



**HAL**  
open science

# Contributions en modélisation de la vision algorithmique et bio-inspirée

Pierre Kornprobst

► **To cite this version:**

Pierre Kornprobst. Contributions en modélisation de la vision algorithmique et bio-inspirée. Interface homme-machine [cs.HC]. Université Nice Sophia Antipolis, 2007. tel-00457491

**HAL Id: tel-00457491**

**<https://theses.hal.science/tel-00457491>**

Submitted on 26 Feb 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nice-Sophia Antipolis

## HABILITATION A DIRIGER DES RECHERCHES

Secteur Disciplinaire : DS9  
(Sciences et Technologies de l'Information et de la Communication)

# Contributions en modélisation de la vision algorithmique et bio-inspirée

par

Pierre Kornprobst



*Le 14 décembre 2007*

Rapporteurs :	<b>Vicent Caselles</b>	Professeur, UPF, Barcelone
	<b>Jean-Pierre Cocquerez</b>	Professeur, UTC
	<b>Simon Thorpe</b>	Directeur de recherche, CNRS, CERCO
Examineurs :	<b>Michel Barlaud</b>	Professeur, UNSA, I3S
	<b>Rachid Deriche</b>	Directeur de recherche, INRIA, ODYSSÉE
	<b>Olivier Faugeras</b>	Directeur de recherche, INRIA, ODYSSÉE
	<b>Guillaume Masson</b>	Directeur de recherche, CNRS, DYVA

 La version électronique de ce document est disponible à l'adresse:  
<ftp://ftp-sop.inria.fr/odyssee/Publications/2007/kornprobst:07.pdf>

# SOMMAIRE

<b>1 Curriculum vitæ et compte-rendu d'activité</b>	<b>7</b>
Parcours professionnel . . . . .	9
Formation . . . . .	9
Activité d'enseignement . . . . .	10
Encadrement d'étudiants . . . . .	10
Participation à la vie de la communauté . . . . .	11
Développement logiciel . . . . .	12
Production scientifique . . . . .	13
<b>2 Rappel du sujet de thèse et des principaux résultats obtenus (1995-1998)</b>	<b>15</b>
Restauration d'images par EDP . . . . .	17
Flot optique avec préservation des discontinuités par approches variationnelles . . . . .	17
Restauration et segmentation de séquences d'images par approche variationnelle . . . . .	18
<b>3 Liste des publications durant la thèse</b>	<b>19</b>
<b>4 Résultats obtenus depuis la thèse et axes de recherches développés (1999-2007)</b>	<b>21</b>
4.1 Approches variationnelles et méthodes par équations aux dérivées partielles (EDP) . . . . .	26
4.1.1 Diffusion des connaissances . . . . .	26
Livre sur les aspects mathématiques des méthodes par EDP en traitement d'images . . . . .	26
Articles d'encyclopédies . . . . .	27
4.1.2 Traitement d'images provenant de modalités spécifiques . . . . .	27
Techniques de super-résolution pour des séquences d'images IRMf . . . . .	27
Restauration et analyse des images SAR . . . . .	30
4.1.3 Autres applications en vision par ordinateur . . . . .	32
Structure spatio-temporelle des séquences d'images et mouvement . . . . .	32
Désoccultation d'images non texturées . . . . .	36
Méthode générique pour la résolution des problèmes variationnels . . . . .	39
4.2 Modélisation bio-inspirée de tâches visuelles . . . . .	41
Modèle de rétine biologique avec contrôle de gain . . . . .	41
Modèles dynamiques des aires V1/MT et catégorisation de mouvement . . . . .	46
Calcul de mouvement avec rétroaction et contrôle par la forme . . . . .	53
4.3 Les méthodes variationnelles en neurosciences ? . . . . .	58
Processus continus et comportements collectifs discrets . . . . .	58
Spécification de processus discrets par une approche variationnelle générale . . . . .	62
Approches variationnelles et segmentation par synchronisation . . . . .	64
<b>5 Liste des publications après la thèse</b>	<b>69</b>
<b>A Principaux développements logiciels</b>	<b>75</b>
Systèmes d'aide à la vision pour les patients atteints de scotomes . . . . .	75
Virtual Retina . . . . .	76
<b>B Documents joints</b>	<b>79</b>



## A propos de la présentation du document

- Les chapitres 2 et 4 présentent mon activité de recherche. Mes publications y sont indiquées et référencées dans les chapitres 3 et 5. Par exemple, **[J 1]** fera référence à un article de journal. Les publications extérieures les plus utiles pour la présentation des résultats seront citées en bas de page au fil du texte.
- Les logos suivants sont utilisés dans le document.

 : indique la présence d'un lien actif dans la version PDF du document.

 (dans la marge) : indique que les papiers mentionnés dans le texte sont disponibles dans la version papier. Voir le chapitre B pour la liste des documents fournis.

 (dans la marge) : indique un paragraphe qui termine en général une contribution, soit pour revenir sur certaines difficultés, soit pour en indiquer quelques perspectives.



# CHAPITRE 1

## Curriculum vitæ et compte-rendu d'activité

### Sommaire

---

Parcours professionnel . . . . .	9
Formation . . . . .	9
Activité d'enseignement . . . . .	10
Encadrement d'étudiants . . . . .	10
Participation à la vie de la communauté . . . . .	11
Développement logiciel . . . . .	12
Production scientifique . . . . .	13

---



## Pierre Kornprobst

INRIA Sophia Antipolis - Méditerranée

Equipe-projet Odysée

2004, route des Lucioles - B.P. 93 06902 Sophia Antipolis Cedex, France

Tél: 04-9238-7979

E-mail: pierre.kornprobst@inria.fr

 <http://www-sop.inria.fr/odyssee/team/Pierre.Kornprobst>

34 ans

Service militaire accompli en tant que coopérant (CSNE)

### Chargé de recherche INRIA 1ère classe, équipe-projet Odysée (INRIA, ENS, ENPC)

Recruté en tant que chargé de recherche dans le projet Robotvis en 2000, j'ai poursuivi mes travaux sur les approches variationnelles avec l'écriture d'un livre<sup>[1]</sup> co-écrit avec Gilles Aubert, et quelques contributions en vision par ordinateur. L'année 2002 marque un tournant : Robotvis devient Odysée, avec un changement thématique important. L'objectif de l'équipe est désormais de s'intéresser à la vision biologique et à l'analyse des mécanismes neuronaux sous-jacents. Suivant cette évolution, je travaille depuis sur la modélisation bio-inspirée de fonctions visuelles (e.g., modèle de rétine impulsionnelle) ou de tâches visuelles (e.g., estimation de mouvement et reconnaissance de mouvements biologiques). Mon objectif est aussi de montrer l'intérêt des approches variationnelles en neurosciences computationnelles (rapport avec les réseaux de neurones, modèles d'architectures corticales, conception d'applications). Depuis, j'ai également exercé des activités d'enseignement et de formation pour différents publics (Université, entreprise et chercheurs), encadré de nombreux étudiants (dont trois thèses), contribué à la diffusion des connaissances (deux chapitres d'encyclopédies), et participé activement à la vie de la communauté (vie de l'institut, comités de programmes, GDR). Ce document est une synthèse de mes contributions.

### Parcours professionnel

Depuis 2003	Chargé de Recherche 1ère classe dans l'équipe-projet Odysée
2000-2003	Chargé de Recherche 2ème classe dans l'équipe-projet Robotvis, devenue Odysée en 2002.
1999-2000	Service national effectué en tant que coopérant (CSNE) dans le laboratoire de vision de l'University of Southern California (Los Angeles), en partenariat avec la société MATRA (département Système et Information), pour une durée de 16 mois.
1998	Collaborateur extérieur dans le projet Robotvis (INRIA Sophia-Antipolis), sous la direction d'Olivier Faugeras, sur l'application des approches variationnelles en magnéto-encéphalographie (MEG), pour une durée de 2 mois.

### Formation

1995-1998	Doctorat en Mathématiques. Titre: " <i>Contribution à la restauration d'images et à l'analyse de séquences: Approches Variationnelles et Solutions de Viscosité</i> ". Directeurs de thèse: Gilles Aubert (Université de Nice Sophia-Antipolis) et Rachid Deriche (projet Robotvis, INRIA Sophia-Antipolis). Soutenance en novembre 1998. <i>Mention très honorable avec les félicitations.</i>
1994-1995	D.E.A. de Mathématiques (UNSA), <i>mention TB</i>
1993-1994	Maîtrise en Ingénierie Mathématique (UNSA), <i>mention TB</i>
1992-1993	Licence de Mathématiques (UNSA), <i>mention TB</i>

---

[1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, January 2002.

- 1990-1991 Classes préparatoires de Mathématiques Supérieures et Spéciales (Lycée Masséna, Nice). Reçu au concours de l'Ecole Nationale Supérieure d'Electronique et de Radioélectricité de Bordeaux (ENSERB Informatique).
- 1989-1990 Terminale C (Lycée Masséna, Nice). *Baccalauréat mention B*

## Activité d'enseignement

- 2007 Vacation donnée à l'Ecole Nationale des Ponts et Chaussées (ENPC), à des élèves de deuxième année d'école, sur le thème: "Quand les mathématiques se mettent en images" (1h).
- 2007 Cours d'une demi-journée donné dans le cadre de la conférence SIGGRAPH sur le thème "A Gentle Introduction to Bilateral Filtering and its Applications," avec Sylvain Paris (MIT), Jack Tumblin (Northwestern University, IL) et Fredo Durand (MIT).
- 2006 Mise en place d'un programme ERASMUS entre l'Universitat Pompeu Fabra (Barcelone) et l'Université de Nice-Sophia Antipolis, pour le master "Image et Géométrie pour le Multimédia et la Modélisation du Vivant" (IGMMV).
- 2003-2007 Cours et Travaux Pratiques de traitement d'images en cinquième année de l'ESINSA, école Polytechnique de l'Université de Nice-Sophia Antipolis, département électronique (module de 30h). Il s'agissait d'une introduction au traitement d'images et aux méthodes par équations aux dérivées partielles (EDP) dans ce domaine.
- 2005 Formation avancée sur trois jours à l'IMRA (filiale de Toyota) sur les méthodes par EDP en traitement d'images. L'objectif était de fournir aux développeurs une vue d'ensemble du potentiel de ce type d'approche par rapport à leurs problèmes, ainsi que les outils nécessaires pour comprendre et tester rapidement des algorithmes de référence.
- 2002-2005 Cours de traitement d'images dans l'option imagerie numérique de la deuxième année du DUT info de l'IUT de Nice (module de 30h). Il s'agissait d'une introduction au traitement d'images. Préparation des travaux dirigés également.
- 2002 Travaux pratiques de traitement d'images en Java/C++ en deuxième année d'ESSI (20h).
- 1995-1998 **Moniteur** : Travaux dirigés de Mathématiques en première année de Deug SVT (Sciences de la Vie et de la Terre) à l'Université de Nice-Sophia Antipolis (64 heures équivalent TD par an). Cette formation comprenait une dizaine de journées de stage par an autour de la pratique enseignante, de la communication, des nouvelles technologies au service de l'enseignement, etc.
- 1997-1998 Travaux dirigés de Mathématiques en formation continue pour cadres à l'IUT de Nice (20h).

## Encadrement d'étudiants

### ► Doctorants

- 2007-2010 **Emilien Tlapale** (Directeur de thèse. Thèse financée par une demi-bourse BDO de la région et par des projets Européens. Titre : "L'utilisation de la rétroaction dans les mécanismes visuels de l'attention, de la segmentation et de l'identification").
- 2006-2009 **María-José Escobar** (Co-directeur. Directeur de thèse : Thierry Viéville. Thèse financée par le contrat COGNICYT de la collaboration Chili-INRIA. Titre : "Etude de la catégorisation de mouvement biologiques par des méthodes bio-inspirées").
- 2004-2007 **Adrien Wohrer** (Directeur de thèse. Thèse financée par l'Ecole Polytechnique. Titre : "La rétine : Etude et modélisation"). Soutenance prévue en janvier 2008.

► Master

- 2007 **Emilien Tlapale** (Master ESSI3-EPU et IGMMV)  
**Gabriel Montero** (Master du Chili)  
**Léonard Gérard** (Master Université Paris 7)
- 2006 **Bassem Bouhleb, Emilien Tlapale, Mohamed Youssef** (Master UNSA I Informatique)  
**Ketty Nguyen** (Master ESSI3-EPU et IGMMV)  
**Sandrine Chemla** (Master ESSI3-EPU et IGMMV)
- 2004 **Ivan Dimov** (Master du Chili)  
**Riadh Layouni** (DEA)  
**Alexandre Reynaud** (ESSI 3<sup>ème</sup> année)  
**Nicolas Rondelet** (ESSI 3<sup>ème</sup> année)
- 2003 **Sophie Di Martino** (DEA)

► Autre

- 2005 **Damien Montigny** (IUT DUT Info)
- 2003 **Eric Boloré** (IUT DUT Info)

## Participation à la vie de la communauté

► Implication dans la vie de l'institut

- Depuis 2007 Membre du comité de suivi doctoral (CSD)
- 2002-2003 Représentant chercheur et secrétaire de séance au *Comité des Projets*
- 2001-2004 Membre élu (suppléant) au *Comité de Centre*
- 2001-2004 Membre nommé (titulaire) au *Comité Local de Formation Permanente*
- 2004 Membre du jury d'admission du concours IT 2004 pour un poste de Gestionnaire au bureau des missions

► Animation de la recherche

- 2008-2011 Responsable INRIA pour le projet Européen SEARISE (FP7, "*Smart Eyes: Attending and Recognizing Instances of Salient Events*")
- 2007-2013 Membre du conseil scientifique du Pôle de Recherche Scientifique et Technique (PRST) intitulé "*Modélisation, informations et systèmes numériques*" (MISN), de la région Lorraine

► Membre de jurys scientifiques

- 2007 Rapporteur de projets du programme ANR "*Jeunes Chercheuses, Jeunes Chercheurs*"
- 2005 Examineur de la thèse de Gloria Haro Ortega (Universitat Pompeu Fabra, Espagne; Directeur de thèse : Vincent Caselles)
- 2004 Examineur de la thèse d'Agustin Trujillo (Universidad de Las Palmas de Gran Canaria, Espagne; Directeur de thèse : Luis Alvarez)
- 2004 Membre du jury pour l'attribution du prix de thèse Telecom Valley 2004

► Implication nationale dans des groupements de recherche (GDR)

- 2007 Séminaire invité dans le cadre de l'AG des 20 ans du GDR ISIS (Jean-Pierre Coquerez)
- 2006 Organisation d'un *workshop* dans le cadre du GDR Mathématiques des Systèmes Perceptifs et Cognitifs (MSPC, Laurent Cohen), ayant pour thème : "*Neurones et réseaux de neurones biologiques : Comment les modéliser?*" ( Lien vers le programme de la journée)
- 1998 Séminaire invité dans le cadre du GDR GT4
- 1996 Séminaire invité dans le cadre du GDR PRC ISIS

- 1996            Séminaire invité dans le cadre du GDR GT1 CHM
- Membre de comités de programme de conférences nationales et internationales
- 2007            PSIVT (The 2007 IEEE Pacific Rim Symposium on Image Video and Technology)
- 2006            ECCV (European conference on computer vision)
- 2005            ORASIS (congrès francophone portant sur les thèmes liés à la vision par ordinateur.)
- 2005            Scale-Space (The 5th international conference on scale space and PDE methods in computer vision)
- 2004            ECCV (European conference on computer vision)
- 2003            Scale Space (The 4th international conference on scale space and PDE methods in computer vision)
- Activité régulière d'évaluation d'articles pour les journaux suivants :
- Journal of Mathematical Imaging and Vision
  - Computer Graphics Journal
  - CVIU Journal
  - IEEE Signal Processing Letters
  - IEEE Transactions on Image Processing
  - IEEE Transactions on Medical Imaging
  - IEEE Transactions on Pattern Analysis Machine Intelligence
  - Image and Vision Computing Journal
  - International Journal of Computer Vision
  - International Journal of Image and Graphics
  - Revue Traitement du Signal
  - SIAM Journal on Scientific Computing
  - IEEE Transactions on Circuits and Systems for Video Technology

## Développement logiciel

- 2007            **SolairePDF** (avec Emilien Tlapale, en collaboration avec Eric Castet de l'Institut de Neurosciences Cognitives de la Méditerranée). Nous nous sommes intéressés au développement d'un outil permettant aux patients atteints de scotomes (i.e., des tâches sombres apparaissant généralement au centre du champ visuel) de lire avec plus de confort des documents. Suite à une étude préliminaire, nous avons développé l'outil SolairePDF qui permet de naviguer dans tous types de PDF et d'obtenir une vision rehaussée d'un ou plusieurs paragraphes choisis par le lecteur. Cet outil a été mis à disposition à l'hôpital de la Timone pour en évaluer l'efficacité avec de vrais patients. Voir l'Appendice A pour plus de détails.
- 2007            **Virtual Retina** (avec Adrien Wohrer et Thierry Viéville) Virtual Retina est un simulateur de rétine qui transforme une vidéo en trains d'impulsions (*spikes*). Le modèle sous-jacent est bio-inspiré : chaque élément correspond à une réalité physiologique. Le modèle comprend une partie de filtrage linéaire non-séparable, un mécanisme de contrôle de gain, et la génération de trains d'impulsions correspondant aux cellules ganglionnaires. Virtual Retina permet de simuler une rétine comportant jusqu'à 100.000 neurones et reproduit fidèlement des mesures faites sur de vraies cellules. Ce développement a été mené dans le cadre du projet FACETS (EC IP project FP6-015879). Voir l'Appendice A pour plus de détails.  
Revue INédit, no 61 : <http://www.inria.fr/actualites/inedit/inedit61.indusb.fr.html>  
 <http://www-sop.inria.fr/odyssee/software/virtualretina>  
*APP logiciel Virtual Retina: IDDN.FR.001.210034.000.S.P.2007.000.31235*

- 2006 **CoolPDE**, plugin de la librairie CImg (voir <http://cimg.sourceforge.net/>) qui propose plusieurs méthodes classiques par équations aux dérivées partielles. Ce plugin est proposé dans le cadre de la seconde version du livre "Mathematical Problems in Image Processing", chez Springer (voir la liste des publications).  
 <http://www-sop.inria.fr/books/imath/>
- 2005 **VisStim** (avec Thierry Viéville, Damien Montigny, Alexandre Reynaud, et Benjamin Caramel, en collaboration avec Eric Castet de l'Institut de Neurosciences Cognitives de la Méditerranée). Visual Stimuli est une mini plateforme open-source de génération des stimuli visuels classiquement utilisés en psychophysique. Cette plateforme permet de disposer d'une base de stimuli ouverte (tout le monde peut en profiter et y contribuer), parfaitement définie (les paramètres du stimulus sont spécifiés de manière formelle). Elle est basée essentiellement sur la technologie S.V.G. de graphisme vectoriel, ce qui permet de générer facilement de nombreux stimuli.  
 <http://www-sop.inria.fr/odyssee/imp/visstim/>  
*APP deposit: IDDN.FR.001.410012.000S.P.2005.000.31235 VISSTIM (software), IDDN.FR.001.410013.000S.P.2005.000.31235 I.M.P. (related middle-ware)*

## Production scientifique

- Livre G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations (Second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.
- HDR P. Kornprobst. *Contributions en modélisation de la vision algorithmique et bio-inspirée*. Habilitation à diriger des recherches, Université de Nice-Sophia Antipolis, dec 2007.
- Thèse P. Kornprobst. *Contributions à la Restauration d'Images et à l'Analyse de Séquences: Approches Variationnelles et Solutions de Viscosité*. PhD thesis, Université de Nice-Sophia Antipolis, 1998.
- Chapitres de livres ou d'encyclopédies: 3
- Publications dans des journaux: 6
- Publications dans des conférences: 28
- Communications avec soumission d'abstract: 18
- Rapports de recherche: 14



## CHAPITRE 2

# Rappel du sujet de thèse et des principaux résultats obtenus (1995-1998)

### Sommaire

---

Restauration d'images par EDP . . . . .	17
Flot optique avec préservation des discontinuités par approches variationnelles	17
Restauration et segmentation de séquences d'images par approche variationnelle	18

---



## Préambule



Le traitement d'images a une longue histoire, à commencer par les approches provenant de techniques du traitement du signal 1-D. Ces approches reposent sur la théorie du filtrage (linéaire ou non), l'analyse spectrale, ou sur des concepts basiques de probabilité et statistiques. Le livre de Jain <sup>[44]</sup> en est un très bon récapitulatif.

Depuis, des approches plus sophistiquées ont été développées, et il est difficile de dire laquelle est la plus naturelle. Trois grandes directions émergent : les approches stochastiques, les ondelettes, et les équations différentielles partielles (EDP). Les approches stochastiques reposent largement sur la théorie des champs de Markov<sup>[51]</sup> et s'appliquent directement sur des images discrètes. Les ondelettes sont issues du traitement du signal et reposent sur des techniques de décomposition<sup>[53]</sup>. Quant aux approches par EDP, elles ont commencé à être appliquées en traitement d'images depuis les années 90, offrant aux mathématiciens un nouveau champ d'applications très prometteur.

C'est dans ce contexte que j'entrepris ma thèse de Mathématiques en 1995, co-encadré par Gilles Aubert (laboratoire J.A. Dieudonné, Université de Nice-Sophia Antipolis) et Rachid Deriche (projet Robotvis, INRIA Sophia Antipolis), avec comme point de départ un certain nombre d'approches existantes sur l'amélioration d'images. Mon objectif était d'apporter une contribution méthodologique à l'analyse de trois applications :

- La restauration d'images.
- L'estimation du mouvement.
- L'analyse des séquences d'images.

Pour chacun de ces trois problèmes, j'ai proposé des modèles par EDP que j'ai étudiés dans le cadre mathématique adapté (solutions de viscosité ou espace des fonctions à variations bornées). Les méthodologies n'étant pas sans rapport avec ce que l'on recherche, mon deuxième objectif était d'explicitier et de justifier les algorithmes utilisés, et bien sûr d'en évaluer les performances.

### Restauration d'images par EDP

Devant la diversité des méthodes par EDP non linéaires, j'ai commencé par une étude comparative des différents opérateurs, d'un point de vue formel mais aussi du point de vue de leur performance. J'ai alors proposé un formalisme commun à l'ensemble de ces méthodes et effectué une série de tests numériques comparatifs **[TH 7]**. Ensuite, j'ai proposé une méthode convergente pour l'élimination simultanée du bruit et du flou **[TH 8]**. Cette approche combine un terme parabolique et un terme hyperbolique. Etant donnée une image  $u_N$ , le schéma d'évolution suivant a été proposé :

$$\frac{\partial u}{\partial t} = -\alpha_a(u - u_N) + \alpha_d(h_{\tau_1}(|G_\sigma * \nabla u|)u_{\eta\eta} + u_{\xi\xi}) - \alpha_r(1 - h_{\tau_2}(|G_\sigma * \nabla u|))\text{sign}((G_{\bar{\sigma}} * u)_{\eta\eta})|\nabla u|,$$

comprenant un terme d'attache aux données, un terme de diffusion et un terme de réaction pour le rehaussement. Dans cette équation, on remarque la présence de la direction  $\eta$  (resp.  $\xi$ ) correspondant à la direction normale (resp. tangentielle) aux isophotes.

D'un point de vue théorique, j'ai justifié ce modèle *via* la théorie des solutions de viscosité, et j'ai proposé un schéma numérique prenant compte la spécificité de chaque opérateur **[TH 1]**.

### Flot optique avec préservation des discontinuités par approches variationnelles

Étant données deux images consécutives dans une séquence d'images, je me suis intéressé au problème de l'estimation du déplacement entre ces deux images, aussi appelé flot optique. J'ai alors proposé une formulation variationnelle s'inspirant de modèles proposés dans le cadre de la restauration d'images, le but étant de

[44] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall International Editions, 1989.

[51] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

[53] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

retrouver des champs de vitesses régularisés, tout en préservant leurs discontinuités **[TH 5]**. La formulation variationnelle suivante a été proposée :

$$\inf_{\sigma} E(\sigma) = \int_{\Omega} |\sigma \cdot \nabla u + u_t| + \sum_{j=1..2} \int_{\Omega} \phi(\nabla \sigma_j) + \int_{\Omega} c(x) |\sigma|^2 dx,$$

où  $\sigma$  est le champ de vitesses à déterminer et  $u$  la séquence d'images d'entrée. Du point de vue théorique, j'ai étudié l'existence et l'unicité d'une solution sur l'espace des fonctions à variations bornées. Le terme de régularisation est à interpréter comme une fonction convexe de mesure. Deux situations ont alors été envisagées suivant la régularité de la donnée  $u$  :

- Pour une donnée  $u$  Lipschitz, j'ai montré que le problème sur BV est bien posé, et j'ai construit un algorithme convergent via des arguments de Gamma-convergence et de dualité **[TH 4]**.
- Pour une donnée  $u$  BV, la formulation est plus complexe, notamment pour l'interprétation de la contrainte du flot optique qui fait intervenir le produit d'une distribution ( $\nabla u$ , en fait  $Du$ ) par une fonction BV ( $\sigma$ ), et j'ai défini une représentation intégrale du problème relaxé **[TH 3]**.

Cette méthode a été implémentée dans le cadre d'une approche multi-échelle, et le logiciel a été transféré dans l'entreprise *Realviz* (start-up INRIA fondée en 1998).

### Restauration et segmentation de séquences d'images par approche variationnelle

Passant à une séquence d'images, j'ai proposé une approche variationnelle originale pour segmenter et restaurer les séquences d'images bruitées avec un fond fixe **[Th 2]**. Étant donnée une séquence d'images bruitées  $u$  prise avec un fond fixe, le problème était de trouver de manière robuste et couplée les objets en mouvement dans la scène (décrits par une fonction  $s \in [0, 1]$ ), ainsi que le fond statique restauré (noté  $b$ ). L'approche variationnelle suivante a été proposée :

$$\inf_{b,s} E(b, s) = \int_t \int_{\Omega} s^2 (b - u)^2 dxdt + \int_t \int_{\Omega} (s - 1)^2 dxdt + \int_{\Omega} \phi_1(|\nabla b|) dx + \int_t \int_{\Omega} \phi_2(|\nabla s|) dxdt.$$

La spécificité de cette approche vient de l'estimation couplée du fond de l'image (parties statiques) et de la détection des parties en mouvement. J'ai alors justifié théoriquement ce modèle sur BV, la spécificité venant de l'aspect vectoriel et non-convexe de l'énergie. De nombreux exemples ont permis de valider cette approche.

