\mathbf{g}$ combining the input $\mathbf{u}_0$ and output $\mathbf{u}^n$.

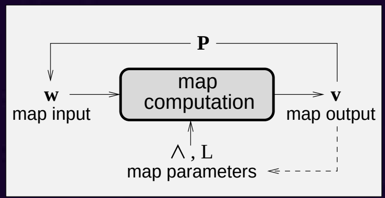
- This means γ small enough and approximately in the right direction

- Non-linear generalization is straight-forward

$\varepsilon = 2\,\cos(\widehat{\gamma, \mathbf{g}})\,|\mathbf{g}|/|\gamma/|\gamma||^2_{\mathbf{C}^T\,\mathbf{C}}$

Implementing variational approaches: Particular methods

- Given an input map w, one look for an output map $\bar{\mathbf{v}}$ verifying



$$\bar{\mathbf{v}} = \underset{\mathbf{v} \in H / \mathbf{c}(\mathbf{v})=0}{\operatorname{argmin}} \; \mathcal{L}(\mathbf{v}), \quad \text{with}$$

$$\mathcal{L}(\mathbf{v}) = \int_\Omega |\hat{\mathbf{w}} - \mathbf{w}|^2_{\mathbf{\Lambda}} + \int_\Omega \phi(|\nabla \mathbf{v}|_{\mathbf{L}}) + \int_\Omega \psi(\mathbf{v}),$$

and $\hat{\mathbf{w}} = \mathbf{P}\,\mathbf{v}$

- Here $|\mathbf{u}|_{\mathrm{M}} = \mathbf{u}^T \mathbf{M} \mathbf{u}$ is defined by a variable symmetric positive matrix $\mathbf{M}$.

- This defined an non-linear unbiased estimation (which includes almost all cases).

Implementing variational approaches: Particular methods

- The solution can be compiled on a "analog" neural network of the form:

$$\dot{v}_i = -\bar{\epsilon}_i(v_i) + \sum_j \bar{\sigma}_{ij}(v_i)\, v_j + \bar{\kappa}_i\, w_i$$

  - The weights $\bar{\sigma}$ corresponds to a discrete integral approximation of the diffusion operator $\mathcal{L}$

$$\Delta_{\mathbf{L}(\mathbf{x})}(\mathbf{f}(\mathbf{x})) \simeq \int_{\mathcal{S}} \bar{\sigma}(\mathbf{x},\mathbf{y})\, \mathbf{f}(\mathbf{y})\, d\mathbf{y} \text{ with } \int_{\mathcal{S}} \bar{\sigma}(\mathbf{x},\mathbf{y})^2\, d\mathbf{y} \text{ minimal}$$

    where $\mathcal{S}$ is a covering of the continuous map by the neuron's fields.
  - The corrective term $\bar{\epsilon}$ includes a leak and a non-linear adjustment of the threshold or delay.

- The compilation of the network parameters is straightforward.

Implementing variational approaches: Particular methods

More precisely, it writes:

$$\left\{ \begin{array}{rcl} \epsilon_i(v) & = & \rho_i\, \mathbf{v} + \xi\, \frac{\partial \mathbf{c}}{\partial \mathbf{v}}^T \mathbf{c} + \psi', \\ \rho_i & = & \sum_j \sigma_{ij} + \mathbf{P}^T \boldsymbol{\Lambda}_i\, \mathbf{P}, \\ \kappa_i & = & \mathbf{P}^T \boldsymbol{\Lambda}_i, \end{array} \right. \quad \text{and} \quad \begin{array}{c} \xi = (1-\lambda)\, |\frac{\partial \mathcal{L}}{\partial \mathbf{v}}| / |\frac{\partial \mathbf{c}}{\partial \mathbf{v}}^T \mathbf{c}| \\ \text{with } \lambda \ll 1. \end{array}$$

Up to order r ($r \geq 2$), at M points, providing $M > \frac{(n+r)!}{n!\,r!} - \frac{n\,(n+1)}{2}$,

