5e Ecole Recherche Multimodale d'Information - TEchniques & Sciences

28 – 30 Septembre 2010
Presqu'île de Giens – Var
http://glotin.univ-tln.fr/ERMITES10



ERMITES 2010
Vision et Cognition

dédiée à l'analyse des processus de vision, à leurs modélisations et à leurs applications en recherche d'information

Intervenants

Jeanny HERAULT (GIPSA/UJF)
Perception Visuelle, faits et modèles

Jean PETITOT (EHESS/X)
Modèles neurogéométriques de Vision

Ugo BOSCAIN (CMAP/X)
Anthropomorphic image reconstruction via hypoelliptic diffusion

Claude TOUZET (LNIA/UNIV. MED)
Modèles cognitifs de l'attention visuelle

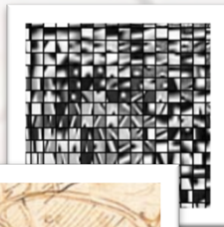
Jean-Paul GAUTHIER (LSIS/USTV)
Sur les mécanismes mis en œuvre par le système nerveux central

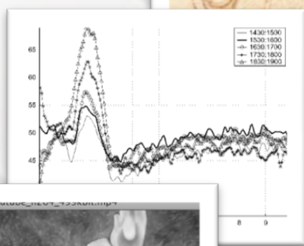
Hervé LE BORGNE (CEA-LIST)
Analyse en composantes indépendantes visuelles

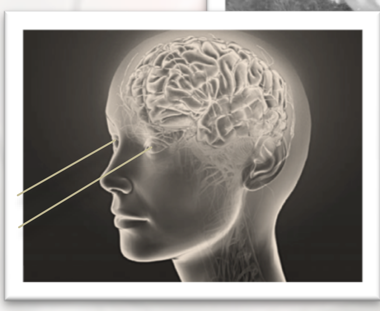
Julien MAIRAL (ENS/INRIA WILLOW)
Sparse Coding and Dictionary Learning

Hervé JEGOU (INRIA/IRISA)
Recherche d'image à grande échelle: procédés d'aggrégation & d'indexation

Sébastien PARIS (LSIS/UNIV. MED)
Reconnaissance de scène & robotique avec vision embarquée







Comités

Hervé GLOTIN
Président du comité de programme

Sébastien PARIS
Président du comité d'organisation

ISBN 2-9524747-3-7
EAN 9782952474733

LSIS – UMR CNRS 6168

INRIA

TOULON PROVENCE MÉDITERRANÉE
COMMUNAUTÉ D'AGGLOMÉRATION

UNIVERSITÉ du SUD Toulon-Var

Imprimé à La Garde en septembre 2010 à l'université Sud-Toulon Var

ISBN 2-9524747-3-7

EAN 9782952474733

Avant-propos

« La réalité se présente à nous sous la forme de phénomènes, de formes, dont nous décelons la présence par leurs discontinuités qualitatives »

Modèles mathématiques de la morphogénèse
René Thom, 1974

Pour sa 5e édition, l'école ERMITES 2010 met l'accent sur le couplage entre vision et cognition ("La Cognivision"), et ses applications en recherche d'information vidéo.

Ainsi sont traités des modèles cognitifs bio-inspirés, des modèles de classification automatique, en passant par des représentations en « mots visuels ». En filigrane de ces méthodes: la recherche d'information dans des espaces de très grande dimension.

Cette école regroupe un bon nombre des experts français de renommée internationale. Le comité de programme et d'organisation les remercient chaleureusement de leur investissement pour cette édition qui a atteint son objectif en regroupant la trentaine de participants.

Nous avons cette année filmé les conférences, qui sont disponibles en ligne (voir table des matières), ce qui permet au lecteur de lire les supports tout en écoutant leur auteur, et donc d'en extraire toute l'information.

Nous remercions l'INRIA, l'UMR CNRS LSIS, TPM et l'USTV sans qui cette école n'existerait pas.

Comité de programme :
H. Glotin (prés.), S. Paris, J Razik, J.-P. Gauthier

Comité d'organisation :
S. Paris (prés.), H. Glotin, J. Razik, A. Zidouni, F. Bénard, M. Chouchane

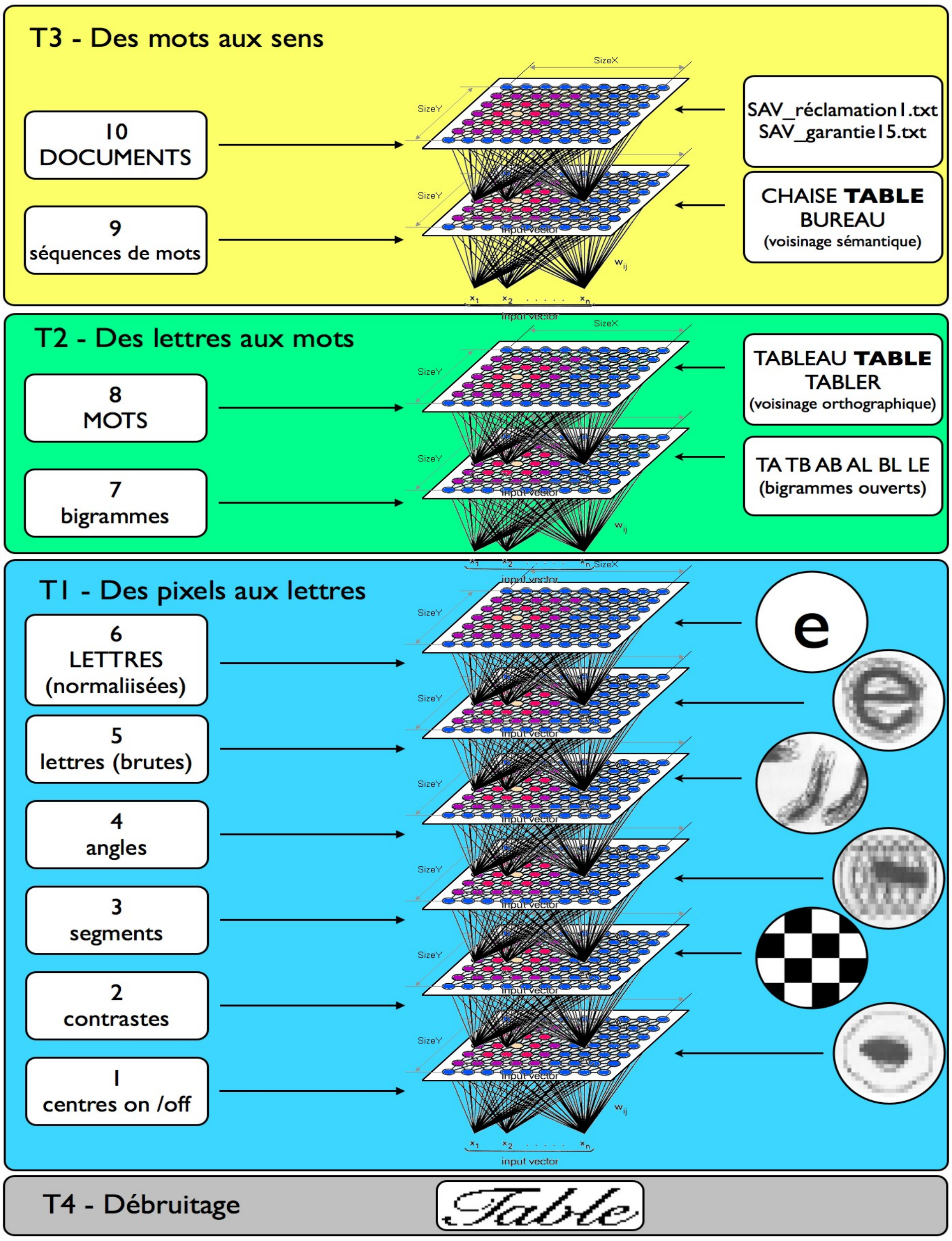
Le 28 septembre 2010,
A La Gardo,
Hervé Glotin & Sébastien Paris

De gauche à droite: J. Petitot, F. Rossi, J. Hérault, J. Mairal, S. Paris, H. Le Borgne, A. Monnin, F. Bouchara, H. Glotin, Y. Wazaefi, O. Caron, C. Thouzet, P. Machart, I. Azarkh, N. Foucault, O. Kleindiest, J. Razik, U. Boscain, R. Delaye, J.-P. Gauthier, C. Maggia, Y. Lacroix, S. Madec, J. Demongeot, H. Jégou, H. Queste
Hors photo: B. Fertil, L. Boutora, D. Merad

Appendice

Le thème de cette année s'inscrit naturellement dans le projet Dynamiques de l'Information du LSIS (http://www.lsis.org/dyni), et dans son projet ANR CONTINT COGNILEGO (http://cognilego.univ-tln.fr) qui est lancé ce mois-ci avec ses partenaires LNIA et A2IA .

Nous y proposons de développer des modèles cognitifs d'intégration de l'information visuelle pour la transcription robuste de documents. Notre approche repose sur des traitements pyramidaux auto-organisés, à différentes échelles sémantiques comme illustrés ci-dessous. Nous espérons qu'ERMITES 2010 suscitera d'autres projets alliant cognition, vision et recherche d'information.



"PERCEPTION, CATEGORISATION VISUELLE & COGNITION"

Programme

Mardi 28 septembre

- **10h** *Café / réception / visite du site*
- **11h** H. Glotin - présentations des journées
- **12h** *Déjeuner*
- **13h30** J. Hérault
- **16h30** *Pause café*
- **17h** H. Le Borgne
- **20h** *Dîner*

Mercredi 29 septembre

- **8h45** J. Petitot
- **10h** P*ause café*
- **10h20** J. Mairal
- **11h20** H. Jégou
- **12h30** *Déjeuner*
- **13h45** J. Petitot (suite)
- **15h** *Pause café*
- **15h30** C. Touzet
- **16h30** U. Boscain
- **18h30** J.-P. Gauthier
- **20h15** *Dîner*
- **21h** C. Touzet (suite)

Jeudi 30 septembre

- **8h30** J. Mairal (suite)
- **9h30** H. Jégou (suite)
- **10h15** *Pause café*
- **10h30** S. Paris
- **12h30** *Déjeuner*
- **13h45** Table ronde
- **15h30** Clôture et visite du site.

J. Hérault..1

« Perception visuelle, faits et modèles »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_HERAULT_1SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_HERAULT_2SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_HERAULT_3SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_HERAULT_4SUR4.mp4

J. Petitot..57

« Modèles neurogéométriques de vision »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PETITOT_1SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PETITOT_2SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PETITOT_3SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PETITOT_4SUR4.mp4

U. Boscain..159

« Anthropomorphic image reconstruction via hypoelliptic diffusion »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_BOSCAIN_1SUR2.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_BOSCAIN_2SUR2.mp4

C. Touzet..191

« Modèles cognitifs de l'attention visuelle »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_TOUZET_1SUR2.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_TOUZET_2SUR2.mp4

H. Le Borgne..203

« Description de scènes naturelles par composantes indépendantes »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_LEBORGNE_1SUR2.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_LEBORGNE_2SUR2.mp4

J. Mairal..235

« Sparse coding and dictionary learning for image analysis »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_MAIRAL_1SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_MAIRAL_2SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_MAIRAL_3SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_MAIRAL_4SUR4.mp4

S. Paris..331

« Reconnaissance de scènes & robotique avec vision embarquée »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PARIS_1SUR2.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_PARIS_2SUR2.mp4

H. Jégou..359

« Recherche d'image à grande échelle: procédés d'agrégation & d'indexation »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_JEGOU_1SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_JEGOU_2SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_JEGOU_3SUR4.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_JEGOU_4SUR4.mp4

J.-P. Gauthier..401

« Sur les mécanismes optimaux mis en œuvre par le système nerveux central »
http://lsis.univ-tln.fr/~glotin/ERMITES_10_GAUTHIER_1SUR2.mp4
http://lsis.univ-tln.fr/~glotin/ERMITES_10_GAUTHIER_2SUR2.mp4

Jeanny HERAULT

GIPSA Lab. / UJF

http://www.gipsa-lab.inpg.fr/page_pro.php?vid=76

"Perception Visuelle, Faits et Modèles"

L'exposé se déroule selon six grandes parties :

1) Illusions visuelles :
Classées par catégories, elles nous permettront, par les questions qu'elles suscitent, d'évoquer les principes de traitement du système visuel. Nous verrons que celui-ci s'est adapté au monde 3D où nous vivons et aux variations des conditions dans lesquelles les objets sont vus.

2) La rétine :
Point d'entrée du système visuel, elle est le siège de prétraitements (filtrages linéaires, non-linéarités adaptatives) qui préparent le signal au mieux pour son analyse et son interprétation par les différentes couches du cortex visuel.

3) Les circuits neuromimétiques :
Le modèle électrique de la rétine a conduit à un ensemble de circuits dits "neuromimétiques" dont nous donnons un exemple pour l'estimation du mouvement en temps réel.

4) Le codage des couleurs :
Ah, si nos ingénieurs avaient connu le traitement rétinien des couleurs, on aurait actuellement une TV de bien meilleure qualité! Nous verrons que c'est le cortex visuel qui décode la couleur et non pas la rétine, et que les principes qu'elle utilise sont économiques et efficaces, surtout de l'échantillonnage aléatoire des photorécepteurs.

5) Les non-linéarités :
Une analyse en détail des non-linéarités et de l'adaptation dans la rétine nous permet de comprendre certaines des illusions visuelles du début et donnent des pistes intéressantes pour le traitement des images (égalisation des niveaux d'intensité, des contrastes et constance des couleurs).

6) Le traitement cortical :
L'analyse des spectres 2D locaux des images ou des scènes par le cortex visuel primaire nous permet d'aborder les aspects de catégorisation, de saillance et d'estimation de la perspective. Quant à l'analyse des aires supérieures (en particulier V4), elle nous conduit aux propriétés d'invariance par rapport à l'échelle, aux rotations des images et aux effets de perspective.

Référence :

"VISION: IMAGES, SIGNALS AND NEURAL NETWORKS Models of Neural Processing in Visual Perception" J. Herault, Ed. Worldscibooks 2010, 308p.

ERMITES Gien, September 2010

VISUAL PERCEPTION

Facts and Models

CONTENTS

I- VISUAL ILLUSIONS

II- THE RETINA

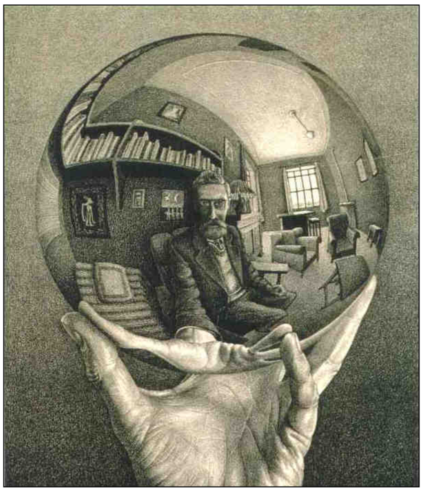
III- NEUROMORPHIC CIRCUITS

IV- COLOR CODING

V- NON-LINEARITY

VI- CORTICAL PROCESSING

VISUAL ILLUSIONS

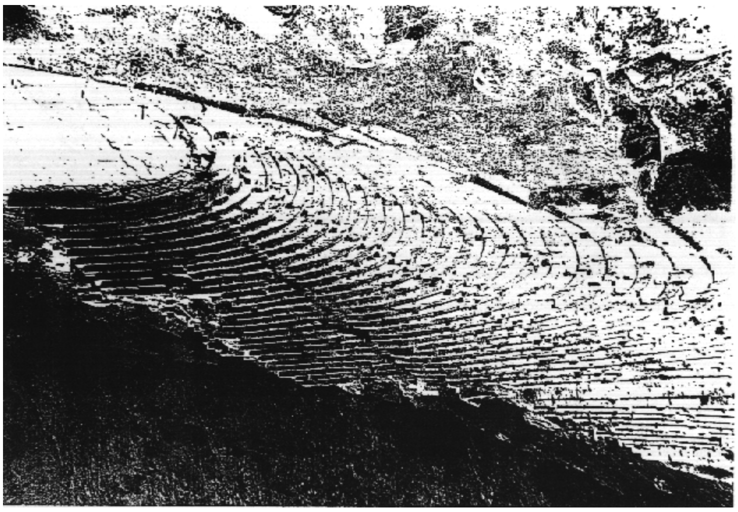


I- SHADOWS AND 3D WORLD

II- ADAPTATION TO SPATIAL CONTEXT

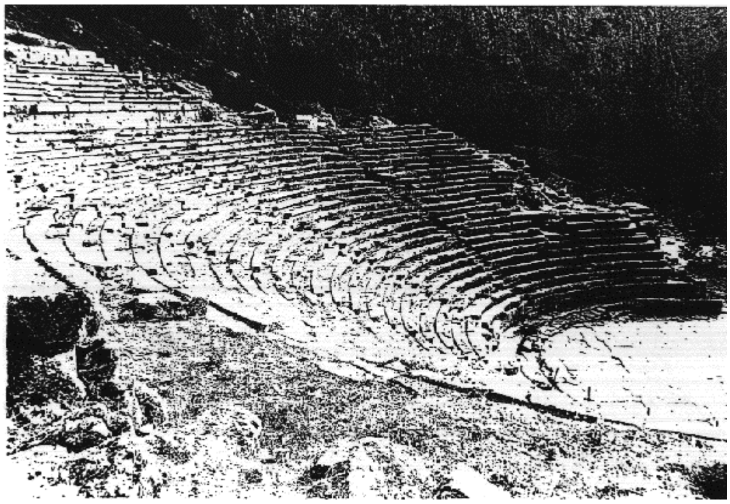
III- ADAPTATION TO TEMPORAL CONTEXT

AN UNKNOWN BUILDING



A WELL KNOWN BUILDING

A WELL KNOWN BUILDING



AN UNKNOWN BUILDING

CAST SHADOW & LEVITATION

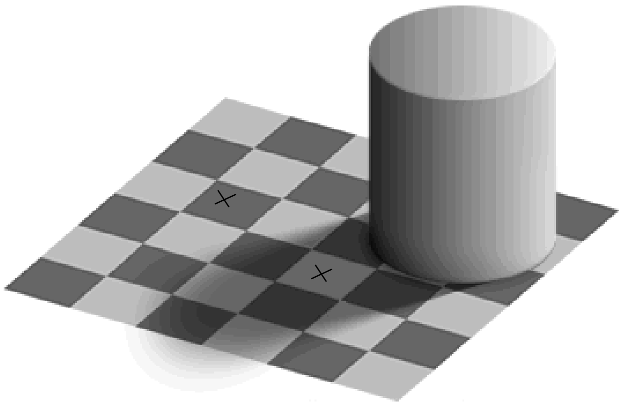
VISUAL ILLUSIONS

I- SHADOWS AND 3D WORLD

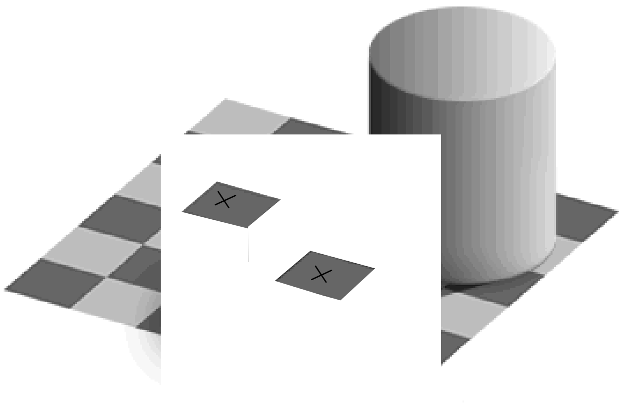
II- ADAPTATION TO SPATIAL CONTEXT
Intensity, contrast, color, shape…

III- ADAPTATION TO TEMPORAL CONTEXT

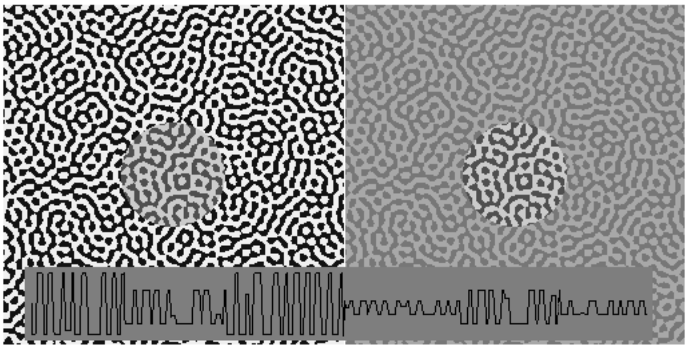
CHECKER BOARD



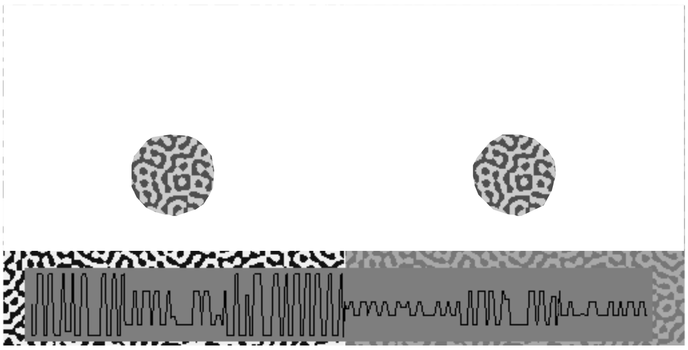
CHECKER BOARD



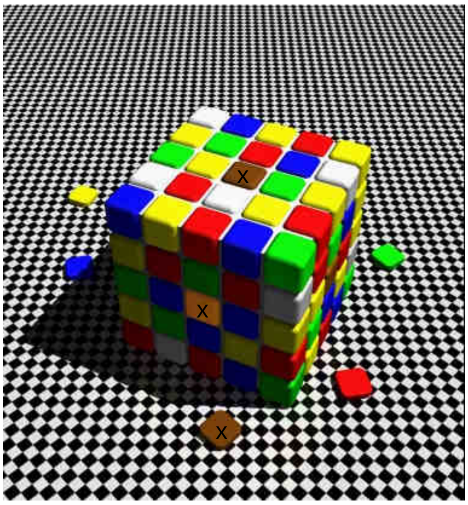
SIMULTANEOUS CONTRASTS



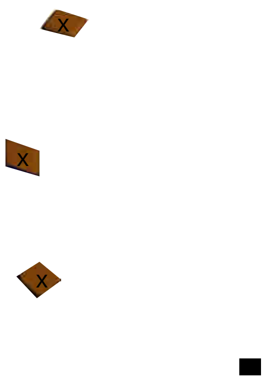
SIMULTANEOUS CONTRASTS



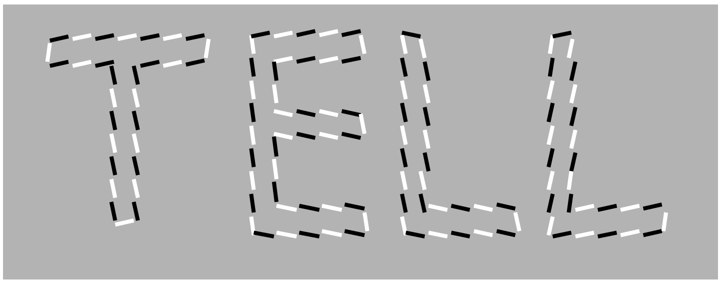
RUBICK 's CUBE



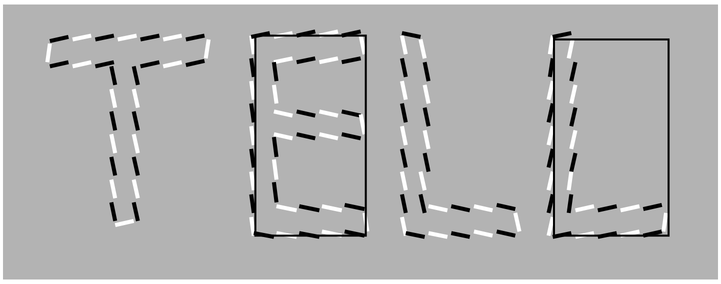
RUBICK 's CUBE



TELL… ME



TELL… ME



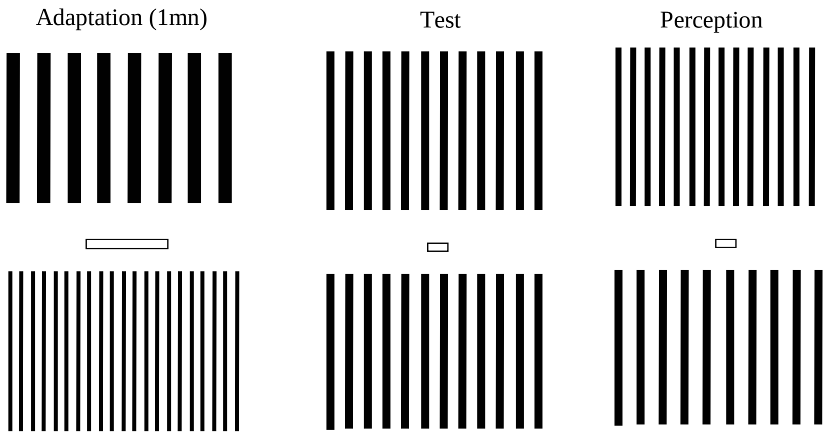
VISUAL ILLUSIONS

I- SHADOWS AND 3D WORLD

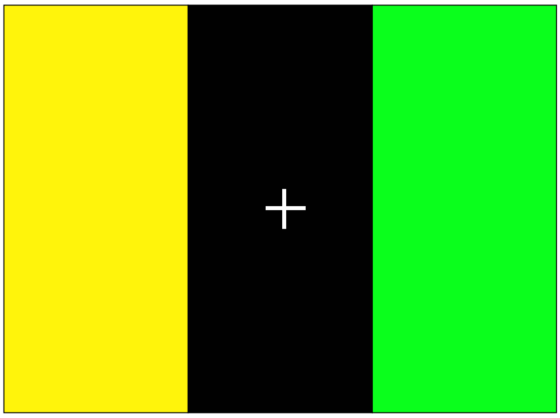
II- ADAPTATION TO SPATIAL CONTEXT

III- ADAPTATION TO TEMPORAL CONTEXT
Orientation, spatial frequency, color, motion...

ADAPTATION TO SPATIAL FREQUENCY



FIXATE THE WHITE CROSS



?...

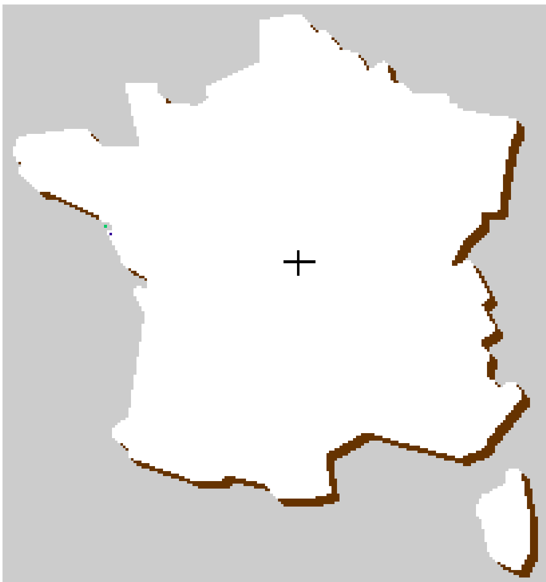
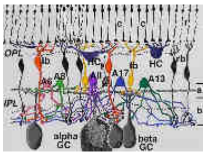


IMAGE PROCESSING IN THE RETINA



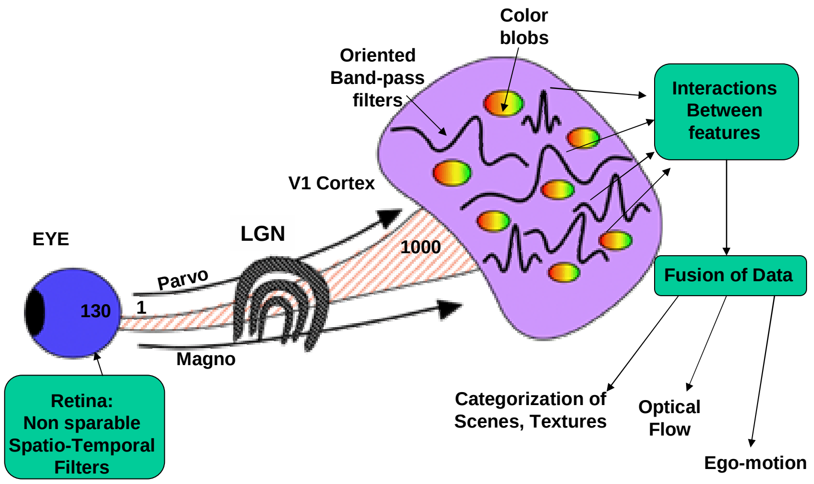
I- RETINAL CELLS & ELECTRICAL MODEL

II- SPATIO-TEMPORAL PROPERTIES

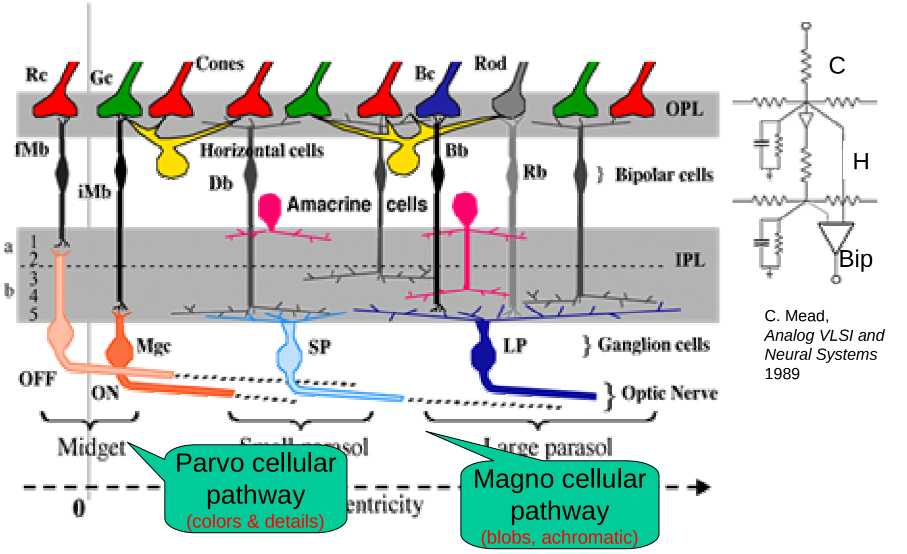
III- EXAMPLES

IV- SYNTHESIS and SILICON RETINAE

Visual System: Gross Anatomy



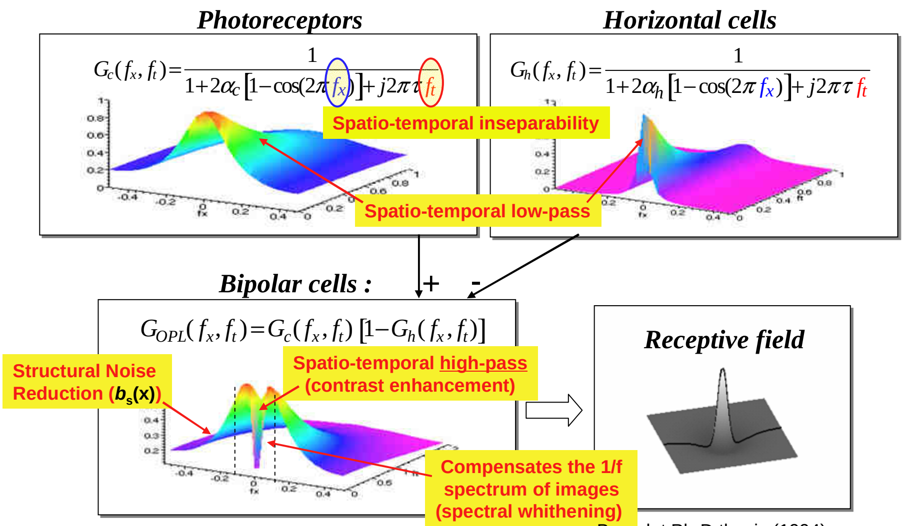
Retina: Linear Model (basics)



C. Mead, *Analog VLSI and Neural Systems* 1989

Retina: Parvo Cellular Pathway

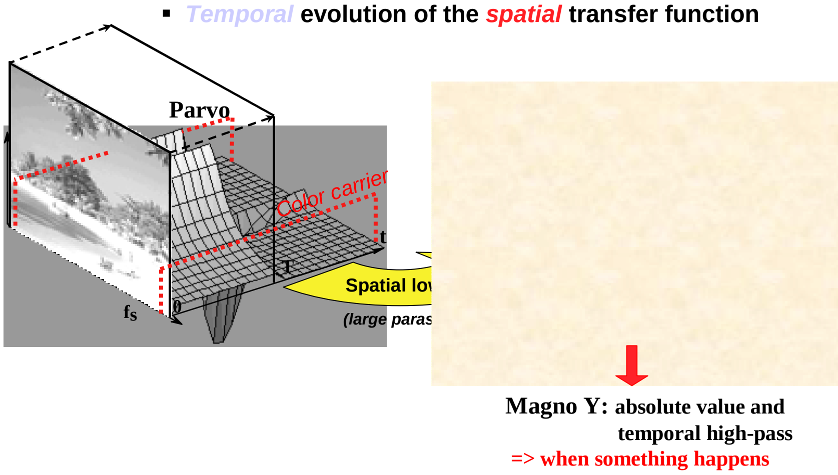
Transfer function (properties)



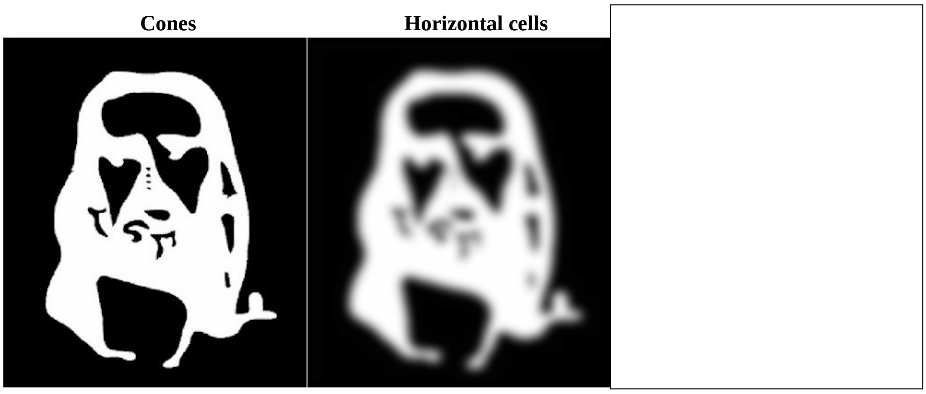
Beaudot Ph-D thesis (1994)

Parvo- & Magnocellular Pathways

- *Temporal* evolution of the *spatial* transfer function



CAN YOU SEE YOUR HORIZONTAL CELLS ?



CAN YOU SEE YOUR HORIZONTAL CELLS ?

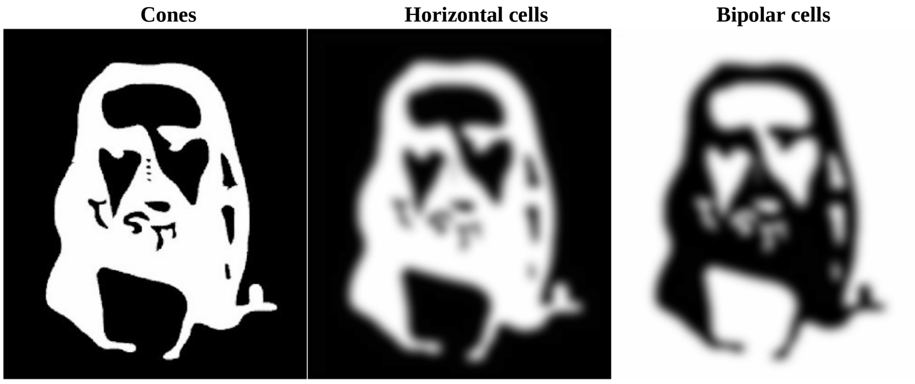


IMAGE PROCESSING IN THE RETINA

I- RETINAL CELLS & ELECTRICAL MODEL

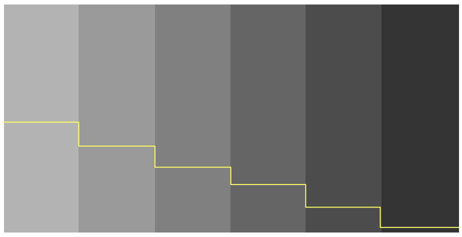
II- SPATIAL & TEMPORAL PROPERTIES

III- EXAMPLES

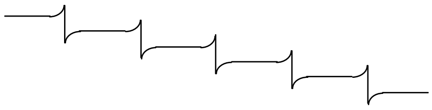
IV- SYNTHESIS and SILICON RETINAE

MACH BANDS

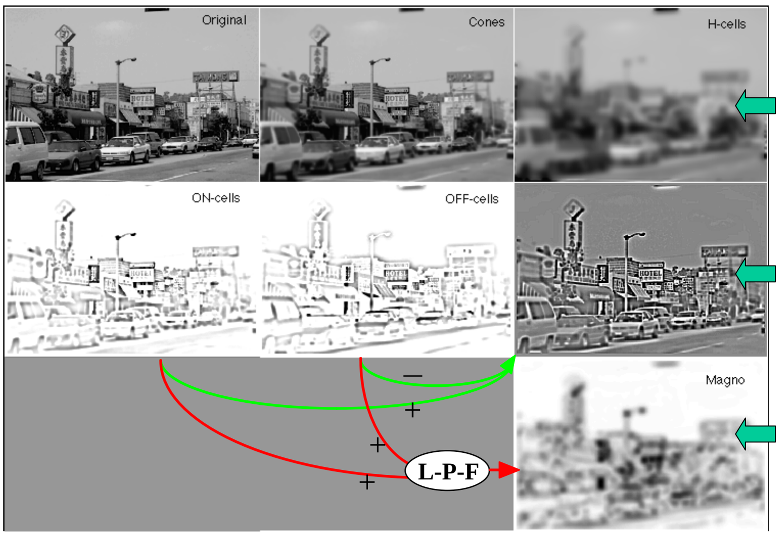
Seen



Perceived



SUMMARY of the SPATIAL PROCESSING



Equiv: «Local mean»

Equiv: «Local centering»

Equiv: «Local Variance»

MOVING vs STATIC SQUARE

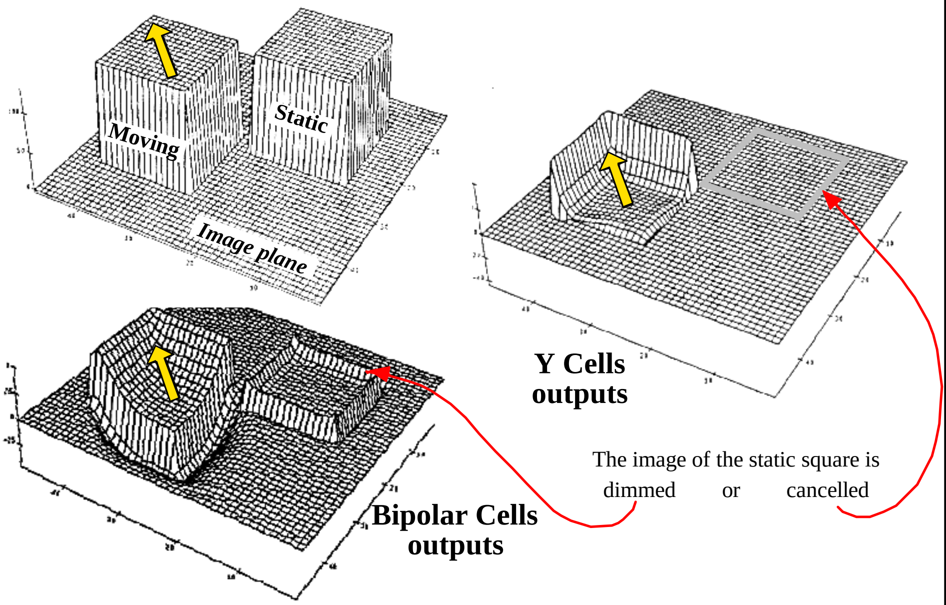


IMAGE PROCESSING IN THE RETINA

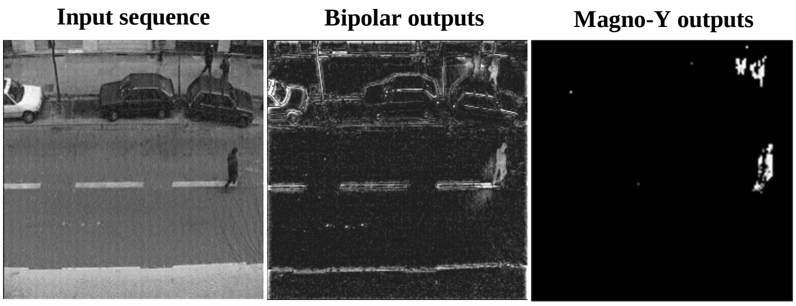
I- RETINAL CELLS & ELECTRICAL MODEL

II- SPATIO-TEMPORAL PROPERTIES

III- EXAMPLES

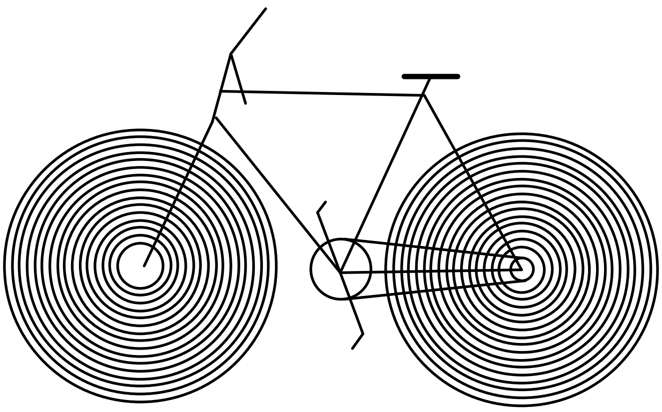
IV- SYNTHESIS and SILICON RETINAE

DYNAMICS (OPL & IPL)



W. Beaudot PhD Thesis (1999)

ILLUSION of RAYS



High- or Low-level effect ?

IMAGE PROCESSING IN THE RETINA

I- INTRODUCTION

II- RETINAL CELLS & ELECTRICAL MODEL

III- SPATIO-TEMPORAL PROPERTIES

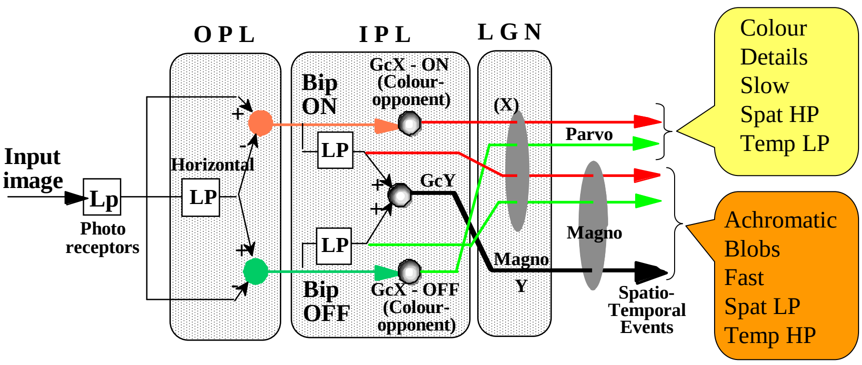
IV- EXAMPLES

V- CONCLUSION

Synthesis

Silicon retina

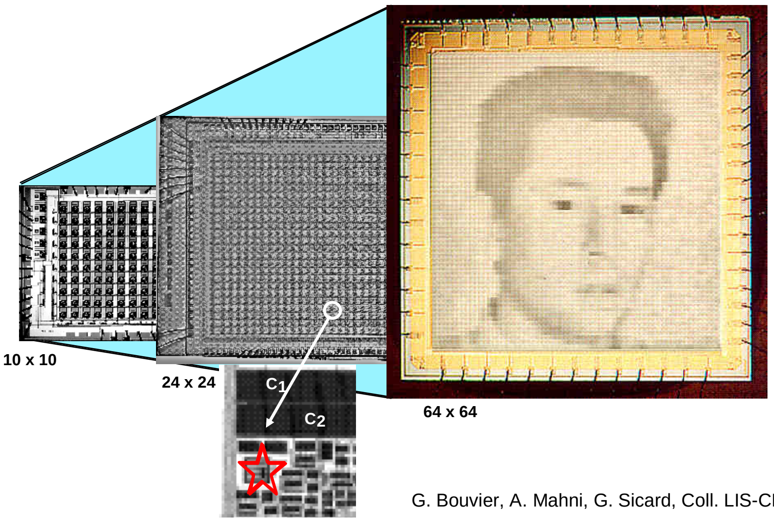
SUMMARY of RETINAL PROCESSING



Extracting the Direction of Light

TRICK : Cross-correlation
of ON and OFF images

LIS 0.5 µ C-MOS SILICON RETINAE



G. Bouvier, A. Mahni, G. Sicard, Coll. LIS-CNET

NEUROMORPHIC CIRCUITS
& MOTION ESTIMATION

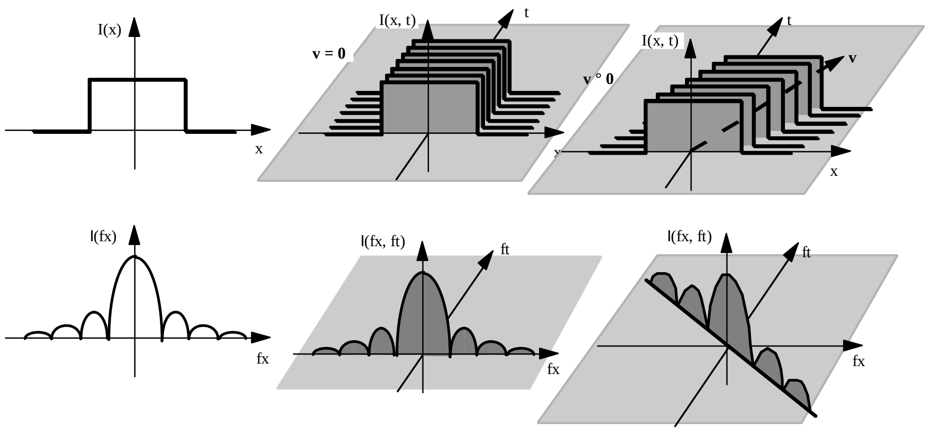
I- MODEL OF 1-D MOTION

II- WHICH CIRCUIT FOR MOTION ?

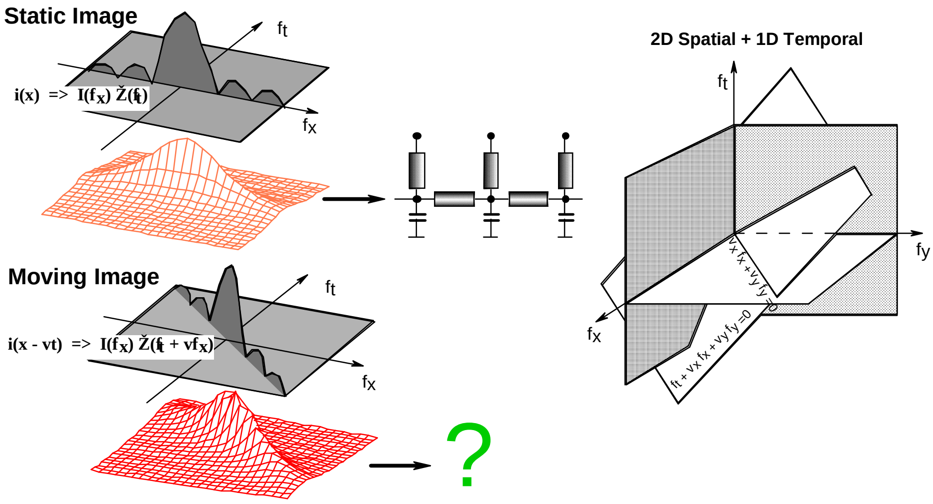
III- VELOCITY ESTIMATION

IV- EXAMPLES

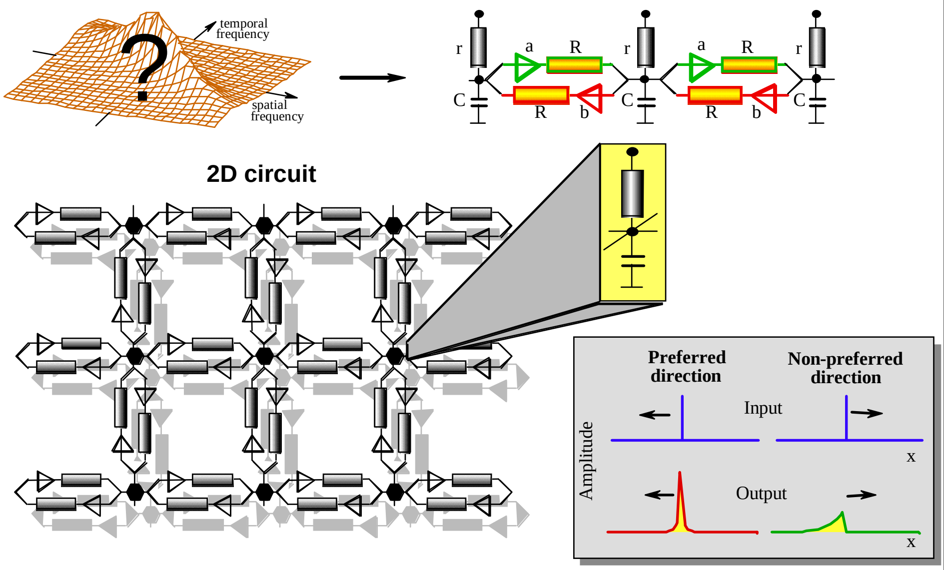
MODEL OF 1-D MOTION



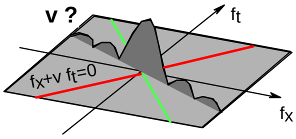
WHICH CIRCUIT FOR MOTION ?

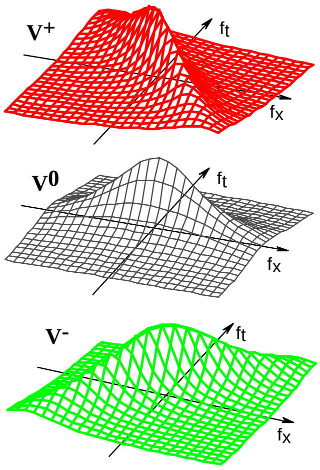


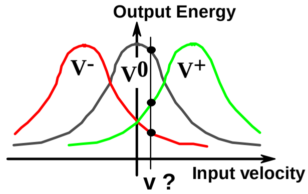
ASYMMETRIC CIRCUIT FOR MOTION



VELOCITY ESTIMATION





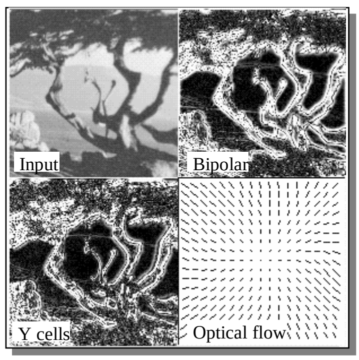


$$v = \frac{v_0}{2} \frac{\frac{1}{E_-^2} - \frac{1}{E_+^2}}{\frac{1}{E_-^2} + \frac{1}{E_+^2} - \frac{2}{E_0^2}}$$

A. Torralba PhD Thesis (19999)

EXAMPLES

Efficient numerical implementation: 1/25 second per (256x256) image

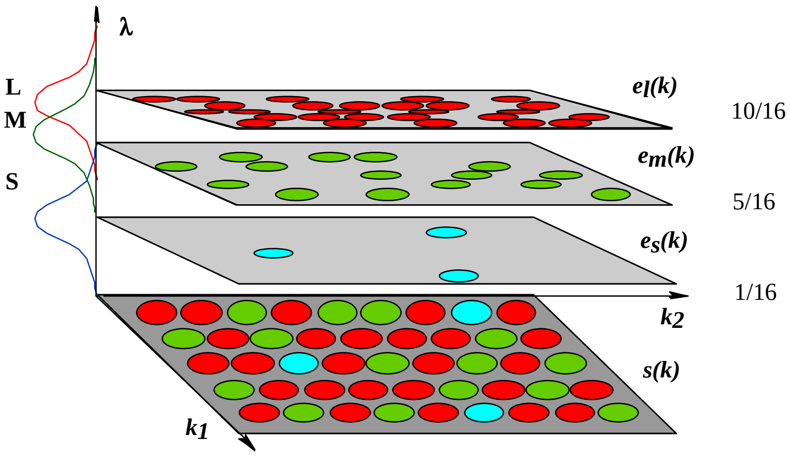


COLOR PERCEPTION

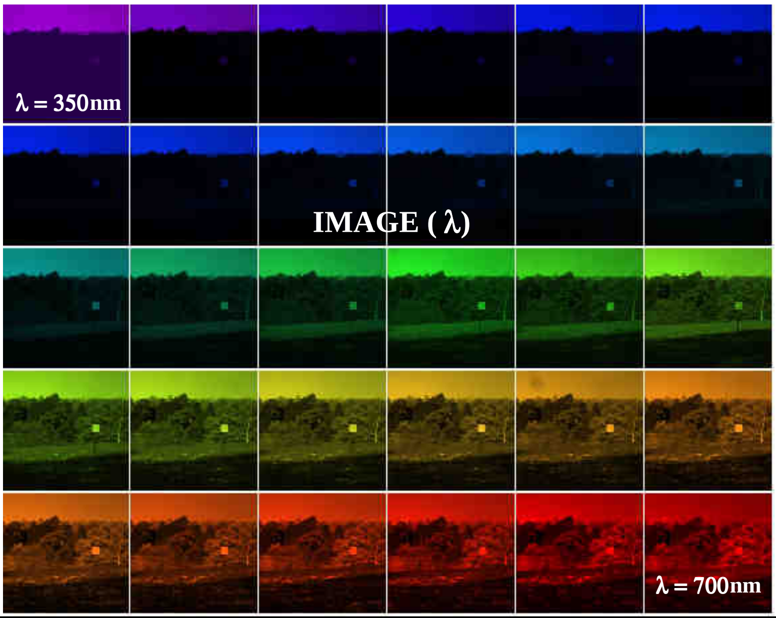
I- PHOTORECPTOR SAMPLING AND COLOR CODING

II- COLOR RECONSTRUCTION

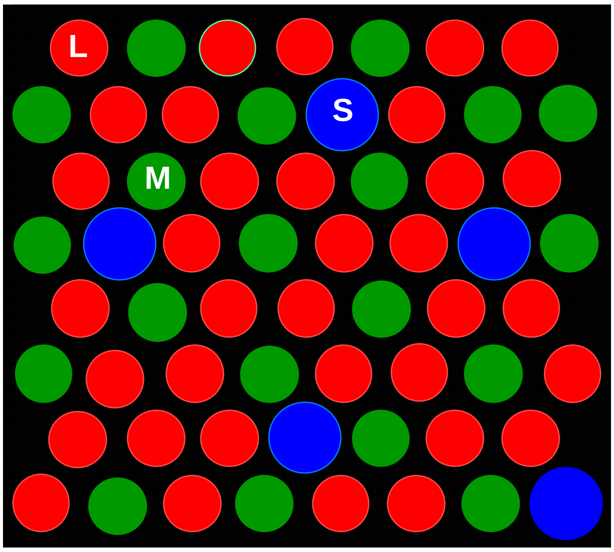
SPATIO-CHROMATIC SAMPLING



COLOUR SPECTRUM AND DETAILS



SAMPLING MODEL

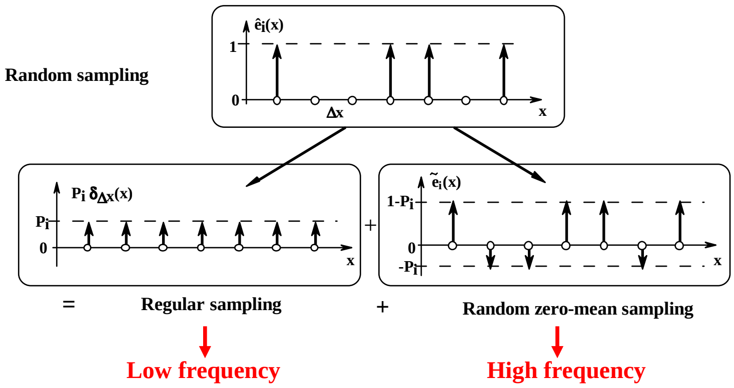


$P_L = {}^{10}/_{16}$

$P_M = 5 / 16$

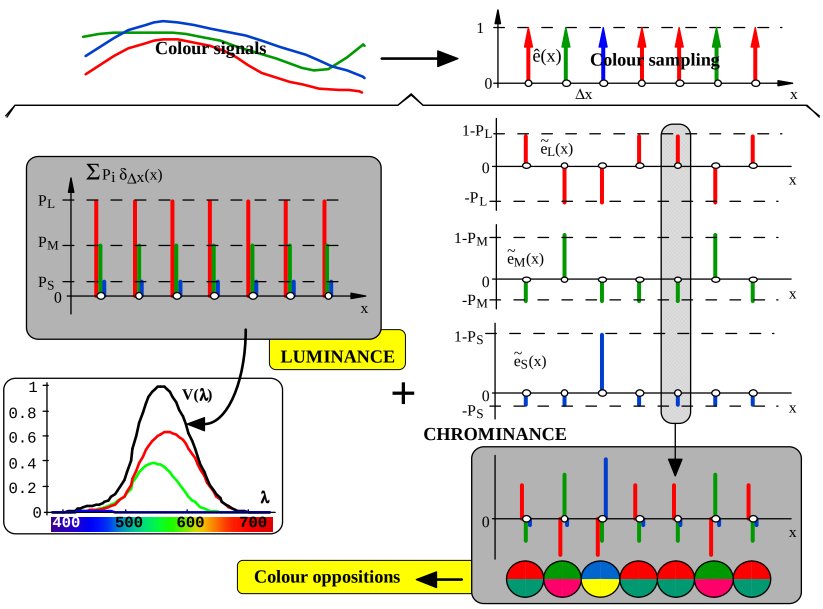
$P_S = {}^{1}/_{16}$

EQUIVALENT for RANDOM SAMPLING

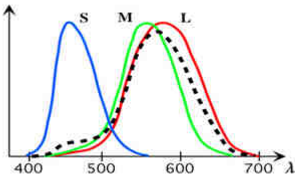


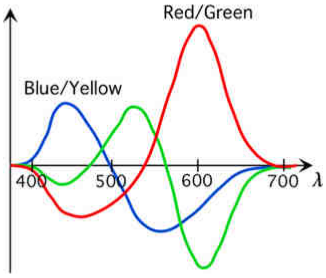
D. Alleysson PhD Thesis (1999)

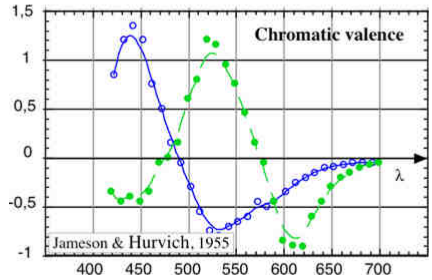
EQUIVALENT for RANDOM SAMPLING



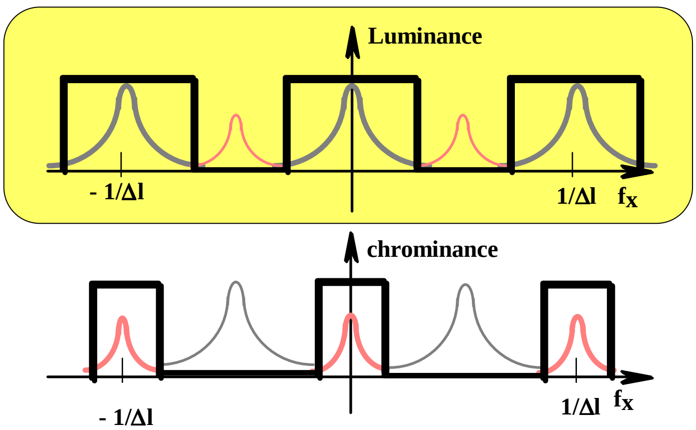
COLOUR OPPOSITION vs HUE CANCELLATION EXPERIMENTS







SPECTRA OF RANDOM SAMPLING



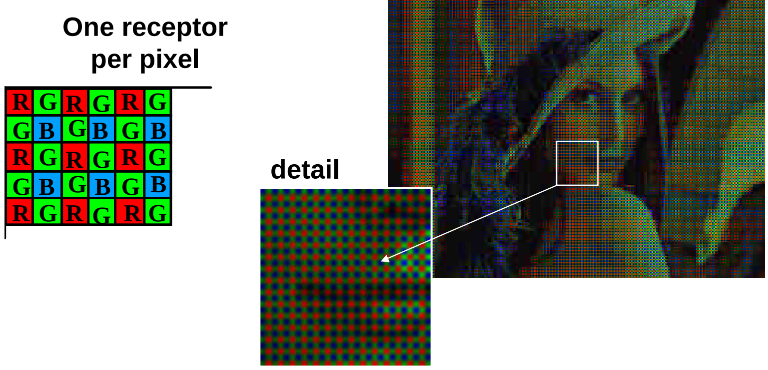
COLOR PERCEPTION

I- PHOTORECPTOR SAMPLING AND COLOR CODING

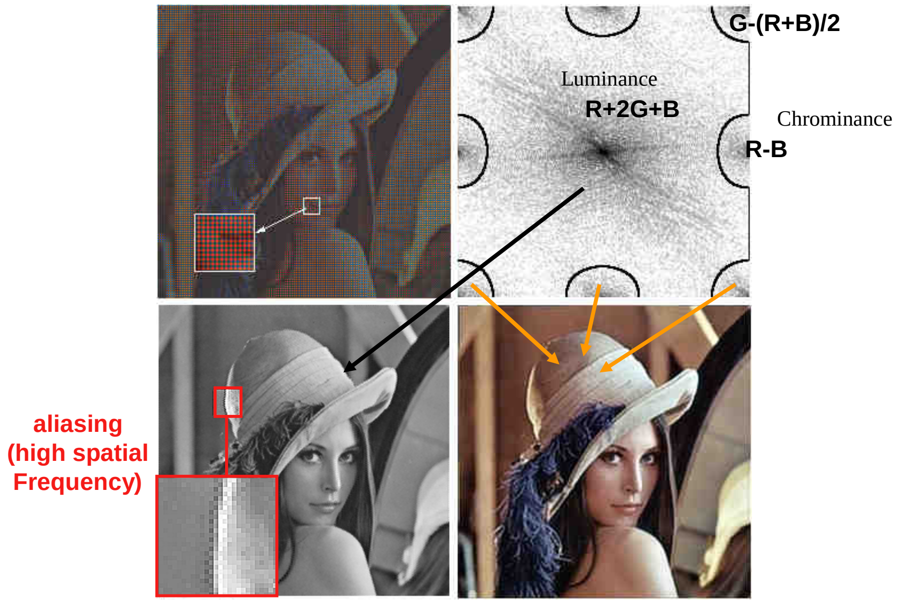
II- COLOR RECONSTRUCTION

Color: Application to CCD Cameras

Bayer Color Filter Array

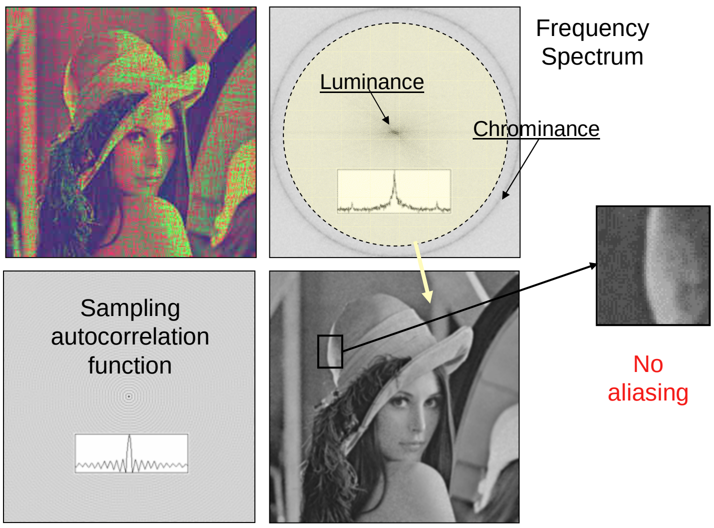


Color: Application to CCD Cameras



Random Sampling of Color

1. Principle



Random Sampling of Color

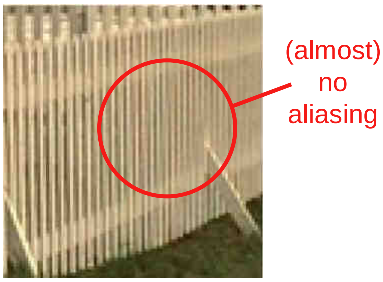
2. Color restitution (fovea: Red & Green cones only)

120 x 120 grid

Random Sampling of Color

3. Application: CCD with random RGB sampling

Regular sampling: Wiener filter restitution



Irregular sampling: Wiener filter restitution

B. Chaix : PhD Thesis (2007); Alleyson, Chaix, Hérault : Patent (2009)

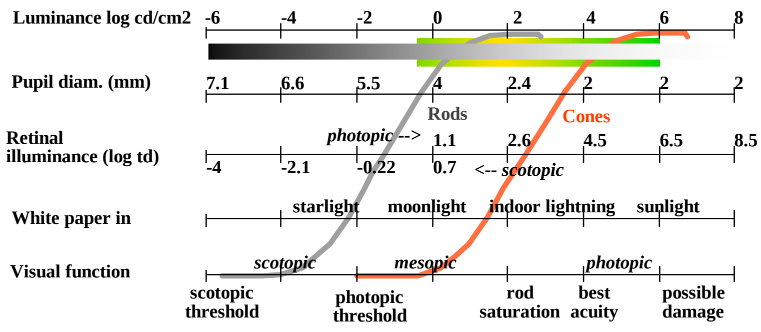
NON-LINEAR PROCESSING

I- PHOTORECEPTORS' NON-LINEARITY

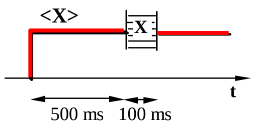
II- NON-LINEARITY IN IPL

III- NON-LINEARITY AND COLOUR

RANGE OF LIGHTENINGS



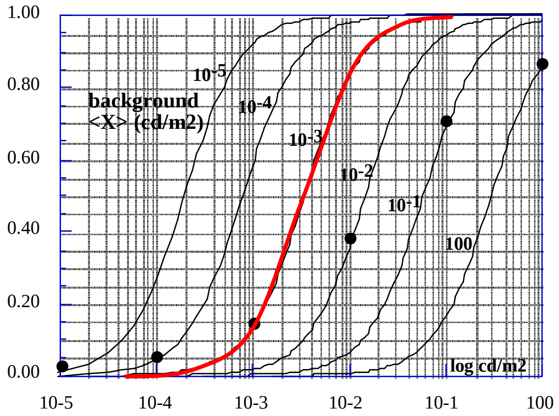
PHOTORECEPTOR SENSITIVITY



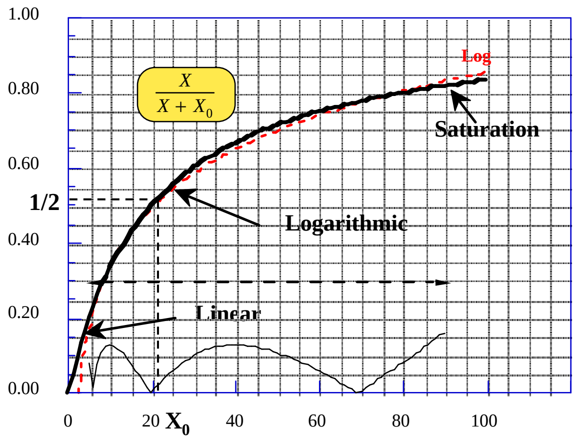
Michaelis-Menten:

$$x = \frac{X^n}{X^n + X_0^n}$$

$$X_0 \approx \langle X \rangle^{\frac{2}{3}}$$

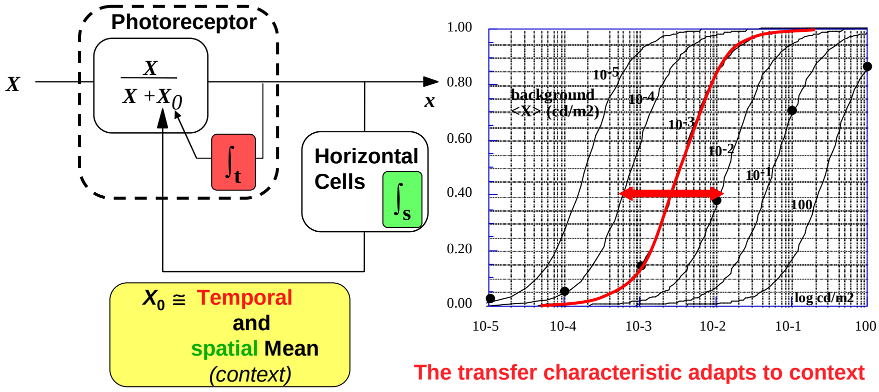


NAKA-RUSHTON LAW



Retina: Non-Linearities

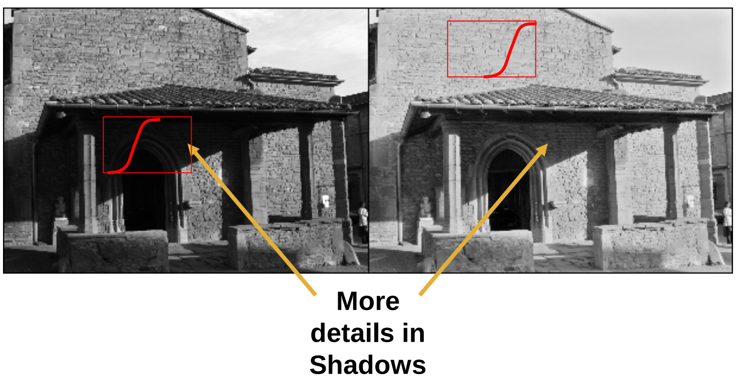
1. Photoreceptors' adaptive compression



The transfer characteristic adapts to context

Retina: Non-Linearities

1. Photoreceptors' adaptive compression: application



NON-LINEAR PROCESSING

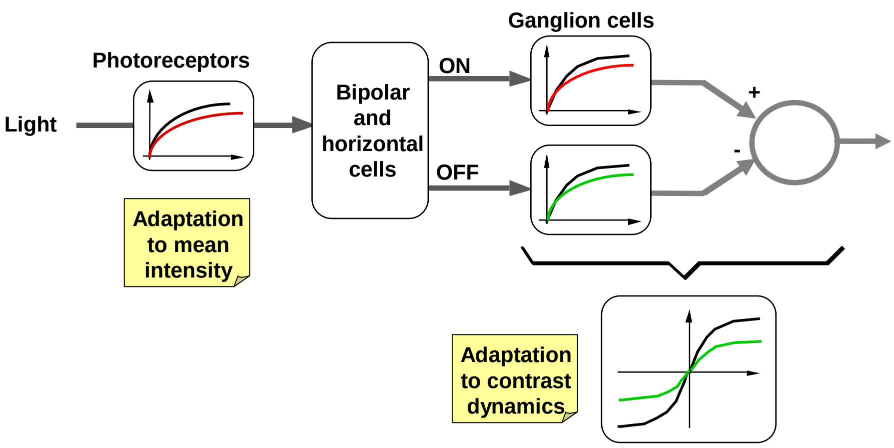
I- PHOTORECEPTORS' NON-LINEARITY

II- NON-LINEARITY IN IPL

III- NON-LINEARITY AND COLOUR

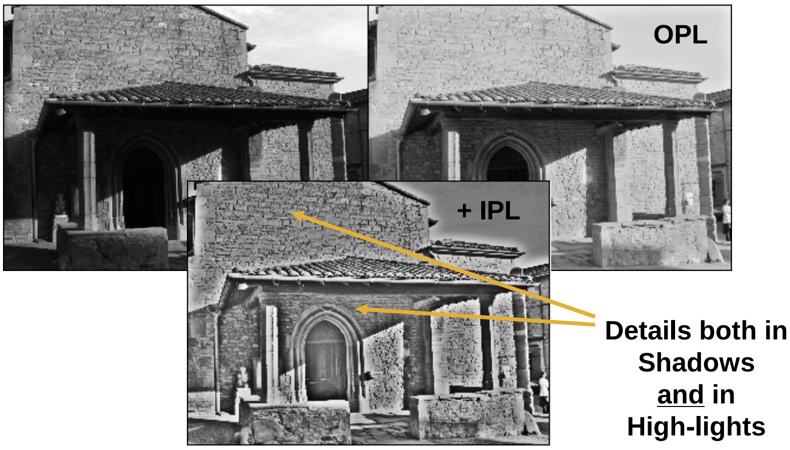
Retina: Non-Linearities

2. Adaptive Compression in IPL



Retina: Non-Linearities

2. Compressive Adaptation in IPL: application

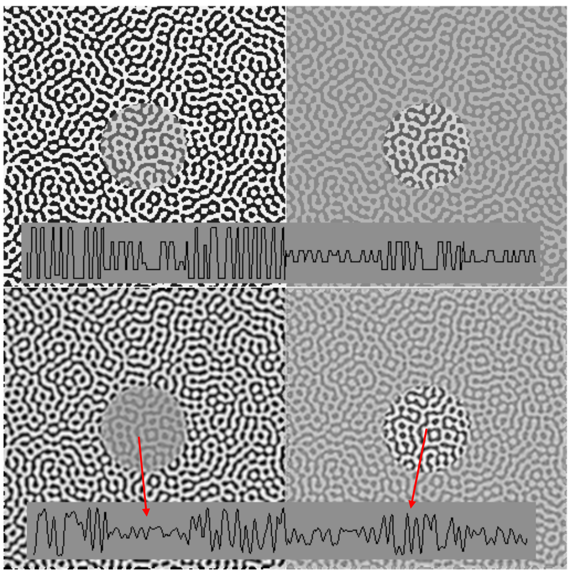


⇒ Elimination of Multiplicative noise $b_m(x)$ = extraction of true objects' reflectances

ADAPTATION IN IPL

Stimulus

Bipolar outputs



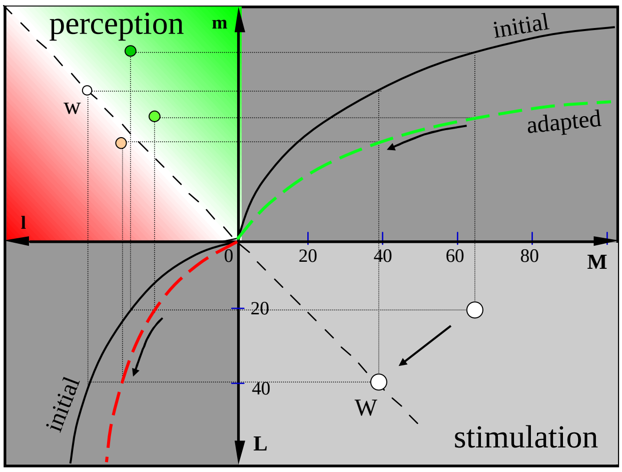
NON-LINEAR PROCESSING

I- PHOTORECEPTORS' NON-LINEARITY

II- NON-LINEARITY IN IPL

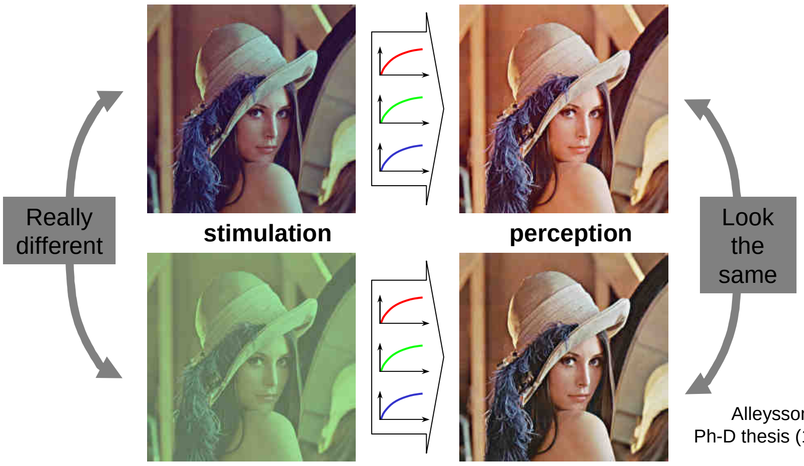
III- NON-LINEARITY AND COLOUR

ADAPTATION TO COLOR



Non-Linearity and Color

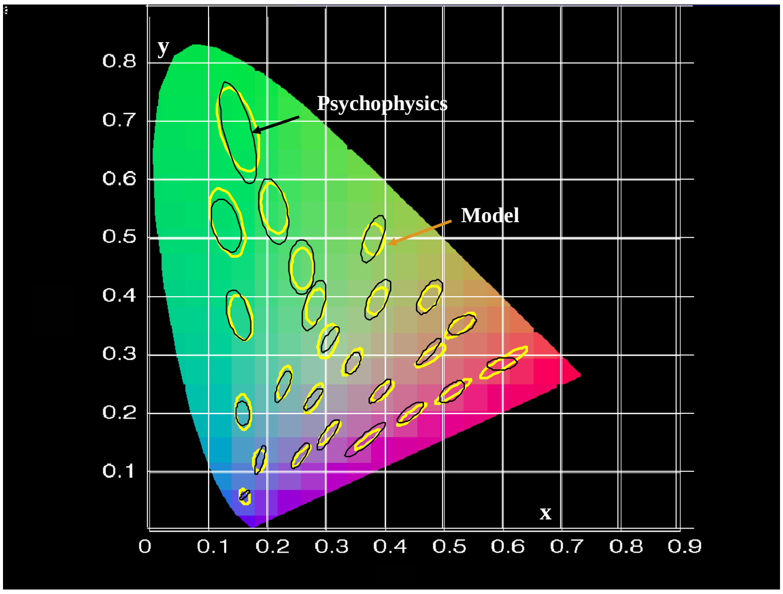
3. Color Constancy => at the photoreceptor level



Alleysson
Ph-D thesis (1999)

⇒ Elimination of Colored Multiplicative noise $b_m(x)$ = Color constancy of objects

MACADAM ELLIPSES



CORTICAL ANALYSIS OF IMAGES

I- IMAGE SPECTRA

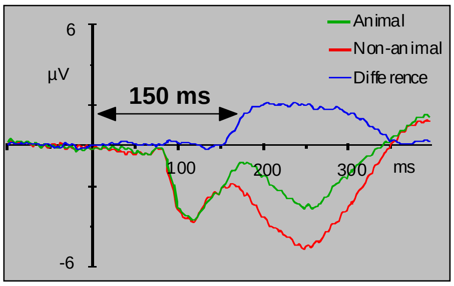
II- CORTICAL ANALYSIS

III- ADAPTATION TO TEMPORAL CONTEXT

150 ms TO RECOGNIZE !

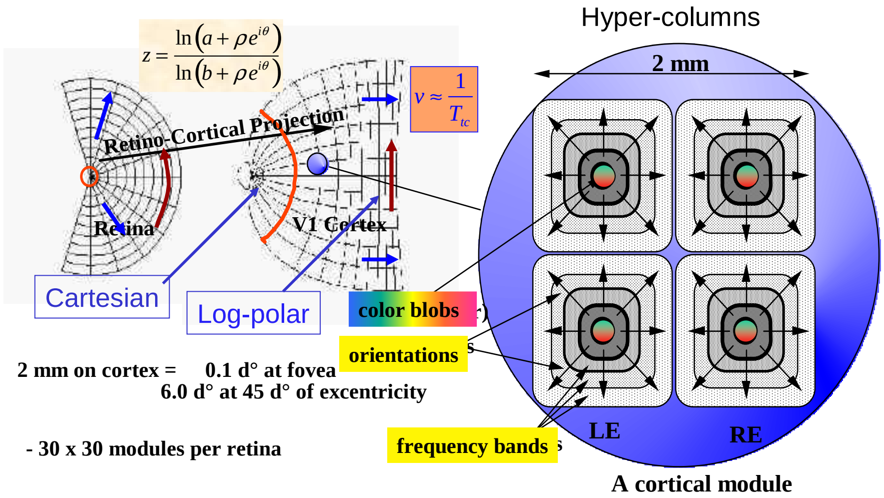
Image database

Visual evoqued potentials



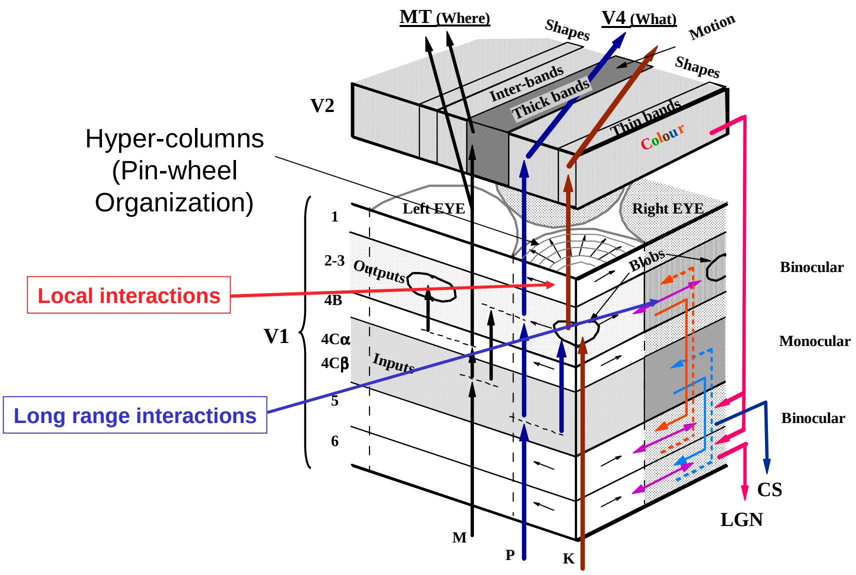
Test:
- An image is flashed during 30 ms.
- Tell whether it contains an animal or not

Retino-Cortical Projections



==> Local spectral analysis

Primary Visual Cortex: Organization



2nd Order Statistics of images and SPECTRAL ANALYSIS

==> An image will be considered as a particular realization of a stochastic process

Statistical Autocorrelation: $R_{12}(x_1, y_1, x_2, y_2) = \mathbf{E}\left[i_1(x_1, y_1) \cdot i_2(x_2, y_2)\right]$ Over **all** possible images, or Over **one** image category

H1: Stationary process $\boxed{R_{12}(x, y) = \mathbf{E}\left[i_1(x_1, y_1) \cdot i_2(x_1 - x, y_1 - y)\right]}$

H2: Ergodicity $\boxed{R(x, y) = \gamma(x, y) = \iint_{x_1, y_1} i(x_1, y_1) \cdot i(x_1 - x, y_1 - y)\, dx_1\, dy_1}$

Statistical Autocorrelation $\equiv$ Spatial Autocorrelation

Th. of Wiener-Kinchine: $F\{R(x, y)\} = S(f_x, f_y)$ ==> Energy Density Spectrum

Properties of the E.D.S. $\boxed{S(f_x, f_y) = F\{\gamma(x, y)\} = \left|F\{i(x, y)\}\right|^2}$

1- The amplitude spectrum of one image is similar to that of its category (comparisons are possible)

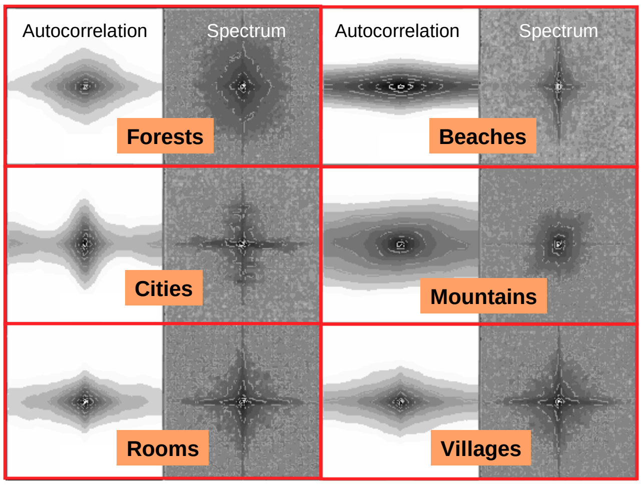
2- The amplitude spectrum of an image is independent of its position (translations are overcome)

Frequency Spectra of Scenes

1. Image data base

Frequency Spectra of Scenes

2. Mean Energy Spectra of image Categories



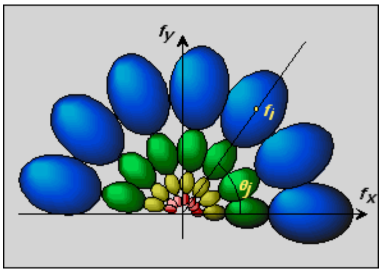
CORTICAL ANALYSIS OF IMAGES

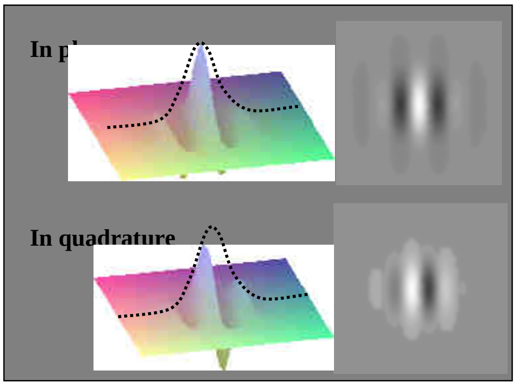
I- IMAGES SPECTRA

II- CORTICAL ANALYSIS

III- ADAPTATION TO TEMPORAL CONTEXT

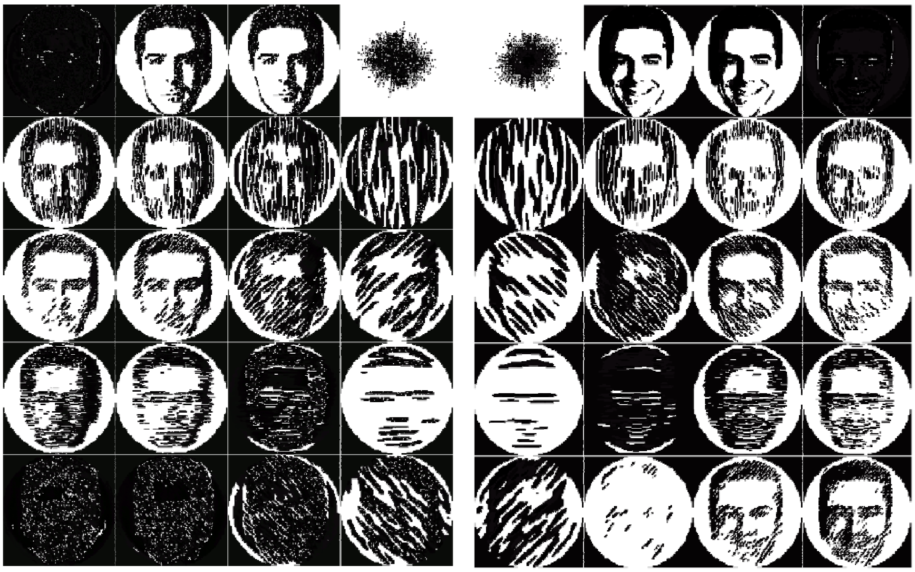
CORTICAL CELLS AND GABOR WAVELETS





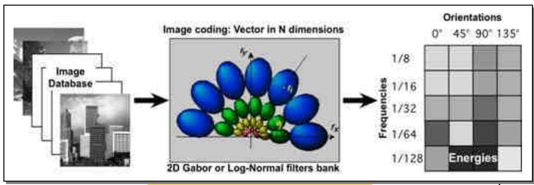
The image spectrum is decomposed into frequency bands and orientations

EXAMPLE OF IMAGE CODING



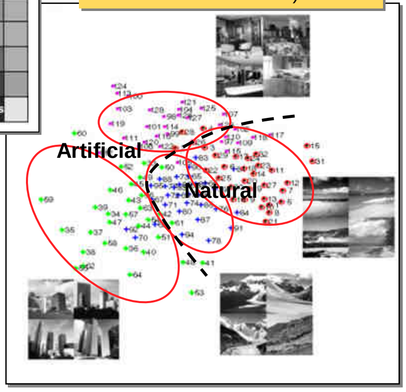
Cortical Model of Scene Analysis

Image Coding as a *high-dimensional vector*



Pin wheel models of V1 Complex cells

Scene Categorization by CCA: (Non-linear MDS with Metrics & Manifolds considerations)

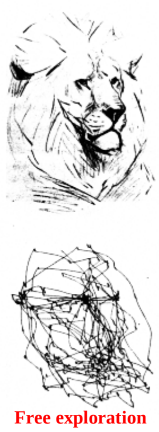


1. **the N-dimensional space is not fully spanned by the image vectors**
2. **the N dimensions are non-linearly correlated**

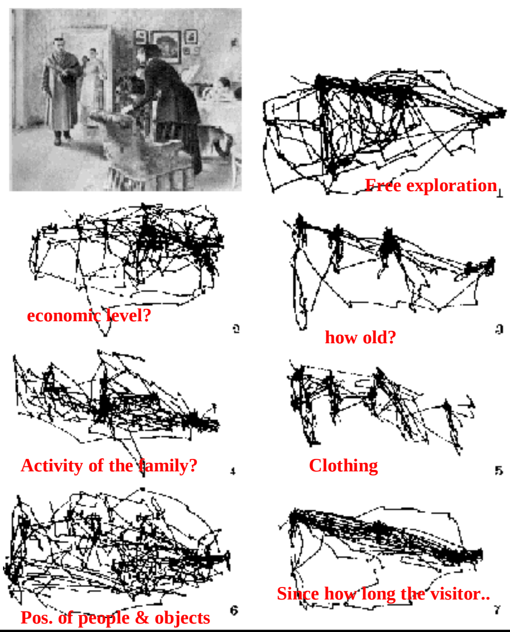
$\Rightarrow$ **use Curvilinear Component Analysis**

Ph-D theses: P. Demartines (1994), A. Oliva (1995), N. Guyader (2004)

OCULAR FIXATIONS

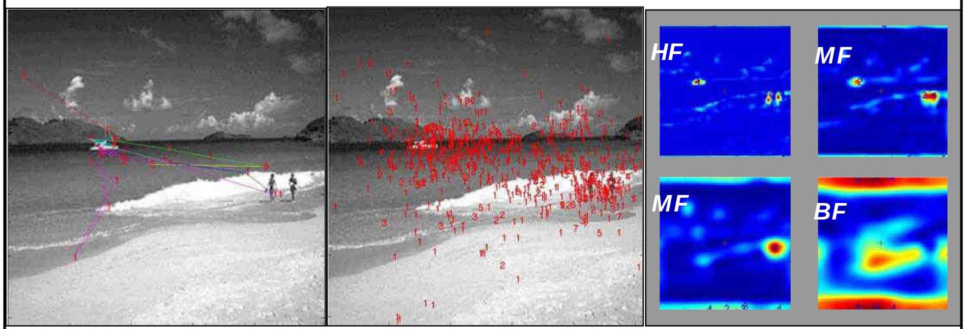


Eye movments



From 1 to 7, the subject is asked to answer a different question, Yarbus (1967).