## CHAPITRE 3

# Liste des publications durant la thèse

### Thèse

- TH 1 P. Kornprobst. *Contributions à la Restauration d'Images et à l'Analyse de Séquences: Approches Variationnelles et Solutions de Viscosité*. PhD thesis, Université de Nice-Sophia Antipolis, 1998.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/1998/kornprobst:98.ps.gz>

### Publications dans des journaux

- TH 2 P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *Journal of Mathematical Imaging and Vision*, 11(1):5–26, October 1999.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/kornprobst-deriche-et-al:99.ps.gz>

- TH 3 G. Aubert and P. Kornprobst. A mathematical study of the relaxed optical flow problem in the space BV. *SIAM Journal on Mathematical Analysis*, 30(6):1282–1308, 1999.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/aubert-kornprobst:99.ps.gz>

- TH 4 G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal of Applied Mathematics*, 60(1):156–182, 1999.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/aubert-deriche-et-al:99.ps.gz>

### Publications dans des conférences

- TH 5 R. Deriche, P. Kornprobst, and G. Aubert. Optical flow estimation while preserving its discontinuities: A variational approach. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume 2, pages 71–80, Singapore, December 1995.

- TH 6 P. Kornprobst, R. Deriche, and G. Aubert. Estimation du flot optique avec préservation des discontinuités: Une approche variationnelle. In *Journées ORASIS'96*, pages 81–86, May 1996.

- TH 7 P. Kornprobst, R. Deriche, and G. Aubert. Nonlinear operators in image restoration. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 325–331, San Juan, Puerto Rico, jun 1997. IEEE Computer Society, IEEE Computer Society.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/kornprobst-deriche-et-al:97.ps.gz>

- TH 8 P. Kornprobst, R. Deriche, and G. Aubert. Image coupling, restoration and enhancement via PDEs. In *Proceedings of the 4th International Conference on Image Processing*, volume 4, pages 458–461. IEEE Computer Society Press, October 1997.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/kornprobst-deriche-et-al:97b.ps.gz>

- TH 9 P. Kornprobst, R. Deriche, and G. Aubert. EDP, débruitage et réhaussement en traitement d'image: Analyse et contributions. In *11ème Congrès RFIA*, volume 1, pages 277–286. AFCET, January 1998.

- TH 10 P. Kornprobst, R. Deriche, and G. Aubert. Image sequence restoration: A PDE based coupled method for image restoration and motion segmentation. In Hans Burkhardt and Bernd Neumann, editors, *Proceedings of the 5th European Conference on Computer Vision*, volume II of *Lecture Notes in Computer Science*, pages 548–562, Freiburg, Germany, June 1998. Springer–Verlag.

📄 <ftp://ftp-sop.inria.fr/robotvis/html/Papers/kornprobst-deriche-etal:98b.ps.gz>

#### Communications avec soumission d'abstract

- TH 11 P. Kornprobst, R. Deriche, and G. Aubert. Image restoration via PDE's. In *First Annual Symposium on Enabling Technologies for Law Enforcement and Security - SPIE Conference 2942 : Investigative Image Processing.*, Boston, Massachusetts, USA, November 1996.
- TH 12 P. Kornprobst, R. Deriche, and G. Aubert. Image sequence restoration: A PDE based coupled method for image restoration and motion segmentation. Technical Report 3308, INRIA, November 1997.
- TH 13 P. Kornprobst, R. Deriche, and G. Aubert. Etude mathématique du problème du flot optique dans l'espace  $BV$ . In *Proceedings of the 29th Congrès Annuel d'Analyse Numérique*, pages 229–230, Larnas, France, November 1997.
- TH 14 P. Kornprobst, R. Deriche, and G. Aubert. Une méthode variationnelle pour l'analyse des séquences d'images. In *Proceedings of the 30th Congrès Annuel d'Analyse Numérique*, page 239, Arles, France, November 1998.

#### Rapports de recherche

- TH 15 G. Aubert, R. Deriche, and P. Kornprobst. A mathematical study of the regularized optical flow problem in the space  $BV(\Omega)$ . Technical Report 503, Université de Nice-Sophia Antipolis, December 1997.
- TH 16 P. Kornprobst, R. Deriche, and G. Aubert. Image sequence restoration: A PDE based coupled method for image restoration and motion segmentation. Technical Report 3308, INRIA, November 1997.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-3308.html>
- TH 17 G. Aubert, R. Deriche, and P. Kornprobst. A variational method and its mathematical study in image sequence analysis. Technical Report 3415, INRIA, April 1998.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-3415.html>

## CHAPITRE 4

# Résultats obtenus depuis la thèse et axes de recherches développés (1999-2007)

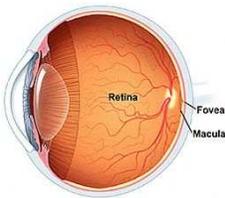
### Sommaire

---

<b>4.1</b>	<b>Approches variationnelles et méthodes par équations aux dérivées partielles (EDP)</b>	<b>26</b>
4.1.1	Diffusion des connaissances	26
	Livre sur les aspects mathématiques des méthodes par EDP en traitement d'images	26
	Articles d'encyclopédies	27
4.1.2	Traitement d'images provenant de modalités spécifiques	27
	Techniques de super-résolution pour des séquences d'images IRMf	27
	Restauration et analyse des images SAR	30
4.1.3	Autres applications en vision par ordinateur	32
	Structure spatio-temporelle des séquences d'images et mouvement	32
	Désoccultation d'images non texturées	36
	Méthode générique pour la résolution des problèmes variationnels	39
<b>4.2</b>	<b>Modélisation bio-inspirée de tâches visuelles</b>	<b>41</b>
	Modèle de rétine biologique avec contrôle de gain	41
	Modèles dynamiques des aires V1/MT et catégorisation de mouvement	46
	Calcul de mouvement avec rétroaction et contrôle par la forme	53
<b>4.3</b>	<b>Les méthodes variationnelles en neurosciences ?</b>	<b>58</b>
	Processus continus et comportements collectifs discrets	58
	Spécification de processus discrets par une approche variationnelle générale	62
	Approches variationnelles et segmentation par synchronisation	64

---





## Préambule

Suite à ma thèse de doctorat soutenue en novembre 1998, j'ai effectué mon service militaire en tant que coopérant, dans le laboratoire de vision par ordinateur de l'University of Southern California (USC, Los Angeles), dirigé par Gérard Médioni, et financé par la société MATRA Systèmes et Information. Cela m'a permis de travailler sur le *tensor voting*, qui est une méthode de groupement perceptuel développée par Gérard Médioni, que j'ai pu appliquer dans le cadre du suivi d'objets en mouvements dans la continuité de ma thèse [C 1].



---

*Recruté dans le projet Robotvis à l'INRIA Sophia Antipolis en 2000, j'ai poursuivi mes travaux sur les approches variationnelles et par équations aux dérivés partielles (EDP), qui avaient connu de nombreux développements en vision par ordinateur.*

---

Mes contributions dans ce domaine sont présentées dans la section 4.1. Il y a eu tout d'abord un effort important sur la diffusion des connaissances (section 4.1.1), mené avec Gilles Aubert (Université de Nice-Sophia Antipolis). Ce fut un travail conséquent pour analyser la littérature, et savoir la présenter de façon unifiée et pédagogique.

- Nous avons écrit un livre sur les aspects mathématiques de ces approches, tout en donnant une vue d'ensemble des applications. Ce livre a été publié par *Springer* en 2002 et une deuxième édition prenant en compte les dernières évolutions a été proposée en 2006.
- Dans cet esprit, nous avons également contribué dans deux encyclopédies (par *Elsevier* et *Vuibert*).

Ensuite, j'ai travaillé sur du traitement d'images venant de modalités spécifiques (section 4.1.2) comme

- La super-résolution de séquences d'IRM fonctionnelles (IRMf).
- La restauration d'images radar (SAR).

Enfin, j'ai travaillé sur d'autres problématiques de traitement d'images (section 4.1.3) :

- L'estimation du mouvement à partir de l'analyse spatio-temporelle des structures.
- La désocclusion (*inpainting*) d'images géométriques.
- La résolution des problèmes variationnels sur les espaces de Sobolev.

---

*2002 marque un tournant : l'équipe-projet Robotvis, dont l'objectif scientifique était la vision algorithmique, devient Odysée (équipe-projet commune entre l'INRIA, l'ENS et l'ENPC), avec un changement thématique important. L'objectif de l'équipe est désormais de s'intéresser à la modélisation de la vision biologique et à l'analyse des mécanismes neuronaux sous-jacents.*

---

Dans la modélisation de la vision biologique, je me suis intéressé à comprendre comment le système visuel fonctionne, principalement de la rétine à l'aire MT, pour proposer des modèles bio-inspirés pouvant servir à effectuer des tâches visuelles de plus haut niveau comme la reconnaissance d'actions dans des séquences d'images (section 4.2).

Bio-inspirés ou bio-plausibles, ces qualificatifs sont de plus en plus utilisés dans la littérature, comme un justificatif irréfutable et absolu du bien fondé d'un modèle, puisqu'il proviendrait de la biologie... Pourtant, définir en quoi un modèle est bio-inspiré n'est pas si clair. On peut penser à plusieurs interprétations :

- *Puisque le système visuel communique au moyen d'impulsions, un modèle est dit bio-inspiré s'il modélise une activité neuronale impulsionnelle et traite l'information contenue dans cette suite d'impulsions* (voir figure 4.1 (a)). Dans ce cas, le modèle est décrit par un réseau de neurones biologiques, dont la structure et la connectivité prend en compte des connaissances de la neurophysiologie, et qui permet également de reproduire des mesures intra ou extra-cellulaires sur des cellules réelles. Une simplification largement répandue est de considérer un neurone biologique comme une unité "intègre-et-tire", décrite notamment

par son potentiel. Un neurone intègre son entrée dans le temps jusqu'à ce que son potentiel atteigne un seuil, alors il "décharge", générant un potentiel d'action (i.e., une impulsion). Le potentiel du neurone est alors réinitialisé et, après une période réfractaire, continue d'intégrer les entrées. Ainsi, une idée également largement utilisée, est de décrire l'activité d'un neurone par sa fréquence de décharge. Mais un neurone n'est jamais isolé : il fait partie d'un large réseau. Décrire son activité par une simple fréquence de décharge est donc réducteur, et de nombreuses autres façons existent pour caractériser l'activité du neurone dans le réseau<sup>1</sup>, prenant mieux en compte la nature discrète temporellement du signal neuronal.

- *Puisque le cortex visuel est organisé de façon hiérarchique en aires corticales spécialisées, un modèle est dit bio-inspiré s'il cherche à reproduire à un niveau macroscopique une organisation fonctionnelle des différentes aires corticales* (voir figure 4.1 (b)). La fonction et la dynamique temporelle des interactions entre les aires sont les deux éléments à reproduire. De nombreuses données sont disponibles, notamment chez le primate, mais aussi chez l'homme grâce à des outils non-invasifs d'imagerie tels que l'IRMf, la MEG et l'EEG (pour l'étude de l'activité électrique du cerveau), ou le dtMRI et les méthodes HARDI (pour l'étude des connectivités entre aires corticales). Dans ce cas, le modèle pourra décrire les activités de plusieurs aires corticales en interaction, au travers de systèmes d'équations différentielles par exemple.
- *Puisque le système visuel produit un percept, un modèle est dit bio-inspiré s'il permet de reproduire ce percept, qui peut différer du stimulus* (voir figure 4.1 (c)). C'est un objectif difficile car le percept résulte l'interaction de nombreux facteurs, dont un certain nombre de grands principes édictés par la théorie de la Gestalt<sup>[46]</sup>. Le percept révèle de manière indirecte des principes de fonctionnement du système visuel, et on peut alors établir des conjectures quant à leur origine. La psychophysique de la perception apporte ainsi un grand nombre d'observations qu'il s'agit de savoir interpréter et modéliser. Dans ce cas, le modèle pourra être simplement phénoménologique, en proposant des mécanismes pour expliquer le percept, sans chercher à vouloir être trop proche des structures et données biologiques sous-jacentes.

Ces propositions de définition du qualificatif bio-inspiré ont permis d'introduire trois échelles d'observation du fonctionnement du cerveau pour la vision. Cette variété des échelles d'analyse<sup>[22]</sup> (voir également la figure 4.2) révèle la difficulté de définir ce qu'est réellement "modéliser le cerveau" de façon bio-inspirée. Au regard de cette extraordinaire complexité, tout modèle sera donc nécessairement une abstraction, qui cherchera à reproduire certains comportements. Mais reproduire n'est pas expliquer : des simplifications sont nécessaires pour rendre compte de mécanismes complexes.

La démarche suivie dans mon travail est de proposer des modèles de fonctions visuelles à partir des connaissances sur le système visuel, en combinant les trois échelles mentionnées ci-dessus, capables de reproduire des mesures de cellules ou des percepts, et permettant de développer des applications concurrentes de la vision par ordinateur. L'ensemble de ce travail a été mené en collaboration avec Thierry Viéville (équipe projet Odyssée/Cortex, INRIA).

Le point de départ de mes recherches a été les travaux autour de Simon Thorpe (Centre de Recherche Cerveau et Cognition, CNRS, Toulouse) sur la catégorisation ultra-rapide d'images, qui a permis de montrer un nouvel aspect du codage neural : le codage par rang<sup>[72]</sup>. Notre but a été d'appliquer et étendre cette méthodologie dans le cas des séquences d'images. Dans des travaux préliminaires [C 10], nous avons montré l'efficacité d'un tel paradigme pour la reconnaissance de mouvements biologiques, mais une limitation importante venait de la manière de générer les trains d'impulsions. Pour une vidéo, la dynamique temporelle joue un rôle important qu'il faut prendre en compte.

- Dans cette perspective, j'ai encadré en thèse Adrien Wohrer dont le sujet portait sur la réalisation d'un simulateur de rétine, capable de transformer une vidéo en un flux continu de trains d'impulsions.
- Ensuite, je me suis de nouveau intéressé à la problématique de la reconnaissance du mouvement, dans le cadre de la thèse de María-José Escobar, en prenant mieux en compte ce que nous connaissons du système visuel et des cellules sensibles au mouvement.

---

<sup>1</sup>Par exemple : temps de latences par rapport à un temps référence, mesures de synchronisation ou de corrélations entre des trains d'impulsions, distributions des intervalles de temps entre impulsions (ISI), répétitions de certains "motifs" spatio-temporels, ou différences de phases en régimes oscillatoires.

[46] G. Kanizsa. *Grammatica Del Vedere*. Bologna: Il Mulino, 1980.

[22] P.S. Churchland and T.J. Sejnowski. *The computational brain*. MIT Press, 1992.

[72] S.J. Thorpe. Spike arrival times: A highly efficient coding scheme for neural networks. *Parallel processing in neural systems and computers*, pages 91–94, 1990.

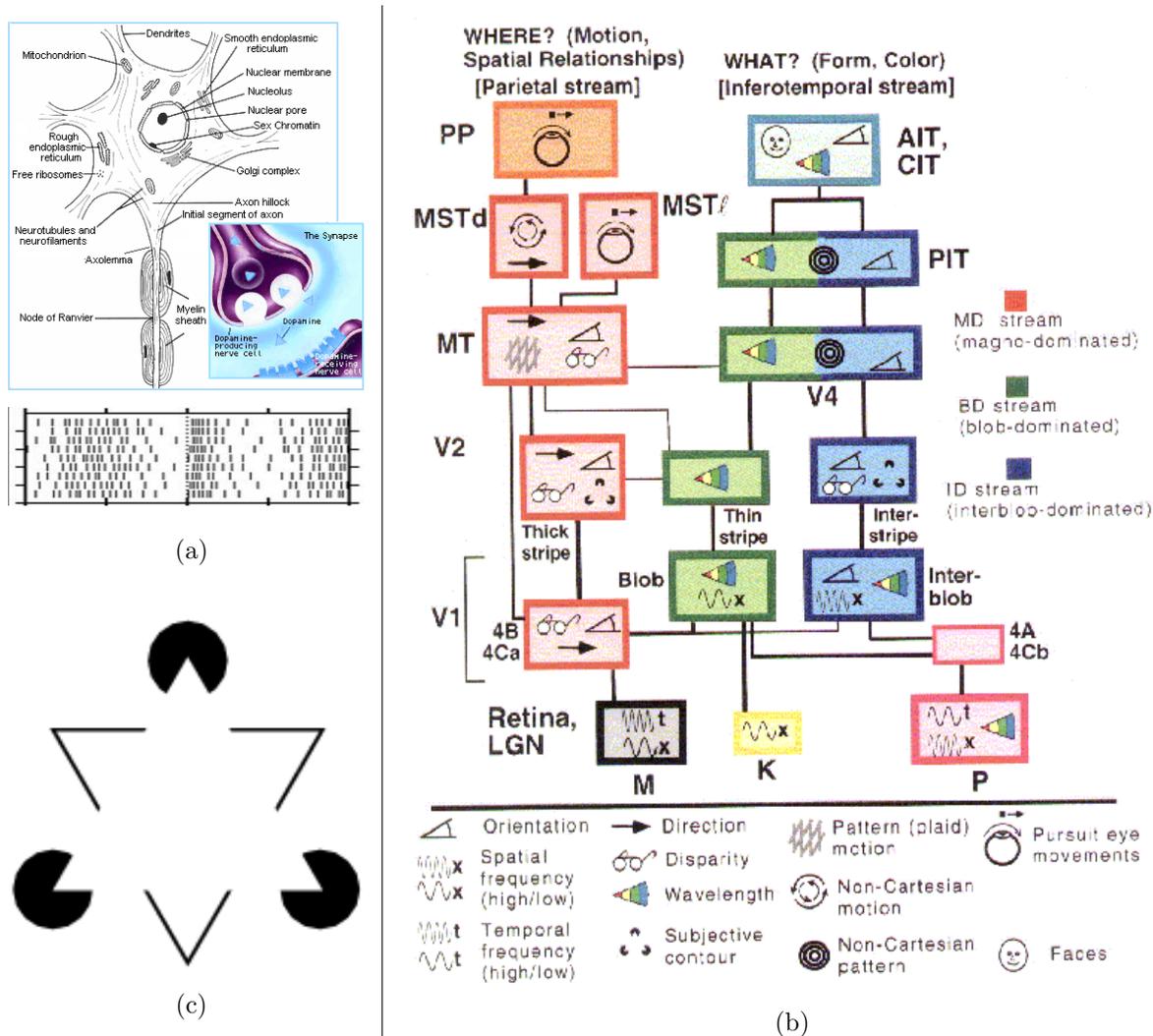


Fig. 4.1: Trois échelles d'observation et de modélisation du cerveau : (a) à l'échelle microscopique, l'entité de calcul est le neurone. Cette figure présente le schéma d'un neurone, et dessous un *raster-plot*, i.e., une représentation des impulsions dans le temps pour un ensemble de neurones. (b) A l'échelle macroscopique, le cortex visuel se compose d'un ensemble hiérarchique d'aires spécialisées et hautement connectées, comme le montre cette représentation de la structure par Van Essen et Gallant (1994). (c) Vu de l'extérieur, nos percepts révèlent certains mécanismes que l'on peut chercher à caractériser : par exemple, le fameux triangle de Kanizsa révèle les contours illusoires d'un triangle blanc en avant de la scène.

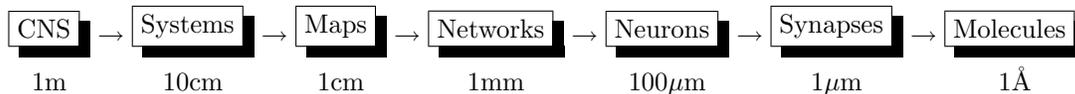


Fig. 4.2: Différentes échelles dans le système nerveux central (CNS), reproduit de Churchland et Sejnowski (1992)

- Enfin, estimer le mouvement met en jeu des mécanismes complexes que nous avons cherché à approfondir avec Emilien Tlapale, en collaboration avec Guillaume Masson (Institut de Neurosciences Cognitives de la Méditerranée, CNRS, Marseille), à la lumière de ce que la psychophysique nous enseigne.

Un deuxième axe de mes recherches porte sur les liens possibles entre le formalisme variationnel et les réseaux de neurones analogiques et impulsifs, en collaboration avec Thierry Viéville (section 4.3). Les trois contributions que je décris dans ce document sont :

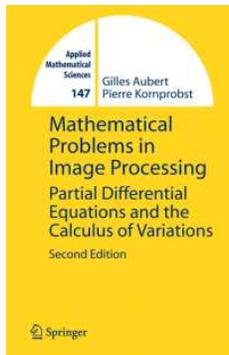
- Etant donné un ensemble d'unités neuronales (e.g., des colonnes d'une carte corticale), nous avons montré comment définir les poids synaptiques permettant d'obtenir une approximation d'un opérateur différentiel elliptique donné.
- Nous avons proposé une formulation variationnelle générale comme un moyen de spécifier le calcul effectué dans une carte corticale, et étudions comment l'interaction entre cartes, et en particulier les récurrences, peuvent-elles être modélisées de manière stable par des ensembles de formulations variationnelles couplées.
- En considérant des neurones cette fois comme des oscillateurs, nous proposons des réseaux de neurones capables de segmenter une image, dont les poids synaptiques sont à partir d'approches variationnelles classiques en traitement d'images. Dans cette étude que nous commençons, on s'intéresse donc aux aspects de synchronisation dans les réseaux de neurones. Plus généralement, cette étude nous conduit à étudier les phénomènes de synchronisation dans les réseaux de neurones biologiques, en fonction du type de connexions, du modèle de neurone, des délais etc.

## 4.1 Approches variationnelles et méthodes par équations aux dérivées partielles (EDP)

### 4.1.1 Diffusion des connaissances

**Livre sur les aspects mathématiques des méthodes par EDP en traitement d'images**

Depuis les années 90, les méthodes par EDP ont connu un réel succès dans le domaine du traitement d'images. Autour de l'année 2000, quelques ouvrages existaient sur le sujet, mais ils étaient souvent restreints à un domaine applicatif, et aucun ne s'intéressait vraiment à expliquer en détail les mathématiques sous-jacentes.



Nous avons donc décidé avec Gilles Aubert d'écrire un livre sur les aspects mathématiques des méthodes par EDP en traitement d'images. Notre objectif était de nous adresser à deux publics : la communauté mathématique, pour démontrer tout l'intérêt de cette méthodologie dans ce contexte, mais aussi la communauté de la vision par ordinateur, pour donner une vision globale et complète des questions mathématiques à se poser lorsqu'on écrit un modèle à base d'EDP.

Ce livre a été publié par Springer en 2002 [LI 1] (315 pages) et sert depuis comme ouvrage de référence dans la communauté. A la demande de l'éditeur, nous avons écrit une nouvelle édition complétée, parue en 2006 [LI 2] (410 pages), dont le sommaire est dans les Annexes.

Le livre comporte cinq chapitres et deux appendices. Le chapitre 1 est une présentation générale. Le chapitre 2 récapitule l'essentiel des notions mathématiques utiles pour ce type de modélisation. Les chapitres 3 et 4 montrent comment les EDP et les approches variationnelles s'appliquent aux problèmes de la restauration et de la segmentation d'images, ce qui permet d'illustrer la plupart des problématiques rencontrées dans ce domaine. Enfin, le chapitre 5 s'intéresse aux nouvelles applications récemment traitées dans ce cadre : nous y présentons la désocclusion des régions manquantes (*inpainting*), la décomposition d'images en géométrie et texture, l'analyse des séquences d'images (pour le calcul du flot optique, la restauration et la segmentation), la classification d'images et le traitement des images vectorielles. Dans les appendices, nous considérons les aspects discrets, à savoir comment implémenter ce type d'approches avec des schémas aux différences finies.

 Lien vers la page web du livre : <http://www-sop.inria.fr/books/imath/>

Cette volonté de diffusion des connaissances est aussi marquée par la rédaction de deux chapitres d'encyclopédies, co-écrits avec Gilles Aubert. Ainsi, nous avons écrit :

- Un chapitre dans l'*Encyclopédie française de l'informatique* éditée par Vuibert [CH 1], en faisant une revue des techniques de base en traitement d'images.  
📄 Lien vers l'encyclopédie : <http://vuibert.com/livre12401.html>
- Un chapitre dans l'*Encyclopedia of Mathematical Physics* éditée par Elsevier [CH 2], dans laquelle nous avons proposé une vue d'ensemble des approches stochastiques, par ondelettes et par EDP en traitement d'images.  
📄 Lien vers l'encyclopédie : <http://www.elsevier.com/locate/emp>

### 4.1.2 Traitement d'images provenant de modalités spécifiques

#### Techniques de super-résolution pour des séquences d'images IRMf

*Cadre contractuel : projet Européen MAPAWAMO ("Mapping visual cortical regions in awake, behaving monkey using functional MRI", projet numéro QLRT1999-30161, 2000–2003). Ce projet avait pour objectif le perfectionnement des techniques existantes d'IRM fonctionnelle pour le singe, afin de permettre la comparaison des images d'IRM fonctionnelle avec une cartographie métabolique standard préexistante chez le singe, et finalement une comparaison directe des réseaux corticaux chez des primates humains et non-humains.*

J'ai travaillé sur la maximisation de la résolution spatiale des séquences d'images fonctionnelles par résonance magnétique (IRMf). Les images fonctionnelles permettent de détecter les variations d'oxygénation sanguine liées à une activation cérébrale (effet BOLD). Augmenter leur résolution est donc un enjeu majeur car cela permet de visualiser les plus petites régions d'activité cérébrale.

Pour y parvenir, la résolution spatiale standard des images IRMf acquises à 1.5 T n'est pas suffisante vu que la condition pour acquérir des images haute résolution est la diminution du SNR, qui est proportionnelle à la décroissance de la taille des voxels. La décroissance du SNR peut être obtenue au moyen de scanners à hauts champs magnétiques, ce qui implique cependant des coûts plus importants, et davantage d'artefacts de distortion dans les images<sup>[33,48]</sup>.