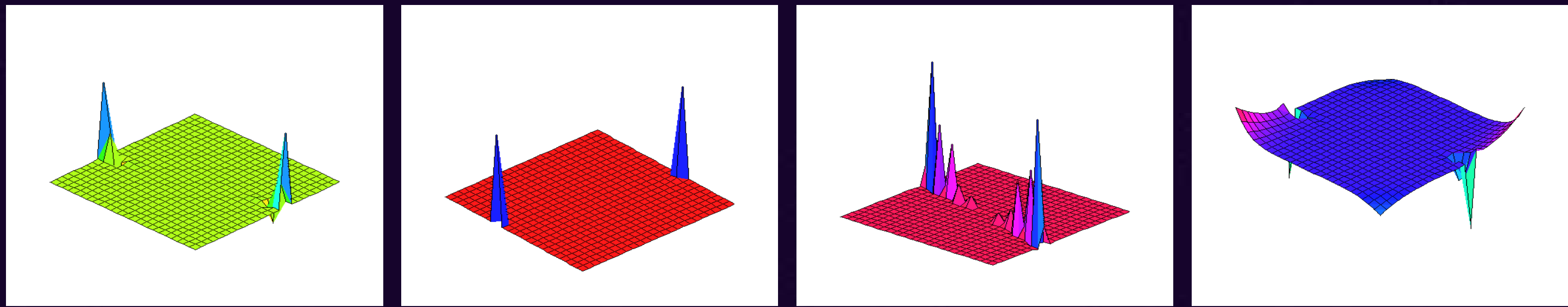
the weights $\sigma = (\sigma_{ij})$ come from:

$$\begin{array}{lrcll} |\alpha| = 2 & \bar{\mathbf{L}}^{kl}(\mathbf{x}) & = & \frac{1}{2} \sum_j \sigma_j\, \bar{\mu}_j^{\mathbf{e}_k+\mathbf{e}_l}(\mathbf{x}), & \text{with } \bar{\mathbf{L}} = \phi'(|\nabla \mathbf{v}|_{\mathbf{L}})\, \mathbf{L} \\ |\alpha| = 1 & \mathbf{div}^k(\bar{\mathbf{L}}(\mathbf{x})) & = & \sum_j \sigma_j\, \bar{\mu}_j^{\mathbf{e}_k}(\mathbf{x}), & \text{while } \sigma_j = (\sigma_{1j} \cdots \sigma_I \cdots) \\ 2 < |\alpha| \leq r & 0 & = & \sum_j \sigma_{ij}\, \bar{\mu}_j^{\alpha}(\mathbf{x}) & \text{(unbiasness)} \\ & & & \min \sum_{ij} \sigma_{ij}^2 & \text{(optimality)} \end{array}$$

which is a quadratic minimization under linear constraints

$\rightarrow$ unique generic closed-form solution

Integral approximations of a diffusion operator: examples



A few examples of operator 1D-profiles, considering an isotropic second-order derivative; from left to right:

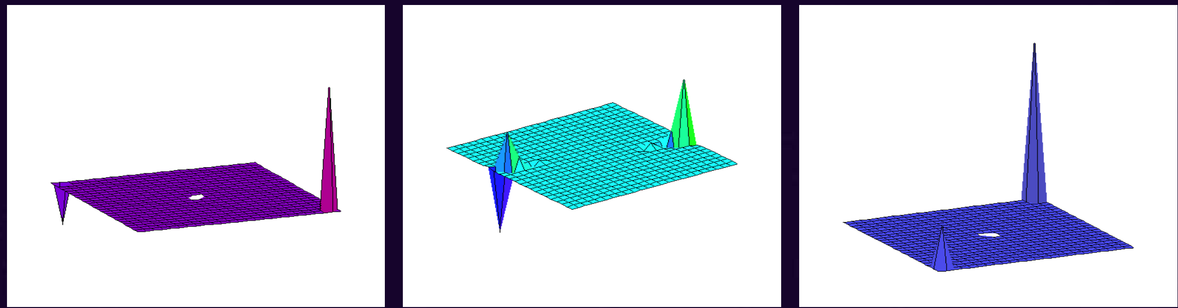
- r = 5, s = 10: we obtain a profile with two poles qualitatively equivalent to the δ'' distribution;

- r = 8, s = 20: increasing the order of correspondence, a profile closer to δ'' is obtained;

- r = 2, s = 3: when the correspondence is insufficient (r is too small) we obtain a profile which is qualitatively correct but very "flat";

- r = 6, s =10: when considering without any redundancy, the approximation may be slightly biased with spurious effects.