OCULAR FIXATIONS AND LOCAL ENERGY



one subject **ensemble of subjects** **local energy**

Ph-D thesis: A. Chauvin (2004)

CORTICAL ANALYSIS OF IMAGES

I- IMAGES SPECTRA

II- CORTICAL ANALYSIS AND SCENE VARIABILITY

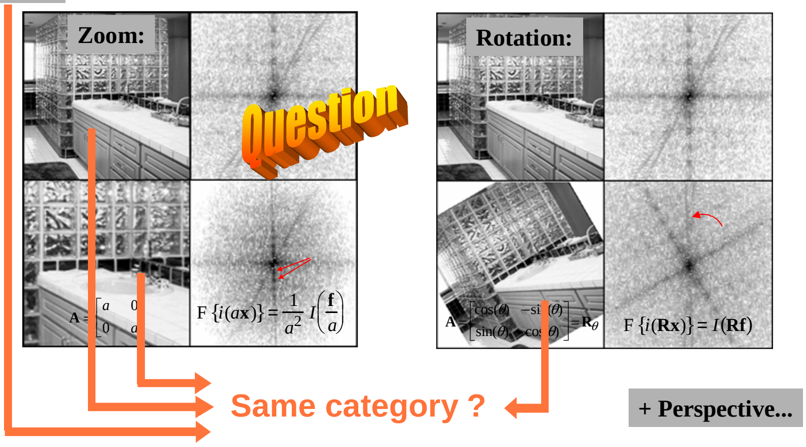
III- ADAPTATION TO TEMPORAL CONTEXT

Variability within a same Category

spectrum of a 2D image transformation:

$$\mathcal{F}\{i(\mathbf{Ax})\} = \iint_{\mathbf{x}} i(\mathbf{Ax}) \exp\left[-j2\pi(\mathbf{x}^{\mathrm{T}}\mathbf{f})\right] d\mathbf{x} \quad \xrightarrow{\mathbf{x}' = \mathbf{Ax}} \quad \mathcal{F}\{i(\mathbf{Ax})\} = \frac{1}{|\det(\mathbf{A})|} \iint_{\mathbf{x}} i(\mathbf{x}') \exp\left[-j2\pi(\mathbf{x}'^{\mathrm{T}} \underbrace{\mathbf{A}^{-\mathrm{T}}\mathbf{f}}_{\mathbf{f}'})\right] d\mathbf{x}'$$

Mirror: $i(x, y) \rightarrow i(-x, y) \quad \Rightarrow \quad I(f_x, f_y) \rightarrow I(-f_x, f_y)$



LOG-POLAR REPRESENTATION
of the FREQUENCY SPECTRUM

Use of the Log-Polar frequency spectrum:

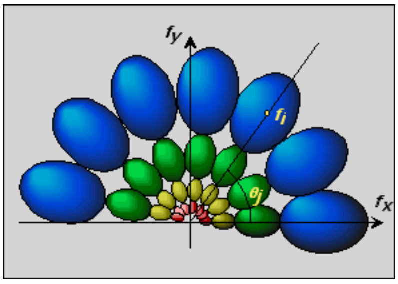
$$\mathcal{F}\{i(a\mathbf{R}_\theta \mathbf{x})\} = \underbrace{\boxed{\frac{1}{a^2} I\left(\frac{\mathbf{R}_\theta \mathbf{f}}{a}\right)}}_{\textit{Cartesian}} = \underbrace{\boxed{\frac{1}{a^2} I\left(e^{\nu - \ln(a)}, \varphi - \theta\right)}}_{\textit{Log-polar}} \quad \text{if} \quad \nu = \ln(f)$$

A zoom factor in the image space ==> a translation in log-frequency

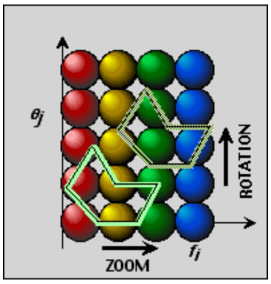
A rotation in the image space ==> a translation in angular frequency

LOG-POLAR REPRESENTATION

Cortical 2D Filter bank



Log-Polar Transform



$$f_{i+1} = k\, f_i, \quad \theta_{j+1} = \theta_j + \frac{k'}{\pi}$$

$$\nu_{i+1} = \nu_i + \ln(k), \quad \theta_{j+1} = \theta_j + \frac{k'}{\pi}$$

- Image zoom and rotation correspond both to log-polar spectral translations

- Image perspective will correspond to a shear in log-polar coordinates

Two kinds of LOG-POLAR Filter banks

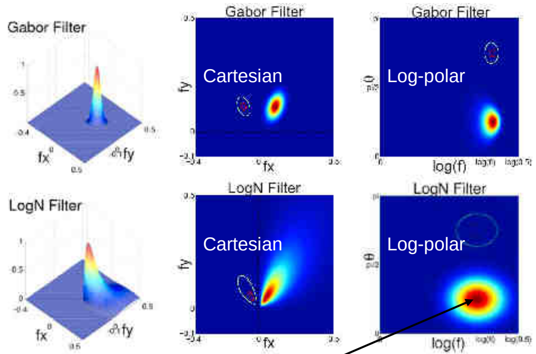
Gabor and Log-Normal filters

Gabor

$$G_i(f) = e^{-\frac{(f-f_i)^2}{2\sigma^2}}$$

Log-Normal

$$G_i(f) = \frac{1}{f} e^{-\frac{[\ln(f/f_i)]^2}{2\sigma'^2}}$$



Gaussian filters in log-polar domain => log-normal filters in cartesian domain

Two kinds of LOG-POLAR Filter banks

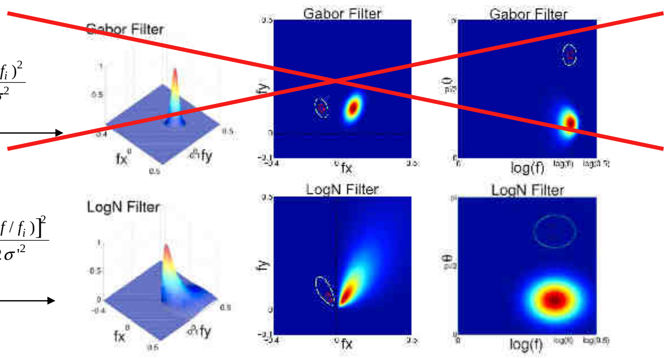
Gabor and Log-Normal filters

Gabor

$$G_i(f) = e^{-\frac{(f-f_i)^2}{2\sigma^2}}$$

Log-Normal

$$G_i(f) = \frac{1}{f} e^{-\frac{[\ln(f/f_i)]^2}{2\sigma'^2}}$$



=> **Log-normal filters are more suitable and more biologically plausible than classical Gabor filters**

Properties OF LOG-NORMAL FILTERS

2D filters w/ separable radial and angular variables: $|G_{ij}|^2 = |G_{i.}|^2 \, |G_{.j}|^2$

Mean energy in each frequency band : $C_{ij} = \iint_{f,\theta} S(f,\theta) \boxed{\frac{1}{f^2} e^{-\frac{\ln^2(f/f_i)}{2\sigma^2}}} \boxed{\left(\cos\left(\theta-\theta_j\right)\right)^{2n}} \, f df d\theta$

$G_{i.}$ are Log-normal filters: $\boxed{|G_{i.}| = \frac{1}{f} e^{-\frac{(\ln(f/f_i))^2}{4\sigma^2}}}$

$G_{.j}$ are π-*periodic Gaussian-like* filters: $\boxed{|G_{.j}| = \left(\cos(\theta-\theta_j)\right)^{2n}}$

with: $\nu = \ln(f/f_0)$

$$\boxed{C_{ij} == \iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-\frac{(\nu-\nu_i)^2}{2\sigma^2}} \left(\cos(\theta-\theta_j)\right)^{2n} d\nu d\theta}$$

For zoom *a* and rotation *φ*:

$i(a\,\mathbf{R}_{\varphi}\mathbf{x}) \Leftrightarrow \frac{1}{a^2} I(\mathbf{R}_{\varphi}\mathbf{f})$

$$\begin{cases} C_{ij} \approx \frac{1}{a^4} \iint_{\nu,\theta} \boxed{S(e^{\nu-\ln(a)},\theta-\varphi)}\, e^{-\frac{(\nu-\nu_i)^2}{2\sigma^2}}\, e^{-\frac{(\theta-\theta_j)^2}{2\sigma'^2}} \, d\nu d\theta \\ C_{ij} \approx \frac{1}{a^4} \iint_{\nu,\theta} S(e^{\nu},\theta) \boxed{e^{-\frac{(\nu-\nu_i+\ln(a))^2}{2\sigma^2}}\, e^{-\frac{(\theta-\theta_j+\varphi)^2}{2\sigma'^2}}} \, d\nu d\theta \end{cases}$$

==> If the log-polar spectrum is *regularly sampled* by the filter bank in ν_i and θ_j, and if the number of filters is *odd*, non-integer translations in ν_i and θ_j can be processed

OVERCOMING CONVOLUTIVE NOISE

Image spectrum: $S_o(f,\theta) = S(f,\theta)\cdot S_{bc}(f,\theta)$

Convolutive noise spectrum

Original image spectrum

Filters' outputs:

$$C_{ij} = \iint_{f\theta} S(f,\theta)\cdot S_{bc}(f)\left|G_{ij}(f,\theta)\right|^2 df\,d\theta \approx S_{bc,i}\cdot\iint_{f\theta} S(f,\theta)\left|G_{ij}(f,\theta)\right|^2 df\,d\theta$$

(because $S_{bc,i}$ is smooth)

By divisive normalization ***(an existing local interaction in V1)*****:**

$$\tilde{C}_{ij} = \frac{C_{ij}}{\sum_j C_{ij}} \approx \frac{\cancel{S_{bc,i}}\cdot\iint_{f\theta} S(f,\theta)\left|G_{ij}(f,\theta)\right|^2 df\,d\theta}{\cancel{S_{bc,i}}\cdot\iint_{f\theta} S(f,\theta)\sum_j\left|G_{ij}(f,\theta)\right|^2 df\,d\theta}$$

==> Overcomes convolutive noise

CORTICAL ANALYSIS OF IMAGES

I- IMAGES SPECTRA

II- CORTICAL ANALYSIS AND PERSPECTIVE ESTIMATION

III- ADAPTATION TO TEMPORAL CONTEXT

LOG-NORMAL FILTERS and PERSPECTIVE

Perspective estimation through local frequency gradients

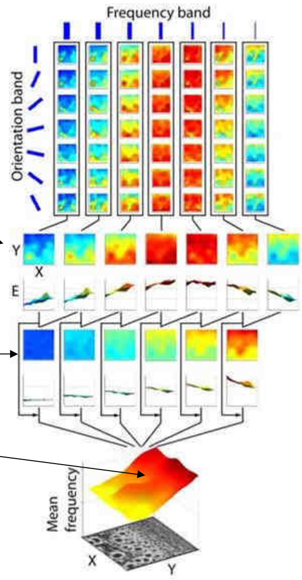
Means across all orientations

narrow-band local frequency

$$\frac{C_{i+1}(x,y)}{C_i(x,y)} = \frac{1}{\sqrt{f_i f_{i+1}}} \langle f \rangle_i (x,y)$$

wide-band local frequency

(similar to long-range inretactions in V1)

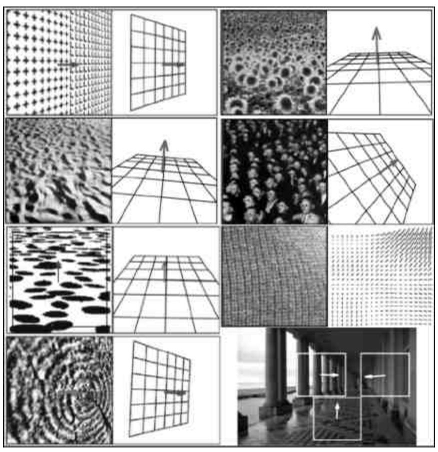


C. Massot Ph-D thesis (2005)

LOG-NORMAL FILTERS and PERSPECTIVE

Perspective estimation through local frequency gradient

Estimates global or local surface's normal with regular textures and with *irregular* textures as well



C. Massot Ph-D thesis (2005)

CORTICAL ANALYSIS OF IMAGES

I- IMAGES SPECTRA

II- CORTICAL ANALYSIS, ZOOM AND ROTATION

III- ADAPTATION TO TEMPORAL CONTEXT

OVERCOMING ZOOM and ROTATION

Fourier transform of the C_{ij} data

$$\mathrm{F}\{C_{ij}\} = \frac{1}{a^4} \iint_{\nu_i,\theta_i} \Bigg\{ \underbrace{\iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-\frac{(\nu-\nu_i+\ln(a))^2}{2\sigma^2}}\, e^{-\frac{(\theta-\theta_j+\varphi)^2}{2\sigma'^2}}\, d\nu\, d\theta}_{\textit{Image spectrum}} \Bigg\} e^{-j2\pi(\xi\nu_i+\eta\theta_j)}\, d\nu_j\, d\theta_j$$

$$\mathrm{F}\{C_{ij}\}(\xi,\eta) = \frac{1}{a^4} \underbrace{e^{-(k_1\xi^2+k_2\eta^2)}}_{\textit{Gaussian envelope}} \underbrace{e^{-j2\pi(\xi\ln(a)+\eta\varphi)}}_{\textit{Zoom (a) \& Rotation }(\varphi)} \underbrace{\iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-j2\pi(\xi\nu+\eta\theta)}\, d\nu\, d\theta}_{\textit{Relative to the scene}}$$

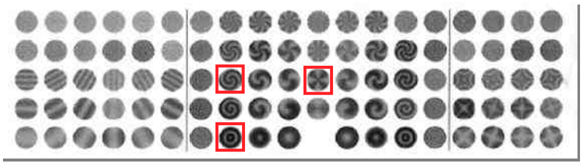
$\left|\mathrm{F}\{C_{ij}\}(\xi,\eta)\right|$ *independent of* (a,φ)

==> Overcomes zoom and rotation variability

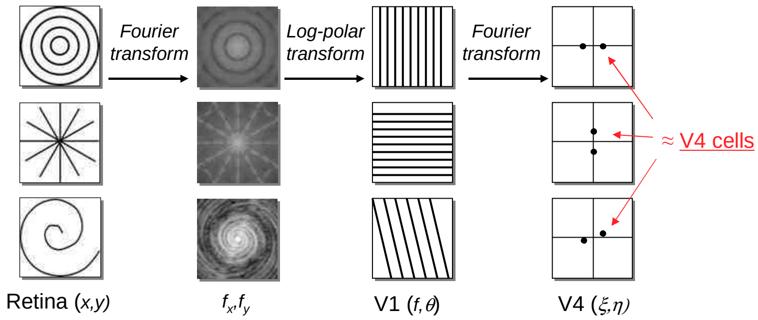
an operation that may exist in visual area V4...

Model of visual area V4

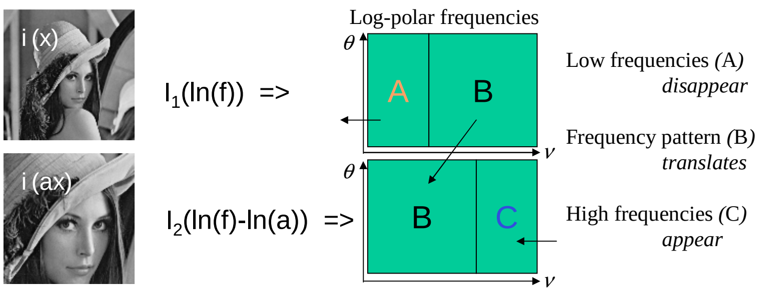
Biology:
Stimuli and
Best responses
in area V4 cells



Model



Problem with different Zoom Ratios



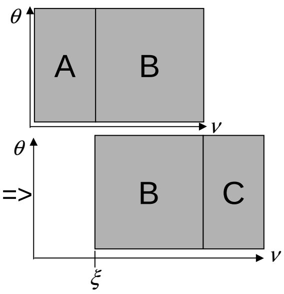
Classical distance: $d_{1,2}^2 = \|I_1\|^2 + \|I_2\|^2 - \gamma_{1,2}(0) - \gamma_{2,1}(0) = \|A\|^2 + 2\|B\|^2 + \|C\|^2 - \gamma_{1,2}(0) - \gamma_{2,1}(0)$

should disappear

Problem with different Zoom Ratios

$I_1(\ln(f))$ =>

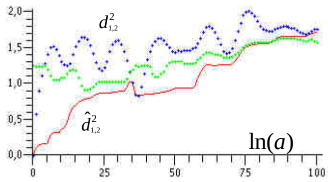
$I_2(\ln(f)-\ln(a))$ =>



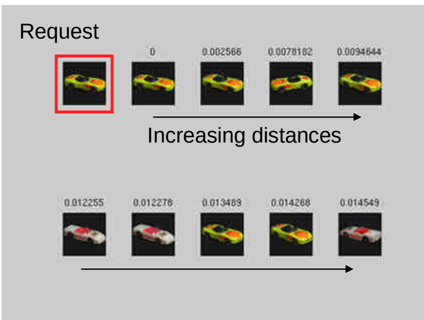
search for maximum Intercorrelation between I_1 and I_2

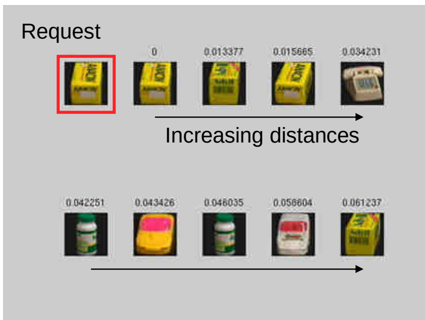
Better distance would be: $\hat{d}^2_{1,2} = \|A\|^2 + \|C\|^2$

That is: $\hat{d}^2_{1,2} = \underbrace{\|I_1\|^2 + \|I_2\|^2}_{\|A\|^2 + 2\|B\|^2 + \|C\|^2} - 2\,\underbrace{Max\left(\gamma_{1,2}(\xi)\right)}_{2\|B\|^2}$



EXAMPLES





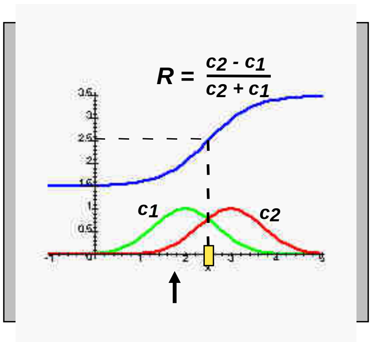
CORTICAL ANALYSIS OF IMAGES

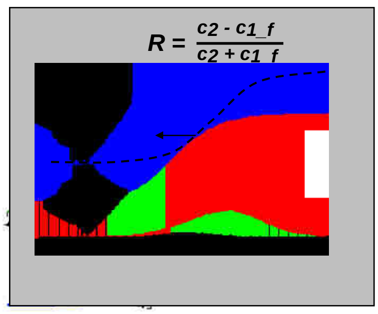
I- IMAGES SPECTRA

II- CORTICAL ANALYSIS

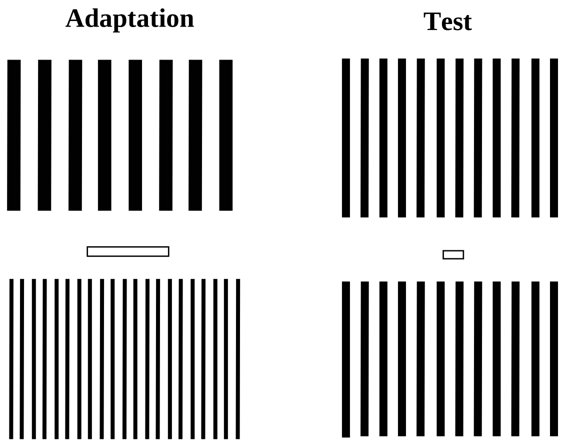
III- ADAPTATION TO TEMPORAL CONTEXT

COMPLEX CELL ADAPTATION

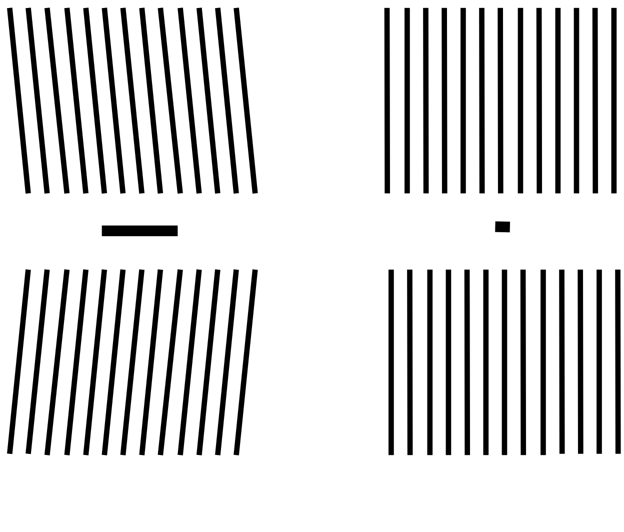




SPATIAL FREQUENCY ADAPTATION



ADAPTATION TO ORIENTATION



Natural vs Computational Vision

Retina

- Spatio-temporal high-pass => whitening (compensates for 1/f spectrum)
- Color coding => efficiency and economy
- Adaptive compression => intensity and contrast equalization
  color constancy
- Random sampling => anti-aliasing
- Neuromorphic circuits => efficient for prostheses & sensory substitution
  efficient algorithm for dense optic flow

Cortex

- Log-polar projection => efficient for ego-motion
- Energy spectrum => independency wrt translation
  scene and texture categorization
- Log-polar frequ. repres. => easy processing of zoom & rotation
  gradients (perspective)
  time derivative (time-to-contact)
- lateral interactions => figure/ground separation, attention

+ other properties:

- Random sampling => ?
- Adaptive compression => ?
- Top-down interactions => ?

$\}$ **next…**

Jean PETITOT

EHESS / X

http://www.crea.polytechnique.fr/JeanPetitot/home.html

"Modèles Neurogéométriques de Vision"

On proposera un modèle géométrique de l'architecture fonctionnelle du cortex visuel primaire (aire V1) et on explicitera les algorithmes géométriques que cette dernière implémente, autrement dit la "neurogéométrie" immanente à la perception visuelle.

1) Le filtrage du signal optique par les neurones visuels s'apparente à une analyse en ondelettes. La structure de contact de l'espace des 1-jets des courbes du plan (ici le plan rétinien) se trouve implémentée par l'architecture fonctionnelle.

2) L'intégration des contours à partir de données sensorielles éventuellement très lacunaires sont modélisables en termes de la géométrie sous-riemannienne associée à cette structure de contact.

Référence :

J. Petitot, « Neurogéomètrie de la vision », Ellipse - Les Editions de l'Ecole Polytechnique, 420 pages, 2008.

ERMITES 2010:
Vision et Cognition
28-30 septembre 2010

Modèles neurogéométriques de vision

Jean Petitot

CAMS, EHESS &
CREA, Ecole Polytechnique, Paris

Introduction

- What I call neurogeometry concerns the neural implementation of geometric structures of visual perception.
- It concerns perceptive geometry "from within" (in the sense of Gromov) and not 3D Euclidean geometry of the outside world .
- The general problem is to understand how the visual system can be a neural geometric engine.

- Many non trivial mathematical structures have been introduced recently to explain natural early vision.
- Contact, symplectic and sub-Riemannian geometry arise naturally in modeling V1 functional architecture.
- Sub-Riemannian geometry provides the simplest model of the horizontal cortico-cortical connections in V1.

- In relation with wavelet analysis this leads naturally to noncommutative harmonic analysis on Heisenberg type groups.

2

Limitations

- We focus on V1, but there are of course many top-down feedbacks from other areas to V1.
- Neural implementation varies with species (rat, ferret, tree shrew, cat, macaque, man, etc.). The same functional architecture can be implemented in different ways.

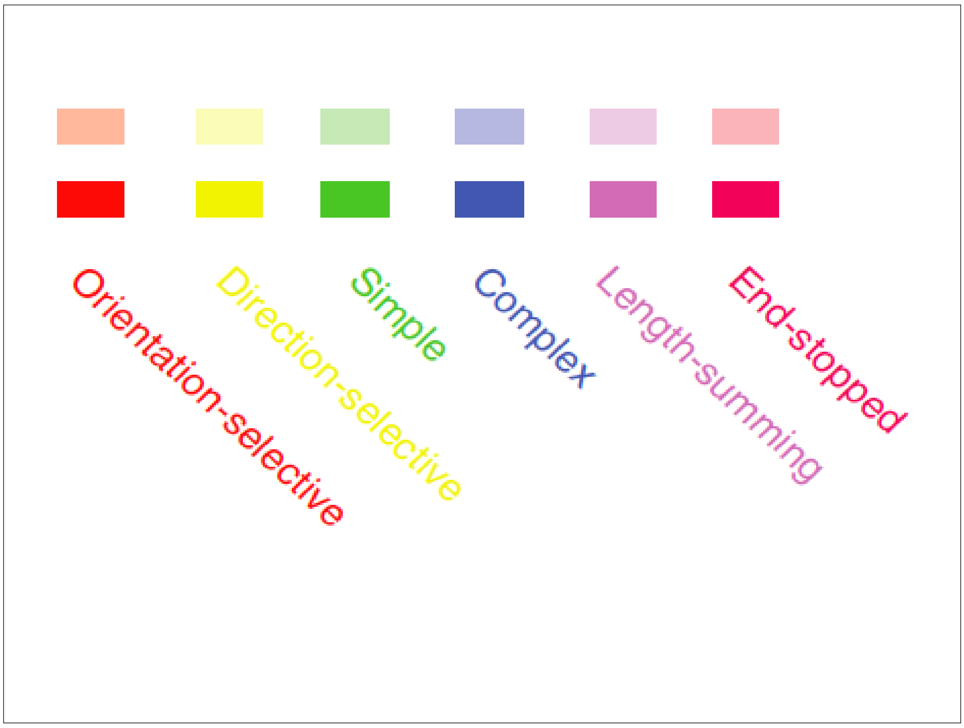
- Stephen van Hooser on “Similarity and diversity" of V1 in mammals (comparative study).
- The gross laminar interconnections and the major functional responses are nearly invariant: 6 layers, LGN projecting mainly on the granular 4th layer.
- Three principal classes of LGN cells: parvocellular (P), magnocellular (M), koniocellular (K), etc.

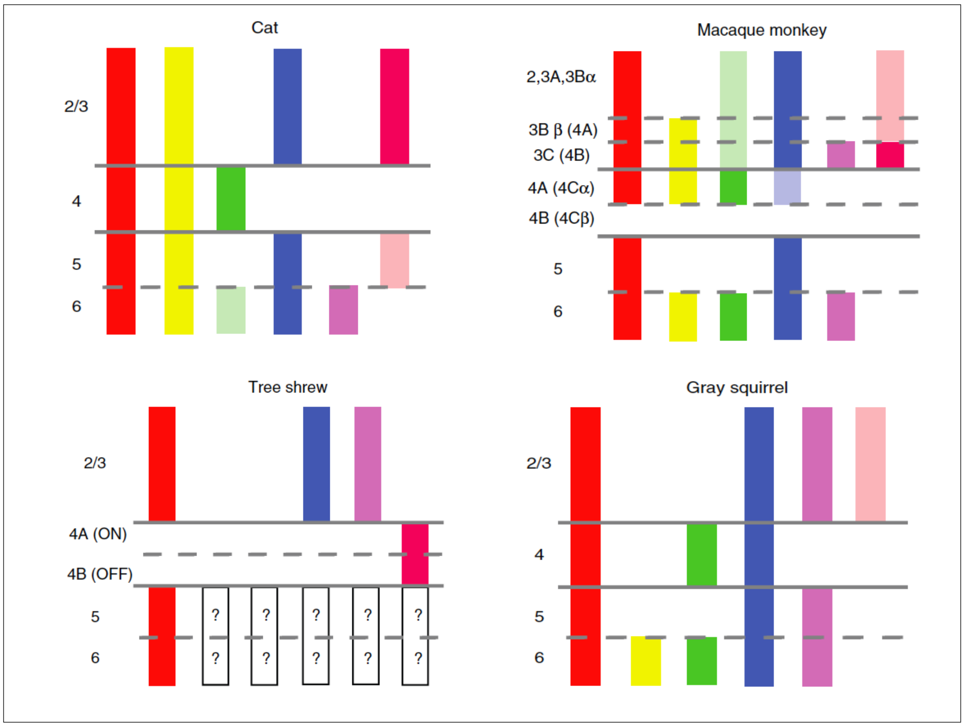
- But the fine laminar structures are quite different.
- Tree shrew (Tupaya), Cat, Macaque have orientation maps with orientation hypercolumns and a functional “horizontal” architecture connecting neurons of similar orientation.
- Rat and Gray squirrel have not.

Figure. Orientation simple cells (red) are absent in macaque 4B and tree shrew layer 4.

[[Direction selectivity dominates in the cat but is only common in specific layers of macaque and squirrel.

End-stopped VS lengthsumming cells : they decrease VS increase their responses as bars or gratings length increases.]]


Orientation-selective
Direction-selective
Simple
Complex
Length-summing
End-stopped

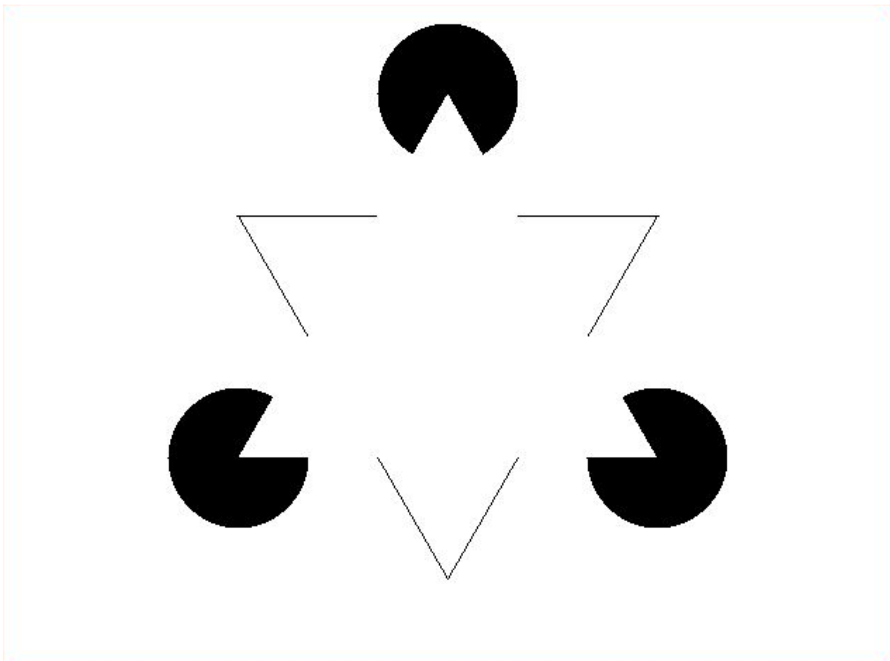

Cat
2/3
4
5
6
Macaque monkey
2,3A,3Bα
3B β (4A)
3C (4B)
4A (4Cα)
4B (4Cβ)
5
6
Tree shrew
2/3
4A (ON)
4B (OFF)
5
6
?
Gray squirrel
2/3
4
5
6

5

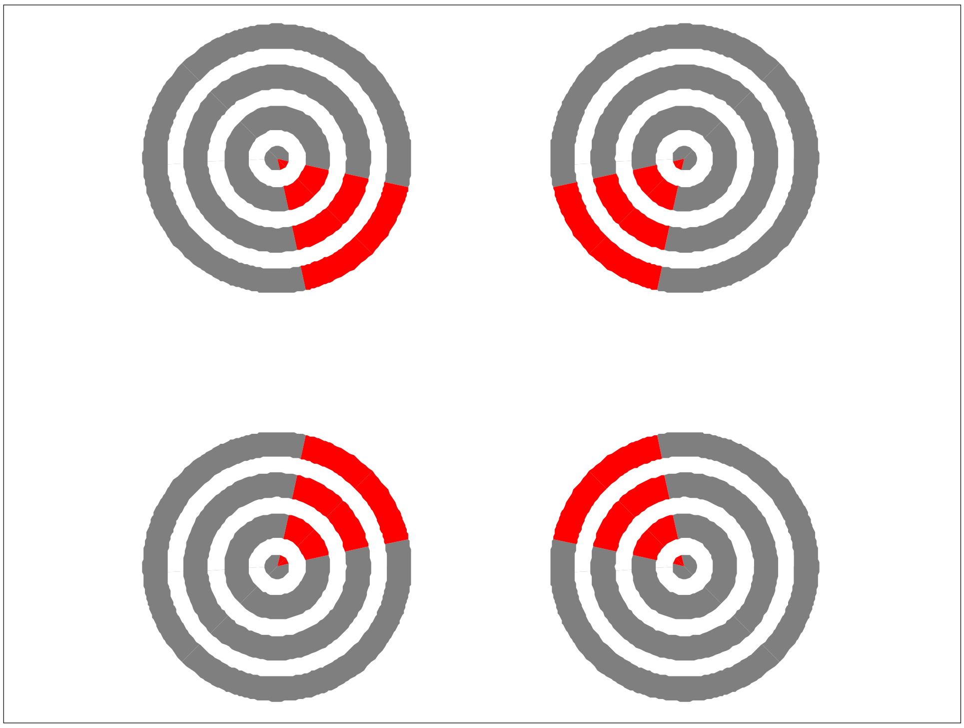
- *Another limitation*. Neural coding is a statistical population coding and, for each elementary computation, a lot of neurons are involved.
- We will not take into account explicitely this redundancy which leads to stochastic models.

A typical example : Kanizsa illusory contours

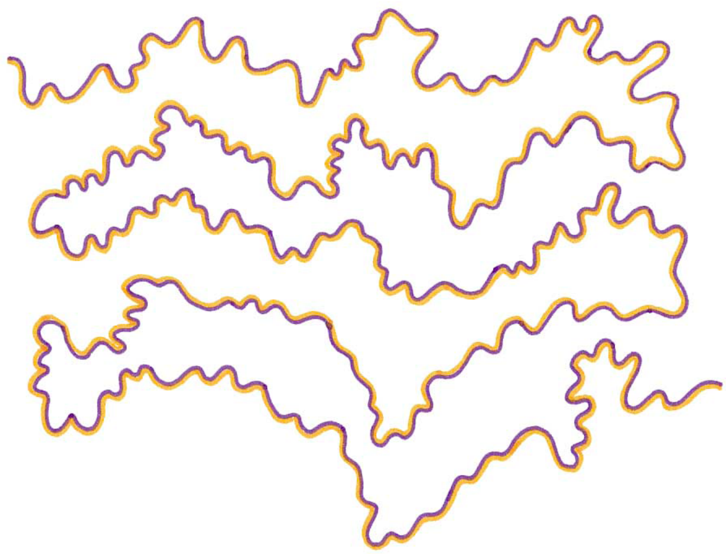
- A typical example of the problems of neurogeometry is given by well known Gestalt phenomena such as Kanizsa illusory contours.
- The visual system (V1 with some feedback from V2) constructs very long range and crisp virtual contours.



- They can even be *curved*.
- With the neon effect (watercolor illusion), virtual contours are boundaries for the diffusion of color inside them.



- B. Pinna, G. Brelstaff, L. Spillmann (*Vision Research*, 41, 2001)): watercolor illusion.

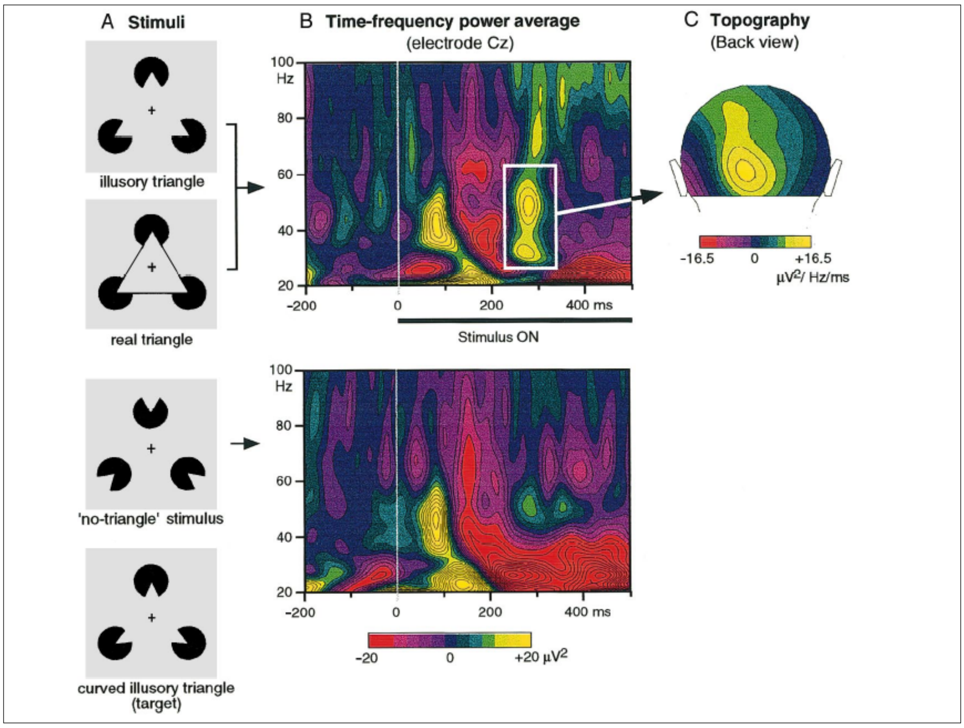


8

- Kanizsa subjective contours manifest a deep neurophysiological phenomenon.
- Here is a result of Catherine Tallon-Baudry in « Oscillatory gamma activity in humans and its role in object representation » (*Trends in Cognitive Science*, 3, 4, 1999).
- Subjects are presented with coherent stimuli (illusory and real triangles) « leading to a coherent percept through a bottom-up feature binding process ».

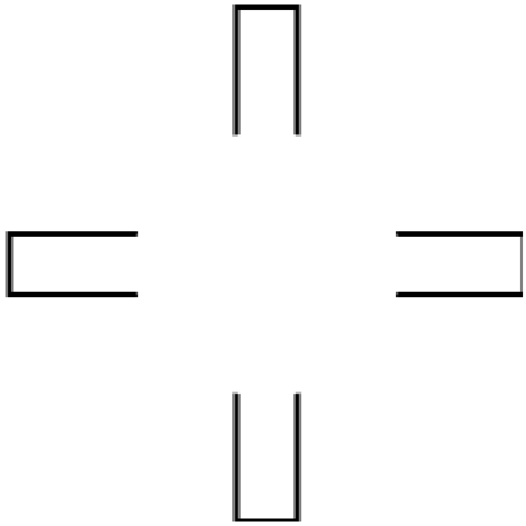
- « Time–frequency power of the EEG at electrode Cz (overall average of 8 subjects), in response to the illusory triangle (top) and to the no-triangle stimulus (bottom ».
- « Two successive bursts of oscillatory activities were observed.
  - A first burst at about 100 ms and 40 Hz. It showed no difference between stimulus types.
  - A second burst around 280 ms and 30-60 Hz. It is most prominent in response to coherent stimuli. »

9

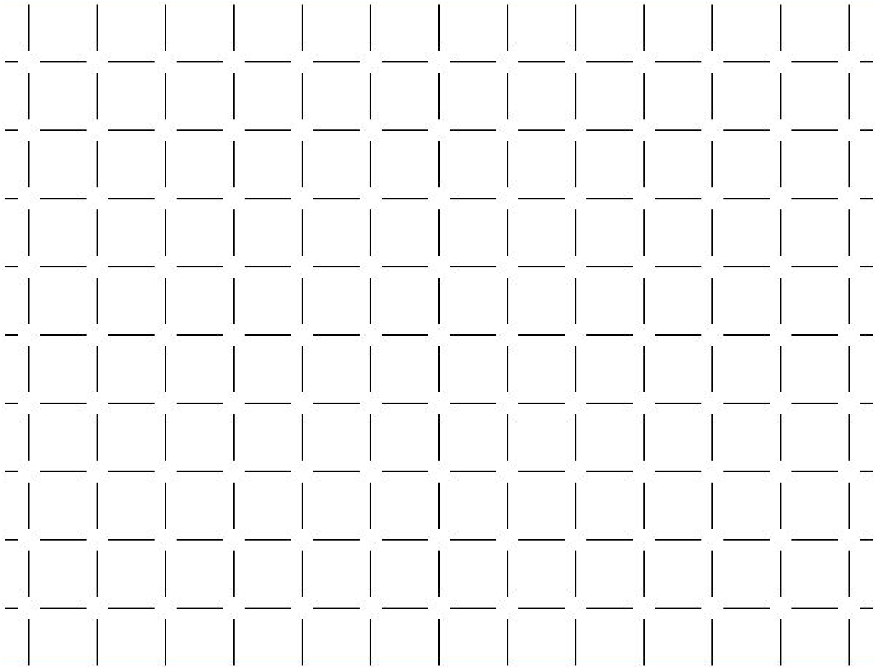


- Many phenomena are striking. E.g. the change of “strategy” between a “diffusion of curvature” strategy and a “piecewise linear” strategy where the whole curvature is concentrated in a singular point.
- It is a variational problem.

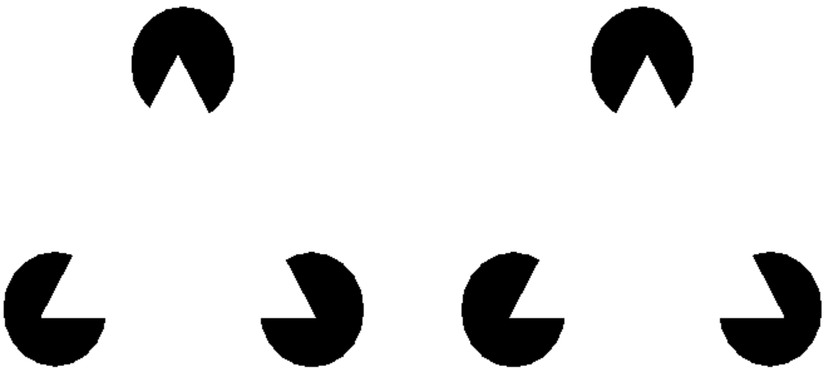
10



- Bistability : the illusory contour is either a circle or a square.
- The example of Ehrenstein illusion:



- Stereo: we must fuse the two figures.

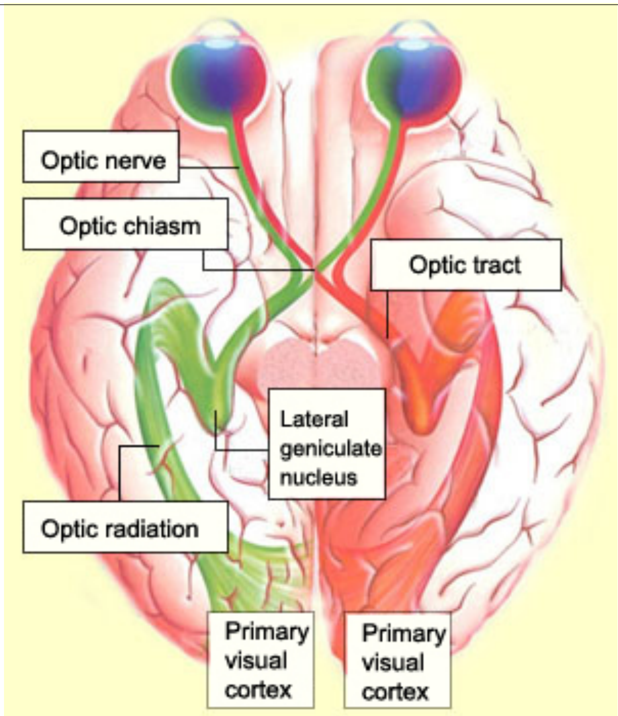


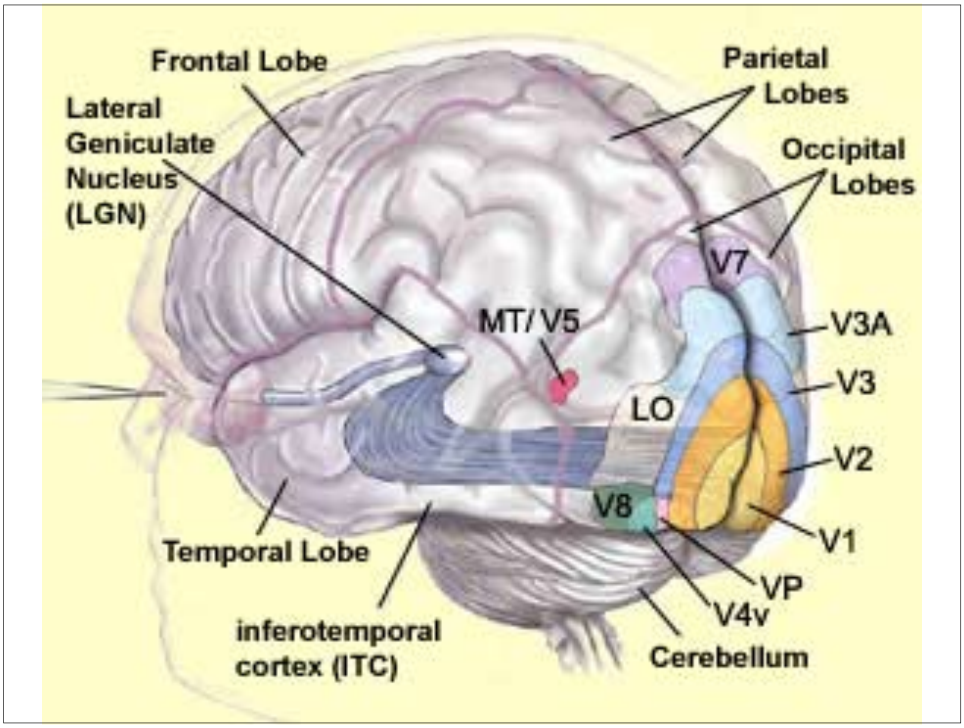
12

- Stereo: curved 2D illusory contours are interpreted as straight boundaries of a curved illusory surface.

The primary visual cortex: area V1

- In mammals (especially higher mammals with frontal eyes), due to the optic chiasm, each visual hemifield projects onto the contralateral hemisphere.
- The fibers from nasal hemiretinae cross the optic chiasm, while the fibers from temporal hemiretinae remain on the ipsilateral side.



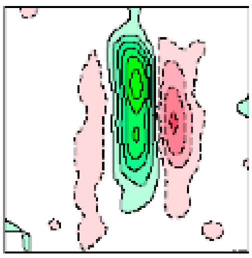


- In the linear approximation, (simple) neurons of V1 operate as filters on the optic signal coming from the retina.
- Their receptive fields (the bundle of photoreceptors they are connected with via the retino-geniculo-cortical pathways) have receptive profiles (transfert function) with a characteristic shape.

15

- We look only at the simplest and most classical definition of the RFs by spiking responses (minimal discharge field).
- We don't take into account the global contextual subthreshold activity of neurons.
- We look at the simplest models.

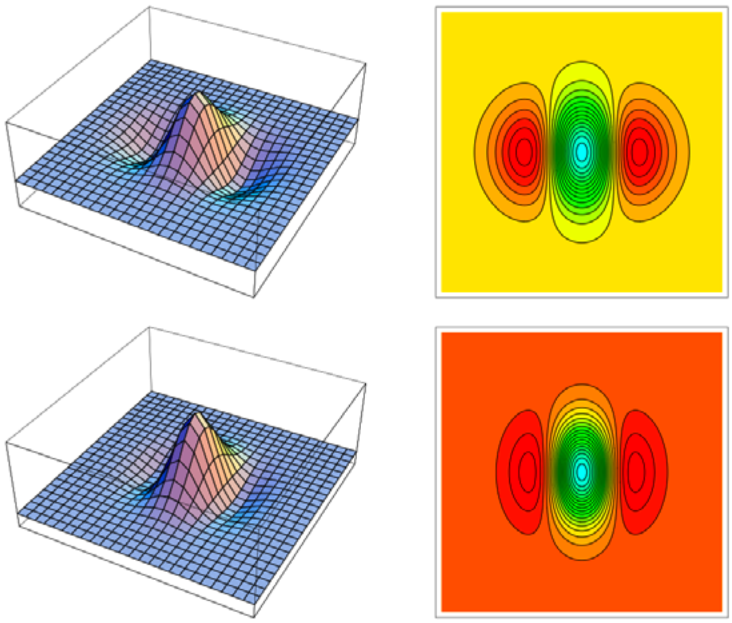
- For simple cells, RFs are highly anisotropic and elongated along a preferential orientation.
- Level curves of the receptive profiles can be recorded :



- The receptive profiles can be modeled either
  - by second order derivatives of Gaussians,
  - or by Gabor wavelets

$$\exp(i2x)\exp\left(-\left(x^2+y^2\right)\right)$$

  (real part).
- See Jeanny Herault's talk.



17

- The RPs operate by <u>convolution</u> on the visual signal.
- Let $I(x, y)$ be the visual signal (x, y are visual coordinates on the retina).

  Let $\varphi(x-x_0, y-y_0)$ be the RP of a neuron N whose RF is defined on a domain D of the retina centered on (x_0, y_0).

- N acts on the signal I as a filter :

$$I_\varphi(x_0, y_0) = \int_D I(x', y')\varphi(x' - x_0, y' - y_0)dx'dy'$$

- A field of such neurons act by convolution on the signal. It is a wavelet analysis.

$$I_\varphi(x, y) = \int_D I(x', y')\varphi(x' - x, y' - y)dx'dy' = (I * \varphi)(x, y)$$

18

- True RF are far more complex. They are adaped to the processing of <u>natural images</u> (and not bars and gratings).
- Joseph Atick, J-P Nadal, have shown that Laplacian RPs can result from "efficiency of information representation".
- An efficient coding must reduce redundancy and maximize the mutual information between visual input and neural response.
- See Hervé Le Borgne's talk.

- The statistic of natural images is very particular because there exist strong correlations between nearby RF.
- Yves Frégnac (UNIC) : 4 statistics. Drifting gratings, dense noise, natural images with eye movements, gratings with EM.
- The variability of spikes decreases with complexity and their temporal precision increases.

- In the linear approximation (convolution $T(I) = I * \varphi$ with a RP $\varphi(x)$), the first thing is to decorrelate the self-correlation of the signal $R(z)$ defined by $R(x-y) = \langle I(x), I(y) \rangle$.
- Fields law (scale invariance of R) : the power spectrum is

$\widehat{R}(\omega) = \frac{1}{|\omega|^2}$ If $\omega = \lambda/\alpha$

$$R(\alpha x) = \int \frac{e^{i\omega\alpha x}}{|\omega|^2} d\omega = \int \frac{\alpha^2 e^{i\lambda x}}{|\lambda|^2} \frac{d\lambda}{\alpha} = \alpha R(x)$$

- Decorrelation = whitening

$$\langle T(I)(x), T(I)(y) \rangle = \delta(x-y)$$

$$\left| \widehat{T(I)}(\omega) \right|^2 = 1$$

- Covariance matrix, with $\varphi'(x) = \varphi(-x)$

$$T(R) = \varphi * R * \varphi'$$

- To get δ, we need

$$\widehat{T(R)}(\omega) = \widehat{\varphi}(\omega)\widehat{R}(\omega)\overline{\widehat{\varphi}}(\omega) = 1$$

20

$$|\widehat{\varphi}(\omega)|^2 = \frac{1}{\widehat{R}(\omega)} \qquad \widehat{R}(\omega) = \frac{1}{|\omega|^2} \qquad |\widehat{\varphi}(\omega)| = |\omega|$$

- This method is not adapted to noise and enhance it at high frequencies where it is already dominant.
- We need a smoothing, hence

$$|\widehat{\varphi}(\omega)|^2 = \frac{\widehat{R}(\omega)+N^2}{\widehat{R}(\omega)^2}$$

- Decorrelation + smoothing leads to Laplacian RPs.

- Complex cells and their non-linearities (Jon Touryan, *Neuron*, 45, 2005, "Spatial Structure of Complex Cell Receptive Fields Measured with Natural Images").
- V1 of cat. Trains of 24 000 natural images (different enough and normalized), every 40ms. Spike-triggered stimulus ensemble.

- Correlation matrix

$$C_{m,n} = \frac{1}{N}\sum_{i=1}^{i=N} S_m(i)\, S_n(i)$$

where $S_n(i)$ is the luminance of the n-th pixel in the stimulus preceding the i-th spike (N = # spikes).

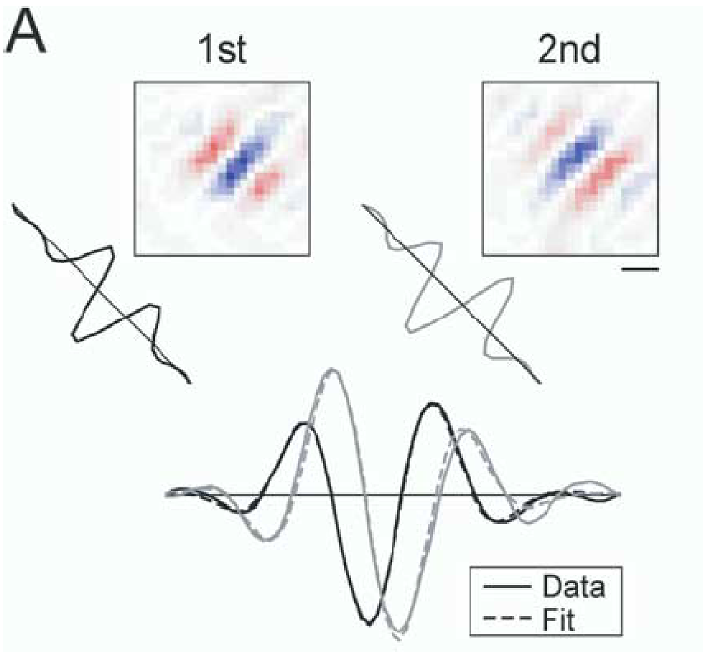
- It is applied to the Atick's transform

$$S_w = SU \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

- The significant eigenvectors are retrieved from

$$V = V_w^T \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix} U^{-1}$$

- In general, there are two significant eigenvectors, which are Gabor RFs with the same spatial frequency and a difference of phase ~ $\pi/2$.



Hypercolumns and pinwheels

- Drastic simplification : simple cells of V1 detect a preferential orientation.
- They measure, at a certain scale, pairs (a, p) of a spatial (retinal) position a and of a local orientation p at a.
- Pairs (a, p) are called in geometry "contact elements".
- Faugeras' generalization (see below).

23

- For a given position $a = (x_0, y_0)$ in R, the simple neurons with variable orientations θ constitute an anatomically definable micromodule called an “hypercolumn”.
- The hypercolumns associate retinotopically to each position a of the retina R a full exemplar P_a of the space P of orientations p at a.

- So, this part of the functional architecture implements the fibration $\pi : R \times P \rightarrow R$ with base R, fiber P, and total space $V = R \times P$.

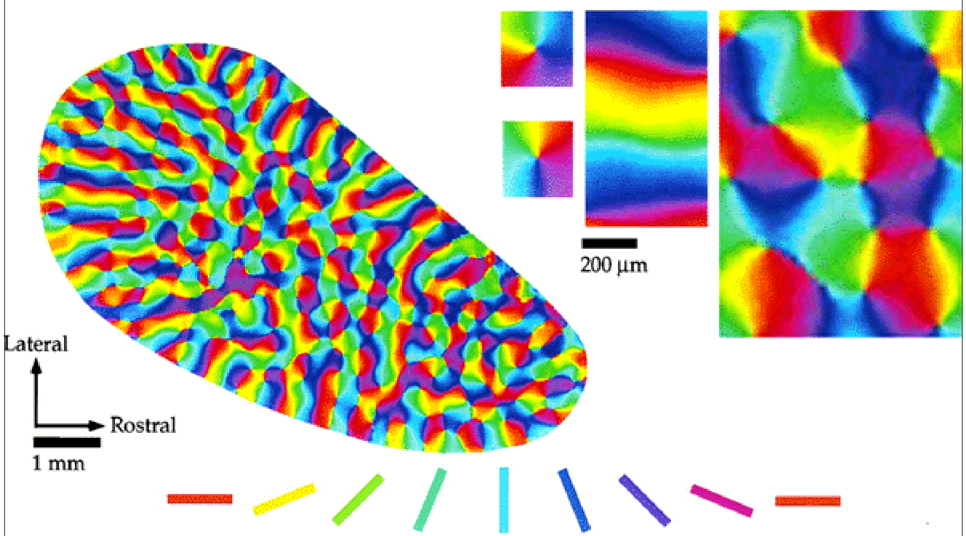
- Hypercolumns are geometrically organized in 2D-pinwheels.
- The cortical layer is reticulated by a network of singular points which are the centers of the pinwheels.
- Locally, around these singular points all the orientations are represented by the rays of a “wheel” and the local wheels are glued together into a global structure.

- The method (Bonhöffer & Grinvald, ~ 1990) of *in vivo optical imaging* based on activity-dependent intrinsic signals allows to acquire images of the activity of the superficial cortical layers.
- Gratings with high contrast are presented many times (20-80) with e.g. a width of 6.25° for the dark strips and of 1.25° for the light ones, a velocity of 22.5°/s, different (8) orientations.

- A window is opened in the skull above V1 and the cortex is illuminated with orange light.
- The concentration of deoxy-hemoglobine increases when neurons are activated. The absorption spectrum of deoxy-hemoglobin is maximal for wave lengths about 600 nm.
- The change is only about 0.2% and the recorded images must therefore be analyzed very carefully.

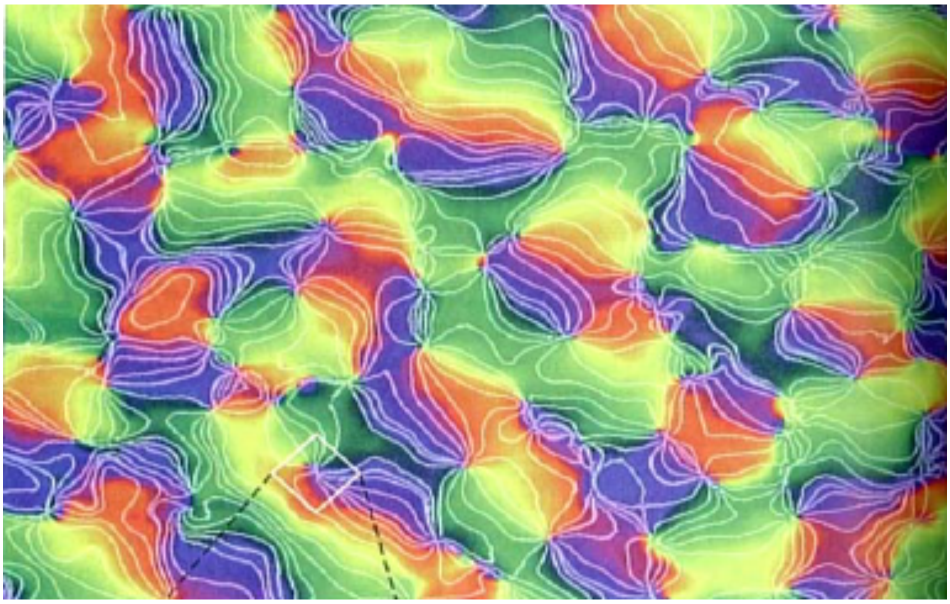
- One does the summation of the images of V1 ’s activity for the different gratings and constructs differential maps (differences between orthogonal gratings).
- The low frequency noise is eliminated.
- The maps are normalized (by dividing the deviation relative to the mean value at each pixel by the global mean deviation).

- In the following picture the orientations are coded by colors and iso-orientation lines are therefore coded by monocolor lines.
- William Bosking, Ying Zhang, Brett Schofield, David Fitzpatrick (Dpt of Neurobiology, Duke) 1997, « Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex », *J. of Neuroscience*, 17, 6, 2112-2127.



27

- There are 2 classes of points :
  - regular points where the orientation field is locally trivial;
  - singular points at the center of the pinwheels;
- Two adjacent singular points are of opposed chirality.



28

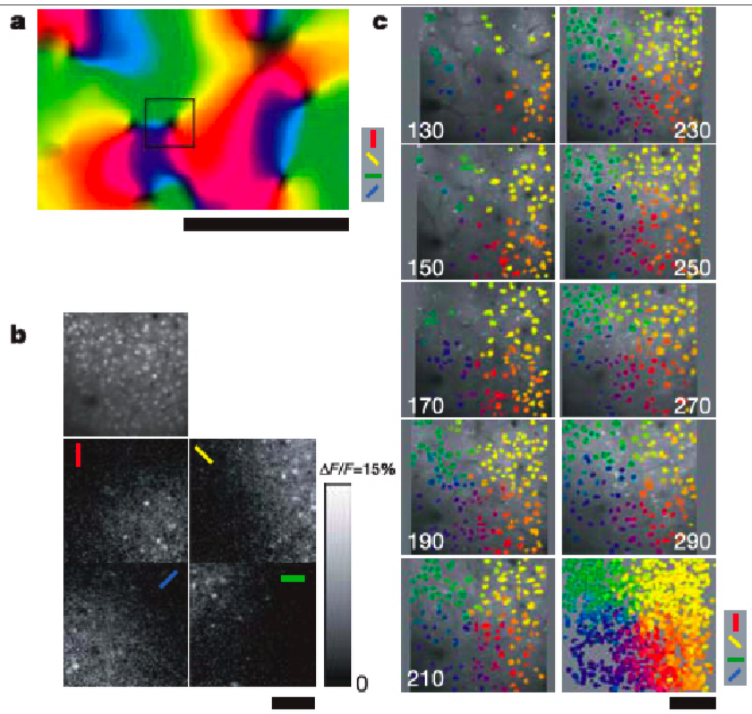
- What is the structure near singularities ?
- The spatial (50μ) and depth resolutions of optical imaging is not sufficient.
- One needs single neuron resolution to understand the micro-structure.

- Two-photon calcium imaging *in vivo* (confocal biphotonic microscopy) provides functional maps at single-cell resolution.
  - Kenichi Ohki, Sooyoung Chung, Prakash Kara, Mark Hübener, Tobias Bonhoeffer and R. Clay Reid:

    Highly ordered arrangement of single neurons in orientation pinwheels, Nature, 442, 925-928 (24 August 2006) .

- (In cat) pinwheels are higly ordered at the micro level and « thus pinwheels centres truly represent singularities in the cortical map ».
- Injection of calcium indicator dye (Oregon Green BAPTA-1 acetoxylmethyl esther) which labels few thousands of neurons in a 300-600μ region.
- Two-photon calcium imaging measures simultaneously calcium signals evoked by visual stimuli on hundreds of such neurons at different depths (from 130 to 290μ by 20μ steps).

- One finds pinwheels with the same orientation wheel.
- « This demonstrates the columnar structure of the orientation map at a very fine spatial scale ».



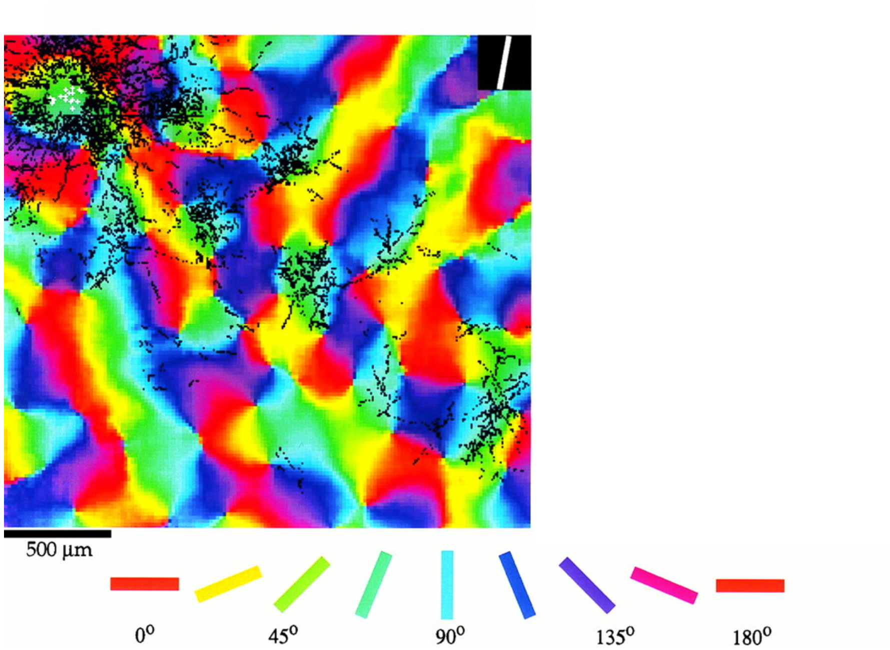
The horizontal structure

- The “vertical” retinotopic structure is not sufficient. To implement a global coherence, the visual system must be able to compare two retinotopically neighboring fibers P_a et P_b over two neighboring points a and b.
- This is a problem of parallel transport. It has been found at the empirical level by the discovery of “horizontal” cortico-cortical connections.

31

- Cortico-cortical connections are slow ($\approx$ 0.2m/s) and weak.
- They connect neurons of almost similar orientation in neighboring hypercolumns.
- This means that the system is able to know, for b near a, if the orientation q at b is the same as the orientation p at a.

- The next slide shows how a marker (biocytin) injected locally in a zone of specific orientation (green-blue) diffuses via horizontal cortico-cortical connections.
- The key fact is that the long range diffusion is highly anisotropic and restricted to zones of the same orientation (the same color) as the initial one.



- Moreover cortico-cortical connections connect neurons coding pairs (a, p) and (b, p) such that p is approximatively the orientation of the axis ab (William Bosking).
  - « The system of long-range horizontal connections can be summarized as preferentially linking neurons with co-oriented, co-axially aligned receptive fields ».
- So, the well known Gestalt law of "good continuation" is neurally implemented.

- In fact, a certain amount of curvature is allowed in alignements.
- These experimental results mean essentially that the contact structure of the fibration $\pi : V = R \times P \rightarrow R$ is neurally implemented.

The contact structure of V1

- The first model : the space of 1-jets of curves C in R.
- It is the beginning of neurogeometry (Hoffman, Koenderink).

- If C is curve in R (a contour), it can be lifted to V. The lifting Γ is the map (1-jet)

$$j : C \to \quad V = R \times P$$

wich associates to every point a of C the pair

(a, p_a) where p_a is the tangent of C at a.

- This Legendrian lift Γ represents C as the enveloppe of its tangents (projective duality).
- In terms of local coordinates (x, y, p) in V, the equation of Γ writes $(x, y, p) = (x, y, y')$.

- If we have an image $I(x, y)$ on R, we can lift it in V by lifting its level curves.

35

Fonctionality of jet spaces

- The functional interest of jet spaces is that they can be implemented by “point processors” (Koenderink) such as neurons.
- But then a functional architecture is needed.
- Functional architectures of point processors can compute features of differential geometry.

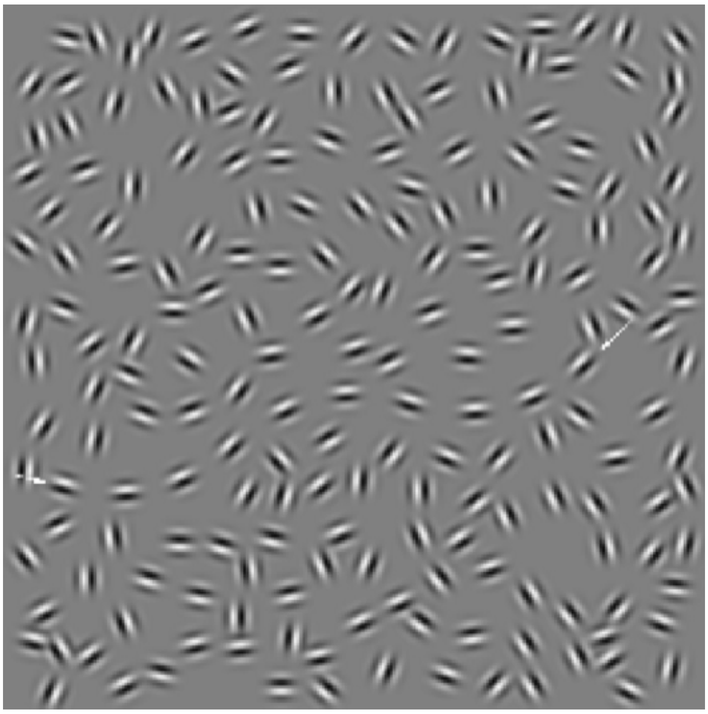
- The key idea is
  - (1) to add new independent variables describing local features such as orientation.
  - (2) to introduce an integrability constraint to integrate them into global structures.
- Neuro-physiologically, this means to add feature detectors and to couple them via a functional architecture in order to ensure binding.

- To every curve C in R is associated a curve Γ in V. But the converse is of course false.
- If $\Gamma = (a, p) = (x, y(x), p(x))$ is a curve in V, the projection $a = (x, y(x))$ of Γ is a curve C in R. But Γ is the lifting of C iff $p(x) = y'(x)$.
- This condition is called a Frobenius integrability condition. It says that to be a coherent curve in V, Γ must be an integral curve of the contact structure of the fibration π.

Frobenius condition and Association field

- Frobenius integrability condition corresponds to the psychophysical experiments on the association field (David Field, Anthony Hayes and Robert Hess).
- They explain experiments on good continuation : pop out of a global curve against a background of randomly distributed distractors

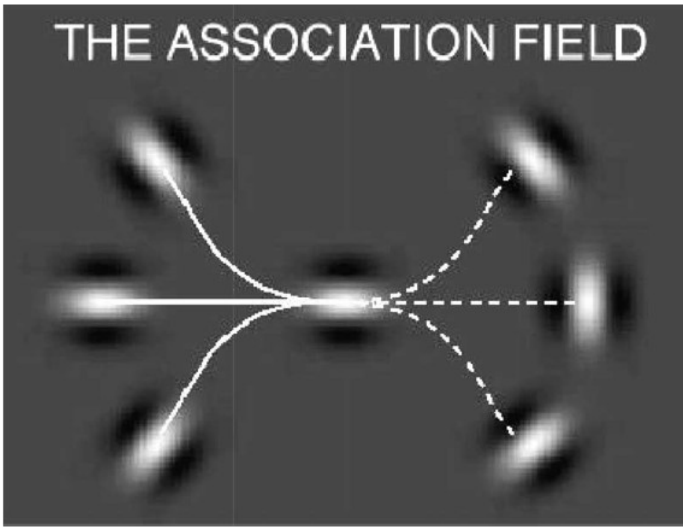
- Let (a_i, p_i) be a set of segments embedded in a background of randomly distributed distractors. The segments generate a perceptively salient curve (pop-out) iff the p_i are tangent to the curve C optimally interpolating between the a_i.



38

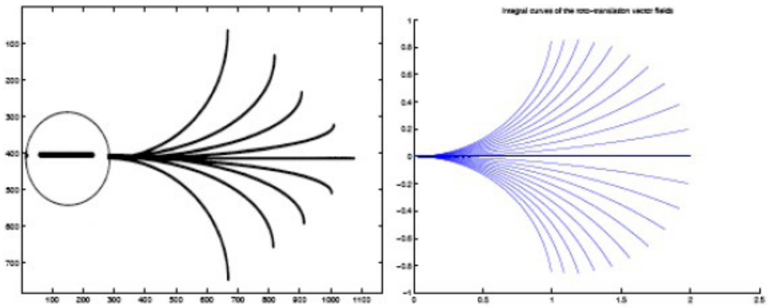
- This is a discretized version of the integrability condition.
- The integrability induces a binding of the local elements. The activities of the neurons detecting them are synchronized and the synchronization produces the pop out.

- One must have the following type of horizontal connectivity :



39

- But this is exactly the integrability condition : the association field (left) correspond to the simplest integral curves of the contact distribution (right).



- Frobenius condition is extremely simple :

$$p = dy/dx$$

- But it contains deep mathematics.

- Frobenius integrability condition is equivalent to the fact that if

$$t = (x, y, p; 1, y', p')$$

is a tangent vector to V at the point (x, y, p), then t is in the kernel of the 1-form

$$\omega = dy - pdx$$

($\omega = 0$ means $p = dy / dx$).

- To compute the value of a 1-form ω on a tangent vector $t = (\xi, \eta, \pi)$ at (x, y, p), one applies the rules

$$dx(t) = \xi, \ dy(t) = \eta, \ dp(t) = \pi.$$

- So the kernel of the 1-form ω is the field of the planes (called the contact planes)

$$\eta - p\xi = 0 .$$

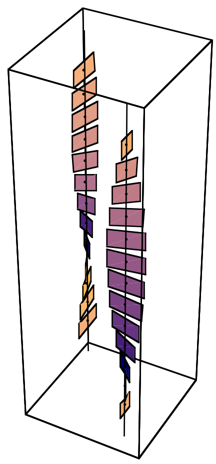
- $X_1 = \partial_x + p\partial_y = (\xi = 1, \ \eta = p, \pi = 0)$, and $X_2 = \partial_p = (\xi = 0, \ \eta = 0, \pi = 1)$ are evident generators.

41

- Moreover, in a Legendrian lift Γ, the vertical component p' of a tangent vector is the curvature of the curve C in the base space R :

$$p = y' \ \Rightarrow \ p' = y''$$

- The field of the contact planes has many integral curves : all the Legendrian lifts. But it has no integral surfaces.
- This is due to the fact that the contact planes "rotate" too fast to be the tangent planes of a surface.



Contact structure and Heisenberg group

- The contact structure on V is left-invariant for a group structure which is isomorphic to the Heisenberg group :

$$(x, y, p).(x', y', p') = (x + x', y + y' + px', p + p')$$

- If $t = (\xi, \eta, \pi)$ are the tangent vectors of $\mathfrak{V} = T_0V$, the Lie algebra of V has the Lie bracket

$$[t, t'] = [(\xi, \eta, \pi), (\xi', \eta', \pi')] = (0, \xi'\pi - \xi\pi', 0)$$

- In matrix terms, $v = (x, y, p)$ and $t = (\xi, \eta, \pi)$

$$\begin{pmatrix} 1 & p & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix}$$

- Inner automorphism :

$$\begin{array}{cccc} A_v : & v' & \mapsto & v.v'.v^{-1} \\ & (x', y', p') & \mapsto & (x', y' + px' - p'x, p') \end{array}$$

- Tangent map of A_v at 0 :

$$Ad_v = \begin{pmatrix} 1 & 0 & 0 \\ p & 1 & -x \\ 0 & 0 & 1 \end{pmatrix}$$

$$Ad_v(t) = (\xi, p\xi + \eta - x\pi, \pi)$$

- This yields the adjoint representation of the Lie group V on its Lie algebra $\mathfrak{V} = T_0V$.

- For the coadjoint representation, take the basis $\{dx, dy, dp\}$ for the 1-forms of $\mathfrak{V}^*$:

$$\theta = \alpha dx + \beta dy + \delta dp = (\alpha, \beta, \delta)$$

- We get $\langle Ad_v^*(\theta), t\rangle = \langle \theta, Ad_{-v}(t)\rangle$

$$Ad_v^*(\theta) = (\alpha - \beta p, \beta, \delta + \beta x)$$

- Orbits :
  - If $\beta \neq 0$, planes $\beta = \text{cst.}$
  - If $\beta = 0$, every point of the $(\alpha, 0, \delta)$ plane.

44

- For $X_1 = \partial_x + p\partial_y = (1, p, 0)$, and $X_2 = \partial_p = (0, 0, 1)$, we have :

$$[X_1, X_2] = X_3 = -\partial_y = (0, -1, 0)$$

(other brackets $= 0$).

- It is essential to understand this geometry since it drives diffusion (heat equation) and propagation (wave equation) in V1.

Sub-Riemannian geometry

- In this neuro-geometrical framework, we can easily interpret the variational process giving rise to illusory contours.
- The key idea is to use a geodesic model in the sub-Riemannian geometry associated to the contact structure.
- This deepens the “elastica” model proposed by David Mumford.
- See Ugo Boscain's and Jean-Paul Gauthier's talks.

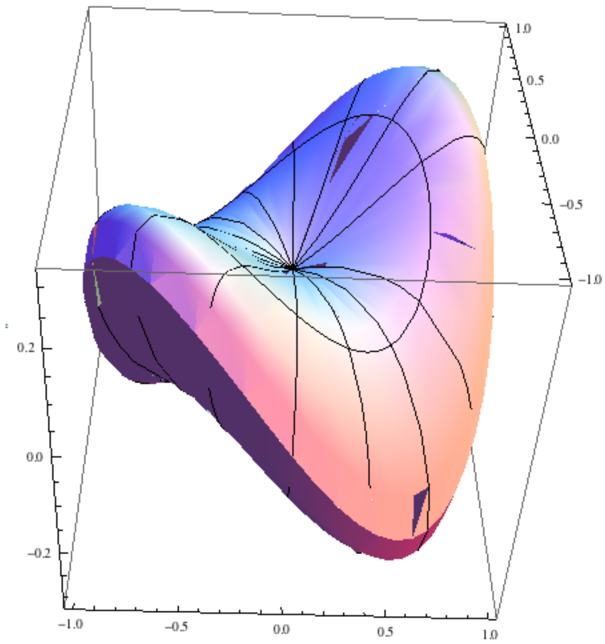
- If $\mathcal{K}$ is the contact structure on V and if one considers only curves Γ in V which are integral curves of $\mathcal{K}$, then metrics $g_{\mathcal{K}}$ defined only on the planes of $\mathcal{K}$ are called sub-Riemannian metrics.

- In a Kanizsa figure, two pacmen of respective centers a and b with a specific aperture angle define two elements (a, p) and (b, q) of V.
- An illusory contour interpolating between (a, p) and (b, q) is
  - 1. a curve C from a to b in R with tangent p at a and tangent q at b;
  - 2. a curve minimizing an "energy" (variational problem), that is a geodesic for some sub-Riemannian metric.

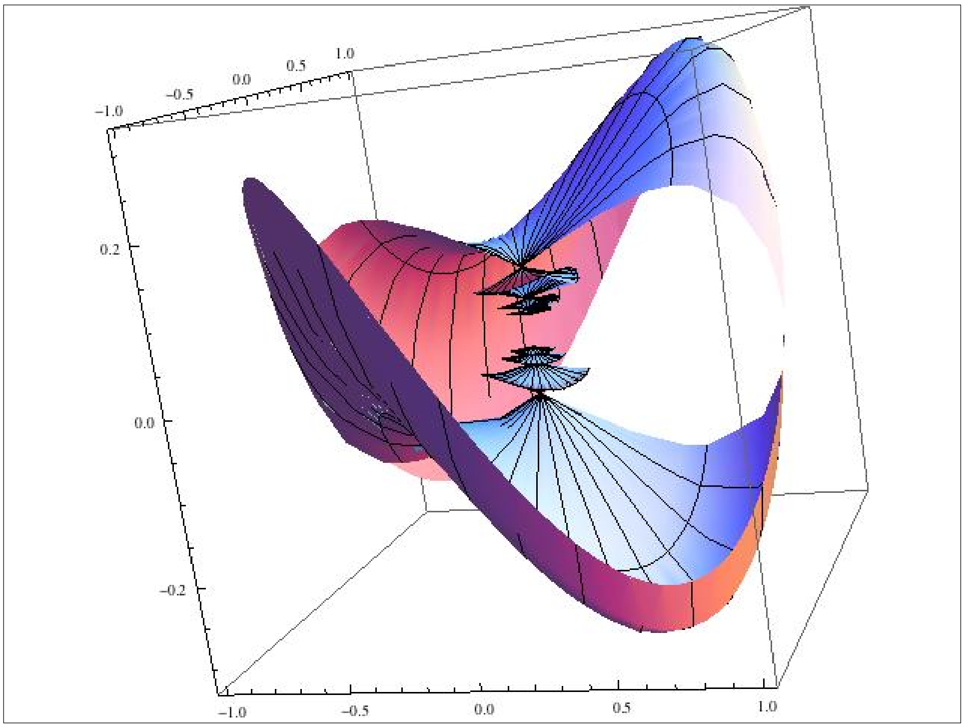
- For a model isomorphic to the space of 1-jets (the Heisenberg group), Richard Beals, Bernard Gaveau and Peter Greiner have solved the problem.
- They claimed :
  - “The results indicate how complicated a control problem can become, even in the simplest situation.”

- In 1977 Bernard Gaveau still said:
  - “Le problème variationnel est le problème de minimiser l'énergie d'une courbe de la variété de base sous la condition de Lagrange que son relèvement horizontal est fixé dans le fibré. Ce problème semble non étudié.'”

- It is natural to take on the contact planes the metric making orthonormal their evident generators : $X_1 = \partial_x + p\partial_y$, $X_2 = \partial_p$ whose Lie bracket is $[X_1, X_2] = -X_3 = -\partial_y$.
- The structure of geodesics implies that the sub-Riemannian sphere S and the wave front W (geodesics of SR length 1) are rather strange. We can compute them explicitely (it is a variant of Beals *et al.* computations).



48



- Geodesics are projections on $V = \mathbb{R}^3$ of Hamiltonian trajectories of an Hamiltonian H defined on the cotangent bundle $T^*\mathbb{R}^3$.
- The fundamental problem (compared with Euclidean geometry) is that the exponential map $\mathscr{E}_v$ (the process of integrating geodesics starting at v) has singularities.

49

- Sphere $S(v, r) = \{\, w : d(v, w) = r$ (geodesics that are global minimizers)$\}$.
- Wave front $W(v, r) = \{\, w : \exists$ a geodesic $\gamma : v \to w$ of length r (not necessarily a global minimizer)}.
- Cut locus of $v = \{\, w : w$ end point of a geodesic $\gamma : v \to w$ which is no longer globally minimizing }.
- Conjugate locus of v = caustic = Σ_v = { singular locus of $\mathscr{E}_v$ }.

- The geodesics are the projections of the trajectories associated to the Hamiltonian

$$H\,(x, y, p, \xi, \eta, \pi) = \frac{1}{2}\left[(\xi + p\eta)^2 + \pi^2\right]$$

which corresponds to the metric making

$$X_1 = (1, p, 0), X_2 = (0, 0, 1)$$

$$\begin{cases} X_1 = \frac{\partial}{\partial x} + p\frac{\partial}{\partial y} \\ X_2 = \frac{\partial}{\partial p} \end{cases}$$

an orthonormal basis.

- Hamilton equations are

$$\begin{cases} \dot{x}(s) = \frac{\partial H}{\partial \xi} = \xi + p\eta \\ \dot{y}(s) = \frac{\partial H}{\partial \eta} = p\left(\xi + p\eta\right) \\ \dot{p}(s) = \frac{\partial H}{\partial \pi} = \pi \\ \dot{\xi}(s) = -\frac{\partial H}{\partial x} = 0 \\ \dot{\eta}(s) = -\frac{\partial H}{\partial y} = 0 \\ \dot{\pi}(s) = -\frac{\partial H}{\partial p} = -\eta\left(\xi + p\eta\right) \end{cases}$$

- The moments ξ and η are constant because H is independent of x and y.