Dans le cadre du projet Européen MAPAWAMO, j'ai donc proposé dans [J 2], [C 4] une technique de super-résolution pour l'amélioration de la résolution spatiale des images IRMf, où la résolution fait ici référence à la taille des voxels. L'idée générale des techniques de super-résolution est d'utiliser plusieurs images basse résolution d'une même scène, avec des décalages entre images, et donc des échantillonnages du signal continu différents, générant une information redondante complémentaire. L'idée était de s'inspirer d'approches déjà proposées en traitement d'images, en imagerie par résonance magnétique anatomique<sup>[37,63]</sup>, et pour le traitement des volumes tridimensionnels<sup>[67]</sup>.

- La première contribution a été de définir et contrôler le processus d'acquisition. Pour cela, j'ai travaillé en étroite collaboration avec Ronald Peeters (Department of Radiology, University Hospitals Leuven, Belgique), qui s'occupait plus spécifiquement de la partie acquisition de données. Étant donné un processus d'acquisition en coupes axiales, suivant l'axe  $z$  (voir figure 4.4 (a) et (b)), nous avons proposé d'acquérir deux volumes basse résolution (i.e., dont la taille des voxel suivant l'axe  $z$  est plus grande) avec un décalage suivant cet axe, produisant ainsi une information redondante échantillonnée différemment (voir figure 4.4 (c) et (d)).

[33] H. Fischer and R. Ladebeck. Echo-planar imaging image artifacts. In et al. F. Schmitt, editor, *Echo-Planar Imaging, Theory, Technique and application.*, pages 191–200. Springer, Heidelberg, 1998.

[48] D.A. Kelley and J.F. Schenck. Very-high-field magnetic resonance imaging: instrumentation and safety issues. *Top Magn Reson Imaging*, 10(1):79–89, February 1999.

[37] H. Greenspan, G. Oz, N. Kiryati, and S. Peled. MRI inter-slice reconstruction using super-resolution. *Magn. Res. Imag.*, 20:437–446, 2002.

[63] S. Peled and Y. Yeshurun. Superresolution in MRI : Application to human white matter fiber tract visualization by diffusion tensor imaging. *Magn. Reson. Med.*, 45:29–35, 2001.

[67] E. Roullot, A. Herment, I. Bloch, M. Nikolova, and E. Mousseaux. Regularized reconstruction of 3D high-resolution magnetic resonance images from acquisitions of anisotropically degraded resolutions. In *Proceedings of the International Conference on Image Processing*, volume 1, pages 350–353. IEEE Signal Processing Society, 2000.

Notons  $\Delta_x, \Delta_y, \Delta_z$ , les pas de discrétisation haute résolution (correspondant à la taille des voxels haute résolution que l'on cherche,  $\Delta_x \times \Delta_y \times R\Delta_z$ ). Chaque voxel  $h_{i,j,k}$  représente l'intensité du signal dans un volume :

$$W_{i,j,k} = [(i-1)\Delta_x, i\Delta_x] \times [(j-1)\Delta_y, j\Delta_y] \times [(k-1)\Delta_z, k\Delta_z].$$

Notons les images basse résolution  $l^r$ , pour  $r = 1, \dots, R$ . Chaque image  $l^r$  représente l'intensité du signal dans un volume :

$$V_{i,j,k}^r = [(i-1)\Delta_x, i\Delta_x] \times [(j-1)\Delta_y, j\Delta_y] \times [(R(k-1)+r)\Delta_z, (Rk+r)\Delta_z].$$

La taille d'un voxel basse résolution est donc  $\Delta_x \times \Delta_y \times R\Delta_z$ .  $R$  correspond donc au facteur de multiplication de la taille en  $z$  que l'on se donne, et donc au nombre de volumes basse résolution décalés qu'il va falloir acquérir (voir l'illustration figure 4.3).

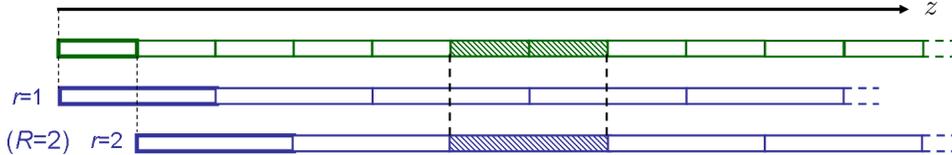


Fig. 4.3: Schéma représentant les acquisitions basse résolution décalées suivant l'axe  $z$ . Les régions hachurées illustrent le principe d'additivité utilisé dans le modèle.

Dans [37], les auteurs proposent un modèle à base de convolution pour décrire la relation entre  $W$  et  $V$ . Ici nous avons supposé un modèle additif : si l'on note  $s(x)$  le signal à une position  $x$ , et le signal mesuré représenté par un voxel  $v$  par  $\int_v s(x)dx$ , alors on a

$$l_{i,j,k}^r = \sum_{n=R(k-1)+r+1}^{Rk+r} \int_{V_{i,j,n}} s(x) dx + n_{i,j,k}^r = \sum_{n=R(k-1)+r+1}^{Rk+r} h_{i,j,n} + n_{i,j,k}^r, \quad (4.1)$$

où  $n$  désigne le bruit de mesure. Cette relation, qui lie  $l^r$  à  $h$ , peut s'écrire sous forme matricielle. Étant donné une position  $(i, j)$ , si l'on note  $h_{i,j}$  et  $\ell_{i,j}^r$  les vecteurs

$$h_{i,j} = [h_{i,j,1}, \dots, h_{i,j,p}]^T \quad \text{et} \quad \ell_{i,j}^r = [\ell_{i,j,1}, \dots, \ell_{i,j,p_r}]^T \quad \text{pour} \quad r = 0, \dots, R-1, \quad (4.2)$$

où  $p_r$  désigne la dimension de  $\ell_{i,j}^r$ , alors (4.1) peut s'écrire

$$\ell_{i,j}^r = B^r h_{i,j} + n_{i,j}^r, \quad \forall i, j \quad \text{avec} \quad B_{k,n}^r = \begin{cases} 1 & \text{si } R(k-1)+r+1 \leq n \leq Rk+r, \\ 0 & \text{sinon.} \end{cases}$$

La matrice  $B^r$  est creuse et le problème de super résolution consiste donc à trouver  $h_{i,j}$ , pour tout  $i, j$ , à partir de  $\ell_{i,j}^r$ , pour tout  $r$  et  $i, j$ . La solution aux moindres carrés consiste à trouver  $\hat{h}$  qui minimise la fonction discrète suivante

$$\Psi(h) = \sum_{i,j} \sum_{r=0}^{R-1} \left| B^r h_{i,j} - \ell_{i,j}^r \right|^2. \quad (4.3)$$

Remarquons que le fait d'avoir considéré un décalage suivant  $z$  uniquement fait que les problèmes ne sont pas liés spatialement. Cela revient à ce stade à considérer des problèmes unidimensionnels.

- On rentre alors dans le cadre classique des problèmes inverses : le problème (4.3) est en fait mal posé<sup>[15]</sup>, si bien que la solution est très sensible au bruit (problème lié à l'inversion de la matrice  $(B^r)^T B^r$ ). L'approche classique en problème inverse est donc de contraindre la solution, par exemple en ajoutant des contraintes de régularité portant sur les dérivées. On cherche alors  $h$  qui minimise une énergie de la forme

$$\mathcal{J}(h) = \Psi(h) + \Phi(h), \quad (4.4)$$

[37] H. Greenspan, G. Oz, N. Kiryati, and S. Peled. MRI inter-slice reconstruction using super-resolution. *Magn. Res. Imag.*, 20:437–446, 2002.

[15] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 600–605, Barcelona, Spain, September 2000. Computer Society Press.

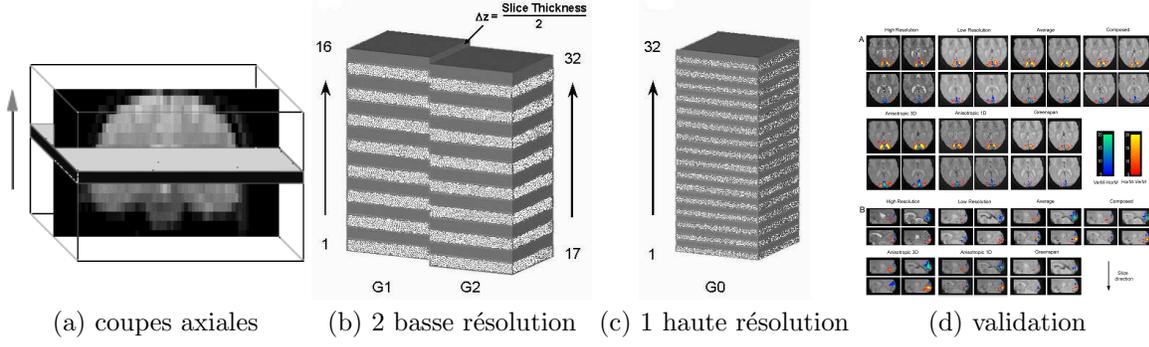


Fig. 4.4: Techniques de super-résolution pour des séquences d'images IRMf : (a) les volumes en IRMf sont obtenus comme un empilement suivant l'axe axial ( $z$ ), (b) le protocole consiste à acquérir plusieurs volumes (ici 2) basse résolution décalés en  $z$ , (c) à partir des volumes en basse résolution décalés suivant  $z$ , on cherche un volume haute résolution avec une taille de voxel en  $z$  plus précise, (d) cette figure illustre le travail effectué pour la validation de cette approche, au travers de l'analyse statistique de séquences d'IRMf, synthétiques et réelles (voir [J 2], [C 4] pour plus de détails).

avec

$$\Phi(h) = \sum_{i,j,k} \beta_x \phi(h_{i,j,k} - h_{i-1,j,k}) + \beta_y \phi(h_{i,j,k} - h_{i,j-1,k}) + \beta_z \phi(h_{i,j,k} - h_{i,j,k-1}), \quad (4.5)$$

où les coefficients  $\beta_*$  sont des scalaires. Chaque terme entre parenthèse est une approximation des dérivées dans chacune des directions et la fonction  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction permettant d'obtenir une régularisation des données avec préservation des discontinuités. Il est classique de choisir une fonction convexe à croissance linéaire à l'infini pour avoir les bonnes propriétés. Nous avons utilisé

$$\phi(t) = \begin{cases} t^2/2 & \text{si } |t| \leq \alpha, \\ \alpha|t| - \alpha^2/2 & \text{si } |t| > \alpha, \end{cases} \quad (\text{fonction de Huber})$$

car ses composantes quadratiques et linéaires permettent des calculs numériques plus efficaces. La procédure d'acquisition se faisant suivant l'axe  $z$ , et après différentes expériences, nous avons choisi un lissage uniquement dans la direction  $z$  :

$$\Phi(h) = \beta \sum_{i,j,k} \phi(h_{i,j,k} - h_{i,j,k-1}) = \beta \sum_{i,j} \sum_k \phi(h_{i,j}(k) - h_{i,j}(k-1)), \quad (4.6)$$

avec les notations (4.2). Ainsi, le problème se ramène à résoudre un ensemble de problèmes d'augmentation de résolution de signaux unidimensionnels, à savoir  $\hat{h}_{i,j} = \arg \min_h J_{i,j}(h)$  pour tout  $i, j$ , où

$$J_{i,j}(h) = \sum_{r=0}^{R-1} \left| B^r h - \ell_{i,j}^r \right|^2 + \beta \sum_k \phi(h(k) - h(k-1)).$$

Parmi les nombreuses approches possibles pour estimer  $\hat{h}$ , en se basant sur l'étude comparative de Nikolova et Ng<sup>[59]</sup>, nous avons choisi la forme additive de la régularisation semi-quadratique, comme j'avais pu l'utiliser dans le cadre de ma thèse également. L'algorithme est donc de calculer de façon itérative  $(h^n, b^n)$  comme suit :

$$\begin{aligned} h^n &= H^{-1}(2B^T \ell_{ij} + \beta D^T b^n) \quad \text{où } H = 2B^T B + \beta D^T D, \\ b_k^n &= \begin{cases} 0 & \text{si } |h^{n-1}(k) - h^{n-1}(k-1)| \leq \alpha \\ h^{n-1}(k) - h^{n-1}(k-1) - \alpha \operatorname{sign}(h^{n-1}(k) - h^{n-1}(k-1)) & \text{sinon} \end{cases} \quad \forall k. \end{aligned}$$

La suite  $(h^n, b^n)$  converge vers  $(\hat{h}, \hat{b})$  tel que  $\hat{h}$  est la solution recherchée. Remarquons que la matrice  $H$  ne doit être inversée qu'une seule fois (avant de commencer la minimisation pour tous les  $i, j$ ), et que le calcul de  $b^n$  est extrêmement simple.

[59] M. Nikolova and M. Ng. Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning. In *Proceedings of the International Conference on Image Processing*, volume 1, pages 277–280. IEEE Signal Processing Society, 2001.

- Une originalité de ce travail vient de la façon dont l'approche a été validée. S'agissant de séquences d'images fonctionnelles, il était important de savoir dans quelle mesure l'augmentation spatiale de résolution de chacune des images permettait une meilleure détection des activités quand on en regarde la succession temporelle de ces images. Après traitement des données, les volumes ont donc été analysés avec le logiciel SPM ( [www.fil.ion.ucl.ac.uk/spm/](http://www.fil.ion.ucl.ac.uk/spm/)) et nous avons comparé les activités obtenues à partir : (i) des volumes basse résolution, (ii) des volumes haute résolution reconstruits, (iii) des volumes haute résolution mesurés (voir l'illustration figure 4.4 (c) et [J 2], [C 4] pour plus de détails). En particulier, à partir d'activités simulées, nous avons montré comment cette approche permet une meilleure localisation et séparation de l'activité.

*Pour conclure cette section, je souhaiterais insister sur la partie expérimentale de cette contribution, qui a nécessité une forte interaction avec le partenaire du département de radiologie. Il a fallu appréhender un volume important de données complexes et de nombreuses expérimentations ont été nécessaires avant d'en arriver à la formulation présentée dans ce document. Ainsi, par exemple :* 

- *Considérer des décalages en  $x, y$  n'apporte pas d'information nouvelle, les informations étant acquises dans ce plan.*
- *Dans l'acquisition des volumes successifs, il faut s'assurer que le sujet n'a pas trop bougé, et enlever l'image si c'est le cas. Il est besoin de recalculer l'ensemble des données après reconstruction pour en faire l'analyse temporelle.*
- *Régulariser dans les trois dimensions s'est avéré intéressant mais a tendance à trop lisser les activités, et nous avons donc choisi une régularisation en  $z$  uniquement, ce qui permet de plus une implémentation très rapide.*

## Restauration et analyse des images SAR

*Ce travail correspond au travail de thèse de Caroline Lacombe (UNSA), co-encadrée par Gilles Aubert (UNSA) et Laure Blanc-Féraud (INRIA, projet ARIANA), avec qui j'ai collaboré.*

L'interférométrie avec des radars à synthèse d'ouverture (SAR) est devenue d'un grand intérêt pour l'obtention de mesures topographiques. Partant d'images d'une même scène prise avec deux angles légèrement différents, le produit interférométrique fournit l'interférogramme et une carte de cohérence.

- L'interférogramme représente la différence de phases entre deux acquisitions, notée  $\varphi$ , et contient uniquement l'information pour déterminer la hauteur du terrain. Cette image de phase est structurée, elle contient des franges et possède des valeurs dans l'intervalle  $[-\pi, \pi]$  (voir figure 4.5 (a)). L'intensité en chaque pixel représente donc la valeur principale de la phase déroulée. Pour reconstruire la géométrie en chaque point de l'image, il faut donc retrouver la phase  $\phi$ . C'est l'étape du déroulement de phase, qui consiste à estimer le nombre de cycles  $k(x, y)$  en chaque point  $(x, y)$  de l'image, tel que :

$$\phi(x, y) = \varphi(x, y) + 2k(x, y)\pi, \quad \text{où } \varphi(x, y) \in [-\pi, \pi[.$$

Les discontinuités présentes dans la phase rendent le déroulement de phase difficile. En outre la modification de certains facteurs d'une image à l'autre introduit une différence importante dans les chatoiements. Une des difficultés de l'étape de déroulement provient du haut niveau de bruit introduit dans la phase enroulée, ce qui induit des erreurs dans la mesure de la hauteur. Il est donc nécessaire de filtrer les images interférométriques avant de les dérouler.

- La cohérence, notée  $\rho$ , est une mesure de la stabilité de la phase entre deux images radar acquises à partir de deux angles de vue différents. L'intensité des pixels dans l'image de cohérence varie entre 0 et 1. Cette image permet d'identifier les zones où les variations de phase entre les acquisitions sont corrélées pour deux pixels voisins. En pratique, plus la cohérence est élevée, plus l'interférogramme est exploitable. En effet, si la nature du sol n'a pas changé entre les deux acquisitions et que les angles d'incidence de prise de vue sont très proches, alors une forte cohérence indique que les deux images sont identiques, y compris au niveau de la distribution statistique du chatoiement. En revanche, lorsque la cohérence prend des valeurs faibles (proches de zéro), cela signifie que des changements sont apparus entre les deux prises de vues, par exemple si la variation entre les deux angles de vue est trop importante. Les images de cohérence fournies dans le produit interférométrique ont généralement un aspect granuleux.

Le travail entrepris a donc comporté deux parties, à savoir comment lisser l'image des phases en prenant en compte la cohérence, puis comment dérouler l'image de phases.

- Dans la première partie décrite dans [C 6], inspirés de l'approche de restauration de Lee et al. [49], nous avons proposé une approche à base d'EDP permettant un lissage préservant au mieux la structure des images, en prenant compte la cohérence. L'approche de Lee et al. comportait déjà ces deux ingrédients, à savoir:

- La prise en compte de la spécificité du bruit interférométrique. La variance du bruit sur la phase dépend de la cohérence, qui représente le meilleur estimateur de qualité de la phase extraite du produit interférométrique.
- L'utilisation d'une fenêtre directionnelle le long des franges, pour lisser le long d'une frange et pas dans la direction orthogonale aux franges.

Notre contribution a été de reformuler cette approche dans le cadre des EDP. Pour cela, compte-tenu de la structure des images de phase, nous sommes partis des approches de diffusion proposées par Weickert [75] qui permettent de définir, via le tenseur de structure, une direction de diffusion traduisant localement les structures de l'image. En effet, le tenseur de structure est défini par

$$J_\mu(\nabla\varphi_\sigma) = K_\mu * \nabla\varphi_\sigma \nabla\varphi_\sigma^T, \quad (4.7)$$

où  $\varphi_\sigma$  est une version lissée de l'image  $\varphi$  avec une variance  $\sigma$ , et  $K_\mu$  est un noyau de convolution de variance  $\mu$ . L'intérêt de cette matrice définie positive est qu'elle possède des éléments propres qui s'interprètent en terme des structures de l'image. Le vecteur propre principal  $w_1$  indique la direction normale aux structures locales. Le vecteur propre  $w_2$  indique la direction de plus faible variation des contrastes. Il est important de remarquer que les deux paramètres  $\sigma$  et  $\mu$  jouent deux rôles bien distincts.

- $\mu$  est un paramètre d'échelle d'intégration qui correspond à la taille caractéristique de la structure. Les valeurs propres décrivent le contraste moyen dans la direction des vecteurs propres à l'intérieur d'un voisinage de taille  $O(\mu)$ .
- $\sigma$  est un paramètre lié au bruit qui rend le descripteur insensible aux détails d'échelle plus petits que  $O(\sigma)$ . C'est à ce niveau qu'intervient la cohérence. Après avoir proposé un lissage de la cohérence par la variation totale, nous avons utilisé la relation décrite dans Lee et al. entre  $\sigma$  et  $\rho$  (voir figure 4.5 (b)).

Partant de (4.7), on a donc la décomposition

$$J_\mu(\nabla\varphi_\sigma) = \lambda_1 w_1 w_1^T + \lambda_2 w_2 w_2^T,$$

où  $\lambda_1, \lambda_2$  sont les valeurs propres. A partir de la connaissance de ces éléments propres, on peut définir des opérateurs de diffusion permettant de préserver la structure des images.

$$\begin{cases} \partial_t \varphi = \text{div}(D(J_\mu(\nabla\varphi_\sigma))\nabla\varphi) & \text{dans } \Omega \times (0, \infty), \\ \langle D(J_\mu(\nabla\varphi_\sigma))\nabla\varphi, n \rangle = 0 & \text{sur } \partial\Omega \times (0, \infty), \\ \varphi(x, 0) = \varphi_0(x) & \text{dans } \Omega, \end{cases} \quad (4.8)$$

où  $D$  est le tenseur que nous avons à définir en fonction des éléments propres de  $J_\mu$ ,  $\langle \cdot, \cdot \rangle$  désigne un produit scalaire, et  $n$  est la normale à  $\partial\Omega$ . Nous faisons référence à [C 6] pour plus de détails.

- Dans la seconde partie décrite dans [C 3], nous avons proposé une approche variationnelle originale pour le déroulement de phase.

[49] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller. Intensity and phase statistic of multilook polarimetric and interferometric SAR imagery. *IEEE Trans. Geo. Rem. Sens.*, 32(5):1017–1028, 1994.

[75] J. Weickert. On discontinuity-preserving optic flow. In *Proc. Computer Vision and Mobile Robotics Workshop*, pages 115–122, Santorini, September 1998.

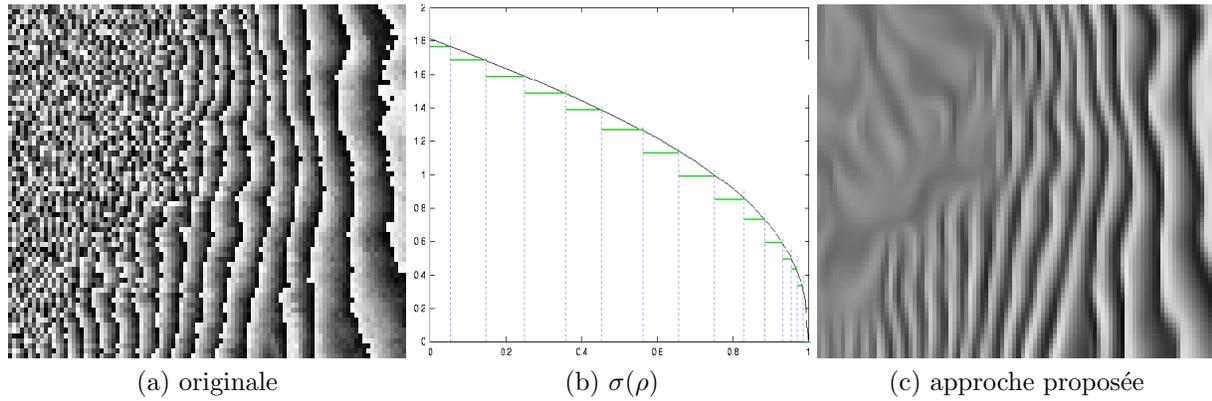


Fig. 4.5: Restauration et analyse des images SAR : (a) exemple d'image SAR, (b) variance  $\sigma$  de la phase en fonction de la cohérence lissée  $\rho$ . Du point de vue numérique, la carte de cohérence filtrée  $\rho$  est d'abord quantifiée, afin d'approcher  $\sigma$  avec une fonction constante par morceaux, (c) résultat obtenu avec notre approche [C 6].

### 4.1.3 Autres applications en vision par ordinateur

#### Structure spatio-temporelle des séquences d'images et mouvement

Toujours intéressé par la thématique mouvement, j'ai travaillé sur l'estimation du mouvement par l'analyse de la structure spatio-temporelle des séquences d'images, vues comme des volumes 3D, le temps étant la troisième dimension. En chaque point d'un volume, l'idée est d'évaluer un tenseur de structure à partir du gradient spatio-temporel de la séquence, et d'en analyser les propriétés. Le tenseur de structure, nous l'avions déjà introduit en 2D pour la restauration d'images SAR (voir l'équation (4.7)). L'objet à étudier et à manipuler est donc un volume de tenseurs, comme c'était le cas pour les collègues de l'équipe travaillant sur les tenseurs de diffusion en DTMRI<sup>[29]</sup>.

Soit une séquence d'images  $u(x_1, x_2, t) : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$  donnée, et  $\sigma(x_1, x_2, t) : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}^2$  le champ de vitesses instantané au temps  $t$  que l'on cherche. Une hypothèse classique consiste à supposer l'intensité constante le long des trajectoires, soit  $u(x_1(t), x_2(t), t) = cte$ , ce que l'on peut dériver formellement pour obtenir la classique contrainte du flot optique (CFO)

$$\nabla u \cdot \sigma + u_t = \nabla_{xt} u \cdot \tilde{\sigma} = 0, \quad (4.9)$$

où  $\tilde{\sigma} = (\sigma_1, \sigma_2, 1)^t$ . Cette équation scalaire ne permet que d'obtenir la composante  $\sigma_\perp$ , normale aux isophotes, i.e., parallèle à  $\hat{n} = \nabla u / |\nabla u|$ , c'est le problème d'ouverture, évoqué plusieurs fois dans ce manuscrit. De nombreuses approches ont été proposées pour incorporer une connaissance *a priori* sur la solution, comme par exemple la prise en compte de critères de régularité sur la solution, afin de restreindre le champ des  $\sigma$  admissibles (voir par exemple mes travaux de thèse [TH 3], [TH 4], chapitre 3).

Une méthode alternative est de regarder ce qu'indique l'ensemble des CFO dans un voisinage. En effet, on s'attend à ce que les structures variées des images indiquent des ensembles de directions possibles, et que leur intersection donne la vitesse réelle. Plus formellement, Lucas et Kanade<sup>[52]</sup> ont fait l'hypothèse que le flot est localement constant sur un voisinage, et ils cherchaient le flot qui vérifie au mieux les équations de CFO dans ce voisinage. Plus précisément, la méthode consiste à résoudre la CFO dans le sens des moindres carrés dans une fenêtre  $W(x) \subset \Omega$  : à une position  $x$  donnée, le flot  $\sigma$  est obtenu comme minimum de l'énergie

$$\min_{\tilde{\sigma}} \int_{W(x)} (\nabla_{xt} u \cdot \tilde{\sigma})^2 dx' = \min_{\tilde{\sigma}} \int_{W(x)} \tilde{\sigma}^t (\nabla_{xt} u) (\nabla_{xt} u)^t \tilde{\sigma} dx' = \min_{\tilde{\sigma}} \tilde{\sigma}^t \left( \int_{W(x)} (\nabla_{xt} u) (\nabla_{xt} u)^t dx' \right) \tilde{\sigma}, \quad (4.10)$$

soit un problème du type

$$\text{Arg min}_{\tilde{\sigma} = (\sigma_1, \sigma_2, 1)^T} \tilde{\sigma}^T S \tilde{\sigma}, \quad (4.11)$$

[29] R. Deriche, D. Tschumperlé, C. Lenglet, and M. Rousson. Variational approaches to the estimation, regularization and segmentation of diffusion tensor images. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, chapter 32, pages 519–532. Springer, 2005 edition, 2005.