Integral approximations of a diffusion operator: examples



A few examples of operator 2D-profiles, with r $=3$, s $=6$, represented in the (x^0, x^1) plane;

- *left view* approximation of 1st order derivative isotropic operator $\partial^{(1,0)}$ qualitatively equivalent to the corresponding continuous operator;

- *middle view* approximation of 2nd order non-isotropic operator $L^{ij}(\mathbf{x}) = \delta^{ij}\, x^0$ and

- *right view* a 2nd-order non-isotropic operator $L^{ij} = \delta^{ij} + i$, both illustrating how solutions adapt to such profiles.

Integral approximations of a diffusion operator: examples

This mechanism not only generates numbers but also formulas !

$$\begin{bmatrix} .1048\,||\hat{\mathbf{n}}||^2 + .3782\,\hat{n}_x\,\hat{n}_y & .5053\,\hat{n}_y^2 - .2511\hat{n}_x^2 & .1048\,||\hat{\mathbf{n}}||^2 - .3782\,\hat{n}_x\,\hat{n}_y \\ .5053\,\hat{n}_x^2 - .2511\hat{n}_y^2 & .07255\,||\hat{\mathbf{n}}||^2 & .5053\,\hat{n}_x^2 - .2511\hat{n}_y^2 \\ .1048\,||\hat{\mathbf{n}}||^2 - .3782\,\hat{n}_x\,\hat{n}_y & .5053\,\hat{n}_y^2 - .2511\hat{n}_x^2 & .1048\,||\hat{\mathbf{n}}||^2 + .3782\,\hat{n}_x\,\hat{n}_y \end{bmatrix}$$

An example of anisotropic 2D-mask in the direction $\hat{\mathbf{n}} = (\hat{n}_x, \hat{n}_y)$ obtained for $r = 2$ or 3 and $s = 1$

The symbolic calculation thus output a piece of code

(automatic generation of Java/C++ code from Maple)

Implementing variational approaches: Particular methods

- The weight/threshold relation is compatible with standard STDP rules
- The architecture of an unit corresponds to an "abstract" cortical column

$\mathbf{w}$	Extra cortical input or intra-cortical input from previous layers
$\mathbf{v}$	Extra cortical or backward intra-cortical output
$\sum_j \sigma_{.j} \mathbf{v}_j$	Local connections
$\mathbf{\Lambda}, \mathbf{L}$	Remote backward connections
Iterative operations	Internal connections

- Stability of several cortical maps in interaction can be established

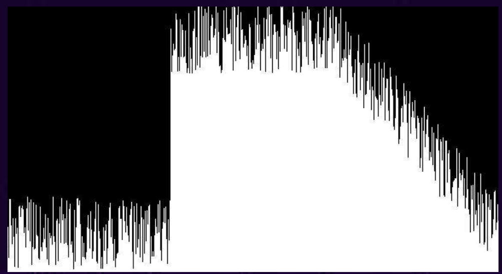
  Objective functions can be combined in this context

  Local minimization yields global optimization

- The method is valid for any local differential operator (here 1-2nd order)

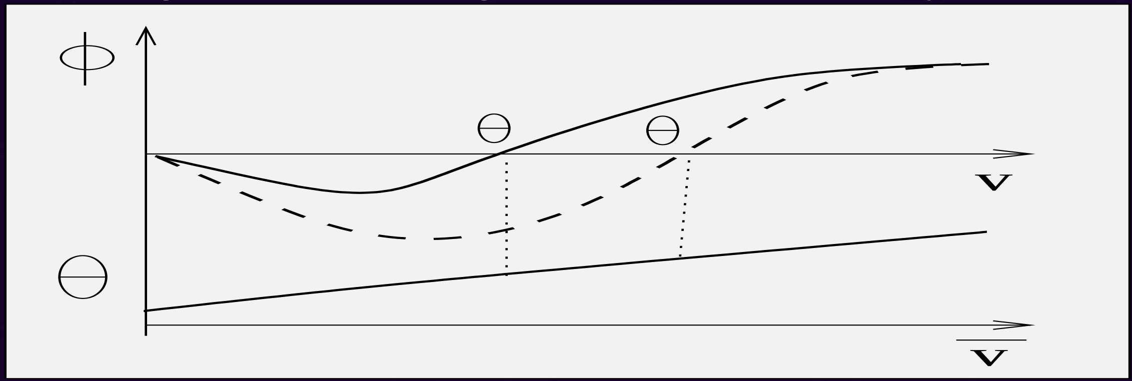
Implementing variational approaches: Particular methods