- The integration of the (x, p) part is the easiest (τ is the end time and x_1, y_1, p_1 the end points of the geodesic starting at 0). It yields :

$$\begin{cases} x\left(s\right) = \frac{\sin\left(\frac{s}{2}\eta_0\right)}{\sin\left(\frac{\tau}{2}\eta_0\right)}\left(\cos\left(\frac{(\tau-s)}{2}\eta_0\right)x_1 - \sin\left(\frac{(\tau-s)}{2}\eta_0\right)p_1\right) \\ p\left(s\right) = \frac{\sin\left(\frac{s}{2}\eta_0\right)}{\sin\left(\frac{\tau}{2}\eta_0\right)}\left(\sin\left(\frac{(\tau-s)}{2}\eta_0\right)x_1 + \cos\left(\frac{(\tau-s)}{2}\eta_0\right)p_1\right) \end{cases}$$

- The integration of y is much more complex. It yields

$$
\begin{aligned}
y(s) - y_0 \quad = \quad & \frac{1}{8(\cos(\eta_0\tau) - 1)}[-2\eta_0 s(x_1^2 + p_1^2) - 4x_1p_1\cos(\eta_0(s-\tau)) + \\
& 2(x_1^2 - p_1^2)\sin(\eta_0(s-\tau)) + \\
& 2x_1p_1\cos(\eta_0(2s-\tau)) - (x_1^2 - p_1^2)\sin(\eta_0(2s-\tau)) + \\
& 2x_1p_1\cos(\eta_0\tau) + (x_1^2 - p_1^2)\sin(\eta_0\tau) + \\
& 2(x_1^2 + p_1^2)\sin(\eta_0 s)]
\end{aligned}
$$

- The key point is that we have, with $z = (x, p)$ and the new variable

$$\varphi = \frac{\eta_0\tau}{2}$$

$$4\left(y_1 - y_0 - \frac{1}{2}x_1p_1\right) = \mu(\varphi)\,\|z_1\|^2$$

with

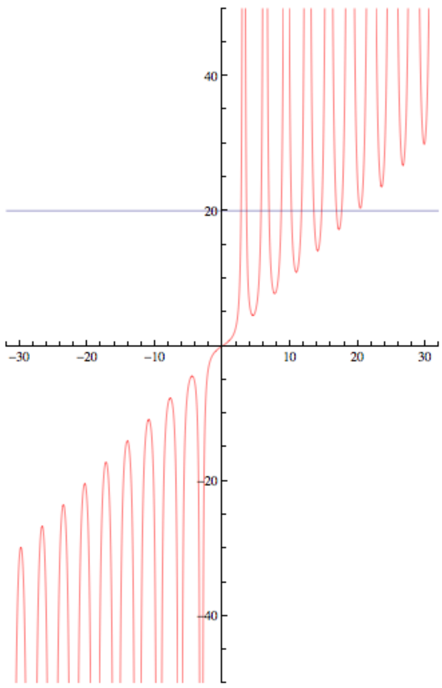
$$\mu(\varphi) = \frac{\varphi}{\sin^2(\varphi)} - \cot(\varphi)$$

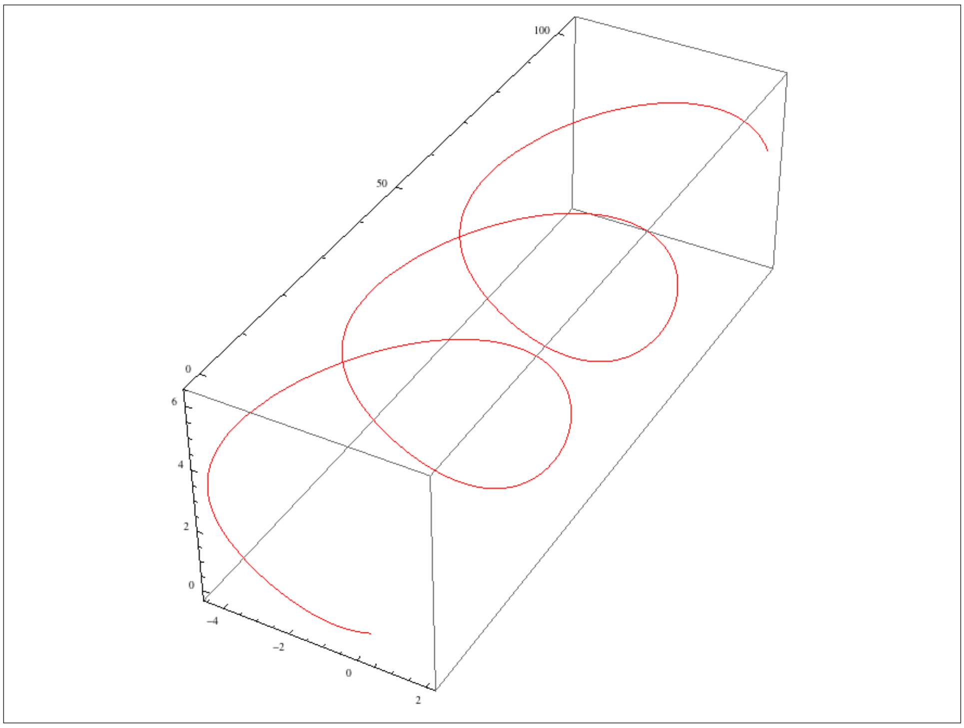
52

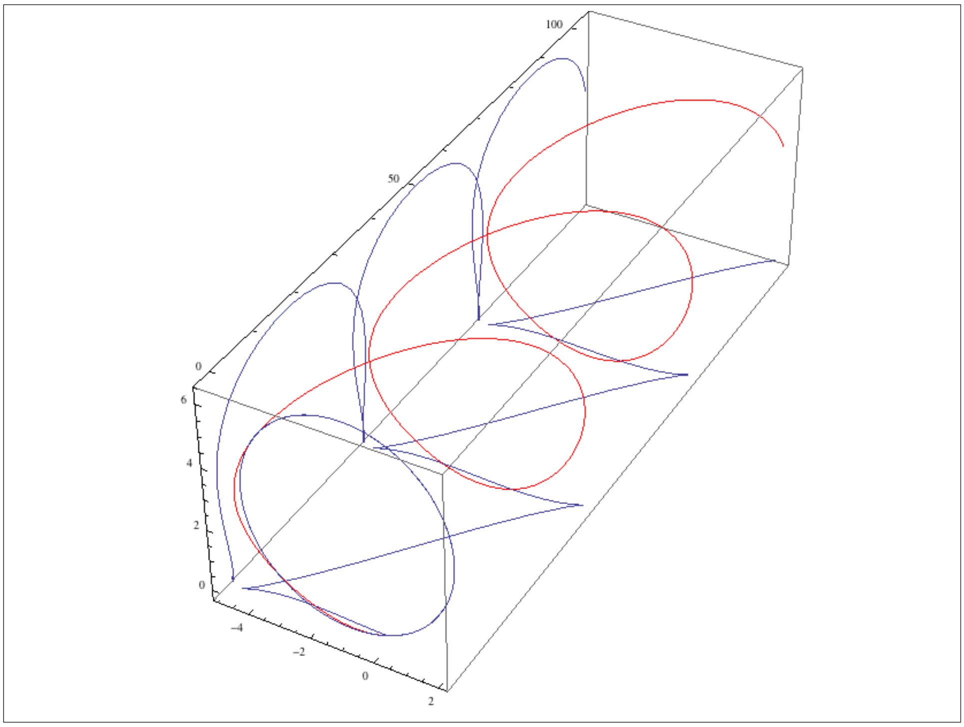
- So geodesics with the <u>same</u> end points correspond to the solutions of the equation

$$\mu(\varphi) = \text{cst}$$

- For instance, for $x_1 = 2$, $p_1 = 4$, $y_1 = 104$, we find $\mu(\varphi) = 20$, which has 11 solutions.







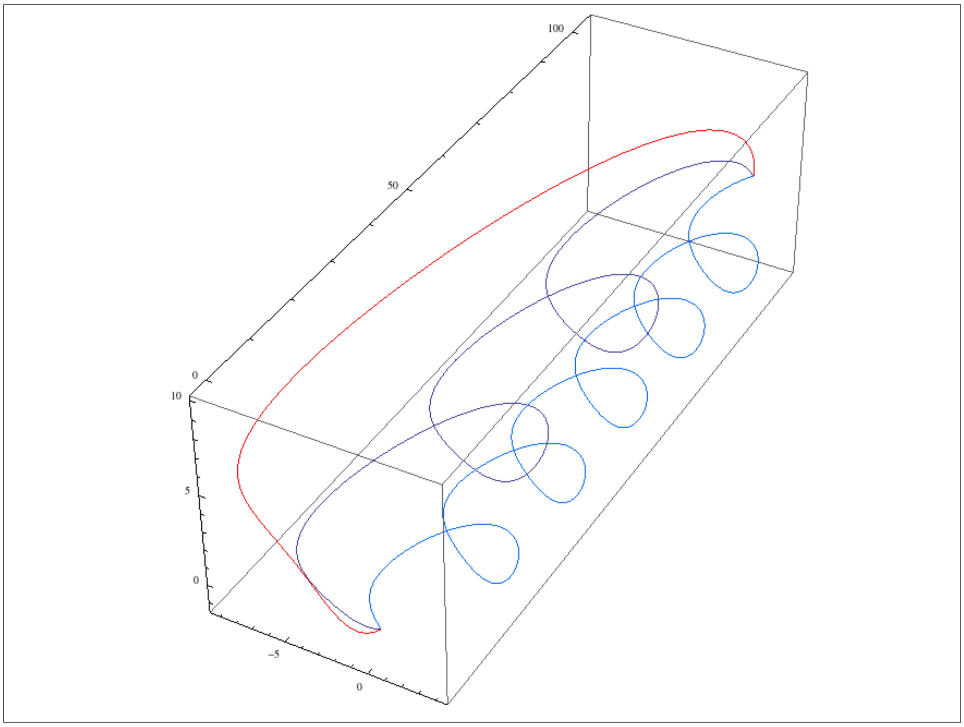
54

- Projections of geodesics on the (x, p) plane are circles

$$x^2 + p^2 - x\left(x_1 + p_1 \cot\left(\frac{\eta_0 \tau}{2}\right)\right) - p\left(p_1 - x_1 \cot\left(\frac{\eta_0 \tau}{2}\right)\right) = 0$$

with center

$$x_c = \frac{1}{2}\left(x_1 + p_1 \cot\left(\frac{\eta_0 \tau}{2}\right)\right), y_c = \frac{1}{2}\left(p_1 - x_1 \cot\left(\frac{\eta_0 \tau}{2}\right)\right)$$



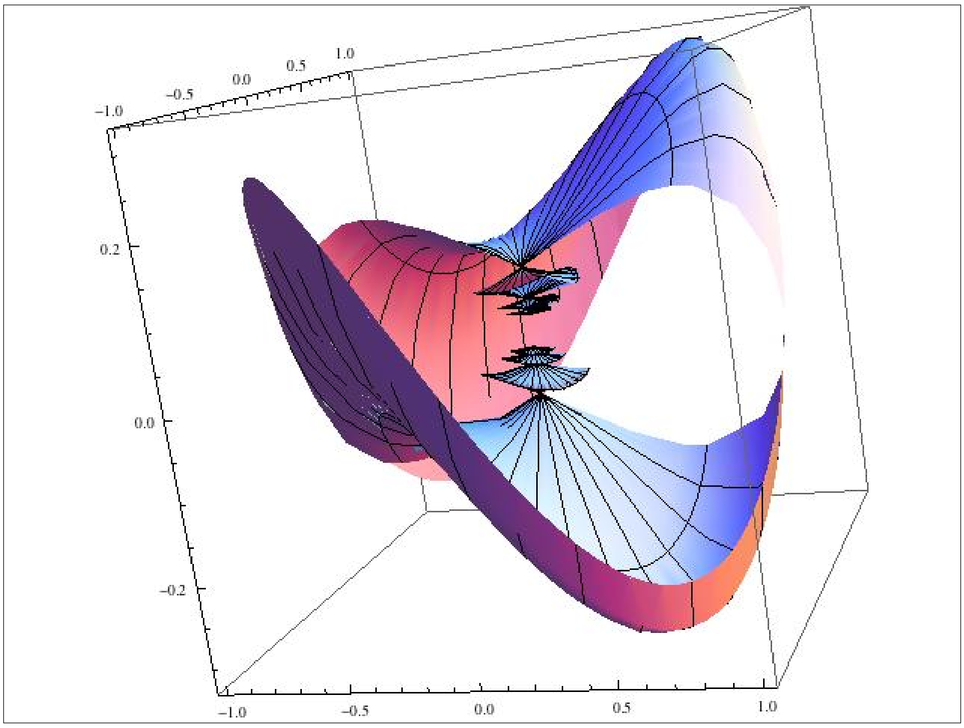
55

- To compute the sub-Riemannian sphere and wave front we must compute the sub-Riemannian length L of geodesics.
- We find the fundamental equation (for L = sqrt(2))

$$|z_1| = \frac{|\sin(\varphi)|}{\varphi}$$

- And therefore

$$x_1 = \frac{|\sin(\varphi)|}{\varphi}\cos(\theta)$$
$$p_1 = \frac{|\sin(\varphi)|}{\varphi}\sin(\theta)$$
$$y_1 = \frac{\varphi + 2\sin^2(\varphi)\cos(\theta)\sin(\theta) - \cos(\varphi)\sin(\varphi)}{4\varphi^2}$$



Unitary irreducible representations

- Since the sub-Riemannian geometry is invariant under the group structure of V, it is essential to know the unitary irreducible representations (unirreps) of this group.
- We can adapt a celebrated result for the Heisenberg group known as the Stone - von Neumann theorem.

- The unirreps of V are either trivial ones of dimension 1 multiplying $z \in \mathbb{C}$ by

$$\pi_{\mu,\nu}(x,y,p) = e^{i(\mu x + \nu p)}$$

or infinite dimensional ones operating in the Hilbert space $L^2(\mathbb{R})$

$$\pi_\lambda(x,y,p)\,u(s) = e^{i\lambda(y+xs)}u(s+p), \text{ with } \lambda \neq 0$$

- Kirillov : they correspond to the orbits of the coadjoint representation of V.

$$Ad_v^*(\theta) = (\alpha - \beta p, \beta, \delta + \beta x)$$

- Planes $\beta = \mathrm{cst} = \lambda$ for $\beta \neq 0$ correspond to

$$\pi_\lambda(x,y,p)\,u(s) = e^{i\lambda(y+xs)}u(s+p), \text{ with } \lambda \neq 0$$

- Points of the $(\alpha=\mu, 0, \delta=\nu)$ plane for $\beta=\lambda=0$ correspond to

$$\pi_{\mu,\nu}(x,y,p) = e^{i(\mu x + \nu p)}$$

Contact structure and Euclidean group

- Alessandro Sarti and Giovanna Citti emphasized the fact that it is more natural to work in the fibration $\pi : V = R \times P \rightarrow R$ with $P = \mathbb{S}^1$ and with the contact form

$$\omega = -\sin(\theta)dx + \cos(\theta)dy$$

which is $\cos(\theta)(dy - pdx)$

- (No privileged x-axis)

- The contact planes are spanned by

$$X_1 = \cos(\theta)\,\partial_x + \sin(\theta)\,\partial_y$$
$$X_2 = \partial_\theta$$

with Lie bracket

$$[X_1, X_2] = \sin(\theta)\,\partial_x - \cos(\theta)\,\partial_y = -X_3$$

- (Tangent vectors are interpreted as oriented derivatives.)
- This is a non-holonomic basis.

- *V* becomes a Lie group isomorphic to the Euclidean group (semi-direct product)

$$E(2) = SO(2) \ltimes \mathbb{R}^2$$

$$\begin{pmatrix} x_1 \\ y_1 \\ \theta_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ y_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \cos(\theta_1) - y_2 \sin(\theta_1) \\ y_1 + x_2 \sin(\theta_1) + y_2 \cos(\theta_1) \\ \theta_1 + \theta_2 \end{pmatrix}$$

- This group is <u>not</u> nilpotent and its tangent cone is the Heisenberg group.

- Left invariance :

$$\{\partial_x, \partial_y, \partial_\theta\}_0$$

left translates into the non-holonomic basis

$$\{\cos(\theta)\,\partial_x + \sin(\theta)\,\partial_y = X_1, -\sin(\theta)\,\partial_x + \cos(\theta)\,\partial_y = X_3, \partial_\theta = X_2\}_q$$

and the covector

$$\omega_0 = dy$$

left translates into the contact form ω.

Curvature and Engel structure

- Some experiments (Steve Zucker) seems to indicate that there exist in the primary visual cortex curvature detectors.
- If we want to model this possibility, we must use 2-jets spaces and add a new independent variable K, which will be interpreted as the curvature of curves C in the (x, y) base plane R.

In the 3D space

$$V = \mathbb{R}^2 \times \mathbb{S}^1$$

we have the contact structure defined by the 1-form K

$$\omega = -\sin(\theta)dx + \cos(\theta)\,dy$$

The (non holonomic) basis for the contact planes is

$$X_1 = \cos(\theta)\,\partial_x + \sin(\theta)\partial_y$$
$$X_2 = \partial_\theta$$

The Lie bracket is

$$[X_1, X_2] = X_3 = -\sin(\theta)\partial_x + \cos(\theta)\,\partial_y$$

We want to add *curvature* K and work in the 4D space

$$W = \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{R}$$

Now, we have a Pfaff system constituted of *two* 1-forms: ω and

$$\tau = d\theta - K ds$$

If we parametrize the curves in the base space (x, y) using the arc length s, the curvature is

$$K = \frac{d\theta}{ds}$$

In cartesian coordinates, for a curve with (local) equation $y = f(x)$, the curvature is

$$K = \frac{f''(x)}{\left(1 + f'(x)^2\right)^{3/2}}$$

The link between the two formulas is easy:

$$\begin{aligned} x'(s) &= \cos(\theta) \\ y'(s) &= \sin(\theta) = f'(x)\, x'(s) \\ y''(s) &= f''(x)\, x'(s)^2 + f'(x)\, x''(s) \end{aligned}$$

62

$$f''(x) = \frac{y''(s) - f'(x)\,x''(s)}{x'(s)^2}$$
$$= \frac{\cos(\theta)\,\theta'(s) + \tan(\theta)\sin(\theta)\,\theta'(s)}{\cos(\theta)^2}$$
$$= K\left(\frac{1}{\cos(\theta)} + \frac{\sin(\theta)^2}{\cos(\theta)^3}\right) = \frac{K}{\cos(\theta)^3}$$

But

$$1 + f'(x)^2 = 1 + \tan(\theta)^2 = \frac{1}{\cos(\theta)^2}$$

and therefore

$$K = f''(x)\cos(\theta)^3 = \frac{f''(x)}{\left(1 + f'(x)^2\right)^{3/2}}$$

63

To express the second 1-form τ, we write

$$dx = \cos(\theta)\, ds$$
$$dy = \sin(\theta) ds$$
$$ds = \cos(\theta)\, dx + \sin(\theta) dy = \left(\cos(\theta)^2 + \sin(\theta)^2\right) ds$$

and therefore

$$\tau = d\theta - K ds$$
$$\tau = d\theta - K\left(\cos(\theta)\, dx + \sin(\theta) dy\right)$$

The kernel of τ is generated by the 3 tangent vectors

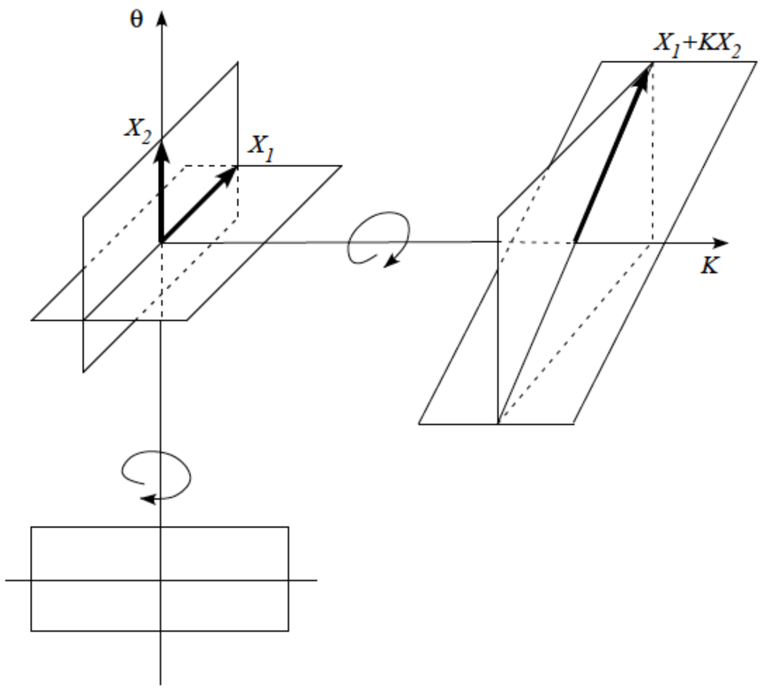
$$X_1^K = \cos(\theta)\, \partial_x + \sin(\theta)\partial_y + K\partial_\theta = X_1 + KX_2$$
$$X_3 = -\sin(\theta)\partial_x + \cos(\theta)\, \partial_y$$
$$X_4^K = \partial_K$$

while the kernel of ω extended to W is generated by X_1, X_2 and X_4^K.

64

In the 4D space W, the tangent vector $X_1 + KX_2$ is helicoidally unfolded along the K-axis.
The distribution of planes is now $\mathrm{Span}\left\{X_1^K, X_4^K\right\}$.
It generates the whole Lie algebra since

$$\left[X_1^K, X_4^K\right] = -X_2 = -\partial_\theta$$
$$\left[\left[X_1^K, X_4^K\right], X_1^K\right] = X_3 = -\sin(\theta)\partial_x + \cos(\theta)\,\partial_y$$



65

Sub-Riemannian geometry of the Euclidean group E(2)

- For the non nilpotent Euclidean group, Andrei Agrachev and his group at the SISSA (Yuri Sachkov, Ugo Boscain, Igor Moiseev) solved the problem of SR geodesics and Sachkov compared it with the Mumford's elastica model.

- One works in the fibration $V = \mathbb{R}^2 \times \mathbb{S}^1$ where the Legendrian lifts are solutions of the control system :

$$\begin{cases} \dot{x} = u_1 \cos(\theta) \\ \dot{y} = u_1 \sin(\theta) \\ \dot{\theta} = u_2 \end{cases}$$

- Let

$$p = (p_x, p_y, p_\theta) \in T_q^* V$$

- The Hamiltonian on T^*V for geodesics is

$$H(p,q) = \frac{1}{2}\left(u_1^2 + u_2^2\right) = \frac{1}{2}\left(\left(p_x \cos(\theta) + p_y \sin(\theta)\right)^2 + p_\theta^2\right)$$

- Hamilton equations are therefore :

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p_x} = p_x \cos^2(\theta) + p_y \cos(\theta)\sin(\theta) \\ \dot{y} = \frac{\partial H}{\partial p_y} = p_y \sin^2(\theta) + p_x \cos(\theta)\sin(\theta) \\ \dot{\theta} = \frac{\partial H}{\partial p_\theta} = p_\theta \end{cases}$$

$$\begin{cases} \dot{p}_x = -\frac{\partial H}{\partial x} = 0 \\ \dot{p}_y = -\frac{\partial H}{\partial y} = 0 \\ \dot{p}_\theta = -\frac{\partial H}{\partial \theta} = \left(p_x \cos(\theta) + p_y \sin(\theta)\right)\left(-p_x \sin(\theta) + p_y \cos(\theta)\right) \end{cases}$$

- The system can be explicitely integrated via elliptic functions.
- The sub-Riemannian geodesics are the projections of the integral curves on V.

Scale-space and symplectic structures

- The contact structure of V is defined as the kernel field of the 1-form ω.
- But this field is only defined up to a scale factor $s = e^{\sigma}$, ω and $s\omega$ having the same kernels.
- It is therefore natural to enlarge the 3 dimensional contact space $V = \mathbb{R}^2 \times \mathbb{S}^1$ to the 4 dimensional space $G = \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{R}$ with coordinates (x, y, θ, σ).

- G is the affine group of the plane and its invariant basis is now

$$\begin{cases} X_1 = e^{\sigma}\left(\cos(\theta)\partial_x + \sin(\theta)\partial_y\right) \\ X_2 = \partial_\theta \\ X_3 = e^{\sigma}\left(-\sin(\theta)\partial_x + \cos(\theta)\partial_y\right) \\ X_4 = \partial_\sigma \end{cases}$$

the invariant 1-form being now

$$\omega = e^{-\sigma}\left(-\sin(\theta)dx + \cos(\theta)dy\right)$$

68

- $d\omega$ is the symplectic 2-form on G

$$\begin{aligned} d\omega &= \left(e^{-\sigma}\cos(\theta)dx + e^{-\sigma}\sin(\theta)dy\right)\wedge d\theta + \\ &\quad \left(-e^{-\sigma}\sin(\theta)dx + e^{-\sigma}\cos(\theta)dy\right)\wedge d\sigma. \end{aligned}$$

deduced via left translations from the canonical symplectic 2-form at 0

$$dx \wedge d\theta + dy \wedge d\sigma.$$

- Indeed, the translated of dx and dy are

$$\begin{cases} \upsilon = e^{-\sigma}\left(\cos(\theta)dx + \sin(\theta)dy\right) \\ \omega = e^{-\sigma}\left(-\sin(\theta)dx + \cos(\theta)dy\right) \end{cases}$$

and $d\omega = \upsilon \wedge d\theta + \omega \wedge d\sigma$

- $d\omega$ can be writen using an antisymmetric matrix B

$$d\omega(X, X') = \langle BX, X' \rangle$$

$$B = e^{-\sigma}\begin{pmatrix} 0 & 0 & -\cos(\theta) & \sin(\theta) \\ 0 & 0 & -\sin(\theta) & -\cos(\theta) \\ \cos(\theta) & \sin(\theta) & 0 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & 0 \end{pmatrix}$$

- $-B^2 = BB^*$ is positive definite $-B^2 = e^{-2\sigma}I$ and we can therefore consider

69

$$P = \sqrt{-B^2} = e^{-\sigma} I$$

- Then, $J = BP^{-1} = e^{\sigma} B$ satisfies $J^2 = -I$

  and defines a complex structure

$$J = \begin{pmatrix} 0 & 0 & -\cos(\theta) & \sin(\theta) \\ 0 & 0 & -\sin(\theta) & -\cos(\theta) \\ \cos(\theta) & \sin(\theta) & 0 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & 0 \end{pmatrix}$$

- If we define a new scalar product by

$$(X|Y) = e^{-\sigma} \langle X|Y \rangle$$

  then $d\omega(X, Y) = (JX|Y)$

- The planes Span$\{ X_1, X_2 \}$, Span$\{ X_3, X_4 \}$ are complex lines (real planes), on which J acts as multiplication by i.

- The equation of the contact curves is

$$\gamma'(t) = X_1(\gamma(t)) + k(t)X_2(\gamma(t)),$$
$$\gamma(0) = (x_0, y_0, \theta_0, \sigma_0),$$

- For $k = \mathrm{cst}$, solutions are

$$\begin{cases} x &= \frac{1}{k}\left(\sin(kt+\theta_0) - \sin(\theta_0) + kx_0\right), \\ y &= \frac{1}{k}\left(-\cos(kt+\theta_0) + \cos(\theta_0) + ky_0\right), \\ \theta &= kt + \theta_0, \\ \sigma &= \sigma_0. \end{cases}$$

- For the distribution $\{X_3, X_4\}$, the equation of the integral curves is

$$\gamma'(t) = X_3(\gamma(t)) + k(t)X_4(\gamma(t)),$$
$$\gamma(0) = (x_0, y_0, \theta_0, \sigma_0).$$

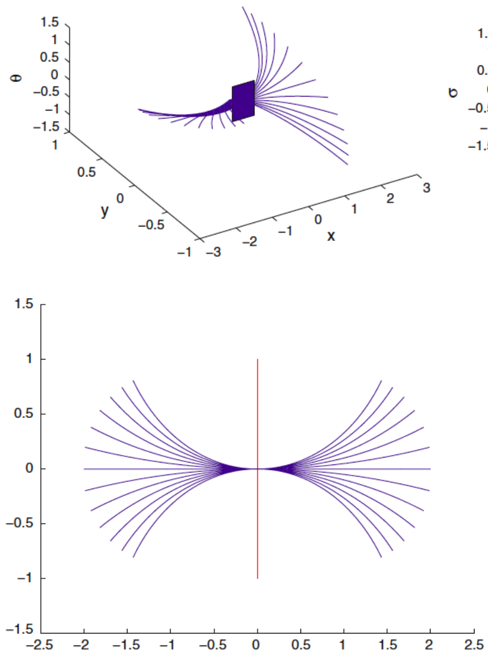
- For $k = \mathrm{cst}$, solutions are

$$\begin{cases} x &= -\frac{\sin(\theta_0)}{k}\mathrm{e}^{\sigma_0}\left(\mathrm{e}^{kt} - 1\right) + x_0, \\ y &= \frac{\cos(\theta_0)}{k}\mathrm{e}^{\sigma_0}\left(\mathrm{e}^{kt} - 1\right) + y_0, \\ \theta &= \theta_0, \\ \sigma &= kt + \sigma_0. \end{cases}$$

- The projections on the (x, y) plane are :

circles of radius $1/k$ tangent to the x-axis

lines <u>independent</u> of k through (x_0, y_0) and orthogonal to the direction θ_0 in the fixed "vertical" plane Span$\{ X_3, X_4 \}$.



- Another (logarithmic) model for the scale-space (zoom and blowing up). See Citti, Sarti, Petitot, *J. of Phys. Paris*, 103, 1-2, 2009.
- We take

$$\omega = \sigma^{-1}\left(-\sin(\theta)dx + \cos(\theta)dy\right)$$

- Then

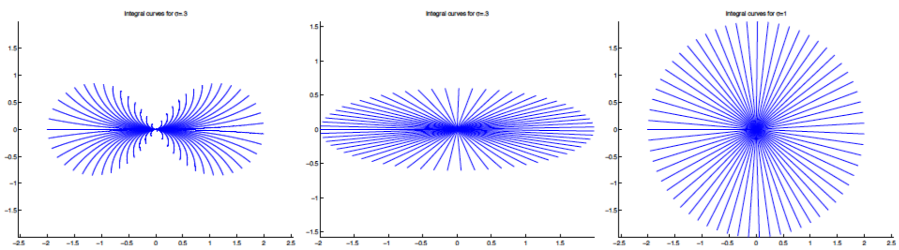
$$\begin{aligned} d\omega &= \sigma^{-1}\left(\cos(\theta)dx + \sin(\theta)dy\right) \wedge d\theta \\ &\quad + \sigma^{-2}\left(-\sin(\theta)dx + \cos(\theta)dy\right) \wedge d\sigma \\ &= \sigma^{-1}\omega_1 \wedge \omega_2 + \sigma^{-2}\omega_3 \wedge \omega_4 \end{aligned}$$

- The scaled tangent vectors are

$$X_1, X_2, \sigma X_3, \sigma X_4.$$

- For $\sigma = 0$ we get the contact subRiemannian geometry and for $\sigma = 1$ we get the Euclidean geometry.

$$\gamma'(t) = X_1(\gamma(t)) + k_2 X_2(\gamma(t)) + k_3 \sigma X_3(\gamma(t)) + k_4 \sigma X_4(\gamma(t))$$



Minimal surfaces in V1

- It seems that illusory contours are in fact boundaries of illusory minimal surfaces in V1.
- The theory of surfaces *S* in a contact manifold endowed with a sub-Riemannian geometry is rather difficult.
- There are in general “characteristic” (generically isolated) points where S is tangent to the contact plane and where the normal vector relative to $\mathcal{K}$ is not defined.
- See Scott Pauls : « Minimal surfaces in the Heisenberg group ».

Coherent states and harmonic analysis on Lie groups

- The natural context of signal analysis in natural vision is therefore that of *coherent states*. We have
  - An Hilbert space $\mathcal{H} = L^2(\mathbb{R}^2)$
  - A (locally compact) Lie group G acting on $\mathcal{H}$ via an irreducible unitary representation π.
  - A well localized « mother » wavelet $\varphi_0 \in \mathcal{H}$

74

- Coherent state = G-orbit $\{\varphi_g\}_{g\in G}$ of φ_0
- Harmonic analysis of a signal f :

$$f(x) = \int_G T_f(g)\, \varphi_g(x)\, d\mu(g)$$

- The transform of f is :

$$T_f(g) = \langle f, \varphi_g \rangle \in L^2(G)$$

- The Gabor transform corresponds to the analysis :

$$G_f(a,\omega) = \int_{\mathbb{R}} f(x) e^{-i\omega(x-a)} g(x-a)^* \, dx \in L^2(\mathbb{R}^2)$$

with the synthesis :

$$f(x) = \frac{1}{2\pi \|g\|^2} \int_{\mathbb{R}} G_f(a,\omega)\, e^{i\omega(x-a)} g(x-a)\, da d\omega.$$

- The coherent states are :

$$g_{a,\omega}(x) = e^{i\omega(x-a)} g(x-a)$$

75

- For classical wavelets, the coherent states are

$$\varphi_{a,s}(x) = \frac{1}{\sqrt{s}}\varphi\left(\frac{x-a}{s}\right)$$

and must satisfy the admissibility condition

$$c_\varphi = \int_R |\widehat{\varphi}(\xi)|^2 \frac{d\xi}{\xi} < \infty$$

- The synthesis is given by the Calderon identity with $T_f(a,s) = \langle f, \varphi_{a,s} \rangle$

$$f(x) = \frac{1}{c_\varphi} \int_{\mathbb{R}_+^* \times \mathbb{R}} T_f(a,s)\, \varphi_{a,s}(x) \frac{ds}{s} \frac{da}{s}$$

- Coherent states enable to represent a signal $f \in \mathcal{H}$ by its transform

$$T_f(g) = \langle f, \varphi_g \rangle \in L^2(G)$$

- It is what is done by V1, the $\langle f, \varphi_g \rangle$ being the measure of f by the receptive profiles φ_g.

76

Harmonic analysis and symmetry axis

- We can apply this to the mother wavelet

$$\varphi_{(0,\sigma)}(x,y) = \frac{1}{e^{2\sigma}} e^{\frac{-(x^2+y^2)}{e^{2\sigma}}} e^{\frac{2iy}{e^{\sigma}}}$$

and look at the associatated coherent state.

- Let C be a closed boundary in the retinal plane $\mathbb{R}^2$ and $a = (x, y)$ a point inside C.

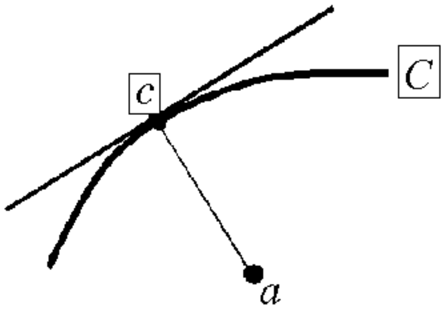
- Citti-Sarti : If we look at the maximal responses of the receptive profiles centered at a, and if c is the nearest point of C relative to a, then

$$d\left((x,y),c\right) = \frac{1}{\sqrt{2}} e^{\bar{\sigma}}$$

and $\bar{\theta}$ is the direction of C at c.

- We can therefore lift $\mathbb{R}^2$ to a surface Σ in G

$$\Sigma = \left\{(x, y, \bar{\theta}(x,y), \bar{\sigma}(x,y))\right\}$$



- The tangent vecteur over $a = (x, y)$

$$X_1 = e^{\bar{\sigma}} \left(\cos(\bar{\theta}) \partial_x + \sin(\bar{\theta}) \partial_y \right)$$

is parallel to C at c which is at <u>minimal</u> distance and therefore, as a derivative, satisfies

$$X_1 \left(\bar{\sigma} \right) = 0$$

- The tangent vector over $a = (x, y)$

$$X_3 = e^{\bar{\sigma}} \left(-\sin(\bar{\theta})\partial_x + \cos(\bar{\theta})\partial_y \right)$$

is orthogonal to C at c and $\bar{\theta}$ is constant along this direction. Therefore

$$X_3 \left(\bar{\theta} \right) = 0$$

- Now, the tangent plane to Σ: $T_{x,y,\bar{\theta},\bar{\sigma}}\Sigma$ is generated by the 2 vectors

$$\begin{cases} X_1 + X_1 \left(\bar{\theta} \right) X_2 + X_1 \left(\bar{\sigma} \right) X_4 \\ X_3 + X_3 \left(\bar{\theta} \right) X_2 + X_3 \left(\bar{\sigma} \right) X_4 \end{cases}$$

- But, since

$$X_1 \left(\bar{\sigma} \right) = X_3 \left(\bar{\theta} \right) = 0$$

it is in fact generated by

$$\begin{cases} X_1 + X_1 \left(\bar{\theta} \right) X_2 \\ X_3 + X_3 \left(\bar{\sigma} \right) X_4 \end{cases}$$

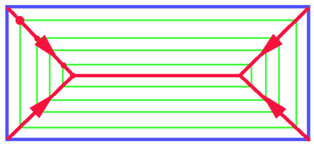
- As $d\omega = \upsilon \wedge d\theta + \omega \wedge d\sigma = \omega_1 \wedge \omega_2 + \omega_3 \wedge \omega_4$

we see that $d\omega$ vanishes on $T\Sigma$: Σ is therefore a *Lagrangian* submanifold of G.

79

- The transform of a closed contour C by this coherent state realizes the propagation of C via the *eikonal equation* of geometrical optics (Huyghens or « grassfire » model).
- The singular locus of this propagation is like the « symmetry axis » or « medial axis » whose role in vision has been strongly emphasized by many authors after Harry Blum : René Thom, David Marr, David Mumford, Steve Zucker, James Damon, Benjamin Kimia, etc.

- MA of a rectangle



- MA of an ellipse computed by A. Sarti using the coherent state

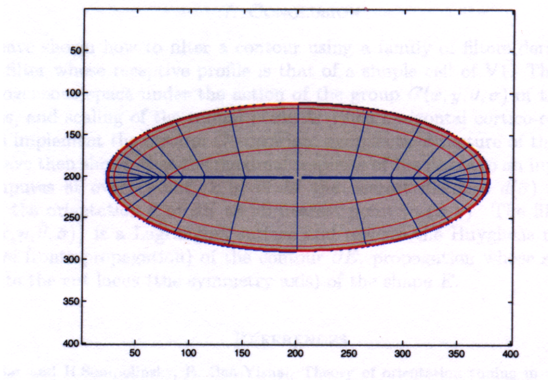


FIGURE 11. Level curves of $\bar{\theta}(x, y)$ (blue) and $\bar{\sigma}(x, y)$ (red).

Non-commutative harmonic analysis

- To a geometry with geodesics are associated diffusion / propagation processes.
- Classical heat kernel :

$$\frac{\partial f(\mathbf{x}, s)}{\partial s} = \Delta f(\mathbf{x}, s)$$

- Elementary solution in $\mathbb{R}^3$:

$$\Delta f(\mathbf{x}) = \delta(\mathbf{x}) \qquad -\frac{1}{4\pi \|\mathbf{x}\|}$$

- Solution of

$$\Delta f(\mathbf{x}) = u(\mathbf{x}) \qquad -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{u(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|} d\mathbf{y}$$

- Links with Euclidean geodesics. These are the projections on $\mathbb{R}^3$ of the bicharacteristics in the cotangent bundle.
- The Hamiltonian of the bicharacteristics is the symbol of the Laplacian (sum of the squares of the moments):

$$H(x_j, \xi_j) = \|\xi\|^2 = \sum_{j=1}^{j=3} \xi_j^2$$

81

- Hamilton equations :

$$\dot{x}_j(s) = \frac{\partial H}{\partial \xi_j} = 2\xi_j \text{ et } \dot{\xi}_j(s) = -\frac{\partial H}{\partial x_j} = 0.$$

- Solutions : $\xi_j = c_j \qquad x_j(s) = 2c_j s + d_j$

- If $\mathbf{x}(0) = 0 \qquad \mathbf{x}(\tau) = \overline{\mathbf{x}}$ then

$$x_j(s) = \frac{\overline{x}_j}{\tau} s, \ \xi_j(s) = \frac{\overline{x}_j}{2\tau}.$$

- Lagrangian (Legendre transform of H) :

$$\sum_{j=1}^{j=3} \xi_j \dot{x}_j - H(x_j, \xi_j) \qquad \sum_{j=1}^{j=3} (\xi_j \dot{x}_j - \xi_j^2)$$

$$L = \sum_{j=1}^{j=3} \left(\frac{\overline{x}_j}{2\tau} \frac{\overline{x}_j}{\tau} - \left(\frac{\overline{x}_j}{2\tau} \right)^2 \right) = \frac{\|\overline{\mathbf{x}}\|^2}{4\tau^2}$$

- Action integral along a geodesic :

$$S = \int_0^\tau L ds = \int_0^\tau \frac{\|\overline{\mathbf{x}}\|^2}{4\tau^2} ds = \frac{\|\overline{\mathbf{x}}\|^2}{4\tau}$$

82

- S is a solution of the Hamilton-Jacobi equation :

$$\frac{\partial S}{\partial \tau} + H\left(\mathbf{x}, \nabla S\right) = 0 \qquad \frac{\partial S}{\partial \tau} = -\left\|\nabla S\right\|^2$$

- Fundamental solution (heat kernel) :

$$P(\mathbf{x}, s) = \frac{1}{\left(2\sqrt{\pi s}\right)^3} e^{-\frac{\|\overline{\mathbf{x}}\|^2}{4s}}$$

- General solution :

$$f(\mathbf{x}, s) = \frac{1}{\left(2\sqrt{\pi s}\right)^3} \int_{\mathbb{R}^3} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4s}} u\left(\mathbf{y}\right) d\mathbf{y}$$

Harmonic analysis and SR geometry

- To understand correctly V1, we would have to correlate harmonic analysis and sub-Riemannian geometry, and in particular investigate the sub-elliptic Laplacian and the heat kernel.
- For the Heisenberg group, there are works of R. Beals, B. Gaveau, P. Greiner, D-Ch Chang.

- The problem is rather difficult since there are cut points in every neighborhood of each point and the classical analysis of heat equation fails at these singular points (B. Gaveau, IHP, 26-10-2005).

The sub-Riemannian Heisenberg case

- Gaveau, Beals, Greiner, Chang.
- The problem is complex because there exists a complicated cut locus.
- Coordinates (z, t) in $\mathbb{R}^3$. Heat equation :

$$\frac{\partial f(z,t,s)}{\partial s} = \Delta_{\mathcal{K}} f(z,t,s)$$

with the sub-Riemannian Laplacian.

- Heat kernel :

$$P(z,t,s) = \frac{1}{(2\pi s)^2}\int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} e^{\left(\frac{i\tau t}{s} - \left(\frac{\|z\|^2}{2s}\right)\frac{2\tau}{\tanh(2\tau)}\right)} d\tau$$
$$= \frac{1}{(2\pi s)^2}\int_{\mathbb{R}} V(\tau)\, e^{\left(-\frac{\Sigma(z,t,\tau)}{s}\right)} d\tau$$

with $V(\tau) = \frac{2\tau}{\sinh(2\tau)}$ $\Sigma(z,t,\tau) = -i\tau t + \|z\|^2 \frac{\tau}{\tanh(2\tau)}$

to be compared with

$$P(\mathbf{x}, s) = \frac{1}{(2\sqrt{\pi s})^3} e^{-\frac{\|\overline{x}\|^2}{4s}}$$

- As the action Σ is complex, P is an oscillatory integral if $t \neq 0$ (especially when $z = 0$).
- One must use techniques such as the stationary phase principle (semi-classical approximation).
- For $s \to 0$, the oscillatory integral

$$I(q,s) = \frac{1}{(2\pi s)^{p/2}} \int e^{i\frac{\varphi(q,\tau)}{s}} a(q,\tau,s) d\tau$$

concentrates on $\frac{\partial \varphi(q,\tau)}{\partial \tau} = 0$

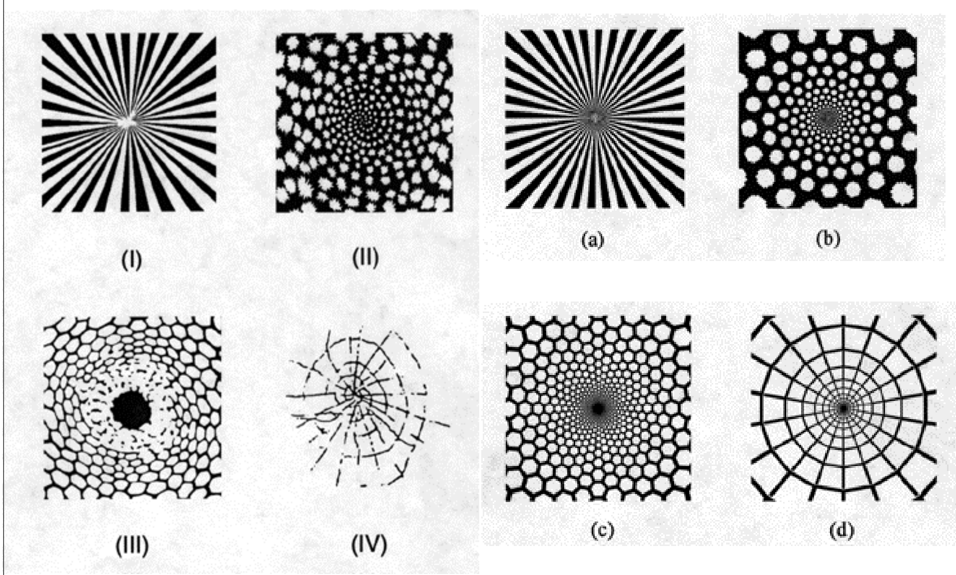
- Recently (2008), Andrei Agrachev, Ugo Boscain, Jean-Paul Gauthier and Francesco Rossi have found the heat kernel for $G = SE(2)$ (and all unimodular Lie groups with a left-invariant sub-Riemannian geometry).
- The hypoelliptic Laplacian is the sum of squares of the bracket generating Lie subalgebra :

$$\Delta_{\mathcal{K}} = X_1^2 + X_2^2$$

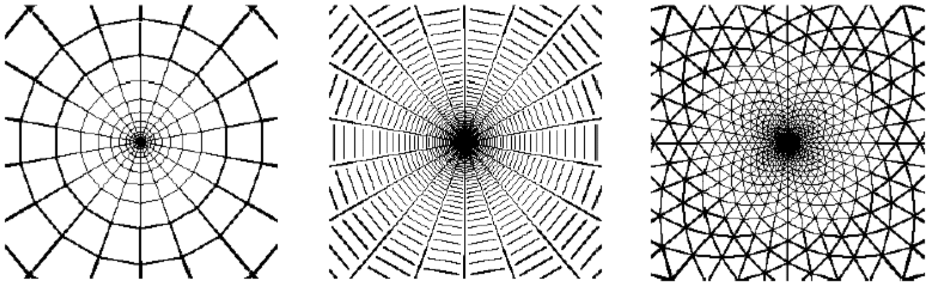
- They use the non-commutative generalized Fourier transform (GFT).

Application to spontaneous geometric visual patterns

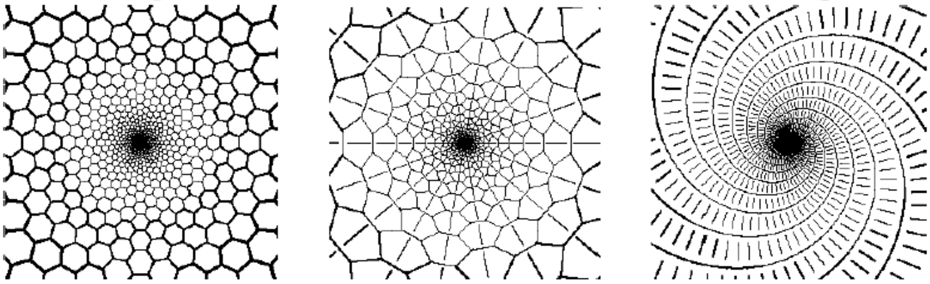
- A beautiful application of these models of functional architecture concerns entoptic vision (hallucinations).
- Paul Bressloff, Jack Cowan, Martin Golubitsky :
- by encoding the functional architecture of V1 into the Hopfield equations of a neural net, one is able to deduce visual morphological patterns.



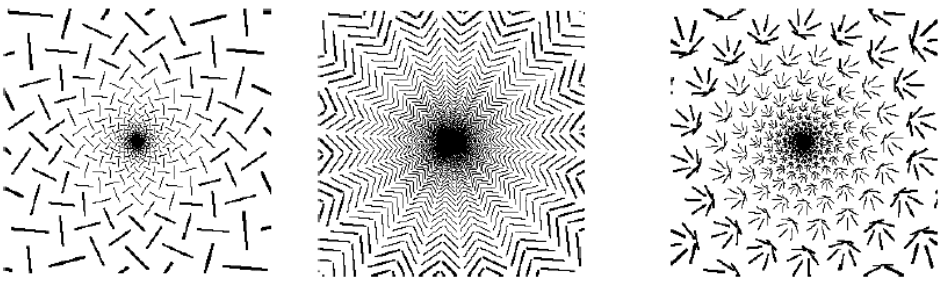
87



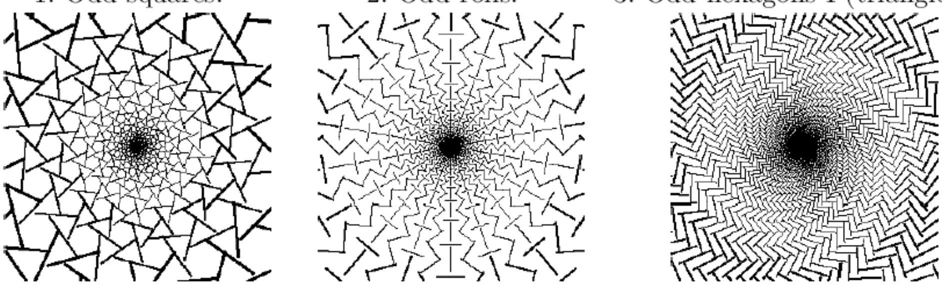
1. Even squares. 2. Even rolls. 3. Even hexagons I.



4. Even hexagons II. 5. Even rhombs. 6. Even rhombic rolls.



1. Odd squares. 2. Odd rolls. 3. Odd hexagons I (triangl



- Let $(\mathbf{x}, \theta)$ be local coordinates in the model V of V1.
- Let $a(\mathbf{x}, \theta, t)$ be the activity of V1. We look for the PDE governing the evolution of a.
- Using standard Hopfield equations for neural nets, we get :

$$\frac{\partial a(\mathbf{x},\theta,t)}{\partial t} = -\alpha a(\mathbf{x},\theta,t) + \frac{\mu}{\pi}\int_0^\pi \int_{\mathbb{R}} w\left\langle \mathbf{x},\theta|\mathbf{x}',\theta'\right\rangle \sigma\left(a(\mathbf{x}',\theta',t)\right) d\mathbf{x}'d\theta' + h(\mathbf{x},\theta,t)$$

where σ is a non linear gain function (with $\sigma(0) = 0$), h an external input and

$$w\left\langle \mathbf{x},\theta|\mathbf{x}',\theta'\right\rangle$$

is the weight of the connection between the neuron $v = (\mathbf{x}, \theta)$ and the neuron $v' = (\mathbf{x}', \theta')$, α a parameter of decay (α can be taken = 1) and μ a parameter of excitability of V1.

- The increasing of μ models an increasing of the excitability of V1 due to the action of substances on the nuclei which produce specific neurotransmitters (such as serotonin or noradrenalin).

Encoding the functional architecture into the synaptic weights

- Bressloff et al. encode only the strictly coaxial alignements. Here again, it is the simplest model.
- The local vertical connections inside a single hypercolumn yield a term:

$$w\left\langle \mathbf{x},\theta|\mathbf{x}',\theta'\right\rangle = w_{\mathrm{loc}}\left(\theta-\theta'\right)\delta\left(\mathbf{x}-\mathbf{x}'\right)$$

where δ is a Dirac function imposing

$$\mathbf{x} = \mathbf{x}'$$

- The lateral horizontal connections between different hypercolumns yield a term:

$$w\left\langle \mathbf{x},\theta|\mathbf{x}',\theta'\right\rangle = w_{\text{lat}}\left(\mathbf{x}-\mathbf{x}',\theta\right)\delta\left(\theta-\theta'\right)$$

where the factor

$$\delta\left(\theta-\theta'\right)$$

imposes $\theta = \theta'$ and expresses the fact that the horizontal cortico-cortical connections connect parallel pairs.

- Moreover, the coaxiality $\theta = \theta' = \mathbf{xx}'$ is expressed by the fact that

$$w_{\text{lat}}\left(\mathbf{x}-\mathbf{x}',\theta\right) = w_{\text{lat}}\left(s\right)\delta\left(\mathbf{x}-\mathbf{x}'-se_{\theta}\right) = \widehat{w}\left(r_{-\theta}\left(\mathbf{x}-\mathbf{x}'\right)\right)$$

where e_{θ} is the unit vector in the direction θ.

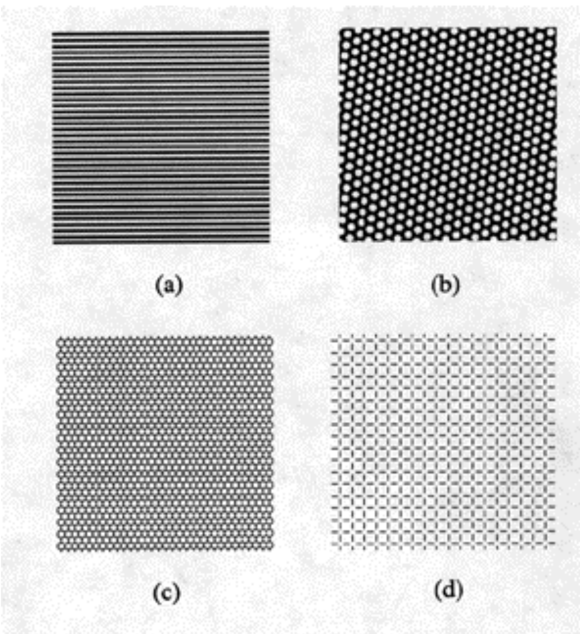
- As the weights w are $E(2)$-invariant, the PDE is itself $E(2)$-equivariant if $h = 0$.

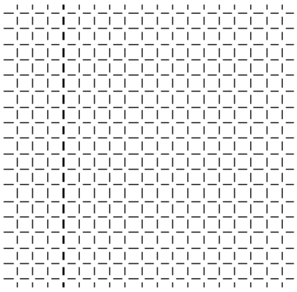
Dynamically emerging morphologies and bifurcations

- We suppose that there exist no external input, that is $h = 0$. For $\mu = 0$, the state $a \equiv 0$ is trivially the state of the network and it is stable.
- $a \equiv 0$ is the "ground state". It can be very complex (endogeneous activity, spontaneous noise, etc.)

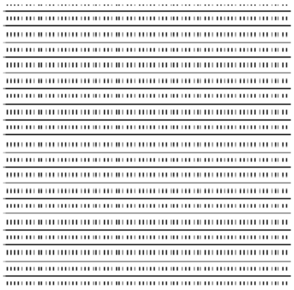
- Now, the analysis of the PDE shows that, as the parameter μ increases, this initial activation state $a \equiv 0$ can become unstable and bifurcate for critical values μ_c of μ.

- The new stable activation states present spatial patterns generated by an $\mathbf{E}(2)$ symmetry breaking.
- The bifurcations can be analyzed using classical methods:
  - Linearization of the PDE near the solution $a \equiv 0$ and the critical value μ_c.
  - Spectral analysis of the linearized equation.
  - Computation of its eigenvectors (eigenmodes).
  - Hypothesis of periodicity w.r.t. a lattice of R.
- Here are some examples of eigenmodes.

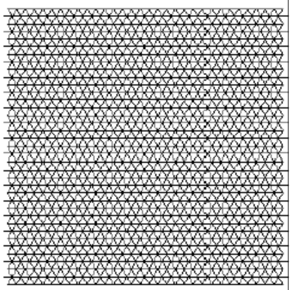




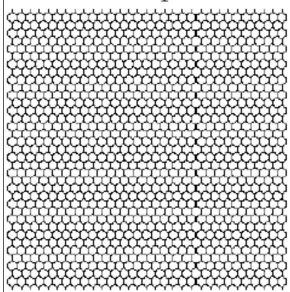
1. Even squares.



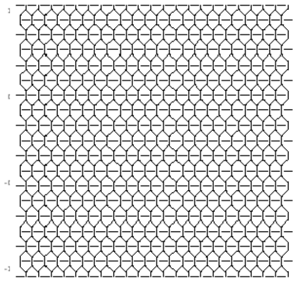
2. Even rolls.



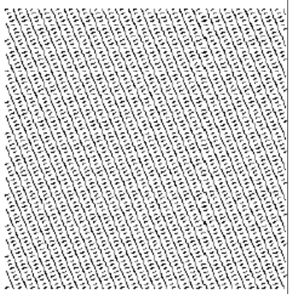
3. Even hexagons I.



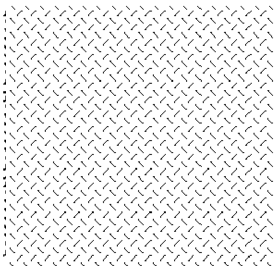
4. Even Hexagons II.



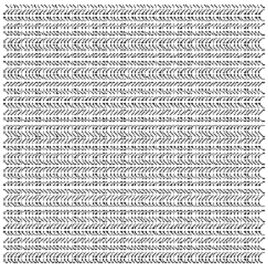
5. Even rhombs.



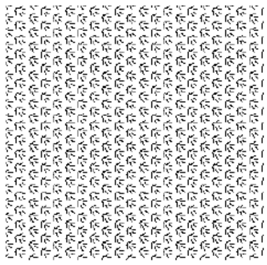
6. Even rhombic rolls.



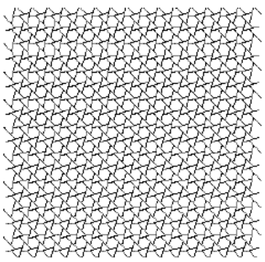
1. Odd squares.



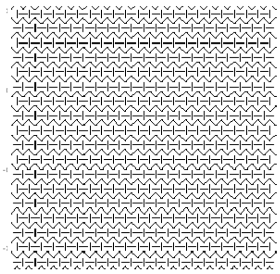
2. Odd rolls.



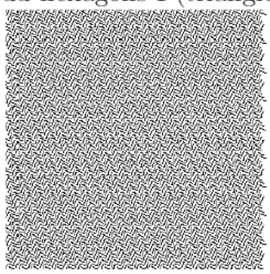
3. Odd hexagons I (triangles).



4. Odd hexagons II.



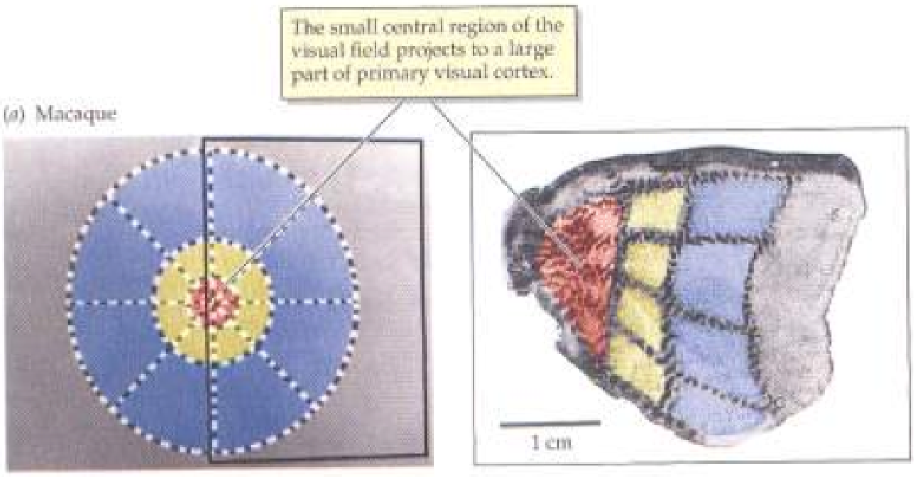
5. Odd rhombs.



6. Odd rhombic rolls.

Patterns as virtual retinal images

- The last step is to reconstruct from eigenmodes in V1 corresponding virtual retinal images.
- For that, we must take into account the retinotopic conformal map mapping the retina R on V1.
- See Jeanny Herault's talk.



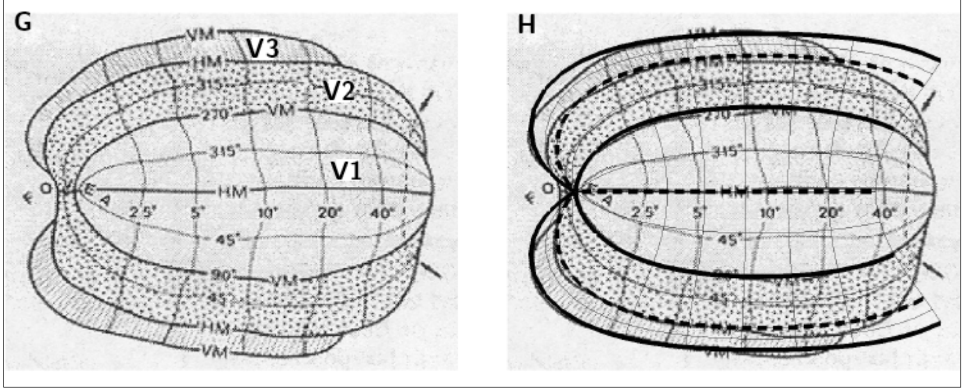
95

- A good model is a wedge-dipole model for V1, V2, and V3

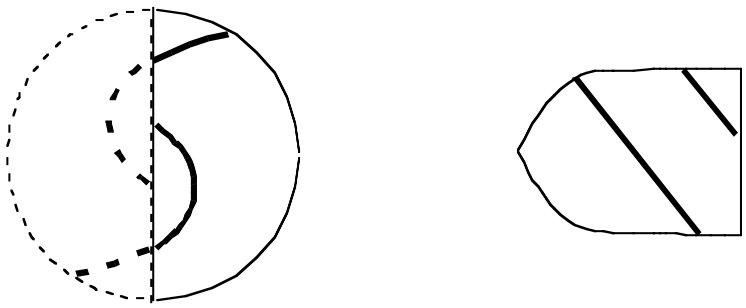
$$\mathrm{Log}[(w(z)+a)/(w(z)+b)]$$

where $w(z)$ wedges the argument.

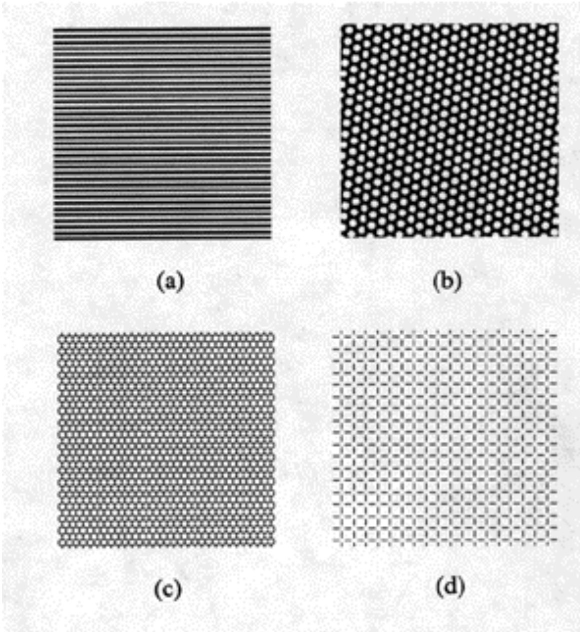
- Left (G) : V1-V2-V3 (Horton & Hoyt 1991).
- Right (H) : fit with a wedge-dipole model (Schwartz 2002).

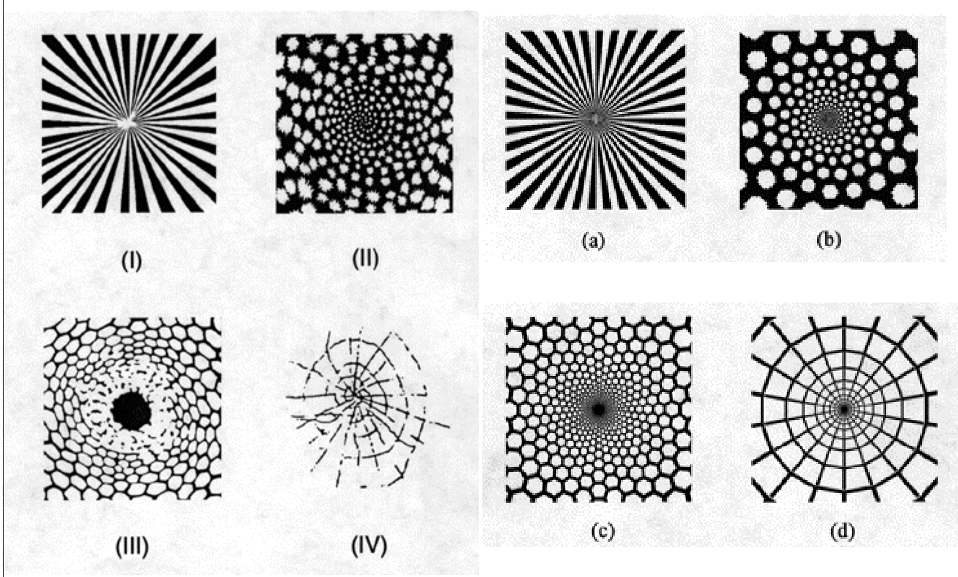


- Lines in V1 correspond qualitatively to spiral on the retina.



- If we apply the inverse of the conformal map to the eigenstates of the PDE (as if V1 activity was induced by a real stimulus) we get quite exact models of Klüver's planforms.





98

Open problems

- Many problems concerning the modeling of experimental results are open. For instance :
  - The very strong correlations between orientation and other variables : spatial frequency (varying along pinwheels rays), phase (varying inside columns), ocular dominance.
  - To take into account binocular vision, stereopsis, depth and 3D shapes.
  - To take into account the subthreshold activities (integration field) and the responses to natural stimuli (not too simple gratings).

99

// Ugo BOSCAIN
SISSA / Italie
http://www.cmapx.polytechnique.fr/~boscain/

"Anthropomorphic Image Reconstruction via Hypoelliptic Diffusion"

We present a model of geometry of vision which generalizes one due to Petitot, Citti and Sarti.

1) One of the main features is that the primary visual cortex V1 lifts the image from R2 to the bundle of directions of the plane P T R2 = R2 × P 1.

2) In this model a corrupted image is reconstructed by minimizing the energy necessary to activate the orientation columns corresponding to regions in which the image is corrupted.

3) The minimization process gives rise to an hypoelliptic heat equation on PTR^2. The hypoelliptic heat equation is studied using generalized Fourier transform.

Références :

Ugo Boscain, Jean Duplaix, Jean-Paul Gauthier, Francesco Rossi, Anthropomorphic image reconstruction via hypoelliptic diffusion. http://arxiv.org/abs/1006.3735

Ugo Boscain, Grégoire Charlot, Francesco Rossi "Existence of planar curves minimizing length and curvature" , to appear in Journal of Mathematical Sciences.

I. Moiseev, Yu. L. Sachkov, Maxwell strata in sub-Riemannian problem on the group of motions of a plane, ESAIM: COCV, to appear

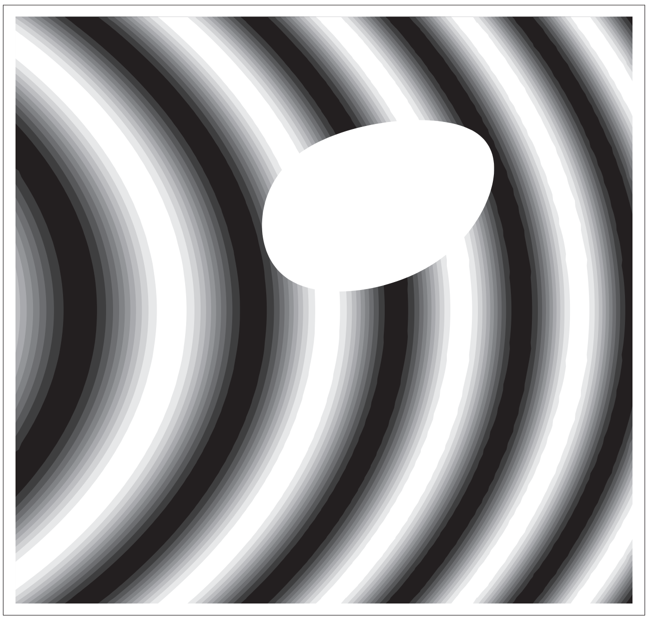
Anthropomorphic image reconstruction via hypoelliptic diffusion

Ugo Boscain (CNRS, CMAP, Ecole Polytechnique, Paris, France)

Jean-Paul Gauthier (Toulon)
Francesco Rossi (Marseille)

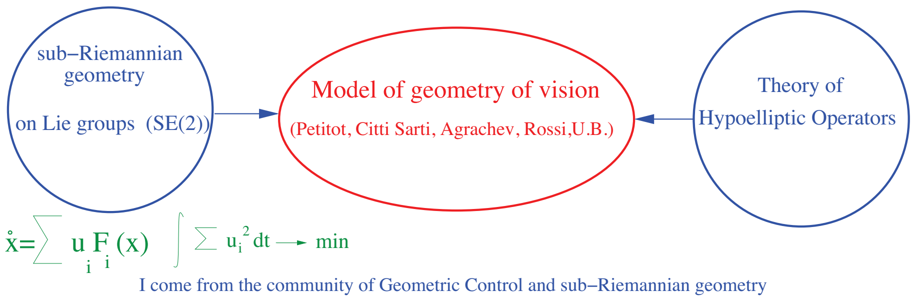
October 4, 2010

The Hole



The two fishes of Kaniza

An algorithm of image reconstruction based on a model of geometry of vision



At the moment our anthropomorphic algorithm does not work better than existing algorithms based on

1. segmentation
2. wavelets
3. exemplar-based methods

But it can be useful:
→in combination with exemplar-based methods
→to validate/invalidate the biological model (via comparison with psychological experiments)

The model

This model was originally introduced by

- Jean Petitot (CREA, Paris)

then refined by

- Giovanna Citti (Bologna), Alessandro Sarti (CNRS, Paris)
- Andrei Agrachev (SISSA, Trieste, Italy)
- Gregoire Charlot (Grenoble), Jean-Paul Gauthier (Toulon), Francesco Rossi (Marseille) and U.B.

I will also talk about the work by

- Yuri Sachkov (Pereslav) and Igor Moiseev

Funding

→ERC StG 2009 GeCoMethods (EU)

→PRIN 2008 ”A geometric approach to controlled dynamics and applications” (IT)

→ANR (FR)

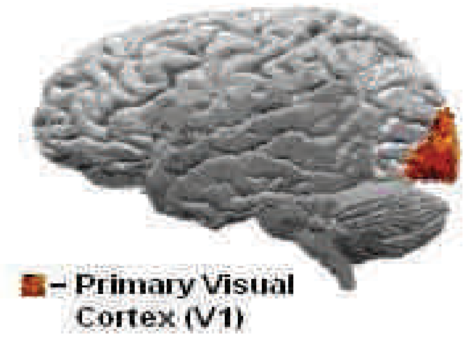
GCM

→DIGITEO “CONGEO” (FR)

→GDRE (INDAM/CNRS) “CONEDP” (IT-FR)

→ESF “OPTPDE” (EU)

two ideas coming from neurophyology of the visual cortex V1

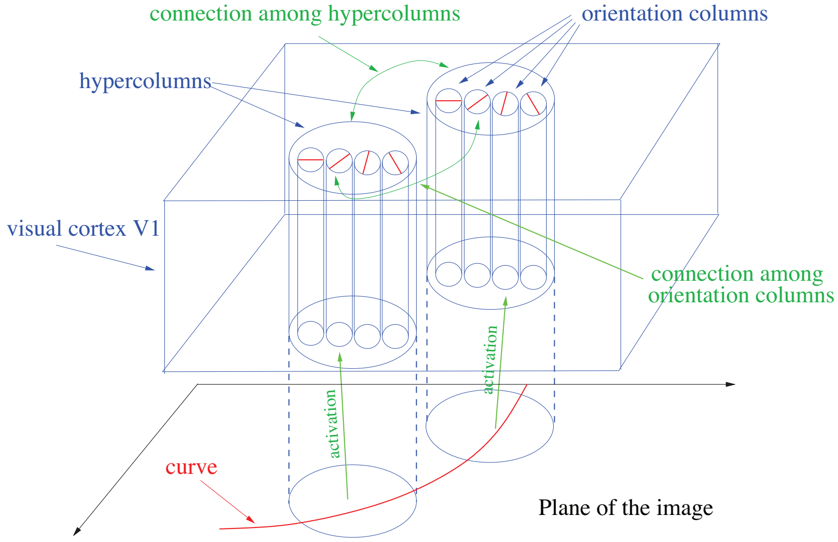


A. In the visual cortex V1, groups of neurons are sensible to both positions and directions. Hence the visual cortex lifts an image on the $PT\mathbf{R}^2 = \mathbf{R}^2 \times P^1$.

B. an image is reconstructed by minimizing the energy necessary to excite groups of neurons that are not excited by the image in $PT\mathbf{R}^2$.

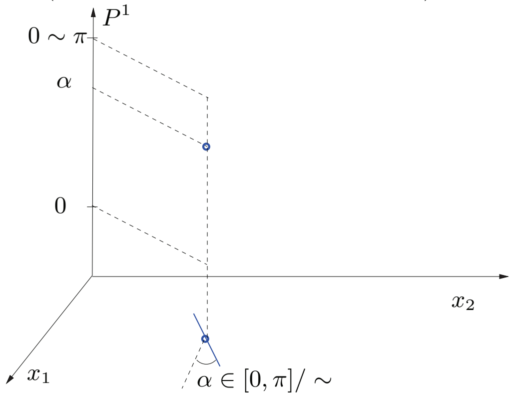
A1. The lift in $PT\mathbf{R}^2$

Hubel and Wiesel (Nobel prize 1981) observed that there are (groups of) neurons sensitive to the direction.



A2. The lift in $PT\mathbf{R}^2$

- the brain stores an image as a set of points and tangent directions. i.e. it makes a lift to $PT\mathbf{R}^2 = \mathbf{R}^2 \times P^1$. The projective tangent bundle of $\mathbf{R}^2$ (or bundle of direction of the plane).



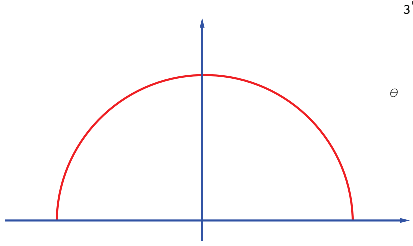
- $PT\mathbf{R}^2$ can be seen as a fiber bundle whose base is $\mathbf{R}^2$ and whose fiber at the point (x_1, x_2) is the set of straight lines (i.e. directions without orientation) $\mathbf{d}_{(x_1,x_2)}$ attached to (x_1, x_2).

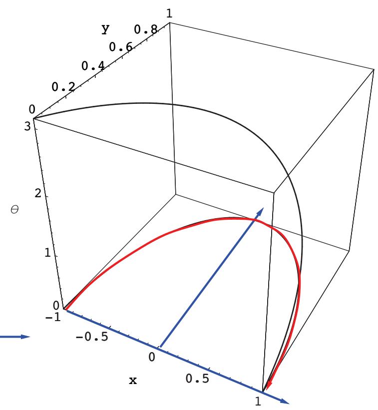
A23 The lift of a curve

$(x_1(t), x_2(t))$ curve in $\mathbf{R}^2$, $\xrightarrow{lift}$ $(x_1(t), x_2(t), \alpha(t))$, curve in $\mathbf{R}^2 \times P^1$

$\alpha(t) = \arctan\left(\frac{\dot{x}_2(t)}{\dot{x}_1(t)}\right) \in [0, \pi]/\sim$

Example: $(\cos(t), \sin(t))$:





$\rightarrow$every $\mathcal{C}^1$ submanifold of $\mathbf{R}^2$ has a lift.

$\rightarrow$not all curves in $\mathbf{R}^2 \times P^1$ are lift of planar curves

A3. Which curves are lift of planar curves?

A curve in $(x_1(t), x_2(t), \alpha(t))$ in $\mathbf{R}^2 \times P^1$ is the lift of a planar curve if

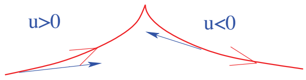
$$\alpha(t) = \arctan\left(\frac{\dot{x}_2(t)}{\dot{x}_1(t)}\right) \leftrightarrow \begin{cases} \dot{x}_1(t) = u(t)\cos(\alpha(t)) \\ \dot{x}_2(t) = u(t)\sin(\alpha(t)) \\ \dot{\alpha}(t) = v(t) \end{cases}, \quad \alpha \in [0,\pi]/\sim \qquad (1)$$

i.e. writing $x = (x_1, x_2, \alpha)$ if

$$\dot{x} = uX_1 + vX_2, \quad X_1 = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

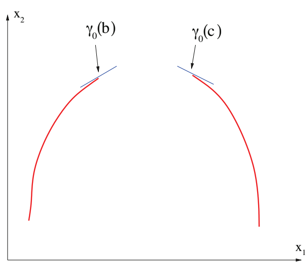
Projective Reed-Shepp car

Remark: If $\theta \in [0, 2\pi]\backslash \sim$ then we have to assume $u \geq 0$ otherwise α is not the angle of $(\dot{x}_1, \dot{x}_2)$



$\rightarrow$but this choice gives problems for existence of minimizers

B1. How V1 reconstruct an interrupted curve



Consider a smooth curve $\gamma_0 : [a, b] \cup [c, d] \to \mathbf{R}^2$, interrupted in $[b, c]$. We want to complete γ_0 by a curve $\gamma : [b, c] \to \mathbf{R}^2$ that is:

- $\gamma(b) = \gamma_0(b)$, $\gamma(c) = \gamma_0(c)$
- $\dot{\gamma}(b) \sim \dot{\gamma}_0(b) \neq 0$, $\dot{\gamma}(c) \sim \dot{\gamma}_0(c) \neq 0$.
- we assume $\gamma(b) \neq \gamma(c)$, $\dot{\gamma}_0(b) \neq 0$, $\dot{\gamma}_0(c) \neq 0$

B1. What to minimize?

there are many models describing situations where the brain takes the decision that minimize some cost (which can be external or internal to the brain)
→when moving objects with the hands, the brain minimizes a compromise between energy and the stress of muscles (external cost).
→For reconstruction of images, the (internal) minimized cost is the energy necessary to activate neurons that are not naturally activated by the image.
Given a neuron that is already active, it is easy to make activation of neurons that are:
-) close to it,
-) sensitive to a similar direction.
i.e. close in $\mathbf{R}^2 \times P^1$.

The most natural cost for lift of planar curves on $PT\mathbf{R}^2$

Riemannian length:

$$\int_b^c \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \beta^2 \dot{\alpha}^2}\ ds = \int_b^c \sqrt{u^2 + \beta^2 v^2}\ ds \to \min$$

on all curves in PTR^2 that are lift of planar curves (non-holonomic constraint)

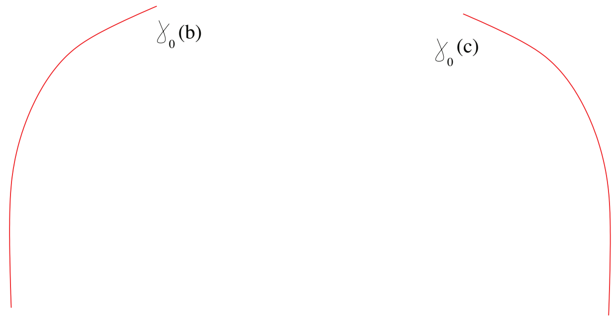
Then we get a problem of sub-Riemannian geometry (on $PT\mathbf{R}^2$):

$$\dot{x} = uX_1 + vX_2, \quad \int_b^c \sqrt{u^2 + \beta^2 v^2} ds \to \min \quad \sim \int_b^c \left(u^2 + \beta^2 v^2\right) ds \to \min,$$

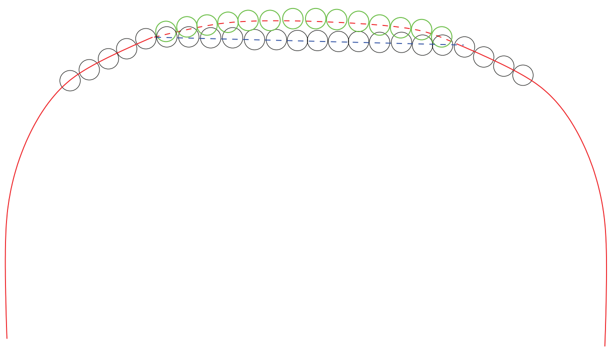
$$x = (x_1, x_2, \alpha), \quad X_1 = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

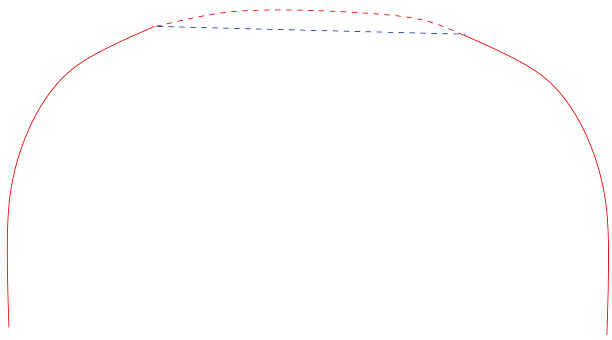
$$\dim\left(\mathrm{span}_x\{X_1, X_2, [X_1, X_2]\}\right) = 3$$

initial and final positions are fixed in $PT\mathbf{R}^2$.



Consider a smooth curve $\gamma_0 : [a, b] \cup [c, d] \to \mathbf{R}^2$, interrupted in $[b, c]$.





Advantages of this cost

1) it is the only sub-Riemannian cost on $PT\mathbf{R}^2$ invariant by rototranslations of the plane (observed by Agrachev)

2)

$$\int_b^c \sqrt{u^2+\beta^2 v^2}\,ds \to \min \quad \sim \quad \int_b^c \big(\underset{\substack{\uparrow \\ \textit{connec. among} \\ \textit{hypercolumns}}}{u^2} + \underset{\substack{\uparrow \\ \textit{connnec. among} \\ \textit{orient.columns}}}{\beta^2 v^2}\big)\,ds \to \min$$

good model for the energy necessary to activate orientation columns which are not directly activated by the image

3) The factor β can be eliminated with the transformation $(x, y) \to (\beta x, \beta y)$, i.e. by a "dilation of the initial conditions"

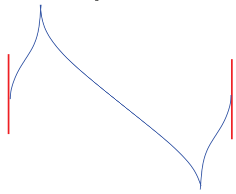
4) It is a compromise between length and curvature:

$$\int_b^c \sqrt{\dot{x}_1^2+\dot{x}_2^2+\beta^2\dot{\alpha}^2}\ ds = \int_b^c \sqrt{u^2+\beta^2 v^2}\ ds = \int_b^c \sqrt{\|\dot{\gamma}\|^2+\beta^2\|\dot{\gamma}\|^2 K^2}\ ds$$

5) there is existence of minimizers in the natural functional space

$$\mathcal{D} := \{\gamma \in \mathcal{C}^2([b,c],\mathbf{R}^2)\ \mid\ \sqrt{\|\dot{\gamma}(t)\|^2+\beta^2\|\dot{\gamma}(t)\|^2 K_\gamma^2(t)} \in L^1([b,c],\mathbf{R}),$$
$$\gamma(b)=x_0,\ \gamma(c)=x_1,\ \dot{\gamma}(b)\approx v_0,\ \dot{\gamma}(c)\approx v_1\}. \qquad (2)$$

($\sim$ for the optimal control formulation to have $u,v \in L^1$)
$\rightarrow$minimizers are analytic functions with no more than two cusps

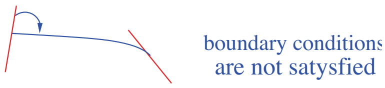


6) Since it is a sub-Riemannian cost, there is a natural hypoelliptic diffusion equation that can be used to reconstruct images

Other costs

In all problems of image reconstruction the curve is reconstructed by minimizing a compromise between length and curvature + invariance by $SE(2)$.
Since $K(t)=\frac{\dot{x}_1\ddot{x}_2-\dot{x}_2\ddot{x}_1}{(\dot{x}_1^2+\dot{x}_2^2)^{3/2}}$ it is natural to require $\gamma \in \mathcal{C}^2$, $\dot{\gamma}\neq 0$,

- Elastica type functionals $\int_0^{L(\gamma)} K^2(\gamma)\ ds$ (Mumford, Cao, Gousseau, Masnou, Pérez, Coope, Linnér)
- $\int_0^{L(\gamma)}(1+K^2(\gamma))\ ds$ (Mumford,Bellettini, Linnér). Minimizers must be studied in sophisticated functional spaces and may present angles (the curvature becomes a measure).
  $\rightarrow$also nonisotropic diffusion (Duits, Franken)
- a model due to Citti and Sarti $\int_0^{L(\gamma)} \sqrt{1+K^2(\gamma)}\ ds$ (based on certain neuro-physiological observations)
  $\rightarrow$no existence (G. Charlot, F. Rossi, U.B., 2010). There exists minimizing sequences whose limit is:



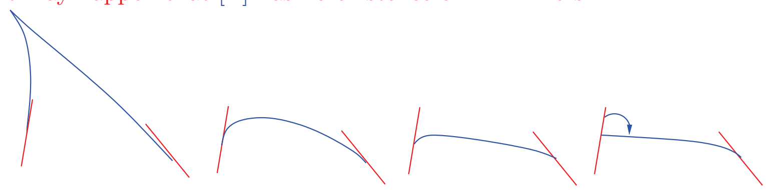
The same happens with our cost and $\alpha \in [0,2\pi]\backslash \sim$ and $u \geq 0$.

[A] $\int_0^T \sqrt{\|\dot\gamma(t)\|^2 + \beta^2\|\dot\gamma(t)\|^2 K(t)^2}\, dt$, [$B$] $\int_0^{L(\gamma)} \sqrt{1+K^2(\gamma)}\, ds$

→if $\dot\gamma \neq 0$ it is just a change of time $\int_0^t \|\dot\gamma(t')\| dt' = s \Rightarrow \|\dot\gamma\| dt = ds$.

→For initial and final conditions for which [A] has no cusps then minimizers of [A] are minimizers of [B].

→For initial and final conditions for which [A] has a cusp then it may happen that [B] has no existence of minimizers



minimizer of [A] minimizing sequence of [B]

→In any case "stationary points of [B] have no cusps. Maybe the right cost is [B] and one has to look only for stationary points.
(but I do not know if there are stationary points without cusps for every initial and final condition)

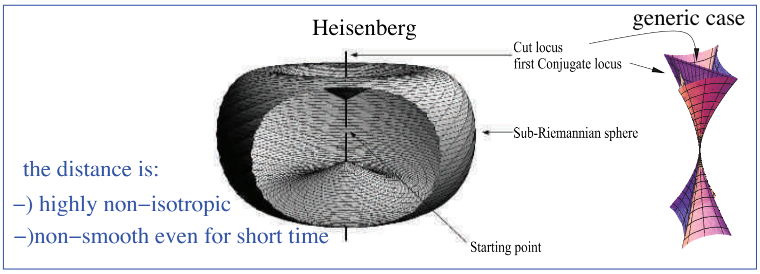
Plan

→reconstruction of curves

→reconstruction of images

Computation of optimal trajectories for curve-reconstruction:

- compute candidate optimal trajectories with the Pontryagin Maximum Principle (they can be computed explicitly in terms of elliptic functions)
- evaluate their optimality (very difficult point)
- the local behavior of optimal trajectory is very complicated (even more complicated than the Heisenberg group)



Pontryagin Maximum Principle

Consider a control problem of the type

$$\dot{x} = uX_1 + vX_2, \qquad \int_0^T (u^2 + v^2)\, ds \to \min$$

where $x \in M^3$, and $\dim(\mathrm{Span}\{X_1, X_2, [X_1, X_2]\}) = 3$. Then optimal trajectories are solutions of the Hamiltonian system (called geodesics):

$$H = \frac{1}{2}(\langle P, X_1\rangle^2 + \langle P, X_2\rangle^2)$$

corresponding to a level set $H = cost(T) \neq 0$.