[52] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

avec

$$S = \int_{W(x)} (\nabla_{xt}u)(\nabla_{xt}u)^t dx'.$$

L'idée naturelle est alors de considérer le même problème (4.11) mais cette fois avec le tenseur de structure, qui n'est autre qu'une moyenne pondérée de  $(\nabla_{xt}u)(\nabla_{xt}u)^t$

$$S = K_\mu * (\nabla_{xt}u_\sigma)(\nabla_{xt}u_\sigma)^t, \quad (4.12)$$

où  $K_\mu$  est un noyau gaussien par exemple, et  $u_\sigma$  une version lissée de  $u$ . Le message important est que l'on voit ainsi apparaître naturellement le tenseur de structure. L'idée est qu'en moyennant ainsi les différentes CFO, on va intégrer suffisamment d'informations pour connaître le vrai flot.

Résoudre (4.11)–(4.12) peut se voir comme un problème aux moindres carrés, ou bien comme une analyse des éléments propres de  $S$ . En effet, on peut considérer le problème voisin

$$\text{Arg min}_{\mu=(\mu_1,\mu_2,\mu_3)^T} \mu^T S \mu \quad (4.13)$$

où la contrainte  $\mu^T \mu = 1$  est ajoutée pour éviter la solution triviale  $\mu = 0$ . Le flot correspondant sera donné par

$$\sigma = \mu_3^{-1} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}.$$

En incorporant la contrainte à l'aide d'un multiplicateur de Lagrange  $\lambda$ , on se ramène à résoudre le problème de minimisation

$$\text{Arg min}_{\mu=(\mu_1,\mu_2,\mu_3)^T} L(\mu, \lambda) = \mu^T S \mu + \lambda(1 - \mu^T \mu). \quad (4.14)$$

La différentiation de (4.14) par rapport à  $\mu$  conduit à la condition d'optimalité

$$S \mu = \lambda \mu, \quad (4.15)$$

i.e.,  $\lambda$  doit être une valeur propre de  $S$ , et  $\mu$  est son vecteur propre correspondant. Pour un tel couple  $(\mu, \lambda)$  on a  $L(\mu, \lambda) = \lambda$  et donc le minimum de (4.13) est atteint quand  $\lambda$  est la valeur propre la plus faible de  $S$ . Puisque  $S$  est le résultat d'une intégration de tenseurs  $3 \times 3$  positifs semi-définis  $T(x, t) = \nabla_{xt}u \nabla_{xt}u^T$ ,  $S$  est de même nature (symétrique réelle, définie positive) et possède donc trois valeurs propres réelles positives  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ . En fonction des valeurs propres, on arrive ainsi à déterminer quelle est la nature du problème localement<sup>[43]</sup> :

- Un tenseur *stick* ( $\lambda_1 \gg \lambda_2, \lambda_3$ ) correspond au cas où les vecteurs  $\nabla_{xt}u$  sont collinéaires dans un voisinage, c'est-à-dire le problème d'ouverture.
- Un tenseur *plate* ( $\lambda_1, \lambda_2 \gg \lambda_3 \approx 0$ ) correspond au cas où l'on a suffisamment d'information (de structure) pour connaître le vrai flot optique.
- Un tenseur *ball* ( $\lambda_1 \approx \lambda_2 \approx \lambda_3$ ) indique la présence de plusieurs mouvements encodés dans le tenseur, et donc l'impossibilité de dégager le flot optique à l'échelle observée.

Pour mieux caractériser les valeurs relatives des valeurs propres dans les différentes situations, il est d'usage d'introduire les quantités suivantes, comprises entre 0 et 1 :

- La *mesure de cohérence totale* définie par

$$c_t = \left( \frac{\lambda_1 - \lambda_3}{\lambda_1 + \lambda_3} \right)^2. \quad (4.16)$$

Elle représente la confiance en l'estimation du mouvement. Quand  $c_t$  est proche de 1, on peut alors estimer un mouvement qui correspondra soit au flot normal soit au flot optique. Si  $c_t$  est proche de 0 cela indique qu'il n'y a pas de mouvement cohérent dans le voisinage de  $(x, t)$ .

[43] B. Jähne. *Handbook of computer vision, vol. 2 : signal processing and pattern recognition*. Academic press, 1999.

- La *mesure de cohérence spatiale* définie par

$$c_s = \left( \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2. \quad (4.17)$$

Elle symbolise la présence du problème d'ouverture. Une valeur de  $c_s$  proche de 1 signifie que le problème d'ouverture est présent. Dans toutes les autres situations,  $c_s$  sera faible.

- La *mesure de coin*  $c_c$  définie par

$$c_c = c_t - c_s. \quad (4.18)$$

Quand  $c_c$  est proche de 1, il n'y a pas de problème d'ouverture et le flot peut être déterminé sans ambiguïté.

Dans ce type d'approches, le principe est d'estimer la vitesse en intégrant une équation de type CFO sur un voisinage, la contrainte étant en général pondérée par une gaussienne. Cela revient à lisser le champ des tenseurs associés, ce qui peut être fait dans le domaine spatial ou spatio-temporel. Ainsi, minimiser ces énergies conduit à des systèmes algébriques, où les coefficients moyennés des tenseurs de structure sont utilisés.

Dans ce cadre, nous avons apporté deux contributions.

- Si l'on envisage le lissage gaussien comme un processus de diffusion linéaire, l'idée est alors de considérer des méthodes non linéaires pour obtenir des coefficients des tenseurs de structure lissés, ce que nous avons proposé dans [C 9] et [CH 3]. L'idée sera alors d'utiliser les coefficients de ces tenseurs lissés à la place de ceux obtenus par convolution gaussienne. ☐☐☐

En particulier, il paraît important d'avoir une mesure de coin élevée pour pouvoir résoudre le problème d'ouverture. Une solution possible pour augmenter la cohérence est d'utiliser des gaussiennes de variance croissante afin de récupérer de la structure (figure 4.6). Pour l'illustrer, nous montrons aussi dans la figure 4.7 (b), l'évolution de la mesure de coin d'un champ de tenseurs, associé à une séquence en fonction de la variance des gaussiennes. La moyenne commence par augmenter, ce qui signifie que l'on récupère de la structure, avant de diminuer car trop d'informations sont prises en compte. Cette évolution montre donc qu'il est important de savoir stopper la diffusion dès que suffisamment d'informations ont été incorporées dans le tenseur de structure. D'autre part, il semble clair que la taille du voisinage que l'on souhaite prendre en compte doit dépendre de la configuration locale de la scène.

Ces remarques peuvent être transcrites en terme d'EDP, et nous avons proposé de diffuser le champ de tenseurs de structure en utilisant la mesure de coin, selon

$$(S_{ij})_t = \operatorname{div}(g(c_c)\nabla S_{ij}), \quad (4.19)$$

où  $g$  est une fonction continue telle que

$$g_\alpha(s) = \begin{cases} 1 & \text{si } s \leq \alpha - \varepsilon, \\ (s - \alpha + \varepsilon)/(2\varepsilon) & \text{si } \alpha - \varepsilon \leq s \leq \alpha + \varepsilon, \\ 0 & \text{si } s \geq \alpha + \varepsilon, \end{cases} \quad (4.20)$$

où  $\varepsilon = 0.1$  et  $\alpha = 0.9$  dans les expériences. Cette approche rentre dans le cadre des méthodes proposées par Brox et Weickert<sup>[76]</sup>. Notre contribution vient ici de l'introduction de la mesure de coin de manière simple et explicite dans le processus de diffusion. L'interprétation de (4.19)–(4.20) est qu'une diffusion isotrope va se poursuivre jusqu'à ce que de la structure permettant de résoudre la CFO soit trouvée. La différence de comportement de l'évolution de la mesure de coin est illustrée dans la figure 4.7 (b). Cette approche a été comparée à de nombreuses approches existantes et a donné de bons résultats dans de nombreux cas (voir [C 9] pour plus de détails).

- La seconde contribution utilise la méthode générale (4.11)–(4.12), et s'intéresse à une limitation connue pour ce type d'approches : l'estimation des grands déplacements. Dans [C 8], nous avons montré comment une approche multi-échelle *coarse-to-fine* pouvait également être mise en œuvre pour appréhender les déplacements importants, en reprenant des idées utilisées pour les approches variationnelles basées sur la contrainte du flot optique. Ainsi nous avons introduit l'idée de champs de tenseurs compensés en mouvement.

Du point de vue numérique, cette approche a de nombreux avantages. En particulier, il y a très peu de paramètres à régler et le gain de temps est important par rapport aux approches variationnelles traditionnelles, car il n'est pas besoin ici d'itérer. Pour les résultats détaillés, voir [C 9]. ☐☐☐

[76] J. Weickert and T. Brox. Diffusion and regularization of vector and matrix-valued images. Technical report, Universitat des Saarlandes, 2002.

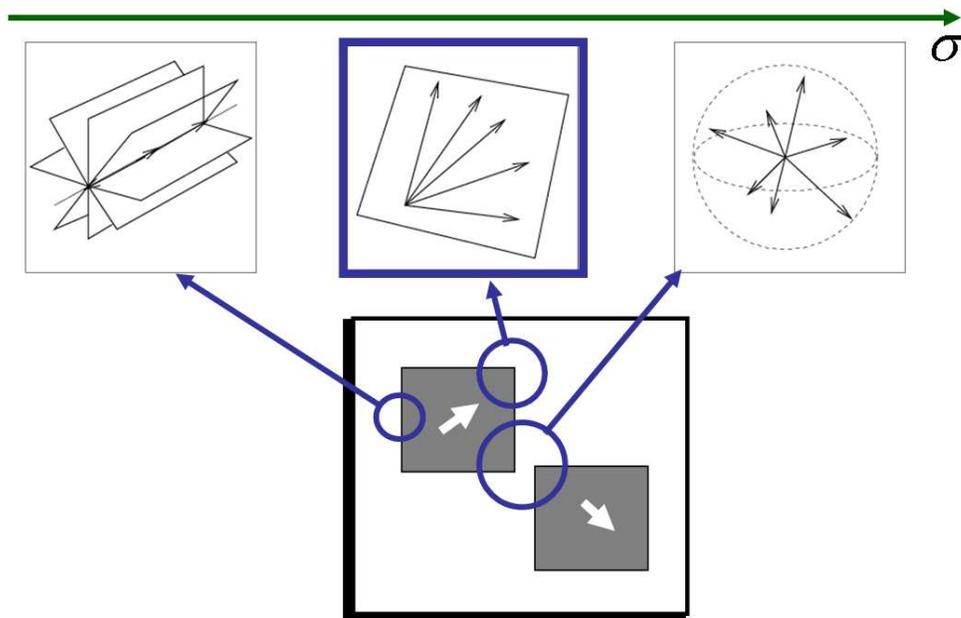


Fig. 4.6: Trouver la bonne échelle pour les tenseurs de structure permet de connaître le flot optique. Dans cette figure on représente les plans contenant  $\nabla_{xt}u$  : la vitesse  $\tilde{\sigma}$  doit être normale à ces plans. Dans le premier cas, avec une variance  $\sigma$  faible et dans un voisinage sans forte courbure, les vecteurs  $\nabla_{xt}u$  sont donc tous identiques. Il existe une infinité de plans contenant  $\nabla_{xt}u$  et donc la vitesse ne peut pas être déterminée de façon unique. Quand  $\sigma$  augmente et qu'on a intégré des informations de structure, alors seul un plan contiendra les  $\nabla_{xt}u$ , permettant de définir le flot optique  $\tilde{\sigma}$ . Enfin, si on intègre dans un voisinage trop grand, on risque de mélanger dans le tenseur des informations non cohérentes, rendant impossible l'estimation du flot optique.

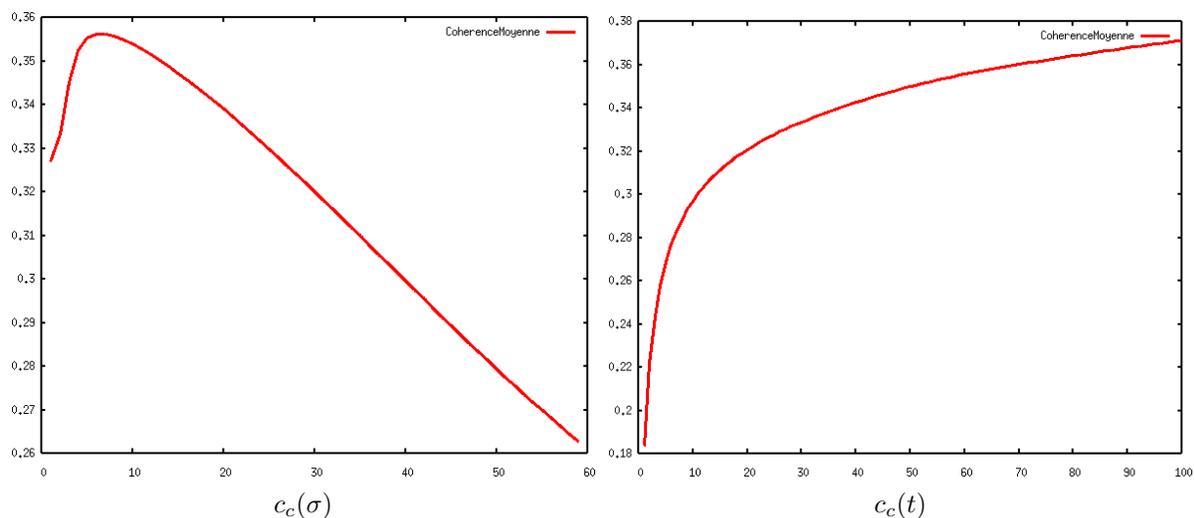


Fig. 4.7: Lisser pour augmenter la mesure de coin et savoir s'arrêter. (a) évolution de  $c_c$  sur des tenseurs de structure avec  $\sigma$  croissant, (b) évolution de  $c_c$  sur les tenseurs estimés par le processus de diffusion par EDP.

Ces résultats m'ont permis de comprendre un autre point de vue sur l'estimation du mouvement. Plutôt que de chercher à régulariser un flot, l'idée était de chercher à comprendre la structure spatio-temporelle d'une séquence d'images. Dans la méthode par EDP que nous avons proposée, on trouve cette idée d'intégrer de l'information autant que nécessaire pour obtenir un flot optique non ambigu. Cette idée générale d'intégration et de propagation des attribus non ambigus, je la retrouverai quelques années plus tard sous l'éclairage d'expériences psychophysiques (voir la section 4.2, page 53). 

## Désoccultation d'images non texturées

La désoccultation d'images (ou *inpainting*) correspond aux techniques de restauration d'images qui permettent de remplir un domaine  $\Omega$  manquant dans une image, étant données les intensités au bord du domaine (voir l'exemple figure 4.8 (b) : les zones en rouge sont à retrouver). On distingue généralement deux situations, faisant appel à des classes d'approches différentes.

- Quand la région à retrouver ne contient pas de textures, ce qui a d'abord été considéré dans la littérature, le problème d'*inpainting* peut se voir comme savoir prolonger au mieux les isophotes qui arrivent au bord du domaine. Cela se décrit bien avec les méthodes par EDP (voir <sup>[7]</sup> et **[LI 2]** pour un état de l'art).
- Quand la région à retrouver contient de la texture, prolonger des isophotes devient impossible, et de nombreuses approches ont récemment été proposées, basées sur la technique de génération de texture d'Efros et Leung<sup>[31]</sup>. Il s'agit d'une méthode de "copier-coller", où l'on cherche à compléter  $\Omega$  par les bouts de texture de l'image les plus adaptés.

Nous avons considéré ici le cas des images géométriques. Le point de départ de nombres approches par EDP est le travail de Nitzberg, Mumford et Shiota (1993) qui ont proposé d'analyser la structure 3D de la scène (profondeur relative des objets) au travers de l'analyse des jonctions en "T" qui sont les points où une frontière coupe une ligne de niveau de  $u_0$ . Partant de ce travail, Masnou et Morel<sup>[55]</sup> ont proposé une méthode pour relier les jonctions en "T" compatibles. En considérant l'ensemble de ces courbes, on peut écrire le problème sous la forme d'un principe variationnel, qui en fait est une généralisation du modèle *elastica* proposé Nitzberg, Mumford et Shiota (1993). Partant de cette idée, on peut également mentionner les travaux de Ballester et al. <sup>[2,3]</sup>, qui ont proposé une interpolation couplée de la direction normale aux isophotes (notée  $\theta$ ) et des niveaux de gris (notés  $u$ ). Le champ de vecteurs  $\theta$  est associé à l'image  $u$  selon  $\theta \cdot \nabla u = |\nabla u|$  et  $|\theta| \leq 1$ . Le problème est donc de résoudre

$$\begin{aligned} \min_{u, \theta} \int_{\Omega} |\operatorname{div}(\theta)|^p (a + b|\nabla k * u|) dx, \\ |\theta| \leq 1, \theta \cdot \nabla u - |\nabla u| = 0 \text{ dans } \Omega, \\ u = u_0 \text{ in } B, \theta \cdot n = \theta_0 \cdot n \text{ sur } \partial\Omega, \end{aligned}$$

où  $p > 1$ ,  $a > 0$ ,  $b \geq 0$ ,  $k$  est un noyau de régularisation de classe  $C^1$  avec  $k(x) > 0$  a.e., et  $n(x)$  est la normale unitaire sortante en  $x \in \partial\Omega$ . L'information est la valeur  $u_0$  sur  $\partial\Omega$  et le champ de vitesses  $\theta_0$  des normales aux courbes de niveau de  $u_0$ .

Il existe également de nombreux modèles ne venant pas d'un principe variationnel. Par exemple, celui proposé par Bertalmio et al. <sup>[8]</sup> est basé sur l'idée de propager de l'information depuis les frontières vers

- 
- [7] M. Bertalmio, V. Caselles, G. Haro, and G. Sapiro. *PDE-based image and surface inpainting*, chapter 3, pages 33–62. Springer, 2005.
- [31] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the 7th International Conference on Computer Vision*, volume 2, pages 1033–1038, Kerkyra, Greece, 1999. IEEE Computer Society, IEEE Computer Society Press.
- [55] S. Masnou and J.M. Morel. Level lines based disocclusion. *International Conference on Image Processing*, III:259–263, 1998.
- [2] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing*, 10(8):1200–1211, 2001.
- [3] C. Ballester, V. Caselles, and J. Verdera. Disocclusion by joint interpolation of vector fields and gray levels. *Multiscale Modeling and Simulation*, 2:80–123, 2003.
- [8] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In Kurt Akeley, editor, *Proceedings of the SIGGRAPH*, pages 417–424. ACM Press, ACM SIGGRAPH, Addison Wesley Longman, 2000.

l'intérieur via une équation de transport, en suivant les isophotes du Laplacien (détecteur de contraste). L'équation du troisième ordre obtenue est<sup>2</sup>

$$\frac{\partial u}{\partial t}(x, y, t) = \nabla(\Delta u(x, y, t))^\perp \cdot \nabla u(x, y, t), \quad (4.21)$$

pour tout  $(x, y) \in \Omega$ , avec la condition initiale  $u(x, y, 0) = u_0^{ext}(x, y)$  où  $u_0^{ext}(x, y)$  est une extension continue de la donnée  $u_0$  (connu seulement sur  $D - \Omega$ ) et  $u(x, y, t) = u_0(x, y)$  pour tout  $(x, y) \in \partial\Omega$  et  $t \geq 0$ .

L'idée qui ressort de ces quelques exemples est le rôle prépondérant des isophotes. Dans les approches variationnelles citées, reconstruire les isophotes est le but essentiel. Dans les approches de type EDP, pour éviter cette reconstruction, des équations diffusant l'intensité vers l'intérieur du domaine ont été proposées. Dans [R 6], nous avons proposé une méthode simple pour estimer les directions des tangentes aux isophotes, suivie d'une reconstruction explicite des niveaux de gris par intégration des trajectoires. Ces deux étapes sont ici découplées. En voici la description :

- Tout d'abord nous estimons la direction des tangentes aux isophotes à reconstruire par une méthode de diffusion des directions définies aux bords de  $\Omega$  (figure 4.8 (c)). Pour cela, définissons tout d'abord

$$\begin{cases} \partial\Omega^{\text{Dir}} = \{(x, y) \in \partial\Omega / |\nabla u_0(x, y)| > 0\}, \\ \partial\Omega^{\text{Neu}} = \partial\Omega - \partial\Omega^{\text{Dir}}. \end{cases}$$

On a alors deux cas.

- Cas 1 : Si  $\partial\Omega^{\text{Dir}} = \emptyset$ , ce qui correspond à la région 3 de la figure 4.8 (b), cela signifie que  $|\nabla u_0(x, y)| = 0$  pour tout  $(x, y) \in \partial\Omega$ , soit  $u_0 = \text{cte}$  sur  $\partial\Omega$ . Dans une telle situation, il n'est plus nécessaire d'estimer des isophotes, remplir le domaine par cette valeur est suffisant.
- Cas 2 : Si  $\partial\Omega^{\text{Dir}} \neq \emptyset$ , nous proposons d'obtenir un champ de tenseurs régularisé  $T$ , qui représente les directions tangentes, en diffusant à l'intérieur de  $\Omega$  les directions connues aux bords. Le problème est donc de déterminer les conditions de bords. Pour  $(x, y)$  dans  $\partial\Omega^{\text{Dir}}$ , où des contrastes sont présents, des conditions de Dirichlet sont imposées, de sorte que  $T$  est défini par

$$T^{\text{Dir}}(x, y) = \begin{cases} \mathcal{F} \left( \frac{\nabla u_0^\perp}{|\nabla u_0^\perp|}(x, y) \right) & \text{si } |\nabla u_0|(x, y) > 0 \\ 0 & \text{(le tenseur nul), sinon} \end{cases} \quad (4.22)$$

où

$$\mathcal{F}(n) = \begin{pmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{pmatrix}, \quad (4.23)$$

ce qui revient en fait à faire un calcul de type tenseur de structure : l'élément propre principal de la matrice va correspondre au vecteur gradient, l'avantage étant qu'on ne considère désormais que la direction (il est plus facile de moyenniser des directions que des vecteurs). Dans (4.22), remarquons que les gradients sont normalisés, ce qui permet aux isophotes reconstruites de ne pas dépendre du contraste local. Ce type de conditions de bords est illustré figure 4.8 (d). Des conditions de bords de type Neumann sont imposées sur  $\partial\Omega^{\text{Neu}}$ . Pour trouver le champ de tenseurs, on a proposé de résoudre l'équation

$$\begin{cases} \frac{\partial T}{\partial t} = \Delta T & (x, y) \in \Omega, t > 0, \\ T(x, y, t) = T^{\text{Dir}}(x, y) & \forall t \geq 0, (x, y) \in \partial\Omega^{\text{Dir}}, \\ \frac{\partial T}{\partial n}(x, y, t) = 0 & \forall t, (x, y) \in \partial\Omega^{\text{Neu}}, \\ T(x, y, 0) = T^0(x, y), & (x, y) \in \Omega, \end{cases} \quad (4.24)$$

où  $T^0(x, y)$ , la condition initiale, est une matrice de  $S^2$  (ensemble des matrices  $(2, 2)$  symétriques non négatives) satisfaisant  $T^0(x, y) = T^{\text{Dir}}(x, y)$  pour tout  $(x, y) \in \partial\Omega^{\text{Dir}}$ , et  $\frac{\partial T}{\partial n}$  est la dérivée directionnelle le long de la normale à  $\partial\Omega$ . On peut prouver qu'avec des conditions initiales

<sup>2</sup>Remarquons que dans leur papier original, les auteurs n'ont pas exactement défini (4.21), mais ont écrit dans le terme de droite  $\nabla(\Delta u) \cdot \nabla u^\perp$  ce qui a une signification différente, mais revient au même à convergence, i.e., quand  $\frac{\partial u}{\partial t}(x, y, t) = 0$ , puisque pour tout vecteur  $V$  et  $W$  on a  $V^\perp \cdot W = -V \cdot W^\perp$ .