- The Maass-Natschläger use of piece-wise linear Gerstner and Kistler S.R.M. allows to derive a implementation on spiking-networks
- The information is coded by the spiking-time w.r.t. to a global clock
- The corrective terms correspond to an adaptive delay (compatible with the neuron biophysic)
- Only preliminary results available:



Implementing variational approaches: a link with the BCM rule

- The Bienenstock, Cooper & Munro rule states that the weight adaptation:
$$\dot{\sigma} = \phi(\mathbf{v}, \theta)\, \mathbf{w}$$
is proportional to the pre-synaptic activity $\mathbf{w}$
and proportional to a non-monotonic function ϕ of the post-synaptic activity $\mathbf{v}$
with some "depression" for low activity and "potentiation" for higher activity
the threshold θ being an increasing function of post-synaptic activity history $\bar{\mathbf{v}}$



Implementing variational approaches: a link with the BCM rule

- The BMC rule can be derived form an energy
  - which can be viewed as a measure of the amount of neuro-transmitter release
- It has been extended to network with feed-forward inhibition
- It has been also (weakly) linked to information theory

The 45mn talk step by step

- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

The brain is . . say . .

a machine to find "causes" ν from inputs u

via a functional equation of the form:

$$\boxed{u = P(\nu, \beta)}$$

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

. . using the Fliess fundamental formula and related Volterra kernels:

$$u(t) = \underbrace{\int_0^t \kappa_1(\tau)\,\nu(t-\tau)\,d\tau}_{\substack{\text{linear influence} \\ \text{from previous causes}}} + \underbrace{\int_0^t \int_0^t \kappa_2(\tau,\tau')\,\nu(t-\tau)\,\nu(t-\tau')\,d\tau d\tau'}_{\text{modulatory influence between causes}} + \cdots$$

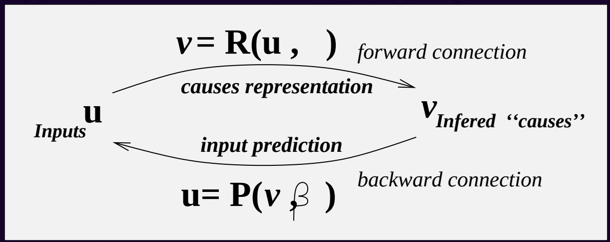
including higher order terms, this causal relationship is parametrized with:

$$\beta = \left[\kappa_1(\tau) = \left.\frac{\partial u(t)}{\partial \nu(t-\tau)}\right|_{t=0}, \kappa_2(\tau,\tau') = \left.\frac{\partial u(t)}{\partial \nu(t-\tau)\,\partial \nu(t-\tau')}\right|_{t=0}, \cdots\right]$$

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

Estimating causes ν from inputs u is -de facto- a forward/backward process:

- *Expectation:* which "infers" the causes from the given inputs (here parametrized by forward connections Φ) and
- *estiMation:* which "predicts" the input from "a-priory" causes (here parametrized by backward connections β)



the inference being coherent if and only if : $u = P(R(u, \Phi), \beta)$.

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

The Bayes approach ("maximally probable" estimation) ν, knowing u, thus:

$$\max_\nu log(p(\nu|u)) = \max_\nu [log(p(u|\nu)) + log(p(\nu))] - log(p(u))$$

(forget $log(p(u))$ constant with respect to the ν)

$\max_\nu$	$log(p(u\|\nu))$	$+$	$log(p(\nu))$
	Conditional information		A priory information
	β tuning : $u = P(\nu, \beta)$		Φ tuning : $\nu = R(u, \Phi)$
$\max_\nu$	$log(p(P(\nu, \beta)\|\nu))$	$+$	$log(p(R(u, \Phi)))$
	Estimation		Expectation

is a canonical instantiation of this architecture $\Rightarrow$ **criterion optimization**.

Cortical maps: interpretation of Grossberg systems

- A Cohen-Grossberg dynamical system is of the form:
$$\dot{u}_i = a_i(u_i)\left[b_i(u_i) - \sum_j c_{ij}\, d_j(u_j)\right]$$
with $a_i() > 0$ and $d'_j() > 0$ (convergence is demonstrated for the case where $c_{ij} = c_{ji}$).