→In the following I fix $H = \frac{1}{2}$ (arclength)
→the initial covector $P(0)$ parameterizes the geodesics

In our case

$$x = (x_1, x_2, \alpha), \quad P = (P_1, P_2, P_3), \quad X_1 = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Hence with $P_1 = P_r \cos P_a$, $P_2 = P_r \sin P_a$, we have

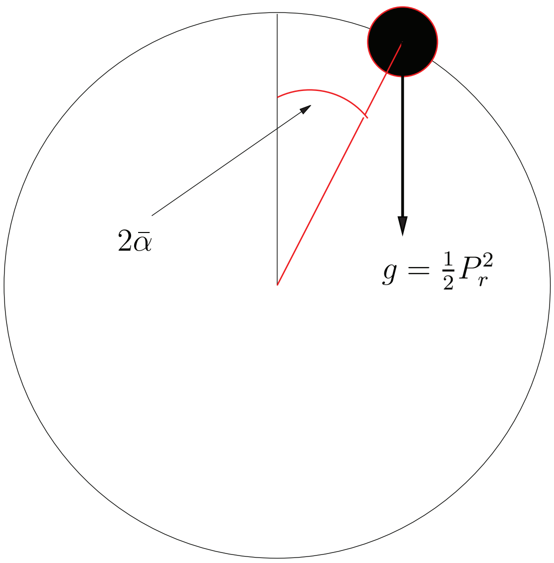
$$H = \frac{1}{2}\left((P_1 \cos\alpha + P_2 \sin\alpha)^2 + P_3^2\right) = \frac{1}{2}\left((P_r^2 \cos^2(\alpha - P_a) + P_3^2\right)$$

$$\begin{aligned}
\dot{x}_1 &= \tfrac{\partial H}{\partial P_1} = P_r \cos(\alpha - P_a)\cos\alpha & \dot{P}_1 &= -\frac{\partial H}{\partial x_1} = 0 \\
\dot{x}_2 &= \tfrac{\partial H}{\partial P_2} = P_r \cos(\alpha - P_a)\sin\alpha & \dot{P}_2 &= -\frac{\partial H}{\partial x_2} = 0 \\
\dot{\alpha} &= \tfrac{\partial H}{\partial P_3} = P_3 & \dot{P}_3 &= -\frac{\partial H}{\partial \alpha} = \frac{1}{2}P_r^2 \sin(2(\alpha - P_a))
\end{aligned}$$

Pendulum equation $\ddot{\alpha} = \frac{1}{2}P_r^2 \sin(2(\alpha - P_a))$

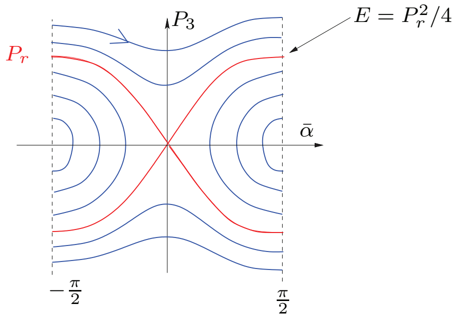
Pendulum equation

$\rightarrow$Setting $\bar{\alpha} = \alpha - P_a$, we get $\ddot{\bar{\alpha}} = \frac{1}{2}P_r^2 \sin(2\bar{\alpha})$



Energy of the Pendulum

We have the constant of the motion $E = T + V = \frac{1}{2}\dot{\bar{\alpha}}^2 + \frac{P_r^2}{4}\cos(2\bar{\alpha})$



$$\dot{x}_1 = P_r \cos\bar{\alpha}(t)\cos(\bar{\alpha}(t) + P_a)$$
$$\dot{x}_2 = P_r \cos\bar{\alpha}(t)\sin(\bar{\alpha}(t) + P_a)$$

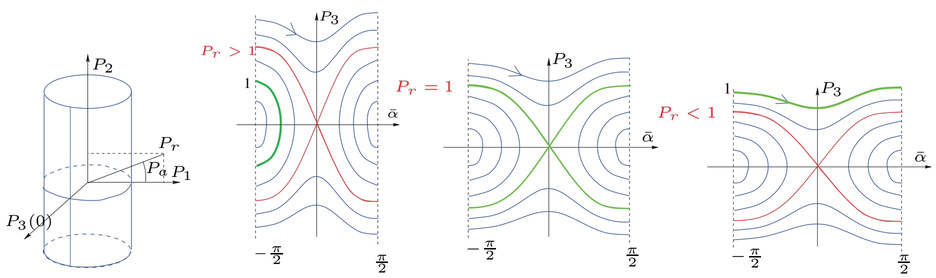
$\rightarrow \bar{\alpha} = \pm\pi/2$ are CUSPS

Initial Conditions

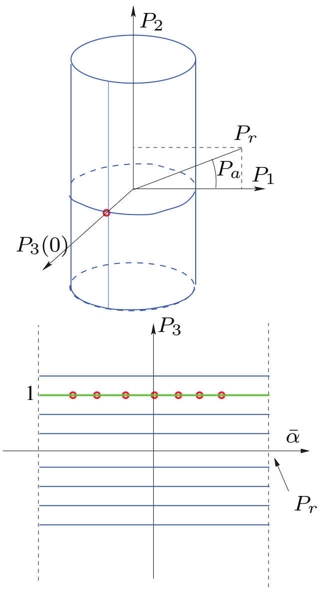
$$x(0) = 0, \quad y(0) = 0, \quad \alpha(0) = 0 \Rightarrow \bar{\alpha}(0) = P_a \tag{3}$$

Geodesics are parameterized by P_1, P_2, P_3 or (P_r, P_a, P_3) but we have the condition that

$\frac{1}{2} = H(0) = \frac{1}{2}(P_1^2 + P_3^2) = \frac{1}{2}(P_r^2\cos^2 P_a + P_3^2)$ (parameterization with velocity 1)



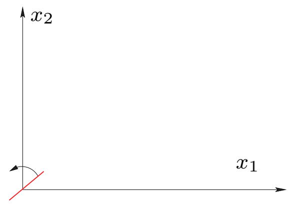
Example 00: $P_r = 0$



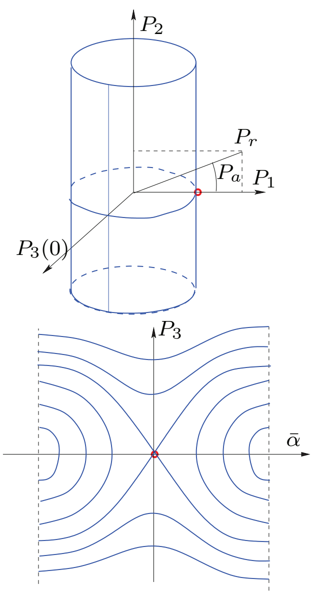
$P_r^2 \cos^2 P_a + P_3^2 = 1$
$P_r = 0, \quad P_3(0) = \dot{\alpha}(0) = 1$

$\ddot{\bar{\alpha}}(t) = \frac{1}{2} P_r^2 \sin(2\bar{\alpha}) = 0$

$\alpha(t) = P_a + \bar{\alpha}(t) = t$
$x_1(t) = 0$
$x_2(t) = 0$



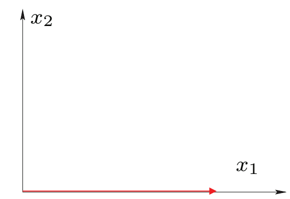
Example 0: $P_r = 1$



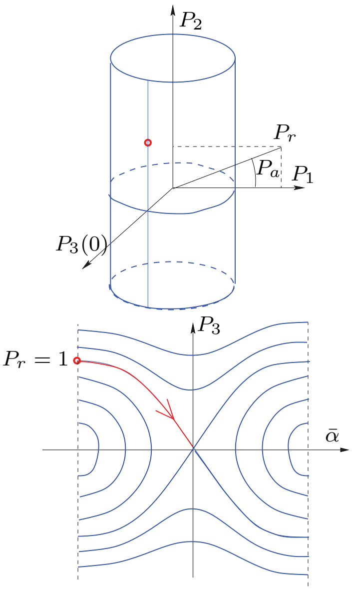
$P_r^2 \cos^2 P_a + P_3^2 = 1$
$P_a = 0, \quad P_r = 1, \quad P_3(0) = \dot{\alpha}(0) = 0$

$\ddot{\bar{\alpha}} = \frac{1}{2} \sin(2\bar{\alpha})$

$\alpha(t) = \bar{\alpha}(t) + P_a = 0$
$x_1(t) = t$
$x_2(t) = 0$

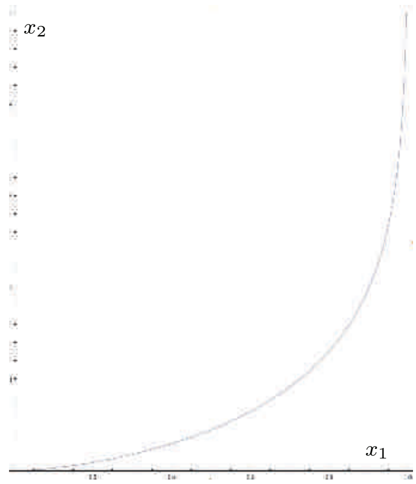


Example 1: $P_r = 1$

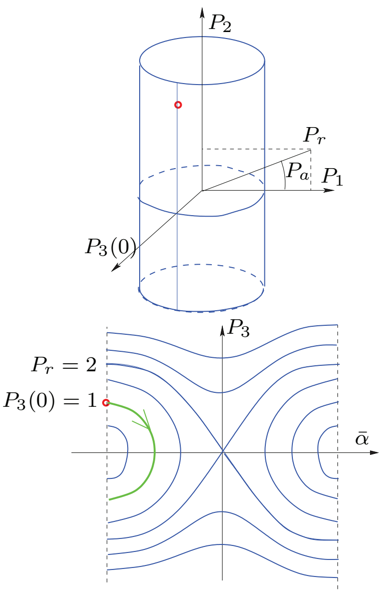


$P_r^2 \cos^2 P_a + P_3^2 = 1$

$P_a = \frac{\pi}{2}, \quad P_r = 1, \quad P_3(0) = \dot{\alpha}(0) = 1$

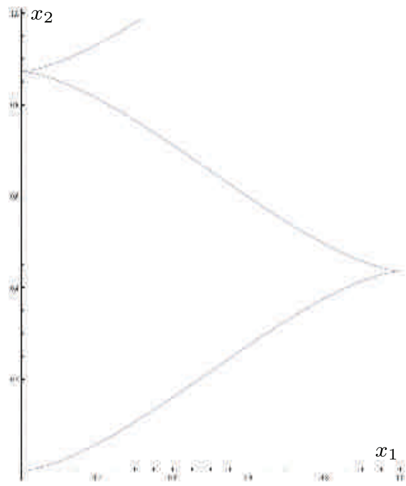


Example 2: $P_r = 2$

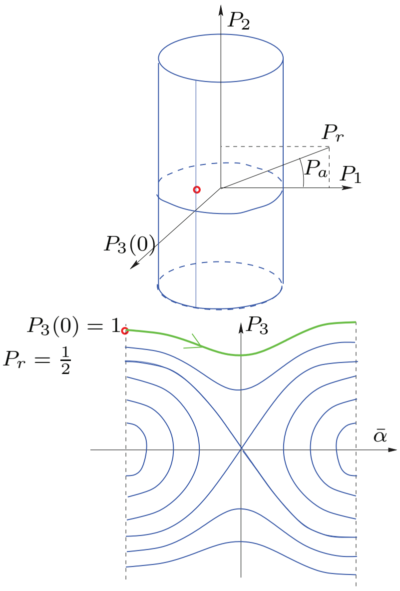


$P_r^2 \cos^2 P_a + P_3^2 = 1$

$P_a = \frac{\pi}{2}, \quad P_r = 2, \quad P_3(0) = \dot{\alpha}(0) = 1$

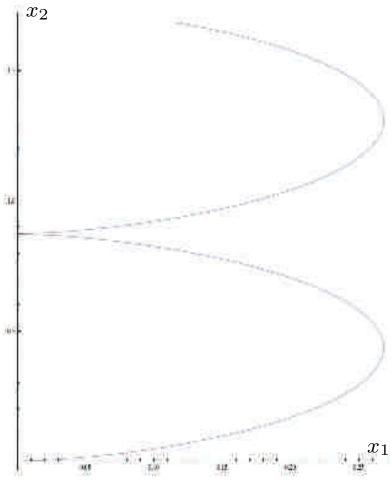


Example 3: $P_r = 1/2$



$P_r^2 \cos^2 P_a + P_3^2 = 1$

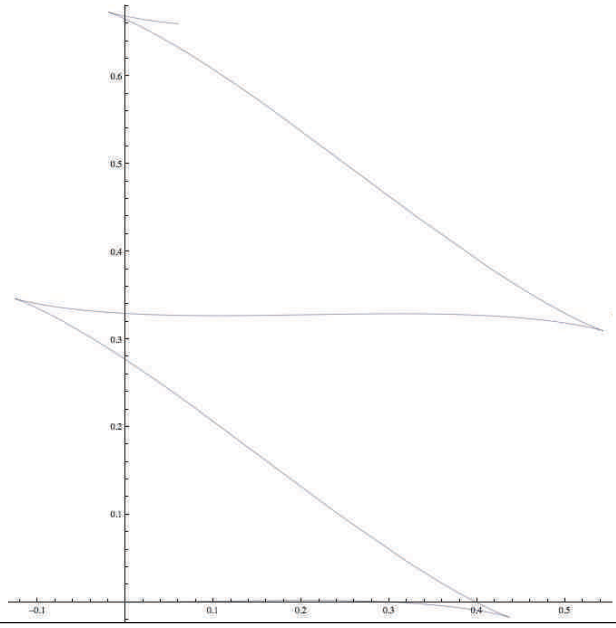
$P_a = \frac{\pi}{2}, \quad P_r = \frac{1}{2}, \quad P_3(0) = \dot{\alpha}(0)$



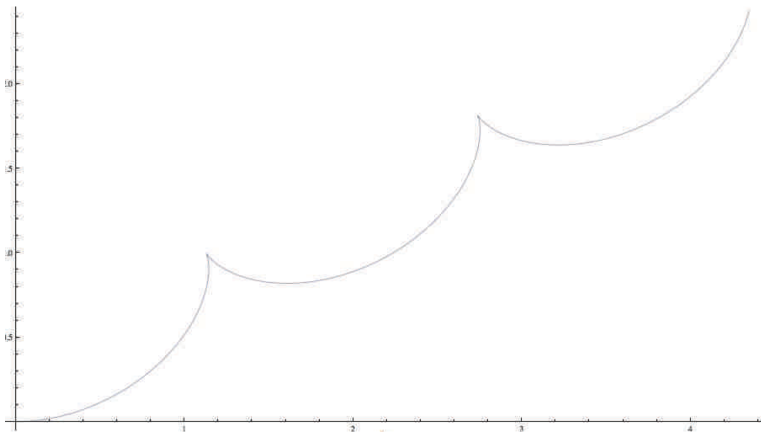
$\rightarrow$changing P_r one changes the type of trajectory (distance of cusps)

$\rightarrow$changing P_a one changes "where to starts on the trajectory"

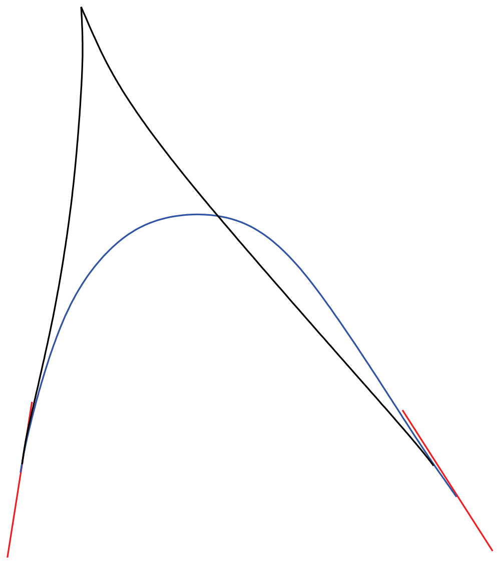
Generic example 1



Generic example 2

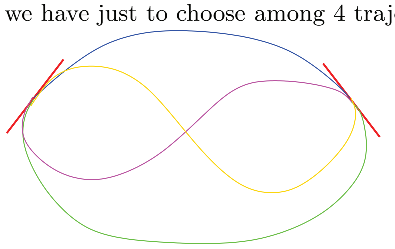


Geodesics are not optimal forever

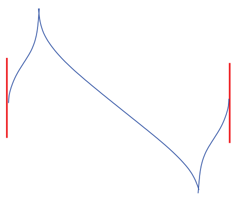


This program has been essentially concluded by Y. Sachkov and I. Moiseev.

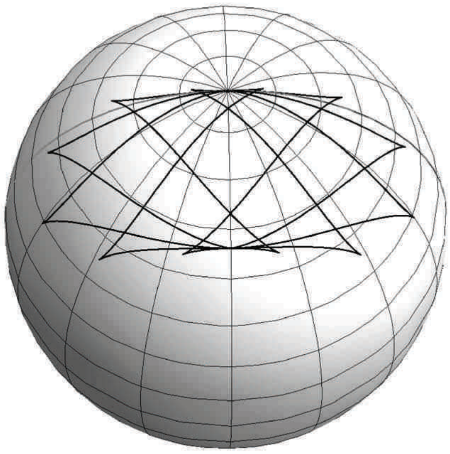
- they computed all optimal trajectories on $\mathbf{R}^2 \times S^1$ (it is not a problem for planar curves)
- we have just to choose among 4 trajectories



Conjecture: there exist optimal trajectories with TWO cusps (and not with three)

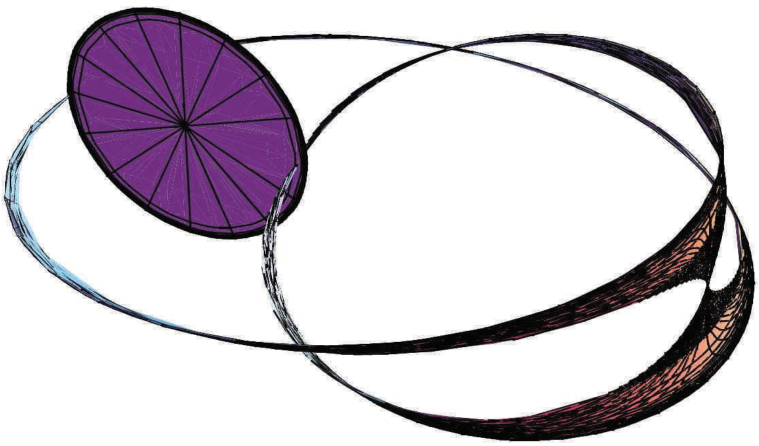


This is true for the problem formulated on PTS^2:



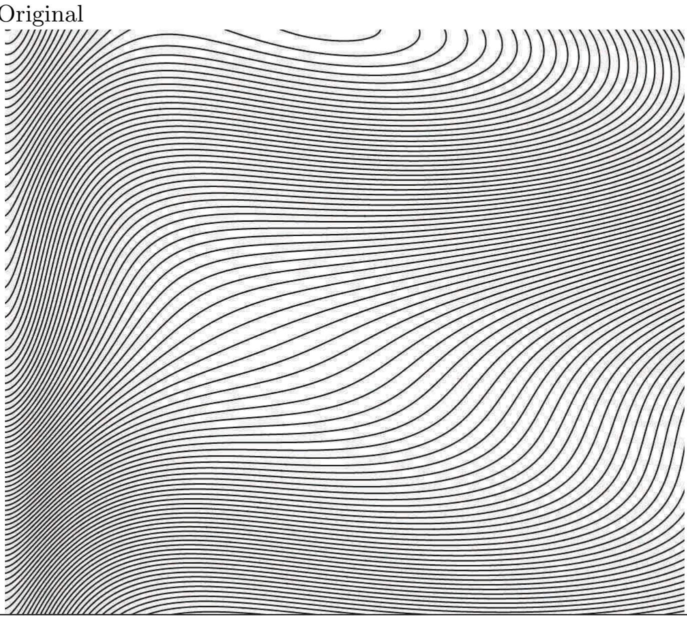
F. Rossi, U. B., COCV 2009

The cut locus for the Sachkov's synthesis



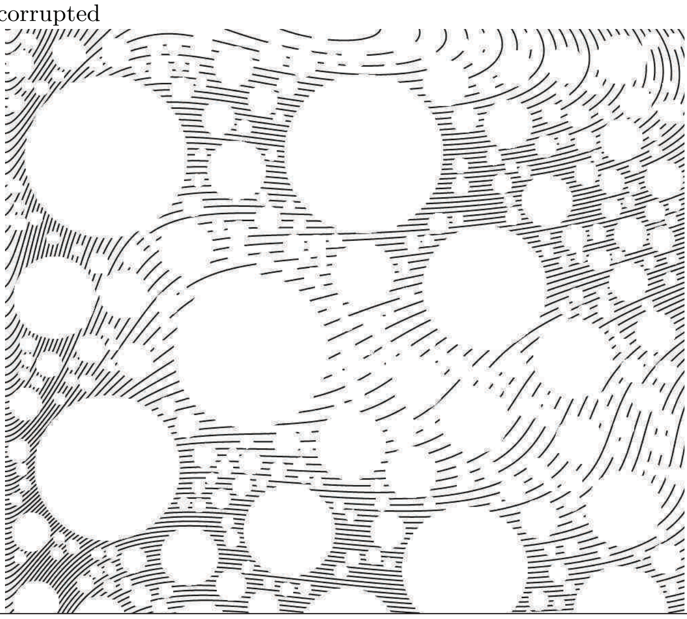
Preliminary results of reconstruction of level sets by Yuri Sachkov

Original



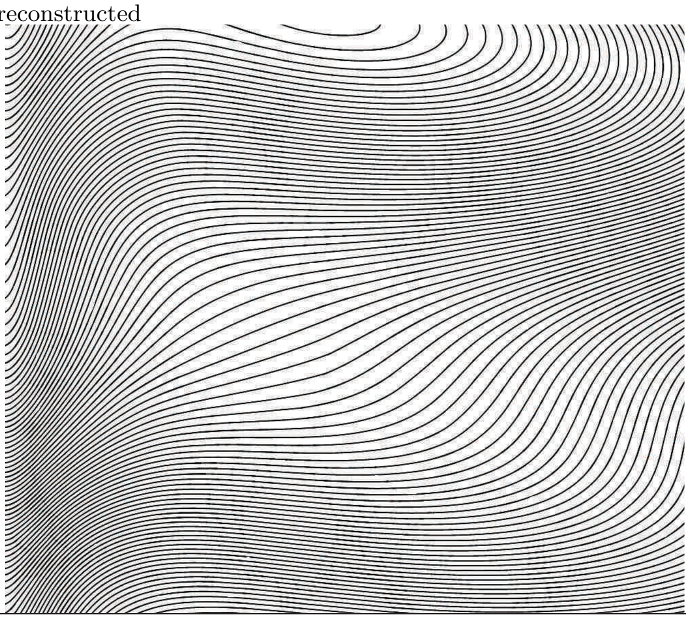
Preliminary results of reconstruction of level sets by Yuri Sachkov

corrupted

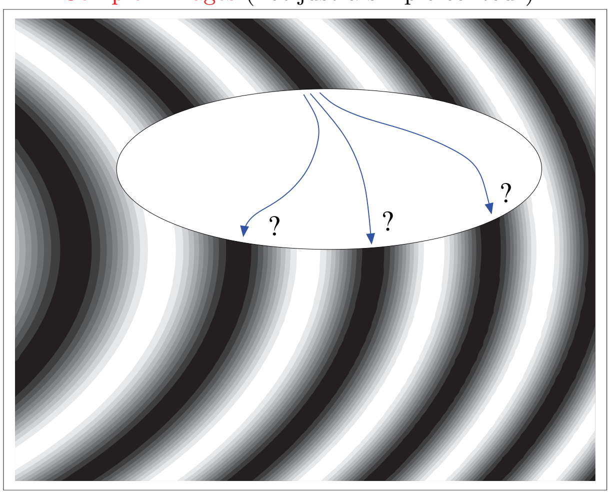


Preliminary results of reconstruction of level sets by Yuri Sachkov

reconstructed



Complex images (not just a simple contour):



all possible paths are activated as a Brownian motion

$$dx = X_1 dW_1 + X_2 dW_2, \to \partial_t \psi(t,x) = (L_{X_1}^2 + L_{X_2}^2)\psi(t,x)$$

$$L_{X_1}^2 + L_{X_2}^2 = (\cos(\alpha)\partial_{x_1} + \sin(\alpha)\partial_{x_2})^2 + \beta^2 \partial_\alpha^2$$

(sub-elliptic Heat equation, under Hörmander condition $\Rightarrow$, solutions are smooth)

$\to$**highly non-isotropic diffusion**

PLAN:

1) lifting the image to $PT\mathbf{R}^2$ and using it as an initial condition for the hypoelliptic heat eq.
2) making convolution with the hypoelliptic heat kernel (Agrachev, Gauthier, Rossi, U.B.)
3) projecting down the image

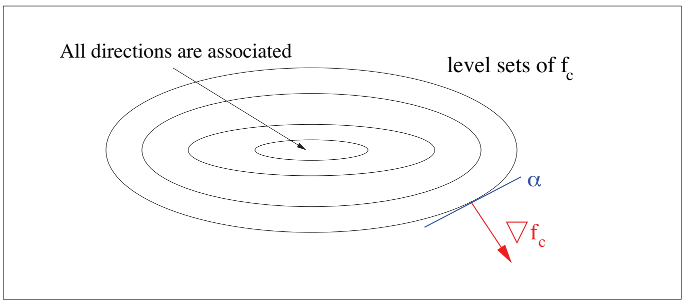
1) lifting an image to a distribution in $PT\mathbf{R}^2$

Assume that the grey level of an corrupted image is described by a function $f_c : \mathbf{R}_c := \mathbf{R}^2 \setminus \Omega \to [0, \infty[$. The set Ω represents the region where the image is corrupted.

(Hyp) the set $\mathbf{R}_c$ is an open subset of $\mathbf{R}^2$ and f_c is $\mathcal{C}^1$.

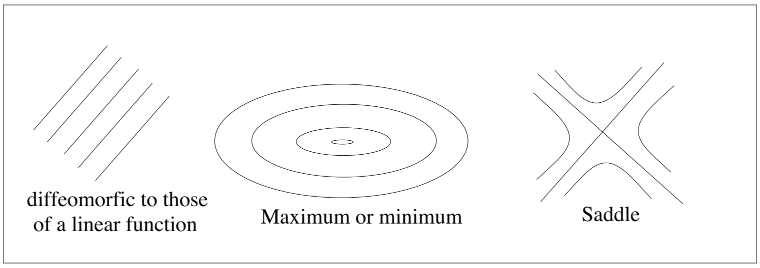
Let us lift the domain $\mathbf{R}_c$ of f_c in $PT\mathbf{R}^2$. This is made by associating to every point (x_1, x_2) of $\mathbf{R}_c$ the direction $\alpha \in \mathbf{R}/\sim$ of the level set of f_c at the point (x_1, x_2).

$$L(f_c) = \{(x_1, x_2, \alpha) \in \mathbf{R}_c^2 \times P^1 \text{ s.t. } \nabla f_c(x_1, x_2) \cdot (\cos(\alpha), \sin(\alpha)) = 0\},$$

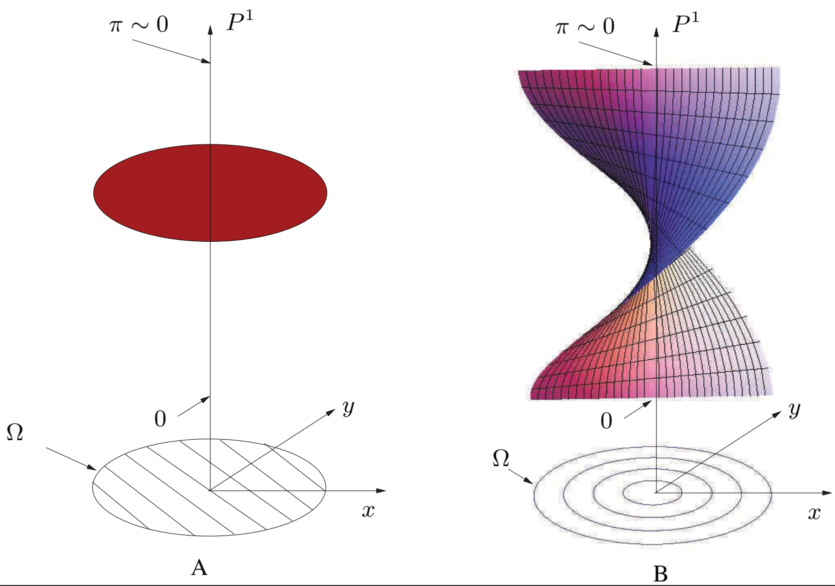


An interesting feature 1

under generic conditions, f_c is a Morse function and $L(f_c)$ is a 2D manifold.
(This is false if $\alpha \in [0, 2\pi] \setminus \sim$)



An interesting feature 2



Why a Morse function?

Describe an Image by $\mathcal{I} \in L^2(D, R)$ (where D is an open bounded domain of R^2)

→even if images are not described by Morse functions, it is widely accepted that the retina approximately smoothes the images by making the convolution with a Gaussian function

[1] L. Peichl, H. Wässle, J Physiol, Vol. 291, 1979, pp. 117-41.

[2] D. Marr; E. Hildreth, Proceedings of the Royal Society of London, Vol. 207, No. 1167. (Feb. 29, 1980), pp. 187-217.

→we have proved that we show that, given $G(\sigma_x, \sigma_y)$, the two dimensional Gaussian centered in $(0, 0)$ with standard deviations $\sigma_x, \sigma_y > 0$, then the smoothed image

$$f = \mathcal{I} * G(\sigma_x, \sigma_y) \in L^2(\mathbf{R}^2, \mathbf{R}) \cap \mathcal{C}^\infty(\mathbf{R}^2, \mathbf{R}),$$

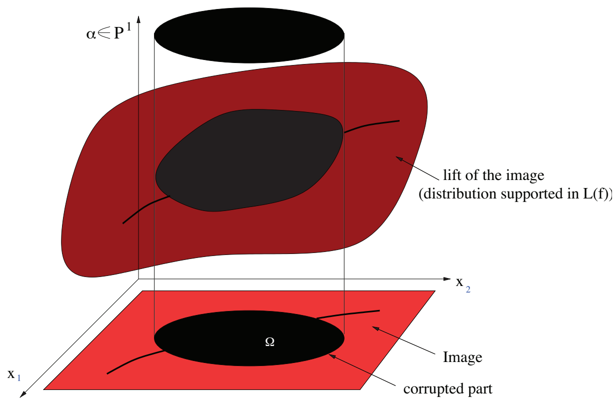
is generically a Morse function.

→i.e. the set $\{\mathcal{I} \in L^2$ such that f is a Morse function$\}$ is a countable intersection of open and dense subsets.

→the set $\{\mathcal{I} \in L^2$ such that $f_{\mathbf{K}}$ is a Morse function$\}$ is open and dense.

Define the distribution on $\mathbf{R}_c \times P^1$:

$$\hat{f}_c(x_1, x_2, \alpha) = f_c(x_1, x_2)\delta(d((x_1, x_2, \alpha), L(f))$$



2) making the convolution with the hypoelliptic heat kernel

- Explicit expressions of Hypoelliptic Heat Kernels are difficult to obtain
- From 1979 up to 2008 the only known expressions were for STEP 2 Nilpotent Lie groups

By using the noncommutative Fourier Transform (based on representation theory) we got the explicit expression for $SU(2), SL(2), SO(3), SE(2)$ and for our problem on $\mathbf{R}^2 \times P^1 \sim SE(2)/Z_2$ (A. Agrachev, F. Rossi, J-P. Gauthier, U. Boscain, JFA, 2009):

The kernel

The kernel on $SE(2)$
$\psi(t,g) = \psi_0 * p_t(g) = \int_G \psi_0(\bar{g}) p_t(\bar{g}^{-1} g) \mu(\bar{g})$ where $g = (x_1, x_2, \alpha)$.

$$\mathbf{p}_t(x_1,x_2,\alpha) = \int_0^\infty \lambda d\lambda \sum_{k=0}^\infty \int_0^{2\pi} du$$

$$\Big(\exp[-a_c(k,\lambda)t]\ C(k,\frac{\lambda^2}{4},u)\ C(k,\frac{\lambda^2}{4},u+\alpha)\ \cos[\lambda[x_1\cos(u) - x_2\sin(u)] +$$

$$\exp[-a_s(k,\lambda)t]\ S(k,\frac{\lambda^2}{4},u)\ S(k,\frac{\lambda^2}{4},u+\alpha) \cos[\lambda[x_1\cos(u) - x_2\sin(u)] \Big)$$

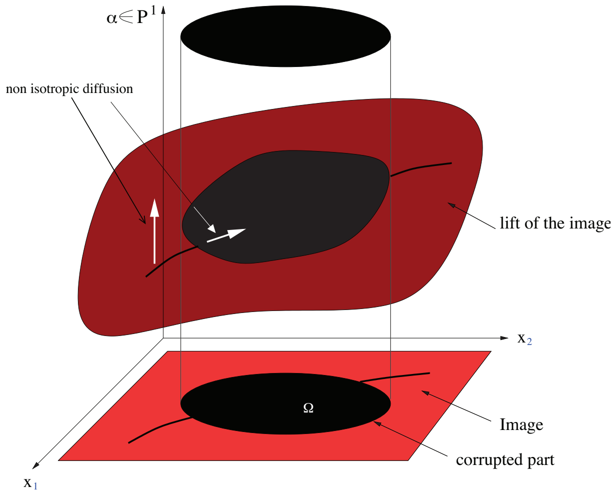
(Periodic Mathieu functions. Here $\beta = 1$.)

The kernel on $PT\mathbf{R}^2$
$\hat{f}_r(t,g) = \hat{f}_c * K_t(g) = \int_G \hat{f}_c(\bar{g}) K_t(g,\bar{g}) \mu(\bar{g})$.

$K_t(x_1,x_2,\alpha,\bar{x}_1,\bar{x}_2,\bar{\alpha}) := p_t((\bar{x},\bar{y},\bar{\alpha})^{-1} \circ (x,y,\alpha)) + p_t((\bar{x},\bar{y},\bar{\alpha})^{-1} \circ (x,y,\alpha+\pi))$.

chirality of pinwheels?

3) Projecting down



$$f_r(t,x_1,x_2) = \int_{P^1} (\mathbf{K}_t * \hat{f}_c)_{(x_1,x_2,\alpha)}\, d\alpha$$

NUMERICAL IMPLEMENTATION

Problems:

- implementing numerically Mathieu functions is difficult
- solving numerically the non-isotropic diffusion equation

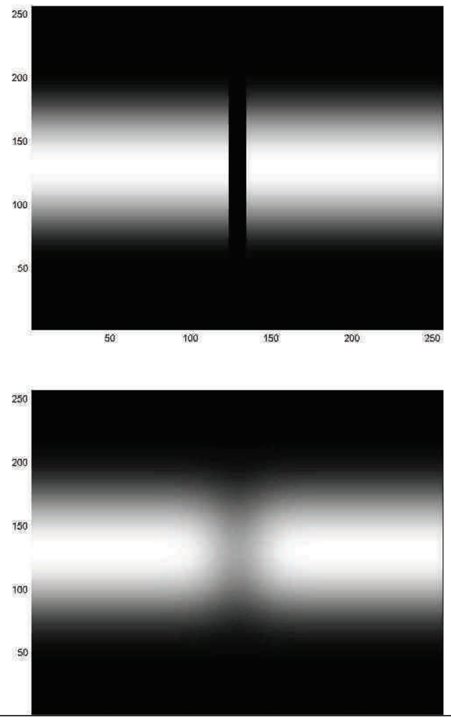
  $\rightarrow$not good because there are two different scales (there are no algorithms to compute numerically non-isotropic diffusion)
- we are solving numerically the (noncommutative) Fourier transform of:

  $$\partial_t \psi(t,x_1,x_2,\alpha) = \left((\cos(\alpha)\partial_{x_1} + \sin(\alpha)\partial_{x_2})^2 + \beta^2 \partial_\alpha^2\right)\psi(t,x_1,x_2,\alpha)$$

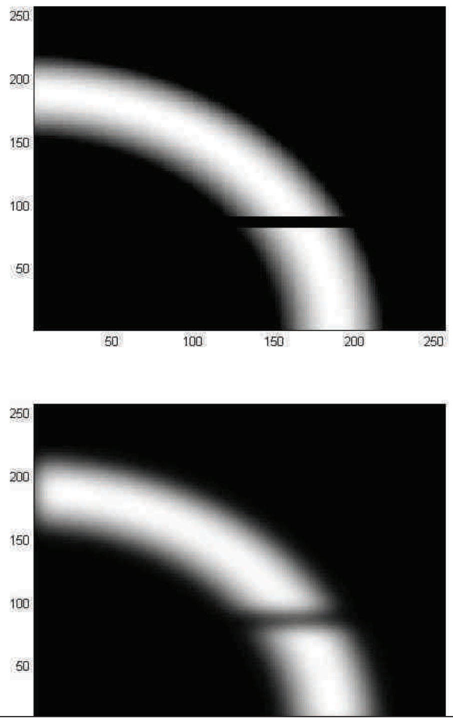
  That is the Mathieu equation

  $$\partial_t \Phi(t,\theta) = (\beta^2 \frac{d^2}{d\theta^2} + \lambda^2 \cos^2(\theta))\Phi(t,\theta)$$

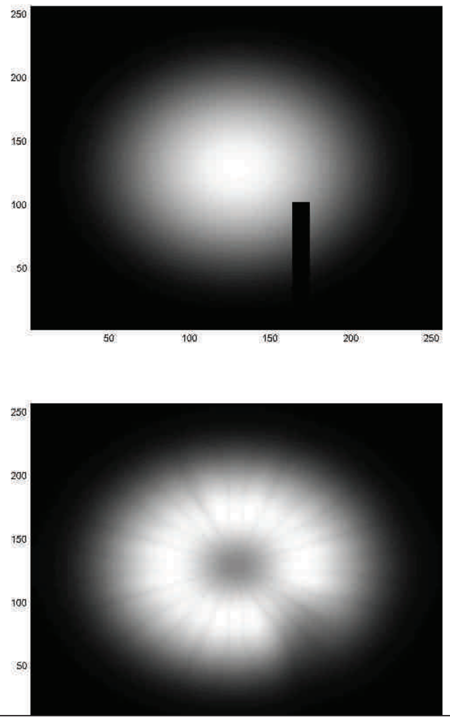
Example 1



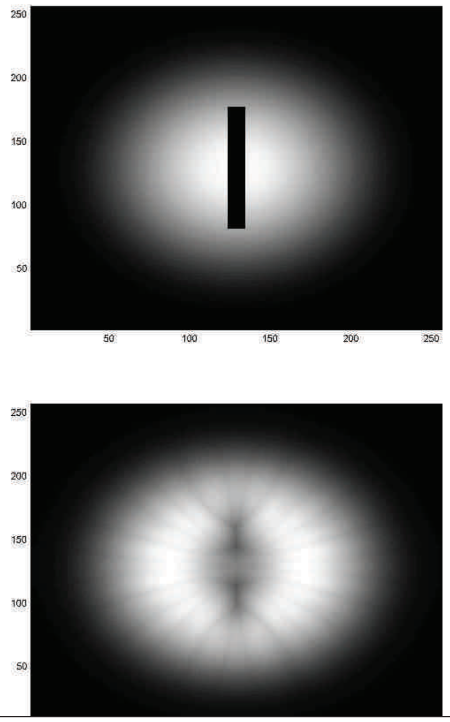
Example 2



Example 3



Example 4



Claude TOUZET

LNIA / Univ. de Provence

http://www.univ-provence.fr/gsite/document.php?project=umr_6149&doc=perso_touzet

"Modèles Cognitifs de l'Attention Visuelle"

Le Monde réel est cohérent et continu. Il présente de fait des régularités - une structuration - que chacun de nous perçoit. Lorsque la perception est visuelle alors le traitement de l'information est appelé la vision.

C'est un processif cognitif au sens où le cortex est impliqué et qu'il permet des traitements hors de portée des autres animaux, tels que la lecture par exemple. La structure neuronale du cortex est connue, mais les liens qui unissent le cortex au Monde réel font débat.

La Théorie neuronale de la Cognition explique comment les phénomènes observés d'attentions endogène et exogène sont produits.

En résumé, le cortex est constitué de multiples cartes corticales organisées au sein de hiérarchies, chaque carte corticale jouant le rôle d'un filtre de nouveauté, passant au niveau hiérarchique suivant (« bottom-up ») les événements imprévus : c'est l'attention exogène.

Des connexions neuronales « top-down » suffisamment nombreuses permettent la mise en place d'une véritable pré-activation de l'ensemble de la hiérarchie en fonction du but identifié à un niveau quelconque de la hiérarchie : c'est l'attention endogène.

Notons en guise de conclusion que ce modèle cognitif de l'attention visuel est généralisable à l'ensemble des processus cognitifs, et donc propose une explication de ce que nous sommes.

Référence :

C. Touzet, Conscience, Intelligence, Libre-Arbitre ? Les réponses de la Théorie neuronale de la Cognition, Editions la Machotte, 2010, 156 pages, ISBN : 978-2-919411-00-9. (www.machotte.fr)

Modèles cognitifs de l'attention visuelle

C. Touzet

claude.touzet@univ-provence.fr

Laboratoire de Neurosciences Intégratives et Adaptatives (LNIA) UMR6149 CNRS-Université de Provence, Marseille

5ème Ecole « Recherches Multimodales d'Informations » (ERMITES'2010) – Vision & Cognition

Résumé

Le Monde réel est cohérent et continu. Il présente de fait des régularités - une structuration - que chacun de nous perçoit. Lorsque la perception est visuelle alors le traitement de l'information est appelé la vision. C'est un processif cognitif au sens où le cortex est impliqué et qu'il permet des traitements hors de portée des autres animaux, tels que la lecture par exemple. La structure neuronale du cortex est connue, mais les liens qui unissent le cortex au Monde réel font débat. La Théorie neuronale de la Cognition explique comment les phénomènes observés d'attentions endogène et exogène sont produits. En résumé, le cortex est constitué de multiples cartes corticales organisées au sein de hiérarchies, chaque carte corticale jouant le rôle d'un filtre de nouveauté, passant au niveau hiérarchique suivant (« bottom-up ») les événements imprévus : c'est l'attention exogène. Des connexions neuronales « top-down » suffisamment nombreuses permettent la mise en place d'une véritable pré activation de l'ensemble de la hiérarchie en fonction du but identifié à un niveau quelconque de la hiérarchie : c'est l'attention endogène. Notons en guise de conclusion que ce modèle cognitif de l'attention visuel est généralisable à l'ensemble des processus cognitifs, et donc propose une explication de ce que nous sommes.

Sommaire

Le Monde réel

Le cortex

Théorie neuronale de la Cognition

Apprentissage

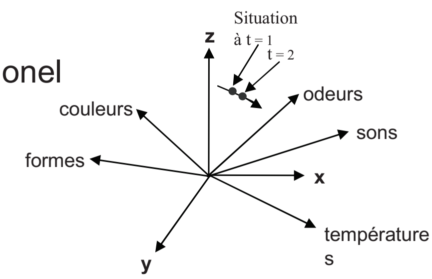
Comportement motivé

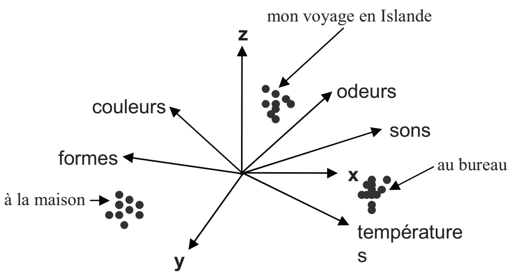
Attention endogène

Attention exogène

Le Monde réel

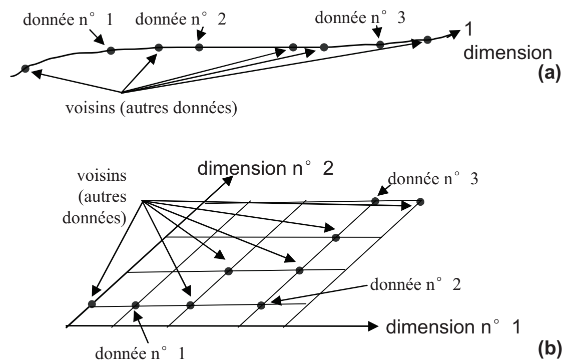
Continu, Cohérent, Multi-dimensionel





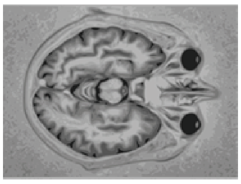
Chaque situation que nous vivons est un point dans un espace multidimensionnel. Le Monde est continu : entre deux instants successifs (t=1 et t=2), la situation change peu. Il est donc possible de discerner des trajectoires qui sont des successions de situations.

Chacun de nous vit certaines situations et pas d'autres. Ces situations ne sont pas réparties de manière homogène, mais regroupées dans certaines régions de l'Espace multidimensionnel. Ces régions définissent notre Environnement personnel et rien d'autre n'existe.

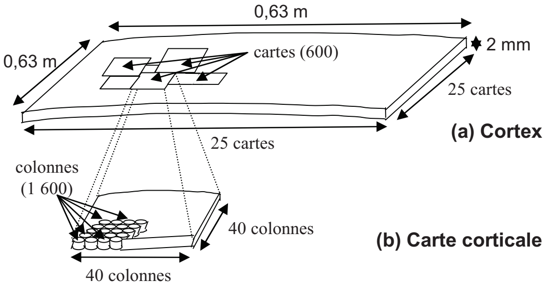


Pour estimer les distances dans un Espace multi-dimensionnel, nous pouvons utiliser le nombre de données (situations) qui séparent deux données particulières. Dans le cas d'un espace de dimension 1 (1-D), la donnée n° 2 est plus proche de la n° 1 que la n° 3. Dans le cas d'un espace de dimension 2 (2-D), la distance qui sépare le n° 1 et le n° 2 est de 2, celle qui sépare le n° 1 et le n° 3 est de 5, celle qui sépare le n° 2 et le n° 3 de 3.

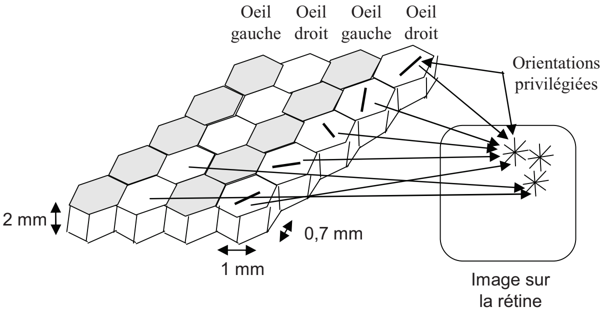
Le cortex



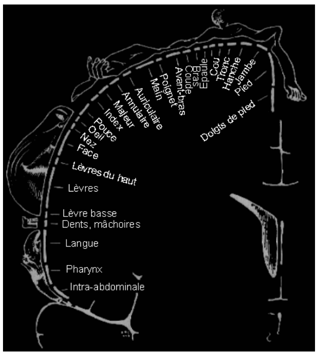
Le cortex (4 000 cm²) est replié sur lui-même pour tenir dans la boite crânienne (IRM). Brain MRI Vector representation, Nevit Dilmen, 2006, licence Creative Commons Attribution ShareAlike 3.0. (Wikipedia).



(a) Le cortex est organisé en centaines de cartes corticales.
(b) Chaque carte corticale contient plus d'un millier de colonnes corticales, chacune comprenant un millier de micro-colonnes.

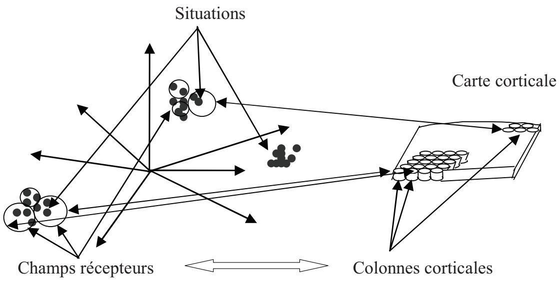


Carte corticale codant les orientations spatiales des stimuli (pluriel de stimulus) présentés sur la rétine. Chaque colonne corticale s'excite principalement pour un contraste dans une orientation précise, en un point précis de la rétine. Sur l'image rétinienne, nous avons dessiné les stimuli pour 3 localisations. Deux de ces localisations génèrent des activations sur des colonnes du morceau de carte dessiné ici. Notez l'alternance œil droit-œil gauche sur la carte. La localisation sur la rétine de la zone d'intérêt est identique entre deux colonnes voisines, même si elles appartiennent à des yeux différents.

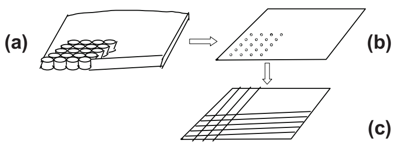


Le cortex

L'Homoncule (de Penfield) est la carte corticale représentant notre corps. Les informations tactiles construisent l'Homoncule sensoriel (hémisphère gauche), tandis que les informations envoyées aux muscles construisent l'Homoncule moteur (hémisphère droit).
D'après W. Penfield, T. Rasmussen, *The cerebral cortex of man*, Macmillan, 1950, pp. 214-215.

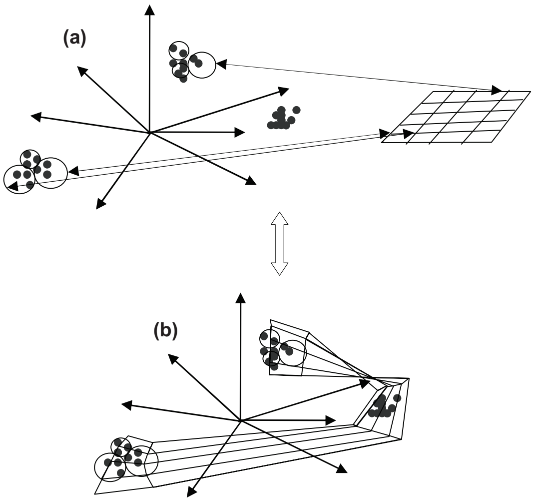


Une carte auto-organisatrice est une projection d'un Espace multi-dimensionnel qui respecte fréquence et voisinage. Ce qui est fréquent dans l'Espace est mieux représenté sur la carte que ce qui ne l'est pas. Ce qui est voisin dans l'Espace est voisin sur la carte. A chaque colonne de la carte est associée une région de l'Espace. Cette région est le « champ récepteur » de la colonne.



Une carte corticale avec ses colonnes (a) peut être représentée par un point au centre de chaque colonne (b). Cependant, il est plus utile de « visualiser » le voisinage de chaque colonne. Les lignes verticales et horizontales en (c) donnent le voisinage. A chaque intersection, il y a une colonne. Chaque colonne a 4 colonnes voisines (sauf sur les bords : 3 voisines, et dans les coins : 2 voisines). De plus, comme les comportements des micro-colonnes d'une colonne sont similaires, nous ne dirons plus « colonne », mais « neurone » (ce neurone en représente environ 110 000).

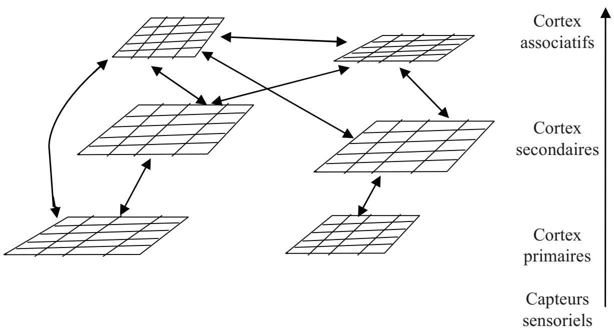
Le cortex



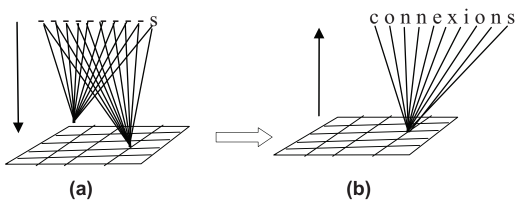
(a) Chaque neurone est associé à un champ récepteur (représenté par un cercle) dans l'Espace multi-dimensionnel. Les champs récepteurs voisins sont associés à des neurones voisins.
(b) Lorsque l'on place les neurones directement au centre de leur champ récepteur et que l'on trace les voisinages, alors l'observateur a l'impression de voir la carte se tordre dans l'Espace multi-dimensionnel. Il s'agit d'une vue de l'esprit. Le cortex ne change pas de forme – mais il est utile de se rendre compte de comment la carte s'organise pour représenter les données avec le minimum d'erreur.

« Si deux neurones A et B interconnectés sont actifs dans une même fenêtre temporelle, alors la force des connexions entre A et B, et aussi entre B et A, est renforcée ». (Loi de Hebb)

Le cortex

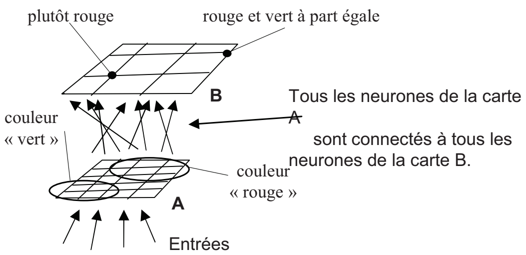


Le cortex est constitué de multiples cartes auto-organisatrices interconnectées. Les cartes les plus proches des capteurs sensoriels sont organisées avant les autres. Elles constituent les cortex primaires (vision, audition, tact, etc.). Elles alimentent des cartes appartenant aux cortex secondaires, puis les cortex associatifs. Il y a de très nombreuses connexions directes et réciproques.

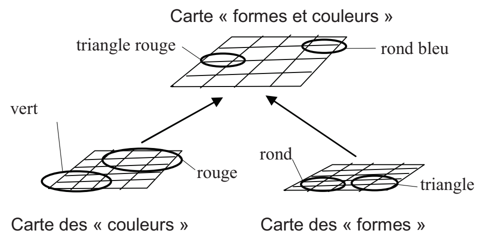


La carte auto-organisatrice est une mémoire associative. Pour obtenir une réponse, il faut fournir une partie de cette réponse.
(a) Si la réponse cherchée est le mot de 10 lettres le plus fréquent de ce livre se terminant par la lettre « s », la question est une activation de la 10ème connexion avec la lettre « s ». Cette activation est envoyée à tous les neurones de la carte. (b) Il suffit alors de décoder les connexions du neurone gagnant vers les entrées pour connaître le mot recherché.

Théorie neuronale de la Cognition

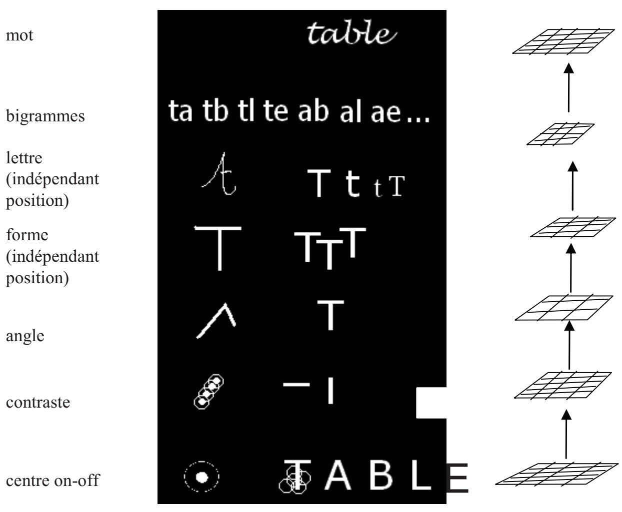


Deux cartes auto-organisatrices (l'une après l'autre) permettent d'extraire des relations complexes entre les données. Ici, B extrait des proportions d'activation entre diverses zones de la carte A.



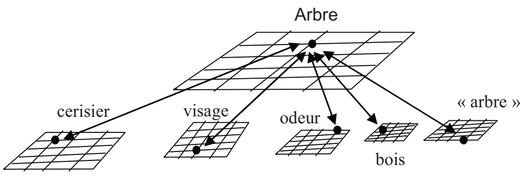
Trois cartes auto-organisatrices permettent la fusion de données. Il suffit que les données traitées par chacune des deux premières cartes appartiennent à des modalités différentes (par exemple, la forme et la couleur), et qu'elles alimentent la troisième carte. C'est cette dernière carte qui réalise la fusion de données.

Théorie neuronale de la Cognition



Six niveaux de cartes auto-organisatrices sont nécessaires pour passer de l'image sur la rétine à la reconnaissance orthographique d'un mot (le 1er niveau est réalisé par l'œil), d'après Dehaene

Théorie neuronale de la Cognition

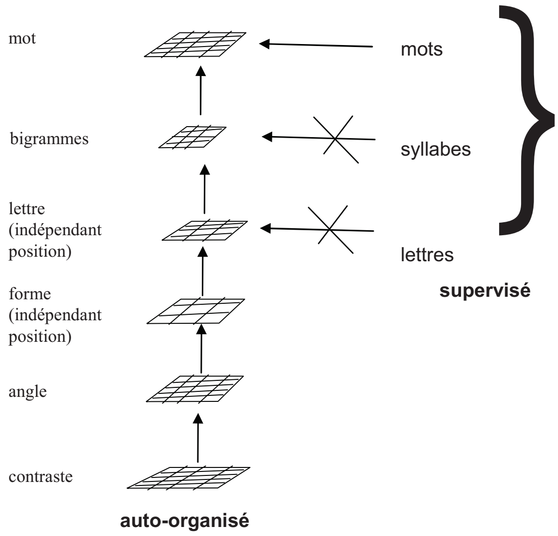


Un « concept » implique l'activation de multiples neurones au sein de plusieurs cartes.

Ordre dans l'organisation

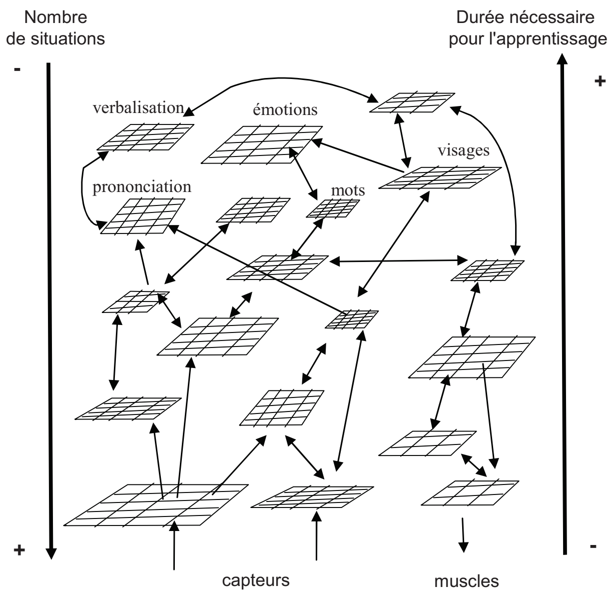
Certitude

Méditation



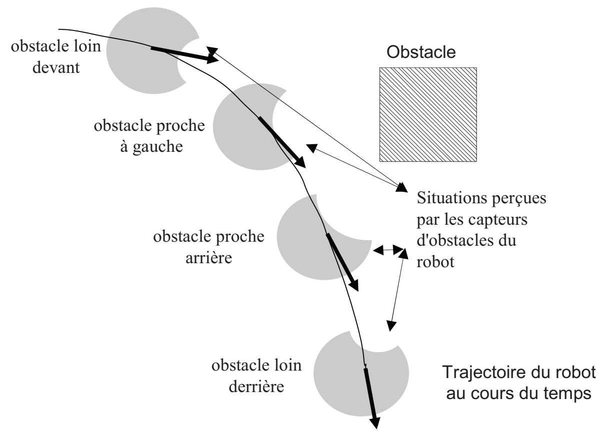
La méthode d'apprentissage de la lecture dite « globale » fait disparaître deux étapes supervisées. L'acquisition de la lecture devient alors pratiquement impossible à l'apprenti lecteur.

Apprentissage



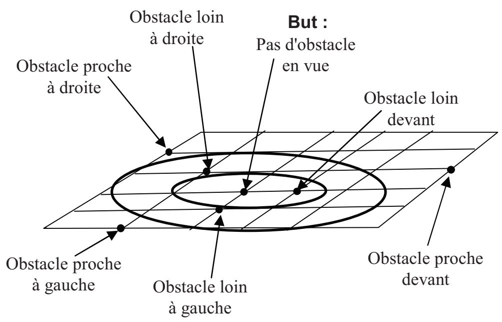
Il y a un ordre et un timing précis dans l'organisation des cartes auto-organisatrices. Les cartes s'organisent en fonction des régularités qu'elles perçoivent dans les données reçues. Certaines cartes doivent s'organiser avant un âge limite (7 ans pour les cartes liées au langage et à la socialisation).

Comportement motivé

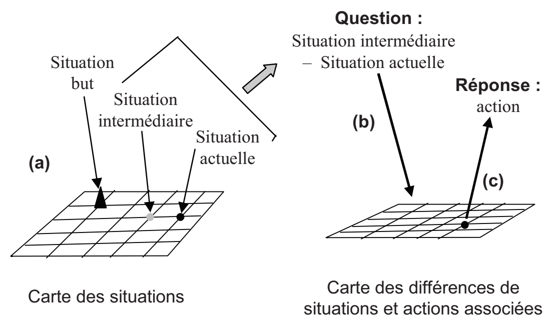


Comportement d'évitement d'obstacles construit à partir d'un but à atteindre « aucun d'obstacle en vue ». Le robot arrive face à l'obstacle, ses capteurs le perçoivent. Il tourne et évite l'obstacle. En grisé, les distances mesurées par les capteurs du robot vis à vis des obstacles éventuels.

Comportement motivé

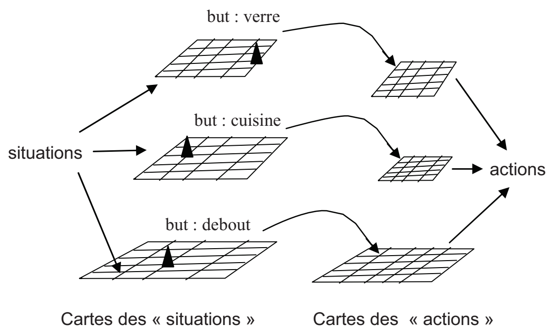


Le voisinage conservé par la carte garantit que les situations voisines sont voisines sur la carte. Il est donc toujours possible de trouver une situation voisine de la situation actuelle qui soit plus proche de la situation « but ».



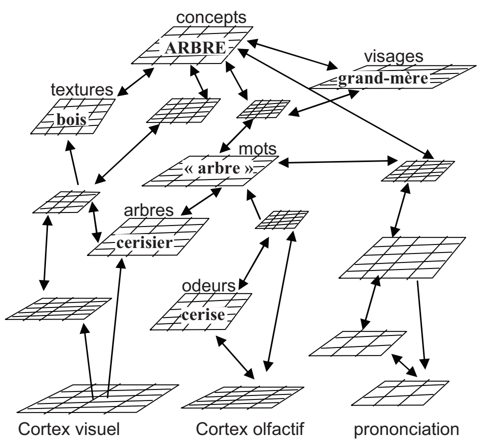
(a) Le but et la situation actuelle définissent la situation intermédiaire désirée.
(b) La variation entre la situation désirée et la situation courante permet de sélectionner un neurone qui code pour l'action correspondante (c).

Comportement motivé



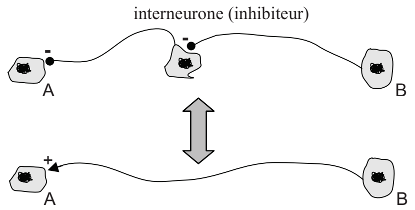
Un comportement complexe : une somme de comportements élémentaires.

Attention endogène

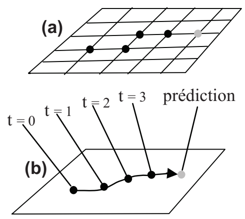


La vue d'un arbre active le concept « Arbre », qui active alors l'ensemble des éléments liés à ce concept.

Attention endogène

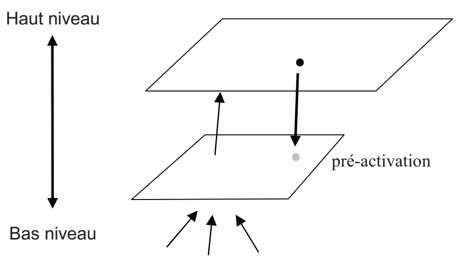


Activation réciproque du neurone B sur A. Ces deux schémas sont équivalents (une flèche représente une synapse excitatrice, un rond une synapse inhibitrice).

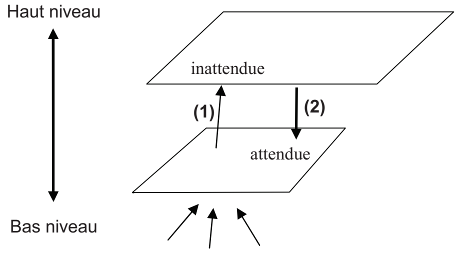


Prédiction de l'évolution de la situation courante.
(a) et (b) représentent la même carte, mais dorénavant nous ne dessinerons plus les voisinages.

Attention exogène

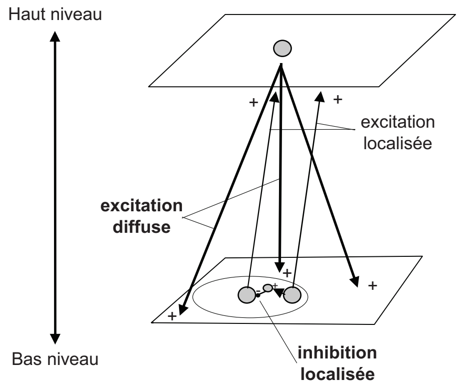


Attention générée par les connexions réciproques : il y a pré-activation des neurones associés à l'activation de plus haut niveau.



Attention exogène (1) et attention endogène (2). En (1) il y a transfert vers les niveaux supérieurs des événements inattendus et arrêt de la transmission pour les événements prévisibles. En (2) il y a facilitation de l'activation de neurones de niveau inférieur appartenant à des événements prévus (attendus) à des niveaux supérieurs.

Synergie des attentions endogène et exogène



Modélisation neuronale des processus attentionnels. Les excitations montantes sont plus localisées et plus fortes que celles en provenance des connexions réciproques (plus diffuses). Les connexions sur une même carte sont inhibitrices et localisées.

Références

C. Touzet, "Modeling and Simulation of Elementary Robot Behaviors using Associative Memories", *International Journal of Advanced Robotic Systems*, Vol 3 n° 2, June 2006.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends Cogn Sci*, 9(7), 335-341.

M. Silver and S. Kastner (2009) – Topographic maps in human frontal and parietal cortex, *Trends in Cognitive Sciences*, Vol. 13, No. 11, 488-495.

Donald Hebb, *The Organization of Behavior : A Neuropsychological Theory,* Wiley, 1949.

Teuvo Kohonen, *Self-Organizing Maps, Springer Series in Information Sciences, Vol. 30*, 2001. Third Ed., 501 pages. ISBN 3-540-67921-9, ISSN 0720-678X

D. Hubel, *Eye, Brain, and Vision*, Freeman, May 1995. ISBN 978-0-7167-6009-2. Disponible sur le Web à l'adresse : http://hubel.med.harvard.edu/book/ch5.pdf

W. Penfield, T. Rasmussen, *The cerebral cortex of man*, Macmillan, 1950.

J. Paillard - (1987) L'ordinateur et le cerveau : un contraste saisissant, *Bulletin de l'AFCET, INTERFACES*, 57, pp 4-9.

C. Touzet, *Conscience, Intelligence, Libre-Arbitre ? Les réponses de la Théorie neuronale de la Cognition*, Editions la Machotte, 2010, 156 pages, ISBN : 978-2-919411-00-9. (www.machotte.fr)

Hervé LE BORGNE

CEA-LIST

http://elm.eeng.dcu.ie/~hlborgne/

"Description de Scènes Naturelles par Composantes Indépendantes"

De nombreuses études montrent que les détecteurs corticaux pourraient résulter de l'application d'un principe de réduction de redondance par indépendance statistique de leurs activités. Nous utilisons l'Analyse en Composantes Indépendantes (ACI) pour générer de tels détecteurs, aboutissant à un codage parcimonieux de l'information.

1) Nous en effectuons une analyse quantitative mettant en valeur l'adaptation des détecteurs aux catégories d'images considérées. Le cadre applicatif concerne la classification de scènes naturelles, pour lequel l'adaptation des descripteurs aux statistiques des catégories à discriminer est attrayante.

2) Nous proposons deux schémas de codage de l'information. Le premier correspond à des modélisations de complexité croissante de la densité résultante du filtrage d'une image par les descripteurs considérés. Le second schéma est une exploitation directe de l'adaptation des filtres aux catégories, qui est semblable aux descripteurs de type "sacs de mots visuels".

Ces travaux [1,2] entrent dans le cadre plus large des rapports entre perception visuelle et sciences de l'ingénieur, dont les apports réciproques permettent une meilleure compréhension de chacun des domaines [3].

Références :

[1] Le Borgne H., Guérin-Dugué A., Antoniadis A. Representation of images for classification with independent features. Pattern Recognition Letters (PRL), vol 25, N° 2, pp 141-154, january 2004.

[2] H. Le Borgne, A. Guérin-Dugué, N.E. O'Connor Learning Mid-level Image Features for Natural Scene and Texture Classification. IEEE transaction on Circuits and Systems for Video Technology, 17(3):286-297, march 2007.

[3] Guyader N., Le Borgne H., Hérault J., Guérin-Dugué A. Towards the introduction of human perception in a natural scene classification system. Proc. of the IEEE International workshop on Neural Networks for Signal Processing (NNSP'2002), pp 385-394, Martigny Valais, Switzerland, September 4-6, 2002

ERMITES 2010

Description de scènes naturelles par composantes indépendantes

Hervé Le Borgne
http://elm.eeng.dcu.ie/~hlborgne/
Septembre 2010

Dans les 120 minutes à venir…

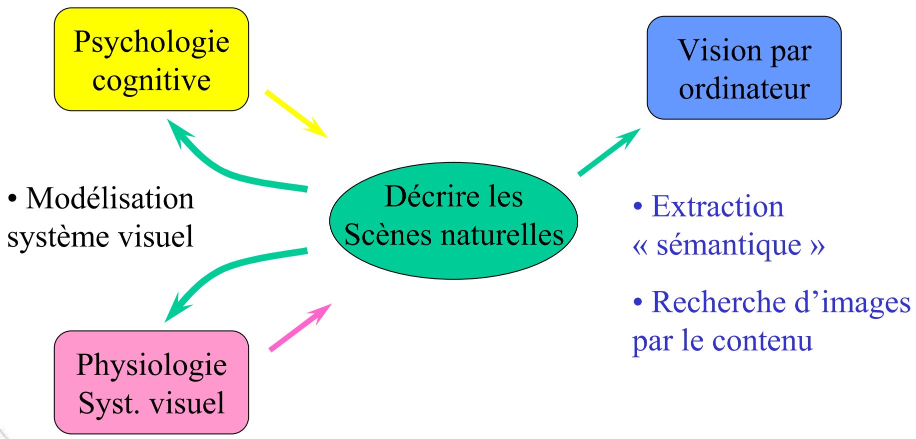
- **Scènes naturelles**
  - Caractéristiques et propriétés
  - Rapport à la vision
  - De la perception au modèle de discrimination
- **Descripteurs *propres* aux scènes naturelles**
  - Rappel sur les « visages propres »
  - Extraction par ACI
  - Propriétés (pour la discrimination)
- **Classification de scènes naturelles**
  - Calcul de signatures
  - Validation expérimentale
- **Conclusion**

2

Rendre à César…

- Travaux réalisés au LIS de Grenoble (→ GIPSA) en 1999-2004
- Sous la direction de **Anne Guérin Dugué**
- En collaboration avec **Anestis Antoniadis**
  - ADCI: analyse discriminante en composantes indépendante (fin exposé)
- Dans l'équipe de **Jeanny Hérault**
  - Approche: vision par ordinateur ↔ perception ↔ physiologie syst. visuel
- En collaboration fréquente avec **Nathalie Guyader**
  - Aspects perception (Alan Chauvin, Christian Marendaz…)

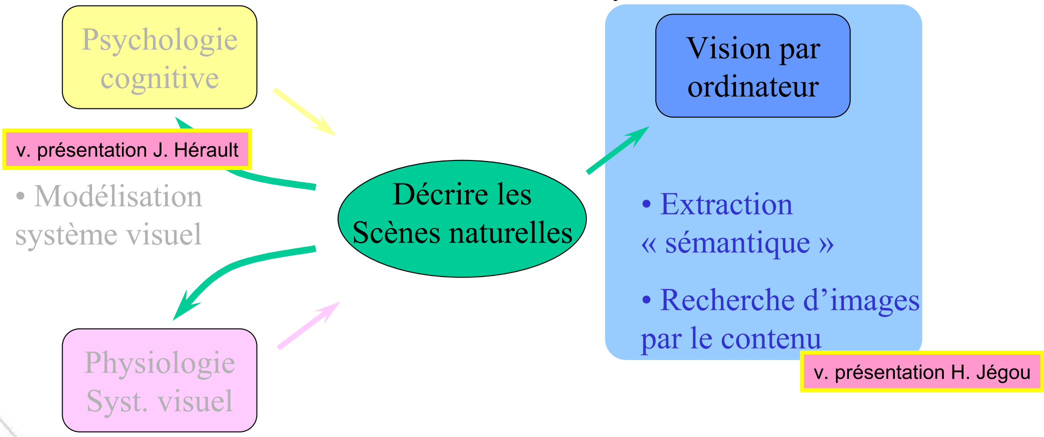
Enjeux scientifiques



Enjeux scientifiques

1 – Catégorisation perceptive

2 – Extraction de caractéristiques
3 – Analyse de scènes naturelles



Scènes naturelles

- **Caractéristiques**
- **Propriétés**

Scènes naturelles

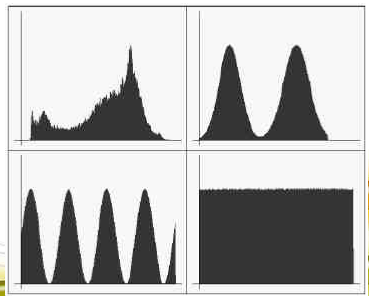
NON

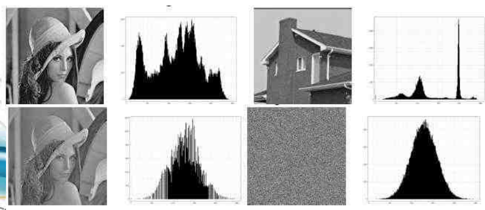
OUI

→ Images du monde réel

Statistiques des scènes naturelles

- **Monde réel → évolution → syst. visuels des êtres vivants**
  - Importance des premiers stimuli [Blackmore & Cooper, 71] [Movshon & van Sluyters 81]
  - Comment?
  - Propriétés du monde réel?
- **Intérêt pratique**
  - Compression [Kretzmer, 1952]
  - Restauration d'image
  - Indexation, classification
- **Statistiques du premier ordre**
  - Très variables d'une image à l'autre
  - Peu informatives sur leur structure

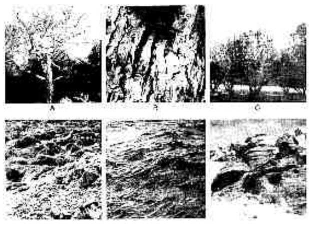




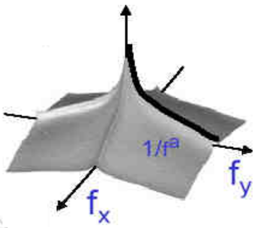
[Gousseau, 2008]

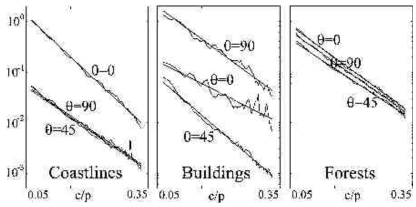
Statistiques des scènes naturelles

- **Statistiques du second ordre**
  - Corrélation
  - Spectre de puissance
- **Invariance à l'échelle**
  - Spectre de puissance: loi en $1/f^a$ (a~2)
  - Anisotropie; liens aux catégories perceptives [Oliva *et al.*, 1999]

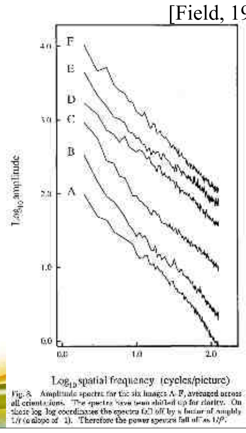


[Field, 1987]



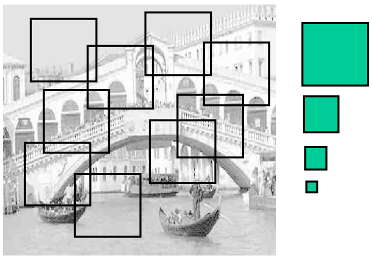


[Oliva Torralba 2001]

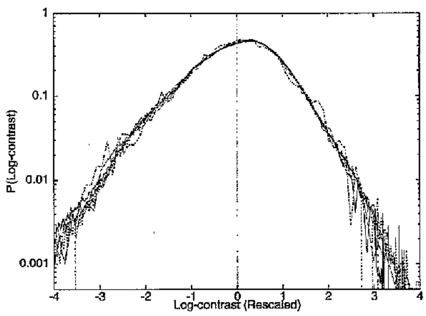


Redondance dans les scènes naturelles

- **Corrélations spatiales→ redondance [Ruderman, 1994]**
  - Moyenne sur Log(niveau de gris) par intégration



$L(x,y)=\ln(I(x,y)/I_0)$
« Log_Contrast »

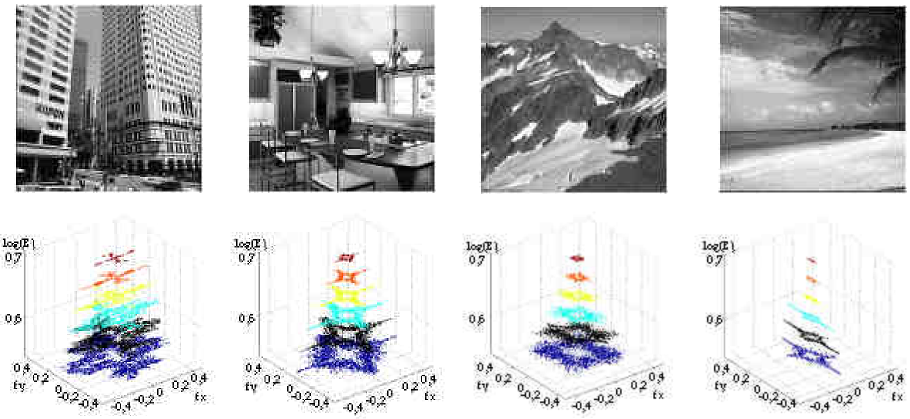


Histogramme (log-contrast) à différentes échelles 1, 2, 4, 8, 16, 32

- **Queues droites de l'histogramme**
  - → Invariance à l'échelle

Statistiques globales

[Oliva *et al.*, 1999]



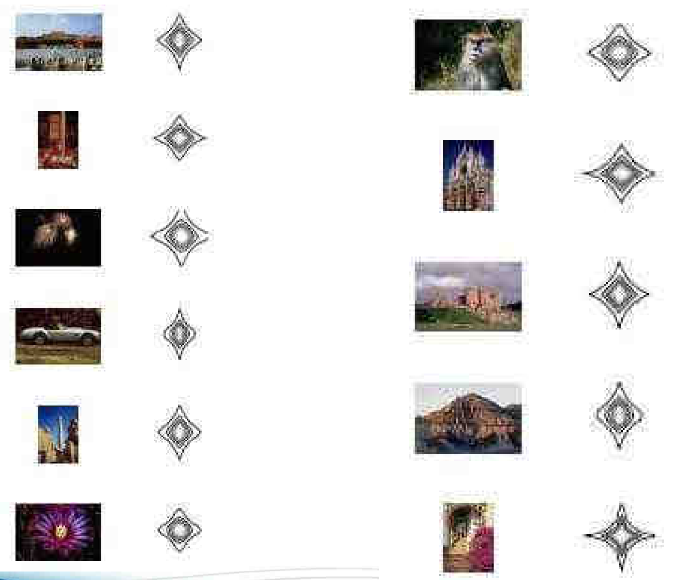
Urbain

Intérieur

Paysage « fermé »

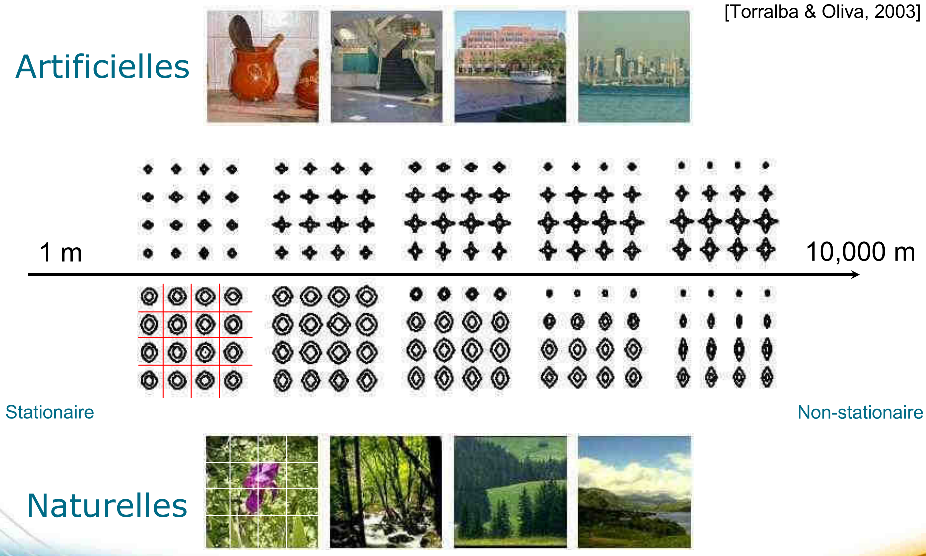
Paysage « ouvert »

Spectres prototypiques



Statistiques locales

[Torralba & Oliva, 2003]



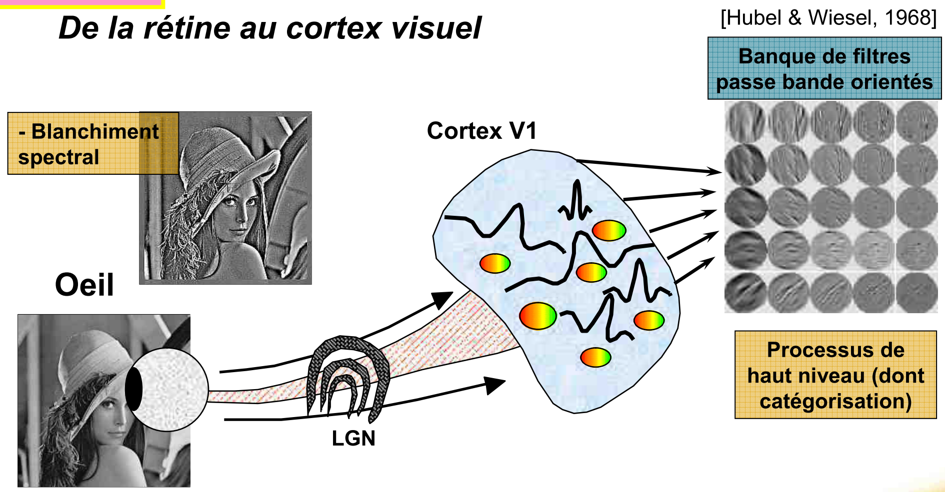
Vers un modèle de discrimination

- **De la rétine à V1**
- **Principes d'un modèle de discimination**
- **Codage**
- **Classification supervisée**

Chemin visuels

v. présentation J. Hérault

De la rétine au cortex visuel



[J.Hérault, rapport interne SCOPIE, 2002]

De la perception au modèle

- **Système sensoriel = tâche de traitement de l’information [Marr, 1982]**
  - Niveau **conceptuel**: qu’est-ce qui est calculé? Pourquoi?
  - Niveau **algorithmique**: comment est-ce calculé?
  - Niveau **implémentation**: plausibilité neurophysiologique?

De la perception au modèle

- **Codage sensoriel: exploiter la structure des images du monde réel**
  - Réduction de la redondance
    - [Barlow, 1961, 1989]

Niveau conceptuel

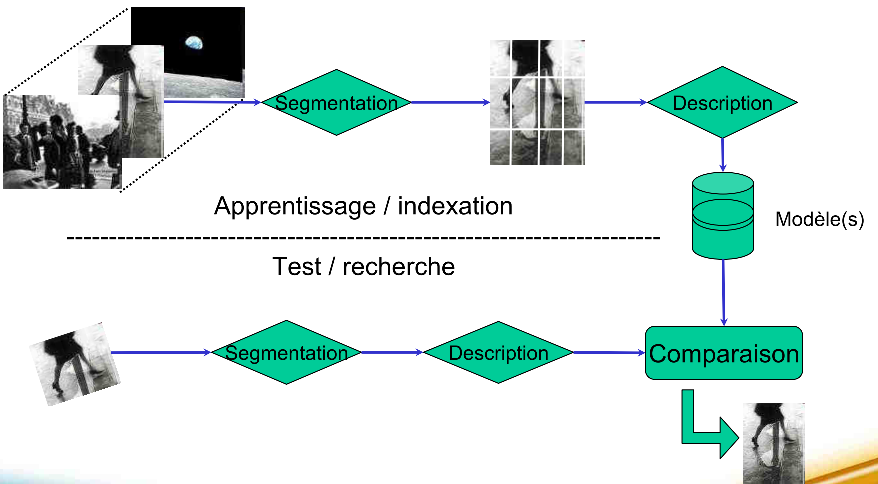
- **Mise en œuvre algorithmique**
  - Décomposition parcimonieuse
    - [Olshausen & Fiedls, 1996]
  - Analyse en composantes indépendantes
    - [Hérault, Jutten, Ans, 1985; Common, 1994]

Niveau algorithmique

Classification supervisée et CBIR

- **Schéma global d'indexation par le contenu**

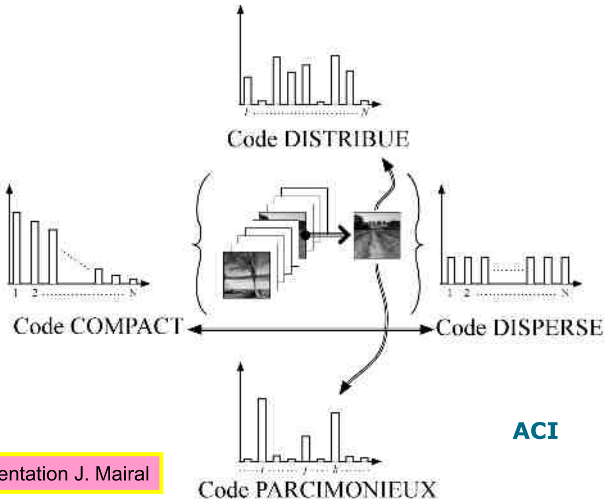
v. présentation H. Jégou



Taxonomie des codes

[Fiedls, 1994]
[Willemore *et al.*, 2000]

ACP



ACI

v. présentation J. Mairal

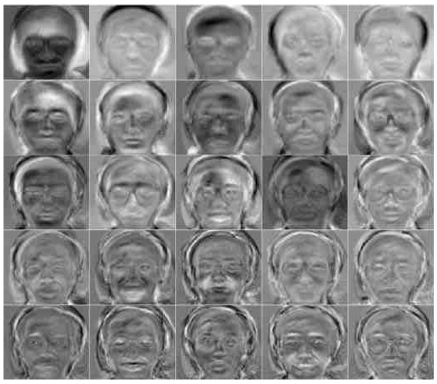
Extraction de composantes indépendantes

- « visages propres » et GIST
- Indépendance statistique
- Analyse en composantes indépendantes

- Modèle génératif
- Méthodes d'extraction
- Propriétés des descripteurs

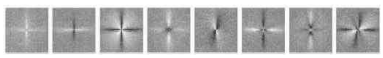
Visages propres

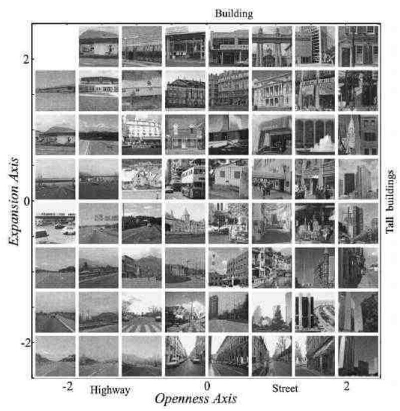
- **Analyse en composante principale**
  - Décorrelation (stat ordre 2)
  - Projections successives sur les axes de variance maximale
- **Algorithmiquement**
  - Matrice de variance/covariance
  - Diagonalisation
  - Vecteur propres
- **Propriétés:**
  - Encodage dégressif de variance
  - → ordre des valeurs propres!
- **Application à la reconnaissance de visages:**
  - [Sirovich & Kirby, 1987]
  - [Turk & Pentland, 1991]



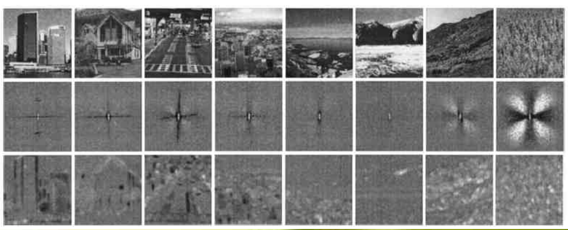
Signature globale de scènes (GIST)

- **Discriminant Spectral Template**
  - Basé sur une décomposition type ACP
  - Analyse des stat. d'ordre 2 (spectre de puissance)
- **En pratique, approximation par filtres de Gabor**
  - Meilleur compromis temps/espace-fréquence (au sens d'Heisenberg) [Gabor,1946]
  - Implémentation
    - 5 fréquences (.02 → 0.32 c/p)
    - 12 orientations
    - 8x8 localisation spatiales
    - ACP → décorrélation





[Oliva *et al.*, 1999]
[Oliva, Torralba, 2001]



Analyse en composantes indépendantes

- **Provient de recherche en neuroscience [Hérault, Jutten Ans, 1984]**
- **Illustration: *cocktail party problem***

- **Problème de la séparation de source [Jutten, Hérault, 1991]**
  - N sources indépendantes
  - P observations via P « micros »
  - But = retrouver les sources indépendantes
- **Indépendance…**
  - Non accoustique (discours « cohérent »…)
  - Non acoustique (même langue…)
  - → Statistique!