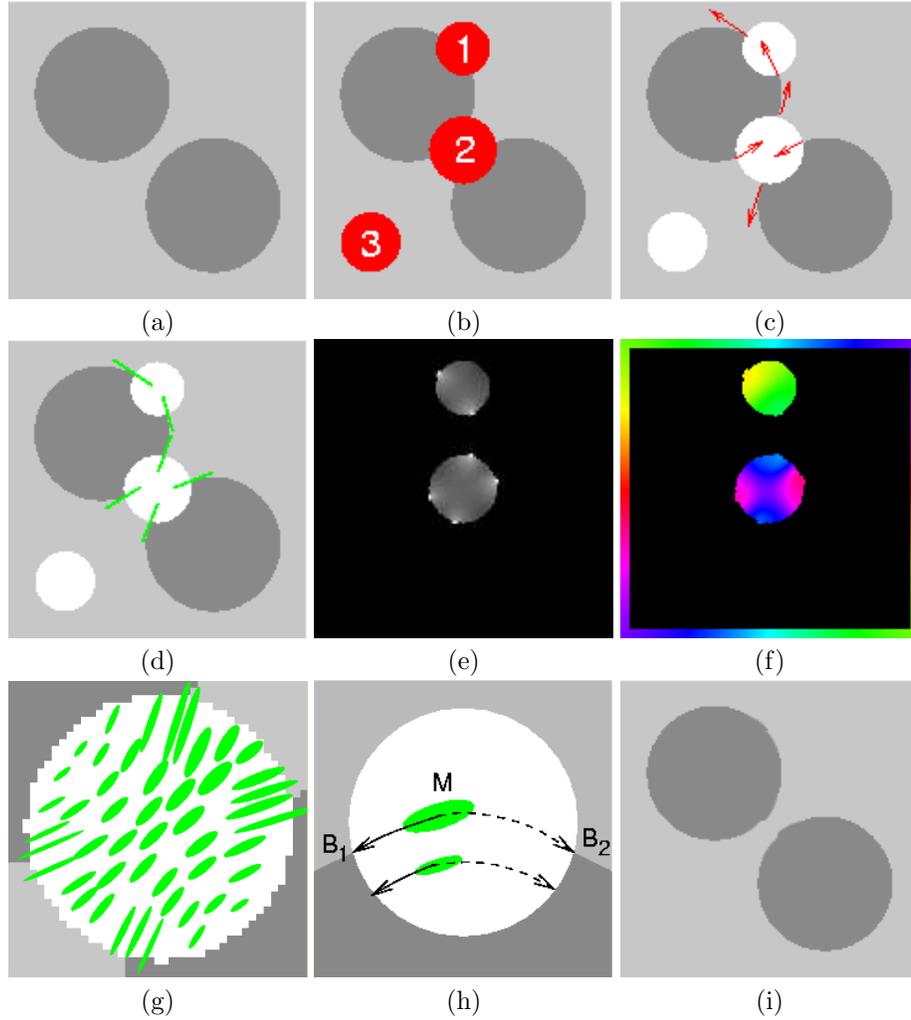


Fig. 4.8: Illustration de l'approche (a) image originale à retrouver, (b) image avec  $\Omega$  indiqué en rouge, i.e., les régions à retrouver, (c) tangentes aux isophotes sur  $\partial\Omega$ , (d) tenseurs définis comme conditions de bords correspondant aux tangentes, (e) salience des tenseurs obtenus après diffusion ( $\lambda_{max} - \lambda_{min}$ ), (f) direction des tenseurs obtenus après diffusion (code couleur indiqué aux bords), (g) illustration du champ de tenseurs pour la région 2, (h) schéma pour l'intégration des trajectoires, (i) résultat obtenu.

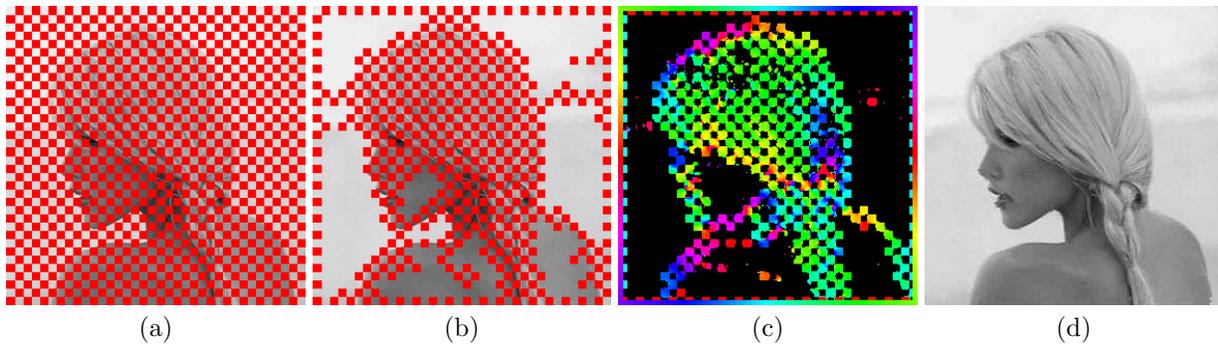


Fig. 4.9: (a) image originale, où les zones en rouge sont à reconstruire, (b) les composantes de  $\Omega$  pour lesquelles aucune courbe de niveau n'arrive sont d'abord remplacées, (c) orientations des matrices après la diffusion, (d) résultat obtenu.

et de frontières appartenant à  $S^2$ , la solution de (4.24) sera dans  $S^2$  pour tout  $t > 0$ . Un exemple de résultat est illustré dans les figures 4.8 (e), (f) et (g). On observe que les matrices ont une forte salience ( $\lambda_{max} - \lambda_{min}$  élevé) près des frontières, et que la direction principale est lissée correctement. En (g), on représente les matrices reconstruites par des ellipsoïdes (cela correspond à la région 2).

- Ensuite, nous retrouvons les niveaux de gris en intégrant le long des trajectoires pour trouver l'intensité correspondant au bord du domaine (figure 4.8 (h) et (i)). Ces trajectoires correspondent aux courbes intégrales associées aux directions propres principales du champ de tenseurs.

Soit  $T$  un champ de tenseurs. Si l'on note  $V : \Omega \rightarrow \mathbb{R}^2$  le champ de vecteurs collinéaires à la direction principale du champ de tenseurs  $T$ , alors l'idée est de regarder la trajectoire  $X(t) = (x(t), y(t))$ :

$$\frac{\partial X}{\partial t}(t) = V(X(t)), \quad X(0) = M. \quad (4.25)$$

Alors, pour  $M = (x, y)$  dans  $\Omega$ , l'idée est de définir  $u(x, y)$  par

$$u(X(t)) = u_0 \left( M + \int_0^{\bar{t}} V(X(t)) dt \right), \quad (4.26)$$

où  $u_0$  est l'intensité connue sur  $D - \Omega$  (et en particulier sur  $\partial\Omega$ , bord de  $\Omega$ ), et  $\bar{t}$  le temps tel que le point  $\left( M + \int_0^{\bar{t}} V(X(t)) dt \right)$  appartient à  $\partial\Omega$ . Alors  $u$  est constant le long des trajectoires. En particulier  $u(M) = u(X(0))$  et

$$\frac{d}{dt} u(X(t)) = \nabla u(X(t)) \cdot V(X(t)) = 0 \text{ pour tout } t,$$

et pour  $t = 0$  on a

$$\nabla u(M) \cdot V(M) = 0.$$

Cette dernière équation signifie que les isophotes de l'image reconstruite sont tangentes au champ de vecteurs  $V$ . Du point de vue numérique, cette intégration le long des trajectoires peut être implémentée par une méthode simple d'Euler, ou de façon plus précise par une méthode de Runge-Kutta.

*Cette approche ne nécessite que très peu de paramètres à régler et nous avons illustré ses performances sur plusieurs exemples classiquement utilisés dans la littérature (voir notamment la figure 4.9).* 

## Méthode générique pour la résolution des problèmes variationnels

De nombreux problèmes de vision par ordinateur peuvent s'écrire comme un problème variationnel dont l'énergie comporte un terme d'attache aux données et un terme de régularisation. Par exemple, le problème de la restauration d'une image  $u_0$  peut s'écrire

$$\inf_u F(u) = \int_{\Omega} |u - u_0|^2 dx + \lambda \int_{\Omega} |\nabla u|^p dx, \quad (4.27)$$

où le premier terme indique que  $u$  doit rester proche de  $u_0$ , et le second terme est une pénalisation sur la norme du gradient dépendant de  $p$ . Par exemple :

- Si  $p = 2$ , on retrouve une pénalisation de type Tikhonov et on s'attend donc à trouver des solutions très régulières. Du point de vue espace fonctionnel, la solution est dans un espace de Sobolev, ne permettant pas de discontinuités le long de courbes.
- Si  $p = 1$ , on retrouve une pénalisation de type variation totale (*total variation, alias TV*), et on obtient alors des solutions régulières par morceaux, avec une préservation des contours. Du point de vue espace fonctionnel, la solution est dans l'espace des fonctions à variations bornées (*bounded variations, alias BV*), et l'étude théorique, mais aussi l'implémentation, deviennent alors plus complexes.

- Si  $p \gg 1$ , l'intérêt est d'approcher les solutions du Laplacien infini, problème numérique difficile.

Le terme de régularisation définit donc notamment quel est l'espace fonctionnel adéquat pour que le problème soit bien défini. En fonction de cet espace, il n'est pas toujours possible de trouver une solution numérique au problème, par exemple dans le cas des fonctions à variations bornées où l'on ne peut pas écrire d'équation d'Euler.

Pour répondre à ce problème, nous avons proposé dans [R 9], [S 1], une approche unificatrice qui permet de traiter toute une classe d'approches variationnelles, à savoir □□□

$$\inf_{u \in W^{1,p}(\Omega)} F(u) = \int_{\Omega} h(x, u(x)) dx + \int_{\Omega} |\nabla u(x)|^p dx, \quad (4.28)$$

avec la même méthodologie ( $p$  quelconque). Notre approche est basée sur la caractérisation des espaces de Sobolev par Bourgain, Brezis et Mironescu<sup>[10]</sup>, qui permet d'approcher les normes dans les espaces de Sobolev par une suite d'opérateurs intégraux, comprenant un quotient différentiel et une suite adaptée de noyaux régularisants radiaux :

$$\lim_{n \rightarrow \infty} \int_{\Omega} \int_{\Omega} \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy = K_{N,p} \int_{\Omega} |\nabla u|^p dx dy.$$

Cette approximation a également été étendue dans le cas des fonctions à variations bornées<sup>[66]</sup>, ce que nous avons utilisé par la suite.

Notre première contribution a été d'utiliser cette caractérisation pour définir la suite de formulations variationnelles

$$\inf_u F_n(u) = \int_{\Omega} h(x, u(x)) dx + \int_{\Omega} \int_{\Omega} \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy. \quad (4.29)$$

Nous avons alors montré l'existence et l'unicité d'une solution d'un minimum de (4.29), puis que la suite des minima de  $F_n$  converge vers la solution du problème variationnel original (4.28). En fait, cette approximation est valide pour tout  $p \geq 1$ , de telle sorte que le cas BV est aussi résolu (grâce aux résultats de Ponce, 2004). Par cette méthode, nous avons donc proposé une alternative pour approcher les problèmes variationnels sur  $BV(\Omega)$  de manière indépendante du terme d'attache aux données contrairement aux travaux de Chambolle<sup>[19]</sup> par exemple.

Notre deuxième contribution est liée à la résolution numérique du problème (4.29). L'équation d'Euler associée, est l'équation intégrale

$$\frac{\partial h(x, u_n(x))}{\partial u} + 2p \int_{\Omega} \frac{|u_n(x) - u_n(y)|^{p-2}}{|x - y|^p} (u_n(x) - u_n(y)) \rho_n(x - y) dy = 0. \quad (4.30)$$

Il s'agit donc de s'avoir discrétiser le terme intégral, noté  $I_{u^k}(x)$ , qui contient un noyau singulier (pour un  $x$  donné, la fonction n'est pas définie quand  $y = x$ ). A cause de cette singularité des schémas classiques de type différences finies sont difficiles à mettre en œuvre. En s'inspirant d'approches développées dans le cadre de la MEG/EEG où des termes similaires sont à approcher<sup>[25]</sup>, nous avons réécrit  $I_{u^k}(x)$  sous la forme

$$I_{u^k}(x) = \int_{\Omega} \frac{J_{u^k}(x, y)}{|x - y|} dy, \quad (4.31)$$

avec

$$J_{u^k}(x, y) = \frac{|u(x) - u(y)|^{p-2} (u(x) - u(y)) \rho_n(|x - y|)}{|x - y|^{p-1}},$$

puis, nous avons

- 
- [10] J. Bourgain, H. Brezis, and P. Mironescu. Another look at sobolev spaces. *Optimal Control and Partial Differential Equations, In honour of Professor Alain Bensoussan's 60th Birthday*, pages 439–455, 2001. J. L. Menaldi, E. Rofman, A. Sulem (eds).
- [66] A. Ponce. A new approach to sobolev spaces and connections to  $\gamma$ -convergence. *Calculus of Variations and Partial Differential Equations*, 19(3):229–255, 2004.
- [19] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, March 2004.
- [25] Eric Darve. *Méthodes multipôles rapides: résolution des équations de Maxwell par formulations intégrales*. PhD thesis, Université de Paris 6, 1999.

- Discrétisé l'espace en utilisant une triangulation, où les nœuds sont ici les pixels.
- Interpolé linéairement la fonction  $J_{u_k}(x, y)$  sur chaque triangle ( $x$  étant fixé).
- Calculé de manière explicite  $\int 1/|x - y|dy$  et  $\int (x - y)/|x - y|dy$  sur chaque triangle, ce qui est nécessaire pour le calcul de (4.31).

Comme exemple d'application, nous avons choisi le problème de la restauration d'images (4.27) en variant le paramètre  $p$ .

*Cette approximation est intéressante mais a un temps de calcul élevé par rapport à des méthodes type différences finies pour les cas réguliers. On pense possible en revanche de mettre en œuvre des méthodes plus efficaces, en utilisant la similarité entre ce type de terme intégral et le filtre bilatéral pour lequel des méthodes numériques très efficaces existent<sup>[21]</sup>. Une autre possibilité à explorer est l'utilisation des méthodes particulières (voir également la section 4.3). Pour conclure, l'intérêt de cette contribution est bien de proposer une méthode d'approximation générale pour toute une classe de problèmes.* 

## 4.2 Modélisation bio-inspirée de tâches visuelles

Cadre contractuel : projet Européen FACETS ("Fast Analog Computing with Emergent Transient States", EC IP project FP6-015879, 2005–2009). FACETS est un grand consortium entre plusieurs communautés scientifiques visant à développer les connaissances sur le système nerveux.

### Modèle de rétine biologique avec contrôle de gain

*Le travail décrit dans cette section correspond au travail de thèse d'Adrien Wohrer (2004–2007) que je co-encadre avec Thierry Viéville.*

Dans son travail décrit dans [C 16], [R 7], Adrien Wohrer a développé une architecture qui possède un lien fort avec l'anatomie des rétines biologiques, tout en restant un modèle et donc une abstraction. Il s'agit donc de comprendre et simuler la transformation effectuée dans la rétine d'un signal lumineux (une vidéo dans notre cas) en un train d'impulsions. La rétine est composée de plusieurs couches, elle est hautement connectée, et possède une grande variété de types de neurones. Les principales caractéristiques de sa structure sont présentées dans la figure 4.10.

Une grande difficulté de ce travail, a été dans l'analyse de l'importante littérature sur la rétine,<sup>3</sup> dont les mécanismes à ce jour ne sont pas encore tous bien connus, et font l'objet de différentes hypothèses comme nous le verrons. En terme de modèle en revanche, on note un certain consensus autour des modèles de type "linéaire, non-linéaires"<sup>[16]</sup> et leurs extensions, qui combinent une étape de filtrage linéaire, suivie d'une non-linéarité statique, et de la génération d'impulsions. Il s'agit essentiellement de modèles phénoménologiques, c'est-à-dire qu'ils cherchent à reproduire des comportements en sortie, sans s'intéresser à la modélisation précise des différentes composantes.

Dans notre contribution, les objectifs étaient de (i) mettre chaque étape en rapport avec la physiologie, (ii) proposer mécanisme générique de contrôle de gain, (iii) développer un simulateur permettant de réaliser des expériences à grande échelle (jusqu'à 100.000 neurones en sortie). Le système proposé modélise les cinq couches de la rétine. La figure 4.11 les illustre et récapitule les formules proposées. En voici quelques commentaires.

- De nombreuses convolutions spatiales (notées  $\overset{x,y}{*}$ ) ou temporelles (notées  $\overset{t}{*}$ ) apparaissent dans le

<sup>3</sup> Voir par exemple la page html très complète Webvision, sur l'organisation de la rétine et du système visuel

[21] J. Chen, S. Paris, and F. Durand. Real-time edge-aware image processing with the bilateral grid. In *Proceedings of the ACM SIGGRAPH'07 conference*, ACM Transactions on Graphics, 2007.

[16] M. Carandini, J. B. Demb, V. Mante, D. J. Tollhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597, November 2005.

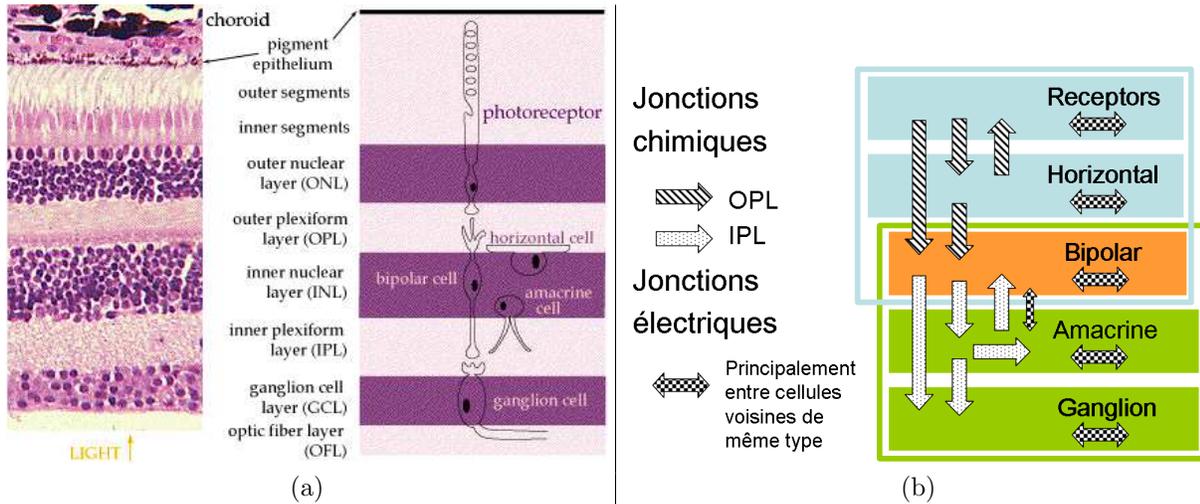


Fig. 4.10: Structure de la rétine. (a) Coupe de la rétine et représentation simplifiée des principaux types de cellules qui vont être modélisés. On remarque en particulier les deux couches, appelées OPL et IPL, siège des connexions synaptiques entre les cellules. (b) Schéma représentant les différentes connexions entre cellules (chimiques et électriques). Seule la couche des cellules ganglionnaires va produire des impulsions, les autres couches étant caractérisées par l'évolution continue de leur potentiel. La rétine, c'est aussi une grande variété de types de cellules. Certaines cellules de la rétine sont "atypiques" : par exemple les cellules bipolaires, qui joueront un rôle particulier dans le système, en possédant des dendrites dans l'OPL et l'IPL, et en se projetant dans l'IPL aussi. Pour d'autres cellules, on ne connaît pas encore précisément leur fonction, comme par exemple les cellules amacrines.

modèle, avec les noyaux

$$G_{\sigma}(x, y) = \exp(-(x^2 + y^2)/(2\sigma^2))/(2\pi\sigma^2), \quad (4.32)$$

$$E_{\tau}(t) = \exp(-t/\tau)/\tau, \text{ si } t > 0, 0 \text{ sinon}, \quad (4.33)$$

$$E_{n,\tau}(t) = (nt)^n \exp(-nt/\tau)/((n-1)!\tau^{n+1}), \text{ si } t > 0, 0 \text{ sinon}, \quad (4.34)$$

$$T_{w,\tau}(t) = \delta_0(t) - wE_{\tau}(t). \quad (4.35)$$

Le noyau (4.32) est un noyau gaussien qui va donc permettre un filtrage passe-bas standard. Les filtres exponentiels (4.33) et (4.34) sont des filtres causaux qui vont permettre un filtrage temporel passe-bas. Notons que le filtre (4.34) atteint son maximum à  $t = \tau$ , et le paramètre  $n$  va servir à faire varier sa forme. Quant au filtre (4.35), il s'agit d'un filtre temporel passe haut, reproduisant ainsi des phénomènes observés dans la rétine à plusieurs échelles. Cet effet provient par exemple des mécanismes cellulaires internes (tels que l'adaptation).

- l'*outer plexiform layer* (OPL) contient les connexions entre les récepteurs, les cellules horizontales et les cellules bipolaires. Comme l'ont montré des mesures faites dans les années 60<sup>[78]</sup>, l'OPL serait l'endroit où se créent les interactions antagonistes centre-périphérie. La plupart des modèles de rétine suivent cette hypothèse.

Classiquement, on peut modéliser l'OPL par un filtrage spatio-temporel, comme suggéré par l'expérimentation<sup>[32]</sup>. Spatialement, l'OPL revient à faire des convolutions par des différences de gaussiennes (DOG). Temporellement, l'OPL revient à faire une convolution par des différences d'exponentielles : son action est biphasique. En sortie des cellules horizontales, on a donc un courant exciteur  $I_{OPL}$ , passe-bande, qui rend compte des variations spatiales mais aussi temporelles du signal.

Dans le modèle proposé, une caractéristique importante est que  $I_{OPL}$  est obtenu par une convolution spatio-temporelle non séparable, ce qui a été également observé expérimentalement<sup>[32]</sup>. En fait, le signal périphérique est plus retardé que le signal central, ce qui permet de détecter des changements

[78] F. S. Werblin and J. E. Dowling. Organization of the retina of the mudpuppy. *Journal of Neurophysiology*, 32(3):339–355, 1969.

[32] C. Enroth-Cugell, J. G. Robson, D. E. Schweitzer-Tong, and A. B. Watson. Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *J Physiol*, 341:279–307, 1983.

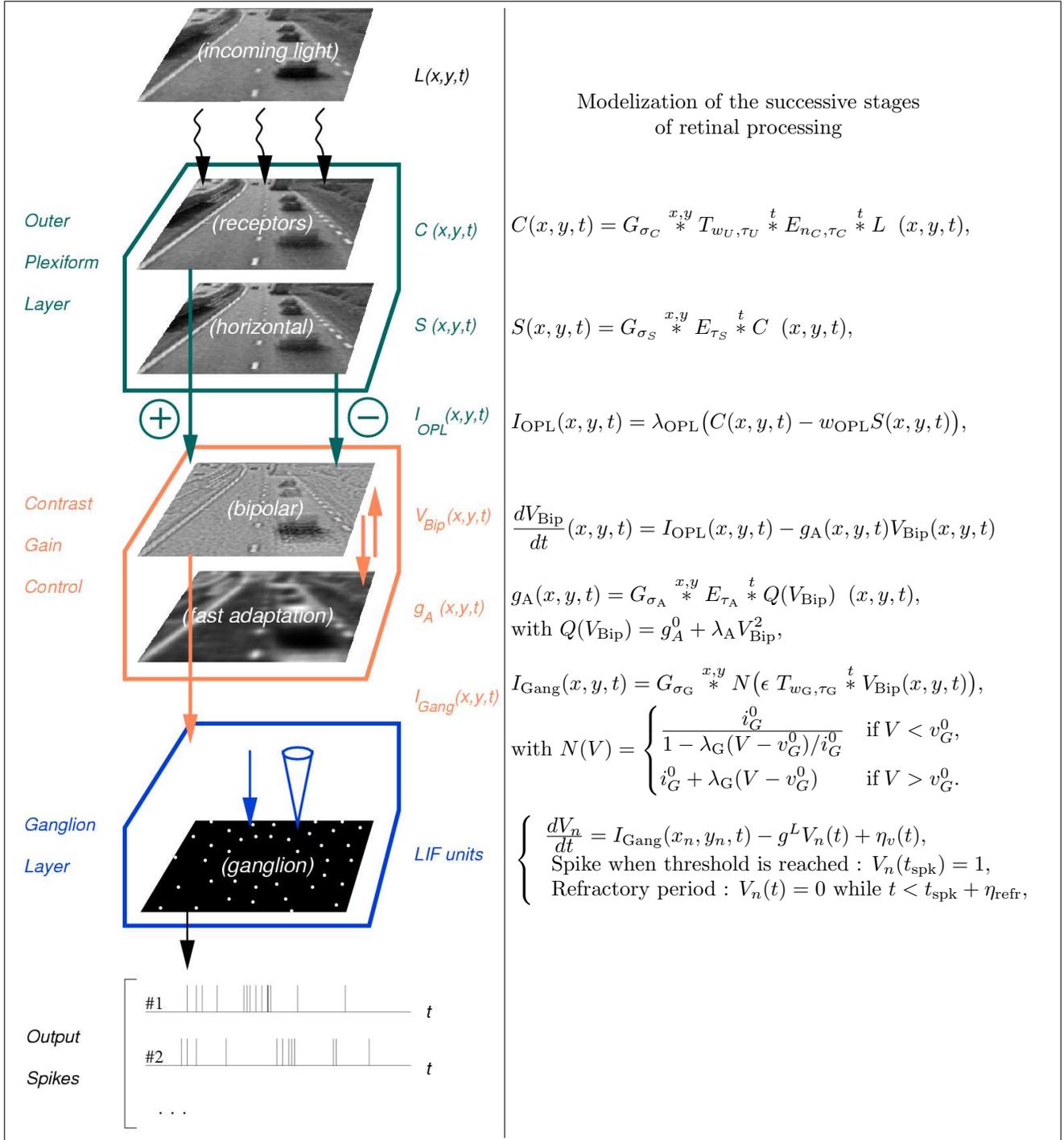


Fig. 4.11: Vue d'ensemble du simulateur de rétine

de luminosité même en l'absence de contraste local. Cela ne serait pas possible avec un filtre séparable, i.e., une simple combinaison d'une convolution spatiale par une DOG et d'une convolution temporelle par une différence d'exponentielles.

- L'*inner plexiform layer* (IPL) contient les connexions entre les cellules bipolaires, amacrines et ganglionnaires. Les cellules bipolaires jouent un rôle particulier en intégrant de l'information venant de l'OPL (via  $I_{OPL}$ ) et des cellules amacrines.

A ce niveau de la rétine, apparaît un phénomène observé en sortie de rétine, à savoir une modulation des réponses des cellules ganglionnaires en fonction du contraste local. On désigne ce phénomène par le contrôle de gain par le contraste, ou *contrast gain control* (CGC). L'idée est que des variations dans l'entrée se traduisent par des variations d'amplitude dans la sortie, dépendant du contraste. Parmi les premiers à observer ce phénomène, nous avons considéré les expériences de Shapley et Victor<sup>[69]</sup> qui ont effectué des mesures sur des cellules ganglionnaires X centre-ON de chat anesthésié. Le stimulus était un *pattern* sinusoïdal fixe modulé temporellement par une somme de sinusoïdes

$$L(x, y, t) = \bar{L} \left( 1 + Gr(x, y) \sum_{i=1}^8 c_i \sin(\xi_i t) \right), \quad (4.36)$$

où  $Gr(x, y)$  est le pattern sinusoïdal. Les  $\xi_i$  sont huit fréquences temporelles qui couvrent en échelle logarithmique l'espace des fréquences (de 0,2Hz à 32Hz). Dans les expériences, les enregistrements effectués avec  $\{c_i\}_{i=1..8}$  donnés, sont analysés dans l'espace de Fourier pour chaque fréquence conduisant ainsi à huit amplitudes et huit phases. Les résultats d'une expérience où  $c_i = c$  sont donnés figure 4.12.