- As soon as c_{ij} is unbiased (in practice local and mainly excitatory)
a Cohen and Grossberg dynamical system locally minimizes,in the general case:
$$\boxed{\frac{1}{2}\int \phi(||\nabla v||^2_L) + 2\,\psi(v)}$$ with $v = d(u)$ while $\psi(v) = -\int b(d^{-1}(v)) + \frac{1}{2}\,\nu\, v^2$
considering, an integral approximation of the diffusion operator $\phi'(||\nabla v||^2_L)\, L$

- Also applicable to Hopfield networks

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

Forward connections	**Backward connections**,
are "driving" for	are "modulatory" for
promulgation and segregation of sensory information	*mediation of contextual effects, co-ordination of processing*
consistent with	
(i) their sparse axonal bifurcation	(i) their frequent bifurcation
(ii) patchy axonal terminations	(ii) diffuse axonal terminations
(iii) topographic projections	(iii) non-topographic projections
(iv) one-to-one / small divergence	(iv) large spatial divergence
	(v) slow time-constants
(vi) define a lattice	(vi) transcend several levels
	(vii) more numerous

INRIA

Cortical maps: the Mumford-Dayan-Abbott-Friston roadmap

- **Where to process:**

  - a rough but fast edge detector feedback which areas have to analyzed in details

  - and automatically tune early-vision parameters

  - large scale (smoothed, eliminating noise) detector tune further process (e.g. figure/background segmentation)

  - low-level focus of attention towards *close*, *mobile* or *textured* feedback from rotational motion

- **What to process:**

  - choose processing modes, configurations of parameters with respect to first recognition,

  - drive visual tasks such as object-background segmentation, using fast categorization.

- **Holistic perception**: Holistic perception may be related to feedback from what has been detected by the "fast-brain".

- **Opportunism** : Feedback in the visual cortex seems to be used to select the relevant attributes, given a task / context.

Beyond visual functions: visual path planning

- Planning is huge abstract problem:

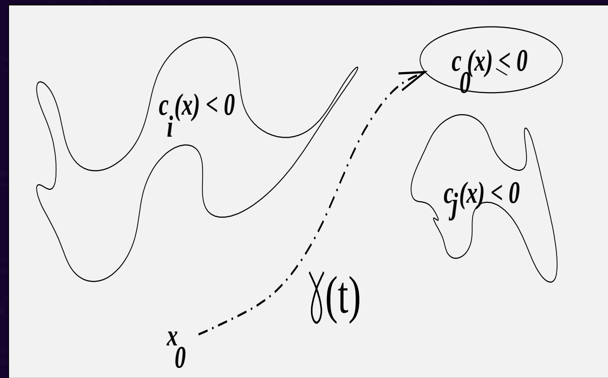
  Let us consider:

  (a) a system, defined by a *state vector* $\mathbf{x} \in \mathcal{R}^n$, $n \geq 2$

  (b) an *initial state*, written $\mathbf{x}_0 \in \mathcal{R}^n$,

  (c) r *constraints* / obstacles $c_i(\mathbf{x}) > 0, i \in \{1..r\}$,

  (d) a *goal* defined by an constraint of the form $c_0(\mathbf{x}) \leq 0$,



- Including: visual navigation, gesture generation, etc...

- Harmonic control introduced by Connolly-Grupen yields a variational solution

Beyond visual functions: visual path planning

- It is solvable by the minimization of an harmonic potential such that:
  - $\mathcal{C}_0$ The goal corresponds to minima of the potential.
  - $\mathcal{C}_i$ Obstacles are maxima of the potential.
  - $\mathcal{C}_c$ There is no local minimum (or flat regions) of the potential
- So that starting at any initial point and moving in the direction of potential decreases leads to the goal
- Such "loci-map" corresponds to hippocampal place fields (sparse representation)
- Other sensori-motor loops have been related to harmonic control

Beyond visual functions: data reduction

- Minimizing energy of the form $|u|^p = [\sum_i u_i^p]^{\frac{1}{p}}$ with $p < 1$ yields sparse solution (many $u_i = 0$, while $\lim_{p \to 0} |u|^p = \# u_i, u_i \neq 0$)
- Object categorization statistical learning is based on margin maximization again specified as a variational problem
- Dimensional reduction is also expressed as an optimization problem, e.g. a Kohonen map is specified via a potential (Fort & Pagès)
- etc . . .

INRIA

The 45mn talk step by step : done !