ACI: indépendance statistique

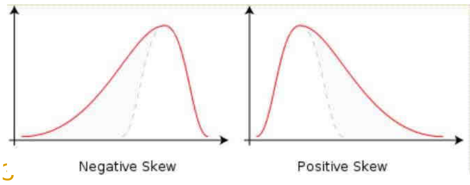
- **Définition: X_1 et X_2 sont indépendantes *ssi* $p(X_1,X_2)=p(X_1)p(X_2)$**
  - Densité jointe = produit densités marginales
- **Fonctions caractéristiques:**
  - Première espèce: $\Phi(t)=E[e^{itx}]$
  - Deuxième espèce: $\Psi(t)=\log(\Phi(t))$

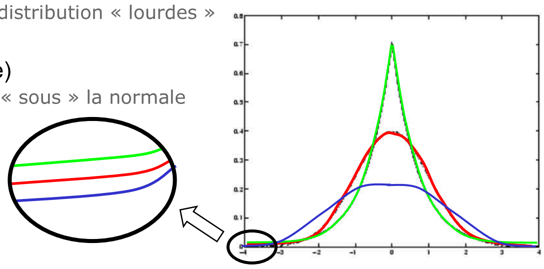
[P. Common, HDR, 1995]

- **Moments et cumulant (développement en série de Mc Laurin)**
  - $\mu_n(X)=(-i)^n\Phi^{(n)}(0)$
  - $\gamma_n(« X »_n)=(-i)^n \Psi^{(n)}(0)$
  - Lien avec $\mu_n(X)=E[X^n]$
  - Définition avec des "fonctions génératrices de moments/cumulants"
    - $M_X(t)=E[e^{tX}]$ et $C_X(t)=\log(E[e^{tX}])$ → $\mu_n(X)=M_X^{(n)}(0)$ et $\gamma_n(X)= C_X^{(n)}(0)$
- **Cas vectoriel: idem (en dimension p)!**
  - Moment croisé d'ordre n: $E[X_1^{k1}X_2^{k2}\ldots X_p^{kp}]$ avec $k_1+\ldots+k_p=n$
    - Ex (p=2, n=4): $E[X^4]$, $E[X^3Y]$,$E[X^2Y^2]$, $E[XY^3]$, $E[Y^4]$
  - Les cumulants croisés s'expriment en fonction des moments croisés:
    - $\mathrm{Cum}(X,X,X,X) = \mu_{1111}-3\mu_{11}\mu_{11} \quad =E[X^4] - 3E[X^2]^2$
    - $\mathrm{Cum}(X,X,X,Y) = \mu_{1112}-3\mu_{11}\mu_{12} \quad =E[X^3Y] - 3E[X^2]E[XY]$
    - $\mathrm{Cum}(X,X,Y,Y) = \mu_{1122}-\mu_{11}\mu_{22}-2\mu_{12}\mu_{12}=E[X^2Y^2] -E[X^2]E[Y^2]-2E[XY]^2$

ACI: indépendance statistique

- **Moment centrés: $m_n=E[(X-\mu_1)^n]$**
- **Quelques moments et cumulants connus**
  - Moyenne: μ_1 (donc $m_1=0$)
  - Variance: $m_2=E[(X-\mu_1)^2]=\sigma^2$
  - Asymétrie (skewness): $m_3/\ \sigma^3$
  - Aplatissement (kurtosis): $\gamma_4\ /\gamma_2^2=m_4/\ \sigma^4$ -3
    - Fonction sur-gausienne = queues de distribution « lourdes »
      - S Kurtosis positif
      - S Leptokurtic (« lepto- » = mince)
    - Fonction sous-gaussiennes = queues « sous » la normale
      - S Kurtosis négatif
      - S Platykurtic (« platy- » = large)
    - Fonction gaussienne
      - S Kurtosis nul
      - S Mesokurtic





- **Si X a une distribution gaussienne, $\gamma_n=0$ pour n>2**
  - Statistiques d'*ordre supérieur* = supérieure à 2
  - Des gaussiennes décorrélées sont indépendantes

ACI: indépendance statistique

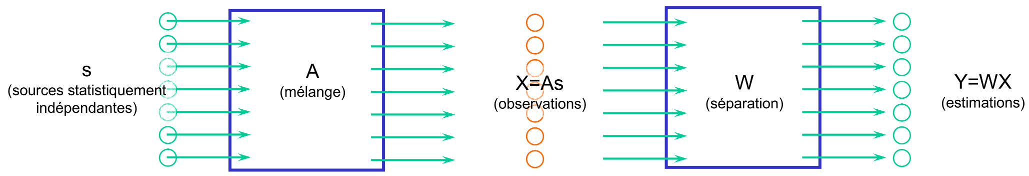
- **X et Y indépendantes → tous les cumulants (et moments) croisés sont nuls**
- **Si X a une distribution gaussienne, $\gamma_n=0$ pour n>2**
  - Statistiques d'*ordre supérieur* = supérieure à 2
  - Des gaussiennes décorrélées sont indépendantes
- **Fonction de contraste (ou *contraste*):**
  - Soit $\mathcal{H}$ les transformation de $\Re^N$ et $\mathcal{T}$ celles qui laissent invariant le sous-ensemble ($\mathcal{S}$) des vecteurs aléatoires à composantes mutuellement indépendantes. $\mathcal{H}s$=image de $\mathcal{S}$ par $\mathcal{H}$.
  - Contraste = application $\mathcal{H} \times \mathcal{H}s \rightarrow$ R
    - Domination: $\forall$ H$\in \mathcal{H}$ et $\forall$ s $\in \mathcal{S}$, $\xi(H,s) \le \xi(Id,s)$
    - Invariance: $\forall$ H$\in \mathcal{T}$, $\xi(H,s) = \xi(Id,s)$
    - Discrimination: $\forall$ s $\in \mathcal{S}$, $\xi(H,s) = \xi(Id,s)$ => H$\in \mathcal{T}$
  - Exemple: information mutuelle = KL(Proba conjointe, produit marginales)

$$I(p_s) = \int f_s(u) \log \frac{f_s(u)}{\prod_{i=1}^{n} f_{s_i}(u_i)} du$$

  - Autres fonctions en pratique…

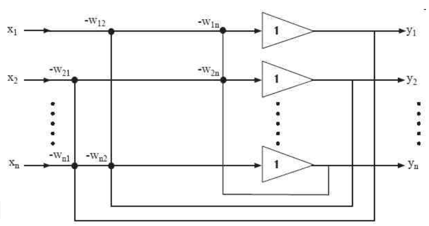
Analyse en composantes indépendantes

- **Provient de recherche en neuroscience [Hérault, Jutten Ans, 1984]**
- **Problème de séparation de source:**



- **Problème mal posé: A *et* s sont inconnues**
  - Ajout de l'hypothèse d'indépendance statistique
- **Cas du mélange linéaire (instantané non bruité) [Common 1994]**
  - Possible si:
    - Au moins autant d'observation que de sources (micros >= discours)
    - Au plus une source est gaussienne
  - Restent deux indéterminations:
    - Ordre des sources
    - Amplitude des sources

ACI: algorithmes



- **Algorithmes**
  - Mesure d'indépendance (contraste)
  - Principe algorithmique
- **Algorithme HJ**
  - Inspiration neuromimétique
  - $Y=(W+I)^{-1}x$ avec $\Delta w_{ij}=f(y_i)g(y_j)$ et $W_{ii}=0$
    - f, g non linéaires impaires. Choix par MV [Pham, 92]
  - Mesure (sous jacente): annulation cumulants
- **Infomax**
  - Implémentation d'un modèle de capacité cognitive par RNA doit maximiser la taux d'information transféré d'une couche à la suivante [Linsker, 1988]
  - Equivalent à la réduction de redondance [Nadal et Parga, 1994]
    - Codage efficace = codage factoriel
  - Règle de type gradient relatif [Bell & Sejnowski, 1998]
    - Gradient: $\Delta W = [I - K\, tanh(y)yT - yyT]W$
    - Equivalent à une approche par MV

ACI: fast ICA [Hyvärinen&Oja, 1997]

- **Mesure d'indépendance par non-gaussianité**
  - Th. Central limite: somme de variable → gaussienne
    - y = Wx = WAs
  - But = maximiser la non-gaussianité de Y
    - Chaque y_i → source indépendante
  - Mesures:
    - Kurtosis: kurt(y) = $E\{y^4\}-2(E\{y^2\})^2$
    - Néguentropie: $J(y)=H(y_{gauss})-H(y)$
    - Approx: $J(y) \approx [E\{G(y)\}-E\{G(u_{gauss})\}]^2$

      Ex: $G(y) = -\exp(-u^2/2)$
- **Algorithme**
  - Point fixe → convergence rapide
    - Dépend des données
  - Package matlab
    - http://www.cis.hut.fi/projects/ica/fastica/
- **Lien avec Poursuite de projection**
  - Recherche d'axes non gaussiens

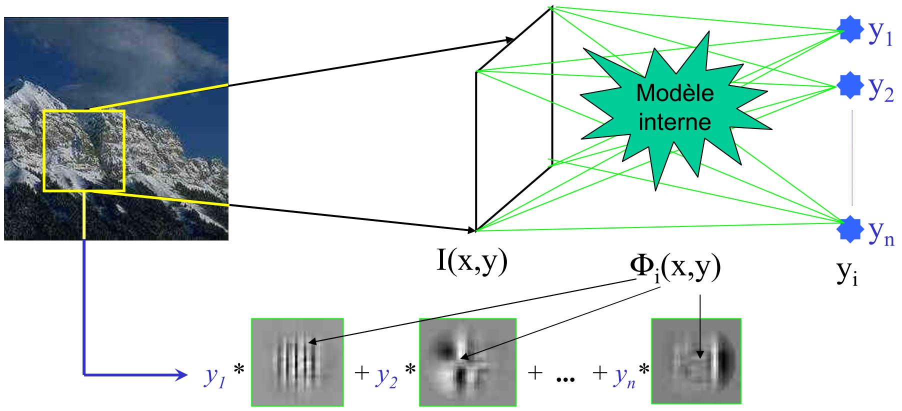
```
W = rand ()                 {Matrice initiale aléatoire}
W = (WW^T)^-1/2 W           {Orthogonalisation}
POUR i : 1 → N_ica
    w_i = w_i / ||w_i||
FIN
t=0
TANT QUE t < t_max          {t_max itérations maximum}
    W_0 = W
    POUR i : 1 → N_ica
        w_i = E{zg(w_i^T z)}-E{g'(w_i^T z)} w_i^T
    FIN
    W = (WW^T)^-1/2 W
    SI 1-min(|diag(W*W_0)|)_i < ε
        RETOUR
    FIN
    t = t + 1;
FIN
```

Application de l'ACI

- **Séparation de signaux**
  - Parole
    - Nécessite des modèles convolutifs
  - Biomédical: signaux EEG, IRMf et MEG
    - [Beckman Smith, 2003]
- **Données financières**
  - La prédiction financière est un travail de longue haleine…
- **Extraction de caractéristiques d'images naturelles**
  - Cf après…
- **Classification d'images**
  - Cf après…
  - « Visages indépendants » [Barlett et al, 1998]
- **Compression d'images**
  - Comparable à JPEG mais moins bien que JPEG2000
- **Débruitage d'image [Hyvarinen et al. 2001]**
- **Estimation de transparence**
  - Modèle additif [Farid & &Adelson, 1999]

Modèle de génération

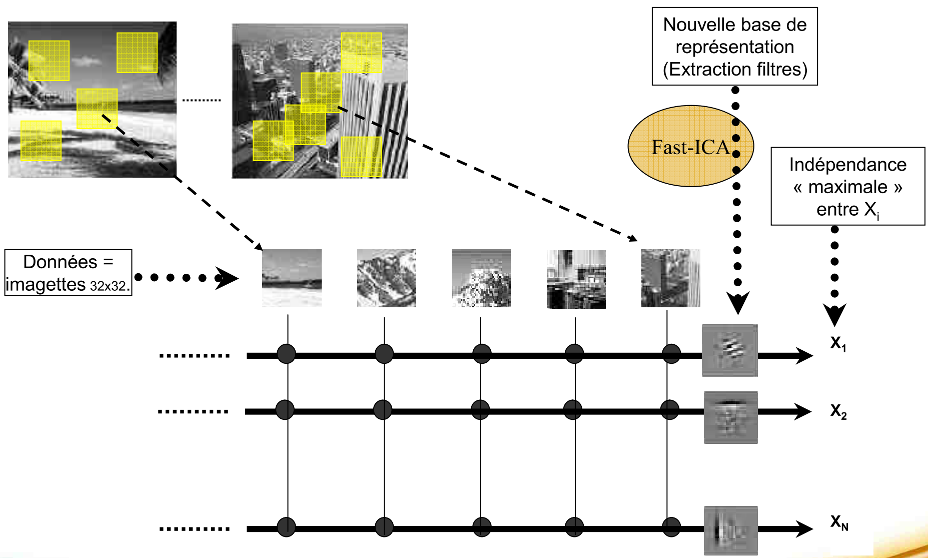
[Olshausen & Field, 1996]



Décomposition linéaire dans la base de fonctions :

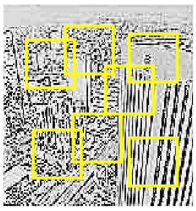
$$I(x,y) \approx \sum_{i=1}^{N} y_i . \phi_i(x,y)$$

ACI appliquée aux images



Indépendance stat → diminution de la redondance

Prétraitements possibles des données



Imagettes

• Limite la dynamique fréquentielle maximale

Apodisation : Focalisation

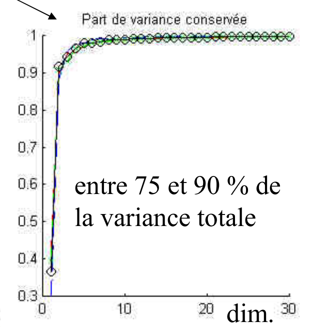
• Évite les effets de bord

1 1024
1
10.000

10.000 observations / 1024 dim.

ACP → ACI

• Dimension significative < 750



entre 75 et 90 % de la variance totale

cea list

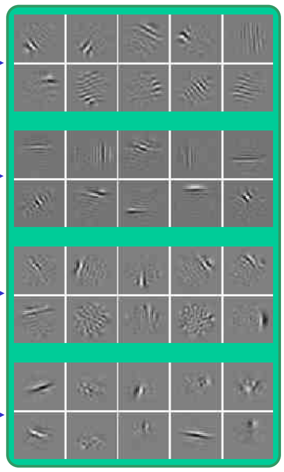
Méthode d'extraction

Prétraitements → ACI

Prétraitements → ACI

Prétraitements → ACI

Prétraitements → ACI



ACI

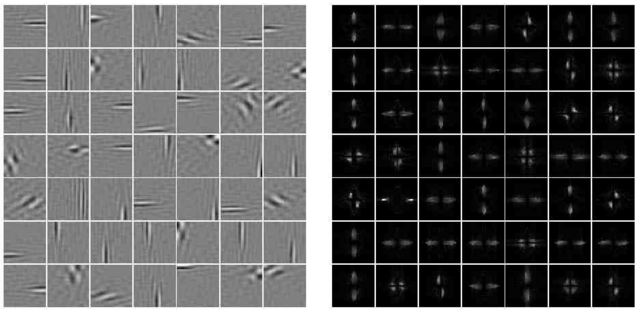
[Bosh & Labbi, 1999,2001]

• Filtres « par catégorie »
• Filtres « toutes catégories »

cea list

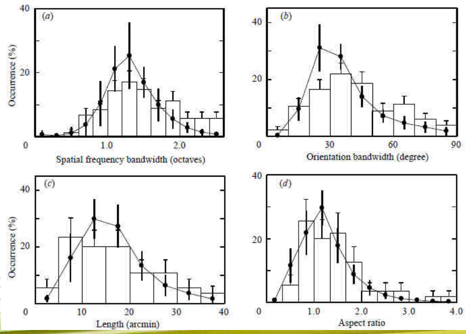
Filtres ACI et physiologie

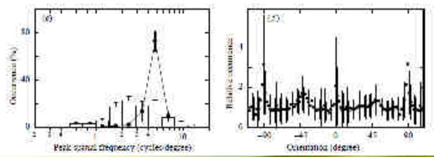
- **Filtres et module de leur TF**
  - Passe-bande
  - Orientés
  - Localisés
- **Ressemble à:**
  - Gabor
  - Cellules simples V1
- **Comparaison physiologique**
  - Donnée du cortex de macaque
  - Congruences
    - Largeur de bande
    - Taille relative / ratio
  - Différences
    - Pic fréquence centrale
    - ±Orientation



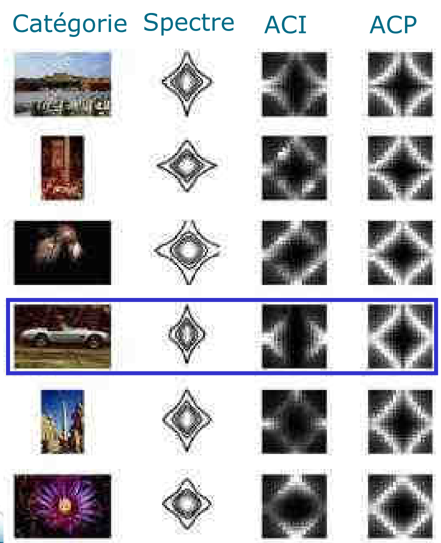
[DeValois et al., 1982]

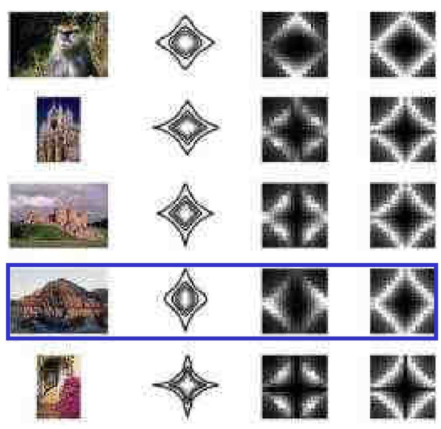
[Van Hateren & Van der Schaaf, 1998]





Adaptation aux spectres: ACP versus ACI

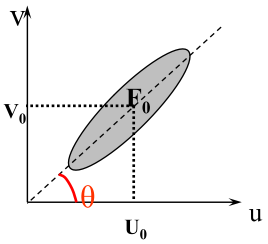




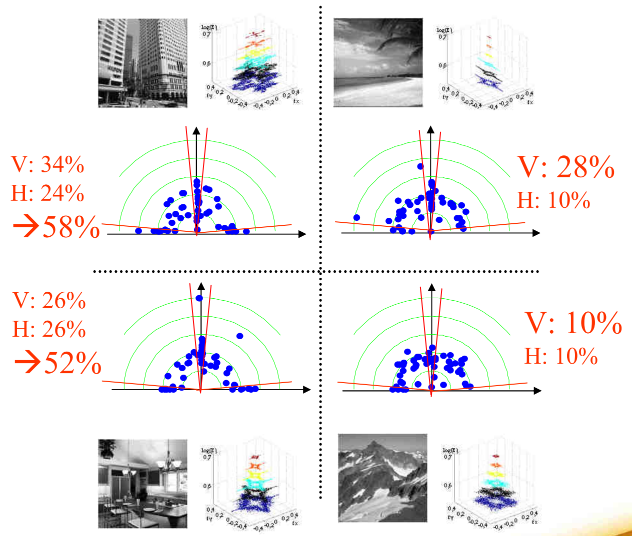
→ Importance des statistiques d'ordre supérieur

Adaptation aux spectres

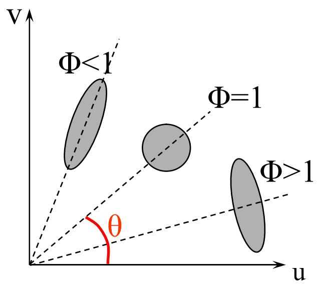
Cellules $V_1 \Leftrightarrow$ Gabor 2D



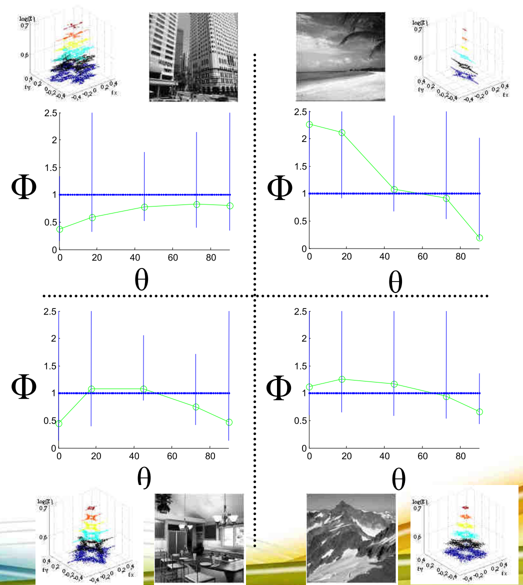
➔Émergence de descripteurs adaptés à la discrimination sans supervision



Sélectivité en orientation

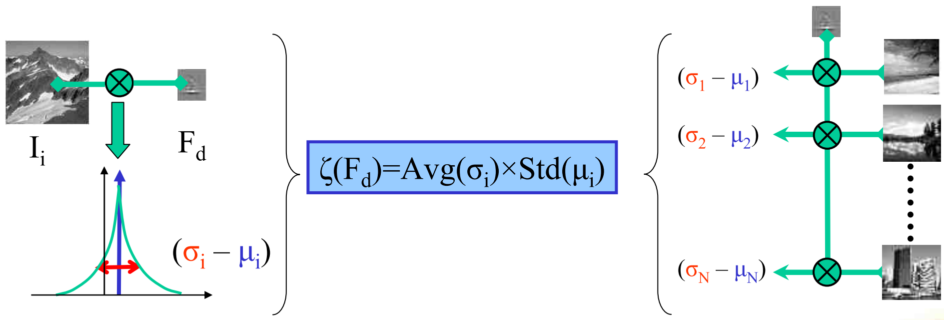


Φ : facteur de forme



Sélection de filtres

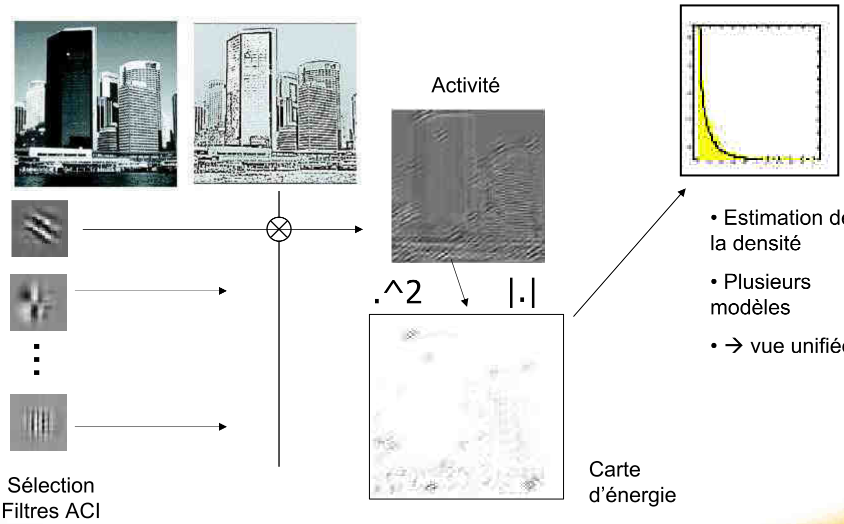
- **Nécessité pour réduire la dimension de description**
- **Recherche exhaustive impossible → critère de sélection ("sous optimal")**
- **<u>Proposition</u>: *dispersal factor***
  - **→ Les filtres les plus utiles sont ceux dont la réponse est la plus variée (sur base d'apprentissage)**



Classification de scènes

- **Définition de signatures d'image**
  - Modélisation de la densité
  - ActivMax
- **Evaluation**

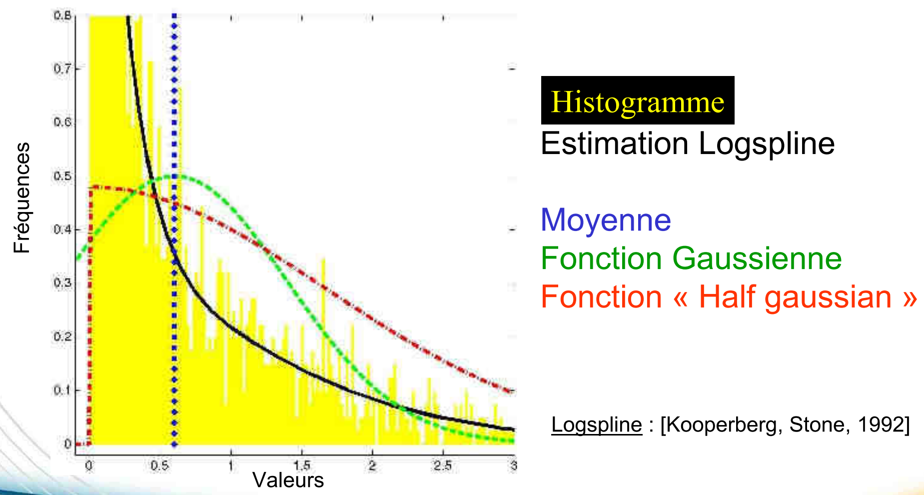
Implémentation densité (1/3)



- Estimation de la densité
- Plusieurs modèles
- → vue unifiée

Implémentation densité (2/3)

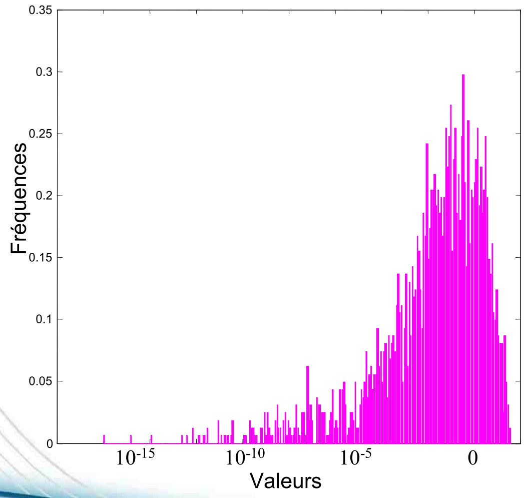
- **Estimation de la densité de probabilité**



Logspline : [Kooperberg, Stone, 1992]

Implémentation densité (2/3)

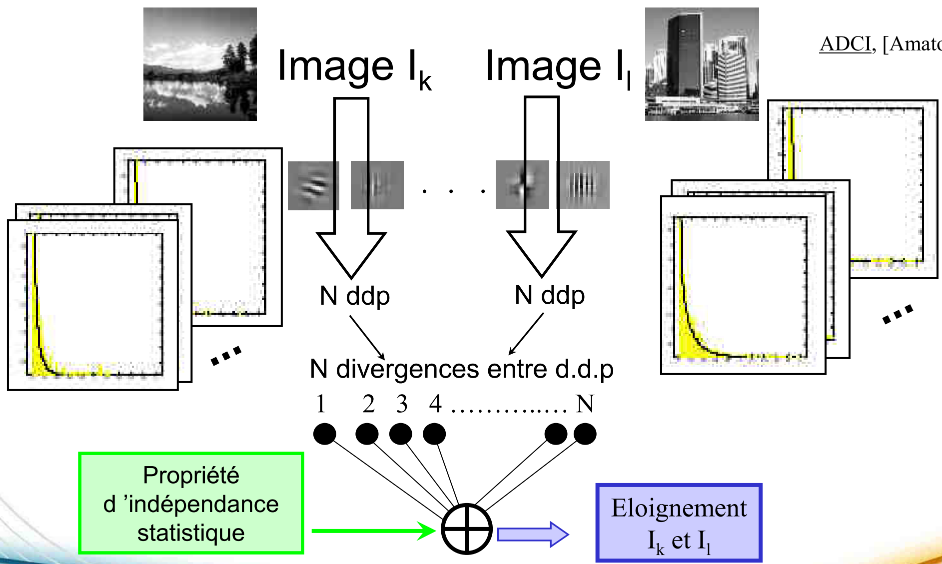
- **Estimation de la densité de probabilité**



Histogramme logarithmique

Implémentation densité (3/3)

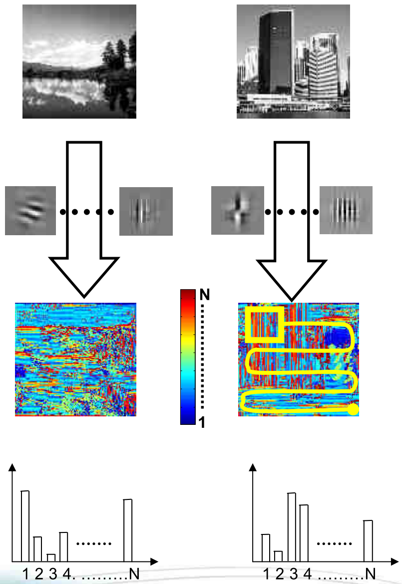
ADCI, [Amato *et al.*, 2002]



Implémentation par activité maximale

Signature Globale

Signature Localisée



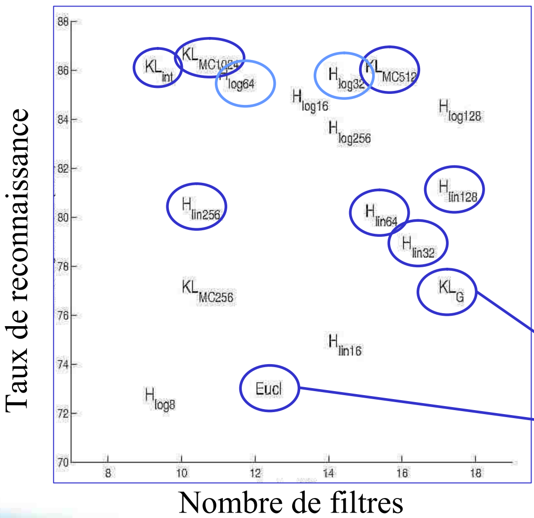
NUMERO de filtre le plus actif en chaque pixel

Fenêtre glissante: reflète la structure des images *localement*

Histogramme
(bag-of-activity)

Adaptation des filtres aux statistiques des catégories

Evaluation: comparaison des signatures



KL_{MCp} – Logspline, Échantillonnage «Monte-Carlo»

Kl_{int} – Logspline, formulation « intégrale »

H_{logN} – histogramme logarithmique

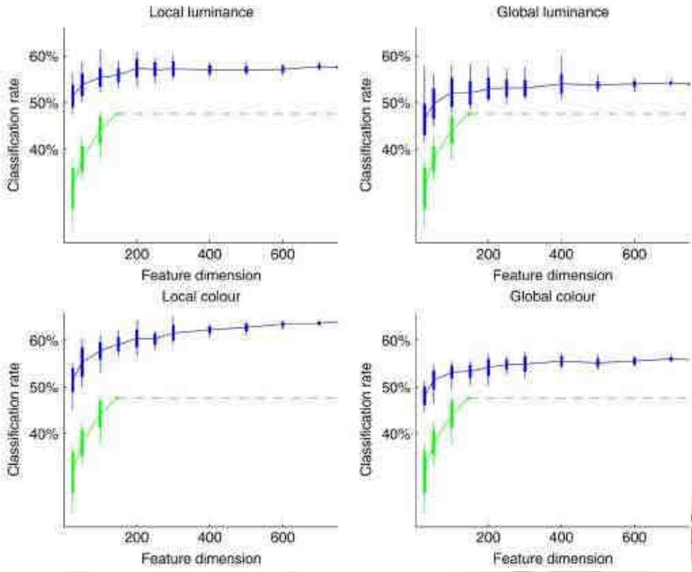
H_{linN} – histogramme linéaire

KL_G – moyenne et écart-type

Eucl – moyenne

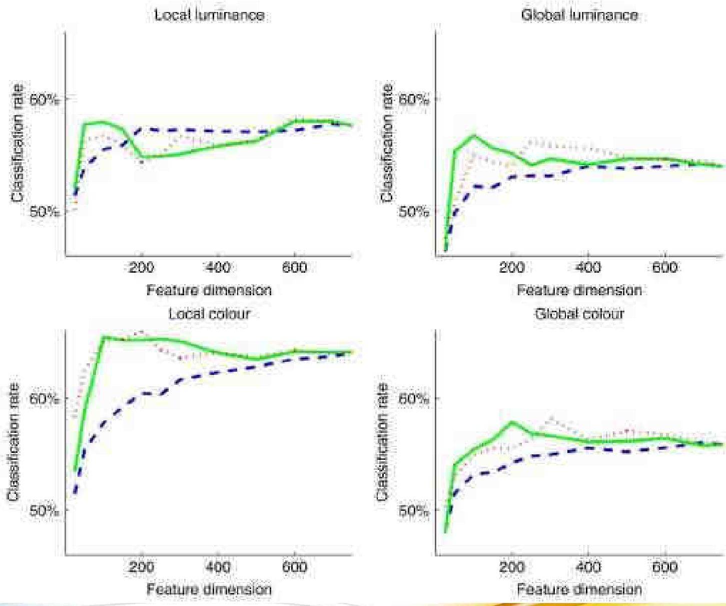
Influence de la taille de signature

- **Sélection aléatoire**
- **20 répétions**
- **Comparaison**
  - ICA
  - EH+SC



Critère de sélection

- **Aléatoire**
- **Dispersal**
- **Critère ζ**



Evaluation

- **Onze catégories COREL (sans recouvrement)**
  - 220 images d'apprentissage
  - 1817 images de test (76 à 521 par catégorie)
- **Description ACI**
  - 50 filtres par catégorie → D=750
  - Ajout couleur (μ et σ de C_b-C_r) → D=754
- **Comparaison**
  - Texture et couleur MPEG-7 (EH, HT, SC, CL)
  - BoSIFT (codebook de taill 50, 100, 200, 1000)
- **Classifieur SVM**
  - libSVM
  - Noyau polynomial d'ordre 3
  - Multiclass: one-versus-one

cea list

Evaluation

Class name	Size	EH	CL	SC	HT	EH + CL	EH + SC	HT + CL	HT + SC	EH + CL + SC	G_{ica}	L_{ica}	G_{ica} + color	L_{ICA} + color
Cities	200	33.9	43.3	19.4	25.0	41.7	42.8	36.7	32.8	51.7	**70.0**	46.1	47.8	69.4
Indoor	541	43.8	25.7	15.4	21.1	39.0	48.0	24.2	21.5	43.4	41.1	51.1	55.3	**66.6**
Firework	100	48.8	85.0	73.8	72.5	**96.3**	**96.3**	73.8	73.8	95.0	47.5	91.3	56.3	88.8
Cars	200	45.0	27.2	26.1	22.8	49.4	51.1	33.9	24.4	52.2	44.4	50.0	50.0	**56.7**
Egypt	100	13.8	31.3	**52.5**	10.0	26.3	25.0	15.0	8.8	31.3	25.0	25.0	36.3	33.8
Flowers	400	41.8	24.5	30.0	30.5	39.5	50.8	33.2	32.9	50.3	**82.9**	81.6	70.5	73.2
Monkeys	100	17.5	25.0	23.8	23.8	25.0	26.3	26.3	28.7	32.5	40.0	40.0	**70.0**	62.5
Churches	96	**46.1**	17.1	15.8	42.1	30.3	42.1	43.4	42.1	34.2	35.5	40.8	42.1	36.8
Castles	100	11.3	17.5	20.0	18.8	16.3	20.0	17.5	15.0	16.3	31.3	**50.0**	20.0	27.5
Mountains	100	37.5	32.5	18.8	23.8	28.7	**45.0**	30.0	33.8	38.8	37.5	36.3	**45.0**	43.8
Doors	100	76.3	60.0	33.8	55.0	72.5	65.0	58.8	56.3	68.8	92.5	91.3	80.0	**93.8**
Total	2037	40.1	31.3	25.6	27.9	41.4	47.7	32.4	30.0	47.1	54.0	57.6	55.6	**63.8**

cea list

Evaluation

Class name	Size	BoK_{50}	BoK_{100}	BoK_{200}	BoK_{1000}	G_{ica}	L_{ica}	G_{ica} + color	L_{ICA} + color
Cities	200	25.0	13.3	11.7	6.7	**70.0**	46.1	47.8	69.4
Indoor	541	21.7	23.8	25.0	6.9	41.1	51.1	55.3	**66.6**
Firework	100	45.0	41.3	37.5	**98.8**	47.5	91.3	56.3	88.8
Cars	200	26.7	36.1	41.7	21.7	44.4	50.0	50.0	**56.7**
Egypt	100	**42.5**	36.3	33.8	28.7	25.0	25.0	36.3	33.8
Flowers	400	20.8	22.6	22.1	3.2	**82.9**	81.6	70.5	73.2
Monkeys	100	15.0	15.0	15.0	2.5	40.0	40.0	**70.0**	62.5
Churches	96	50.0	51.3	51.3	44.7	35.5	40.8	**42.1**	36.8
Castles	100	37.5	36.3	38.8	18.8	31.3	**50.0**	20.0	27.5
Mountains	100	20.0	13.8	17.5	1.3	37.5	36.3	**45.0**	43.8
Doors	100	42.5	41.3	38.8	25.0	92.5	91.3	80.0	**93.8**
Total	2037	26.7	26.7	27.2	15.0	54.0	57.6	55.6	**63.8**

- **Probable problème dans BoSIFT**
  - Echantillonnage « dense » plus adapté

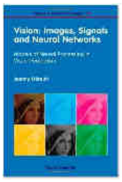
Conclusion

- **Les scènes naturelles ont une structure très contrainte**
- **Un codage efficace des scènes naturelles peut être obtenu par un principe de réduction de redondance**
- **L'ACI permet de mettre en œuvre un tel principe**
- **Les unités codantes résultantes d'adaptent aux statistiques des (catégories des) images**
- **Plusieurs modèles de signatures possibles**

Références: statistiques des scènes naturelles

- Kretzmer, E.R., Statistics of Television Signals, Bell System Tech.(31), No. 4, July 1952, pp. 7551-763
- De Valois, R.L. and De Valois, K.K., 1988. Spatial vision. , Oxford University Press, New York pp. 1–381 .
- A. Oliva, A. Torralba « Modeling the shape of the scene: a holistic representation of the spatial envelope » International Journal of Computer Vision, Vol. 42(3): 145-175, 2001.
- Olshausen BA, Field DJ (1996). « Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images » Nature, 381: 607-609
- D. L. Ruderman, "The statistics of natural images," Network 5, 517–548 (1994)
- E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. 24, 1193–1216 (2001)
- A. Torralba, A. Oliva « Semantic organization of scenes using discriminant structural templates » ICCV, pp 1253-1258, Korfu, Grece, 1999.
- A. Torralba, A.Oliva, Statistics of Natural Image Categories. Network: Computation in Neural Systems 14 (2003) 391-412

Références... De référence!



- J. Hérault. « Vision: images, signals and Neural networks. Models of neural processing in neural networks ». World scientific, 2010
- P. Comon, ``Quelques Développements Récents en Traitement du Signal,'' Habilitation a Diriger les Recherches, Université de Nice, UFR Sciences, 18 sept 1995

Références: eigenfaces

- M. Kirby and L. Sirovich (1990). "Application of the Karhunen-Loeve procedure for the characterization of human faces". *IEEE Transactions on Pattern analysis and Machine Intelligence* 12 (1): 103–108.
- L. Sirovich and M. Kirby (1987). "Low-dimensional procedure for the characterization of human faces". *Journal of the Optical Society of America A* 4: 519–524. doi:10.1364/JOSAA.4.000519.
- M. Turk and A. Pentland (1991). "Face recognition using eigenfaces". *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. pp. 586–591. http://www.cs.ucsb.edu/~mturk/Papers/mturk-CVPR91.pdf.
- M. Turk and A. Pentland (1991). "Eigenfaces for recognition". *Journal of Cognitive Neuroscience* 3 (1): 71–86.

Références: ACI

- Hérault, C. Jutten, and B. Ans, “Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé,” in Actes du Xième colloque GRETSI, vol. 2, Nice, France, Mai 1985, pp. 1017–1022.
- C. Jutten and H. Hérault, “Blind separation of sources, part i: an adaptative algorithm based on neuromimetic architecture,” Signal Processing, vol. 24, pp. 1–10, 1991
- Jutten, Ch. and Taleb, A. “Source separation: From dusk till dawn *ICA 2000, pages 15-26 (invited paper), Helsinki, Finland, June 2000.*
- P. Comon, “Indépendant component analysais - a new concept?” Signal Processing, vol. 36, no. 3, pp. 287–314, 1994.
- A. Hyvärinen, J. Karhunen, and E. Oja, “Independent Component Analysis”. John Wiley and Son, 2001
- P.Comon, C. Jutten and H. Hérault, “Blind separation of sources, part II: problem statement,” Signal Processing, vol. 24, pp. 11-20, 1991
- D.T. Pham, P. Garat. IEEE T Signal Processing, 45(7):1712-1725, 1997
- Cardoso J.-F, Souloumiac A., *Blind beamforming for non-gaussian signals*, IEE proceedings-F, 140(6):362-370, 1993
- Linsker R., "Self-organisation in a perceptual network", *IEEE Computer*, 21:105-117, 1988
- Nadal J;-P., Parga N., *Network: computation in neural systems*, 5:565-581, 1994
- Barlow H.B., Sensory Communication, ed. WA Rosenblith, pp 217-34. Cambridge, MA:MIT press, 1961
- Bell T, Sejnowsky T.J., *Neural Computation*, 7:1129-1159, 1995

Références: ACI & physiologie

- DeValois, R. L.,Yund, E.W & Hepler, N. The orientation and direction selectivity of cells in macaque visual cortex. Vision Research. 22, 531-544, 1982.
- DeValois, R. L., Albrecht, D. G. & Thorell, L. G. Spatial frequency selectivity of cells in macaque visual cortex. Vision Research. 22, 545-559, 1982.
- J. H. van Hateren* and A. van der Schaaf, Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B 7, vol. 265, no. 1394, pp 359-366, March 1998*

Références: ACI appliquée à la discrimination de scènes

- Le Borgne H., Guérin-Dugué A. Sparse-Dispersed Coding and Images Discrimination with Independent Component Analysis. Third International Conference on Independent Component Analysis and Signal Separation, San Diego, California, December 9-13, 2001
- U. Amato, A. Antoniadis, and Grefoire G. Independent component discriminant analysis. International Mathematical Journal, pages 735–753, 2002.
- H. Le Borgne, A. Guérin-Dugué, N.E. O'Connor Learning Mid-level Image Features for Natural Scene and Texture Classification IEEE transaction on Circuits and Systems for Video Technology, 17(3):286-297, march 2007
- Le Borgne H., Guérin-Dugué A., Antoniadis A. Representation of images for classification with independent features Pattern Recognition Letters, vol 25, N° 2, pp 141-154, january 2004.

Julien MAIRAL

ENS / INRIA WILLOW

http://www.di.ens.fr/~mairal

"Sparse Coding and Dictionary Learning"

La modélisation parcimonieuse de signaux consiste à représenter des données vectorielles comme une combinaison linéaire d'un petit nombre d'éléments d'un dictionnaire. Définir un dictionnaire adapté à une classe de signaux telle que les images naturelles, a donné lieu à de nombreux travaux. Nous nous intéresserons ici à une approche récente qui consiste à apprendre le dictionnaire à partir de données d'entraînement. Nous présenterons de récentes avancées utilisant cette technique en traitement d'image, apprentissage statistique et vision par ordinateur pour la reconnaissance d'objets. Ce tutoriel est structuré en 4 parties :

1) Sparse coding and dictionary learning for Image processing
   - Image denoising
   - Inpainting, demosaicking
   - Extensions to video processing
   - Other applications, deblurring, inverse haltoning

2) Sparse linear models and the dictionary learning formulation
   - Why does the l1-norm induce sparsity?
   - Sparsity-inducing norms and group-sparsity.
   - Dictionary learning and matrix factorization (PCA, NMF, hard/soft clustering)
   - New sparse models and structured sparsity.

3) Applications to computer vision
   - Learning codebooks for image classification
   - Modelling the local appearance of image patches

4) Optimization for sparse methods
   - Greedy algorithms
   - l1-optimization
   - stochastic optimization for dictionary learning

Références :

J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman. Supervised Dictionary Learning. Advances Neural Information Processing Systems, 2008. Vancouver. Canada.

J. Mairal, G. Sapiro and M. Elad. Learning multiscale sparse representations for image and video restoration. SIAM Multiscale Modeling and Simulation. Vol 7. No 1.2008. p 214-241.

J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online Learning for Matrix Factorization and Sparse Coding Journal of Machine Learning Research, volume 11. pages 19-60. 2010.

Sparse Coding and Dictionary Learning for Image Analysis

Julien Mairal

Willow group - INRIA - ENS - Paris

ERMITES 2010, September 2010

ERMITES 2010
At Hyères, September 28 - 30, 2010

Acknowledgements


Francis Bach


Jean Ponce


Rodolphe Jenatton


Guillaume Obozinski


Guillermo Sapiro


Florent Couzinié-Devy


Y-Lan Boureau

What this lecture is about?

- **Why sparsity, what for and how?**
- **Signal and image processing**: Restoration, reconstruction.
- **Machine learning**: Selecting relevant features.
- **Computer vision**: Modelling the local appearance of image patches.
- **Computer vision**: Recent (and intriguing) results in bags of words models.
- **Optimization**: Solving challenging problems.

1. Image Processing Applications
2. Sparse Linear Models and Dictionary Learning
3. Computer Vision Applications
4. Optimization for sparse methods

1. Image Processing Applications
   - Image Denoising
   - Inpainting, Demosaicking
   - Video Processing
   - Other Applications
2. Sparse Linear Models and Dictionary Learning
3. Computer Vision Applications
4. Optimization for sparse methods

The Image Denoising Problem


$$\underbrace{\mathbf{y}}_{\text{measurements}} = \underbrace{\mathbf{x}_{orig}}_{\text{original image}} + \underbrace{\mathbf{w}}_{\text{noise}}$$

Sparse representations for image restoration

$$\underbrace{\mathbf{y}}_{\text{measurements}} = \underbrace{\mathbf{x}_{orig}}_{\text{original image}} + \underbrace{\mathbf{w}}_{\text{noise}}$$

Energy minimization problem - MAP estimation

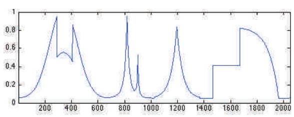
$$E(\mathbf{x}) = \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2}_{\text{relation to measurements}} + \underbrace{Pr(\mathbf{x})}_{\text{image model (-log prior)}}$$

Some classical priors

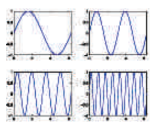
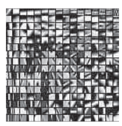
- Smoothness $\lambda\|\mathcal{L}\mathbf{x}\|_2^2$
- Total variation $\lambda\|\nabla\mathbf{x}\|_1^2$
- MRF priors
- ...

What is a Sparse Linear Model?

Let $\mathbf{x}$ in $\mathbb{R}^m$ be a signal.



Let $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_p] \in \mathbb{R}^{m \times p}$ be a set of normalized "basis vectors".
We call it **dictionary**.

$\mathbf{D}$ is "adapted" to $\mathbf{y}$ if it can represent it with a few basis vectors—that is, there exists a **sparse vector** $\boldsymbol{\alpha}$ in $\mathbb{R}^p$ such that $\mathbf{y} \approx \mathbf{D}\boldsymbol{\alpha}$. We call $\boldsymbol{\alpha}$ the **sparse code**.

$$\underbrace{\begin{pmatrix} \mathbf{y} \end{pmatrix}}_{\mathbf{y}\in\mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}_{\boldsymbol{\alpha}\in\mathbb{R}^p,\ \textbf{sparse}}$$

First Important Idea

Why Sparsity?

A dictionary can be good for representing a class of signals, but not for representing white Gaussian noise.

The Sparse Decomposition Problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\boldsymbol{\alpha})}_{\text{sparsity-inducing regularization}}$$

ψ induces sparsity in $\boldsymbol{\alpha}$. It can be

- the ℓ_0 "pseudo-norm". $\|\boldsymbol{\alpha}\|_0 \triangleq \#\{i \text{ s.t. } \boldsymbol{\alpha}[i] \neq 0\}$ (NP-hard)
- the ℓ_1 norm. $\|\boldsymbol{\alpha}\|_1 \triangleq \sum_{i=1}^p |\boldsymbol{\alpha}[i]|$ (convex),
- ...

This is a selection problem. When ψ is the ℓ_1-norm, the problem is called Lasso [Tibshirani, 1996] or basis pursuit [Chen et al., 1999]

Sparse representations for image restoration

Designed dictionaries

[Haar, 1910], [Zweig, Morlet, Grossman ~70s], [Meyer, Mallat, Daubechies, Coifman, Donoho, Candes ~80s-today]... (see [Mallat, 1999])
Wavelets, Curvelets, Wedgelets, Bandlets, ... lets

Learned dictionaries of patches

[Olshausen and Field, 1997], [Engan et al., 1999], [Lewicki and Sejnowski, 2000], [Aharon et al., 2006] , [Roth and Black, 2005], [Lee et al., 2007]

$$\min_{\boldsymbol{\alpha}_i, \mathbf{D} \in \mathcal{C}} \sum_i \underbrace{\frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2}_{\text{reconstruction}} + \underbrace{\lambda \psi(\boldsymbol{\alpha}_i)}_{\text{sparsity}}$$

- $\psi(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_0$ ("ℓ_0 pseudo-norm")
- $\psi(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$ (ℓ_1 norm)

Sparse representations for image restoration

Solving the denoising problem

[Elad and Aharon, 2006]

- Extract all overlapping 8×8 patches $\mathbf{y}_i$.
- Solve a matrix factorization problem:

$$\min_{\boldsymbol{\alpha}_i, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^{n} \underbrace{\frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2}_{\text{reconstruction}} + \underbrace{\lambda \psi(\boldsymbol{\alpha}_i)}_{\text{sparsity}},$$

  with $n > 100,000$
- Average the reconstruction of each patch.

Sparse representations for image restoration

K-SVD: [Elad and Aharon, 2006]

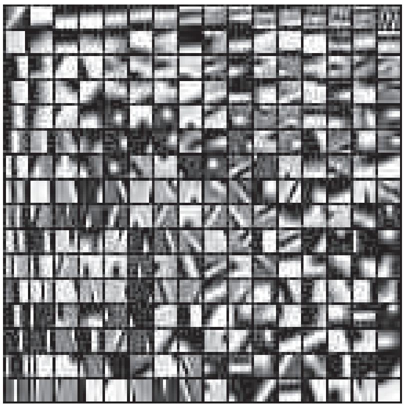
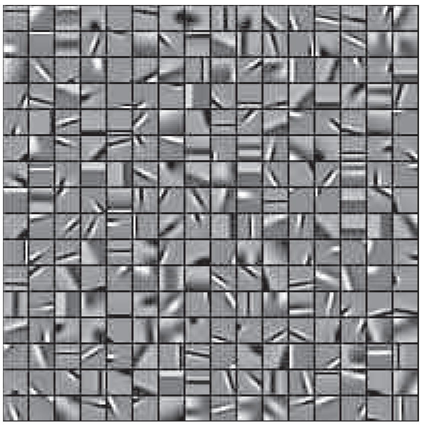
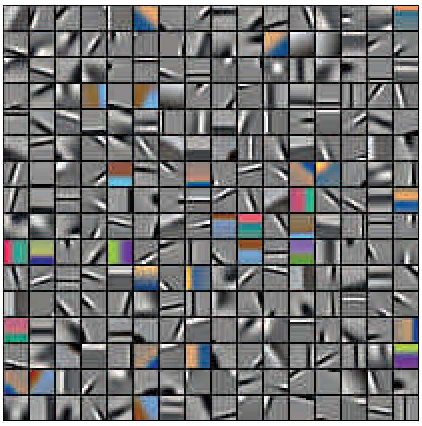


Figure: Dictionary trained on a noisy version of the image boat.

Sparse representations for image restoration

Grayscale vs color image patches



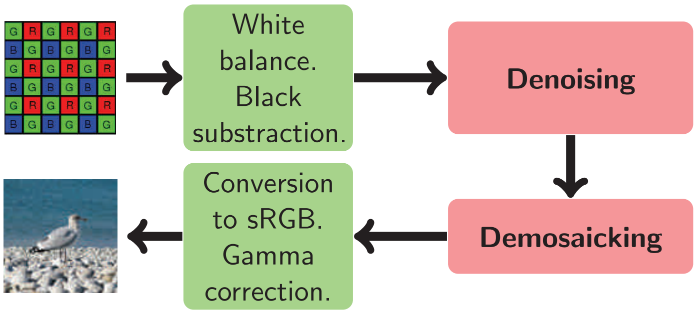


Sparse representations for image restoration

Inpainting, Demosaicking

$$\min_{\mathbf{D}\in\mathcal{C},\boldsymbol{\alpha}} \sum_i \frac{1}{2}\|\boldsymbol{\beta}_i(\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i)\|_2^2 + \lambda_i\psi(\boldsymbol{\alpha}_i)$$

RAW Image Processing



Sparse representations for image restoration

[Mairal, Sapiro, and Elad, 2008d]

Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008b]

Sparse representations for image restoration

Inpainting, [Mairal, Elad, and Sapiro, 2008b]

Sparse representations for video restoration

Key ideas for video processing

[Protter and Elad, 2009]

- Using a 3D dictionary.
- Processing of many frames at the same time.
- Dictionary propagation.

Sparse representations for image restoration

Inpainting, [Mairal, Sapiro, and Elad, 2008d]

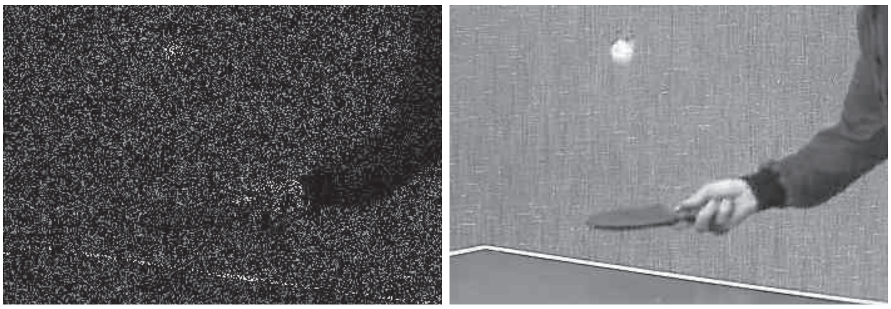


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008d]

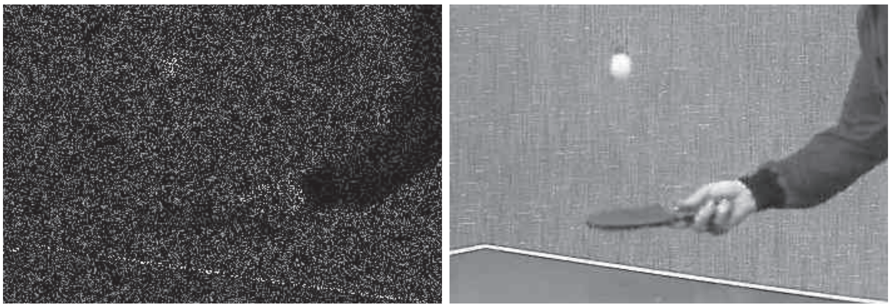


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008d]

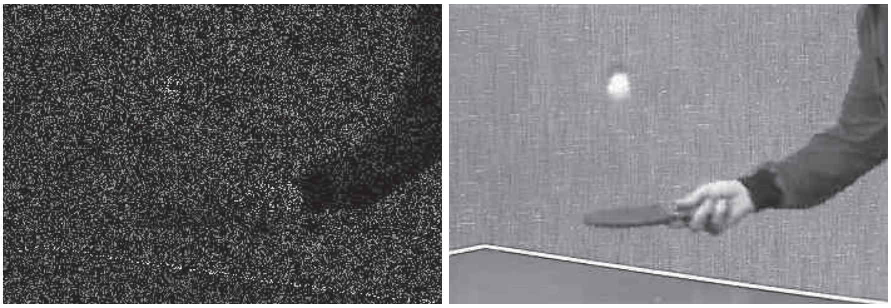


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008d]

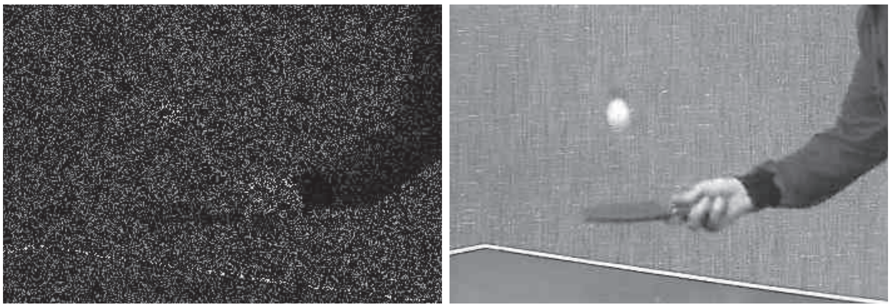


Figure: Inpainting results.

Sparse representations for image restoration
Inpainting, [Mairal, Sapiro, and Elad, 2008d]

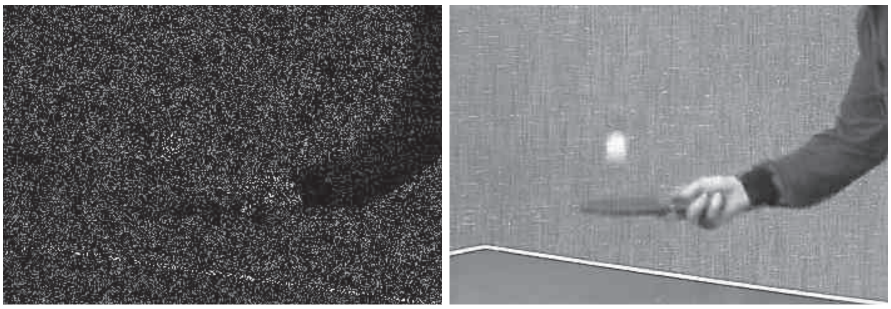


Figure: Inpainting results.

Sparse representations for image restoration
Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration
Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

Figure: Denoising results. $\sigma = 25$

Sparse representations for image restoration

Color video denoising, [Mairal, Sapiro, and Elad, 2008d]

Figure: Denoising results. $\sigma = 25$

Digital Zooming

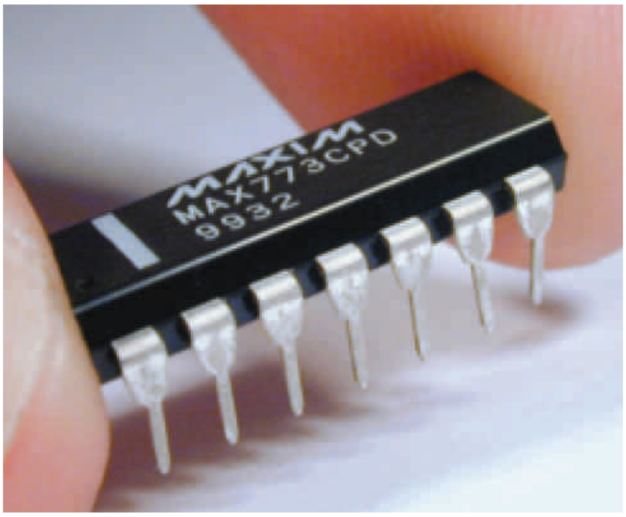
Couzinie-Devy, 2010, Original

Digital Zooming
Couzinie-Devy, 2010, Bicubic

Digital Zooming
Couzinie-Devy, 2010, Proposed method

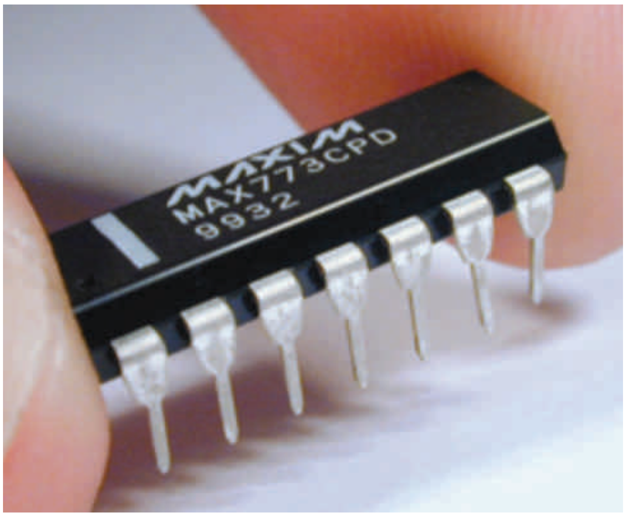
Digital Zooming

Couzinie-Devy, 2010, Original



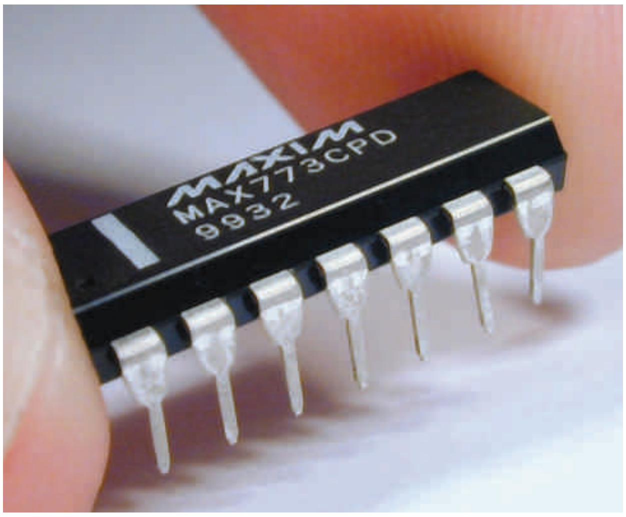
Digital Zooming

Couzinie-Devy, 2010, Bicubic



Digital Zooming

Couzinie-Devy, 2010, Proposed approach



Inverse half-toning

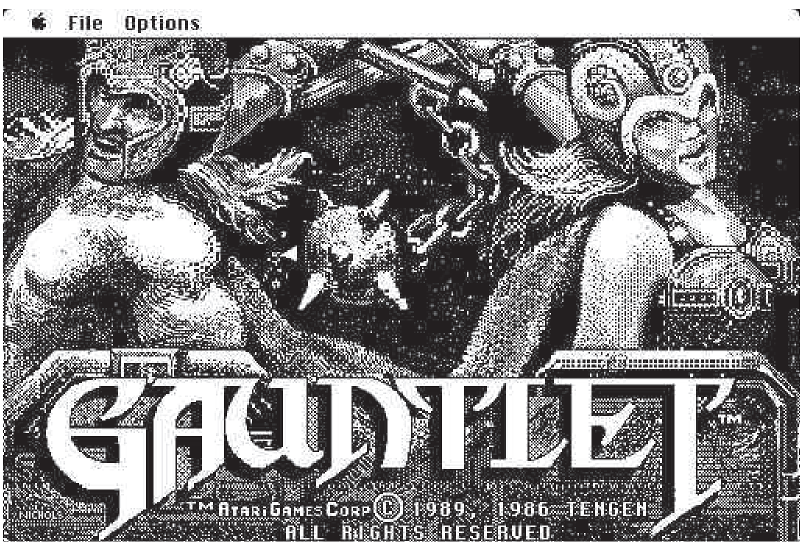
Original

Inverse half-toning

Reconstructed image

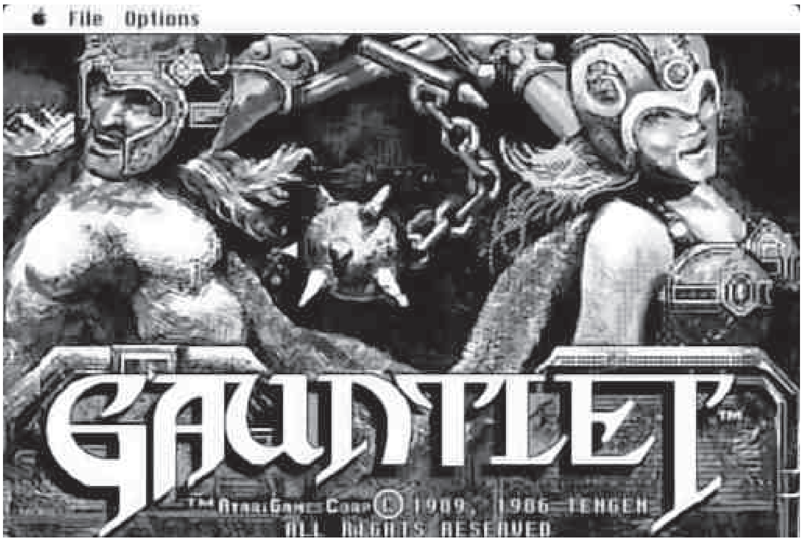
Inverse half-toning

Original



Inverse half-toning

Reconstructed image



Inverse half-toning

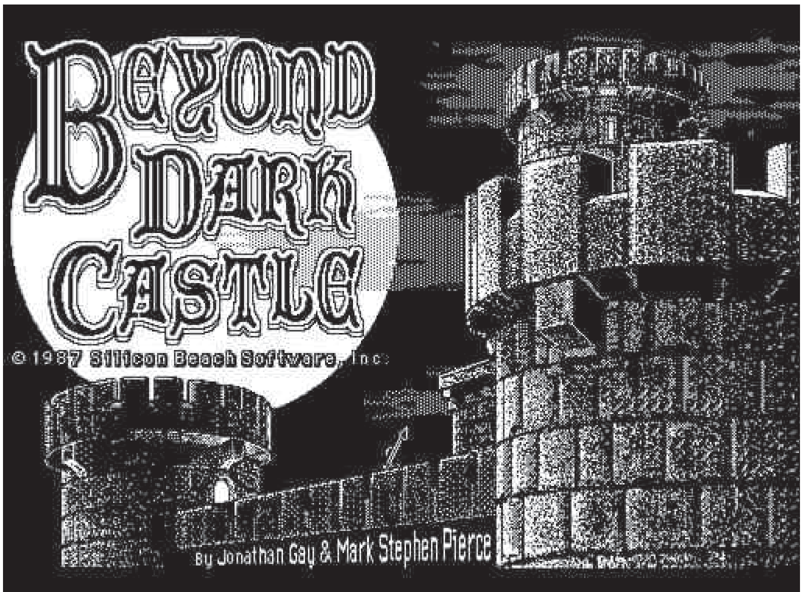
Original

Inverse half-toning

Reconstructed image

Inverse half-toning

Original



Inverse half-toning

Reconstructed image

Inverse half-toning

Original

Inverse half-toning

Reconstructed image

One short slide on compressed sensing

Important message

Sparse coding is not "compressed sensing".

Compressed sensing is a theory [see Candes, 2006] saying that a sparse signal can be recovered from a few linear measurements under some conditions.

- Signal Acquisition: $\mathbf{W}^\top \mathbf{y}$, where $\mathbf{W} \in \mathbb{R}^{m \times s}$ is a "sensing" matrix with $s \ll m$.
- Signal Decoding: $\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \|\boldsymbol{\alpha}\|_1$ s.t. $\mathbf{W}^\top \mathbf{y} = \mathbf{W}^\top \mathbf{D}\boldsymbol{\alpha}$.

with extensions to approximately sparse signals, noisy measurements.

Remark

The dictionaries we are using in this lecture do not satisfy the recovery assumptions of compressed sensing.

Important messages

- Patch-based approaches are achieving state-of-the-art results for many image processing task.
- Dictionary Learning adapts to the data you want to restore.
- Dictionary Learning is well adapted to data that admit sparse representation. **Sparsity is for sparse data only**.

Next topics

- A bit of machine learning.
- Why does the ℓ_1-norm induce sparsity?
- Some properties of the Lasso.
- Links between dictionary learning and matrix factorization techniques.
- A simple algorithm for learning dictionaries.
- Beyond sparsity: Group-sparsity, Structured Sparsity

1. Image Processing Applications
2. Sparse Linear Models and Dictionary Learning
   - The machine learning point of view
   - Why does the ℓ_1-norm induce sparsity?
   - Dictionary Learning and Matrix Factorization
   - Group Sparsity
   - Structured Sparsity
3. Computer Vision Applications
4. Optimization for sparse methods

Sparse Linear Model: Machine Learning Point of View

Let $(y^i, \mathbf{x}^i)_{i=1}^n$ be a training set, where the vectors $\mathbf{x}^i$ are in $\mathbb{R}^p$ and are called features. The scalars y^i are in

- $\{-1, +1\}$ for **binary** classification problems.
- $\{1, \ldots, N\}$ for **multiclass** classification problems.
- $\mathbb{R}$ for **regression** problems.

In a linear model, on assumes a relation $y \approx \mathbf{w}^\top \mathbf{x}$ (or $y \approx \text{sign}(\mathbf{w}^\top \mathbf{x})$), and solves

$$\min_{\mathbf{w} \in \mathbb{R}^p} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(y^i, \mathbf{w}^\top \mathbf{x}^i)}_{\text{data-fitting}} + \underbrace{\lambda \Omega(\mathbf{w})}_{\text{regularization}} .$$

Sparse Linear Models: Machine Learning Point of View

A few examples:

Ridge regression: $$\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{2n}\sum_{i=1}^{n}(y^i - \mathbf{w}^\top \mathbf{x}^i)^2 + \lambda\|\mathbf{w}\|_2^2.$$

Linear SVM: $$\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}\max(0, 1 - y^i\mathbf{w}^\top \mathbf{x}^i) + \lambda\|\mathbf{w}\|_2^2.$$

Logistic regression: $$\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^{n}\log\left(1 + e^{-y^i\mathbf{w}^\top \mathbf{x}^i}\right) + \lambda\|\mathbf{w}\|_2^2.$$

The squared ℓ_2-norm induces **smoothness** in $\mathbf{w}$. When one knows in advance that $\mathbf{w}$ should be sparse, one should use a **sparsity-inducing** regularization such as the ℓ_1-norm. [Chen et al., 1999, Tibshirani, 1996]

The purpose of the regularization is to add **additional a-priori knowledge** in the regularization.

Sparse Linear Models: the Lasso

- Signal processing: $\mathbf{D}$ is a dictionary in $\mathbb{R}^{n\times p}$,

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1.$$

- Machine Learning:

$$\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{2n}\sum_{i=1}^{n}(y^i - \mathbf{x}^{i\top}\mathbf{w})^2 + \lambda\|\mathbf{w}\|_1 = \min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{2n}\|\mathbf{y} - \mathbf{X}^\top\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1,$$

with $\mathbf{X} \triangleq [\mathbf{x}^1, \ldots, \mathbf{x}^n]$, and $\mathbf{y} \triangleq [y^1, \ldots, y^n]^\top$.

Useful tool in signal processing, machine learning, statistics,... as long as one wishes to **select** features.

Why does the ℓ_1-norm induce sparsity?

Exemple: quadratic problem in 1D

$$\min_{\alpha \in \mathbb{R}} \frac{1}{2}(y-\alpha)^2 + \lambda|\alpha|$$

Piecewise quadratic function with a kink at zero.

Derivative at 0_+: $g_+ = -y + \lambda$ and 0_-: $g_- = -y - \lambda$.

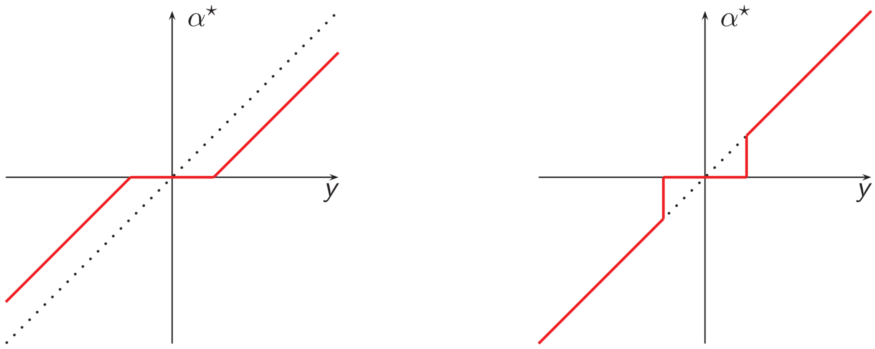
Optimality conditions. α is optimal iff:

- $|\alpha| > 0$ and $(y-\alpha) + \lambda\,\text{sign}(\alpha) = 0$
- $\alpha = 0$ and $g_+ \geq 0$ and $g_- \leq 0$

The solution is a **soft-thresholding**:

$$\alpha^\star = \text{sign}(y)(|y| - \lambda)^+.$$

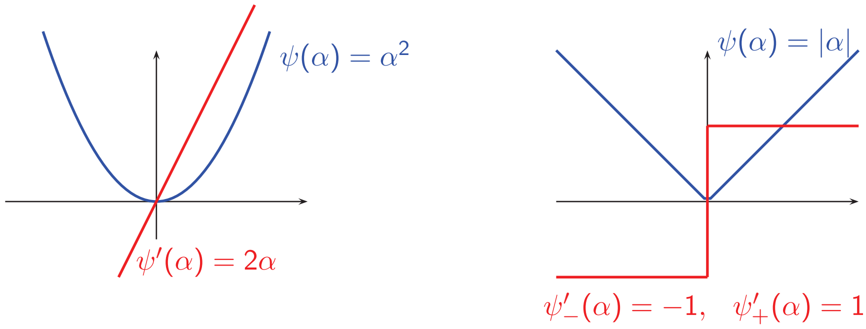
Why does the ℓ_1-norm induce sparsity?



(a) soft-thresholding operator (b) hard-thresholding operator

Why does the ℓ_1-norm induce sparsity?

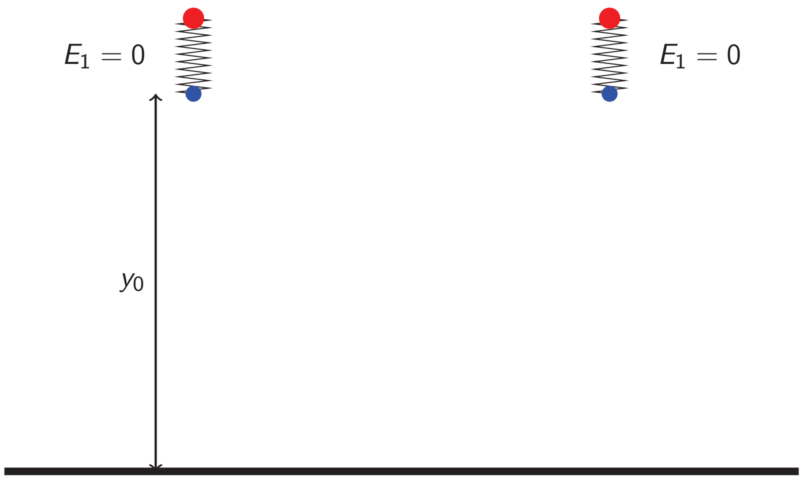
Analysis of the norms in 1D



The gradient of the ℓ_2-norm vanishes when α get close to 0. On its differentiable part, the norm of the gradient of the ℓ_1-norm is constant.

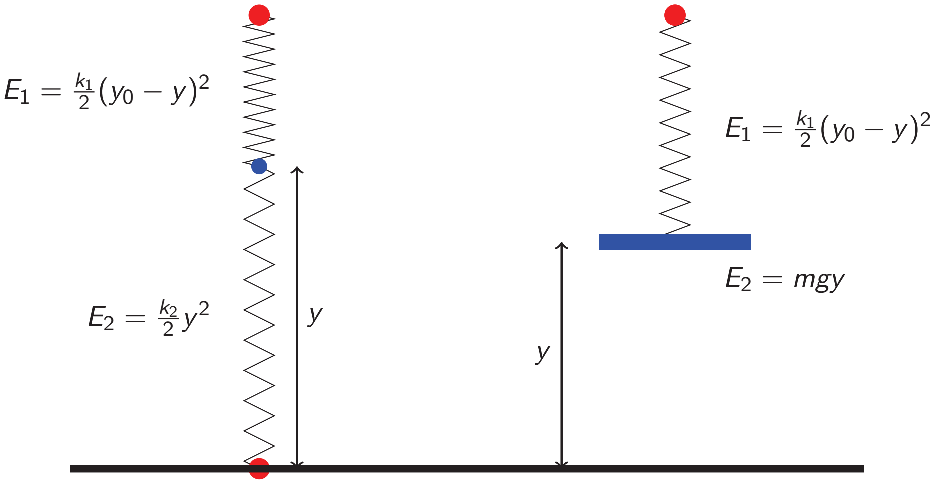
Why does the ℓ_1-norm induce sparsity?

Physical illustration



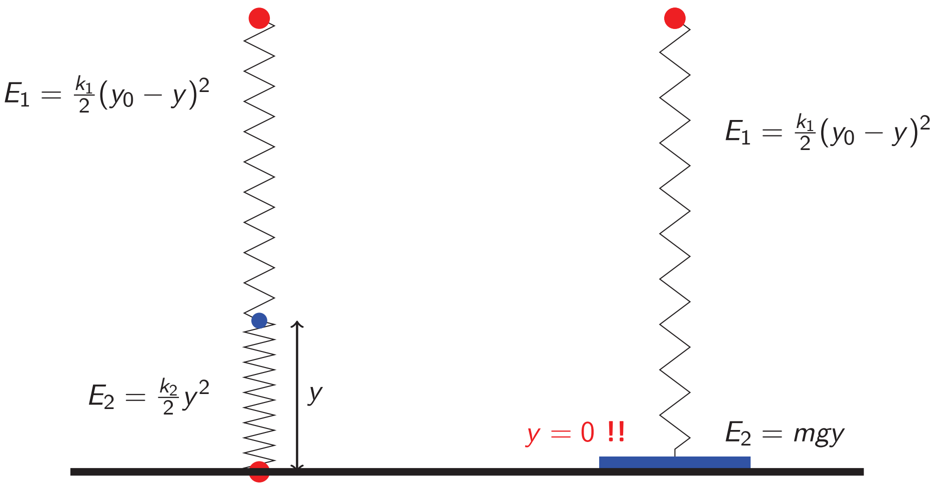
Why does the ℓ_1-norm induce sparsity?

Physical illustration



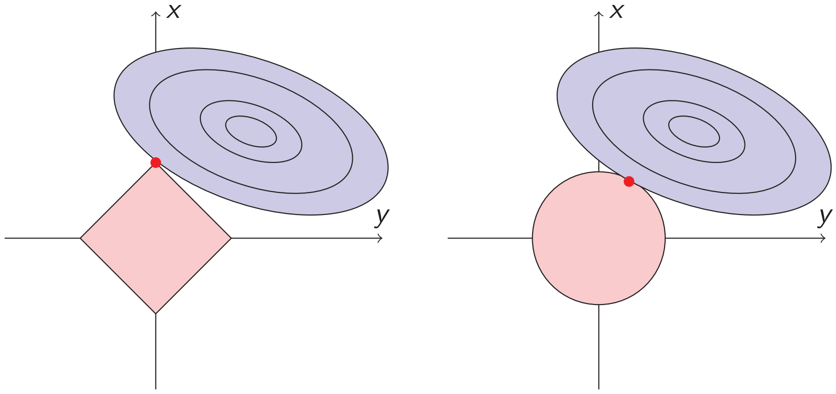
Why does the ℓ_1-norm induce sparsity?

Physical illustration



Why does the ℓ_1-norm induce sparsity?

Geometric explanation



$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \ \text{ s.t. } \ \|\boldsymbol{\alpha}\|_1 \leq T.$$

Important property of the Lasso

Piecewise linearity of the regularization path

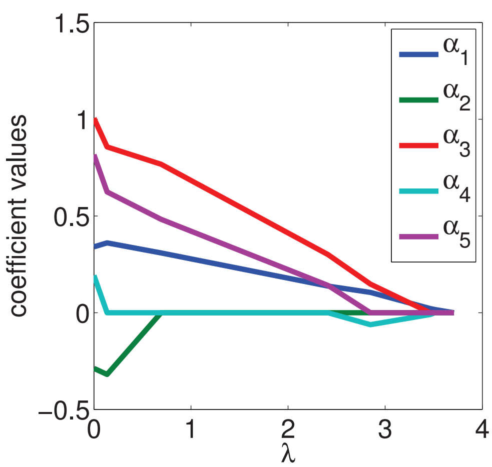


Figure: Regularization path of the Lasso

Optimization for Dictionary Learning

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{p \times n} \\ \mathbf{D} \in \mathcal{C}}} \sum_{i=1}^{n} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

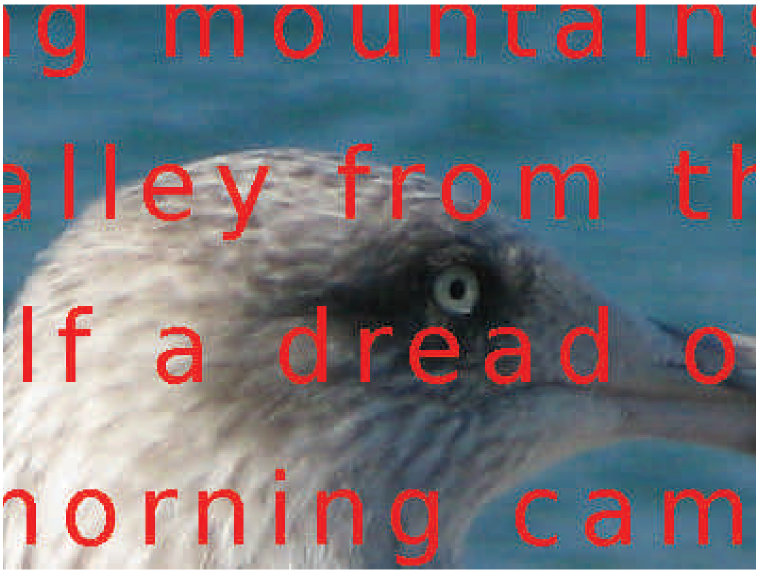
$$\mathcal{C} \triangleq \{\mathbf{D} \in \mathbb{R}^{m \times p} \;\; \text{s.t.} \;\; \forall j = 1, \ldots, p, \;\; \|\mathbf{d}_j\|_2 \leq 1\}.$$

- Classical optimization alternates between $\mathbf{D}$ and $\boldsymbol{\alpha}$.
- Good results, but slow!
- Instead use online learning [Mairal et al., 2009a]

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph



Optimization for Dictionary Learning
Inpainting a 12-Mpixel photograph

Matrix Factorization Problems and Dictionary Learning

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathcal{C}}} \sum_{i=1}^{n} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1$$

can be rewritten

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathcal{C}}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\boldsymbol{\alpha}\|_F^2 + \lambda\|\boldsymbol{\alpha}\|_1,$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n]$.

Matrix Factorization Problems and Dictionary Learning

PCA

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{p\times n}\\ \mathbf{D}\in\mathbb{R}^{m\times p}}} \frac{1}{2}\|\mathbf{Y}-\mathbf{D}\boldsymbol{\alpha}\|_F^2 \;\; \text{s.t.} \;\; \mathbf{D}^\top\mathbf{D}=\mathbf{I} \text{ and } \boldsymbol{\alpha}\boldsymbol{\alpha}^\top \text{ is diagonal.}$$

$\mathbf{D}=[\mathbf{d}_1,\ldots,\mathbf{d}_p]$ are the principal components.

Matrix Factorization Problems and Dictionary Learning

Hard clustering

$$\min_{\substack{\boldsymbol{\alpha}\in\{0,1\}^{p\times n}\\ \mathbf{D}\in\mathbb{R}^{m\times p}}} \frac{1}{2}\|\mathbf{Y}-\mathbf{D}\boldsymbol{\alpha}\|_F^2 \;\; \text{s.t.} \;\; \forall i\in\{1,\ldots,p\},\; \sum_{j=1}^{p}\boldsymbol{\alpha}_i[j]=1.$$

$\mathbf{D}=[\mathbf{d}_1,\ldots,\mathbf{d}_p]$ are the centroids of the p clusters.

Matrix Factorization Problems and Dictionary Learning

Soft clustering

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}_+^{p \times n} \\ \mathbf{D} \in \mathbb{R}^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\alpha}\|_F^2, \quad \text{s.t.} \quad \forall i \in \{1, \ldots, p\}, \quad \sum_{j=1}^{p} \boldsymbol{\alpha}_i[j] = 1.$$

$\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_p]$ are the centroids of the p clusters.

Matrix Factorization Problems and Dictionary Learning

Non-negative matrix factorization [Lee and Seung, 2001]

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}_+^{p \times n} \\ \mathbf{D} \in \mathbb{R}_+^{m \times p}}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\boldsymbol{\alpha}\|_F^2$$

Matrix Factorization Problems and Dictionary Learning

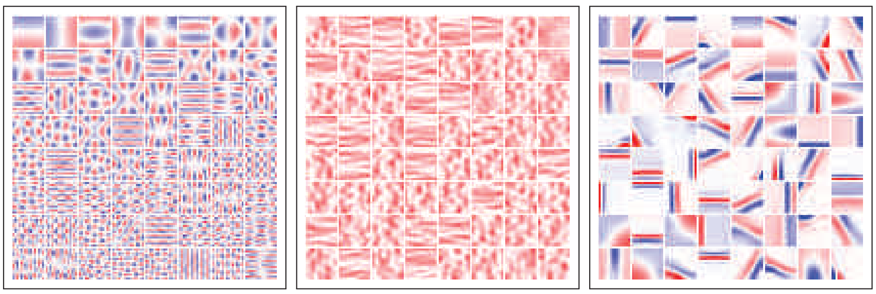
NMF+sparsity?

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}_+^{p\times n}\\ \mathbf{D}\in\mathbb{R}_+^{m\times p}}} \frac{1}{2}\|\mathbf{Y}-\mathbf{D}\boldsymbol{\alpha}\|_F^2 + \lambda\|\boldsymbol{\alpha}\|_1.$$

Most of these formulations can be addressed the same types of algorithms.

Matrix Factorization Problems and Dictionary Learning

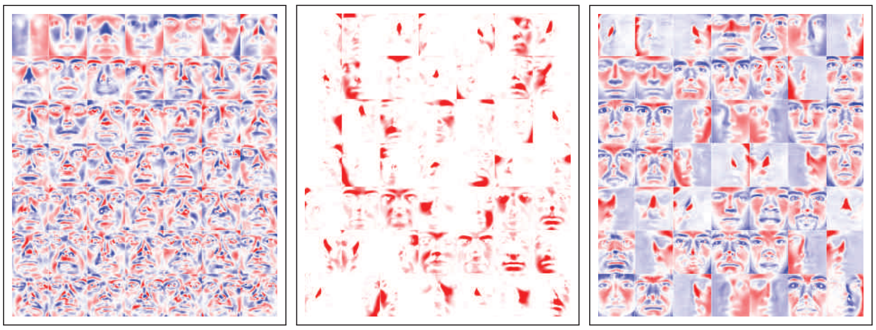
Natural Patches



(a) PCA (b) NNMF (c) DL

Matrix Factorization Problems and Dictionary Learning

Faces



(d) PCA (e) NNMF (f) DL

Sparsity-Inducing Norms (1/2)

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \overbrace{f(\boldsymbol{\alpha})}^{\text{data fitting term}} + \lambda \underbrace{\Omega(\boldsymbol{\alpha})}_{\text{sparsity-inducing norm}}$$

Standard approach to enforce sparsity in learning procedures:

- Regularizing by a **sparsity-inducing norm** ψ.
- The effect of Ω is to set some $\boldsymbol{\alpha}[j]$'s to zero, depending on the regularization parameter $\lambda \geq 0$.

The most popular choice for ψ:

- The ℓ_1 norm, $\|\boldsymbol{\alpha}\|_1 = \sum_{j=1}^p |\boldsymbol{\alpha}[j]|$.
- For the square loss, Lasso [Tibshirani, 1996].
- However, the ℓ_1 norm encodes poor information, just **cardinality**!