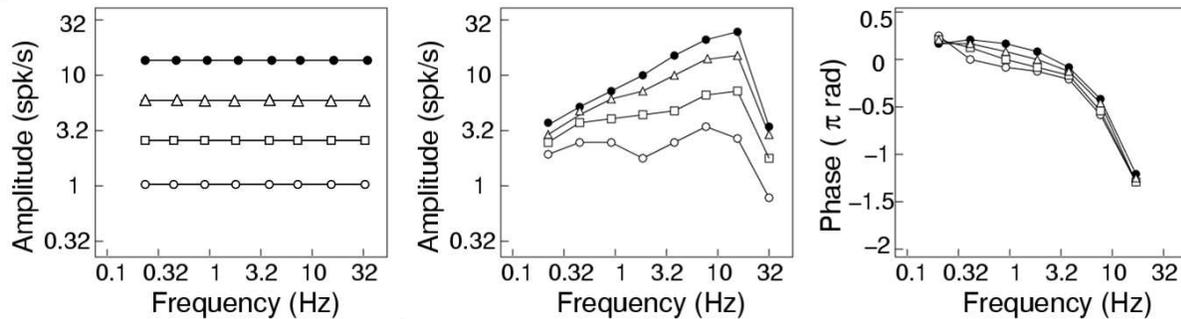


Fig. 4.12: Mesures de cellules ganglionnaires de chat centre ON, suivant l'expérience de Shapley et Victor (stimulus décrit en (4.36)) : (a) amplitude des entrées pour les quatre stimuli, (b) et (c) sont l'amplitude et la phase des réponses en sortie.

Chaque courbe correspond à un contraste doublé, ce qui permet de mettre en évidence le CGC : en l'absence de CGC, toutes les courbes d'amplitude de sortie devraient être parallèles et espacées de  $\log(2)$ , tandis que les courbes de phase devraient être confondues. Dans la figure 4.12 on observe que (i) les cellules répondent de manière sous-linéaire pour les basses fréquences (l'écart est inférieur à  $\log(2)$ ), (ii) pour les forts contrastes, il y a une avance de phase.

De nombreuses hypothèses ont été avancées pour expliquer ce phénomène. Par exemple certains ont proposé que le CGC proviendrait d'une action des cellules amacrines dont le rôle n'est pas bien identifié. Récemment, des expériences menées par Beaudoin et al. <sup>[5]</sup> montrent que le CGC aurait pour origine les propriétés des connexions présynaptiques des cellules bipolaires, sans aucune intervention des cellules amacrines. Il a aussi été mis en évidence deux types de CGC : un lent et un rapide. C'est le CGC rapide que nous avons cherché à modéliser.

Une contribution originale a été d'avoir modélisé le CGC rapide par une équation d'évolution qui va moduler le potentiel  $V_{Bip}$  au moyen d'une conductance de fuite dépendant d'une activité

[69] R. M. Shapley and J. D. Victor. The effect of contrast on the transfer properties of cat retinal ganglion cells. *The Journal of Physiology*, 285(1):275–298, 1978.

[5] D.L. Beaudoin, B.G. Borghuis, and J.B. Demb. Cellular basis for contrast gain control over the receptive field center of mammalian retinal ganglion cells. *Journal of Neuroscience*, 27:2636–2645, 2007.

moyenne locale. Cette conductance est elle-même calculée en fonction de  $V_{\text{Bip}}$  avec une dépendance quadratique<sup>4</sup>. Ce modèle a été mis en relation avec d'autres modèles de contrôle de gain existant. Par ce mécanisme, nous avons ainsi reproduit les enregistrements décrits dans la figure 4.12 ainsi que d'autres expériences.

- La couche des cellules ganglionnaires (de l'ordre d'un million chez les primates), qui va intégrer le courant  $I_{\text{Gang}}$  pour générer des trains d'impulsions. Il y a une grande variété de cellules ganglionnaires dont les fonctions ne sont pas toutes connues, et dont la dénomination dépend des espèces. Chez le primate, deux types en particulier ont été plus étudiés, les cellules *Midget* et *Parasol*.
  - Les cellules Midget (correspondant aux cellules X ou "toniques" identifiées chez le chat) sont les plus nombreuses (75 à 90%) : très précises spatialement (petits champs récepteurs, connectées à un sol cône à la fovéa), elles codent pour les oppositions de couleurs. Elles ne présente pas beaucoup d'effet de CGC<sup>[47]</sup>. Elles sont connectées à la voie Parvocellulaire du LGN, supposée en charge de la détection précise des formes. La vitesse de conduction est faible (environ 20m/s).
  - Les cellules Parasol (correspondant aux cellules Y ou "phasiques" identifiées chez le chat) ont des champs récepteurs plus larges, ne sont pas sensibles à la couleur mais sont très sensibles au contraste et présentent un fort effet de CGC. Elles sont connectées à la voie Magno-cellulaire du LGN, supposée en charge de la détection du mouvement et de l'analyse grossière de la scène. La vitesse de conduction est élevée (environ 40m/s).

Il semblerait que les cellules Midget constituent une nouvelle voie de traitement de l'information visuelle possédée uniquement par les primates, alors que les cellules Parasol sont un trait commun partagé par de nombreux mammifères.

Dans le modèle proposé, les deux types de cellules peuvent être simulées, excepté que pour les cellules Midget les oppositions de couleurs n'ont pas été considérées. Les cellules ganglionnaires peuvent être réparties sur la surface de la rétine en prenant en compte les caractéristiques physiologiques connues (fovéa, dépendance de la densité et du champ récepteur des cellules en fonction de l'excentricité).

De nombreuses expériences ont été réalisées, comme par exemple dans la figure 4.13 : dans cette simulation, les zones de fortes luminosité commencent à apparaître, les contrastes apparaissant ensuite quand la modulation par la périphérie des cellules horizontales entre en jeu.

*La rétine est donc loin d'être un simple écran d'enregistrement passif du flux lumineux qui lui parvient, à l'imitation du tapis de diodes d'un caméscope. Il s'agit d'un véritable centre de traitement de l'information, qui, à partir de l'analyse poussée de la lumière qui lui parvient, élabore un signal complexe codé en fonction des capacités de réception des centres supérieurs. Plusieurs canaux séparés traitent vraisemblablement des différentes caractéristiques de l'image, par exemple la perception des formes, l'information colorée, la détection des mouvements et l'organisation spatiale.*



*Le simulateur dont nous venons de décrire les principes, appelé Virtual Retina, a été déposé en 2007 en code ouvert avec la licence CeCillC (voir aussi l'Appendice A pour plus d'informations sur le simulateur). Nous espérons qu'il sera un outil utile à la communauté pour la réalisation de simulations à grande échelle, à des fins de recherche ou pédagogiques.*

**Q** <http://www-sop.inria.fr/odyssee/software/virtualretina>  
 APP logiciel Virtual Retina : IDDN.FR.001.210034.000.S.P.2007.000.31235

<sup>4</sup>Remarque : la dépendance non-linéaire de la transmission synaptique sur le voltage est une hypothèse raisonnable du point de vue bio-chimique. Cela a été utilisé par exemple dans le modèle d'Izhikevich, réduction du modèle d'Hodgkin et Huxley. Ce comportement quadratique pourrait aussi s'apparenter à l'existence de canaux ioniques qui ne s'ouvrent que pour les forts potentiels (*voltage gated*). Dans le modèle, si on remplace la fonction quadratique par la fonction paire  $Q(V) = 0$  si  $0 \leq V \leq S$  et  $Q(V) = \alpha(V - S)$  si  $V \geq 0$ , alors on observe des comportements similaires quelque-soit la pente  $\alpha$ .

[47] Ehud Kaplan and Ethan Bernadete. The dynamics of primate retinal ganglion cells. *Progress in Brain Research*, 134:1–18, 2001.



Fig. 4.13: Représentation de l'activité en sortie de rétine. Le stimulus, représenté sur la première ligne, est une séquence d'images de bruits de même luminosité moyenne que l'image *Lenna* présentée ensuite. Dans cet exemple, la fovéa occupe tout l'image : à chaque pixel correspond une cellule ganglionnaire. Le niveau de gris dans la deuxième ligne représente le nombre d'impulsions dans des fenêtres temporelles de 5ms.

### Modèles dynamiques des aires V1/MT et catégorisation de mouvement

*Le travail décrit dans cette section correspond au travail de thèse de María-José Escobar (2006–2009) que je co-encadre avec Thierry Viéville.*

Le point de départ de mes recherches a été les travaux autour de Simon Thorpe sur la catégorisation ultra-rapide d'images. Dans ses travaux, Simon Thorpe a montré que la classification d'images statiques peut être réalisée par le cortex visuel avec des latences d'environ 150 ms<sup>[71]</sup> et même plus vite. Pourtant, si l'on considère les temps de latence de la voie visuelle<sup>[60]</sup>, de tels temps de réponse ne peuvent s'expliquer que par une architecture et des mécanismes de calculs spécifiques.

Comme explication à cette extraordinaire performance, l'idée du codage par rang a été proposée<sup>[72,35]</sup> : l'information neuronale serait codée par l'ordre relatif des temps d'impulsions. Les neurones les plus excités seraient donc aussi les premiers à émettre leurs impulsions. En regardant les temps d'arrivée des impulsions, on a donc une idée du message contenu dans le train d'impulsions. Plus généralement, c'est une théorie du traitement rapide de l'information visuelle qui a été proposée<sup>[73]</sup>, confirmant des intuitions des années 60 selon lesquelles des informations pertinentes seraient véhiculées par les impulsions individuelles<sup>[64]</sup> (et pas seulement par les fréquences de décharge). Depuis, cette idée a été confirmée par de nombreuses autres expériences psychophysiques. Preuve de son efficacité, elle permet de faire des modèles capables d'effectuer des tâches de reconnaissances complexes<sup>[28]5</sup>.

Cette interprétation du codage neuronal étant particulièrement efficace, notre objectif a été de comprendre comment nous pourrions l'étendre dans le cas des séquences d'images et de la reconnaissance de mouvements biologiques<sup>6</sup>. Ce fut l'objectif du travail présenté dans [C 11], où nous avons expérimenté

<sup>5</sup>L'idée du codage par rang est aussi à la base de la société *Spikenet technology* qui propose plusieurs produits d'analyse d'images.

<sup>6</sup>Dans VanRullen, Thorpe (2002), les auteurs abordent l'extension de leur théorie quand on a un flux continu en entrée. L'idée repose sur des mécanismes de "réinitialisation du temps", permettant de redonner un sens à des temps de latence, regroupant les trains d'impulsion en différentes vagues. Une origine possible d'un tel mécanisme de "réinitialisation" serait l'existence des rythmes oscillatoires corticaux ou sous-corticaux, qui définiraient des intervalles de temps séparés dans lesquels les latences pourraient être interprétées.

[71] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[60] L.G. Nowak and J. Bullier. *The Timing of Information Transfer in the Visual System*, volume 12 of *Cerebral Cortex*, chapter 5, pages 205–241. Plenum Press, New York, 1997.

[72] S.J. Thorpe. Spike arrival times: A highly efficient coding scheme for neural networks. *Parallel processing in neural systems and computers*, pages 91–94, 1990.

[35] J. Gautrais and S. Thorpe. Rate coding vs temporal order coding : a theoretical approach. *Biosystems*, 48:57–65, 1998.

[73] R. VanRullen and S. J. Thorpe. Surfing a spike wave down the ventral stream. *Vision Research*, 42:2593–2615, 2002.

[64] D. H. Perkel and T. H. Bullock. Neural coding. *Neurosciences Research Program Bulletin*, 6:221–348, 1968.

[28] A. Delorme and S. Thorpe. Face processing using one spike per neuron: resistance to image degradation. *Neural Networks*, 14:795–804, 2001.

le principe du codage par rang de manière simple dans cette application. Etant donnée une séquence d'images centrées sur l'action observée, la méthode comportait trois étapes.

- Un codage éparsé de chaque image basé sur un algorithme de poursuite gourmande (*matching pursuit*<sup>[54]</sup>), comme présenté par Perrinet et al. <sup>[65]</sup> (voir aussi les travaux d'Olshausen<sup>[61]</sup>). Etant donné un dictionnaire de filtres  $\mathcal{D} = \{w_i\}$  (par exemple des filtres de Gabor), et une image  $u^{(0)}$ , on calcule l'ensemble des produits scalaires

$$C_i^0 = \langle u^{(0)}, w_i \rangle = \sum_x u^{(0)}(x)w_i(x), \quad (4.37)$$

on sélectionne alors le filtre  $i^0$  donnant la plus forte réponse

$$i^0 = \underset{i}{\operatorname{argmax}} |C_i^0|,$$

et on définit alors la nouvelle image

$$u^{(1)} = u^{(0)} - \frac{\langle u^{(0)}, w_{i^0} \rangle}{|w_{i^0}|^1} w_{i^0}.$$

On peut alors recommencer le calcul des coefficients (4.37) sur  $u^{(1)}$ . En fait, ce calcul peut-être évité en mettant à jour les coefficients par

$$C_i^1 = \langle u^{(1)}, w_i \rangle = C_i^0 - C_{i^0}^0 \frac{\langle w_{i^0}, w_i \rangle}{|w_{i^0}|^2}. \quad (4.38)$$

On décompose ainsi le signal en enlevant à chaque fois la partie qui vient d'être codée : cela permet d'obtenir un codage éparsé (*sparse coding*). En d'autres termes, les interactions latérales inhibent les filtres voisins suivant leur corrélation. Il est important de noter qu'à chaque étape, l'énergie maximale du signal détecté va décroître, le reste perdant au fur et à mesure du contenu.

On peut également utiliser cette décomposition pour reconstruire progressivement un signal

$$v^T = \sum_{t=1}^T \frac{C_{i^t}^t}{|w_{i^t}|^2} w_{i^t}. \quad (4.39)$$

Cette reconstruction progressive peut s'interpréter en terme de codage neuronal de l'information. Un filtre correspondant à un neurone et à son champ récepteur, quand le filtre répond au maximum, c'est-à-dire quand le neurone est activé fortement, on compte sa contribution dans (4.39). Dans le temps, cette reconstruction permet donc de visualiser, au fur et à mesure, quels sont les neurones les plus activés, avec une énergie décroissante : les informations les plus salientes vont être prises en compte les premières. Une telle reconstruction est illustrée dans la figure 4.14 : il est avancé qu'un tel codage intervient dans la transmission du signal de rétine vers le cortex, expliquant nos performances en vision précoce (voir également la figure 4.13 pour comparaison).

- De l'étape précédente, on a donc un ensemble ordonné d'évènements qui codent l'information présente dans la scène. Sur l'ensemble de la séquence, et pour chaque image, nous n'avons donc retenu que les positions des filtres donnant une réponse supérieure à un certain seuil.
- Ensuite, nous avons utilisé une méthode de classification supervisée standard en utilisant ce vecteur de positions comme caractérisant la séquence et donc l'action.

*Testée sur des séquences d'images comportant deux classes, cette approche avait déjà donné des résultats encourageants. L'idée intéressante était qu'en ne considérant que les évènements les plus salients dans les images, cela permettait une reconnaissance satisfaisante des actions (pour le jeu de données considéré).* 

[54] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3414, 1993.

[65] L. Perrinet, M. Samuelides, and S. Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit. *Neurocomputing*, 57:125–134, 2003.

[61] B.A. Olshausen. Learning sparse, overcomplete representations of time-varying natural images. In *IEEE International Conference on Image Processing*, 2003.



Fig. 4.14: Illustration d’une reconstruction d’image par (4.39), avec l’algorithme de poursuite gourmande, suivant l’approche présentée dans Perrinet et al. (2003). La librairie *matlabPyrTools* développée par Simoncelli a été utilisée pour construire et manipuler les pyramides Laplaciennes. Il apparaît qu’avec seulement 1% des coefficients (les 1% des cellules les plus excitées, correspondant à l’avant dernière image), la reconstruction est déjà suffisamment précise pour des tâches de reconnaissance.

*Mais cette approche ouvrait un certain nombre de questions que nous avons abordées par la suite. Tout d’abord, coder individuellement les images quand on considère une séquence, c’est ignorer la dynamique temporelle de la séquence et la causalité : les attributs que l’on extrait par cette méthode n’ont pas de vraie signification temporelle. Ensuite, considérer comme vecteur d’attribut pour la classification un vecteur contenant la position des “neurones les plus activés”, comporte de nombreuses limitations (longueur de la séquence, sensibilité au temps initial, pas d’information sur la vitesse) : on préférerait savoir définir une carte de mouvement qui soit représentative d’une action. Enfin, la vraie question était finalement de déterminer quels attributs utilisés dans le système visuel doivent être modélisés pour cette tâche de reconnaissance. Nous avons donc poursuivi ce travail avec ces nouveaux objectifs*

En particulier, les psychophysiciens ont proposé plusieurs mécanismes cérébraux impliqués dans cette reconnaissance de ces mouvements biologiques. En se basant sur des stimuli simples, montrant simplement des points qui se déplacent en suivant par exemple les jonctions principales d’une personne en train de marcher (voir un exemple dans la figure 4.15), il a été montré comment la forme et l’estimation du mouvement pouvaient aider à la reconnaissance d’une action<sup>[6,17]</sup>. Ces interprétations ont été confirmées par la suite par de nombreuses études en IMR fonctionnelle, montrant quelles aires sont impliquées dans le processus de reconnaissance. Dans une étude récente<sup>[18]</sup>, il apparaît que la connaissance d’une localisation approximative des éléments en déplacement et de leur mouvement est suffisante pour la reconnaissance. Le mouvement joue donc un rôle primordial pour la reconnaissance.



Fig. 4.15: Exemple de stimulus *point light* : les points dans ces images représentent les positions des jonctions principales. Il est remarquable qu’un stimulus aussi simple nous évoque la présence d’une personne en train d’effectuer une action. Ce type de stimulus a été largement utilisé en psychophysique pour expliquer les mécanismes sous-jacents dans l’interprétation du mouvement. Il contient en fait beaucoup d’informations : voir par exemple <http://www.biomotionlab.ca/Demos/BMLwalker.html>

Dans cette perspective, nous avons proposé dans [C 13], [R 10], une méthode plus aboutie de catégorisation supervisée de mouvement (voir figure 4.16). La méthode est supervisée, c’est-à-dire que l’on se donne un ensemble de  $N_s$  séquences d’images

$$I_n : (t, x) \in [0, T_n] \times \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R},$$

- 
- [6] J.A. Beintema and M. Lappe. Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the USA*, 99(8):5661–5663, 2002.
- [17] A. Casile and M. Giese. Roles of motion and form in biological motion recognition. *Artificial Networks and Neural Information Processing, Lecture Notes in Computer Science 2714*, pages 854–862, 2003.
- [18] A. Casile and M. Giese. Critical features for the recognition of biological motion. *Journal of Vision*, 5:348–360, 2005.

contenant une action connue et donnée au système par l'utilisateur (la base d'apprentissage). Le problème que l'on cherche à résoudre est, étant donnée une séquence d'images

$$I : (t, x) \in \mathbb{R}^+ \times \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R},$$

nouvelle, de savoir quelle action est présente dans la fenêtre temporelle  $[t - \delta t, t]$  ( $\forall t$ ) où  $\delta t$  est l'intervalle de temps utilisé par le système pour analyser le mouvement<sup>7</sup>. La méthode proposée comprend trois axes principaux que nous allons détailler : estimation de mouvement, représentation et catégorisation de mouvement, et évaluation sur des bases de données comprenant plusieurs actions.

- La première contribution est de proposer un modèle bio-inspiré de mouvement qui, contrairement aux modèles existants, cherche à reproduire l'activité impulsionnelle des couches V1 et MT. L'inspiration biologique est donc à deux niveaux :
  - D'abord dans l'architecture : nous modélisons une couche qui va estimer localement du mouvement à partir de la séquence d'images  $I$  (couche V1), puis une couche qui va intégrer cette information en temps et en espace (couche MT). Chaque couche correspondra à la donnée de  $N_n$  neurones répartis spatialement sur  $\Omega$ , selon une densité et avec un champ récepteur dépendant de la distance au point de fixation (ici le centre de l'image). En outre, ces neurones seront aussi regroupés en colonnes d'orientations : à une position donnée, on aura un ensemble de neurones avec des sélectivités à l'orientation différentes. Chaque neurone  $i$  sera caractérisé par son potentiel  $v_i$ .
  - Ensuite par la façon de représenter l'activité du réseau : l'activité sera représentée par une suite d'événements discrets (les impulsions), dont on pourra extraire de l'information de plusieurs façons. Pour le neurone  $i$ , on notera  $(t_i^{(1)}, t_i^{(2)}, \dots, t_i^{(f)}, \dots)$  la liste des temps d'émission d'impulsions. Informatiquement, cela demande des outils spécifiques pour effectuer les simulations. De nombreux simulateurs de réseaux de neurones ont été proposés<sup>[11]</sup>. Nous avons choisi le simulateur événementiel développé par Olivier Rochel. Dans notre cadre, l'intérêt conceptuel de choisir un simulateur événementiel est l'efficacité : il n'y a pas de calcul à effectuer tant qu'un neurone reste silencieux, et la seule fonction à connaître est le temps d'émission de la prochaine impulsion. Cela permet de faire avancer l'horloge jusqu'à la prochaine impulsion, en évitant les erreurs d'approximation qu'on aurait eu en discrétisant le temps.

Ainsi, nous avons choisi de modéliser l'activité des neurones des couches V1 et MT par des neurones à conductance intègre-et-tire<sup>[36]</sup>. Pour un neurone  $i$ , l'évolution temporelle de son potentiel  $v_i(t)$  varie selon

$$\begin{cases} \frac{dv_i(t)}{dt} = -G_i^{exc}(t)(v_i(t) - E^{exc}) - G_i^{inh}(t)(v_i(t) - E^{inh}) - g^L(v_i(t) - E^L) + J(t), \text{ tant que } v_i(t) < Th, \\ v_i(t) = 0, t_i^{(k)} = t, k = k + 1, \text{ si } v_i(t) = Th, \end{cases}$$

où la conductance excitatrice  $G_i^{exc}(t)$  ou le courant  $J(t)$  va décrire l'entrée du neurone,  $G_i^{inh}(t)$  va servir notamment à prendre en compte les inhibitions venant des voisins,  $g^L$  correspond aux fuites au niveau de la membrane,  $J(t)$  désigne un courant externe en entrée du neurone, et  $E^{exc}$ ,  $E^{inh}$  et  $E^L$  sont des constantes (appelées potentiels de renversement, valant 0mv, -75mv et 0mV respectivement). Nous devons donc définir  $G_i^{exc}(t)$  et  $G_i^{inh}(t)$  pour les deux couches, V1 et MT.

- Pour V1, nous avons d'abord suivi une démarche classique. L'entrée du neurone va être définie par le courant  $J(t)$  à partir des réponses des cellules simples puis complexes (étape 2).
  - \* Les cellules simples, organisées en colonnes d'orientations. Une colonne d'orientation est définie par un ensemble de cellules (simples), définies à la même position, sensibles aux mêmes fréquences spatio-temporelles, mais chacune étant sélective à une orientation propre. Ces caractéristiques définissent un filtre spatiotemporel  $F_{\theta_i, f_i}$ . La réponse d'une cellule simple est obtenue par simple convolution du signal d'entrée  $I(t, x)$  avec  $F_{\theta_i, f_i}$ .

<sup>7</sup>Hypothèse : on considère pour le moment des séquences dans lesquelles le sujet est centré dans l'image, i.e., suivi dans son mouvement quand il y a déplacement.

[11] R. Brette, M. Rudolph, T. Carnevale, M. Hines, D. Beeman, J. M. Bower, M. Diesmann, A. Morrison, P. H. Goodman, F. C. Harris Jr., M. Zirpe, T. Natschläger, D. Pecevski, B. Ermentrout, M. Djurfeldt, A. Lansner, O. Rochel, T. Vieville, E. Muller, A. P. Davison, S. El Boustani, and A. Destexhe. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of Computational Neuroscience*, 23(3):349–398, 2007.