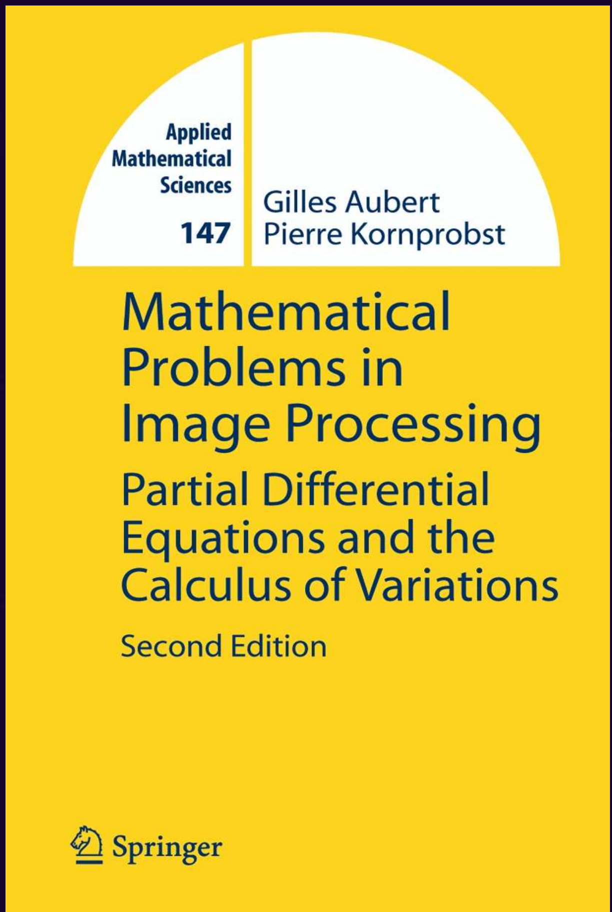
- (10mn) An introductory example
- (10mn) Specification of visual functions
- (05mn) All what you do not want to know about hidden maths
- (15mn) Implementing variational approaches
- (10mn) Generalization to other sensori-motor functions

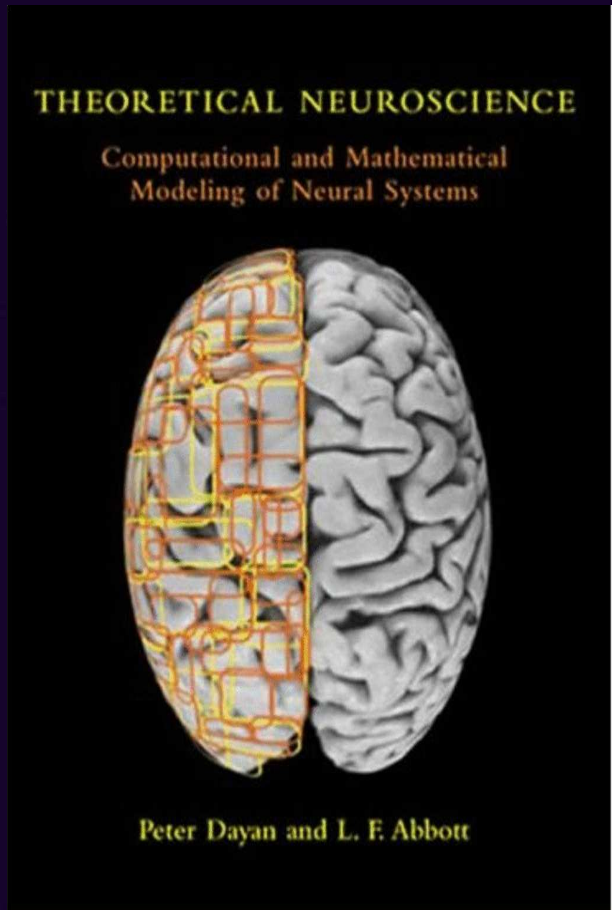
More . .

Three accessible documents and . . one software:

- Image Analysis and P.D.E.'s F Guichard et J-M Morel
- PDE-Based Regularization of Multivalued Images and Applications D Tshumperle
- Level Set methods S. Osher et R. Fedwik
- The CImg middle-ware open-source

More . .





FACETS contributions:

- Kornprobst et al. (cortical maps)
- Escobar et al. (high-level function)
- Kornprobst, Masson et al. (transparent motion)
- Deriche et al. (segmentation)
- etc..