Sparsity-Inducing Norms (2/2)

Another popular choice for ψ:

- The ℓ_1-ℓ_2 norm,

$$\sum_{G \in \mathcal{G}} \|\boldsymbol{\alpha}_G\|_2 = \sum_{G \in \mathcal{G}} \Big(\sum_{j \in \mathcal{G}} \boldsymbol{\alpha}_j^2 \Big)^{1/2}, \text{ with } \mathcal{G} \text{ a } \textbf{partition} \text{ of } \{1, \ldots, p\}.$$

- The ℓ_1-ℓ_2 norm sets to zero **groups of non-overlapping variables** (as opposed to single variables for the ℓ_1 norm).
- For the square loss, group Lasso [Yuan and Lin, 2006].
- However, the ℓ_1-ℓ_2 norm encodes fixed/static prior information, requires to know in advance how to group the variables !

Applications:

- Selecting groups of features instead of individual variables.
- Multi-task learning, multiple kernel learning.

Non-local Sparse Image Models

Image Self-Similarities, [Buades et al., 2006, Efros and Leung, 1999, Dabov et al., 2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^{n} \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^{n} K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \tag{1}$$

Non-local Sparse Image Models

Image Self-Similarities, [Buades et al., 2006, Efros and Leung, 1999, Dabov et al., 2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^{n} \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^{n} K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \tag{1}$$

Successful application to texture synthesis: Efros and Leung [1999]
...to image denoising **(Non-Local Means)**: Buades et al. [2006]
...to image demosaicking: Buades et al. [2009]

Non-local Sparse Image Models

Image Self-Similarities, [Buades et al., 2006, Efros and Leung, 1999, Dabov et al., 2007]

Image pixels are well explained by a Nadaraya-Watson estimator:

$$\hat{\mathbf{x}}[i] = \sum_{j=1}^{n} \frac{K_h(\mathbf{y}_i - \mathbf{y}_j)}{\sum_{l=1}^{n} K_h(\mathbf{y}_i - \mathbf{y}_l)} \mathbf{y}[j], \tag{1}$$

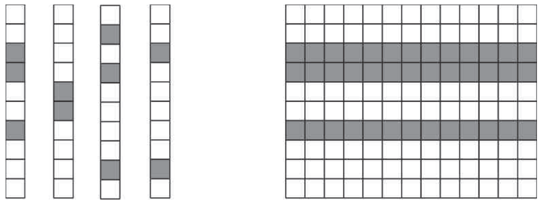
Successful application to texture synthesis: Efros and Leung [1999]
...to image denoising **(Non-Local Means)**: Buades et al. [2006]
...to image demosaicking: Buades et al. [2009]

Block-Matching with 3D filtering (BM3D) Dabov et al. [2007],
Similar patches are jointly denoised with orthogonal wavelet thresholding + several (good) heuristics: $\Longrightarrow$ state-of-the-art denoising results, less artefacts, higher PSNR.

Non-local Sparse Image Models

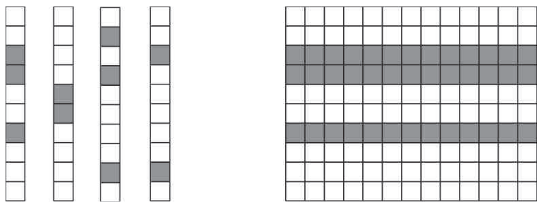
- **non-local means**: **stable** estimator. Can fail when there are no self-similarities.
- **sparse representations**: "unique" patches also admit a sparse approximation on the learned dictionary. potentially **unstable** decompositions.

Improving the stability of sparse decompositions is a current topic of research in statistics Bach [2008], Meinshausen and Buehlmann [2010].
Mairal et al. [2009b]: Similar patches should admit similar patterns:



Sparsity vs. joint sparsity

Non-local Sparse Image Models



Sparsity vs. joint sparsity

Joint sparsity is achieved through specific regularizers such as

$$\begin{aligned} \|\mathbf{A}\|_{0,\infty} &\triangleq \sum_{i=1}^{p} \|\boldsymbol{\alpha}^i\|_0, \quad \text{(not convex, not a norm)} \\ \|\mathbf{A}\|_{1,2} &\triangleq \sum_{i=1}^{p} \|\boldsymbol{\alpha}^i\|_2. \quad \text{(convex norm)} \end{aligned} \tag{2}$$

Non-local Sparse Image Models

Basic scheme for image denoising:

1. Cluster patches

$$S_i \triangleq \{j = 1, \ldots, n \ \text{ s.t. } \ \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq \xi\}, \tag{3}$$

2. Learn a dictionary with group-sparsity regularization

$$\min_{(\mathbf{A}_i)_{i=1}^n, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^{n} \frac{\|\mathbf{A}_i\|_{1,2}}{|S_i|} \ \text{ s.t. } \ \forall i \sum_{j \in S_i} \|\mathbf{y}_j - \mathbf{D}\boldsymbol{\alpha}_{ij}\|_2^2 \leq \varepsilon_i \tag{4}$$

3. Estimate the final image by averaging the representations

Non-local Sparse Image Models

Basic scheme for image denoising:

1. Cluster patches

$$S_i \triangleq \{j = 1, \ldots, n \ \text{ s.t. } \ \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq \xi\}, \tag{3}$$

2. Learn a dictionary with group-sparsity regularization

$$\min_{(\mathbf{A}_i)_{i=1}^n, \mathbf{D} \in \mathcal{C}} \sum_{i=1}^{n} \frac{\|\mathbf{A}_i\|_{1,2}}{|S_i|} \ \text{ s.t. } \ \forall i \sum_{j \in S_i} \|\mathbf{y}_j - \mathbf{D}\boldsymbol{\alpha}_{ij}\|_2^2 \leq \varepsilon_i \tag{4}$$

3. Estimate the final image by averaging the representations

Details:

- Greedy clustering (linear time) and online learning.
- Eventually use two passes.
- Use non-convex regularization for the final reconstruction.

Non-local Sparse Image Models

Denoising results, synthetic noise

Average PSNR on 10 standard images (higher is better)

σ	GSM	FOE	KSVD	BM3D	SC	LSC	LSSC
5	37.05	37.03	37.42	37.62	37.46	37.66	**37.67**
10	33.34	33.11	33.62	34.00	33.76	33.98	**34.06**
15	31.31	30.99	31.58	32.05	31.72	31.99	**32.12**
20	29.91	29.62	30.18	30.73	30.29	30.60	**30.78**
25	28.84	28.36	29.10	29.72	29.18	29.52	**29.74**
50	25.66	24.36	25.61	26.38	25.83	26.18	**26.57**
100	22.80	21.36	22.10	23.25	22.46	22.62	**23.39**

Improvement over BM3D is significant only for large values of σ. The comparison is made with GSM (Gaussian Scale Mixture) Portilla et al. [2003], FOE (Field of Experts) Roth and Black [2005], KSVD Elad and Aharon [2006] and BM3D Dabov et al. [2007].

Non-local Sparse Image Models

Denoising results, synthetic noise

Non-local Sparse Image Models

Denoising results, synthetic noise

Non-local Sparse Image Models

Demosaicking results, Kodak database

Average PSNR on the Kodak dataset (24 images)

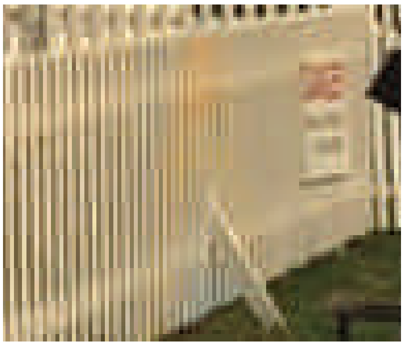
Im.	AP	DL	LPA	SC	LSC	LSSC
Av.	39.21	40.05	40.52	40.88	41.13	**41.39**

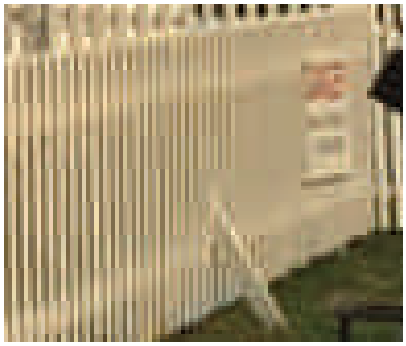
The comparison is made with AP (Alternative Projections) Gunturk et al. [2002], DL Zhang and Wu [2005] and LPA Paliy et al. [2007] (best known result on this database).

Non-local Sparse Image Models

Demosaicking results, Kodak database

More importantly than a PSNR improvement:





Regular sparsity on the left, Joint-sparsity on the right

Structured Sparsity

[Jenatton et al., 2009]

Case of general overlapping groups.

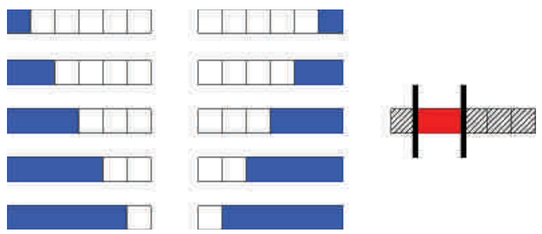
When penalizing by the ℓ_1-ℓ_2 norm,

$$\sum_{G \in \mathcal{G}} \|\boldsymbol{\alpha}_G\|_2 = \sum_{G \in \mathcal{G}} \Big(\sum_{j \in G} \boldsymbol{\alpha}_j^2 \Big)^{1/2}$$

- The ℓ_1 norm induces sparsity at the group level:
  - Some $\boldsymbol{\alpha}_G$'s are set to zero.
- Inside the groups, the ℓ_2 norm does not promote sparsity.
- Intuitively, variables belonging to the same groups are encouraged to be set to zero together.
- Optimization via reweighted least-squares, proximal methods, etc...

Examples of set of groups $\mathcal{G}$ (1/3)

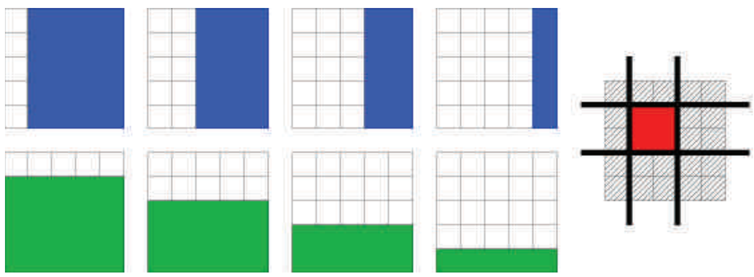
Selection of contiguous patterns on a sequence, $p = 6$.



- $\mathcal{G}$ is the set of blue groups.
- Any union of blue groups set to zero leads to the selection of a contiguous pattern.

Examples of set of groups $\mathcal{G}$ (2/3)

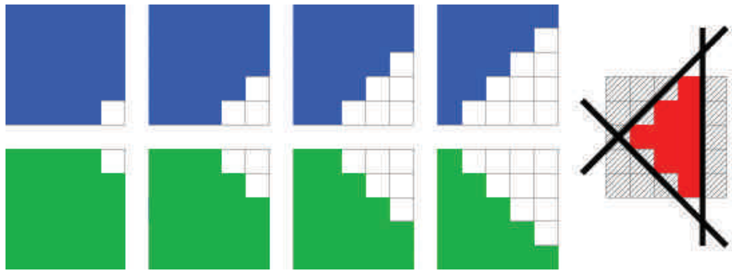
Selection of rectangles on a 2-D grids, $p = 25$.



- $\mathcal{G}$ is the set of blue/green groups (with their not displayed complements).
- Any union of blue/green groups set to zero leads to the selection of a rectangle.

Examples of set of groups $\mathcal{G}$ (3/3)

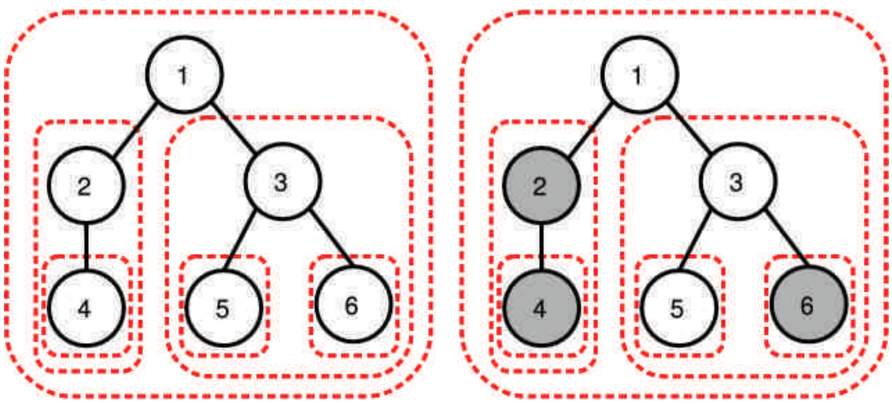
Selection of diamond-shaped patterns on a 2-D grids, $p = 25$.



- It is possible to extent such settings to 3-D space, or more complex topologies.

Hierarchical Norms

[Jenatton, Mairal, Obozinski, and Bach, 2010a]



A node can be active only if its **ancestors are active**.
The selected patterns are **rooted subtrees.**

Optimization via efficient proximal methods (same cost as ℓ_1)

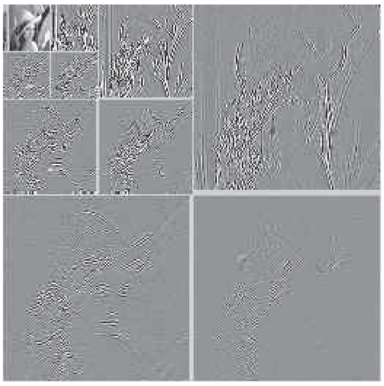
Wavelet denoising with hierarchical norms

[Jenatton, Mairal, Obozinski, and Bach, 2010b]

Classical wavelet denoising [Donoho and Johnstone, 1995]:

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1,$$

When $\mathbf{D}$ is orthogonal, the solution is obtained via **soft-thresholding**.



Wavelet denoising with hierarchical norms

[Jenatton, Mairal, Obozinski, and Bach, 2010b]

Wavelet with hierarchical norm: Add **a-priori knowledge** that the coefficients are embedded in a tree.

(g) Barb., $\sigma = 50$, ℓ_1

(h) Barb., $\sigma = 50$, tree

Wavelet denoising with hierarchical norms

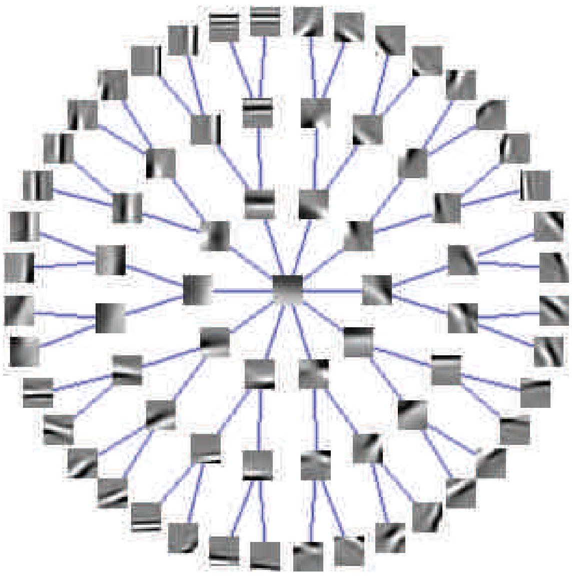
[Jenatton, Mairal, Obozinski, and Bach, 2010b]

Benchmark on a database of 12 standard images:

	σ	Haar ℓ_0	Haar ℓ_1	Haar Ω_{ℓ_2}	Haar Ω_{ℓ_∞}
PSNR	5	34.48	35.52	**35.89**	35.79
	10	29.63	30.74	**31.40**	31.23
	25	24.44	25.30	**26.41**	26.14
	50	21.53	20.42	**23.41**	23.05
	100	19.27	19.43	**20.97**	20.58
IPSNR	5	-	$1.04 \pm .31$	$\mathbf{1.41 \pm .45}$	$1.31 \pm .41$
	10	-	$1.10 \pm .22$	$\mathbf{1.76 \pm .26}$	$1.59 \pm .22$
	25	-	$.86 \pm .35$	$\mathbf{1.96 \pm .22}$	$1.69 \pm .21$
	50	-	$.46 \pm .28$	$\mathbf{1.87 \pm .20}$	$1.51 \pm .20$
	100	-	$.15 \pm .23$	$\mathbf{1.69 \pm .19}$	$1.30 \pm .19$

Hierarchical Dictionaries

[Jenatton, Mairal, Obozinski, and Bach, 2010a]

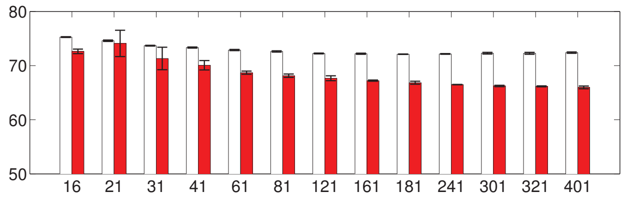


Application to patch reconstrution

[Jenatton, Mairal, Obozinski, and Bach, 2010a]

- Reconstruction of 100,000 8×8 natural images patches
  - Remove randomly subsampled pixels
  - Reconstruct with matrix factorization and structured sparsity

noise	50 %	60 %	70 %	80 %	90 %
flat	19.3 ± 0.1	26.8 ± 0.1	36.7 ± 0.1	50.6 ± 0.0	72.1 ± 0.0
tree	18.6 ± 0.1	25.7 ± 0.1	35.0 ± 0.1	48.0 ± 0.0	65.9 ± 0.3

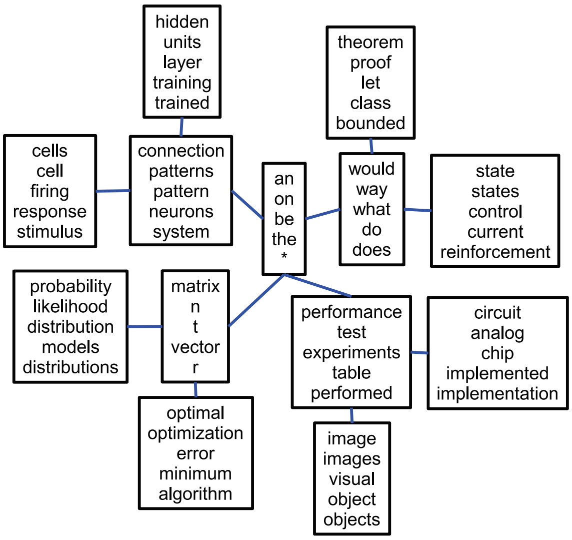


Hierarchical Topic Models for text corpora

[Jenatton, Mairal, Obozinski, and Bach, 2010a]

- Each document is modeled through word counts
- Low-rank matrix factorization of word-document matrix
- Probabilistic topic models such as Latent Dirichlet Allocation [Blei et al., 2003]
- Organise the topics in a tree.
- Previously approached using non-parametric Bayesian methods (Hierarchical Chinese Restaurant Process and nested Dirichlet Process): [Blei et al., 2010]
- **Can we achieve similar performance with simple matrix factorization formulation?**

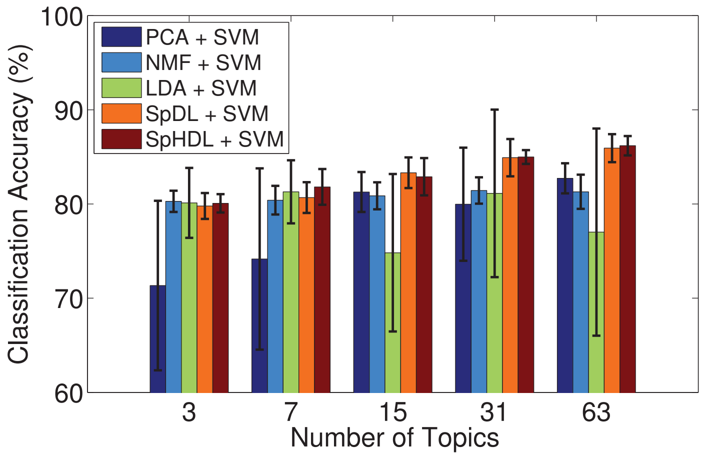
Tree of Topics



Classification based on topics

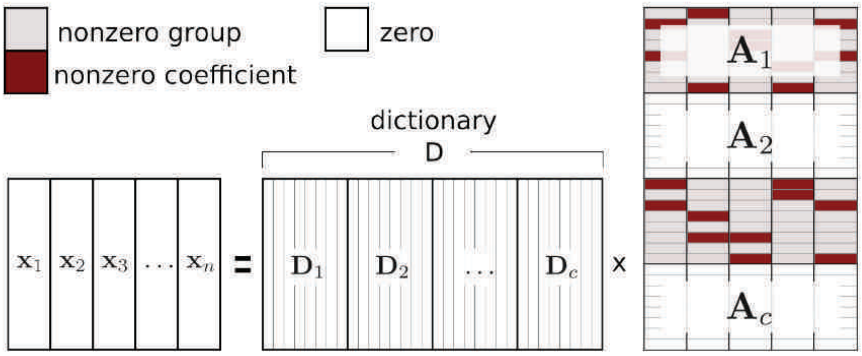
Comparison on predicting newsgroup article subjects

- 20 newsgroup articles (1425 documents, 13312 words)



Group Lasso + Sparsity

[Sprechmann et al., 2010a]



Important messages

- The ℓ_1-norm induces sparsity and shrinks the coefficients (soft-thresholding)
- The regularization path of the Lasso is piecewise linear.
- Learning the dictionary is simple, fast and scalable.
- Dictionary learning is related to several matrix factorization problems.
- Sparsity can be induced at the group level.
- Structured sparsity opens a whole range of new applications.

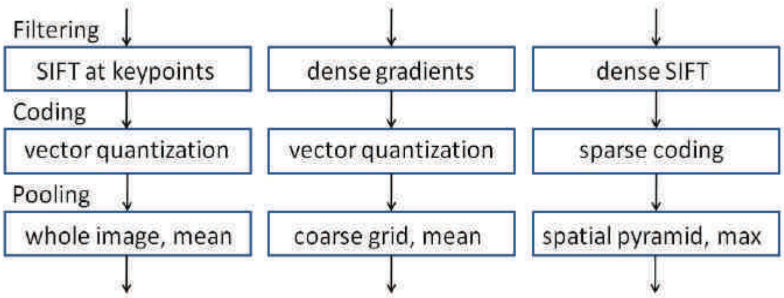
Software SPAMS is available for all of this:
`www.di.ens.fr/willow/SPAMS/`.

Next topics: Computer Vision

- Intriguing results on the use of dictionary learning for bags of words.
- Modelling the local appearance of image patches.
- Preliminary applications of structured sparsity.

1. Image Processing Applications
2. Sparse Linear Models and Dictionary Learning
3. Computer Vision Applications
   - Learning codebooks for image classification
   - Modelling the local appearance of image patches
   - Background subtraction with structured sparsity
4. Optimization for sparse methods

Learning Codebooks for Image Classification



Idea

Replacing Vector Quantization by Learned Dictionaries!

- unsupervised: [Yang et al., 2009]
- supervised: [Boureau et al., 2010, Yang et al., 2010]

Learning Codebooks for Image Classification

Let an image be represented by a set of low-level descriptors $\mathbf{y}_i$ at N locations identified with their indices $i = 1, \ldots, N$.

- hard-quantization:

$$\mathbf{y}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i \in \{0,1\}^p \text{ and } \sum_{j=1}^{p} \boldsymbol{\alpha}_i[j] = 1$$

- soft-quantization:

$$\boldsymbol{\alpha}_i[j] = \frac{e^{-\beta\|\mathbf{y}_i - \mathbf{d}_j\|_2^2}}{\sum_{k=1}^{p} e^{-\beta\|\mathbf{y}_i - \mathbf{d}_k\|_2^2}}$$

- sparse coding:

$$\mathbf{y}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

Learning Codebooks for Image Classification

Table from Boureau et al. [2010]

Method	Caltech-101, 30 training examples: Average Pool	Caltech-101, 30 training examples: Max Pool	15 Scenes, 100 training examples: Average Pool	15 Scenes, 100 training examples: Max Pool
	Results with basic features, SIFT extracted each 8 pixels			
Hard quantization, linear kernel	51.4 ± 0.9 [256]	64.3 ± 0.9 [256]	73.9 ± 0.9 [1024]	80.1 ± 0.6 [1024]
Hard quantization, intersection kernel	64.2 ± 1.0 [256] (1)	64.3 ± 0.9 [256]	80.8 ± 0.4 [256] (1)	80.1 ± 0.6 [1024]
Soft quantization, linear kernel	57.9 ± 1.5 [1024]	69.0 ± 0.8 [256]	75.6 ± 0.5 [1024]	81.4 ± 0.6 [1024]
Soft quantization, intersection kernel	66.1 ± 1.2 [512] (2)	70.6 ± 1.0 [1024]	81.2 ± 0.4 [1024] (2)	83.0 ± 0.7 [1024]
Sparse codes, linear kernel	61.3 ± 1.3 [1024]	$\mathbf{71.5 \pm 1.1}$ [1024] (3)	76.9 ± 0.6 [1024]	83.1 ± 0.6 [1024] (3)
Sparse codes, intersection kernel	70.3 ± 1.3 [1024]	$\mathbf{71.8 \pm 1.0}$ [1024] (4)	83.2 ± 0.4 [1024]	$\mathbf{84.1 \pm 0.5}$ [1024] (4)
	Results with macrofeatures and denser SIFT sampling			
Hard quantization, linear kernel	55.6 ± 1.6 [256]	70.9 ± 1.0 [1024]	74.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Hard quantization, intersection kernel	68.8 ± 1.4 [512]	70.9 ± 1.0 [1024]	81.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Soft quantization, linear kernel	61.6 ± 1.6 [1024]	71.5 ± 1.0 [1024]	76.4 ± 0.7 [1024]	81.5 ± 0.4 [1024]
Soft quantization, intersection kernel	70.1 ± 1.3 [1024]	73.2 ± 1.0 [1024]	81.8 ± 0.4 [1024]	83.0 ± 0.4 [1024]
Sparse codes, linear kernel	65.7 ± 1.4 [1024]	$\mathbf{75.1 \pm 0.9}$ [1024]	78.2 ± 0.7 [1024]	83.6 ± 0.4 [1024]
Sparse codes, intersection kernel	73.7 ± 1.3 [1024]	$\mathbf{75.7 \pm 1.1}$ [1024]	83.5 ± 0.4 [1024]	$\mathbf{84.3 \pm 0.5}$ [1024]

	Unsup	Discr
Linear	83.6 ± 0.4	$\mathbf{84.9 \pm 0.3}$
Intersect	84.3 ± 0.5	$\mathbf{84.7 \pm 0.4}$

Yang et al. [2009] have won the PASCAL VOC'09 challenge using this kind of techniques.

Learning dictionaries with a discriminative cost function

Idea:

Let us consider 2 sets S_-, S_+ of signals representing 2 different classes. Each set should admit a dictionary best adapted to its reconstruction.

Classification procedure for a signal $\mathbf{y} \in \mathbb{R}^n$:

$$\min(\mathbf{R}^\star(\mathbf{y}, \mathbf{D}_-), \mathbf{R}^\star(\mathbf{y}, \mathbf{D}_+))$$

where

$$\mathbf{R}^\star(\mathbf{y}, \mathbf{D}) = \min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_0 \leq L.$$

"Reconstructive" training

$$\begin{cases} \min_{\mathbf{D}_-} \sum_{i \in S_-} \mathbf{R}^\star(\mathbf{y}_i, \mathbf{D}_-) \\ \min_{\mathbf{D}_+} \sum_{i \in S_+} \mathbf{R}^\star(\mathbf{y}_i, \mathbf{D}_+) \end{cases}$$

[Grosse et al., 2007], [Huang and Aviyente, 2006],
[Sprechmann et al., 2010b] for unsupervised clustering (CVPR '10)

Learning dictionaries with a discriminative cost function

"Discriminative" training

[Mairal, Bach, Ponce, Sapiro, and Zisserman, 2008a]

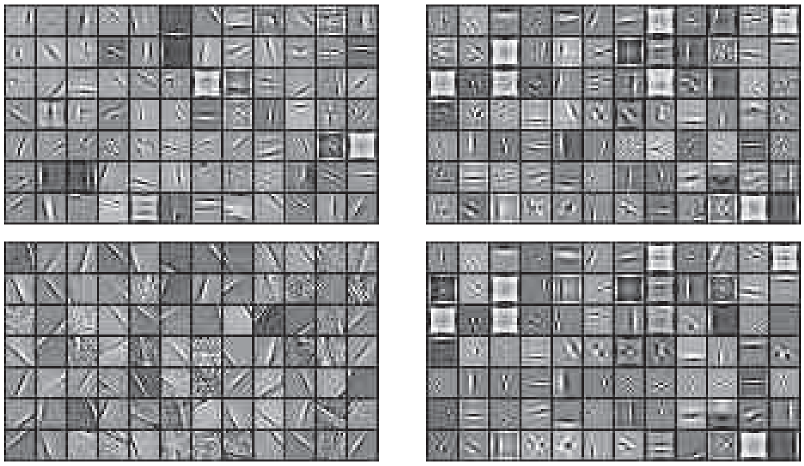
$$\min_{\mathbf{D}_-,\mathbf{D}_+} \sum_i \mathcal{C}\Big(\lambda z_i \big(\mathbf{R}^\star(\mathbf{y}_i, \mathbf{D}_-) - \mathbf{R}^\star(\mathbf{y}_i, \mathbf{D}_+)\big)\Big),$$

where $z_i \in \{-1, +1\}$ is the label of $\mathbf{y}_i$.

Logistic regression function

Learning dictionaries with a discriminative cost function

Examples of dictionaries

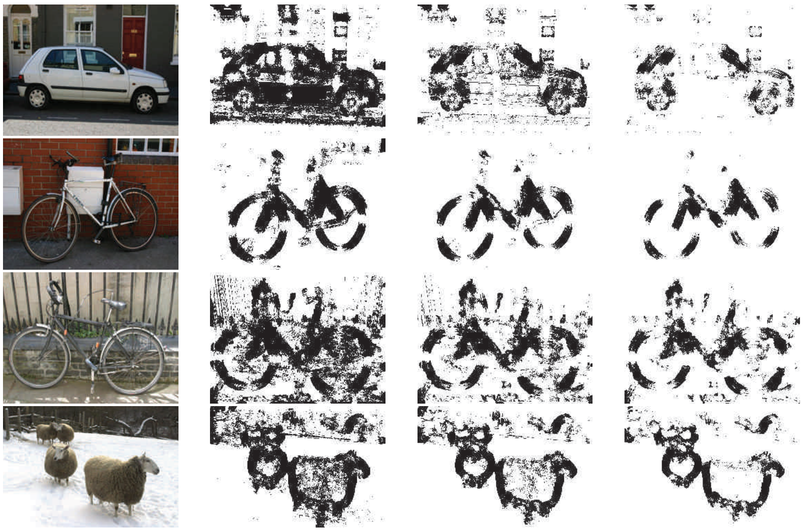


Top: reconstructive, Bottom: discriminative, Left: Bicycle, Right: Background.

Learning dictionaries with a discriminative cost function
Texture segmentation

Learning dictionaries with a discriminative cost function
Texture segmentation

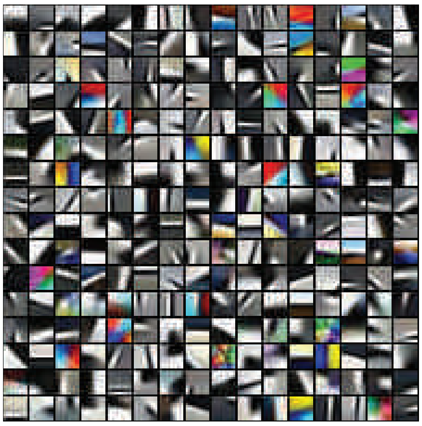
Learning dictionaries with a discriminative cost function
Pixelwise classification



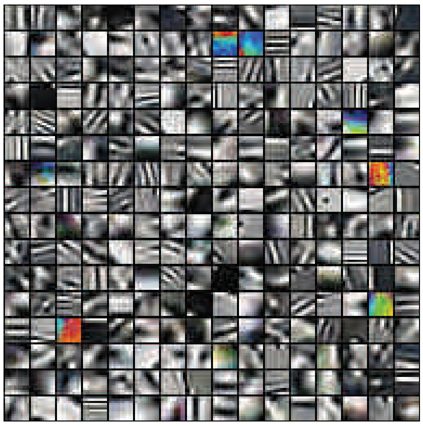
Learning dictionaries with a discriminative cost function
weakly-supervised pixel classification

Application to edge detection and classification

[Mairal, Leordeanu, Bach, Hebert, and Ponce, 2008c]



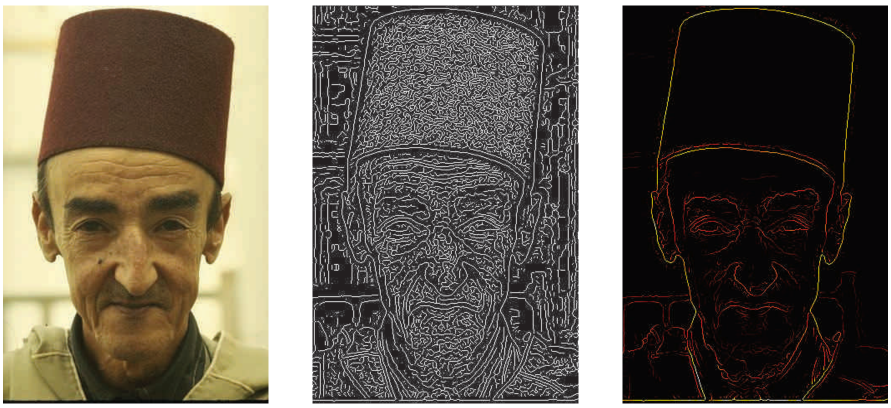
Good edges



Bad edges

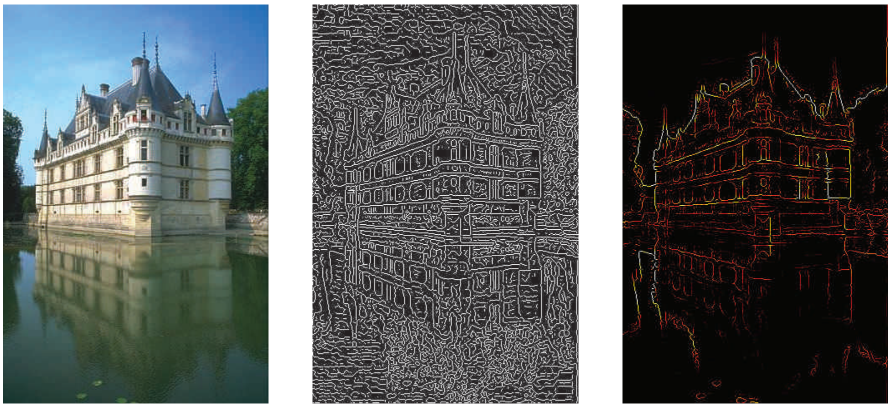
Application to edge detection and classification

Berkeley segmentation benchmark



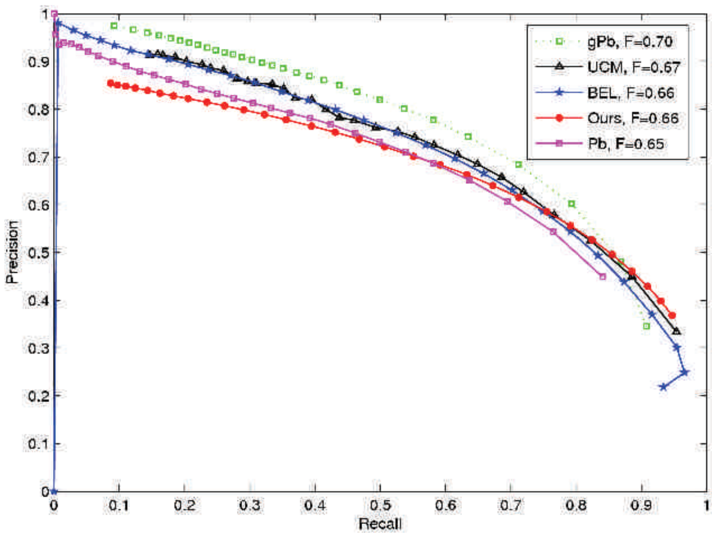
Raw edge detection on the right

Application to edge detection and classification
Berkeley segmentation benchmark



Raw edge detection on the right

Application to edge detection and classification
Berkeley segmentation benchmark

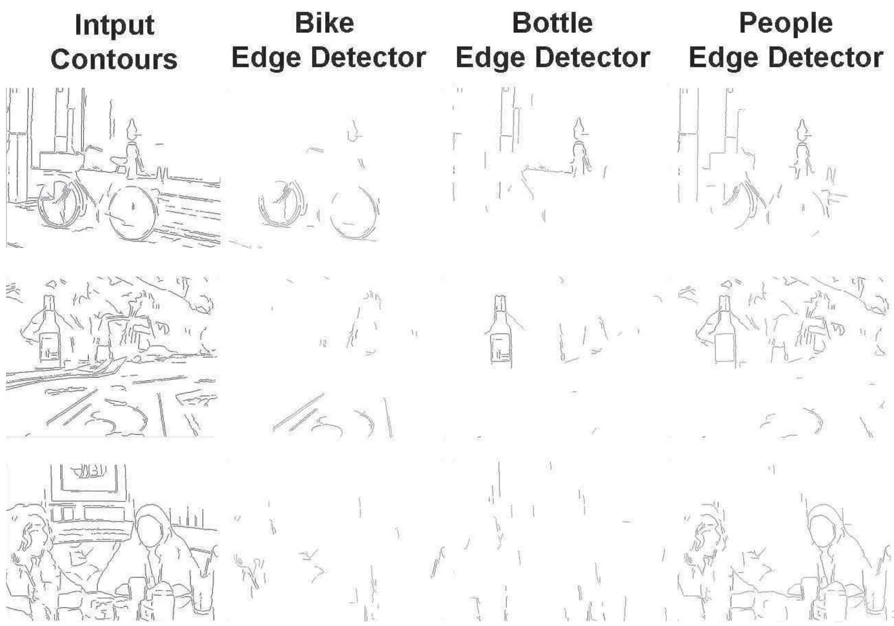


Application to edge detection and classification

Contour-based classifier: [Leordeanu, Hebert, and Sukthankar, 2007]

Is there a bike, a motorbike, a car or a person on this image?

Application to edge detection and classification



Application to edge detection and classification

Performance gain due to the prefiltering

Ours + [Leordeanu '07]	[Leordeanu '07]	[Winn '05]
96.8%	89.4%	76.9%

Recognition rates for the same experiment as [Winn et al., 2005] on VOC 2005.

Category	Ours+[Leordeanu '07]	[Leordeanu '07]
Aeroplane	71.9%	61.9%
Boat	67.1%	56.4%
Cat	82.6%	53.4%
Cow	68.7%	59.2%
Horse	76.0%	67%
Motorbike	80.6%	73.6%
Sheep	72.9%	58.4%
Tvmonitor	87.7%	83.8%
Average	75.9%	64.2 %

Recognition performance at equal error rate for 8 classes on a subset of images from Pascal 07.

Digital Art Authentification

Data Courtesy of Hugues, Graham, and Rockmore [2009]

Authentic

Fake

Given a pair of paintings, Which one is the fake?

Digital Art Authentification

Data Courtesy of Hugues, Graham, and Rockmore [2009]

Authentic

Fake

Fake

Digital Art Authentification

Data Courtesy of Hugues, Graham, and Rockmore [2009]

Authentic

Fake

Authentic

Background Subtraction

Given a video sequence, how can we remove foreground objects?

video sequence 1 video sequence 2

Background Subtraction

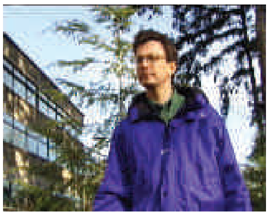
$$\underbrace{\mathbf{y}}_{\text{frame}} \approx \underbrace{\mathbf{D}\boldsymbol{\alpha}}_{\text{linear combination of background frames}} + \underbrace{\mathbf{e}}_{\text{error term}}.$$

Solved by

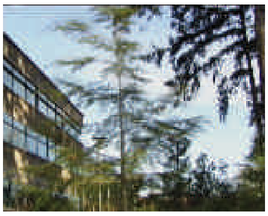
$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p,\mathbf{e}\in\mathbb{R}^m} \frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}-\mathbf{e}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\| + \lambda_2\psi(\mathbf{e}).$$

Same idea used by Wright et al. '09 for robust face recognition with $\psi = \ell_1$.

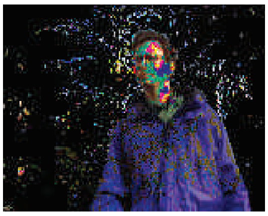
Background Subtraction



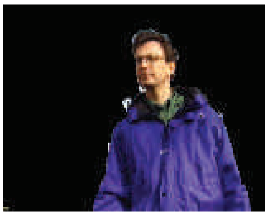
(a) input



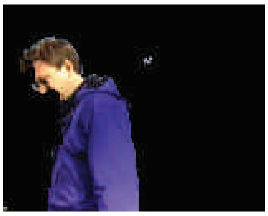
(b) estimated background



(c) foreground, ℓ_1

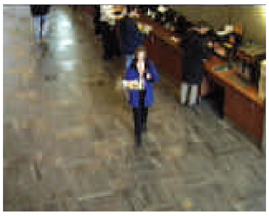


(d) foreground, ℓ_1+struct

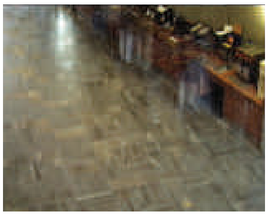


(e) other example

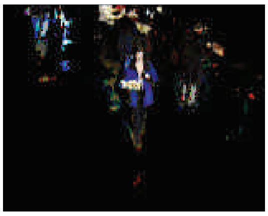
Background Subtraction



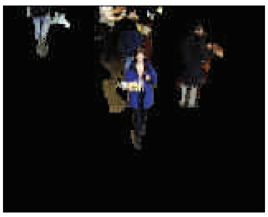
(a) input



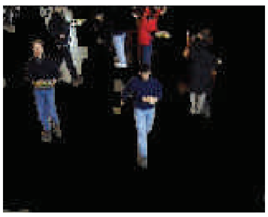
(b) estimated background



(c) foreground, ℓ_1



(d) foreground, ℓ_1+struct



(e) other example

Important messages

- Learned dictionaries are well adapted to model the local appearance of images and edges.
- They can be used to learn dictionaries of SIFT features.
- New applications coming with structured sparsity?

Next topics

- Optimization for solving sparse decomposition problems
- Optimization for dictionary learning

1. Image Processing Applications
2. Sparse Linear Models and Dictionary Learning
3. Computer Vision Applications
4. Optimization for sparse methods
   - Greedy algorithms
   - ℓ_1 optimization
   - online dictionary learning

Recall: The Sparse Decomposition Problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\boldsymbol{\alpha})}_{\text{sparsity-inducing regularization}}$$

ψ induces sparsity in $\boldsymbol{\alpha}$. It can be

- the ℓ_0 "pseudo-norm". $\|\boldsymbol{\alpha}\|_0 \triangleq \#\{i \text{ s.t. } \boldsymbol{\alpha}[i] \neq 0\}$ (NP-hard)
- the ℓ_1 norm. $\|\boldsymbol{\alpha}\|_1 \triangleq \sum_{i=1}^p |\boldsymbol{\alpha}[i]|$ (convex)
- ...

This is a selection problem.

Finding your way in the sparse coding literature...

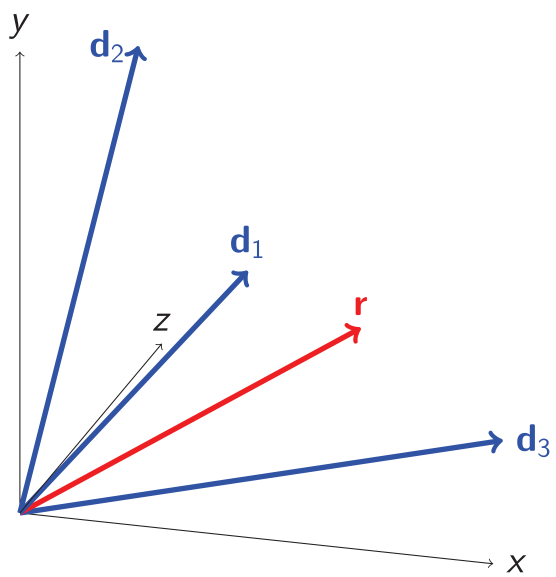
...is not easy. The literature is vast, redundant, sometimes confusing and many papers are claiming victory...

The main class of methods are

- greedy procedures [Mallat and Zhang, 1993], [Weisberg, 1980]
- homotopy [Osborne et al., 2000], [Efron et al., 2004], [Markowitz, 1956]
- soft-thresholding based methods [Fu, 1998], [Daubechies et al., 2004], [Friedman et al., 2007], [Nesterov, 2007], [Beck and Teboulle, 2009], ...
- reweighted-ℓ_2 methods [Daubechies et al., 2009],...
- active-set methods [Roth and Fischer, 2008].
- ...

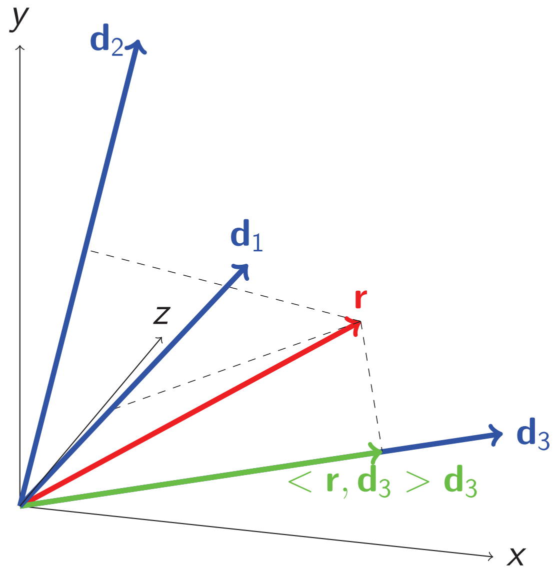
Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



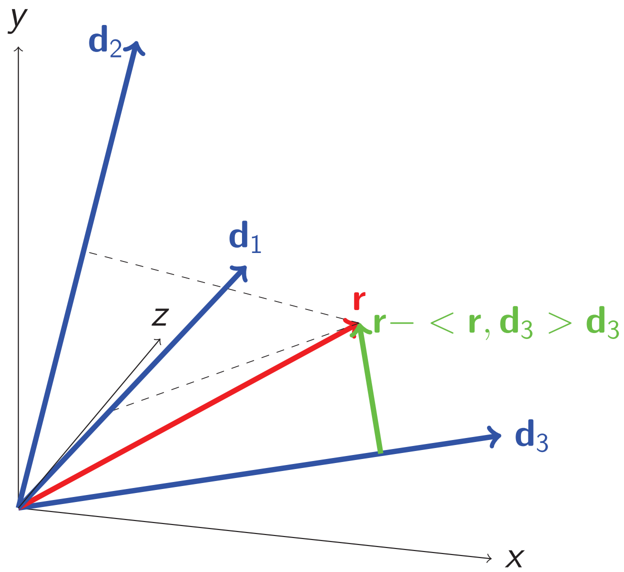
Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



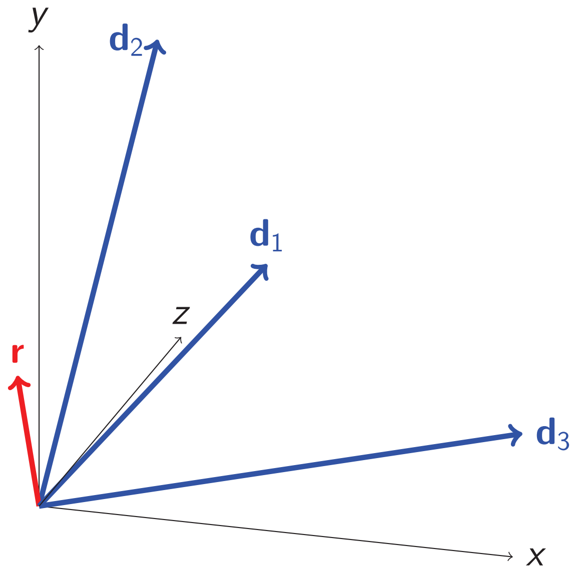
Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



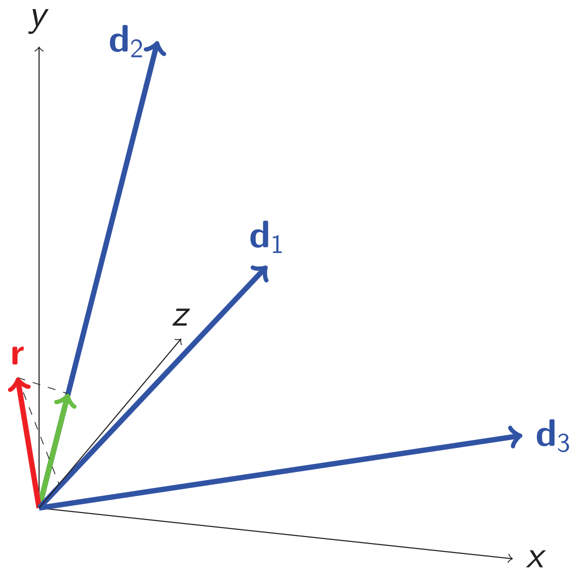
Matching Pursuit

$\alpha = (0, 0, 0.75)$



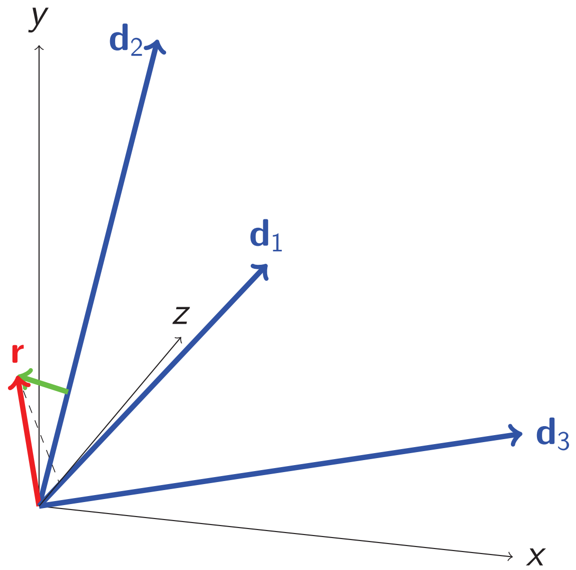
Matching Pursuit

$\alpha = (0, 0, 0.75)$



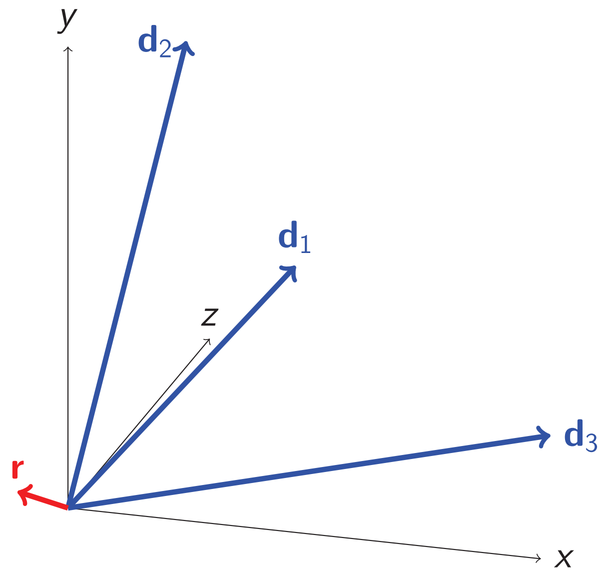
Matching Pursuit

$\boldsymbol{\alpha} = (0, 0, 0.75)$



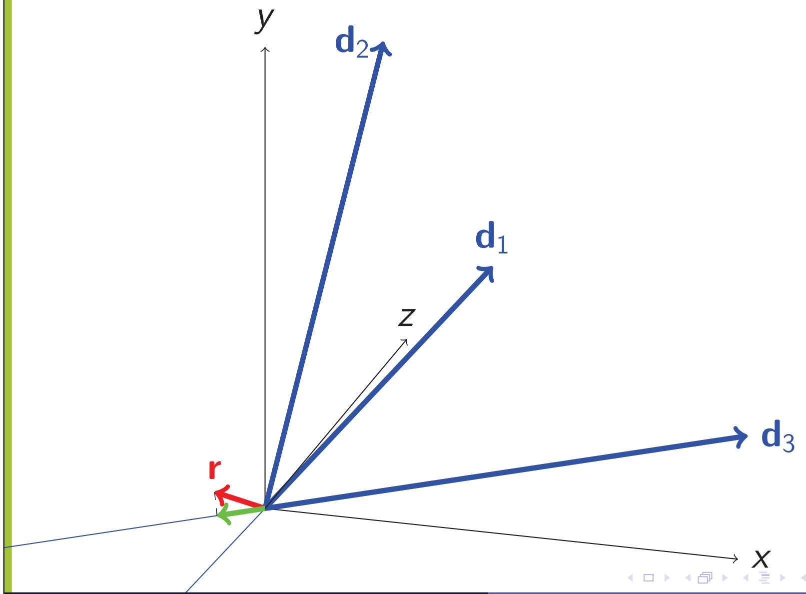
Matching Pursuit

$\boldsymbol{\alpha} = (0, 0.24, 0.75)$



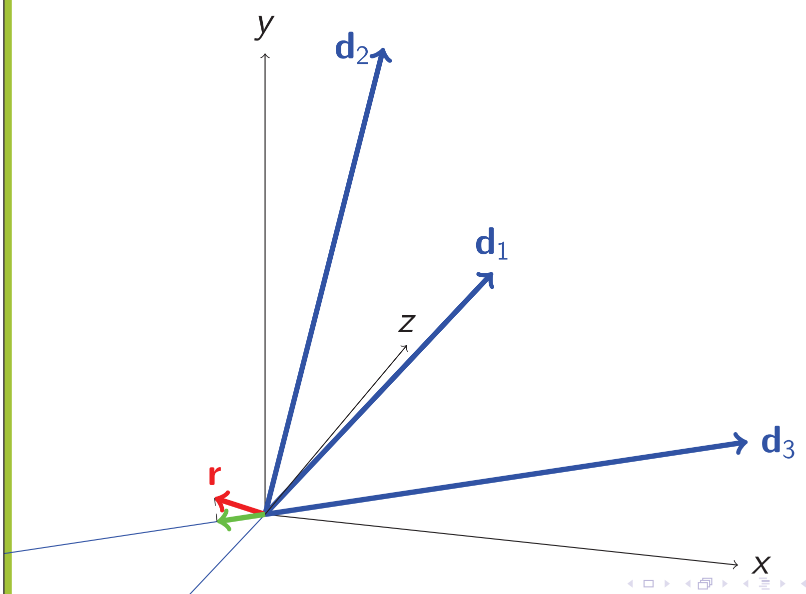
Matching Pursuit

$$\alpha = (0, 0.24, 0.75)$$



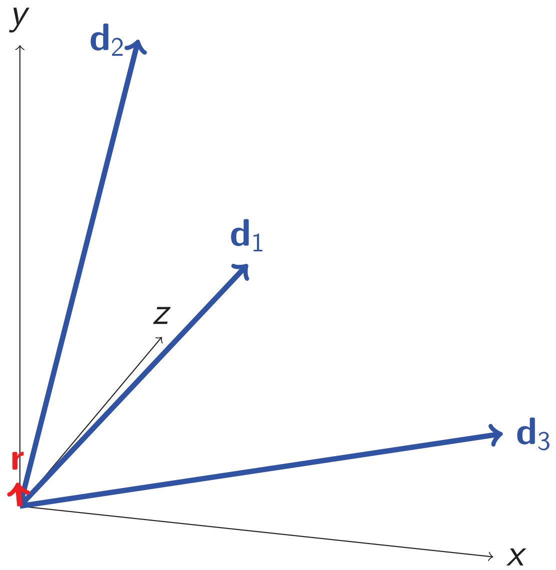
Matching Pursuit

$$\alpha = (0, 0.24, 0.75)$$



Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0.24, 0.65)$$



Matching Pursuit

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \|\underbrace{\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}}_{\mathbf{r}}\|_2^2 \text{ s.t. } \|\boldsymbol{\alpha}\|_0 \leq L$$

1: $\boldsymbol{\alpha} \leftarrow 0$
2: $\mathbf{r} \leftarrow \mathbf{y}$ (residual).
3: **while** $\|\boldsymbol{\alpha}\|_0 < L$ **do**
4: Select the atom with maximum correlation with the residual

$$\hat{\imath} \leftarrow \arg\max_{i=1,\ldots,p} |\mathbf{d}_i^T \mathbf{r}|$$

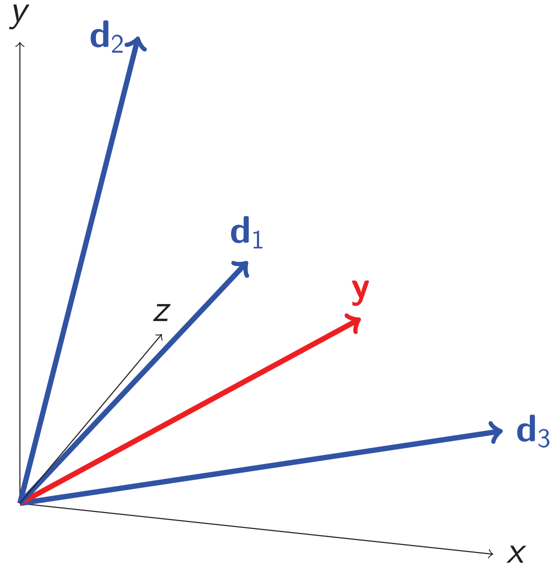
5: Update the residual and the coefficients

$$\begin{aligned} \boldsymbol{\alpha}[\hat{\imath}] &\leftarrow \boldsymbol{\alpha}[\hat{\imath}] + \mathbf{d}_{\hat{\imath}}^T \mathbf{r} \\ \mathbf{r} &\leftarrow \mathbf{r} - (\mathbf{d}_{\hat{\imath}}^T \mathbf{r}) \mathbf{d}_{\hat{\imath}} \end{aligned}$$

6: **end while**

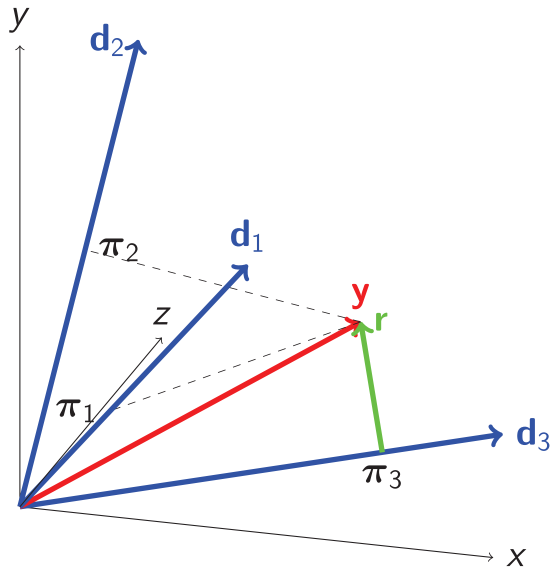
Orthogonal Matching Pursuit

$$\alpha = (0, 0, 0)$$
$$\Gamma = \emptyset$$



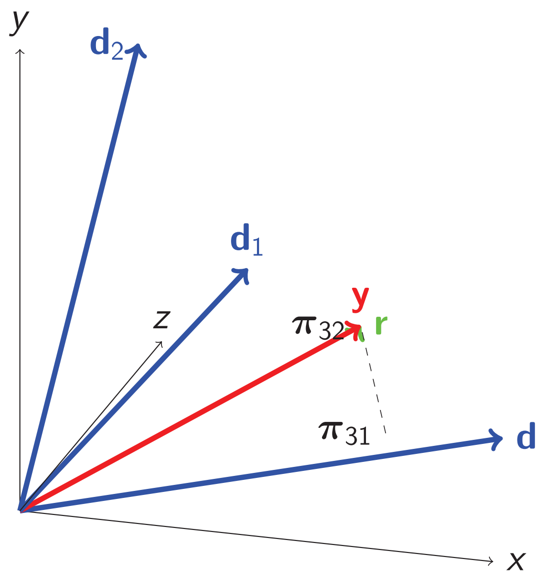
Orthogonal Matching Pursuit

$$\alpha = (0, 0, 0.75)$$
$$\Gamma = \{3\}$$



Orthogonal Matching Pursuit

$$\alpha = (0, 0.29, 0.63)$$
$$\Gamma = \{3, 2\}$$



Orthogonal Matching Pursuit

$$\min_{\alpha \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 \text{ s.t. } \|\alpha\|_0 \leq L$$

1: $\Gamma = \emptyset$.
2: **for** *iter* $= 1, \ldots, L$ **do**
3: Select the atom which most reduces the objective

$$\hat{\imath} \leftarrow \arg\min_{i \in \Gamma^C} \left\{ \min_{\alpha'} \|\mathbf{y} - \mathbf{D}_{\Gamma \cup \{i\}} \alpha'\|_2^2 \right\}$$

4: Update the active set: $\Gamma \leftarrow \Gamma \cup \{\hat{\imath}\}$.
5: Update the residual (orthogonal projection)

$$\mathbf{r} \leftarrow (\mathbf{I} - \mathbf{D}_\Gamma(\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma)^{-1} \mathbf{D}_\Gamma^T)\mathbf{y}.$$

6: Update the coefficients

$$\alpha_\Gamma \leftarrow (\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma)^{-1} \mathbf{D}_\Gamma^T \mathbf{y}.$$

7: **end for**

Orthogonal Matching Pursuit

Contrary to MP, an atom can only be selected one time with OMP. It is, however, more difficult to implement efficiently. The keys for a good implementation in the case of a large number of signals are

- Precompute the Gram matrix $\mathbf{G} = \mathbf{D}^T\mathbf{D}$ once in for all,
- Maintain the computation of $\mathbf{D}^T\mathbf{r}$ for each signal,
- Maintain a Cholesky decomposition of $(\mathbf{D}_\Gamma^T\mathbf{D}_\Gamma)^{-1}$ for each signal.

The total complexity for decomposing n L-sparse signals of size m with a dictionary of size p is

$$\underbrace{O(p^2m)}_{\text{Gram matrix}} + \underbrace{O(nL^3)}_{\text{Cholesky}} + \underbrace{O(n(pm + pL^2))}_{\mathbf{D}^T\mathbf{r}} = O(np(m + L^2))$$

It is also possible to use the matrix inversion lemma instead of a Cholesky decomposition (same complexity, but less numerical stability)

Example with the software SPAMS

Software available at http://www.di.ens.fr/willow/SPAMS/

```
>> I=double(imread('data/lena.eps'))/255;
>> %extract all patches of I
>> X=im2col(I,[8 8],'sliding');
>> %load a dictionary of size 64 x 256
>> D=load('dict.mat');
>>
>> %set the sparsity parameter L to 10
>> param.L=10;
>> alpha=mexOMP(X,D,param);
```

On a 8-cores 2.83Ghz machine: **230000 signals processed per second!**

Optimality conditions of the Lasso

Nonsmooth optimization

Directional derivatives and subgradients are useful tools for studying ℓ_1-decomposition problems:

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

In this tutorial, we use the **directional derivatives** to derive simple optimality conditions of the Lasso.

For more information on convex analysis and nonsmooth optimization, see the following books: [Boyd and Vandenberghe, 2004], [Nocedal and Wright, 2006], [Borwein and Lewis, 2006], [Bonnans et al., 2006], [Bertsekas, 1999].

Optimality conditions of the Lasso

Directional derivatives

- **Directional derivative** in the direction $\mathbf{u}$ at $\boldsymbol{\alpha}$:

$$\nabla f(\boldsymbol{\alpha}, \mathbf{u}) = \lim_{t \to 0^+} \frac{f(\boldsymbol{\alpha} + t\mathbf{u}) - f(\boldsymbol{\alpha})}{t}$$

- Main idea: in non smooth situations, one may need to look at all directions $\mathbf{u}$ and not simply p independent ones!
- **Proposition 1:** if f is differentiable in $\boldsymbol{\alpha}$, $\nabla f(\boldsymbol{\alpha}, \mathbf{u}) = \nabla f(\boldsymbol{\alpha})^T \mathbf{u}$.
- **Proposition 2:** $\boldsymbol{\alpha}$ is optimal iff for all $\mathbf{u}$ in $\mathbb{R}^p$, $\nabla f(\boldsymbol{\alpha}, \mathbf{u}) \geq 0$.

Optimality conditions of the Lasso

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

$\boldsymbol{\alpha}^\star$ is optimal iff for all $\mathbf{u}$ in $\mathbb{R}^p$, $\nabla f(\boldsymbol{\alpha},\mathbf{u}) \geq 0$—that is,

$$-\mathbf{u}^T\mathbf{D}^T(\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}^\star) + \lambda \sum_{i,\boldsymbol{\alpha}^\star[i]\neq 0} \text{sign}(\boldsymbol{\alpha}^\star[i])\mathbf{u}[i] + \lambda \sum_{i,\boldsymbol{\alpha}^\star[i]=0} |\mathbf{u}_i| \geq 0,$$

which is equivalent to the following conditions:

$$\forall i = 1,\ldots,p, \quad \left\{ \begin{array}{lcll} |\mathbf{d}_i^T(\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}^\star)| & \leq & \lambda & \text{if } \boldsymbol{\alpha}^\star[i]=0 \\ \mathbf{d}_i^T(\mathbf{y}-\mathbf{D}\boldsymbol{\alpha}^\star) & = & \lambda\,\text{sign}(\boldsymbol{\alpha}^\star[i]) & \text{if } \boldsymbol{\alpha}^\star[i]\neq 0 \end{array} \right.$$

Homotopy

- A homotopy method provides a set of solutions indexed by a parameter.
- The regularization path $(\lambda, \boldsymbol{\alpha}^\star(\lambda))$ for instance!!
- It can be useful when the path has some “nice” properties (piecewise linear, piecewise quadratic).
- LARS [Efron et al., 2004] starts from a trivial solution, and follows the regularization path of the Lasso, which is is **piecewise linear**.

Homotopy, LARS

[Osborne et al., 2000], [Efron et al., 2004]

$$\forall i = 1, \ldots, p, \quad \begin{cases} |\mathbf{d}_i^T(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}^\star)| & \leq & \lambda & \text{if } \boldsymbol{\alpha}^\star[i] = 0 \\ \mathbf{d}_i^T(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}^\star) & = & \lambda \,\text{sign}(\boldsymbol{\alpha}^\star[i]) & \text{if } \boldsymbol{\alpha}^\star[i] \neq 0 \end{cases} \tag{5}$$

The regularization path is piecewise linear:

$$\mathbf{D}_\Gamma^T(\mathbf{y} - \mathbf{D}_\Gamma \boldsymbol{\alpha}_\Gamma^\star) = \lambda \,\text{sign}(\boldsymbol{\alpha}_\Gamma^\star)$$
$$\boldsymbol{\alpha}_\Gamma^\star(\lambda) = (\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma)^{-1}(\mathbf{D}_\Gamma^T \mathbf{y} - \lambda \,\text{sign}(\boldsymbol{\alpha}_\Gamma^\star)) = \mathbf{A} + \lambda \mathbf{B}$$

A simple interpretation of LARS

- Start from the trivial solution ($\lambda = \|\mathbf{D}^T\mathbf{y}\|_\infty, \boldsymbol{\alpha}^\star(\lambda) = 0$).
- Maintain the computations of $|\mathbf{d}_i^T(\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}^\star(\lambda))|$ for all i.
- Maintain the computation of the current direction $\mathbf{B}$.
- Follow the path by reducing λ until the next kink.

Example with the software SPAMS

http://www.di.ens.fr/willow/SPAMS/

```
>> I=double(imread('data/lena.eps'))/255;
>> %extract all patches of I
>> X=normalize(im2col(I,[8 8],'sliding'));
>> %load a dictionary of size 64 x 256
>> D=load('dict.mat');
>>
>> %set the sparsity parameter lambda to 0.15
>> param.lambda=0.15;
>> alpha=mexLasso(X,D,param);
```

On a 8-cores 2.83Ghz machine: **77000 signals processed per second!**
Note that it can also solve **constrained** version of the problem. The complexity is more or less the same as OMP and uses the same tricks (Cholesky decomposition).

Coordinate Descent

- Coordinate descent + nonsmooth objective: **WARNING: not convergent in general**
- Here, the problem is equivalent to a convex smooth optimization problem with **separable** constraints

$$\min_{\boldsymbol{\alpha}_+,\boldsymbol{\alpha}_-} \frac{1}{2}\|\mathbf{y}-\mathbf{D}_+\boldsymbol{\alpha}_+ + \mathbf{D}_-\boldsymbol{\alpha}_-\|_2^2 + \lambda\boldsymbol{\alpha}_+^T\mathbf{1} + \lambda\boldsymbol{\alpha}_-^T\mathbf{1} \;\text{ s.t. }\; \boldsymbol{\alpha}_-,\boldsymbol{\alpha}_+ \geq 0.$$

- For this **specific** problem, coordinate descent is **convergent**.
- Supposing $\|\mathbf{d}_i\|_2 = 1$, updating the coordinate i:

$$\boldsymbol{\alpha}[i] \leftarrow \arg\min_{\beta} \frac{1}{2}\|\underbrace{\mathbf{y} - \sum_{j\neq i}\boldsymbol{\alpha}[j]\mathbf{d}_j}_{\mathbf{r}} - \beta\mathbf{d}_i\|_2^2 + \lambda|\beta|$$

$$\leftarrow \mathrm{sign}(\mathbf{d}_i^T\mathbf{r})(|\mathbf{d}_i^T\mathbf{r}| - \lambda)^+$$

- $\Rightarrow$ **soft-thresholding**!

Example with the software SPAMS

http://www.di.ens.fr/willow/SPAMS/

```
>> I=double(imread('data/lena.eps'))/255;
>> %extract all patches of I
>> X=normalize(im2col(I,[8 8],'sliding'));
>> %load a dictionary of size 64 x 256
>> D=load('dict.mat');
>>
>> %set the sparsity parameter lambda to 0.15
>> param.lambda=0.15;
>> param.tol=1e-2;
>> param.itermax=200;
>> alpha=mexCD(X,D,param);
```

On a 8-cores 2.83Ghz machine: **93000 signals processed per second!**

First-order/proximal methods

$$\min_{\alpha \in \mathbb{R}^p} f(\alpha) + \lambda\Omega(\alpha)$$

- f is strictly convex and differentiable with a Lipshitz gradient.
- Generalizes the idea of gradient descent

$$\alpha^{k+1} \leftarrow \arg\min_{\alpha \in \mathbb{R}^p} \underbrace{f(\alpha^k) + \nabla f(\alpha^k)^\top(\alpha - \alpha^k)}_{\text{linear approximation}} + \underbrace{\frac{L}{2}\|\alpha - \alpha^k\|_2^2}_{\text{quadratic term}} + \lambda\Omega(\alpha)$$

$$\leftarrow \arg\min_{\alpha \in \mathbb{R}^p} \frac{1}{2}\|\alpha - (\alpha^k - \frac{1}{L}\nabla f(\alpha^k))\|_2^2 + \frac{\lambda}{L}\Omega(\alpha)$$

When $\lambda = 0$, $\alpha^{k+1} \leftarrow \alpha^k - \frac{1}{L}\nabla f(\alpha^k)$, this is equivalent to a classical gradient descent step.

First-order/proximal methods

- They require solving efficiently the proximal operator

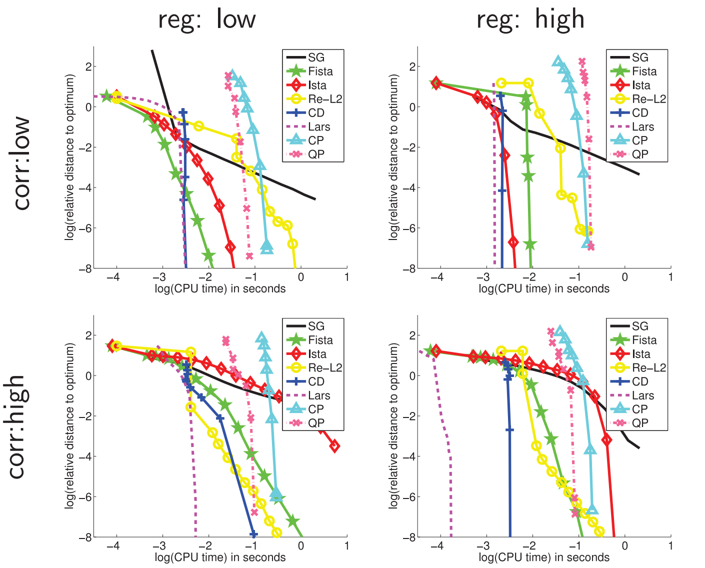
$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{u} - \alpha\|_2^2 + \lambda\Omega(\alpha)$$

- For the ℓ_1-norm, this amounts to a soft-thresholding:

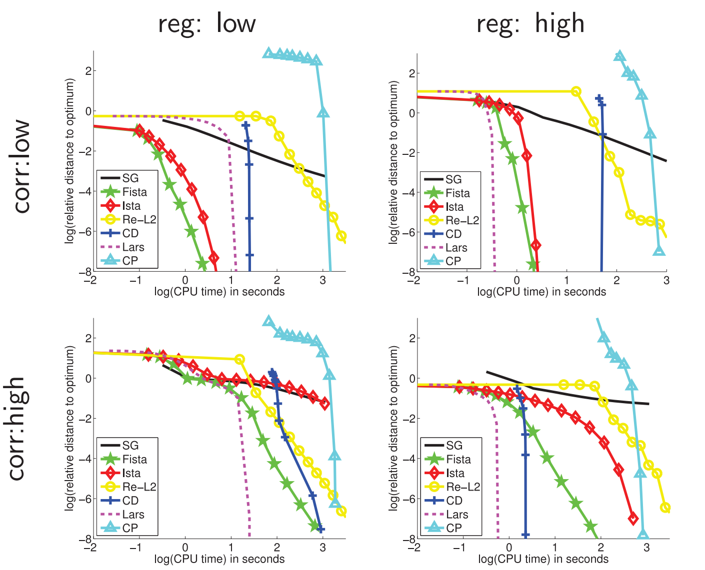
$$\alpha_i^\star = \text{sign}(\mathbf{u}_i)(\mathbf{u}_i - \lambda)^+.$$

- There exists accelerated versions based on Nesterov optimal first-order method (gradient method with "extrapolation") [Beck and Teboulle, 2009, Nesterov, 2007, 1983]
- suited for large-scale experiments.

Lasso Empirical comparison: Lasso, small scale ($n = 200, p = 200$)



Empirical comparison: Lasso, medium scale ($n = 2000, p = 10000$)



Empirical comparison: conclusions

Lasso

- Generic methods very slow
- LARS fastest in **low dimension** or for **high correlation**
- Proximal methods competitive
  - esp. larger setting with weak corr. + weak reg.
- Coordinate descent
  - Dominated by the LARS
  - Would benefit from an offline computation of the matrix

Smooth Losses

- LARS not available $\rightarrow$ CD and proximal methods good candidates

Optimization for Grouped Sparsity

The formulation:

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \underbrace{\frac{1}{2}\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2}_{\text{data fitting term}} + \underbrace{\lambda \sum_{g \in \mathcal{G}} \|\boldsymbol{\alpha}_g\|_q}_{\textbf{group-sparsity-inducing regularization}}$$

The main class of algorithms for solving grouped-sparsity problems are

- Greedy approaches
- Block-coordinate descent
- Proximal methods

Optimization for Grouped Sparsity

The proximal operator:

$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{u} - \alpha\|_2^2 + \lambda \sum_{g \in \mathcal{G}} \|\alpha_g\|_q$$

For $q = 2$,

$$\alpha_g^\star = \frac{\mathbf{u}_g}{\|\mathbf{u}_g\|_2}(\|\mathbf{u}_g\|_2 - \lambda)^+, \quad \forall g \in \mathcal{G}$$

For $q = \infty$,

$$\alpha_g^\star = \mathbf{u}_g - \Pi_{\|.\|_1 \leq \lambda}[\mathbf{u}_g], \quad \forall g \in \mathcal{G}$$

These formula generalize soft-thrsholding to groups of variables. They are used in block-coordinate descent and proximal algorithms.

Reweighted ℓ_2

Let us start from something simple

$$a^2 - 2ab + b^2 \geq 0.$$

Then

$$a \leq \frac{1}{2}\left(\frac{a^2}{b} + b\right) \quad \text{with equality iff} \quad a = b$$

and

$$\|\alpha\|_1 = \min_{\eta_j \geq 0} \frac{1}{2} \sum_{j=1}^{p} \frac{\alpha[j]^2}{\eta_j} + \eta_j.$$

The formulation becomes

$$\min_{\alpha, \eta_j \geq \varepsilon} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \frac{\lambda}{2} \sum_{j=1}^{p} \frac{\alpha[j]^2}{\eta_j} + \eta_j.$$

Important messages

- Greedy methods directly address the NP-hard ℓ_0-decomposition problem.
- Homotopy methods can be extremely efficient for small or medium-sized problems, or when the solution is very sparse.
- Coordinate descent provides in general quickly a solution with a small/medium precision, but gets slower when there is a lot of correlation in the dictionary.
- First order methods are very attractive in the large scale setting.
- Other good alternatives exists, active-set, reweighted ℓ_2 methods, stochastic variants, variants of OMP,...

Optimization for Dictionary Learning

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{p \times n} \\ \mathbf{D} \in \mathcal{C}}} \sum_{i=1}^{n} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

$$\mathcal{C} \triangleq \{\mathbf{D} \in \mathbb{R}^{m \times p} \ \text{ s.t. } \ \forall j = 1, \ldots, p, \ \ \|\mathbf{d}_j\|_2 \leq 1\}.$$

- Classical optimization alternates between $\mathbf{D}$ and $\boldsymbol{\alpha}$.
- Good results, but very slow!

Optimization for Dictionary Learning

[Mairal, Bach, Ponce, and Sapiro, 2009a]

Classical formulation of dictionary learning

$$\min_{\mathbf{D}\in\mathcal{C}} f_n(\mathbf{D}) = \min_{\mathbf{D}\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} l(\mathbf{y}_i, \mathbf{D}),$$

where

$$l(\mathbf{x}, \mathbf{D}) \triangleq \min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1.$$

Which formulation are we interested in?

$$\min_{\mathbf{D}\in\mathcal{C}} \left\{ f(\mathbf{D}) = \mathbb{E}_{\mathbf{y}}[l(\mathbf{y}, \mathbf{D})] \approx \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} l(\mathbf{y}_i, \mathbf{D}) \right\}$$

[Bottou and Bousquet, 2008]: Online learning can

- handle potentially infinite or dynamic datasets,
- be dramatically faster than batch algorithms.

Optimization for Dictionary Learning

Require: $\mathbf{D}_0 \in \mathbb{R}^{m\times p}$ (initial dictionary); $\lambda \in \mathbb{R}$

1: $\mathbf{A}_0 = 0$, $\mathbf{B}_0 = 0$.

2: **for** t=1,...,T **do**

3: Draw $\mathbf{y}_t$

4: Sparse Coding: $\boldsymbol{\alpha}_t \leftarrow \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1,$

5: Aggregate sufficient statistics
$\mathbf{A}_t \leftarrow \mathbf{A}_{t-1} + \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^T$, $\mathbf{B}_t \leftarrow \mathbf{B}_{t-1} + \mathbf{y}_t\boldsymbol{\alpha}_t^T$

6: Dictionary Update (block-coordinate descent)

$$\begin{aligned}\mathbf{D}_t &\leftarrow \arg\min_{\mathbf{D}\in\mathcal{C}} \tfrac{1}{t}\textstyle\sum_{i=1}^{t} \left(\tfrac{1}{2}\|\mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1\right). && (6)\\ &= \arg\min_{\mathbf{D}\in\mathcal{C}} \tfrac{1}{t}\left(\tfrac{1}{2}\operatorname{Tr}(\mathbf{D}^T\mathbf{D}\mathbf{A}_t) - \operatorname{Tr}(\mathbf{D}^T\mathbf{B}_t)\right). && (7)\end{aligned}$$

7: **end for**

Optimization for Dictionary Learning

Which guarantees do we have?

Under a few reasonable assumptions,

- we build a surrogate function $\hat{f}_t$ of the expected cost f verifying

$$\lim_{t \to +\infty} \hat{f}_t(\mathbf{D}_t) - f(\mathbf{D}_t) = 0,$$

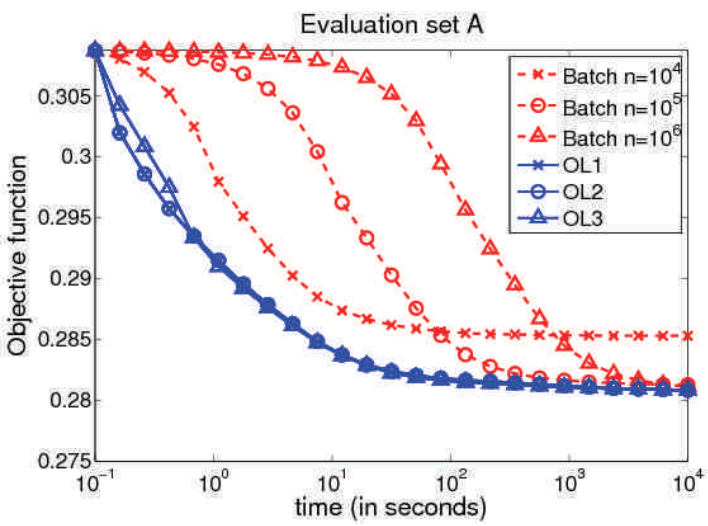
- $\mathbf{D}_t$ is asymptotically close to a stationary point.

Extensions (all implemented in SPAMS)

- non-negative matrix decompositions.
- sparse PCA (sparse dictionaries).
- fused-lasso regularizations (piecewise constant dictionaries)

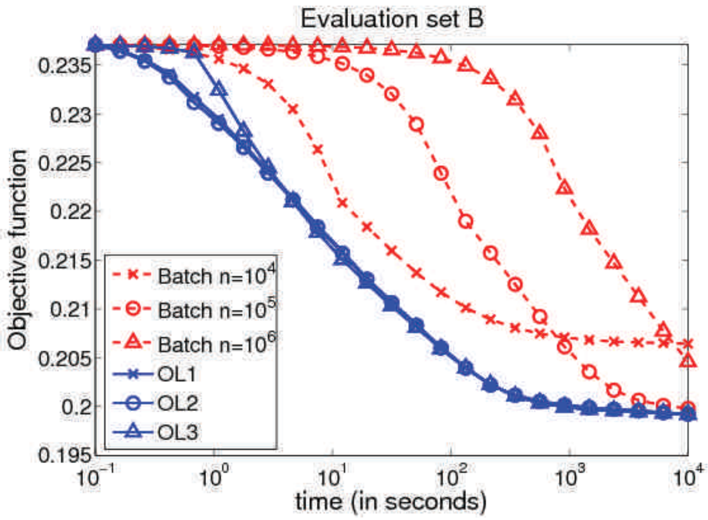
Optimization for Dictionary Learning

Experimental results, batch vs online



$m = 8 \times 8$, $p = 256$

Optimization for Dictionary Learning
Experimental results, batch vs online



$m = 12 \times 12 \times 3,\ p = 512$

References I

M. Aharon, M. Elad, and A. M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.

F. Bach. Bolasso: model consistent lasso estimation through the bootstrap. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

D. P. Bertsekas. *Nonlinear programming*. Athena Scientific Belmont, Mass, 1999.

D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

D. Blei, T. Griffiths, and M. Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2): 1–30, 2010.

J.F. Bonnans, J.C. Gilbert, C. Lemarechal, and C.A. Sagastizabal. *Numerical optimization: theoretical and practical aspects*. Springer-Verlag New York Inc, 2006.

References II

J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization: Theory and examples*. Springer, 2006.

L. Bottou and O. Bousquet. The trade-offs of large scale learning. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20, pages 161–168. MIT Press, Cambridge, MA, 2008.

Y-L. Boureau, F. Bach, Y. Lecun, and J. Ponce. Learning mid-level features for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

A. Buades, B. Coll, and J.M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modelling and Simulation*, 4(2):490–530, 2006.

A. Buades, B. Coll, J.-M. Morel, and C Sbert. Self-similarity driven color demosaicking. *IEEE Transactions on Image Processing*, 18(6):1192–1202, 2009.

E. Candes. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, volume 3, 2006.

S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1999.

References III

K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57: 1413–1457, 2004.

I. Daubechies, R. DeVore, M. Fornasier, and S. Gunturk. Iteratively re-weighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math*, 2009.

D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of statistics*, 32(2):407–499, 2004.

A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1999.

References IV

M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 54(12): 3736–3745, December 2006.

K. Engan, S. O. Aase, and J. H. Husoy. Frame based signal compression using method of optimal directions (MOD). In *Proceedings of the 1999 IEEE International Symposium on Circuits Systems*, volume 4, 1999.

J. Friedman, T. Hastie, H. Hölfling, and R. Tibshirani. Pathwise coordinate optimization. *Annals of statistics*, 1(2):302–332, 2007.

W. J. Fu. Penalized regressions: The bridge versus the Lasso. *Journal of computational and graphical statistics*, 7:397–416, 1998.

R. Grosse, R. Raina, H. Kwong, and A. Y. Ng. Shift-invariant sparse coding for audio classification. In *Proceedings of the Twenty-third Conference on Uncertainty in Artificial Intelligence*, 2007.

BK Gunturk, Y. Altunbasak, and RM Mersereau. Color plane interpolation using alternating projections. *IEEE Transactions on Image Processing*, 11(9):997–1013, 2002.

A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.

References V

K. Huang and S. Aviyente. Sparse representation for signal classification. In *Advances in Neural Information Processing Systems*, Vancouver, Canada, December 2006.

J. M. Hugues, D. J. Graham, and D. N. Rockmore. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proceedings of the National Academy of Science, TODO USA*, 107(4):1279–1283, 2009.

R. Jenatton, J-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. Technical report, 2009. preprint arXiv:0904.3523v1.

R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2010a.

R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. Technical report, 2010b. submitted, arXiv:1009.3139.

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, 2001.

H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 801–808. MIT Press, Cambridge, MA, 2007.

References VI

M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000.

J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008a.

J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, January 2008b.

J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce. Discriminative sparse image models for class-specific edge detection and image interpretation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008c.

J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modelling and Simulation*, 7(1): 214–241, April 2008d.

References VII

J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009a.

J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009b.

S. Mallat. *A Wavelet Tour of Signal Processing, Second Edition*. Academic Press, New York, September 1999.

S. Mallat and Z. Zhang. Matching pursuit in a time-frequency dictionary. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

H. M. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, 3:111–133, 1956.

N. Meinshausen and P. Buehlmann. Stability selection. *Journal of the Royal Statistical Society, Series B*, 2010. to appear.

Y. Nesterov. A method for solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math. Dokl.*, 27:372–376, 1983.

Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, CORE, 2007.

References VIII

J. Nocedal and SJ Wright. *Numerical Optimization*. Springer: New York, 2006. 2nd Edition.

B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

M. R. Osborne, B. Presnell, and B. A. Turlach. On the Lasso and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–37, 2000.

D. Paliy, V. Katkovnik, R. Bilcu, S. Alenius, and K. Egiazarian. Spatially adaptive color filter array interpolation for noiseless and noisy data. *Intern. J. of Imaging Sys. and Tech.*, 17(3), 2007.

J. Portilla, V. Strela, MJ Wainwright, and EP Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003.

M. Protter and M. Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing*, 18(1):27–36, 2009.

S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

References IX

V. Roth and B. Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. Collaborative hierarchical sparse modeling. Technical report, 2010a. Preprint arXiv:1003.0400v1.

P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar. Collaborative hierarchical sparse modeling. Technical report, 2010b. Preprint arXiv:1003.0400v1.

R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.

S. Weisberg. *Applied Linear Regression*. Wiley, New York, 1980.

J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2005.