[36] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

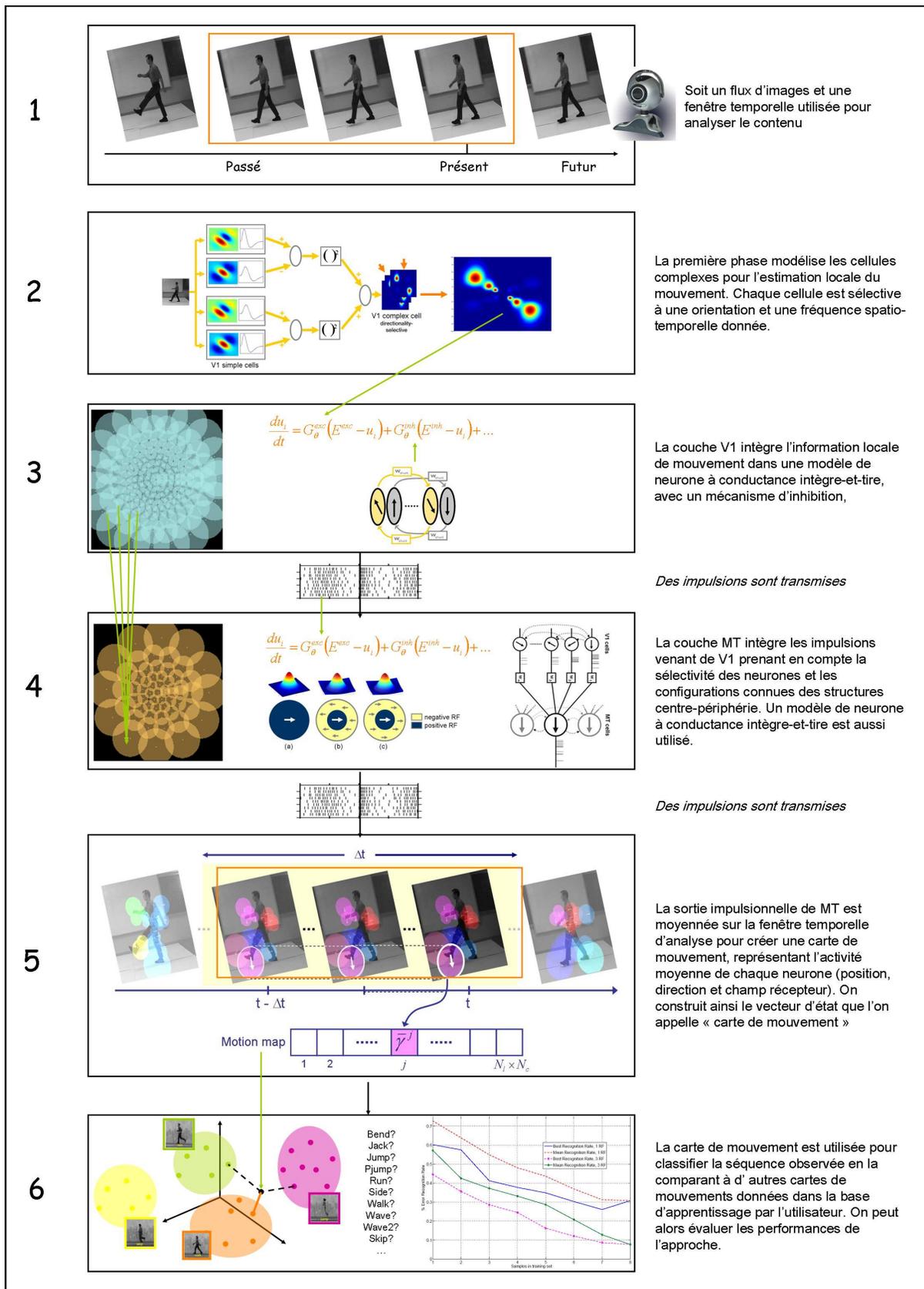


Fig. 4.16: Système de catégorisation de mouvement basé sur des calculs impulsionnels

- \* les cellules complexes, en combinant les réponses des cellules simples suivant un modèle de type énergie<sup>[40]</sup>. Si on note  $C_{\mathbf{x}_i, \theta_i, f_i}(t)$  la réponse de la cellule complexe  $i$  située à la position  $\mathbf{x}_i = (x_i, y_i)$ , sélective à l'orientation  $\theta_i$  et à la fréquence spatio-temporelle  $f_i = (\bar{\xi}_i, \bar{\omega}_i)$ , alors, on a

$$C_{\mathbf{x}_i, \theta_i, f_i}(t) = [(F_{\theta_i, f_i}^a * I)(x_i, y_i, t)]^2 + [(F_{\theta_i, f_i}^b * I)(x_i, y_i, t)]^2, \quad (4.40)$$

où  $*$  représente une convolution spatiotemporelle, et  $F_{\theta_i, f_i}^a, F_{\theta_i, f_i}^b$  sont des filtres en quadrature classiquement utilisés pour modéliser les réponses des cellules simples de V1.

Le courant  $J(t)$  va est défini à partir de l'activité des cellules complexes  $C_{\mathbf{x}_i, \theta_i, f_i}$  par

$$J(t)(t) = k_{exc} \Lambda_i(t) C_{\mathbf{x}_i, \theta_i, f_i}(t),$$

où  $\Lambda_i(t)$  va réaliser une modulation (normalisation) de  $C_{\mathbf{x}_i, \theta_i, f_i}(t)$  de la forme

$$\Lambda_i(t) = \frac{1}{\sum_{j|\theta_j=\theta_i} G_\sigma(|\mathbf{x}_i - \mathbf{x}_j|) C_{\mathbf{x}_j, \theta_j, f_j}(t)},$$

où  $G_\sigma$  est un noyau gaussien. Cette modulation s'interprète comme une inhibition extra-colonne : l'activité du neurone  $i$  sélectif à l'orientation  $\theta_i$  va être normalisée en fonction des neurones voisins sélectifs à la même orientation.

La conductance inhibitrice  $G_i^{inh}(t)$  va traduire l'inhibition intra-colonne entre la cellule  $i$  et une cellule  $j$  de la même colonne d'orientation, mais sélective à la direction opposée ( $\theta_j = \theta_i + \pi[2\pi]$ ). On a

$$G_i^{inh}(t) = w_{shunt} \sum_f \alpha(t - t_j^{(f)}),$$

où  $\alpha$  représente le courant post-synaptique qui peut être modélisé par une exponentielle décroissante<sup>8</sup>.

- Pour MT (étape 4), l'entrée sera les trains d'impulsions de V1.  $G_i^{exc}(t)$  et  $G_i^{inh}(t)$  seront donc définis comme des sommes de potentiels post-synaptiques, en fonction des connexions entre les différents neurones. On distingue :

- \* Les connexions entre V1 et une cellule  $i$  de MT, pouvant être excitatrices ou inhibitrices avec des poids respectifs de  $w_{ij}^+ \geq 0$  et  $w_{ij}^- \leq 0$ . On a

$$G_i^{excV1}(t) = \sum_{j \in V1} w_{ij}^+ \sum_f \alpha(t - t_j^{(f)}), \quad G_i^{inhV1}(t) = \sum_{j \in V1} w_{ij}^- \sum_f \alpha(t - t_j^{(f)}),$$

avec

$$w_{ij} = \begin{cases} k_c w_{cs}(\mathbf{x}_j - \mathbf{x}_i) \cos(\alpha_{ij}) & \text{si } 0 \leq \alpha_{ij} \leq \frac{\pi}{2} \\ 0 & \text{si } \frac{\pi}{2} \leq \alpha_{ij} \leq \pi \end{cases}, \quad w_{ij}^+ = \max(w_{ij}, 0), \quad w_{ij}^- = \min(w_{ij}, 0)$$

où  $\alpha_{ij}$  est la différence entre les directions préférées des cellules  $i$  et  $j$  (en valeur absolue), et  $w_{cs}(x, y)$  est une fonction qui va permettre de décrire les interactions centre-périphérie caractéristiques de cellules de MT, par exemple une différence de gaussiennes. Mais en réalité, il existe une grande variété dans les structures centre-périphéries des cellules de MT<sup>[79]</sup>, et seules 20% des cellules de MT ont une périphérie symétrique, 50% sont asymétriques et concentrent la suppression en seulement une position dans la direction préférée, et 25% ont deux zones de suppression symétriques le long d'un axe. Dans notre cas, nous avons modélisé trois types de cellules : un type avec périphérie excitatrice pour la même vitesse, et deux types de cellules avec la périphérie inhibitrice (par rapport à la même vitesse et la vitesse opposée<sup>[9]</sup>). Comme nous le verrons par la suite, ces trois types permettent déjà d'obtenir des résultats de reconnaissance intéressants.

<sup>8</sup>Cette inhibition intra-colonne interviendrait en fait au niveau de MT selon certaines études, comme Quian et Andersen (1994)

[40] Norberto Grzywacz and A.L. Yuille. A model for the estimate of local image velocity by cells on the visual cortex. *Proc R Soc Lond B Biol Sci.*, 239(1295):129–161, mar 1990.

[79] D. K. Xiao, S. Raiguel, V. Marcar, and G. A. Orban. The spatial distribution of the antagonistic surround of mt/v5 neurons. *Cereb Cortex*, 7(7):662–77, 1997.

[9] R. T. Born. Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*, 84:2658–2669, 2000.

\* Les connexions horizontales, vont également modifier les conductances

$$G_i^{exc_{MT}}(t) = \sum_{j \in MT} \xi_{ij}^+ \sum_f \alpha(t - t_j^{(f)}), \quad G_i^{inh_{MT}}(t) = \sum_{j \in MT} \xi_{ij}^- \sum_f \alpha(t - t_j^{(f)}),$$

où  $\xi_{ij}^\pm$ , définissent les interactions horizontales. Dans nos simulations, nous utilisons seulement une diffusion, c'est-à-dire  $\xi_{ij}^- = G_\sigma(|\mathbf{x}_j - \mathbf{x}_i|)$  et  $\xi_{ij}^- = 0$ .

La dynamique des cellules de MT va donc être gouvernée par les conductances

$$G_i^{exc}(t) = G_i^{exc_{V1}}(t) + G_i^{exc_{MT}}(t), \\ G_i^{inh}(t) = G_i^{inh_{V1}}(t) + G_i^{inh_{MT}}(t).$$

La contribution principale de cette partie est d'avoir proposé une couche de MT complètement impulsioneuse, prenant des trains d'impulsions en entrée, et délivrant des impulsions en sortie. En développant ce modèle, cela devrait nous permettre d'étudier plus finement la dynamique des cellules de MT par rapport à des mesures cellulaires, comme cela a pu être fait dans le travail sur la rétine.

- La deuxième contribution originale est de comprendre comment exploiter des trains d'impulsions pour effectuer la catégorisation des actions présentes dans une vidéo. La difficulté vient ici de l'aspect continu de ces trains d'impulsions qui arrivent comme un flux : on ne peut plus parler directement de codage par rang au sens où il n'y a plus de temps de référence. L'activité d'un neurone au temps  $t$  est le résultat de l'image en cours mais aussi de l'historique.

Notre approche consiste à mesurer une activité moyenne de chaque neurone sur la fenêtre temporelle  $\Delta t$ , ce qui revient à calculer un taux de décharge moyen (étape 5)

$$\gamma_I^j(t, \Delta t) = \frac{1}{\delta t \Delta t} \sum_{i=t-\Delta t}^t \alpha(t - t_j^{(i)}),$$

où  $\gamma^j(t)$  est le taux de décharge de la cellule MT  $j$ , mesuré sur un intervalle de temps  $\Delta t$ , et  $\delta t$  représente l'intervalle de temps entre deux images consécutives. La reconnaissance de l'action qui se déroule dans la fenêtre temporelle  $[t - \Delta t, t]$  va reposer sur l'ensemble de ces mesures d'activités (à une position donnée, pour une orientation donnée, et avec une taille de champ récepteur donnée), qui permettent de définir un vecteur d'état, que nous appellerons carte de mouvement :

$$\mathcal{M}_I(t, \Delta t) = \left\{ \gamma_I^j(t) \right\}_{j=1, \dots, N_{MT}},$$

où  $N_{MT}$  est le nombre total de cellules de MT.  $\mathcal{M}_I(t, \Delta t)$  représente donc l'activité globale des cellules de MT avec une certaine mémoire temporelle.

En revenant à notre problème de catégorisation, le but est donc de savoir si  $\mathcal{M}_I(t, \Delta t)$  correspond à une action apprise dans une séquence  $I_n$  de la base d'apprentissage, décrite par  $\mathcal{M}_{I_n}(t', \Delta t')$ . C'est l'étape de classification (étape 6). Nous avons choisi une classification simple basée sur le plus proche voisin, avec une distance entre cartes de mouvement définie par

$$D(\mathcal{M}_I(t, \Delta t), \mathcal{M}_{I_n}(t', \Delta t')) = \frac{1}{N_{MT}} \sum_{l=1}^{N_{MT}} \frac{\left( \gamma_I^j(t, \Delta t) - \gamma_{I_n}^j(t', \Delta t') \right)^2}{\gamma_I^j(t, \Delta t) + \gamma_{I_n}^j(t', \Delta t')}.$$

- La troisième contribution est d'analyser les principes et les performances de notre approche par rapport à des méthodes développées en vision par ordinateur<sup>[80]</sup>. C'était l'un des objectifs que l'équipe Odyssée s'était fixé, à savoir comment une meilleure compréhension et prise en compte du système visuel pouvait conduire à des approches comparables ou meilleures que celles développées en vision par ordinateur. Cette tendance apparaît de plus en plus dans la communauté de la vision par ordinateur, comme l'attestent un certain nombre de publications récentes (voir par exemple

[80] L. Zelnik-Manor and M. Irani. Event-based analysis of video. In *Proceedings of CVPR'01*, volume 2, pages 123–128, 2001.

[45]). Aussi, dans ce travail, nous avons utilisé les mêmes séquences qu'en vision par ordinateur. Les résultats obtenus sont de très bon niveau et permettent également de justifier certains choix dans le modèle, motivés par la biologie (comme par exemple le fait de choisir plusieurs types de cellules MT).

*Ce travail se poursuit avec plusieurs perspectives, qui vont dans le sens d'améliorer les performances du système, tout en donnant plus de sens biologique au modèle.* 

- *La première perspective est de mieux exploiter les trains d'impulsions : à présent, nous avons dans la dernière étape calculé des fréquences de décharge, si bien que la plupart des calculs pourraient sans doute se faire en analogique, i.e., en continu. L'idée est de chercher à mieux prendre en compte la nature discrète des trains d'impulsions, en regardant par exemple les corrélations pouvant exister entre les différents neurones.*
- *La deuxième perspective est d'introduire des rétroactions dans le système. Une rétroaction possible serait déjà celle de MT vers V1 (voir contribution suivante). Une autre rétroaction à laquelle nous pensons, est une rétroaction de l'aire IT effectuant la reconnaissance, vers MT : l'idée est que quand on commence à reconnaître une action, les neurones dont l'activité ne serait pas en rapport avec cette action seraient inhibés, produisant ainsi une segmentation de la scène. Comme application nous pensons tester notre approche avec des éléments perturbateurs dans la scène, par exemple quand plusieurs actions sont présentes. L'attention pourrait aussi être considérée, comme moyen d'introduire un biais dans le processus de reconnaissance.*

### Calcul de mouvement avec rétroaction et contrôle par la forme

*Le travail décrit dans cette section correspond au travail de stage de Master 2 d'Émilien Tlapale (Master IGMMV, UNSA), dont la thèse a démarré en octobre 2007, que je co-encadre avec Guillaume Masson (INCM).*

Le travail entrepris sur la création de modèles impulsionnels des aires V1 et MT pour la catégorisation de mouvement a révélé plusieurs aspects sur le calcul du mouvement lui-même que nous avons souhaité approfondir. En particulier, nous avons cherché à mieux comprendre les mécanismes impliqués dans l'estimation de mouvement dans les aires V1 et MT, avec la modulation de la forme décrite dans l'aire V2, et les rétroactions de MT vers V1. Le but est de proposer un modèle d'estimation de mouvement prenant en compte ces mécanismes et permettant de reproduire certains percepts. Pour y parvenir, nous nous sommes intéressés aux résultats des études psychophysiques sur la perception du mouvement, et aux données physiologiques des aires corticales impliquées. Ce travail est présenté dans **[R 11]**. 

De la psychophysique sur la perception du mouvement, je souhaite insister sur trois éléments :

- Le local est ambigu, ce qui est aussi bien décrit en vision par ordinateur par le problème d'ouverture. Localement, seule la vitesse normale à un motif peut-être estimée et perçue (figure 4.17 (a)), et on a besoin de l'intervention d'informations non ambiguës pour supprimer l'ambiguïté du mouvement. Dans la figure 4.17 (b), les terminaisons des barres vont permettre de connaître le vrai mouvement : les informations non ambiguës vont se propager pour indiquer le bon mouvement. Cette propagation possède également une dynamique que l'on peut mesurer<sup>[56]</sup>. Si on demande à des sujets de fixer le centre d'une barre en déplacement, on observe d'abord une poursuite dans la direction normale, qui se rectifie ensuite pour suivre la bonne direction du déplacement.
- La diffusion engendre la perception. Comme nous venons de l'évoquer, il semble exister une propagation des indices non ambigus. Ce phénomène est aussi observable avec des "patterns" sinusoidaux en translation, dont la perception du mouvement va dépendre de la forme de l'ouverture au travers de laquelle ils sont observés. Une illustration est donnée dans la figure 4.18. De cette expérience, on pourrait penser que la perception globale du mouvement dépend de la quantité des mouvements non ambigus, les plus nombreux l'emportant.

[45] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In *Proceedings of the 11th International Conference on Computer Vision*. iee-comp-soc, iee-comp-soc-press, 2007.

[56] A. Montagnini, M. Spering, and G. S. Masson. Predicting 2d target velocity cannot help 2d motion integration for smooth pursuit initiation. *J Neurophysiol*, 96:3545–3550, 2006.

- La forme contrôle la diffusion. En fait, la "quantité" des mouvements non ambigus n'est pas une explication complètement satisfaisante, comme l'atteste l'expérience décrite dans la figure 4.19. Dans cette expérience, apparaît une propriété liée aux jonctions, qui pourront être considérées comme intrinsèques ou extrinsèques en fonction d'une vue globale de la scène. Les jonctions définies comme extrinsèques n'ont pas de réalité physique et doivent être ignorées. En revanche, les jonctions définies comme intrinsèques sont interprétées comme des jonctions à prendre en compte dans le percept. Dans le cas de la séquence de la figure 4.19 (a), les fins de lignes indiquent les bons mouvements et sont intégrées. La jonction correspondant au croisement (ici extrinsèque), qui a localement un mouvement vertical est ignorée. Inversement, pour la séquence de la figure 4.19 (b), représentant le même stimulus mais en ajoutant des zones noires en haut et en bas, le percept change : on perçoit désormais un mouvement rigide vers le haut du croisement. Les fins de lignes (ici extrinsèques) ont été ignorées et la jonction centrale (devenue intrinsèque) est à présent prise en compte. La forme, ici les jonctions, est donc un élément déterminant du calcul.

D'autres études sur la forme [42] et le contraste [62] révèlent l'importance de ces facteurs dans l'intégration.

En physiologie et en neurophysiologie, une masse de connaissances existe sur les propriétés des aires V1 et MT. Je souhaite ici insister sur deux éléments :

- Il y a bien sûr les connaissances générales sur V1 et MT : la rétinotopie des aires corticales, la taille des champs récepteurs qui augmente en fonction de l'excentricité et en suivant la hiérarchie des aires corticales (figure 4.20 (a)), la sélectivité à la direction des cellules de V1 et MT (plus marquée dans MT), la forte inhibition dans MT vers les cellules préférant la direction opposée, la sélectivité à la vitesse des cellules de MT, les différentes structures centre-périphérie des cellules de MT (géométrie<sup>[79]</sup>), et sélectivité des vitesses de la périphérie par rapport au centre : identique ou opposée<sup>[9]</sup>, l'existence possible de sous-unités dans MT, etc.
- L'existence de rétroactions (*feedbacks*) importantes de MT vers V1. Les rétroactions représentent une part très importante des échanges dans le cortex et la plupart des aires visuelles possèdent des rétroactions sur V1 qui se comporte alors comme un "black-board", intégrant les informations. La figure 4.20 (b) indique l'extension des rétroactions vers V1. Une conséquence directe est que la diffusion longue distance des informations devient possible et plus efficace grâce aux rétroactions. En effet, les connexions horizontales dans l'aire V1 sont lentes (0,1 à 0,2 m/s), par opposition aux connexions entre V1 et MT myélinisées donc beaucoup plus rapides (2 à 6 m/s) [38]. C'est donc un enjeu important que de modéliser ces rétroactions, ce qui va correspondre aux comportements de type centre-périphérie.

Suite à cette étude, nous avons cherché à intégrer ces différents éléments dans un modèle en le comparant aux modèles existants, et en l'évaluant sur des exemples complexes proposés en psychophysique et en vision par ordinateur. Nous présentons dans [R 11] un rappel de quelques approches relatives à notre approche, et notamment l'approche de Bayerl et al. [4] de laquelle nous nous sommes inspirés.

Dans leur approche, Bayerl et al. ont cherché à établir les "probabilités" de présence d'une vitesse  $v$  à une position  $x \in \Omega \subset \mathbb{R}^2$  et un temps  $t$  donnés, à différents niveaux du traitement de l'information, simulant ainsi les traitements successifs dans V1 et MT. Ces probabilités peuvent également se voir comme des activités moyennes des neurones sélectifs à une direction  $v$ . Nous sommes partis de la même formulation, en cherchant à étendre le modèle original.

---

[42] X. Huang, T. D. Albright, and G. R. Stoner. Adaptive surround modulation in cortical area mt. *Neuron*, 53:761–770, March 2007.

[62] C. C. Pack, J. N. Hunter, and R. T. Born. Contrast dependence of suppressive influences in cortical area mt of alert macaque. *Journal of Neurophysiology*, 93(3):1809–1815, Mar 2005.

[79] D. K. Xiao, S. Raviguel, V. Marcar, and G. A. Orban. The spatial distribution of the antagonistic surround of mt/v5 neurons. *Cereb Cortex*, 7(7):662–77, 1997.

[9] R. T. Born. Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*, 84:2658–2669, 2000.

[38] A. Grinvald, E. E. Lieke, R. D. Frostig, and R. Hildesheim. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14(5):2545–2568, 1994.

[4] P. Bayerl and H. Neumann. Disambiguating visual motion by form-motion interaction—a computational model. *International Journal of Computer Vision*, 72(1):27–45, 2007.

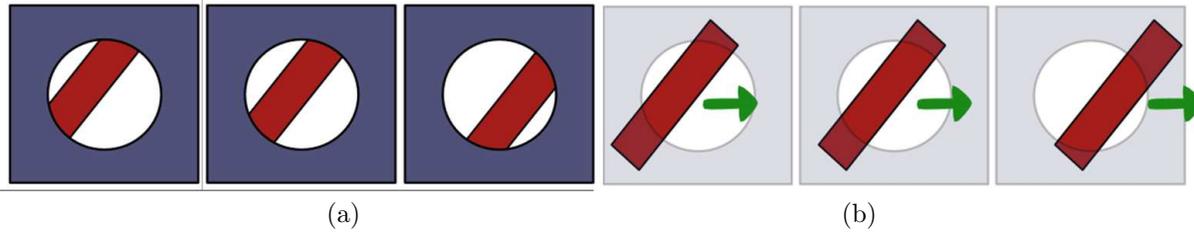


Fig. 4.17: Le problème d'ouverture : en l'absence d'indices 2D non ambigus comme dans (a), le mouvement réel (b) ne peut pas être connu. On perçoit alors le mouvement normal à la structure, correspondant au mouvement le plus faible.

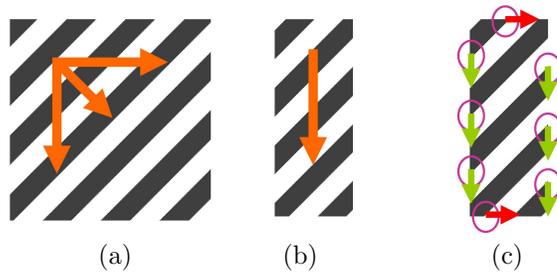


Fig. 4.18: Le *barber-pole* : (a) soient des barres se déplaçant dans la direction normale, dont le mouvement peut également s'interpréter de plusieurs façons. (b) Le même stimulus vu au travers d'une ouverture rectangulaire donne un percept différent. (c) Une explication vient du nombre d'indices 2D dans la scène.

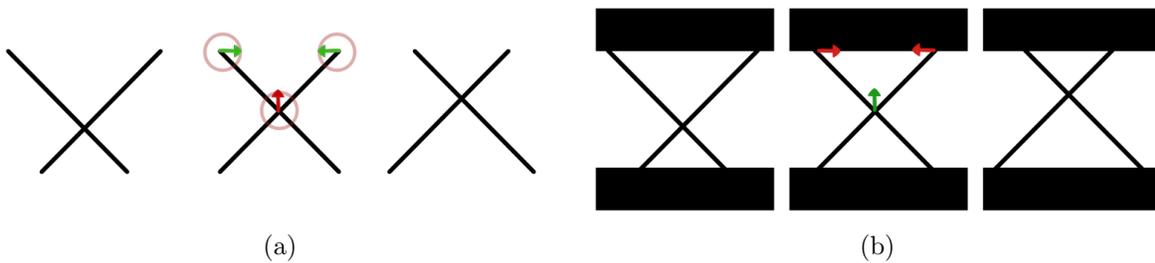


Fig. 4.19: L'illusion des *chopsticks* et la nature des jonctions (rouge code pour extrinsèque, vert code pour intrinsèque) : (a) deux barres se croisent et cela correspond au percept. (b) Le même stimulus mais en ajoutant des zones noires en haut et en bas, et le percept change : on perçoit désormais un mouvement rigide d'un croisement vers le haut.

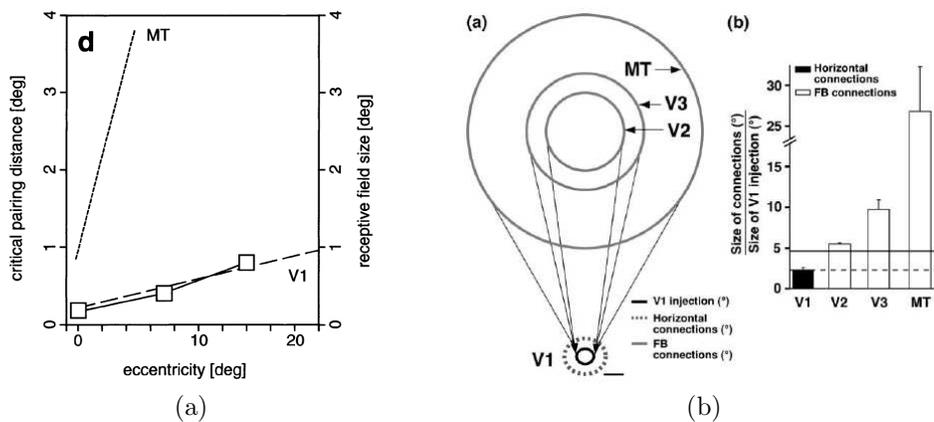


Fig. 4.20: (a) [schéma de Mestre, Masson et al., 2001] En pointillés : champs récepteurs de V1 et MT en fonction de l'excentricité; en trait plein : distance critique entre des points en déplacement contraire donnant une suppression du percept de mouvement. Ce schéma indique donc que les phénomènes de suppression semble arriver au niveau de V1. (b) [schéma d'Angelucci, Bullier, 2003] Représentation de la large étendue des cellules en rétroaction sur V1, à comparer avec l'étendue faible des connexions horizontales de V1.

Le modèle proposé repose sur une suite de traitements que l'on attribue à différentes couches. Soit une séquence d'images  $I : (t, x) \in \mathbb{R}^+ \times \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ . Soit une couche  $i$  définie sur un domaine spatial  $\Omega$ . On définit la fonction

$$p_i : (t, x, v) \in (\mathbb{R}^+, \Omega, V) \rightarrow p_i(t, x, v) \in \mathbb{R}^+, \quad (4.41)$$

comme la "probabilité"<sup>9</sup> au temps  $t$  et à la position  $x$  d'avoir un déplacement dans la direction  $v$ , où  $V = \{v^n\}$  est un ensemble de vitesses.