J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

References X

J. Yang, K. Yu, , and T. Huang. Supervised translation-invariant sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, 68:49–67, 2006.

L. Zhang and X. Wu. Color demosaicking via directional linear minimum mean square-error estimation. *IEEE Transactions on Image Processing*, 14(12): 2167–2178, 2005.

Sébastien PARIS

LSIS / Univ. Med. / DYNI

http://www.lsis.org/spip.php?id_rubrique=28

http://www.mathworks.com/matlabcentral/fileexchange/authors/13308

" Reconnaissance de Scènes & Robotique avec Vision Embarquée "

Cette présentation liée au challenge RobotVision@ICPR 2010 fait un tour d'horizon de la chaîne complète de traitement et des différentes techniques (basées sur de la vision et du machine learning) dédiées à la tache de catégorisation (dynamique) de scènes d'intérieur.

Ce tour d'horizon commencera par une revue des descripteurs locaux/globaux utilisés (LBP, spHOG, SIFT, etc...), des techniques d'encodage des descripteurs en dictionnaires visuels (VQ, soft VQ, Sparse Coding, etc...), en passant par les classifieurs à vastes marges dédiés aux grandes échelles (FastIKSVM, Liblinear, etc...) sur noyaux potentiellement multiples (MKL, GMKL, etc...).

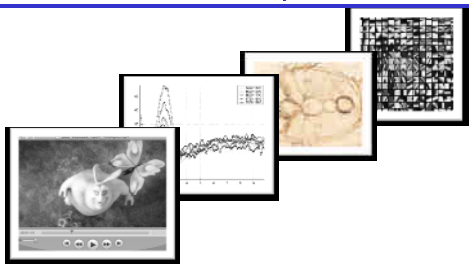
Nous montrons comment intégrer à cet ensemble la dynamique de l'état (approche HMM, particulaire, etc...), et abouti au meilleur modèle de localisation robotique dans le challenge ImageClef ICPR 2010.

Références :

Sébastien Paris, Hervé Glotin, "PyramidalMulti-Level Features for the RobotVision@ICPR 2010 Challenge" , pp. 1-4, ICPR'2010, Turkey, 2010

Sébastien Paris, Hervé Glotin, "Linear SVM for new Pyramidal Multi-Level Visual only Concept" , CLEF (Notebook Papers/LABs/Workshops), http://clef2010.org/resources/proceedings/clef2010labs_submission_118.pdf, 2010

Catégorisation de scènes par vision appliquée à la robotique



ERMITES 2010, 28-30 septembre

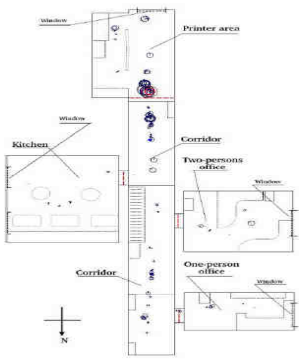
Sébastien Paris, LSIS/DYNI

Introduction (1/3)

- Robovision@ICPR 2010 challenge [1] : « which room I am ? »
  - ✓ Binocular cameras
  - ✓ 9 visited rooms in training 'easy' & 'hard' databases
  - ✓ 'validation' database to tune the system
  - ✓ 'test' sequence with different conditions (people, moved furnitures, illumination, unseen part of the rooms, unknown visited rooms)

Left camera

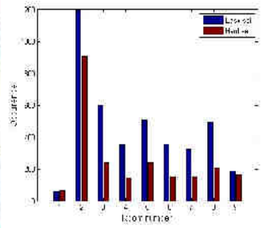
Right camera



Floorplan

Introduction (2/3)

Rooms : Elevator(1), Corridor(2), Kitchen(3), LargeOffice1(4), LargeOffice2(5), SmallOffice2(6), StudentOffice(7), Lab(8), PrinterArea(9), Unknown(10)

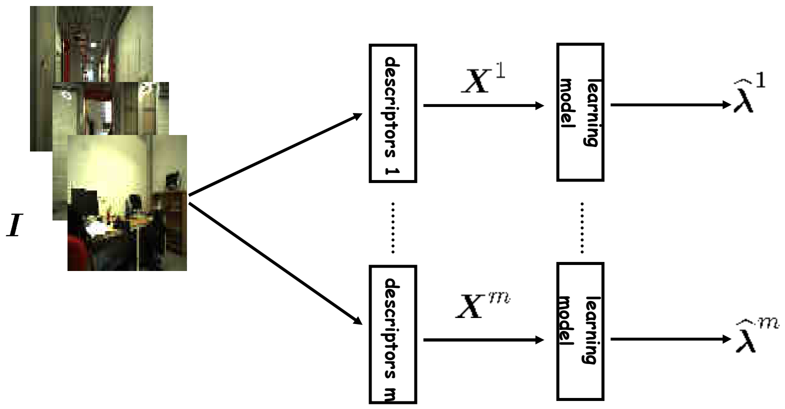


Introduction (3/3)

- ❑ Fundamentally a machine learning problem: scene/object recognition/matching task
  - ✓ Numerous databases available: Caltech 101/256 [2,3], Graz-02 [4], VOC [5], 15 class scene [6], …
  - ✓ Thousand references and algorithms
- ❑ but with some specificities
  - ✓ stereoscopic vision
  - ✓ robot's motion
  - ✓ medium size of training databases ([4500-6000] stereoscopic frames)
- ❑ Several architectures
  - ✓ Machine Learning architecture
  - ✓ Matching architecture
  - ✓…

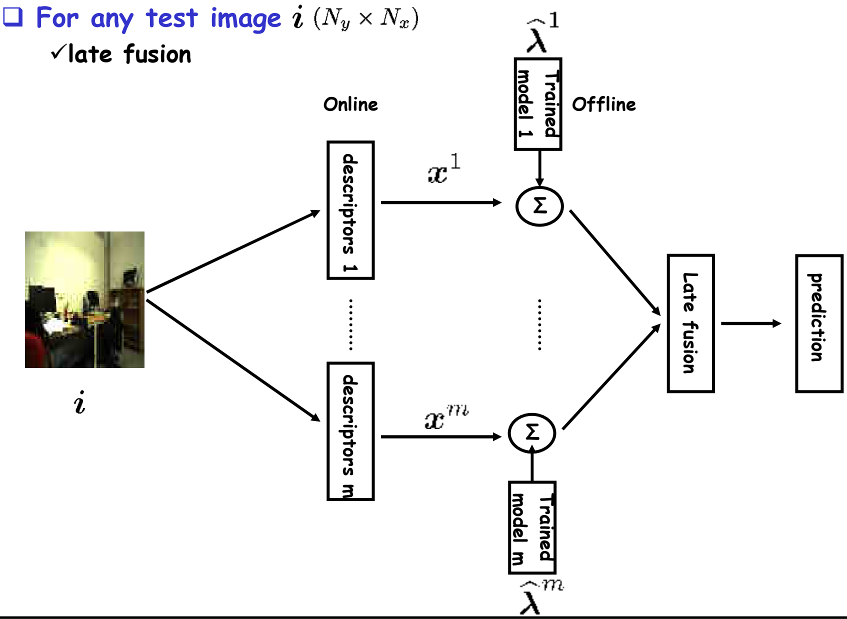
Late Fusion architecture (1/2)

- Given a set of training images $I = \{i_1, \ldots, i_N\} \subset \mathcal{I}$
- Classical machine learning architecture in vision
  - ✓ descriptors computed on images set X^l $\{x_1^l, \ldots, x_N^l\} \subset \mathcal{X}^l$.
  - ✓1 model per descriptor (classifier parameters)



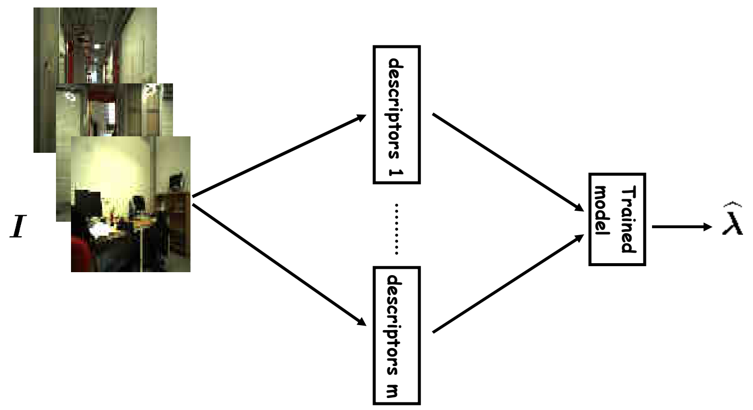
Late Fusion architecture (2/2)

- For any test image i $(N_y \times N_x)$
  - ✓late fusion



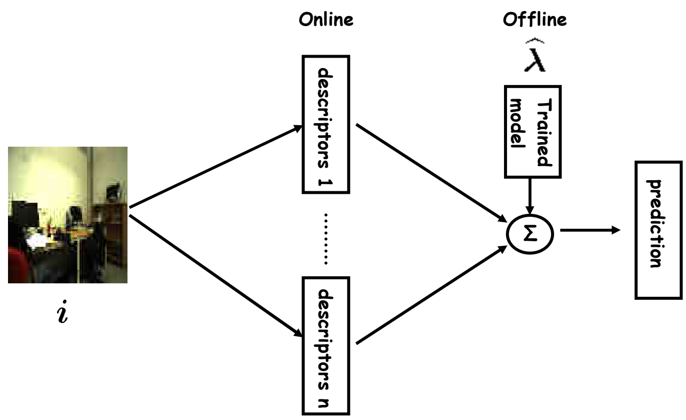
Early Fusion architecture (1/2)

- ❑ Early fusion machine learning architecture in vision
  - ✓1 fused model (descriptors concatenation, MKL, etc …)



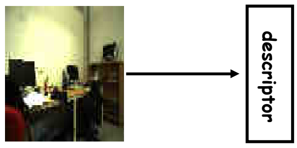
Early Fusion architecture (2/2)

- ❑ For any test image i $(N_y \times N_x)$



Descriptors in Vision

- ❑ Role of image descriptors in vision
  - ✓Extract information from image (shape, color, texture, moments)
  - ✓Usually more informative than the image itself
  - ✓Reduce considerably the size of input vectors



$I \subset \mathcal{I}$ $\qquad$ $x \subset \mathcal{X}$

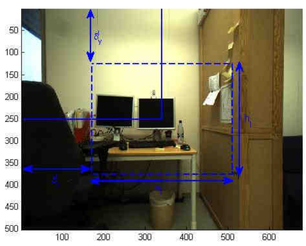
- ❑ descriptors properties requiered
  - ✓Invariant to image transformation (scale, rotation, illumination, etc…)
  - ✓Robust to noise
  - ✓Capture global & local informations
  - ✓Dimension of up to date descriptors can be $d \geq 10000$
  - ✓Used directly as input of classfiers or matching procedures

Spatial pyramidal architecture

❑ Capture local information at different scales

❑ Each image i is divided in N_s subwindows $(h_l \times w_l)$ according to a L levels pyramid where $h_l = \lfloor Ny.r_{y,l} \rfloor$ and $w_l = \lfloor Nx.r_{x,l} \rfloor$

❑ Subwindows shifts are defined by $\delta_{y,l} = \lfloor Ny.d_{y,l} \rfloor$ and $\delta_{x,l} = \lfloor Nx.d_{x,l} \rfloor$

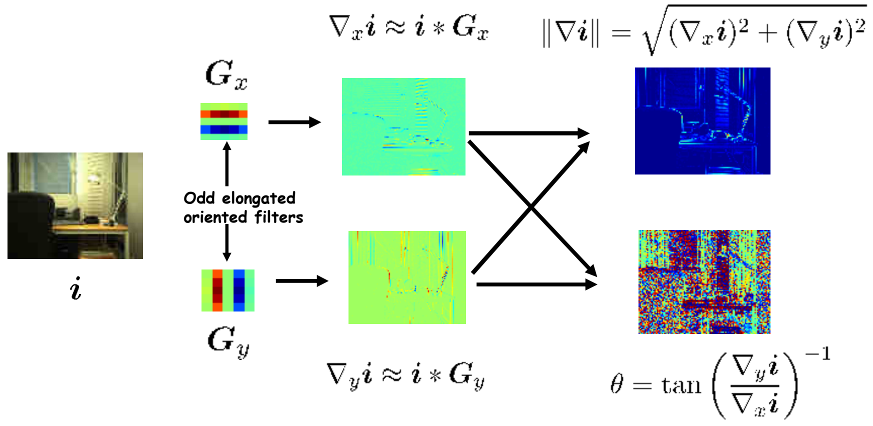


$$Ns = \sum_{l=1,\dots,L} \lfloor \frac{(1-r_{y,l})}{(d_{y,l}+1)} \rfloor . \lfloor \frac{(1-r_{x,l})}{(d_{x,l}+1)} \rfloor$$

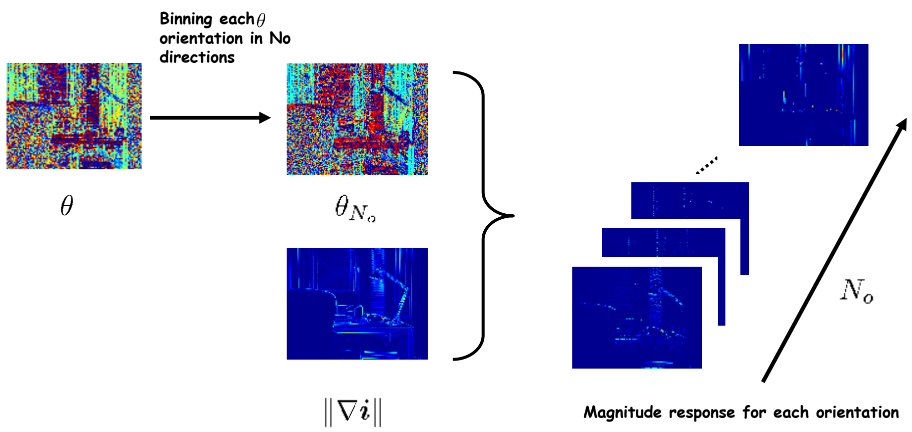
❑ Histograms built for each subwindows are weighted according their level l

Histogram of Oriented Edge Energy (1/3)

- Variant of Histogram of Gradient (HOG) introduced by Maiji [7]

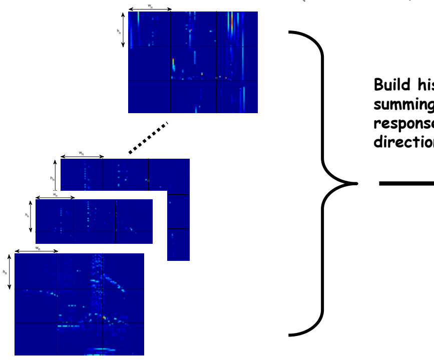


Histogram of Oriented Edge Energy (2/3)



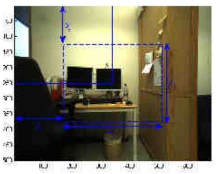
Histogram of Oriented Edge Energy (3/3)

L1 normalization with block of size $(h_n \times w_n)$



Build histogram by summing edge energy responses in each direction

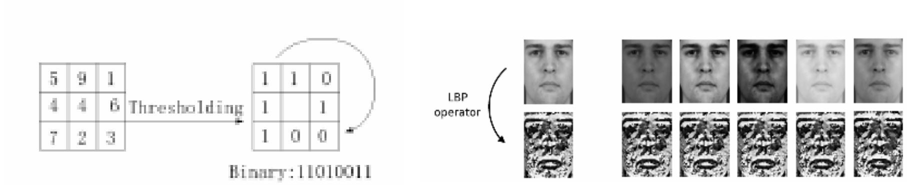
$$\boldsymbol{x} = [x_1, \ldots, x_{N_o}]^T$$



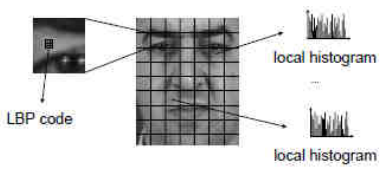
With the spatial pyramid architecture, each histogram is built for each subwindows with the Integral Histogram method $\longrightarrow \boldsymbol{x} \in \mathcal{R}^{N_s N_o}$

Local Binary Pattern (1/4)

- ❑ LBP
  - ✓ Encode relationship between central pixel and its neighbors
  - ✓ LBP can be considered as a parametric visual word
  - ✓ Robust to illumination variations

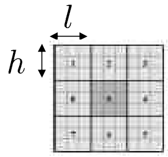


- ❑ Local Histogram of LBP for each subwindows



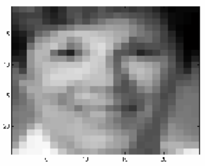
Local Binary Pattern (2/4)

- Variant of LBP
  - ✓ Multiblock LBP with a block of size $(h \times l)$ [11]. Thanks to integral image to accelerate computations
  - ✓ Center Symmetric LBP [12]
  - ✓ Gabor LBP [13]
  - ✓ Color LBP [14]
  - ✓ Local Derivative Pattern [15]

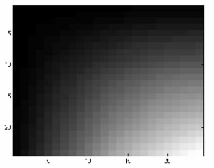


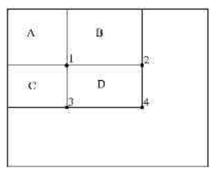
Integral Image $I'(x, y) = \sum_{x'=1}^{x'=n_x} \sum_{y'=1}^{y'=n_y} I(x', y')$

I



I'





$D = 4 - (3+2) + 1$

Local Binary Pattern (3/4)

The object category classification results in AUC for the SIFT descriptor and for eight different CS-LBP descriptors

Class	SIFT	CS-LBP: $W = W_U$, $T = 0.01$							
		$M = 4$				$M = 3$			
		$R,N = 2,8$	$R,N = 1,8$	$R,N = 2,6$	$R,N = 1,6$	$R,N = 2,8$	$R,N = 1,8$	$R,N = 2,6$	$R,N = 1,6$
Bicycle	0.9191	0.9167	0.9171	0.9029	0.9007	0.9220	0.9143	0.9067	0.9077
Bus	0.9726	0.9731	0.9745	0.9738	0.9712	0.9727	0.9740	0.9699	0.9690
Car	0.9595	0.9666	0.9665	0.9682	0.9672	0.9645	0.9675	0.9644	0.9660
Cat	0.8824	0.8883	0.8838	0.8829	0.8921	0.8853	0.8822	0.8827	0.8845
Cow	0.8967	0.9155	0.9113	0.9077	0.9138	0.9059	0.9128	0.9113	0.9091
Dog	0.8192	0.8317	0.8303	0.8254	0.8350	0.8363	0.8384	0.8274	0.8299
Horse	0.8449	0.8869	0.8932	0.8879	0.8948	0.9036	0.8794	0.8911	0.8763
Motorbike	0.9391	0.9502	0.9523	0.9346	0.9419	0.9397	0.9515	0.9264	0.9409
Person	0.8068	0.8193	0.8295	0.8079	0.8172	0.8131	0.8200	0.8083	0.8118
Sheep	0.8959	0.9197	0.9241	0.9207	0.9176	0.9231	0.9197	0.9235	0.9199
Mean	0.8936	0.9068	0.9083	0.9012	0.9052	0.9066	0.9060	0.9012	0.9015

Abbreviations: M, Cartesian grid size; W, weighting method; W_U, uniform; (R, N, T), CS-LBP operator parameters.

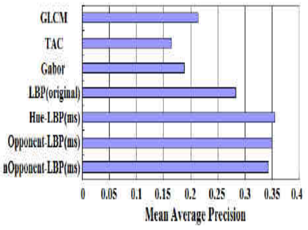
The object category classification results in AUC for eight different LBP descriptors

Class	LBP: $W = W_U$, $T = 0.01$							
	$M = 4$				$M = 3$			
	$R,N = 2,4$	$R,N = 1,4$	$R,N = 2,3$	$R,N = 1,3$	$R,N = 2,4$	$R,N = 1,4$	$R,N = 2,3$	$R,N = 1,3$
Bicycle	0.9268	0.9303	0.9075	0.9199	0.9187	0.9238	0.9138	0.9219
Bus	0.9741	0.9717	0.9711	0.9664	0.9724	0.9718	0.9656	0.9629
Car	0.9695	0.9701	0.9652	0.9656	0.9684	0.9716	0.9632	0.9663
Cat	0.8969	0.9014	0.8960	0.8909	0.8917	0.8982	0.8988	0.8962
Cow	0.9156	0.9215	0.9132	0.9156	0.9236	0.9239	0.9159	0.9167
Dog	0.8357	0.8351	0.8410	0.8339	0.8402	0.8449	0.8324	0.8397
Horse	0.8877	0.8877	0.8794	0.8848	0.8923	0.8941	0.8753	0.8766
Motorbike	0.9431	0.9524	0.9411	0.9459	0.9426	0.9477	0.9366	0.9333
Person	0.8248	0.8328	0.8022	0.8128	0.8217	0.8277	0.7914	0.8071
Sheep	0.9273	0.9294	0.9223	0.9219	0.9308	0.9316	0.9239	0.9241
Mean	0.9102	0.9132	0.9039	0.9058	0.9102	0.9135	0.9017	0.9045

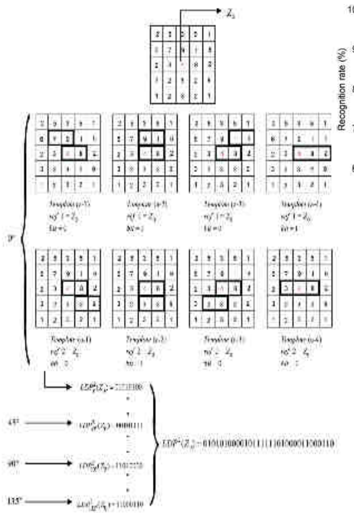
Abbreviations: M, Cartesian grid size; W, weighting method; W_U, uniform; (R, N, T), LBP operator parameters.

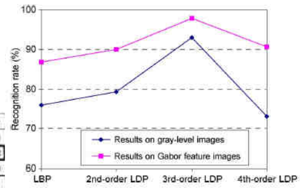
Local Binary Pattern (4/4)

Colors LBP



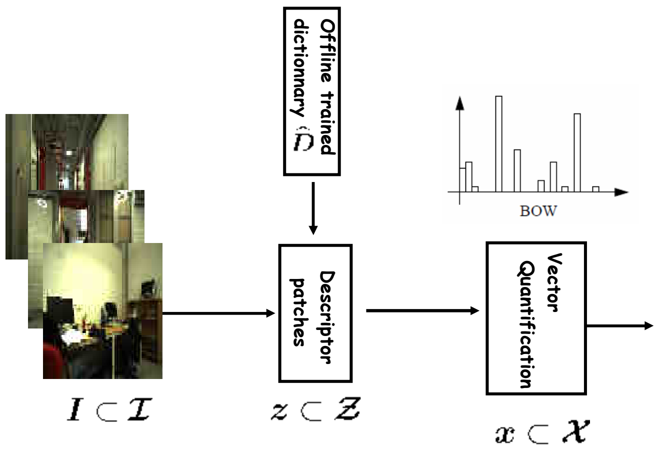
LDP





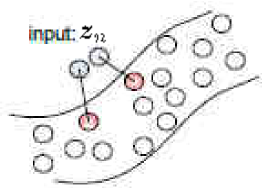
Traditional Bag of Word (1/3)

- Bag of (visual) Word (BoW)
  - ✓Quantize local patches into discrete visual words with a trained codebook
  - ✓Compute a compact histogram representation



Traditional Bag of Word (2/3)

- Visual words (or codebook) are usually trained by K-means or GMM
- Given a collection of descriptors computed on each training image $\boldsymbol{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\} \subset \mathcal{Z}$
- the codebook $\boldsymbol{D}$ with K words is computed by:



$$\min_{\boldsymbol{D}} \sum_{n=1}^{N} \min_{k=1,\ldots,K} \|\boldsymbol{z}_n - \boldsymbol{d}_k\|^2$$

or reformulated by: $\min_{\boldsymbol{U},\boldsymbol{D}} \sum_{n=1}^{N} \|\boldsymbol{z}_n - \boldsymbol{u}_n \boldsymbol{D}\|^2$ with $Card(\boldsymbol{u}_n) = 1, |\boldsymbol{u}_n| = 1, \succeq 0$

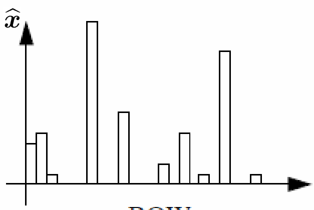
With N>1000000, fast Kmeans is required [20]
(thanks to yael https://gforge.inria.fr/projects/yael)

Traditional Bag of Word (3/3)

With a test image with $P \ll N$ patches, when $\widehat{\boldsymbol{D}}$ is trained, $\widehat{\boldsymbol{U}}$ is retrieved by:

$$\min_{\{\boldsymbol{u}_i\}_{i=1,\ldots,P}} \sum_{p=1}^{P} \|\boldsymbol{z}_p - \boldsymbol{u}_p \widehat{\boldsymbol{D}}\|^2$$

Histogram of a test image is obtained by: $\widehat{\boldsymbol{x}} = \frac{1}{P} \sum_{p=1}^{P} \widehat{\boldsymbol{u}}_p$ and normalized



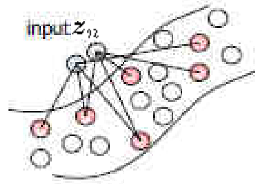
Sparse coding (1/3)

- Better patches (sparse) representation can be obtained with sparse coding (SC)

$$\min_{\{\boldsymbol{u}_i\}_{i=1,\dots,N},\boldsymbol{D}} \sum_{n=1}^{N} \|\boldsymbol{z}_n - \boldsymbol{u}_n \boldsymbol{D}\|^2 + \lambda \|\boldsymbol{u}_i\|_{l_1} \quad \text{with} \quad \|\boldsymbol{d}_k\|^2 \leq 1,\ k = 1,\dots,K$$

enforcing sparsity

Tradeoff between representation fidelity and sparsity



U, D are optimized alternatively

i) Fixing D, U is found by linear programming

ii) Fixing U, D is found by quadratically constrained quadratic programming

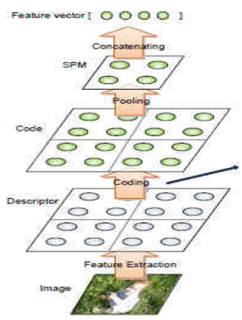
Sparse coding (2/3)

i) With a test image, when $\widehat{D}$ is trained, $\widehat{U}$ is retrieved sparse coding with [21]:

$$\min_{U} \|Z - \widehat{D}U\|^2 + \lambda \|U\|_{l_1}$$

ii) The feature can be computed for example with a max pooling strategy

$$\widehat{\boldsymbol{x}} = \max\{|\widehat{U}|\}$$



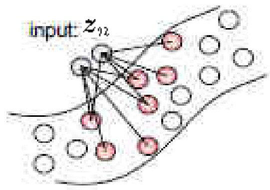
1. Less quantization error than VQ
2. Image patches are sparse in nature
3. Max pooling is more salient and robust to local translations

Sparse coding (3/3)

- ❑ Variant of sparse coding with Locality constrained Linear Coding [22]

$$\min_{\{u_i\}_{i=1,\dots,P}} \sum_{p=1}^{P} \|z_p - u_p \hat{D}\|^2 + \lambda \|u_p \odot f(z_p, \hat{D})\|^2$$ with $|u_n| \quad 1$

similarity function

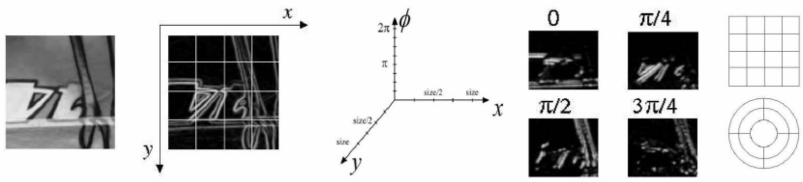


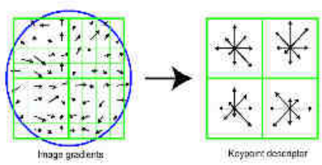
With Sparse Coding, linear classifiers are very efficient (due to the sparsity)

- ❑ Supervised dictionary learning [23] : learn jointly $\hat{D}$ and $\hat{w}$

Histogram of Gradient/Dense SIFT (1/3)

- ❑ Simplified version of SIFT
  - ✓ no keypoints detection, SIFT [8] patches evaluated on dense grid
  - ✓ no dominant orientation alignment
  - ✓ single scale
  - ✓ 3D histogram of gradient locations and orientations
  - ✓ A (4*4*8) vector

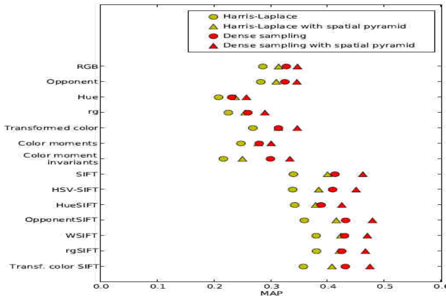




Histogram of Gradient/Dense SIFT (2/3)

- Variant of SIFT
  - ✓ SURF
  - ✓ Color SIFT [9,10]
  - ✓ …

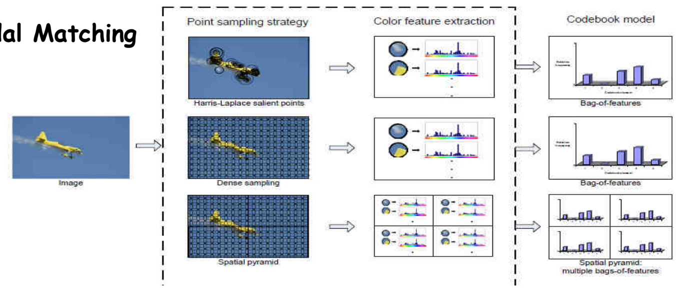
	Light intensity change	Light intensity shift	Light intensity change and shift	Light color change	Light color change and shift
RGB Histogram	-	-	-	-	-
O_1, O_2	-	+	-	-	-
O_3, Intensity					
Hue	+	+	+	-	-
Saturation	+	+	+		
r, g	+				
Transformed color	+	+	+	+	+
Color moments		+			
Moment invariants	+	+	+	+	+
SIFT (∇I)	+	+	+	+	+
HSV-SIFT	+	+	+	+/-	+/-
HueSIFT	+	+	+	+/-	+/-
OpponentSIFT	+/-	+	+/-	+/-	+/-
W-SIFT	+	+	+	+/-	+/-
rgSIFT	+	+	+	+/-	+/-
Transf. color SIFT	+	+	+	+	+

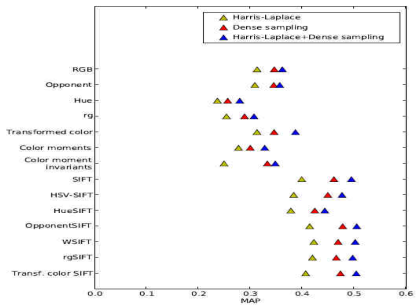


Pascal VOC 2007

Histogram of Gradient/Dense SIFT (3/3)

With Spatial Pyramidal Matching





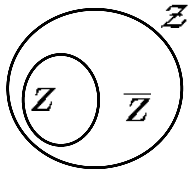
Statistical Machine Learning (1/3)

Training databases $\boldsymbol{I} = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N\} \subset \boldsymbol{\mathcal{I}}$ with corresponding labels $Y = \{y_1, \ldots, y_N\} \subset \mathcal{Y}$, where $y_i = 1, \ldots, L,\ i = 1, \ldots, N.$

Given the descriptors computed over $\boldsymbol{I}$

$$\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \boldsymbol{\mathcal{X}}.$$

We define: $\boldsymbol{Z} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\} \subset \boldsymbol{\mathcal{Z}}.$



$\overline{\boldsymbol{Z}} = \boldsymbol{\mathcal{Z}} \cap \boldsymbol{Z}$

We assume than examples are drawn i.i.d. according to $p(\boldsymbol{x}, y)$

Expected Risk $\quad E(f) = \displaystyle\int_{\boldsymbol{x}, y \in \boldsymbol{\mathcal{X}} \times \mathcal{Y}} \mathcal{L}(y, f(\boldsymbol{x})) p(\boldsymbol{x}, y) d\boldsymbol{x} dy$

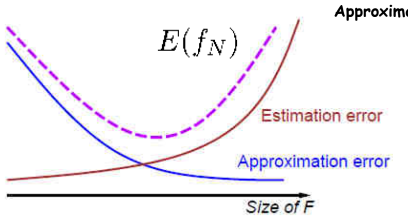
Empirical Risk $\quad E_N(f) = \displaystyle\sum_{i=1}^{N} \mathcal{L}(y_i, f(\boldsymbol{x}_i)),\ \boldsymbol{x}_i \in \boldsymbol{X}, y_i \in Y$

Statistical Machine Learning (2/3)

- We would like to find $f^\star$ that minimize $E(f)$ among all functions

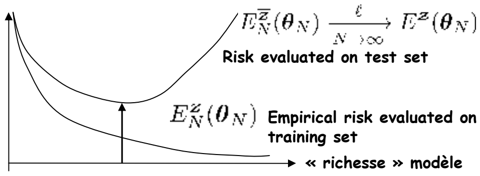
But in general $f^\star \notin \mathcal{F}$ and the best we can find $f^\star_{\mathcal{F}} \in \mathcal{F}$ than minimize $E(f)$

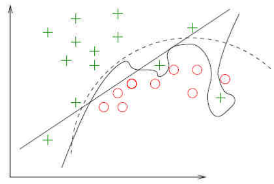
- Since $p(\boldsymbol{x}, y)$ is unknown, we compute $f_N \in \mathcal{F}$ that minimize $E_N(f)$.
- A natural idea is to find $\quad f_N = \arg\min_{f \in \mathcal{F}} \{E_N(f)\}$ but often costly
- Error $\quad E(f_N) - E(f^\star) = \underbrace{E(f^\star_{\mathcal{F}}) - E(f^\star)}_{\text{Approximation error}} + \underbrace{E(f_N) - E(f^\star_{\mathcal{F}})}_{\text{Estimation error}}$



Statistical Machine Learning (3/3)

❑ Usually family $\mathcal{F}$ functions depend on parameter $\boldsymbol{\theta}$ i.e. $f_N = f(\boldsymbol{\theta}_N)$





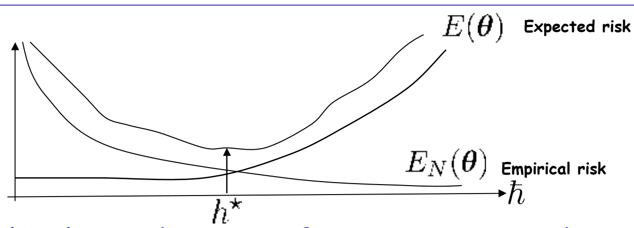
❑ First attempt was to stop model optimization when test error start to increase

❑ When $E_N(\boldsymbol{\theta}_N) \xrightarrow[N \to \infty]{\ell} \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \{E(\boldsymbol{\theta})\}$?. Vapnik theory with Structural Error Minimization

❑ Under strong PAC, with a probability $1-\eta$

$$E(\boldsymbol{\theta}) \leq E_N(\boldsymbol{\theta}) + \underbrace{\sqrt{\frac{\hbar(\log(\frac{2N}{\hbar}+1)) - \log(\frac{\eta}{4})}{N}}}_{\text{Confidence interval } C_N(h)}$$

Structural Risk Minimisation (SRM)



❑ Empirical Risk is a decreasing function versus VC-dimension $.\hbar$

❑ $C_N(h)$ is a increasing function when:

- ✓ Number of training examples decrease
- ✓ VC $\hbar$ dimension increase (usually proportional to feature dimension size, number of neurons, etc…)

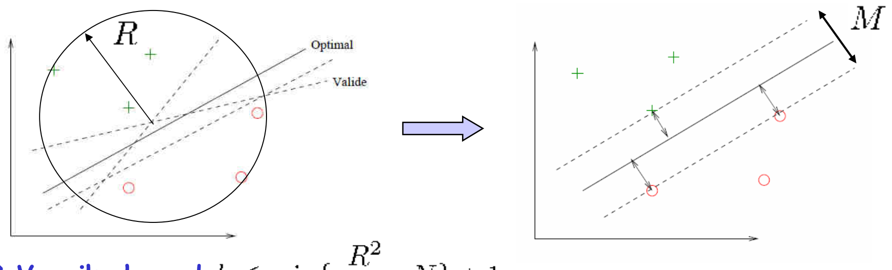
$h = d+1$ for linear Kernel

⇨ Minimization of the Structural Risk:

$$\{\widehat{\boldsymbol{\theta}}^N, \widehat{\hbar}^N\} = \arg \min_{\{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \hbar \in h\}} \{E_N(\boldsymbol{\theta}) + C_N(\hbar)\}$$

SRM and Large Margin classifiers

Large Margin classifier objective (e.g. SVM) : find θ minimizing Empirical Risk with the largest margin M



- Vapnik showed $h \leq \min\{\frac{R^2}{4M^2}, N\} + 1$

$\Rightarrow$ Find θ with the largest M decreases $C_N(h)$

- For LVQ approaches

$$E(\boldsymbol{\theta}) \leq E_N(\boldsymbol{\theta}) + \frac{8K \cdot U(U-1)B(B+1)(B+2)\sqrt{N}}{M^{\boldsymbol{\lambda}}(\boldsymbol{x};\boldsymbol{\theta})N} + \left(\frac{8K \cdot U(U-1)}{M^{\boldsymbol{\lambda}}(\boldsymbol{x};\boldsymbol{\theta})N}\right)\sqrt{\frac{\ln(\frac{4}{\delta})}{2N}}.$$

Linear SVM formulation

- SVM learning rule (primal form)

$$\min_{\boldsymbol{w}} \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{N}\sum_{i=1}^{N}\max\{0, 1 - y_i\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle\}$$

- Can be written as QP problem:

$$\min_{\boldsymbol{w},\boldsymbol{\xi}} \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{N}\sum_{i=1}^{N}\xi_i \quad \text{s.t.} \quad \forall i,\ 1 - y_i\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle \leq \xi_i,\ \xi_i \geq 0$$

- Dual form:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle \quad \text{s.t.}\ 0 \leq \alpha_i \leq \frac{1}{\lambda n}$$

- Dual form initially preferred
  - ✓ Easy non-linear version with Kernel Trick
  - ✓ Easier to handle constraints (variables lie in a bounded interval)

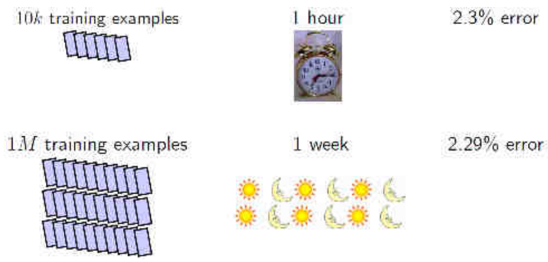
Classic SVM solvers

- Original QP solver introduced by Vapnik is in the worst case $O(N^3)$ in runtime (involve all non-zero Lagrange multipliers)

- SMO technique (Platt, 1999) break up the SVM optimization problem in smaller subproblems. Optimize pair of Lagrange multipliers. The worst case in runtime is $O(dN^2)$

  ⇨ Still infeasible on large datasets

- With large training sets:



⇨ Can we approximate f_N to accelerate learning ?

Large Scale problems

- For Large Scale problems, the active budget constraint is the computing time T

- Let's assume our solver returns $\widetilde{f}_N$ such that (ρ approximation)

$$E_N(\widetilde{f}_N) \leq E_N(f_N) + \rho$$

Stop iterative optimization long before convergence

- Error

$$E(\widetilde{f}_N) - E(f^\star) = \underbrace{E(f^\star_{\mathcal{F}}) - E(f^\star)}_{\text{Approximation error}} + \underbrace{E(f_N) - E(f^\star_{\mathcal{F}})}_{\text{Estimation error}} + \underbrace{E(\widetilde{f}_N) - E(f_N)}_{\text{Optimization error}}$$

- Computation time depends on 3 variables $T(\mathcal{F}, N, \rho)$

- Exact tradeoff depends on optimization algorithm

Linear Large Scale solvers

- Fixed $\mathcal{F} = f(x;\boldsymbol{\theta}) = f_{\boldsymbol{w}}(x)$
- 2 solvers are commonly used [18]
  - ✓ PEGASOS: online primal optimization form with SGD/SGP
  - ✓ Liblinear: batch dual optimization form with dual coordinate descent
- Pegasos runtime $O(\frac{d}{\lambda\rho})$ [16]
  - ✓ pick random training samples
  - ✓ independent of the number of examples
  - ✓ slow convergence
  - ✓ can be extended to nonlinear kernels thanks to the representer theorem
- Liblinear runtime $O(Nd\log(\frac{1}{\rho}))$ [17]
  - ✓ Series of univariate optimizations
  - ✓ Achieves lower generalization error quicker than Pegasos
  - ✓ faster convergence
  - ✓ but runtime decreases on larger training sets

Pegasos

for $t = 1 \ldots T$ **do**

Pick random $A_t \subseteq \mathcal{T}$ such that $|A_t| = k$

$\mathcal{M} := \{(x, y) \in A_t : 1 - y(w \cdot x) > 0\}$

$\nabla_t := \lambda w_t - \frac{1}{|\mathcal{M}|}\sum_{(x,y)\in\mathcal{M}} yx$

Update $w_{t+\frac{1}{2}} \leftarrow w_t - \frac{1}{\lambda t} \cdot \nabla_t$

Let $w_{t+1} \leftarrow \min\left(1, \frac{1}{\sqrt{\lambda}\|w_{t+\frac{1}{2}}\|}\right) w_{t+\frac{1}{2}}$

end for

return w_{T+1}

Liblinear

Let $\alpha \leftarrow$ zero vector in $\mathbb{R}^n$
Let $w \leftarrow$ zero vector in $\mathbb{R}^d$

while α is not optimal **do**
  Pick an index $i \in \{1, \ldots, n\}$
  $\alpha_i^{\text{old}} \leftarrow \alpha_i$
  $\nabla \leftarrow y_i(w \cdot x_i) - 1 + D_{ii}\alpha_i$
  $\nabla_P \leftarrow \min(\max(\nabla, 0), U)$

  if $\nabla_P \neq 0$ **then**
    $\alpha_i \leftarrow \min(\max(\alpha_i - \nabla / Q_{ii}, 0), U)$
    $w \leftarrow w + (\alpha_i - \alpha_i^{\text{old}}) y_i x_i$
  end if
end while

Large Scale SVM solvers …

Algorithm	Citation	SVM type	Optimization type	Style	Runtime
SMO	[Platt, 1999]	Kernel	Dual QP	Batch	$\Omega(n^2 d)$
SVM^{light}	[Joachims, 1999]	Kernel	Dual QP	Batch	$\Omega(n^2 d)$
Core Vector Machine	[Tsang et al., 2005, 2007]	SL Kernel	Dual geometry	Batch	$O(s/\rho^4)$
SVM^{perf}	[Joachims, 2006]	Linear	Dual QP	Batch	$O(ns/\lambda\rho^2)$
NORMA	[Kivinen et al., 2004]	Kernel	Primal SGD	Online(-style)	$\tilde{O}(s/\rho^2)$
SVM-SGD	[Bottou, 2007]	Linear	Primal SGD	Online-style	Unknown
Pegasos	[Shalev-Shwartz et al., 2007]	Kernel	Primal SGD/SGP	Online-style	$\tilde{O}(s/\lambda\rho)$
LibLinear	[Hsieh et al., 2008]	Linear	Dual coordinate descent	Batch	$O(nd \cdot \log(1/\rho))$
SGD-QN	[Bordes and Bottou, 2008]	Linear	Primal 2SGD	Online-style	Unknown
FOLOS	[Duchi and Singer, 2008]	Linear	Primal SGP	Online-style	$\tilde{O}(s/\lambda\rho)$
BMRM	[Smola et al., 2007]	Linear	Dual QP	Batch	$O(d/\lambda\rho)$
OCAS	[Franc and Sonnenburg, 2008]	Linear	Primal QP	Batch	$O(nd)$

Efficient Additive Kernels (1/3)

- Use fast solvers for linear kernel with feature map approximating additive kernels. The framework is [19]:
  1. provide explicit feature maps for all homogenous additive kernels,
  2. derive corresponding approximate finite dimensional feature maps based on Fourier sampling theorem
  3. quantify the extend of the approximation
- Homogeneous additive kernel:
  - ✓ Chi2
  - ✓ Intersection Hellinger
- Additive kernel for finite distributions (histograms)

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{b=1}^{B} k(\boldsymbol{x}_b, \boldsymbol{y}_b)$$

- Homogeneous kernel $\forall c \geq 0: \; k(c\boldsymbol{x}, c\boldsymbol{y}) = ck(\boldsymbol{x}, \boldsymbol{y})$

Efficient Additive Kernels (2/3)

- Homogeneous kernel with $c = \sqrt{\boldsymbol{xy}}$

$k(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\boldsymbol{xy}}k\left(\sqrt{\frac{\boldsymbol{x}}{\boldsymbol{y}}}, \sqrt{\frac{\boldsymbol{y}}{\boldsymbol{x}}}\right) = \sqrt{\boldsymbol{xy}}\mathcal{K}\left(\log(\frac{\boldsymbol{y}}{\boldsymbol{x}})\right)$ where $\mathcal{K}(\omega) \quad k(e^{\quad \omega/2}, e^{\omega/2})\omega = \log(\frac{\boldsymbol{y}}{\boldsymbol{x}})$

- There exist a symmetric non negative measure $\kappa(\lambda)\, d\lambda$

$$k(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\boldsymbol{xy}} \int_{-\infty}^{+\infty} e^{-i\lambda \log(\frac{\boldsymbol{y}}{\boldsymbol{x}})} \kappa(\lambda) d\lambda \qquad \mathcal{K}(\omega) = \int_{-\infty}^{+\infty} e^{-i\lambda\omega} \kappa(\lambda) d\lambda$$

- Analytic form of feature maps

$$k(\boldsymbol{x}, \boldsymbol{y}) = \int_{-\infty}^{+\infty} [\Psi(\boldsymbol{x})]_\lambda^* [\Psi(\boldsymbol{y})]_\lambda \, d\lambda \qquad [\Psi(\boldsymbol{x})]_\lambda \quad e^{i\lambda \log(\boldsymbol{x})} \sqrt{\boldsymbol{x}\kappa(\lambda)}$$

kernel	$k(x, y)$	$\mathcal{K}(\omega)$	$\kappa(\lambda)$	feature $[\Psi(x)]_\lambda$	γ-homogeneous variant
Hellinger's	$\sqrt{xy}$	1	$\delta(\lambda)$	$\sqrt{x}$	$(xy)^{\frac{\gamma}{2}}$
χ^2	$2\frac{xy}{x+y}$	$\mathrm{sech}(\omega/2)$	$\mathrm{sech}(\pi\lambda)$	$e^{i\lambda \log x}\sqrt{x \,\mathrm{sech}(\pi\lambda)}$	$2\frac{(xy)^{\frac{\gamma+1}{2}}}{x+y}$
intersection	$\min\{x, y\}$	$e^{-\lvert\omega\rvert/2}$	$\frac{2}{\pi}\frac{1}{1+4\lambda^2}$	$e^{i\lambda \log x}\sqrt{\frac{2x}{\pi}\frac{1}{1+4\lambda^2}}$	$\min\left\{x^{\frac{\gamma+1}{2}}y^{\frac{\gamma-1}{2}}, x^{\frac{\gamma-1}{2}}y^{\frac{\gamma+1}{2}}\right\}$

Efficient Additive Kernels (3/3)

- Approximated finite feature maps
  - ✓ Sampling at $\lambda \quad nL, (\quad n + 1)L, \ldots, nL$

$$
\frac{[\Psi(x)]_j}{\sqrt{xL}} = \begin{cases} \sqrt{\kappa(0)}, \ j = 0 \\ \sqrt{2\kappa(\frac{j+1}{2}L)}\cos(\frac{j+1}{2}L\log(x)), \ j > 0 \ odd \\ \sqrt{2\kappa(\frac{j}{2}L)}\sin(\frac{j}{2}L\log(x)), \ j > 0 \ even \end{cases}
$$

Approximate signature is Fourier transform of sampled and truncated signal $\kappa(jL), \ j \quad n \quad 1, \ldots, n + 1$

$$
\hat{K}(\log(\frac{y}{x})) = \sum_{j=-n}^{n} L\kappa(jL)e^{ijL\log(\frac{y}{x})}
$$

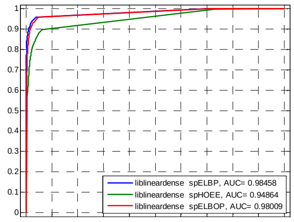
Caltech 101

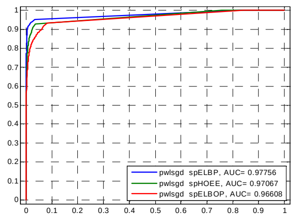
		linear kernel acc.	linear kernel time	Hellinger's kernel acc.	Hellinger's kernel time	χ^2 kernel acc.	χ^2 kernel time	inters. kernel acc.	inters. kernel time
		49.0±1.5	29.2±0.9	63.7±1.9	19.9±0.4	64.2±1.7	388.4±8.7	62.2±1.8	354.7±24.4
appr.	1					62.4±1.6	**20.7**±0.3	62.0±1.4	22.9±0.7
appr.	3					64.2±1.5	58.4±7.2	63.9±1.2	66.5±2.3
appr.	5					64.0±1.6	101.3±0.7	64.0±1.7	105.8±6.5
appr-γ	3					**65.8**±1.5	54.7±6.2	65.7±1.5	52.6±7.7
MB	1					–	–	55.9±0.9	26.9±0.8
MB	3					–	–	60.5±1.3	25.5±1.2
MB	5					–	–	61.3±1.1	22.1±3.3

Application to Robotvision@ICPR 2010

- For spHOEE, we choose $L = 4$ $r_x = r_y = \left[1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}\right]^T$ $d_x = d_y = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\right]^T$ $\sigma = 2$ $N_o = 12$ leading to $N_s = 540$ and $d = 6480$
- For spELBP, spELBOP, we choose $L = 3$ $r_x = r_y = d_x = d_y = \left[1, \frac{1}{2}, \frac{1}{4}\right]^T$ leading to $N_s = 42$ and $d = 10752$

Set|Classifier	ER	SRV	AUC
spHOEE ($d = 6480, \eta = 0.30$)			
Easy | *TRON*	0.328	1292.5	0.9251
Easy | *PWLSGD*	0.216	1724.5	0.9687
Hard | *TRON*	0.398	1025.5	0.8471
Hard | *PWLSGD*	0.371	1129.0	0.8394
spELBP ($d = 10752, \eta = 0.35$)			
Set|Classifier	*ER*	*SRV*	*AUC*
Easy | *TRON*	0.224	1691.5	0.9683
Easy | *PWLSGD*	**0.156**	**1952.5**	0.9783
Hard | *TRON*	**0.338**	**1255.0**	0.8855
Hard | *PWLSGD*	0.350	1211.5	0.8452
spELBOP ($d = 10752, \eta = 0.35$)			
Set|Classifier	*ER*	*SRV*	*AUC*
Easy | *TRON*	0.219	1711.0	0.9699
Easy | *PWLSGD*	0.215	1727.5	0.9621
Hard | *TRON*	0.388	1066.0	0.8416
Hard | *PWLSGD*	0.393	1043.5	0.8395





Application to Robotvision@ICPR 2010

- Integrate the dynamic of the robot by smoothing the raw labels sequence via a Forward-Backward algorithm.
- Corridor is connected to all other rooms. The state transition probabilities matrix A (10×10) is defined by:

$$a_{ij} = \Pr(y_k = i | y_{k-1} = j) = \begin{cases} 1-\lambda, & i = j \\ \lambda, & i = 2, j \neq i \\ \frac{\lambda}{9}, & i \neq j, j = 2 \\ 0, & else \end{cases}$$

where $\lambda = \frac{1}{\tau}$ and τ the mean sejour in current state.

Conditional measurements probabilities are proportional to SVM outputs

Late Fusion

ER	*Easy*	*Hard*	
Task 1	0.1477	0.2994	
Task 2	0.0619	0.2567	

SRV	*Easy*	*Hard*	*Easy+Hard*
Task 1	1985.5	1405.0	**3390.5**
Task 2	2314.0	1568.5	**3882.5**

Demo

References

[1] http://www.imageclef.org/2010/ICPR/RobotVision

[2] Fei-Fei, L., Fergus, R., and Perona, P. Learning Generative VisualModels from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. In CVPRW '04:

[3] Grin, G., Holub, A., and Perona, P. Caltech-256 Object Category, Dataset. Tech. Rep. 7694, California Institute of Technology, 2007

[4] Opelt, A., Fussenegger, M., and Auer, P. Generic Object Recognition with Boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(3):416{431, 2006

[5] http://pascallin.ecs.soton.ac.uk/challenges/VOC/

[6] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001

References

[7] Subhransu Maji and Alexander C. Berg and Jitendra Malik. Classification Using Intersection Kernel Support Vector Machines is efficient. CVPR 2008, Anchorage, Alaska.

[8] Lowe, David G,"Object recognition from local scale-invariant features". pp. 1150–1157, ICCV.1999

[9] Koen E. A. van de Sande, Theo Gevers, Cees G. M. Snoek, A Comparison of Color Features for Visual Concept Classification ACM International Conference on Image and Video Retrieval, page 141--150, 2008

[10] Koen E. A. van de Sande, Theo Gevers and Cees G. M. Snoek, Evaluating Color Descriptors for Object and Scene Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 32 (9), pages 1582-1596, 2010.

[11] Liao, S.C and all, Learning Multi-scale Block Local Binary Patterns for Face Recognition, ICB07(828-837).

References

[12] Heikkilä, M., Pietikäinen, M. and Schmid, C. (2009), Description of Interest Regions with Local Binary Patterns. Pattern Recognition 42(3):425-436

[13] Sanqiang Zhao and all, Performance Evaluation of Micropattern Representation on Gabor Features for Face Recognition, ICPR 2010

[14] Chao ZHU Multi-scale Color Local Binary Patterns for Visual Object Classes Recognition, ICPR 2010

[15] Zhang, Baochang and Gao, Yongsheng and Zhao, Sanqiang and Liu, Jianzhuang, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, IEEE Transactions on Image Processing, Vol 19, Feb 2010

[16] Shai Shalev-Shwartz, Yoram Singer, and Nathan Srebro. Pe gasos:Primal Estimated sub-GrAdient SOlver for SVM. In ICML, pages 807–814, New York, NY, USA, 2007

References

[17] Cho-Jui Hsieh, Kai-Wei Chang, Chih-Jen Lin, S. Sathiya Keerthi, and Sellamanickam Sundararajan. A dual coordinate descent method for large-scale linear SVM. In ICML,2008.

[18] Aditya Krishna Menon, Large-Scale Support Vector Machines: Algorithms and Theory. Aditya Krishna Menon. Research Exam, University of California, San Diego. 2009

[19] A. Vedaldi and A. Zisserman, Efficient Additive Kernels via Explicit Feature Maps, CVPR 2010

[20] Hervé Jégou, Matthijs Douze, Cordelia Schmid, Product quantization for nearest neighbor search, IEEE Transactions on Pattern Analysis & Machine Intelligence - 2010

[21] Jianchao Yang, Kai Yu, Yihong Gong, Thomas Huang, Linear Spatial Pyramid Matching using Sparse Coding for Image Classification (CVPR'09)

[22] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2010

[23] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman. Supervised Dictionary Learning. Advances Neural Information Processing Systems, 2008. Vancouver. Canada

[24] J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online Dictionary Learning for Sparse Coding. International Conference on Machine Learning, Montreal, Canada, 2009

// Hervé JEGOU
INRIA / IRISA
http://www.irisa.fr/texmex/people/jegou/

"Recherche d'Image à Grande Echelle : Procédés d'Aggrégation & d'Indexation"

Cet exposé sera axé sur la recherche d'image dans de très grandes bases d'images et de vidéos, pour lesquelles de nombreuses approches ont récemment été proposées tant du point de vue de la description que des stratégies d'indexation associées. Du point de vue de la description, nous nous intéresserons en particulier :

1) A la comparaison de techniques d'aggrégation vectorielles de descripteurs locaux, où des alternatives aux approches par sac-de-mots ont récemment émergées pour la recherche et la classification d'images,

2) Et aux techniques d'indexation récentes permettant d'indexer ces représentations.

Mots clefs : Méthode d'aggrégation vectorielle, Bag-of-words, Fisher Kernel et approximation, Recherche approximative, Locality-sensitive hashing, Méthodes basées codage de source, Recherche à grande échelle.

Références :

F. Perronnin and C. R. Dance. Fisher kernels on visual vocabularies for image categorization. CVPR 2007.

F. Perronnin, Y. Liu, J. Sanchez, and H. Poirier. Large-scale image retrieval with compressed Fisher vectors. CVPR 2010

Jegou, Douze, Schmid & Perez. Aggregating local descriptors into a compact representation, CVPR 2010.

Jegou, Douze & Schmid. Product quantization for nearest neighbor search. T. PAMI, to appear.

L. Paulevé, H. Jégou & Laurent Amsaleg. Locality sensitive hashing: a comparison of hash function types and querying mechanisms. Pattern Recognition Letters, August 2010

M. Muja and D. G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", VISAPP'09

Recherche d'image à grande échelle: procédés d'agrégation & d'indexation

École d'été ERMITES

2010

Hervé Jégou, INRIA

thanks to M. Douze, P. Pérez, F. Perronnin, H. Sandhawalia, C. Schmid

Image indexing: what is the problem we want to address?

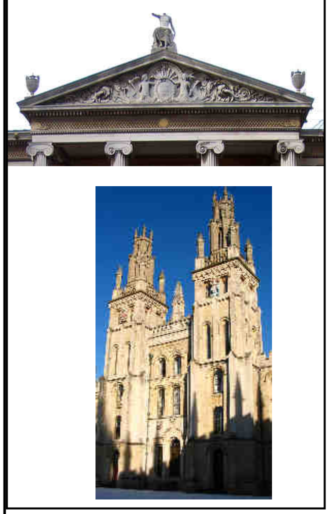
http://labs.ideeinc.com/multicolour/

Image search: what we might want to do

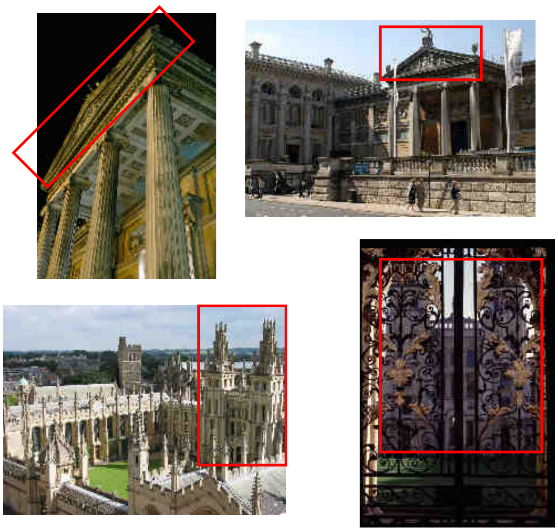
- The problem must be posed (=there is no obvious definition)
  - what are we looking for? Trees ? Locations? Plants that share 99% of DNA with the query ? Reproduce human interpretation?!
  - Machine learning approach: training set + evaluation set

Image search: what we might want to do

Queries (objects)



Sample images to be returned



http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html

Image search: proposed framework

- In the following : visual and geometrical similarity, no interpretation
  - restrictive setup, but adapted to certain application setup
  - ex: pirated image detection, place recognition

queries

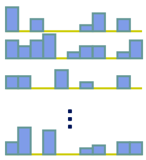
relevant answers

Image indexing: description step

input image

descriptors' extraction

vector(s) of features



- Image processing : analysis step (=description)
  - produce 1 or several descriptors (vectors) associated with the image
    ⇨ convey useful information in exploitable form
  - proximity between vectors: reflects visual similarity
- I won't insist on this description stage
  - see the talk of Sébastien Paris
- SIFT (scale invariant feature transform)
  - detection of regions of interest
  - spatial description of intensity gradients
  - invariance to geometrical and photometrical transformations

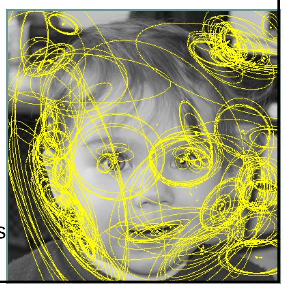
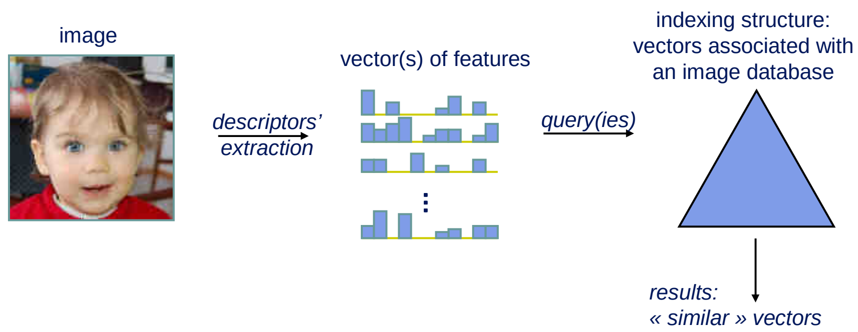
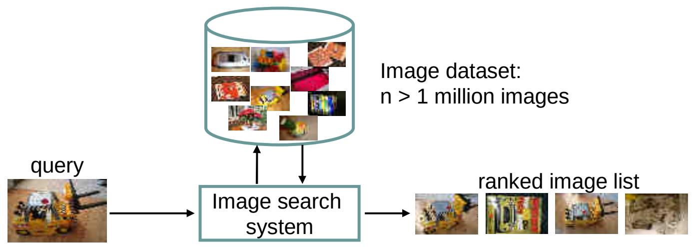


Image indexing: the search step



- Searching an image =
  - **finding quickly, in a large set of vectors, the ones that most resemble to the vectors associated with our query image**
  - for a given comparison metric between vectors (or between sets of vectors)

Large scale object/scene recognition: the complexity issue



- An image described by approximately m=2000 descriptors (dimension d=128)
  - $n=2.10^9$ descriptors to index!
- Database representation in RAM:
  - Raw size of descriptors : 1 TB, memory intractable
- Search: n_q * n * d elementary operations = 5. 1014
  - computationally not tractable

The complexity issue: Mimick text retrieval

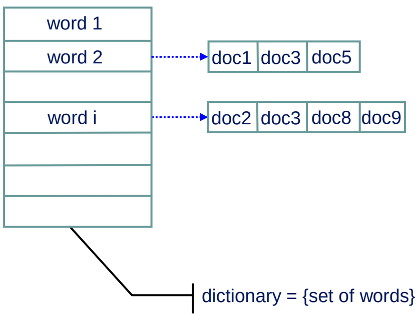
- Seminal work: the « Video-Google » approach, Sivic & Zisserman'03
- Key idea: aggregation of vectors

n local descriptor describing the image $\rightarrow$ 1 vector

- This is « the vector model »
  - sparse vectors ⇨ efficient comparison
  - inherits invariance of the local descriptors

Inverted file

- In database systems, this structure is a set of lists
  - which groups the entities that share the same value for a given attribute
  - in practice: is used for secondary keys (if limited set of possible attribute values)
- Typical application case : queries for textual documents (e-mails, …)



- The query consists in retrieving the list of documents containing the word
  - complexity = O(number of document to be returned)

Parenthesis: searching/comparing textual documents

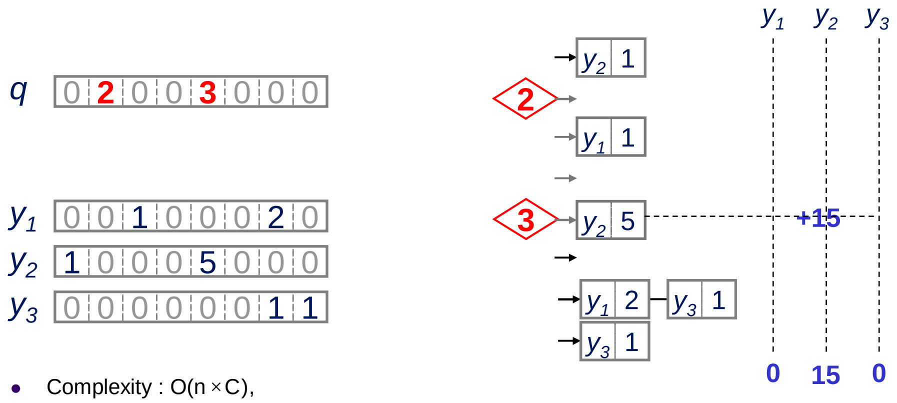
- Vector model
  - consider a dictionary with d words
  - a textual document is represented by a vector f=(f1,…,fi,…fd) $\in \mathbb{R}^d$
  - each component i corresponds to a dictionary word
  - f_i = word frequency within the document
  - Non discriminative words are removed (stopping lists)
    - i french, “le”, “la”, “est”, “a”, etc, are not considered
  - sparse vectors: dictionary size large compared to the number of words in a document

- Example (in french)
  - dictionary = {“vélo”, “voiture”, “déplace”, “travail”, “école”, “Rennes”}
  - representation space: $\mathbb{R}^6$
  - “Rennes est une belle ville. Je me déplace à vélo dans Rennes”
    ⇨ $f=(1/4,0,1/4,0,0,1/2)^t$

- comparing textual documents = comparing their vector representations

Back to our inverted file: distance between sparse vectors

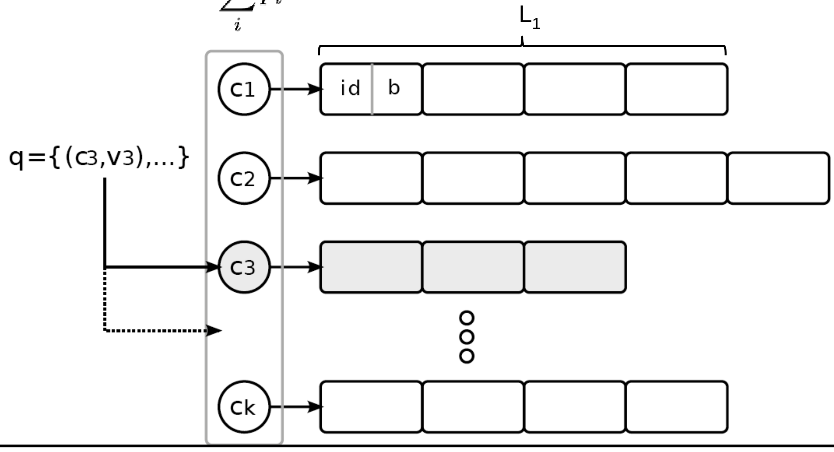
- query q and database $Y=\{y_1,y_2,\dots y_n\}$: sparse vectors
- Inverted file: efficient calculation of inner product (in fact, any L_p distance, Chi-2,etc)
- Example for the inner product:



- Complexity : $O(n \times C)$,
  - C : expectation of the number of shared non-null component positions
  - linear complexity does not mean inefficient!

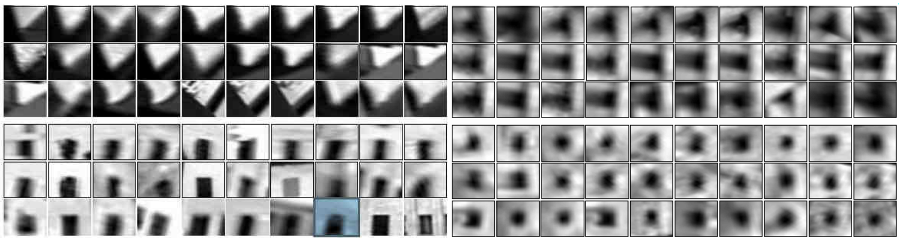
Comparing sparse vectors with an inverted file: complexity

- p_i = P(position i is non zero)
- Li = list of non zeros elements for component i
- Expectations of total shared non zero positions = $n\, n_q \sum_i {p_i}^2$
- Optimum reached if p_i=1/k
- Imbalance factor: $k \sum_i {p_i}^2$

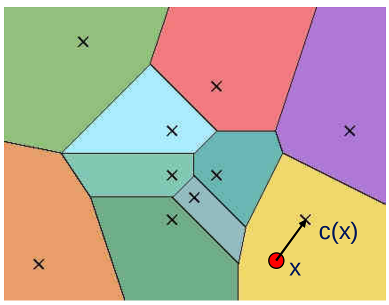


Video-Google : a textual approach to image search

- The goal: "put the images into words", visual words

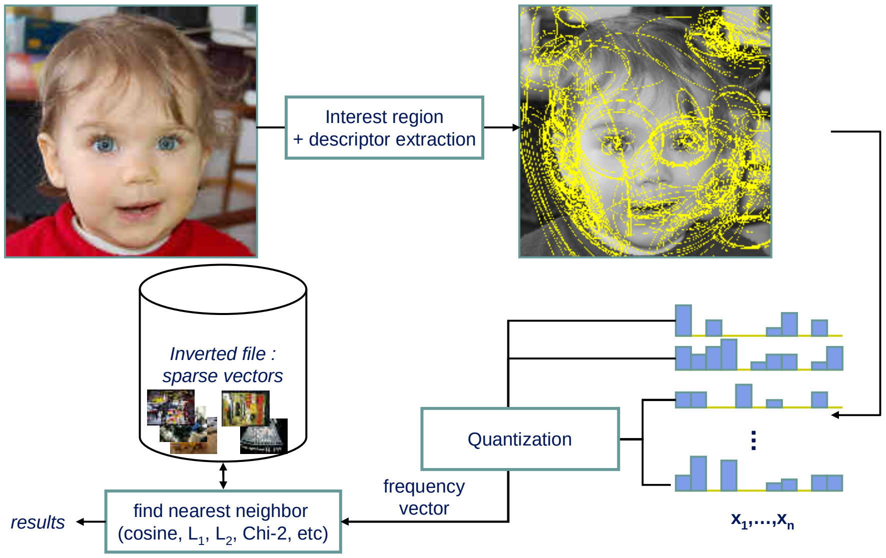


- Quantizer function: $\mathbb{R}^d \to \omega$
  $x \to c(x) \in \omega$, dictionary of "visual words"
- Most the state-of-the-art image search algorithms are built upon this approach
- *idf* (inverse document frequency) weights the discriminative power of the visual word



Vidéo-Google : mimicking textual retrieval systems

- vector model = vector of empirical frequency of visual word in an image



Complexity

- Frequency vectors (Bag-of-words) are sparse
  - on purpose! Unlike in text, we decide the vocabulary size

  → construction of a very large dictionary [Nister 06, Philbin 07]
  - yields higher efficiency when search performed with an inverted file

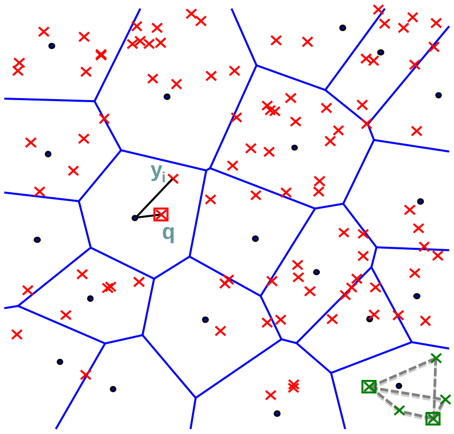
- Complexity: $n\, n_q \sum_i {p_i}^2$
  - linear w.r.t the number of images n
  - but small constant: typically 0.005 for a million-sized dictionary

- Memory usage
  - sparse vector representation: $\{(pos_i, val_i)\}_{i=1..nz}$
  - better choice in this context: $\{pos_i\}_{i=1..nz}$
  - 1 million images: ≈ 8 GB

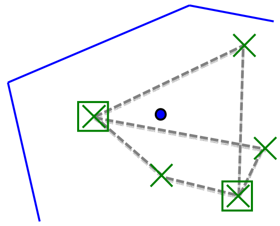
Bag-of-word : another interpretation

- « Visual words » are a view of mind
  - another interpretation: approximate nearest neighbor search+voting
  - or exact search for a particular definition of the neighborhood

$$N(q) = \{ y_i \text{ in } Y : c(y_i) = c(q) \}$$



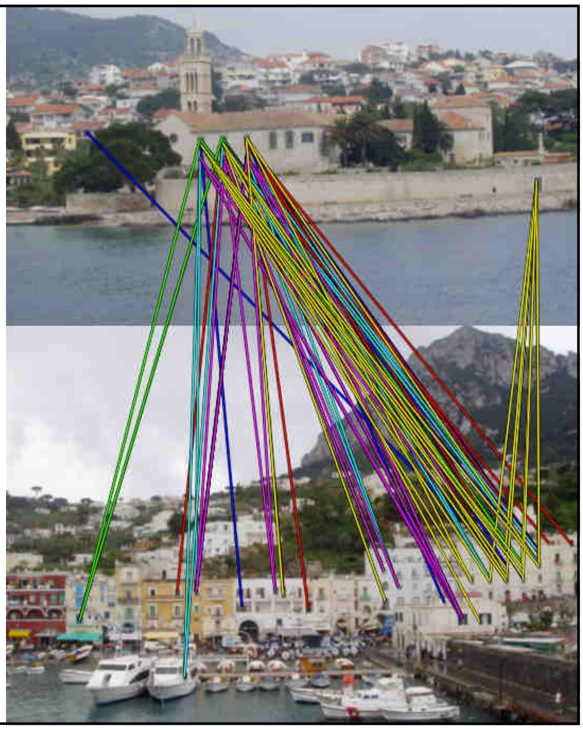
- increasing the dictionary size = reduce the cell size
- no disambiguation:



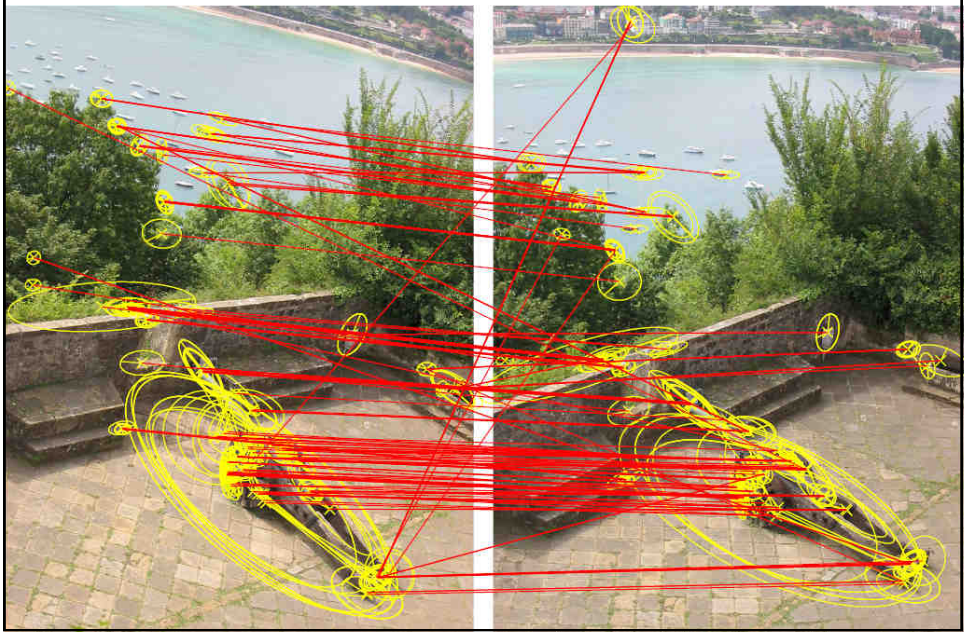
⇨ 6 votes for 2x3 descriptors
= inner product!

No disambiguation

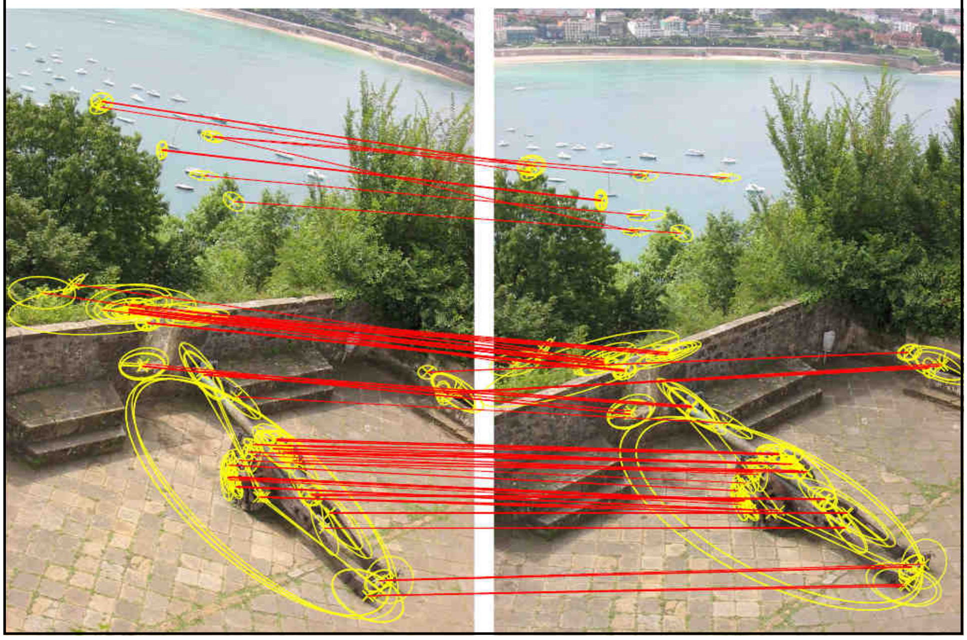
- Problem for the scoring



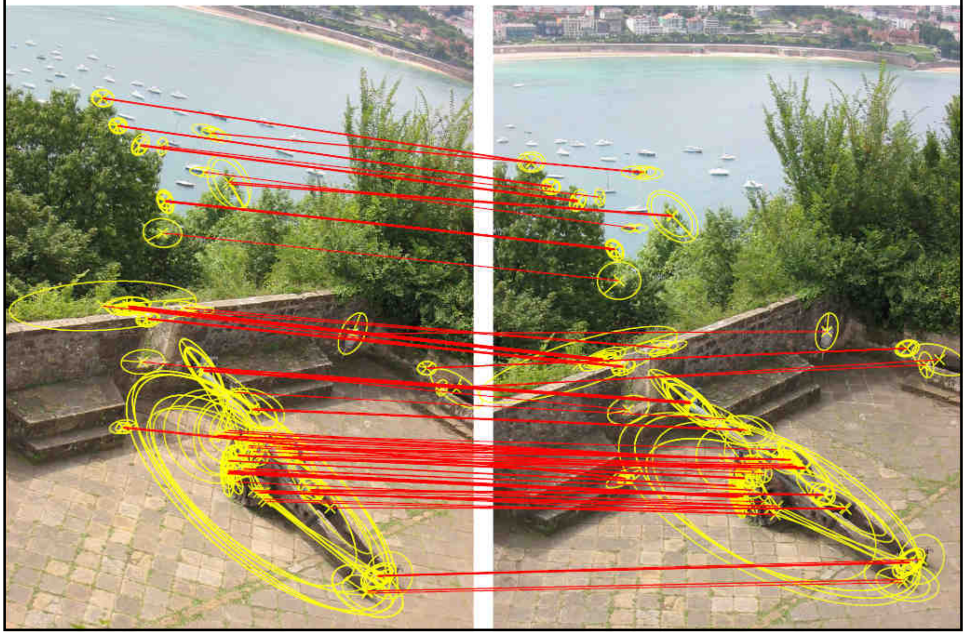
Compromise on vocabulary size: k=20000



Compromise on vocabulary size: k=200000



Hamming Embedding: ANN extension of bag-of-words

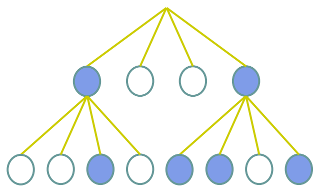


Extension to the bag-of-words approach

- Large vocabularies, **Hierarchical quantization** [Nister and Stewenius 2006]
- Soft Quantization [Philbin 2008, but used before]
- Post treatment: geometrical verification [Lowe 2004, Philbin 2007, …]
- **Integration of geometry within the index**
- Voting interpretation: refinement of descriptor comparison [Jegou 2008]
  → see the part on approximate nearest neighbor search
- Query expansion [Chum et al 2007]
- **Handling bursts** [Jegou 2009]
- …

Hierarchical quantization

- Vocabulary trees [Nister 06]
- Using large vocabularies means high quantization cost



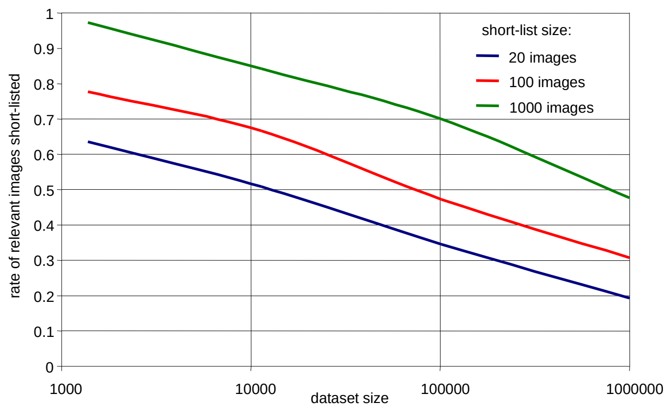
- Complexity of hierarchical quantization for k visual words and l levels

$$l \times k^{\frac{1}{l}}$$

- Better approach: quantification based on ANN [Philbin 07]
- We will come back on this point later

Geometrical verification

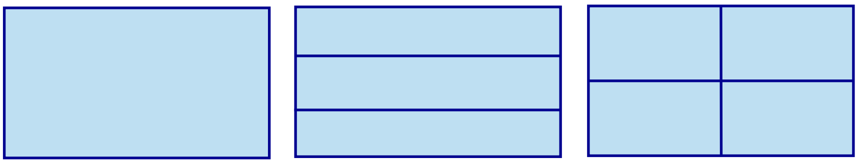
- Re-ranking based on full geometric verification [Lowe 04, Chum & al 2007]
  - works very well
  - but inefficient: performed on a short-list only (typically, 100 images)
  - → for very large datasets, the number of distracting images is so high that relevant images are not even short-listed!



The graal: geometric verification on a large scale

- Still a long way to go…

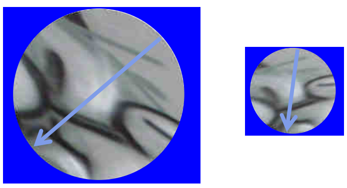
- Important activity on the topic
  - Geometrical Min-hash [Chum 09]
  - Bundling features [Wu 09]
  - Weak geometry consistency [Jegou 08]
  - Spatial inverted file [Lin 10]
  - …
- In general, these representations does not correspond to a vector model
  - not useable for classification
- except for the spatial pyramid method [Lazbenik 2006]
  - not invariant to translation, zoom, rotation, … → for classification only



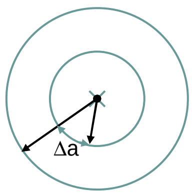
Weak geometry consistency [Jegou 2008]

Geometrical information used for **all** images (not only the short-list)

Each invariant interest region detection has a **scale** and rotation **angle** associated, here characteristic scale and dominant gradient orientation



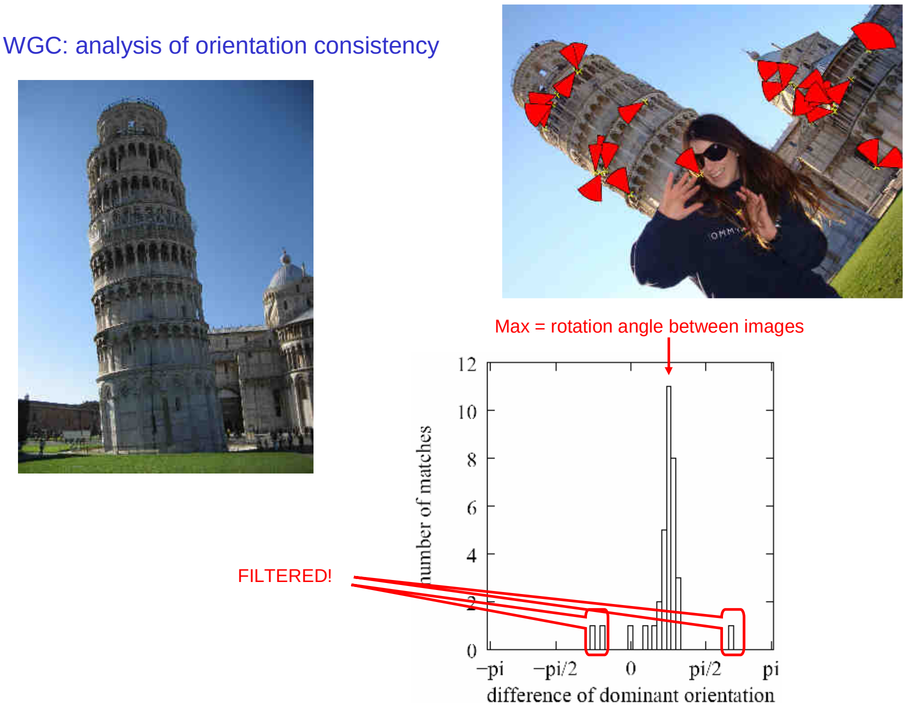
Scale change ≈ 2
Rotation: Δa ≈ 20 degrees

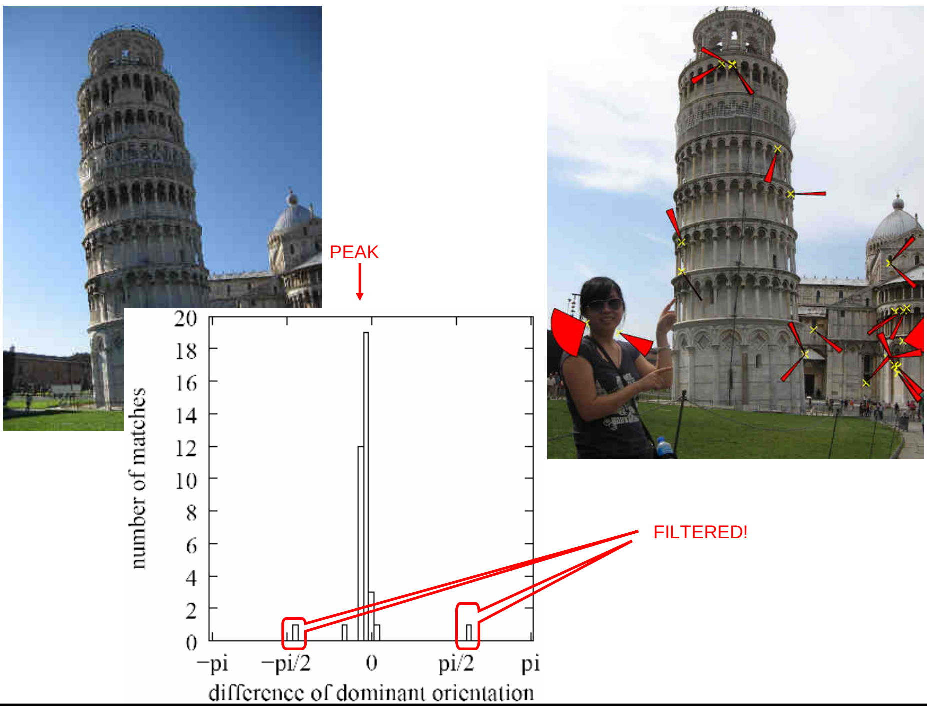


Each matching pair results in logscale and angle difference: Δs and Δa

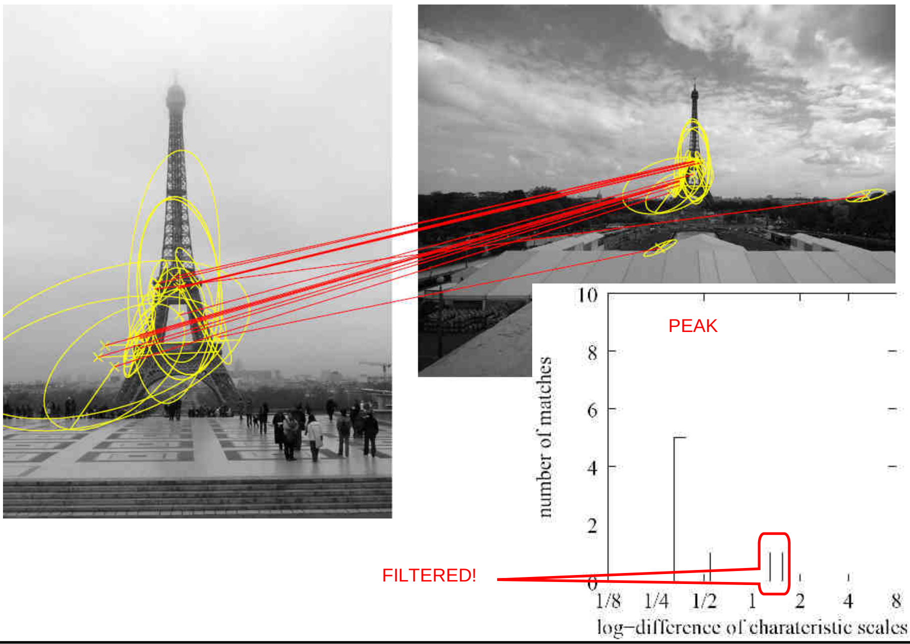
For the global image scale and rotation, changes are roughly consistent

WGC: analysis of orientation consistency

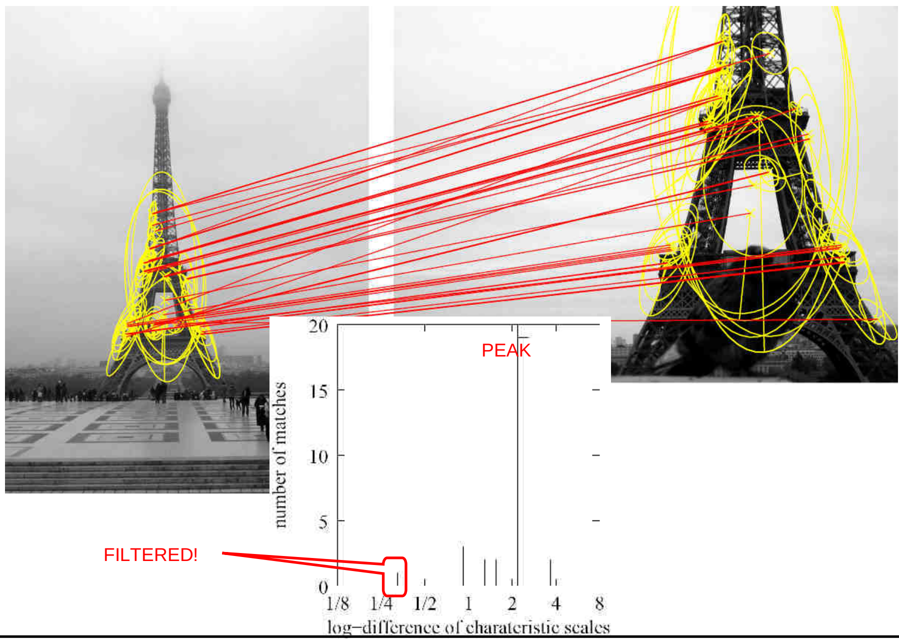




WGC: scale consistency

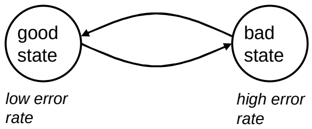


WGC: scale consistency



What is burstiness?

- In numerical communications, error bursts are taken into account by specific channel models, e.g., the Gilbert-Elliot channel:



- In text retrieval: if a word occurs in a document, then it is more likely to occur again than what you would expect from an independent model
  - see (for instance) [Church 95] and [Katz 96]

Example: “Oxford” is the 2772$^{\text{th}}$ most frequent word in english

Pr (word = “Oxford”) = 0.0000280204

Source: Gutenberg project, http://en.wiktionary.org/wiki/Wikitionary:Frequency_lists

Probability of “Oxford ” assuming an independent model ?

Welcome to Burlington House Hotel in Oxford

Nestling in leafy Oxford, a short bus ride away from the bustling City Centre, Burlington House offers high quality and peaceful luxury bed & breakfast accommodation to Oxford's many visitors.

Oxford Hotel Credentials

Burlington House has been consistently awarded the highest accommodation rating and is now Five Stars Highly Commended under the new AA scheme. Burlington House has been recommended by both the Which? Guide to Good Hotels and its sister publication, the Which? Guide to Good Bed & Breakfasts.

For more information on what you can expect from a five star establishment such as Burlington House, please visit our accreditations page.

History of Burlington House, Quality Oxford B&B,

Burlington House is a large, Victorian merchant's house, dating from 1889. Without disturbing its Victorian elegance, the house has been totally renovated, resulting in a stylish and luxurious bed and breakfast hotel with 12 rooms, a residents lounge, dining room, on-site car parking facilities and landscaped gardens.

Our Oxford Hotel location,

Burlington House is situated in Oxford's premier residential area of Summertown, Oxford, Oxfordshire - less than five minutes ride by car or bus to the centre of the city. Public transport is frequent and conveniently close to the hotel. Within walking distance is the Summertown Parade of Shops which has many quality restaurants, bars and retail outlets.

Independent model, 224 words: Pr(“Oxford” appears ≥ 8 times) ≈ **5.2 10^{-23}**

Source: http://burlington-house.co.uk/

Solution in text retrieval: burstiness

Handling burstiness has been shown to be important in text retrieval
→ this phenomenon was shown to disturb the document comparison performed by the cosine similarity when using the vector space model

The vector space model is the one used in large scale image retrieval !

- "Video-Google: …", [Josef and Andrew, ICCV'2003]

Now, the obvious question:

do images contain bursts ?

☺ Yes!

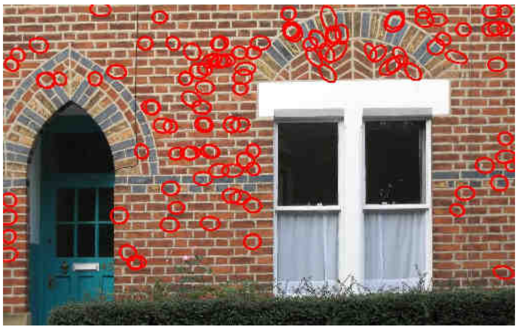
Many descriptors in an image are assigned to the same visual word

Bursts on a building image

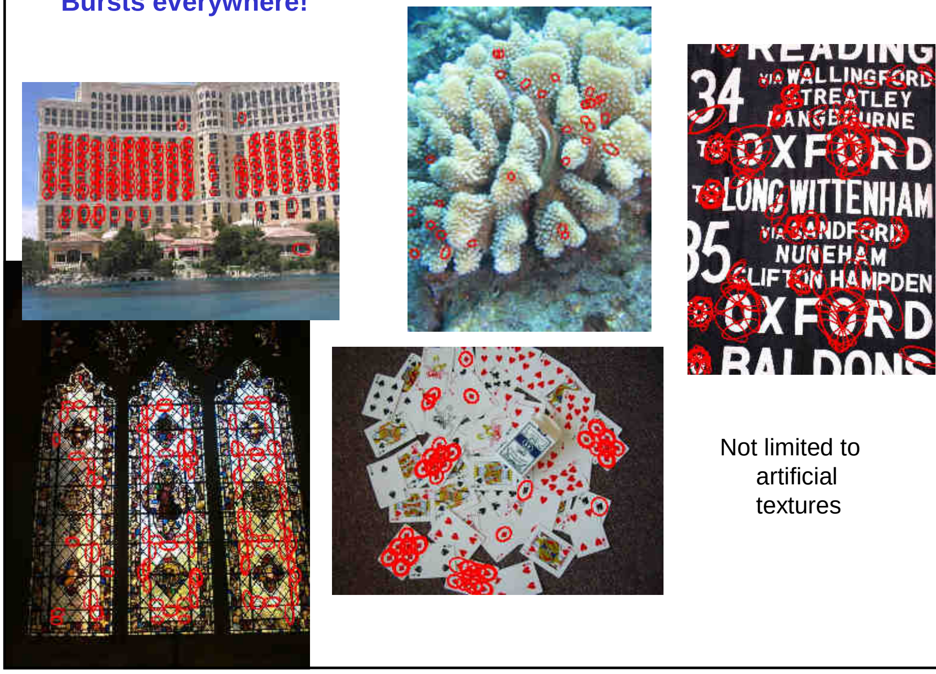
Most occurring visual word

Second most occurring visual word

Etc

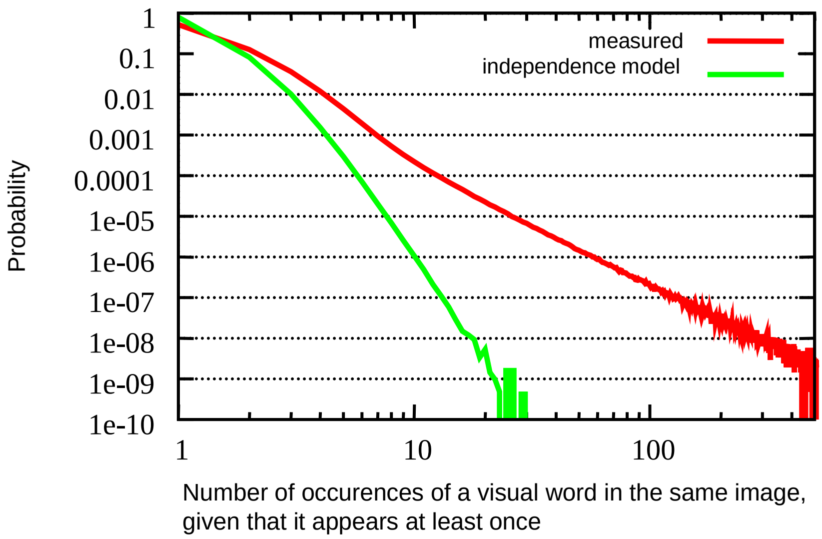


Bursts everywhere!



Not limited to artificial textures

Quantifying burstiness in images



So, what is the problem with bursts ?

corruption of the scoring measure

	Oxford (mAP)	Holidays (mAP)	Kentucky (score: x/4)
BOF	0.338	0.469	2.99
HE +WGC	0.542	0.751	3.36
HE +WGC +burstiness	0.596	0.807	3.47
HE +WGC + burstiness + Multiple assignment	0.647	0.839	3.54
HE +WGC +burstiness +MA +SP (Oxford: +QE)	0.747	0.848	3.64
State-of-the-art	Soft +Query exp. +SP 0.718 [Philbin & al, 08]	HE+WGC 0.751 [Jegou & al, 08]	CDM (scalable) 3.68 (3.40) [Jegou & al, 07, 09]

Bag-of-words: the ultimate solution?

- Interesting for two tasks
  - large scale indexing: efficient search inherited from inverted file
  - classification: vector model → useable with strong classifiers, in particular SVM
- A practical solution: mimick text → the same recipes/ingredients can be used
  - stopping words, query expansion, handling burstiness, etc
- However, imprecise approximation of a set of descriptors
- Concurrent approaches: non vector methods
  - using approximate nearest neighbor search
    → see later
- « Emerging » aggregation method: Fisher kernels
  - excellent results in international competitions (PASCAL VOC, Imagenet,…)
  - simplified version: VLAD

VLAD : Vector of Locally Aggregated Descriptors

- Learning: *k*-means
  - output: *k* centroids : $c_1,\ldots,c_i,\ldots c_k$
- VLAD computation:

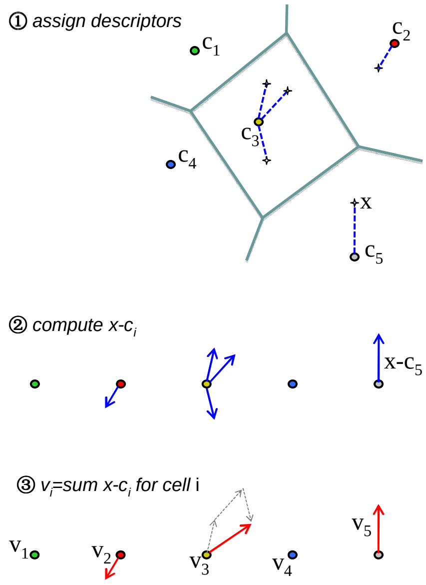
① ▸ $c(x) = \arg\min_{c_i} ||c_i - x||^2$

②③ ▸ $v_i = \sum_{x:c(x)=c_i} x - c_i$

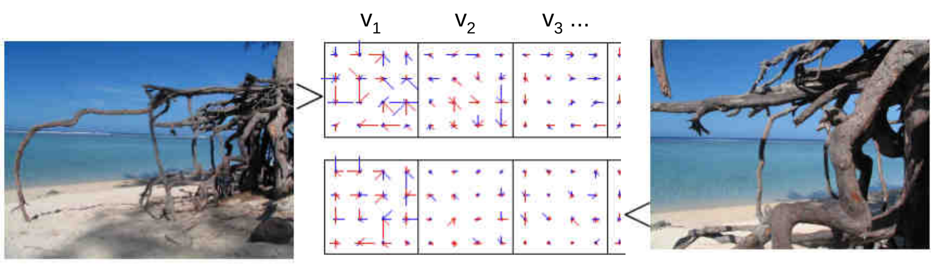
▸ $v = [v_1, \ldots, v_i, \ldots, v_k]\ ,\ v_i \in \mathbb{R}^{128}$

□dimension $D = k*128$

- L2-normalized
- Typical parameter: k=64 (D=8192)



VLADs for corresponding images



SIFT-like representation per centroid (+ components: blue, - components: red)

- good coincidence of energy & orientations

VLAD performance and dimensionality reduction

- We compare VLAD descriptors with BoF: INRIA Holidays Dataset (mAP,%)
- Dimension is reduced to from D to D' dimensions with PCA

Aggregator	k	D	D'=D (no reduction)	D'=128	D'=64
BoF	1,000	1,000	41.4	44.4	43.4
BoF	20,000	20,000	44.6	45.2	44.5
BoF	200,000	200,000	54.9	43.2	41.6
VLAD	16	2,048	49.6	49.5	**49.4**
VLAD	64	8,192	52.6	**51.0**	47.7
VLAD	256	32,768	**57.5**	50.8	47.6

- Observations:
  - VLAD better than BoF for a given descriptor size
    → comparable to Fisher kernels for these operating points
  - Choose a small D if output dimension D' is small

The Fisher kernel principle [Jaakkola & Haussler 99]

- Exploiting generative models in discriminative classifiers
- Feature vector is derivative w.r.t. probabilistic model

$$\boxed{G_\lambda^X = \nabla_\lambda \log p(X|\lambda)}$$

- Measure the similarity using the "Fisher kernel":

$$\boxed{K(X,Y) = G_\lambda^{X\prime} F_\lambda^{-1} G_\lambda^Y}$$

where F_λ is the Fisher information matrix

$$F_\lambda = E_{x\sim p}\left[\nabla_\lambda \log p(x|\lambda) \nabla_\lambda \log p(x|\lambda)'\right]$$

- Equivalently, perform a dot product on "normalized" Fisher vectors:

$$\boxed{\mathcal{G}_\lambda^X = L_\lambda G_\lambda^X} \quad \text{with} \quad F_\lambda = L_\lambda' L_\lambda$$

For classification: classifiers learned on these normalized fisher vectors

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Representing images with Fisher vectors [Perronnin & Dance 07]

- First introduced for image categorization
- set of D-dimensional local local descriptors $X = \{x_t, t = 1 \dots T\}$
- Gaussian Mixture Model $p(x) = \sum_{i=1}^{N} w_i p_i(x)$

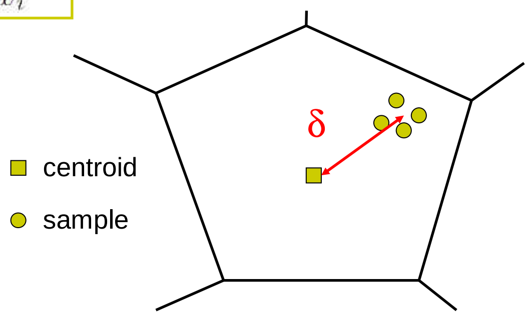
  with parameters $\lambda = \{w_i, \mu_i, \Sigma_i, i = 1 \dots N\}$
  trained on a large set of local descriptors: visual vocabulary
- Gradient with respect to mean vector: $\boxed{\mathcal{G}_i^X = \frac{1}{\sqrt{w_i}} \sum_{t=1}^{T} \gamma_t(i) \left(\frac{x_t - \mu_i}{\sigma_i} \right)}$

  with $\gamma_t(i) = \frac{w_i p_i(x_t)}{\sum_{j=1}^{N} w_j p_j(x_t)}$
- Concatenation of N gradient vectors ⇨ ND-dimensional vector

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Fisher vectors: other parameters can be taken into account

- We can rewrite: $\boxed{\frac{1}{T}\mathcal{G}_i^X = \frac{w_i^X}{\sqrt{w_i}} \delta_i^X}$
- "BOV part":
  $$w_i^X = \frac{1}{T} \sum_{t=1}^{T} \gamma_t(i)$$
- "delta part":
  $$\delta_i^X = \frac{\mu_i^X - \mu_i}{\sigma_i}$$
  $$\mu_i^X = \frac{\sum_{t=1}^{T} \gamma_t(i) x_t}{\sum_{t=1}^{T} \gamma_t(i)}$$



- Similarity between two images proportional to:
  $$\boxed{\sum_{i=1}^{N} \frac{w_i^X w_i^Y}{w_i} \delta_i^{X\prime} \delta_i^Y}$$

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Comparison between Bag-of-words and Fisher vectors

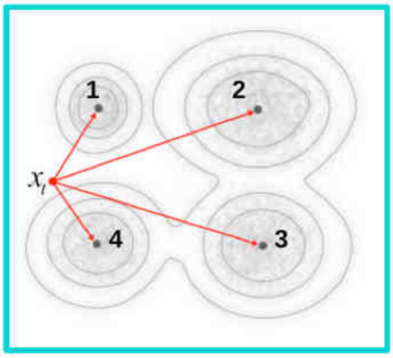
BOV

Hard Assignment

[0 0 0 1]

Soft Assignment

[.3 .1 .1 .5]



Fisher Vector

Gradient wrt w

[.15 -.2 -.35 .2]

Gradient wrt mean

[.8 -1.5 -3.7 -1.3 -3.8 1.2 -.9 1.4]

Gradient wrt var

[-1.2 -.9 1.4 -.8 1.5 -3.7 1.3 -3.8]

BOV Histogram has size: K
Fisher Vector (wrt to mean and var): 2 * D * K

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Properties of the Fisher vector

- Hard assignment assumption: $\gamma_t(i) \approx 1$

- "Background" (i.e. frequent) descriptors: large $p(x_t) \approx w_i p_i(x_t)$
- large w_i value: corresponds to frequent visual word
- large $p_i(x_t)$ value, i.e. small $||\frac{x_t - \mu_i}{\sigma_i}||^2$ value

- Frequent descriptors are automatically discounted:
  ⇨ extension of tf-idf to continuous variables
  ⇨ an image is described by what makes it different from others

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

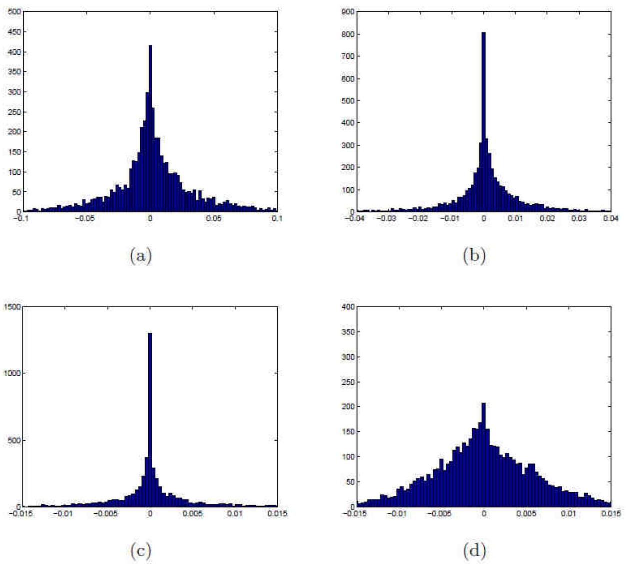
Fisher vector normalization

- L2 normalization:
  - ensures that the query always come first
  - normalizes with respect to amount of background
- Power normalization: $z \rightarrow \text{sign}(z)\ |z|\alpha$
  - “unsparsifies” sparse vector
  - Fisher vector can be viewed as a sample randomly drawn from a compound Poisson distribution ⇨ variance stabilization

cf F. Perronnin, Y. Liu, J. Sanchez and H. Poirier, “Large-Scale Image Retrieval with Compressed Fisher Vectors, CVPR 2010.

Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Improving the Fisher Vector



Distribution of Fisher vector values (first dimension) as estimated on VOC 07.

(a) 1 Gaussian

(b) 16 Gaussians

(c) 256 Gaussians

(d) 256 Gaussians with pow. norm.

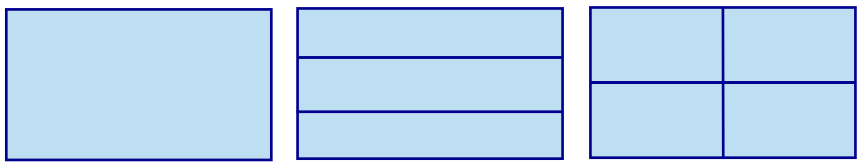
Slide by courtesy of Florent Perronnin (XRCE Grenoble)

Other extensions of the Fisher kernel representation

- Explicit embedding
  - Objective: using better comparison kernels
  - but keep linear SVM (or distance L2 for indexing purpose)
  - Idea: explicit approximation of the implicit space (infinite dimensionality)
  - See Perronnin'10

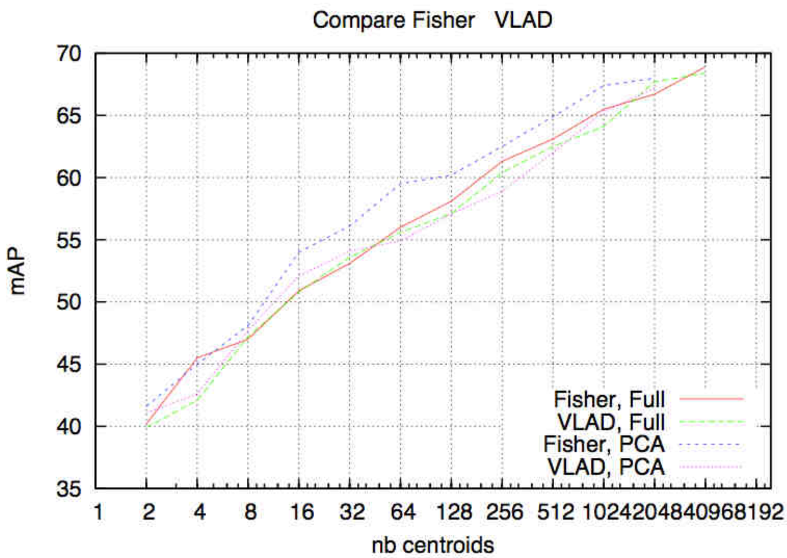
- Spatial pyramid for categorization (similar to BOF ones)
  - Account for the rough geometry of a scene:
    - extract one Fisher vector per region
    - concatenate Lp-normalized vectors

- Use partition of the winning VOC 2008 system:



Performance as a function of the descriptor size

Fisher need prior local PCA (pb with GMM estimation?)



Comparison of aggregation techniques for image search

- The goal here: optimizing the trade-off between
  - search quality
  - search speed
  - memory usage
- Approach: joint optimization of three stages
  - local descriptor aggregation
  - dimension reduction
  - indexing algorithm

$$\xrightarrow[\text{SIFT}]{\text{extract}} \underbrace{\begin{bmatrix}5\\0\\\vdots\\35\end{bmatrix}\begin{bmatrix}34\\6\\\vdots\\12\end{bmatrix}\cdots\begin{bmatrix}14\\20\\\vdots\\47\end{bmatrix}}_{n \text{ SIFTs (128 dim)}} \xrightarrow[\text{descriptors}]{\text{aggregate}} \underbrace{\begin{bmatrix}-0.234\\0.452\\0.134\\\vdots\\0.001\end{bmatrix}}_{D} \xrightarrow[\text{reduction}]{\text{dimension}} \underbrace{\begin{bmatrix}0.155\\0.230\\\vdots\\0.435\end{bmatrix}}_{D'} \xrightarrow[\text{encoding /indexing}]{\text{vector}} \text{code}$$

BOF, VLAD, Fisher [Perronnin 2010]

- Dimension reduced to from D to D' dimensions with PCA

Descriptor	k	D	Holidays (mAP)					
			$D'=D$	$\rightarrow D'$=2048	$\rightarrow D'$=512	$\rightarrow D'$=128	$\rightarrow D'$=64	$\rightarrow D'$=32
BOF	1 000	1 000	40.1		43.5	44.4	43.4	40.8
	20 000	20 000	43.7	41.8	44.9	45.2	44.4	41.8
Fisher (μ)	16	1 024	54.0		54.6	52.3	49.9	46.6
	64	4 096	59.5	60.7	61.0	56.5	52.0	48.0
	256	16 384	62.5	62.6	57.0	53.8	50.6	48.6
VLAD	16	1 024	52.0		52.7	52.6	50.5	47.7
	64	4 096	55.6	57.6	59.8	55.7	52.3	48.4
	256	16 384	58.7	62.1	56.7	54.2	51.3	48.1

- Observations:
  - VLAD better than BoF for a given descriptor size
    → comparable to Fisher kernels for these operating points
  - Choose a small D if output dimension D' is small

A few additional results

- see LSVRC'2010 challenge (Imagenet dataset)

Approximate nearest neighbor search

Approximate nearest neighbor search: old but active problem

- several community involved
  - algorithmic geometry
  - database
  - data compression
  - image processing
  - computer vision
  - …

Preliminaries

- vector dataset: $Y = \{ y_i \in \mathbb{R}^d \}_{i=1..n}$

- query vector: q in $\mathbb{R}^d$

- neighbors of q: N(q)□Y
  → "similar" vectors

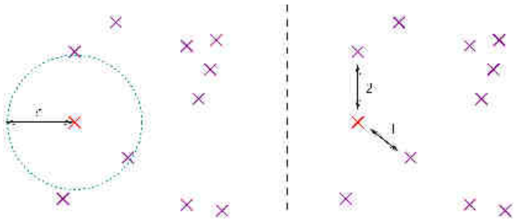
- The neighborhood is induced by a distance (or similarity/dissimilarity)

→ focus on the Euclidean case

$$d(x,y) = ||x-y||_2$$

But still several definitions

- ε-neighborhood
- k-neighborhood
- distance ratio criterion:
  ε-neighborhood,
  with $\varepsilon = \alpha\, d(q,NN(q))$



Naïve approach

- if Y is not organized, (=not indexed),
  - computing N_k ou N_ε requires the set of distances $d(q,y_i)$
  - $O(n \times d)$

- For several query vectors
  - Q set of query $[q_1 \dots q_m]$ and database $B=[b_1 \dots b_n]$
  - Inner product trick:

$$d(a-b)^2 = ||a||^2 + ||b||^2 - 2\langle a|b\rangle$$

  - $||a||^2$ and $||b||^2$ pre-computed: $O(n+m)$
  - The set of inner products $\langle a|b\rangle$ computed with BLAS3 operations as

$$Q' \times B$$

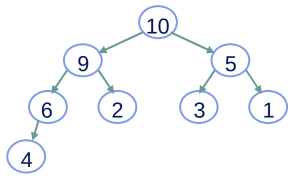
  - typical efficiency improvement of this "naïve" implementation: x10
  - for very highly dimensional, competitive efficiency

Binary heap

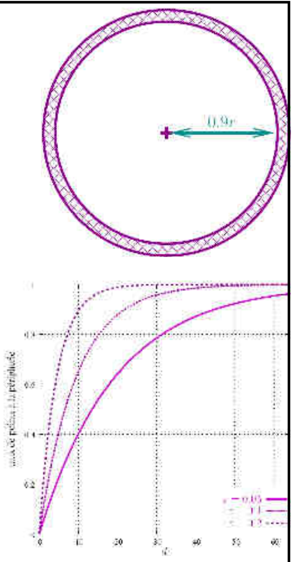
- For $N_k(q)$ neighborhood, we need k-arg-min$_i$ $d(q,y_i)$

- Example: we want 3-argmin {1,3,9,4,6,2,10,5}
  - Naive method 1: sort $\rightarrow O(n \log n)$. BERK.
  - Naïve method 2: maintain the set of k smallest values in a table $\rightarrow O(n\ k)$

- Better: binary MinHeap (see Knuth)
  - Complexity : O(n log k) in the worst case
  - much better in practice (when k << n)
  - O(n) in practice



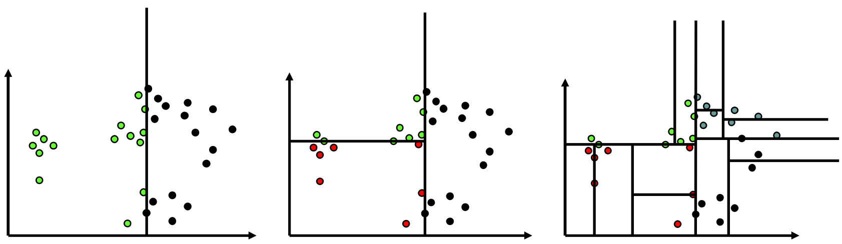
Indexing in high dimensional spaces



- For d $\geq$ 2, no total order
  - no logarithmic search?
- Need to organize the space: indexing
  - Kd-tree
  - Locality Sensitive Hashing
  - Spectral Hashing
  - FLANN
  - Searching with quantization
- Before going further:
  - indexing highly dimensional space is difficult
    - void space phenomenon
    - vanishing variance, hubs, …
  - the naïve approach is the best for exact search (for d>10-20)
  - focus on *approximate* nearest neighbor search

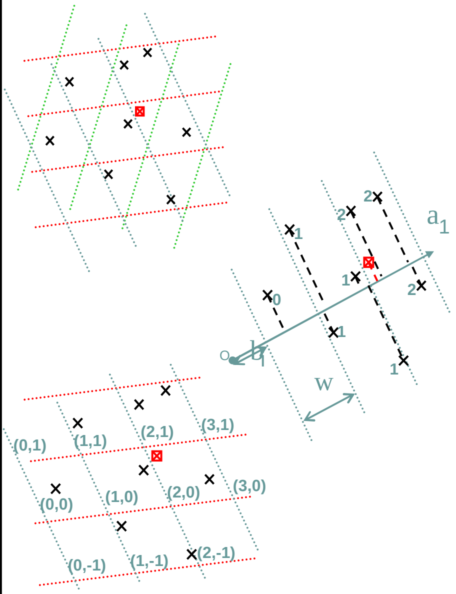
A (old) popular high-dimensional indexing scheme: Kd-Tree

- Partitioning technique
  - space recursively cut using hyperplans (on natural axis)
  - cutting position = median value of the considered axis
  - choice of next axis: highest residual variance



- Two ways to use it
  - exact search (not efficient in practice in high dimensional spaces)
  - approximate search

Euclidean Locality Sensitive Hashing (E2LSH)



1) Projection on *m* random directions

$$h_i^r(x) = \frac{\langle x|a_i\rangle - b_i}{w}$$

$$h_i(x) = \lfloor h_i^r(x) \rfloor$$

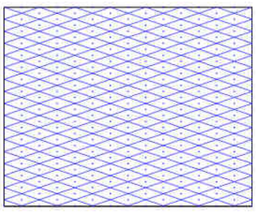
2) Construction of l hash functions: concatenate *k* indexes h_i per hash function

$$g_j(x) = (h_{j;1}(x); \dots; h_{j;k(x)})$$

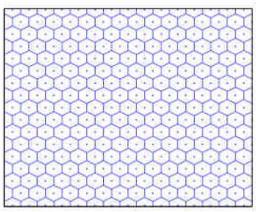
- 3) For each g_j, compute two hash values
  - universal hash functions: $u_1(.)$, $u_2(.)$
  - store the vector *id* in a hash table

Impact of the hash functions [

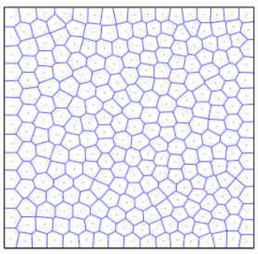
- LSH is theoretically nice, but hash functions not adapted to data distribution



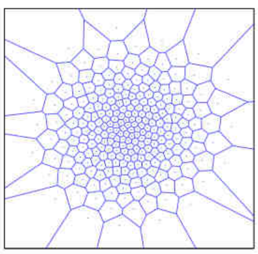
(a) Random projections



(b) A_2 lattice

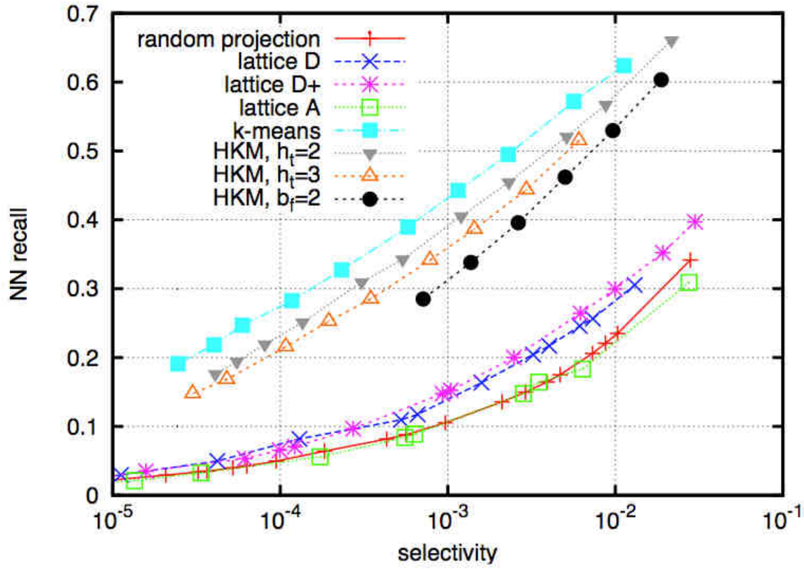


(c) k-means
Uniform distribution



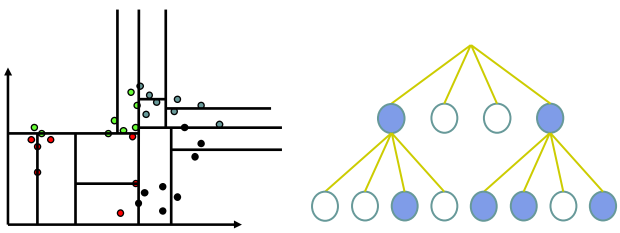
(d) k-means
Gaussian distribution

Impact of hash function



FLANN [Muja 2009]

- « Fast Library for Approximate Nearest Neighbors »
- Key contribution: automatic tuning (by cross-validation)
  - of parameters
  - including of partitioning function
- Two types of partitioning: KD-tree and hierarchical k-means
- Use multiple (randomized) partition
  - in the spirit of LSH

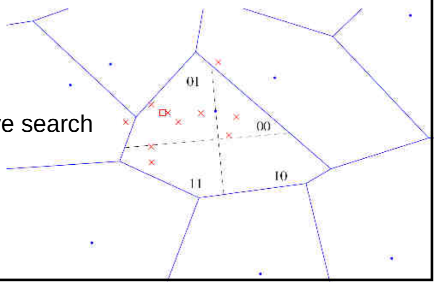


Hamming Embedding [Jegou 08] & Spectral hashing [Weiss 08]

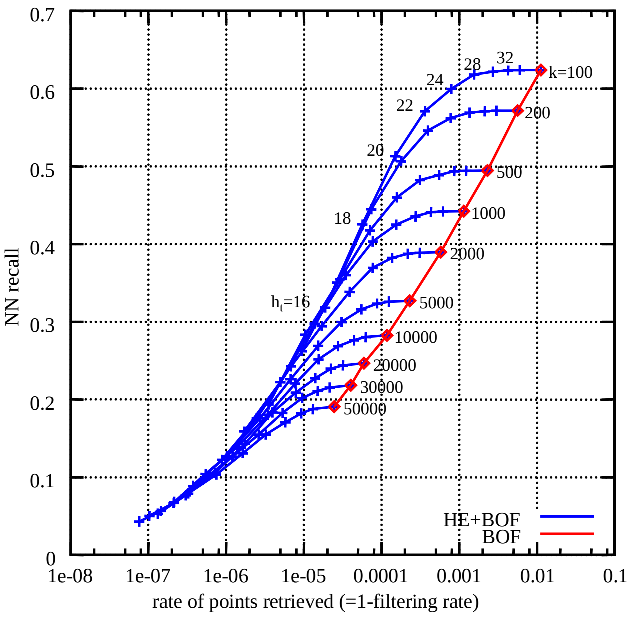
- Idea: $\mathbb{R}^d \rightarrow \{0,1\}^{d'}$
- Why: computing Hamming distances is (very) efficient
- Goal: finding an embedding function such that the Hamming neighbor of an embedded point reflects the Euclidean neighborhood of the original vector
- Spectral Hashing
  - find encoding which minimizes sum of Hamming distances between binary codes weighted by Gaussian kernel between original vectors

  ⇨ learning (PCA), tune number of bits
- Hamming Embedding [Jegou 08]
  - coarse quantization for non exhaustive search

  ⇨ use an inverted file
  - local binary signature (LSH-like) for improved comparison



ANN evaluation of Hamming Embedding



compared to BOF: at least 10 times less points in the short-list for the same level of accuracy

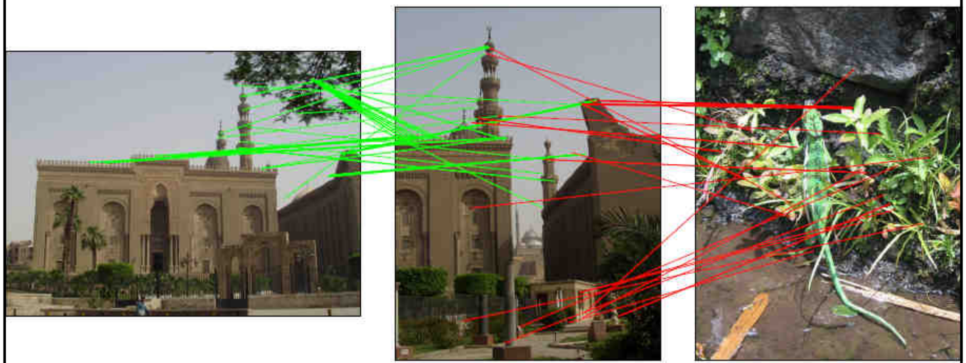
Hamming Embedding provides a much better trade-off between recall and ambiguity removal

Hamming Embedding vs BOF. Impact on matches.

BOF

69 votes

35 votes

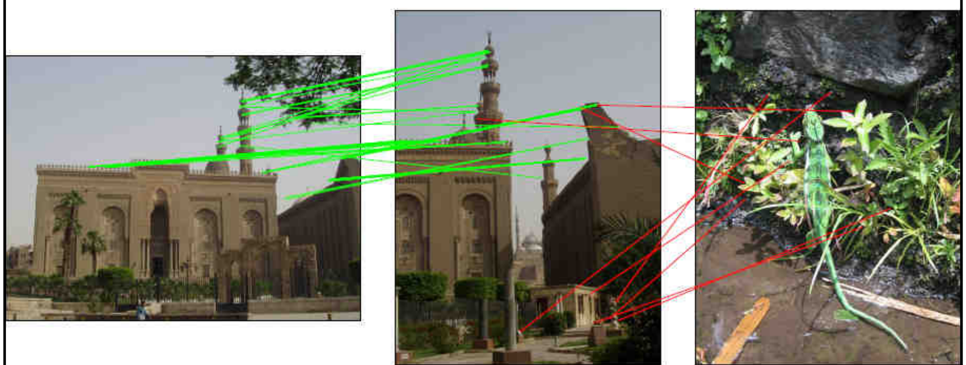


Hamming Embedding vs BOF. Impact on matches.

HE

83 votes

8 votes



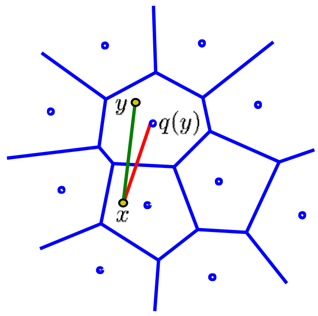
Searching with quantization [Jegou et al. 10]

- Search/Indexing = distance approximation problem
- The distance between a query vector *x* and a database vector y is estimated by

  $$d(x, y) \approx d(x, q(y))$$

  where q(.) is a quantizer

→ vector-to-code distance



- The choice of the quantizer is critical
  - needs many centroids
  - regular k-means and approximate k-means can not be used
    → we typically want k=2^{64} for 64-bit codes

Product quantization for nearest neighbor search

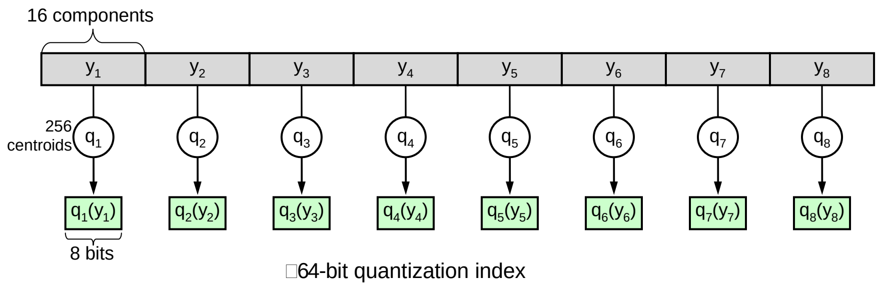
- Vector split into *m* subvectors: $y \rightarrow [y_1|\dots|y_m]$

- Subvectors are quantized separately by quantizers

  $$q(y) = [q_1(y_1)|\dots|q_m(y_m)]$$

  where each q_i is learned by *k*-means with a limited number of centroids

- Example: y = 128-dim vector split in 8 subvectors of dimension 16



Product quantizer: asymmetric distance computation (ADC)

- Compute the square distance approximation in the compressed domain

$$d(x,y)^2 \approx \sum_{i=1}^{m} d(x_i, q_i(y_i))^2$$

- To compute distance between query x and many codes
  - compute $d(x_i, c_{i,j})^2$ for each subvector x_i and all possible centroids

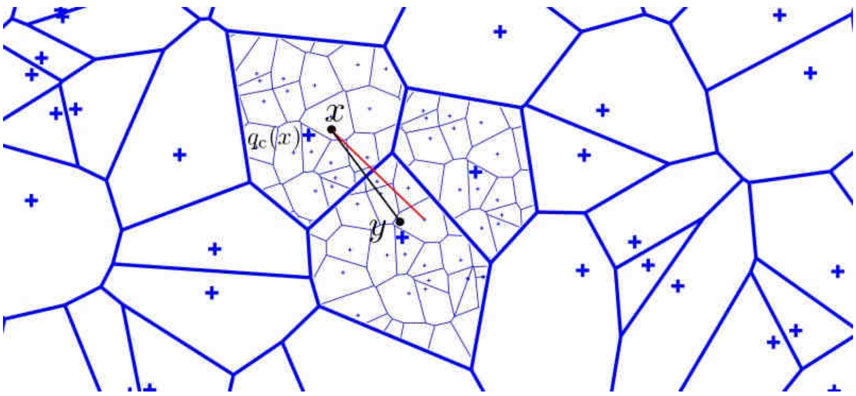
→ stored in look-up tables

  - for each database code: sum the elementary square distances

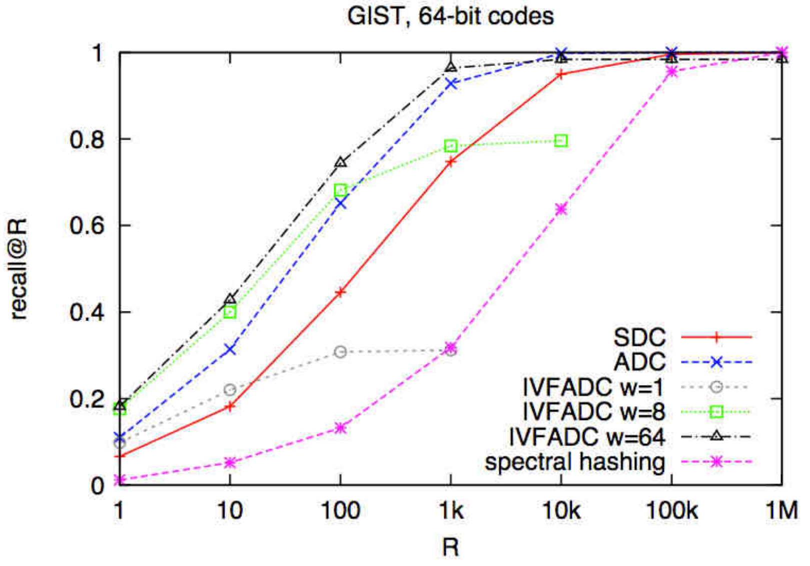
- Each 8x8=64-bits code requires only **m=8 additions per distance**!

- IVFADC: combination with an inverted file to avoid exhaustive search

Combination with an inverted file



Searching with quantization: comparison with spectral Hashing



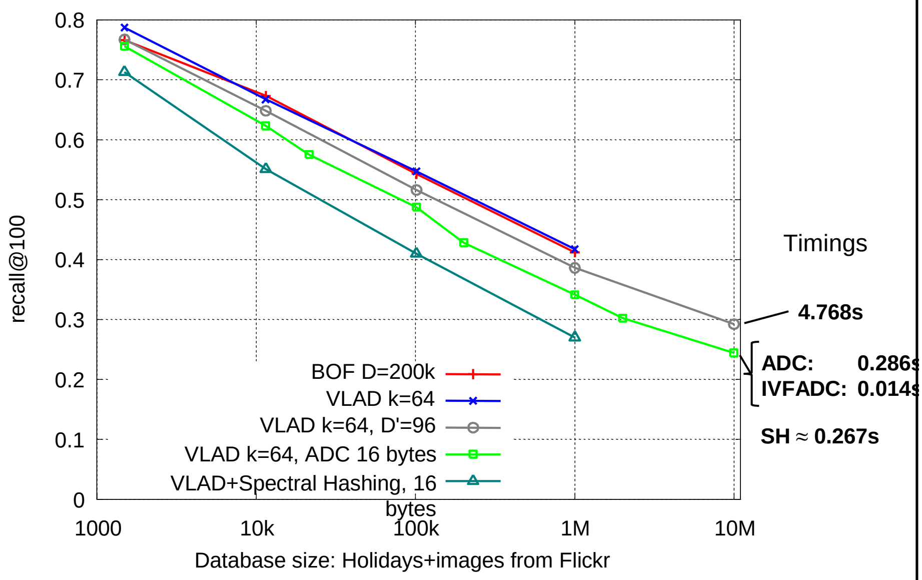
Results on standard datasets

- Datasets
  - University of Kentucky benchmark — score: nb relevant images, max: 4
  - INRIA Holidays dataset — score: mAP (%)

Method	bytes	UKB	Holidays
BoF, k=20,000	10K	2.92	44.6
BoF, k=200,000	12K	3.06	54.9
miniBOF	20	2.07	25.5
miniBOF	160	2.72	40.3
VLAD k=16, ADC	**16**	**2.88**	**46.0**
VLAD k=64, ADC	**40**	**3.10**	**49.5**

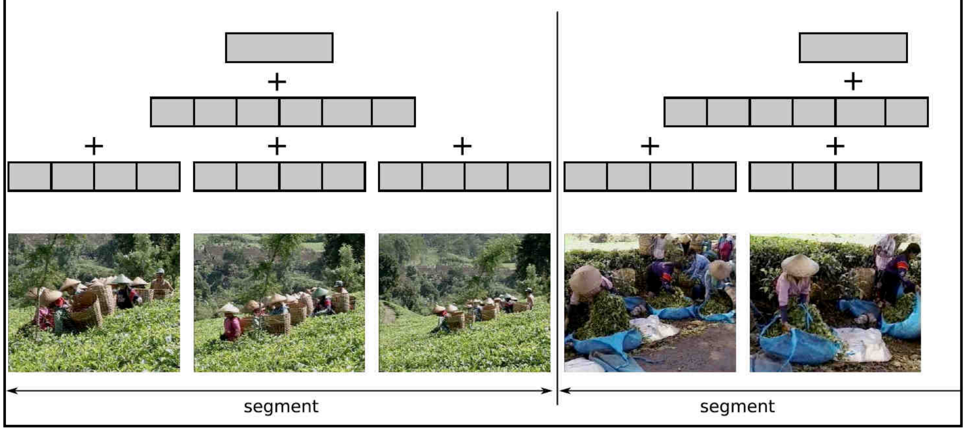
miniBOF: "Packing Bag-of-Features", ICCV'09

Large scale experiments (10 million images)



Extension to videos

- Specificity of video: temporal redundancy to be exploited
  - in the spirit of compression-based approaches
- Hierarchical coding of frame vector representation
  - segment representation + refinement
  - frame representation (=segment representation+refinement)



A few results

structure/algorithm	total mem	C_{mem}	C_{dist}	AP	time
none/brute-force search	62 GB	62 GB	19 M	96.7	89 ×
Levels 1+2, $r = 1$	122 MB	25 MB	18432	90.1	2.52 ×
Levels 1+2, no refinement	12 MB	2.71 MB	18432	74.1	1.10 ×
Levels 1+2+3 ($m_{\mathrm{r}} = 16$)	43 MB	2.73 MB	19710	91.2	1.33 ×
Levels 1+2+3 ($m_{\mathrm{r}} = 32$)	73 MB	2.75 MB	19710	90.5	1.42 ×
Levels 1+2+3 ($m_{\mathrm{r}} = 64$)	134 MB	2.79 MB	19710	91.7	1.43 ×

TRECVID'08 copy detection task

no	transformation	best	second	ours	rank (/23)
1	camcording	0.079	0.363	0.224	2
2	picture in picture	0.015	0.148	0.321	4
3	insertion of patterns	0.015	0.076	0.079	3
4	strong re-encoding	0.023	0.095	0.064	2
5	change of gamma	0.000	0.000	0.023	3
6	photometric attacks	0.038	0.192	0.064	2
7	geometric attacks	0.065	0.436	0.140	2
8	3 random transformations from 6/7	0.045	0.076	0.437	5
9	5 random transformations from 6/7	0.038	0.173	0.693	5
10	5 random transformations	0.201	0.558	0.537	2

Thank you for your attention

DEMO!

Jean-Paul GAUTHIER
LSIS / USTV - ESCODI
http://www.lsis.org/gauthierjp/

"Sur les Mécanismes Optimaux mis en oeuvre par le Système Nerveux Central"

On prouve le "Théorème" suivant: Le système nerveux central minimise quelque chose comme le travail absolu, c'est à dire la dépense effective d'énergie non signée.

Cette preuve repose sur 3 points: L'observation systématique de l'apparition de périodes de silence de l'activité musculaire (agoniste et simultanément antagoniste) dans les mouvements de pointage; le principe du maximum de Pontriaguin ; et le théorème de transversalité (Thom).

Une discussion est alors posée sur le contrôle optimal oculaire, à savoir si ce principe d'optimalité est applicable ou non, si oui dans quelle mesure.

Références :

Gauthier et al., "A biomechanical inactivation principle ", Proceedings of the Steklov Mathematical Institute, Vol 268, 2010.

Gauthier et al., " The Inactivation Principle: Mathematical Solutions minimizing the Absolute Work and Biological Implications for the Planning of Arm Movements ", PLoS Comput. Biol. 4 (2008), N10, http://www.lsis.org/gauthierjp/papers/114.pdf "

An Inactivation Principle In Biomechanics

By J.P. Gauthier, B.Berret, F.Jean
Universities of Toulon,
Dijon, and E.N.S.T.A. (France).
Supported by CNES (National Center for Spatial Studies)

1

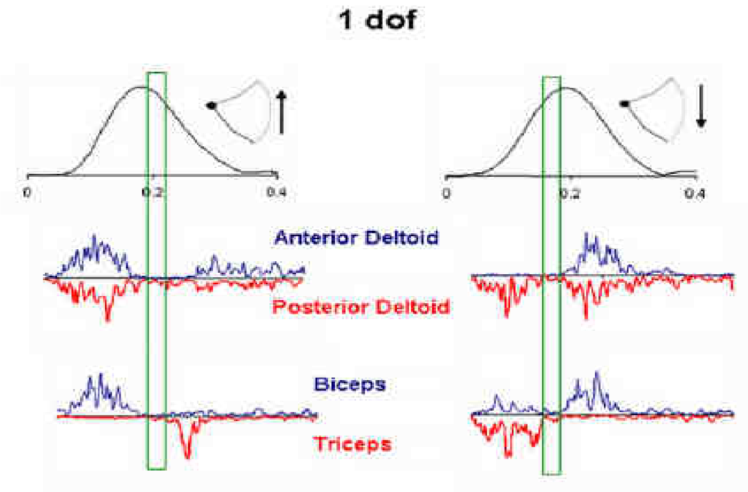
I. Description of practical experiments.

- **Conditions of the experiments:**

- **We consider « pointing movements » of human beings, i.e. movements of the arms, in (specified) time T, the endpoint of the finger starting (at time 0) from an initial point p0, and ending (at time T) at a terminal point pT.**
- **The movement is learned by the subject (he learns how to realize the specified movement within the given time T). In order to learn the duration T of the movement, the subject is given bip sounds at initial and terminal times.**
- **He learns by repetition.**
- **Duration T of the movements is short (about 0.5 sec.)**
- **Although the theory is general, experiments were made for one or two degrees of freedom only, in vertical planes. In the case of a single d.o.f., the elbow joint is attached (constrained to remain at 180°).**
- **Experiments were done in normal gravity, and also in zero gravity.**
- **Study in the perspective of long spatial flights.**

2

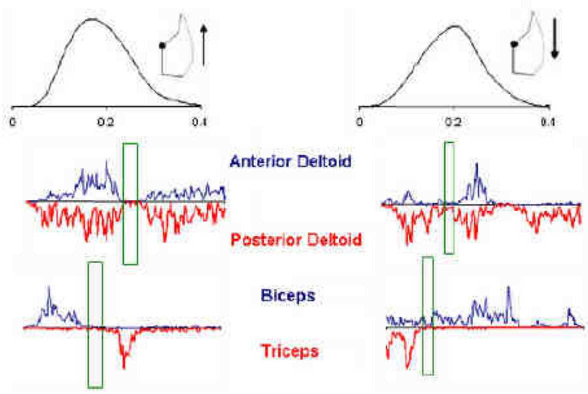
Experiments



3

experiments

Two-degree of freedom experiment:



4

Observations

- **The experiments show the following.**

- **-Inactivations :**
  - **Always simultaneous for the agonistic and antagonistic muscles.**
  - **In the one d.o.f. case, simultaneous for all musles of both joints.**

- **Asymmetries:**
  - **Maximum velocity is not reached at the middle of the time inteval, but a bit before (for upward motion).**

5

The systems under consideration

$$(1.1) \qquad (\Sigma)\ \ddot{x} = \phi(x, \dot{x}, u),$$

Lagrangian: $L(x, \dot{x}) = \frac{1}{2}\dot{x}^T M(x)\dot{x} - V(x),$

Let $\tau = S(x)u$ represent the generalized force resulting from the input u and define $N(x, \dot{x})$ to be any other forces acting on the system, for instance friction forces. *We assume that the control acts on every degree of freedom, that is, $u \in \mathbb{R}^n$ and $S(x)$ is invertible.* Moreover, in the "exactly-fully-actuated" case that we consider first, $S(x) = Id$, which means that we control directly via the generalized forces. **This is always possible up to some feedback.**

The equations of motion are given by substituting into Lagrange's equation,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = S(x)u + N(x, \dot{x}) = \tau + N(x, \dot{x}),$$

and are exactly of the form (1.1), with

$$(3.1) \qquad \phi(x, \dot{x}, u) = M(x)^{-1}(N(x, \dot{x}) - \nabla V(x) - C(x, \dot{x})\dot{x} + \tau),$$

where the Coriolis matrix $C(x, \dot{x}) \in M_n(\mathbb{R})$ is defined as

$$C_{ij}(x, \dot{x}) = \frac{1}{2}\sum_{k=1}^{n}\left(\frac{\partial M_{ij}}{\partial x_k} + \frac{\partial M_{ik}}{\partial x_j} - \frac{\partial M_{kj}}{\partial x_i}\right)\dot{x}_k.$$

6

Systems under consideration

Also, using the Legendre transform: $(x, \dot{x}) \to (x, p)$, where $p = \frac{\partial L}{\partial \dot{x}}$, considering the Hamiltonian $h(x, p)$of the problem:

$$h(x,p) = < p, \dot{x} > -L(x,p),$$

we get the equations of the motion via the characteristic field of the Hamilton-Jacobi equation:

$$\dot{x} = \frac{\partial h}{\partial p}, \dot{p} = -\frac{\partial h}{\partial x} + \tau + N(x,p),$$

from what follows that the work w of external forces, $w = \int (\tau + N(x,p))dx$ is just equal to the variation of the Hamiltonian:

$$\dot{w} = \dot{h}.$$

In particular, if there is no friction, the variation of the Hamiltonian is equal to the work of the external forces τ during the motion. Of course, in this last equality, the work of external forces is counted algebraically: a motion in one direction followed by a motion in opposite direction may give zero work.

7

The absolute work

In the paper, we consider the Absolute-Work Aw of the external forces, which corresponds actually to the energy spent to control the system:

$$(3.2) \qquad Aw = \int |\tau \dot{x}| dt = \int \sum_{i=1}^{n} |\tau_i \dot{x}_i| dt.$$

As a consequence, $\dot{Aw}$ denotes the function of y, τ :

$$\dot{Aw} = |\tau \dot{x}| = \sum_{i=1}^{n} |\tau_i y_i|.$$

Then, in the next sections we shall consider our controlled mechanical system Σ (with $u = \tau$) :

$$(\Sigma)\ \dot{X} = \Phi(X,u), \qquad X \in \mathbb{R}^{2n}, \quad u \in U \subset \mathbb{R}^n,$$

and we shall minimize a cost that will be a compromise of the form:

$$(3.3) \qquad J(u) = Aw + \int_0^T M(X,u)dt,$$

in which $M(X,u)$ is a "comfort term" that for technical reasons we will assume to be smooth and strictly convex w.r.t. the control u.

8

Agonistic-antagonistic action

- u belongs to a subset U of $\mathbb{R}^m$ with $0 \in int\, U$. Here in most cases, U is a product of intervals of the type:

$$U = [u_1^-, u_1^+] \times ... \times [u_n^-, u_n^+],$$

when the system is exactly-fully-actuated, or:

$$U = [0, u_1^+] \times ... \times [0, u_n^+] \times [u_1^-, 0] \times ... \times [u_n^-, 0],$$

in the case of a pair of agonistic-antagonistic muscles for each degree of freedom. In both cases $u_i^- < 0$, $u_i^+ > 0$, $i = 1, ..., n$.
- $\phi \in C^\infty(\mathbb{R}^{3n}, \mathbb{R}^n)$ is such that $\frac{\partial \phi}{\partial u}(x, \dot{x}, u)$ is always invertible.

Setting $X = (x, y) = (x, \dot{x})$, we rewrite the system as $\dot{X} = \Phi(X, u)$, $\quad X \in \mathbb{R}^{2n}$, $\quad u \in U \subset \mathbb{R}^n$.

$$(1.2) \qquad (\Sigma)\ \dot{X} = \Phi(X, u), \qquad X \in \mathbb{R}^{2n}, \quad u \in U \subset \mathbb{R}^n.$$

9

Agonistic-antagonistic action

In the case of agonistic-antagonistic action, we set $u = u_1 - u_2$, where $0 \le u_{1,i} \le u_i^+$ and $0 \le u_{2,i} \le -u_i^-$. Then $u_{1,i}$ (resp. $u_{2,i}$) are the agonistic (resp. antagonistic) generalized force applied at the i^{th} degree of freedom.

We consider a compromise of the type (1.4), i.e. $J(u) = \int_0^T f(x, y, u)dt + Aw$, in which

$$Aw = \int_0^T \sum_{i=1}^n |u_i y_i| dt$$

for total actions. It means that, for agonistic-antagonistic actions, we shall minimize:

$$J'(u_1, u_2) = \int_0^T f(x, y, u_1 - u_2)dt + Aw',$$

where,

$$Aw' = \int_0^T (\sum_{i=1}^n |u_{1i} y_i| + \sum_{i=1}^n |u_{2i} y_i|)dt,$$

is the **total** absolute work of external forces.

10

Choice of the compromise to be minized:

- **Typical choice of f(x,y,u) in the compromise:**

Compromise « absolute work » versus « energy of acceleration »:

We chose f(x,y,u)= $$\sum_{i=1}^{n} \alpha_i \, (\ddot{x}_i)^2,$$

Such a compromise means that human being don't like large accelerations.

11

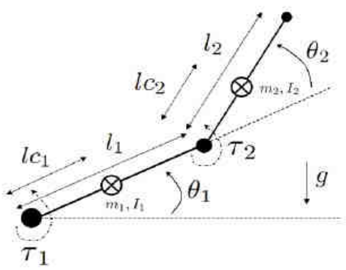
Examples

- **1 dof:**

$$(2.1) \qquad (\Sigma_{1d}) \; \dot{x} = y$$
$$\dot{y} = u - k\cos(x).$$

Here, the constant k reflects the action of the gravity field, $u \in \mathbb{R}$ is a torque, and depending on the context u is bounded ($u^- \leq u \leq u^+$ but $u^- < 0 < u^+$) or $\dot{u}$ is bounded ($\dot{u}^- \leq \dot{u} \leq \dot{u}^+$).

. **2 dof**



12

examples

$$(2.2) \qquad \tau = H(\theta).\ddot{\theta} - \hat{h}(\theta).r(\dot{\theta}) + G(\theta) + B.\dot{\theta},$$

In which H is the (symmetric positive definite) matrix of principal inertia moments, $\hat{h}(\theta).r(\dot{\theta})$ is the Coriolis term, G is the vector of gravitational torques and B is the matrix of friction terms (a constant here). The term τ is the vector of external torques (the controls in our case), i.e. $\tau = u$. Details show:

$$(2.3) \qquad \begin{aligned} \tau_1 &= H_{11}\ddot{\theta}_1 + H_{12}\ddot{\theta}_2 - \hat{h}\ \dot{\theta}_2^2 - 2\hat{h}\ \dot{\theta}_1\dot{\theta}_2 + G_1 + B_{11}\dot{\theta}_1 + B_{12}\dot{\theta}_2, \\ \tau_2 &= H_{22}\ddot{\theta}_1 + H_{21}\ddot{\theta}_2 + \hat{h}\ \dot{\theta}_1^2 + G_2 + B_{21}\dot{\theta}_1 + B_{22}\dot{\theta}_2, \end{aligned}$$

with

$$\begin{aligned} H_{11} &= m_1 l_{c1}^2 + I_1 + m_2 l_{c2}^2 + I_2 + m_2(l_1^2 + 2 l_1 l_{c2}\cos\theta_2), \\ H_{12} &= m_2 l_{c2}^2 + I_2 + m_2 l_1 l_{c2}\cos\theta_2, \\ H_{21} &= H_{12}, \\ H_{22} &= m_2 l_{c2}^2 + I_2, \\ \hat{h} &= m_2 l_1 l_{c2}\sin\theta_2, \\ G_1 &= g\{m_1 l_{c1}\cos\theta_1 + m_2\ (l_{c2}\cos(\theta_1+\theta_2) + l_1\cos\theta_1)\}, \\ G_2 &= g m_2 l_{c2}\cos(\theta_1+\theta_2), \end{aligned}$$

13

Examples: 2 dof

In the following, the variables are denoted as follows: $x_1 = \theta_1, y_1 = \dot{\theta}_1, x_2 = \theta_2,\ y_2 = \dot{\theta}_2, \tau_1 = u_1, \tau_2 = u_2$.

Let $H(x_2)$ denote the matrix

$$H(x_2) = \begin{pmatrix} m_1 l_{c1}^2 + I_1 + m_2 l_{c2}^2 + I_2 + m_2(l_1^2 + 2 l_1 l_{c2}\cos x_2) & m_2 l_{c2}^2 + I_2 + m_2 l_1 l_{c2}\cos x_2 \\ m_2 l_{c2}^2 + I_2 + m_2 l_1 l_{c2}\cos x_2 & m_2 l_{c2}^2 + I_2 \end{pmatrix}.$$

Then finally our control system can be rewritten as:

(2.4)

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ H^{-1}(x_2).\begin{pmatrix} u_1 - G_1(x_1,x_2) + \hat{h}(x_2).(y_2^2 + 2y_1y_2) - B_{11}y_1 - B_{12}y_2 \\ u_2 - G_2(x_1,x_2) - \hat{h}(x_2).y_1^2 - B_{21}y_1 - B_{22}y_2 \end{pmatrix} \end{pmatrix}$$

14

Examples: small angles assumption

Example 3. *(One degree of freedom, small angles assumption).*

The first order approximation in the equation 2.1 just consists of setting $\cos(x) = \cos(x_0)$. Assuming the initial condition to be the "horizontal arm", we get $\cos(x) = 1$ and the system is just the following standard linear[3] system:

$$(2.5) \qquad (\Sigma_{1dl})\ \dot{x} = y, \quad \dot{y} = u - k.$$

Example 4. *(Two degrees of freedom, small angles assumption).*

Here, as in the previous example, we neglect the friction terms. Therefore, in the linearization around an equilibrium point $(x, \dot{x}) = (x, y) = (x_0, 0)$, we get no occurrence of y : linear part is zero and quadratic part in y disappears at $y = 0$. Therefore, our linearized[4] system is of the following form, setting $X = \begin{pmatrix} x \\ y \end{pmatrix}$:

$$(2.6) \qquad (\Sigma_{2dl})\ \dot{X} = AX + Bu + F,$$

where A, B, F are of the form:

$$A = \begin{pmatrix} 0 & Id_2 \\ \tilde{A} & 0 \end{pmatrix},\ B = \begin{pmatrix} 0 \\ \tilde{B} \end{pmatrix}, F = \begin{pmatrix} 0 \\ \tilde{F} \end{pmatrix}.$$

Under this assumption, all computations can be made explicitely.

15

Non existence of Inactivations in the smooth case.

- **Using transversality theorems, we prove the following;**
- **The system is given and fixed, ans we assume that we minimize a smooth cost:**

$$J(u) = \int_0^T f(x, y, u)dt \cdot$$

We also assume that u=0 (inactivation) is not in the boundary of the control set.
Note that this is not true for agonistic-antagonistic action. But in that case, we proceed in a different way to show the result.

Theorem: The set of costs f(x,y,u) such that no inactivation occurs is open-dense.
(In the Whitney topology).

Then, there should be some nonsmoothness in the cost: something like the absolute work.

16

Idea for the proof (of nonexistence of inactivations in smooth case).

- **We give the idea for total inactivations. Caution: for partial inactivations, it is more complicated. In the one dof case, it is the same.**

(P1) $\dot{P}(t)^T = -\frac{\partial H}{\partial X}(X(t), P(t), \lambda, 0)$,
(P2) $H(X(t), P(t), \lambda, 0) = \max_{v \in U} H(X(t), P(t), \lambda, v)$,

where the Hamiltonian of the problem is

$$H(X, P, \lambda, u) = p^T y + q^T \phi(X, u) - \lambda f(X, u).$$

Note that, since $0 \in int\,U$, property (P2) implies $\frac{\partial H}{\partial u}(X(t), P(t), \lambda, 0) = 0$. It follows,

$$q(t)^T = \lambda \frac{\partial f}{\partial u}(X(t), 0)\frac{\partial \phi}{\partial u}(X(t), 0)^{-1}.$$

If $\lambda = 0$, then $q \equiv 0$. From $\dot{q} \equiv 0$ and (P1) we deduce $p \equiv 0$ and then $(P, \lambda) \equiv 0$, which is impossible. Thus λ is positive and a standard argument of homogeneity allows to normalize it to $\lambda = 1$. Finally, from respectively (P1) and (P2), the following holds on the interval I:

$$\dot{p}^T = -q^T \frac{\partial \phi}{\partial x}(X, 0) - \frac{\partial f}{\partial x}(X, 0),$$

$$\dot{q}^T = -p^T - q^T \frac{\partial \phi}{\partial y}(X, 0) - \frac{\partial f}{\partial y}(X, 0) \tag{5.1}$$

and,

$$q^T = \frac{\partial f}{\partial u}(X, 0)\frac{\partial \phi}{\partial u}(X, 0)^{-1}. \tag{5.2}$$

17

Idea for the proof (of nonexistence of inactivations in smooth case).

Now, recall that on I the dynamic is $\dot{X} = F(X)$. Since $X^0 = 0$ is not an equilibrium point of F, we assume, up to a local change of the coordinates $X = (X_1, \ldots, X_{2n})$ on $\mathbb{R}^{2n}$, that $F = \frac{\partial}{\partial X_1}$. Differentiating (5.1) with respect to time leads to

$$\begin{aligned}\ddot{q}^T &= -\dot{p}^T - \dot{q}^T \frac{\partial \phi}{\partial y} - q^T \frac{\partial}{\partial X_1}\frac{\partial \phi}{\partial y} - \frac{\partial}{\partial X_1}\frac{\partial f}{\partial y} \\ &= -q^T \frac{\partial \phi}{\partial x} - \frac{\partial f}{\partial x} - \dot{q}^T \frac{\partial \phi}{\partial y} - q^T \frac{\partial}{\partial X_1}\frac{\partial \phi}{\partial y} - \frac{\partial}{\partial X_1}\frac{\partial f}{\partial y},\end{aligned} \tag{5.3}$$

in which we omit the evaluation at $(X, 0)$.

On the other hand, we can also obtain $\dot{q}^T$ and $\ddot{q}^T$ by differentiation of (5.2):

$$\dot{q}^T = \frac{\partial}{\partial X_1}\frac{\partial f}{\partial u} \times (\frac{\partial \phi}{\partial u})^{-1} + \frac{\partial f}{\partial u} \times \frac{\partial}{\partial X_1}(\frac{\partial \phi}{\partial u})^{-1}$$

$$\ddot{q}^T = \frac{\partial^2}{\partial X_1^2}\frac{\partial f}{\partial u} \times (\frac{\partial \phi}{\partial u})^{-1} + 2\frac{\partial}{\partial X_1}\frac{\partial f}{\partial u} \times \frac{\partial}{\partial X_1}(\frac{\partial \phi}{\partial u})^{-1} + \frac{\partial f}{\partial u} \times \frac{\partial^2}{\partial X_1^2}(\frac{\partial \phi}{\partial u})^{-1}.$$

Substituting these expressions and (5.2) into (5.3), we eliminate q^T, $\dot{q}^T$, and $\ddot{q}^T$ and we obtain:

$$\frac{\partial^2}{\partial X_1^2}\frac{\partial f}{\partial u} + R_X(\frac{\partial}{\partial X_1}\frac{\partial f}{\partial u}, \frac{\partial f}{\partial u}, \frac{\partial}{\partial X_1}\frac{\partial f}{\partial X_i}, \frac{\partial f}{\partial X_i}) = 0 \qquad \text{on } I,$$

where, for every X, R_X is a linear mapping and $X \mapsto R_X$ is smooth. Successive derivations and evaluation of the derivatives at $t = 0$ (remind that $X(0) = 0$) lead to a system of equations of the form:

$$\frac{\partial^k}{\partial X_1^k}\frac{\partial f}{\partial u}(0) + R^k\Big(\frac{\partial^j}{\partial X_1^j}\frac{\partial f}{\partial u}(0), \frac{\partial^j}{\partial X_1^j}\frac{\partial f}{\partial X_i}(0); j < k, 1 \le i \le 2n\Big) = 0, \quad k \ge 2,$$

where each R^k is a linear mapping.

Thus we have proved $\mathcal{A}^m(0) \subset \ker \psi$, where $\psi : \mathcal{J}_0^m \longrightarrow \mathbb{R}^{n(m-2)}$ is the linear mapping which to a m-jet $j_0^m f$ associates

$$\Bigg(\frac{\partial^k}{\partial X_1^k}\frac{\partial f}{\partial u}(0) + R^k\Big(\frac{\partial^j}{\partial X_1^j}\frac{\partial f}{\partial u}(0), \frac{\partial^j}{\partial X_1^j}\frac{\partial f}{\partial X_i}(0); j < k, 1 \le i \le 2n\Big)\Bigg)_{2 \le k \le m-1}.$$

This linear mapping being obviously surjective, the conclusion follows. $\square$ .8

Inactivation principle (for total actions).

We minimise a compromise « absolute work » versus « comfort term ».

Let us consider some optimal trajectory defined on $[0, T]$, and meeting the following two technical assumptions (H_1, H_2) :

(H_1) **Continuity of optimal control:** the corresponding optimal control $u^*(t)$ is continuous on $[0, T]$,

(H_2) **Change of sign for optimal control:** some component u_i^* of the optimal control changes sign at some time $t_c \in]0, T[$, while $y_i(t)$ keeps constant sign. It means that there are some times t_1, t_2, $t_1 < t_c < t_2$, such that $u_i^*(t_1)u_i^*(t_2) < 0$ and $y_i(t) \neq 0$ for $t_1 \leq t \leq t_2$.

Theorem 4. *(inactivation principle) Along a regular optimal trajectory meeting (H_1, H_2) there are partial inactivations. There are also some regular extremal trajectories along which total inactivations do exist, through any $X \in \mathbb{R}^{2n}$.*

The proof relies on some (trivial) nonsmooth analysis argument. But some of its generalizations (dynamics on the muscles) really use a nonsmooth version of PMP.

19

Validity of assumptions (H1) and (H2).

- **Assumption (H1) is automatically reached (for regular trajectories) as soon as the comfort term is strictly convex with respect to the control. In particular, this is the case for a compromise of type: « absolute work » vs « energy of acceleration ».**

- **Optimal trajectories are not singular.**

- **Assumption (H2) holds in general if T is not too large in front of minimum time:**

The reason is that for minimum time, control is bang-bang (with changes of sign, due to initial and terminal condition of zero velocity).

This fits with practical observations.

20

Idea of the proof of the inactivation principle

Proof. The proof is very simple: along the optimal trajectory the Hamiltonian h of the optimal control problem has to be maximum. which means by (4.8) that $0 \in \partial_{u_i} h$ for all $i = 1, ..., p$. But, $h(\lambda, X(t), P(t), u^*(t)) = \lambda\varphi(Aw, X, u^*) + P.\Phi(X, u^*)$ and $\lambda < 0$ since we consider regular trajectories only.

The maximum condition for the Hamiltonian gives:

$$0 \in \partial_{u_i} h(P(t), X(t), u^*(t)). \tag{4.10}$$

The variables $P(t)$ and $X(t)$ being also continuous, the quantity $\partial_{u_i} h(P(t), X(t), u^*(t))$ is an interval $I(t)$ (degenerating to a point as soon as $u_i^*(t) \neq 0$) and **moving continuously** with the time t. At a time t_c when $u_i^*(t_c) = 0$, it is a nontrivial time interval $I(t_c)$, since $\frac{\partial \varphi}{\partial Aw}$ and λ are both different from zero. Hence, since $u_i^*(t)$ changes sign at t_c, it takes a certain strictly positive amount of time to cross $I(t_c)$. Then $u_i^*(t)$ remains exactly equal to zero during some nontrivial time interval. This is partial inactivation.

Second, we take an arbitrary $X = (x, y)$, with $y_i \neq 0$ for all $i = 1, ..., n$ and $\lambda = -1$. We denote by $(M(x)^{-1})_i$ the i^{th} column of the invertible matrix $M(x)^{-1}$. Then, for $u = 0$, we compute the set $\mathcal{S} = \partial_u h(P, X, u)$. If we set $P = (p, q)$, then due to the fact that $\frac{\partial P.\Phi(X,u)}{\partial u_i} = q\frac{\partial \phi(x,y,0)}{\partial u_i} = q(M(x)^{-1})_i$, we can chose q for 0 be exactly the center of the set $\mathcal{S} \subset \mathbb{R}^n$, which is a hypercube with nonempty interior. It is clear by construction that the extremals starting from this point $(X, P, 0)$ have total inactivations. □

21

Results: one dof case

- **In that case, with small angles assumptions, explicit formulas are easily obtained.**

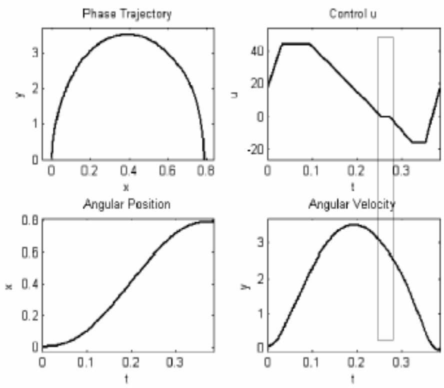


FIGURE 3. The results we get for an upward motion. The phase of inactivation of the muscles is shown.

22

Results in 2-dof case

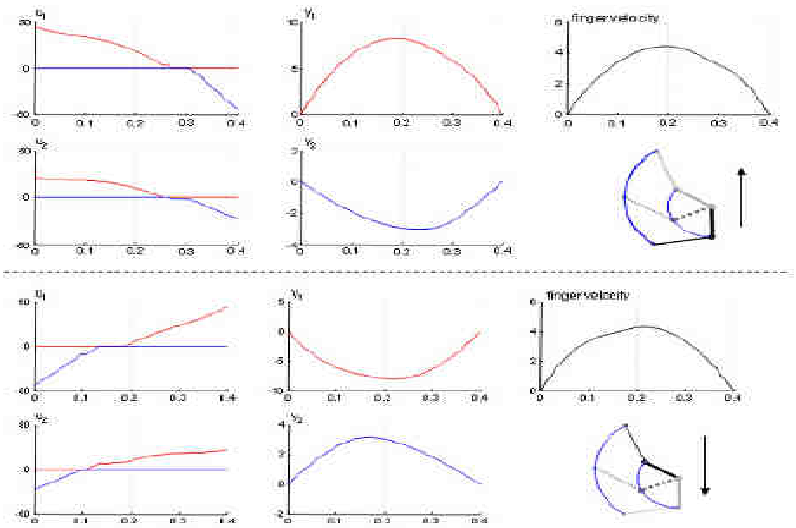


FIGURE 7. 2-degree of freedom theoretical result.

23

Gradient constraints on the inputs.

- **One has to apply Clarke's maximum principle (in a trivial way).**

$$\dot{x} = y, \ \dot{y} = \varphi(x, y) + u, \ \dot{u} = v$$
$$v^- \leq v \leq v^+, \ v^- < 0, \ v^+ > 0,$$
$$\min_v J(v) = Aw + \int_0^T f(X, u)dt.$$

Now the cost is not differentiable anymore with respect to the state (in place of the control previously). Therefore, the Clarke's nonsmooth version of the maximum principle is needed ([3]).

If (p, q, r) denote now the adjoint variables, we get:

$$\tilde{H} = -\lambda(y|u| + f(x, y, u)) + py + q(u - k\cos x) + rv.$$

Once again, x, y, p, q, r, u are continuous (by nature now, just as classical solutions of differential equations). The a-priori fact that y remains ≥ 0 is just checked numerically. Although, it is expectable from the results in the case without gradient constraints.

Also, for similar reasons as in Section (4.2.3), the "abnormal extremal may be excluded, i.e. v not "bang" and $\lambda = 0$: maximality of the Hamiltonian for non-bang implies that r is identically zero, which, with two successive differentiations imply that q and p respectively are also identically zero. Total adjoint vector is zero, which contradicts the maximum principle. Hence we may assume $\lambda = 1$.

We assume that the gradient constraints v^-, v^+ are large enough for the optimal control be of the following type: gradient constraints are active only at the beginning and at the end of the motion. If we refer to the scenarii occurring in Section 4.5.1, it should be what happens: without the gradient constraints, gradient is never large. Then, there will be saturation of the gradient constraints only because of the jumps at the beginning and at the end of the motion. Numerical computations confirm this scenario.

24

Gradient on controls: Clarke's maximum principle

Then, to connect (in optimal way) the source $(x, y, u) = (x_0, 0, u_0)$ to the target $(x_T, 0, u_T)$, where u_0 and u_T are the stationary controls corresponding respectively to $x_0 < x_T$,the strategy must be as follows: $v = v^+$, for $0 \leq t \leq T_1$; $v^- < v < v^+$ for $T_1 < t \leq T_2$; $v = v^-$ for $T_2 < t \leq T$.

Therefore, inside the interval $[T_1, T_2]$, the Hamiltonian being maximum w.r.t. v, we must have $r(t) = 0$. therefore also $\frac{dr}{dt} = 0$. But by the Clarke's maximum principle, it means that $\frac{dr}{dt} \in -\partial_u \tilde{H} = yI + \frac{\partial f}{\partial u} - q$, in which I is the Clarke's gradient of the absolute value function at zero, i.e. $I = [-1, 1]$.

Since $\frac{dr}{dt} = 0$, we conclude:

$$0 \in -\partial_u \tilde{H} = yI + \frac{\partial f}{\partial u} - q.$$

This equation was exactly the cause of inactivations in the non-constrained case.

25

Inactivations for agonistic-antagonistic muscles

For this analysis, we consider that $u = u_1 - u_2$, where $0 \leq u_{1,i} \leq u_i^+$ and $0 \leq u_{2,i} \leq -u_i^-$. Then $u_{1,i}$ (resp. $u_{2,i}$) are the agonistic (resp. antagonistic) generalized force applied at the i^{th} degree of freedom.

$$J'(u_1, u_2) = \int_0^T f(x, y, u_1 - u_2)dt + Aw',$$

where,

$$Aw' = \int_0^T \left(\sum_{i=1}^n |u_{1i} y_i| + \sum_{i=1}^n |u_{2i} y_i|\right) dt,$$

is the **total** absolute work of external forces.

Theorem 5. *(Total inactivation for agonistic-antagonistic actions) In the case of agonistic-antagonistic actions, minimizing a compromise containing the absolute work leads to total (simultaneous) inactivations of both actions, exactly where the total optimal action is inactive.*

26

Proof of inactivation principle for agonistic-antagonistic muscles

First let us assume that $u_1(t)$, $u_2(t)$, minimize J', with optimal value $J^{*\prime}$. Consider $u(t) = u_1(t) - u_2(t)$. Clearly, $u(t)$ applied to the system:

$$\ddot{x} = \phi(x, \dot{x}, u), \tag{4.13}$$

and $u_1(t), u_2(t)$ applied to the system:

$$\ddot{x} = \phi(x, \dot{x}, u_1(t) - u_2(t)), \tag{4.14}$$

produce the same x-trajectories. Therefore,

$$\begin{aligned} J(u) &= \int_0^T (f(x, y, u_1 - u_2) + \sum_{i=1}^{n} |(u_{1i} - u_{2i}) y_i|) dt, \\ &\leq \int_0^T (f(x, y, u_1 - u_2) + \sum_{i=1}^{n} |u_{1i} y_i| + \sum_{i=1}^{n} |u_{2i} y_i|)) dt, \\ &= J'(u_1, u_2) = J^{*'} \end{aligned}$$

This shows that the minimum $J^* = \min_u J(u) \leq J^{*'}$.

27

Proof of inactivation principle for agonistic-antagonistic muscles

Conversely, assume that u attains the minimum J^* of $J(u)$. We define u_1, u_2 from u as follows:

$$\begin{aligned} u_1(t) &= u(t) \text{ if } u(t) > 0, \\ &= 0 \text{ elsewhere, and} \\ u_2(t) &= -u(t) \text{ if } u(t) < 0, \\ &= 0 \text{ elsewhere.} \end{aligned} \tag{4.15}$$

Again $u_1 - u_2 = u$, hence applying u to (4.13) produces the same x-trajectory than applying $u_1 - u_2$ to 4.14. therefore:

$$\begin{aligned} J'(u_1, u_2) &= \int_0^T (f(x, y, u_1 - u_2) + \sum_{i=1}^{n} |u_{1i} y_i| + \sum_{i=1}^{n} |u_{2i} y_i|)) dt, \\ &= \int_0^T (f(x, y, u_1 - u_2) + \sum_{i=1}^{n} |(u_{1i} - u_{2i}) y_i|) dt, \end{aligned}$$

by definition of u_1, u_2. It means that:

$$J'(u_1, u_2) = J^*, \tag{4.16}$$

This shows the converse.

28

Dynamics on the muscles and the « triphasic pattern »

- **Dynamics on the muscles:**

$$(4.17)\qquad \begin{cases} \dot{x} = y \\ \dot{y} = u_1 - u_2 - k \\ \dot{u}_1 = -\frac{u_1}{\sigma_1} + v_1 \\ \dot{u}_2 = -\frac{u_2}{\sigma_2} + v_2 \end{cases}$$

with again $u_1,\, u_2 \geq 0$, and we look for the minimum $\min_{u_1,u_2} \int_0^T \; yu_1 + yu_2 \,+\, \alpha \ddot{y}^2 \; dt$.

For this, we use the a-priori fact (which is checked numerically) that, as in the case of total action ([2]), y remains positive during the upward motion. The Hamiltonian may be written as:

$$\mathcal{H} = -y(u_1+u_2) - \alpha(u_1-u_2-k)^2 + py + q(u_1-u_2-k) + r_1\left(-\frac{u_1}{\sigma_1}+v_1\right) + r_2\left(-\frac{u_2}{\sigma_2}+v_2\right)$$

At this point, there is an important technical detail that physiologically makes sense. It can be understood as "muscular coactivation" at the end of the motion, a well know phenomenon: due to the first order linear dynamic on the muscles, and the constraints $u_i \geq 0$, we can go back to zero asymptotically only. Therefore, the terminal condition $u_T = k$ is impossible. Hence we require, with $\varepsilon > 0$:

$$(4.18)\qquad \begin{aligned} &(I)\ u_1^0 = k \text{ and } u_2^0 = 0, \\ &(II)\ u_1^T = k + \varepsilon \text{ and } u_2^T = \varepsilon. \end{aligned}$$

Requirement (II) is coactivation at terminal time T. Then, explicit computations with the maximum principle, together with a numerical research of the commutation times show that the optimal scenario is like shown on Figure 8

29

Dynamics on the muscles and the « triphasic pattern »

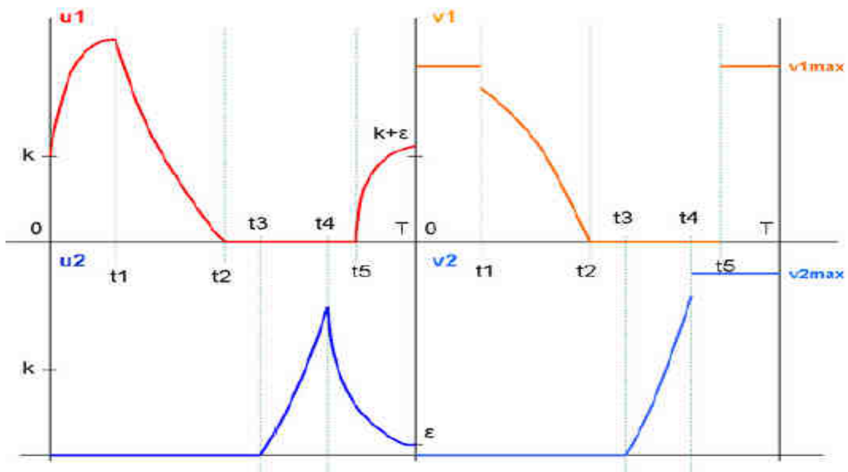


FIGURE 8. Optimal Triphasic Pattern.

Agonistic burst, followed by antagonistic burst, followed against by agonistic burst. The scenario ends with co-contraction of the muscles.

30

Extensions to vision

- **Two directions:**

- **1. Saccades of the eye:**
  In that case the equations consist of the Euler equations with control terms.
  Results should completely similar. But there is a problem for validation (need to go to Guantanamo again).
  But, absolute work is very small.

- **2. Listing's law for saccades:**
  Such a behaviour could be studied by the same approach.

I shall discuss this shortly!!

31

Imprimé à La Garde en septembre 2010 à l'université Sud-Toulon Var

ISBN   2-9524747-3-7

EAN   9782952474733