Le modèle que nous avons proposé comporte les sept couches suivantes

$$\begin{aligned} p_0(t, x, v) &= \text{détecteurs de Reichardt}, \\ p_1(t, x, v) &= G_\sigma(|v|) p_0(t, x, v), \\ p_2(t, x, v) &= \frac{p_1^2(t, x, v)}{\sum_u p_1^2(t, x, u) * G_\sigma} \\ p_3(t, x, v) &= p_2(t, x, v)(1 + c_f p_6(t - 1, x, v)), \\ p_4(t, x, v) &= \frac{p_3^2(t, x, v)}{\sum_u p_3^2(t, x, u) * G_\sigma} \\ p_5(t, x, v) &= \int_X G_\sigma(x - \xi) \pi(t, \xi, x\xi) p_4(t, \xi, v) d\xi, \\ p_6(t, x, v) &= \frac{p_5^2(t, x, v)}{\sum_u p_5^2(t, x, u) * G_\sigma}, \end{aligned}$$

où

- $p_0$  est calculé à partir d'un détecteur de mouvement classique, par exemple le détecteur de Reichardt comme utilisé dans Bayerl et Neumann (2007).
- $p_1$  représente l'activité des cellules V1 "components", qui ne réagissent qu'au mouvement local, avec le problème d'ouverture. Un *a priori* a été ajouté sur les faibles vitesses, ce qui permet de choisir une vitesse dans les cas ambigus mais aussi d'expliquer certains percepts (cas du rhombus par exemple<sup>[77]</sup>).
- $p_2, p_4$  et  $p_6$  correspondent à une normalisation, aussi appelée inhibition divisive. Elle est classiquement utilisée dans de nombreux modèles.
- $p_3$  représente l'activité des cellules V1 "pattern" pour lesquelles le problème d'ouverture est résolu. L'existence de telles cellules dans V1 a été révélée dans des études récentes<sup>[41]</sup>. L'activité de cette couche reçoit la sortie de  $p_2$ , mais aussi une rétroaction de la dernière couche  $p_6$ . Cette rétroaction dont nous avons parlé précédemment (voir aussi figure 4.20 (b)) va jouer un rôle très important dans la diffusion et l'intégration de l'information.
- $p_5$  représente l'activité des cellules de MT, qui ont intégré l'information venant de V1 en prenant en compte la forme décrite par la fonction  $\pi$ . Chaque activité correspond à des champs récepteurs plus larges que ceux de V1 (puisque'il y a une intégration), et elle est retournée dans  $p_3$ .

Cette modulation de l'intégration du mouvement par la forme est une contribution originale qui permet d'expliquer un certain nombre de phénomènes observés en psychophysique. La fonction  $\pi(t, x, d)$  va définir dans quelles directions les cellules de MT vont intégrer l'information, en fonction des structures présentes dans l'image. Pour un temps  $t$  et une position  $x$  donnés,  $\pi(t, x, d)$  indique si la direction  $d$  est une bonne direction d'intégration :

$$\pi(t, x, d) = \int_\Omega G_{\sigma_s}(\xi - x) G_{\sigma_\theta}(\widehat{d, x\xi}) G_{\sigma_i}(S(\xi, x)) d\xi,$$

<sup>9</sup>Ce ne sera pas une probabilité dans le sens où la somme des  $p_i$  ne sera pas égale à 1. On peut voir cette fonction comme une activité moyenne d'une unité neuronale au temps  $t$  et à la position  $x$ , sensible à la direction  $v$ .

[77] Y. Weiss and E. H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. *Center for Biological and Computational Learning Paper*, 1998.

[41] K. Guo, P. J. Benson, and C. Blakemore. Pattern motion is present in V1 of awake but not anaesthetized monkeys. *European Journal of Neuroscience*, 19(4):1055–1066, 2004.

où  $G_\sigma$  sont des gaussiennes,  $\widehat{d, \vec{x}\xi}$  est l'angle entre les vecteurs  $d$  et  $\vec{x}\xi$ , et  $S(\xi, x)$  est une mesure de similarité entre les intensités  $I(t, x)$  et  $I(t, \xi)$ . Cette mesure est donc une mesure des points  $\xi$  dans  $\Omega$  tels que  $\xi$  soit proche de  $x$ , l'angle formé par  $d$  et  $\vec{x}\xi$  est faible, et les intensités de l'image en  $x$  et  $\xi$  sont proches. Une illustration de cette fonction est donnée dans la figure 4.21.

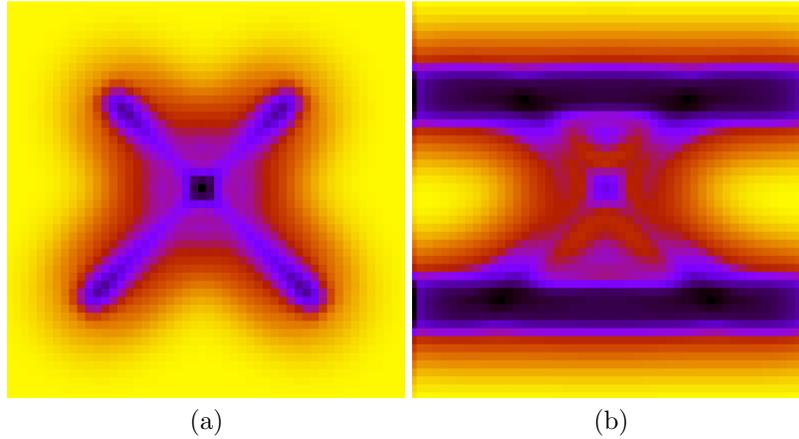


Fig. 4.21: Représentation du coefficient de diffusion défini par la somme des  $\pi(t, x, d)$  pour tout  $d$  : (a) dans le cas des deux barres qui se croisent, et (b) avec ces mêmes barres et les zones noires (voir la figure 4.19). On observe que le coefficient de diffusion est le plus faible pour les zones qui seront considérées comme extrinsèques : la vitesse portée par ces zones va donc moins se diffuser dans l'image.

Ce modèle a été testé sur de nombreuses séquences, et il permet de reproduire des phénomènes observés en psychophysique, comme ceux décrits dans la figure 4.19 : voir le résultat 4.22 et [R 11] pour plus de détails.

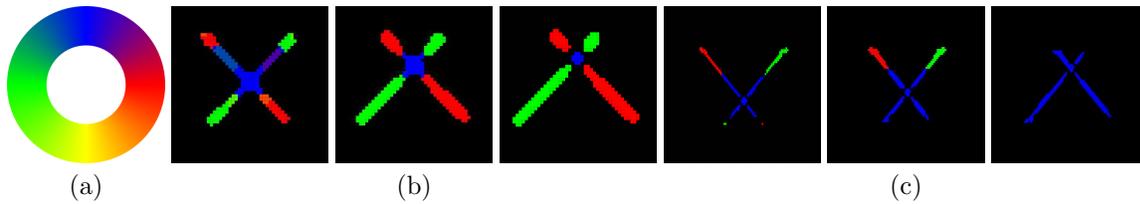


Fig. 4.22: Exemple de résultat sur les séquences décrites dans la figure 4.19. Seule la direction des cellules les plus activées de V1 (fonction  $p_4(t, x, v)$ ) est affichée suivant le code couleur indiqué en (a). Ces résultats mettent en évidence la différence de comportement de l'approche en fonction des occultations : (b) sans occultation, on retrouve le mouvement des deux barres, quand le temps évolue, (c) avec occultation, le mouvement du croisement l'emporte.

*Nous avons plusieurs pistes pour améliorer ce modèle. D'abord, l'améliorer pour prendre davantage en compte les caractéristiques du système visuel. Ensuite, considérer une mise à jour de l'activité ayant sa propre dynamique, qui ne soit pas calée avec les images. Enfin, l'un des défauts de l'approche actuelle est son coût, venant notamment des intégrales pour  $p_5$  en  $\mathcal{O}(N^2)$ . Une possibilité à envisager, vu que c'est parallélisable et compatible avec notre modèle, est l'utilisation du GPU.*

*A l'avenir, ce travail s'intégrera dans le cadre d'un nouveau projet Européen SEARISE venant d'être accepté ("Smart Eyes : Attending and Recognizing Instances of Salient Events", 2008–2011), que je coordonne au niveau de l'INRIA. L'objectif de ce projet est de développer un système de vision active pour la détection, le suivi et la catégorisation d'évènements saillants. Contrairement aux systèmes de vidéo surveillance existants, l'objectif sera d'y intégrer des approches bio-inspirées reproduisant des comportements de type humain sur l'attention, l'adaptabilité, l'apprentissage, et les stratégies d'observation d'une scène.*

### 4.3 Les méthodes variationnelles en neurosciences ?

*Cadre contractuel : projet Européen FACETS ("Fast Analog Computing with Emergent Transient States", EC IP project FP6-015879, 2005–2009). FACETS est un grand consortium entre plusieurs communautés scientifiques visant à développer les connaissances sur le système nerveux. Ce travail est mené en collaboration avec Thierry Viéville.*

#### Processus continus et comportements collectifs discrets

L'introduction des approches variationnelles dans le contexte des réseaux de neurones n'est pas nouvelle. Elle a notamment été introduite par Cottet [23]. Dans ses travaux, il s'agissait essentiellement de montrer comment se traduisent les interactions entre voisins, en continu et en discret.

- En continu, dans le cadre des équations aux dérivées partielles (EDP), les interactions entre voisins se font au travers d'opérateurs de diffusion, qui sont des opérateurs différentiels elliptiques d'ordre 2. Il s'agit d'une opération très locale.
- En discret, dans le cadre des équations différentielles ordinaires (EDO), les interactions entre voisins se font par une somme pondérée des valeurs des voisins (échantillons), les poids correspondant aux poids synaptiques. L'opération peut se faire sur un voisinage arbitrairement grand.

Le problème est donc de démontrer la correspondance entre un opérateur différentiel de diffusion et une somme pondérée de valeurs.

Par exemple, considérons un réseau de Hopfield

$$\frac{du_i}{dt} = \frac{1}{n_i} \sum_j J_{ij} v_j - \alpha_i u_i, \quad (4.42)$$

$$v_j = g(\lambda u_j), \quad (4.43)$$

où  $u_i$  s'interprète comme le montant duquel le potentiel de membrane  $v_j$  dépasse son seuil, i.e., le taux de décharge, et les coefficients  $J_{ij}$  sont les poids synaptiques. Dans (4.42), le premier terme correspond à la diffusion, et le second à la fuite. Dans (4.43),  $g$  est une fonction strictement croissante, de type sigmoïde, permettant le passage du potentiel au taux de décharge.  $n_i$  est le nombre de neurones du système connectés au neurone  $i$ . Le résultat classique est que si les  $J_{ij}$  sont constants et symétriques, alors on peut montrer la convergence du système en montrant l'existence d'une fonction de Lyapunov.

On peut également ré-écrire (4.42)–(4.43) comme un système d'équations différentielles ordinaires, avec uniquement la variable  $v_i$  :

$$\frac{dv_i}{dt} = \frac{\lambda}{G'(v_i)} \left( \frac{1}{n_i} \sum_j J_{ij} v_j - \frac{\alpha_i}{\lambda} G(v_i) \right), \quad (4.44)$$

où  $G$  est la fonction réciproque de  $g$ . Avec certains choix de  $J_{ij}$  dans (4.44), Cottet et Ayyadi<sup>[24]</sup> ont montré que (4.44) devient équivalent à une EDP. L'équivalence signifie que les solutions discrètes en espace de (4.44) vont être arbitrairement proches de la solution de l'EDP correspondante, quand la distance entre les neurones tend vers zéro, et que le nombre de neurones tend vers l'infini (voir l'illustration de la figure 4.23). C'est un résultat limite.

Regardons un exemple pour montrer le type de calculs à mener pour faire ce passage discret-continu. Dans cet exemple, on va considérer des poids synaptiques ne dépendant que de la distance entre neurones (pas de leur état), ce qui va correspondre, en continu, à un opérateur de diffusion linéaire isotrope.

**Exemple** Soient  $N$  neurones à la position  $x_i \in \mathfrak{R}$  suivant la dynamique (4.44) avec les poids synaptiques constants et symétriques

$$J_{ij} = \eta \left( \frac{x_i - x_j}{\varepsilon} \right), \quad (4.45)$$

[23] G.-H. Cottet. Neural networks: continuous approach and applications to image processing. *J. Biological Systems*, 3, 1995.

[24] G.-H. Cottet and M. El Ayyadi. A Volterra type model for image processing. *IEEE Transactions on Image Processing*, 7(3), March 1998.

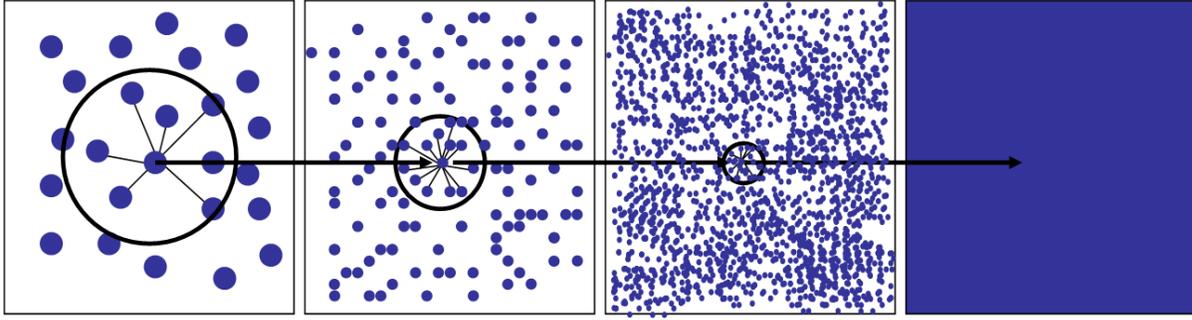


Fig. 4.23: Du discret au continu, des résultats limites : des voisinages de plus en plus petits, contenant de plus en plus d'échantillons.

où  $\eta$  est une fonction paire et  $\varepsilon$  indique l'étendue de connectivité du voisinage. On définit les moments

$$\tau_0 = \int \eta(y)dy, \quad \tau_2 = \int y^2 \eta(y)dy, \quad (4.46)$$

le moment  $\tau_1$  étant nul. Alors, les solutions de l'équation de réaction diffusion définie par

$$\frac{\partial v}{\partial t} = \left( \varepsilon \frac{\lambda \tau_2}{G'(v)} \right) \frac{\partial^2 v}{\partial x^2}(t, x) + \frac{1}{G'(v)} (\lambda \tau_0 v - \alpha G(v)), \quad (4.47)$$

est une approximation de (4.44). ■

**Démonstration** En effet, avec un développement de Taylor de  $v(y)$  autour de  $x$ , on a

$$v(t, y) = v(t, x) + (y - x) \frac{\partial v}{\partial x}(t, x) + \frac{(y - x)^2}{2} \frac{\partial^2 v}{\partial x^2}(t, x) + \frac{(y - x)^3}{3!} \frac{\partial^3 v}{\partial x^3}(t, x) + o((y - x)^3),$$

soit

$$\frac{(y - x)^2}{2} \frac{\partial^2 v}{\partial x^2}(t, x) = v(y) - v(x) - (y - x) \frac{\partial v}{\partial x}(t, x) - \frac{(y - x)^3}{3!} \frac{\partial^3 v}{\partial x^3}(t, x) + o((y - x)^3). \quad (4.48)$$

En multipliant (4.48) par  $\eta\left(\frac{y-x}{\varepsilon}\right)$  et en intégrant, on a

$$\varepsilon^3 \frac{\tau_2}{2} \frac{\partial^2 v}{\partial x^2}(t, x) = \int \eta\left(\frac{y-x}{\varepsilon}\right) v(y) dy - \varepsilon^2 \tau_0 v(x) + o((y-x)^3). \quad (4.49)$$

A présent, si on discrétise (4.49) avec une méthode de quadrature (voir aussi les méthodes particulières) avec des échantillons réguliers  $x_i = ih$  ( $h \in Z$ ), où  $h$  désigne la distance entre les échantillons ( $\varepsilon$  désignait la taille de la fenêtre d'intégration, au travers de la fonction  $\eta$ ), on a

$$\begin{aligned} \frac{\tau_2}{2} \frac{\partial^2 v}{\partial x^2}(t, x_i) &= \frac{\varepsilon^{-1}}{n} \sum_j \eta\left(\frac{x_j - x_i}{\varepsilon}\right) v(x_j) - \varepsilon^{-1} \tau_0 v(x_i) + o\left(\frac{h}{\varepsilon^3}\right) + o\left(\frac{h^3}{\varepsilon^3}\right), \\ &\approx \frac{\varepsilon^{-1}}{n} \sum_j J_{ij} v_j - \varepsilon^{-1} \tau_0 v_i, \end{aligned} \quad (4.50)$$

où  $n = 1/h$ . L'approximation (4.50) met en relation un opérateur différentiel (ici une dérivée seconde) avec une somme pondérée de valeurs de  $v$  aux échantillons  $x_i$ . Comme nous l'avons annoncé, c'est un résultat valable à convergence, quand le pas de discrétisation  $h$  tend vers zéro, et quand le voisinage  $\varepsilon$  tend aussi vers zéro, mais moins vite que  $h$ , i.e., un nombre d'échantillons tendant vers l'infini dans un voisinage d'intégration tendant vers zéro (figure 4.23).

Mais cela ne suffit pas : nous souhaitons savoir dans quelle mesure la solution discrète du réseau de Hopfield est une bonne approximation de la solution continue de l'EDP (4.47). Formellement, si on reprend (4.47) en prenant en compte (4.50), on obtient

$$\frac{\partial v}{\partial t}(t, x_i) = \frac{\lambda}{G'(v_i)} \left( \frac{1}{n} \sum_j J_{ij} v_j - \frac{\alpha}{\lambda} G(v_i) \right),$$

à savoir l'équation de Hopfield (4.44), laissant donc deviner que les solutions vont rester proches (au moins en temps fini). De façon plus rigoureuse, montrer que les solutions de l'EDP et de l'EDO sont proches passe par le Théorème 2 proposé par Edwards<sup>[30]</sup>. Comme précédemment, il s'agit d'un résultat limite. ■

De façon plus générale, comme cela a été étudié par Degond et Mas Gallic<sup>[27]</sup> puis Edwards<sup>[30]</sup>, l'objectif est de trouver des correspondances entre les poids d'une formulation discrète additive et un opérateur de diffusion. Comme il est démontré dans Edwards, le point clé est de montrer comment un opérateur de diffusion

$$D_2 v(x) = \operatorname{div}(L(x)\nabla v), \quad (4.51)$$

où  $L$  est une matrice, peut être discrétisé par un opérateur intégral sous la forme

$$Qv(x) = \int \sigma(x, y)(v(y) - v(x))dy, \quad (4.52)$$

avec (en dimension 2)

$$\sigma(x, y) = \frac{1}{\varepsilon^2} \sum_{i,j} M_{ij}(x, y) \eta_{ij} \left( \frac{y-x}{\varepsilon} \right). \quad (4.53)$$

Une solution consiste à se donner la fonction  $\eta_{ij}$  qui va définir le voisinage d'interactions synaptiques et de trouver  $M_{ij}$  vérifiant des conditions de compatibilité avec  $L$  du type

$$\sum_{i,j=1..N} m_{ij}(x) \tau_{e_l+e_k}^{ij} = 2L_{kl} \quad \forall k, l = 1..N,$$

où  $m_{ij}(x) = M_{ij}(x, x)$ , et  $\tau_{\alpha}^{ij}(x) = \int (y-x)^{\alpha} \eta_{ij}(y) dy$  (la notation multi-indices standard est utilisée ici, i.e., pour un vecteur d'indices entiers  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathcal{N}^n$ , on définit  $|\alpha| = \alpha_1 + \dots + \alpha_n$  et  $x^{\alpha} = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ ). Cette solution n'est pas unique (voir Edwards (1996) pour plus de détails).

*Les preuves de ces résultats sont techniques et notre objectif a été de les revisiter en prenant en compte dès le début la structure d'une carte corticale, avec son échantillonnage naturellement défini par les colonnes corticales, leur champ récepteur et leur connectivité. Ce travail a été présenté dans [C 15], [J 3].* 

Comme point de départ, donnons nous une architecture neuronale simple décrivant des interactions entre colonnes corticales à l'intérieur d'une même carte ou entre deux cartes en connexion montante (*feed-forward*). Les cartes corticales que l'on considère sont définies de façon rétinotopique sur un domaine  $\Omega$ . Dans  $\Omega$ , on considère un ensemble de colonnes corticales  $j$  à la position  $y_j$  (les échantillons). Chaque colonne possède un champ récepteur  $\mathcal{S}_j$  et elle est caractérisée par une densité  $\mu_j$  (par exemple,  $\mu_j(y) = \delta(y - y_j)$  ou  $\mu_j(y) = \text{cste}$ ). Soit  $v$  une fonction définie sur  $\Omega$ . On définit alors l'activité moyenne de la colonne  $j$  par

$$A[v(y)]_{\mu_j}^y = \int_{\mathcal{S}_j} v(y) \mu_j(y) dy, \quad (4.54)$$

où la variable  $y$  en exposant indique la variable d'intégration pour la fonction indiquée entre crochets.

Soit maintenant une colonne corticale à la position  $x$ , ayant un champ récepteur contenant un ensemble de champs récepteurs  $\mathcal{S}_j$ , dans la même carte ou en leur équivalent dans la carte suivante. (voir par exemple la figure 4.24).

Nous avons alors montré le résultat suivant.

**Proposition 4.3.1** *Etant donné un opérateur différentiel (4.51), on cherche un approximation intégrale du type (4.52) avec*

$$\sigma(x, y) = \sum_j \chi_{\mathcal{S}_j}(y) \sigma(x, y_j) \mu_j(y), \quad (4.55)$$

où  $\chi_{\mathcal{S}_j}$  est la fonction indicatrice de l'ensemble  $\mathcal{S}_j$ . Alors, les coefficients  $\sigma(x, y_j)$  doivent vérifier les conditions de compatibilités suivantes

$$\begin{aligned} \sum_j \sigma(x, y_j) \tau_{e_k+e_l}^j(x) &= 2L_{kl}(x), \\ \sum_j \sigma(x, y_j) \tau_{e_k}^j(x) &= \mathbf{div}_k(L(x)), \end{aligned} \quad (4.56)$$

[30] R. Edwards. Approximation of neural network dynamics by reaction-difusion equations. *Mathematical Methods in the Applied Sciences*, 19:651–677, 1996.

[27] P. Degond and S. Mas-Gallic. The weighted particle method for convection-diffusion equations. *Mathematics of Computation*, 53(188):485–525, 1989.

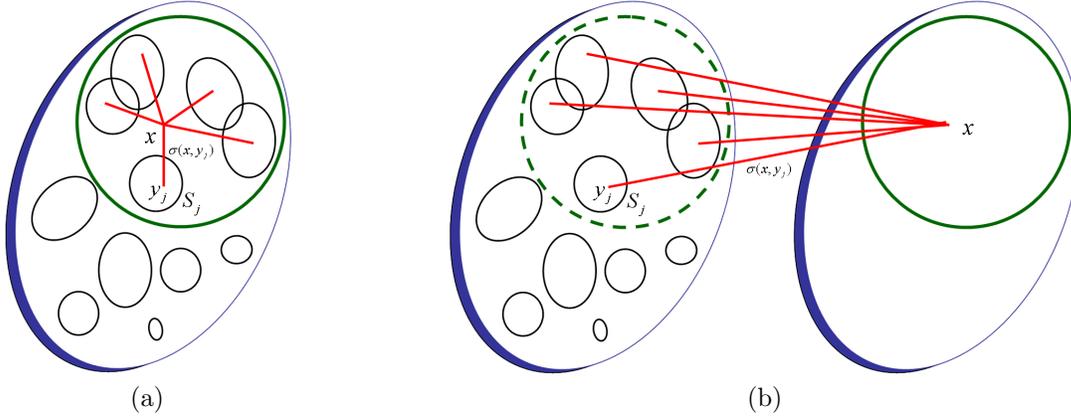


Fig. 4.24: Structure des cartes corticales considérées. Les calculs présentés pourront correspondre aux deux situations suivantes : (a) une seule carte corticale composée de colonnes corticales et d'unités de calculs dans la même carte, prenant en compte les colonnes corticales dans leur champ récepteur, (b) deux cartes corticales en correspondance par rétinotopie, la seconde carte intégrant de l'information de la première.

où  $\tau_\alpha^j(x) = \int_{S_j} (y-x)^\alpha \mu_j(y) dy$ .

**Démonstration** Partant de l'approximation (4.52), effectuons un développement de Taylor de  $v$  au voisinage de  $x$

$$v(y) - v(x) = \sum_{|\alpha|=1}^r \frac{\partial^\alpha v}{\alpha!}(x) \Big|_{y=x} (y-x)^\alpha + o(|y-x|^r) = \left[ \sum_{|\alpha|=1}^r (y-x)^\alpha \frac{\partial^\alpha}{\alpha!} \right] v(x) + o(|y-x|^r). \quad (4.57)$$

On a alors

$$\begin{aligned} Qv(x) &= \int_S \sigma(x,y) (v(y) - v(x)) dy \\ &= \sum_j \int_S \chi_{S_j} \sigma(x,y_j) \mu_j(y) (v(y) - v(x)) dy \\ &= \sum_j \sigma(x,y_j) \int_{S_j} \mu_j(y) (v(y) - v(x)) dy \\ &= \sum_j \sigma(x,y_j) \int_{S_j} \mu_j(y) \left[ \sum_{|\alpha|=1}^r (y-x)^\alpha \frac{\partial^\alpha}{\alpha!} \right] v(x) dy + \mathbf{R}^r v \\ &= \left[ \sum_{|\alpha|=1}^r \sum_j \sigma(x,y_j) \int_{S_j} (y-x)^\alpha \mu_j(y) dy \frac{\partial^\alpha}{\alpha!} \right] v(x) + \mathbf{R}^r v \\ &= \left[ \sum_{|\alpha|=1}^r \sum_j \sigma(x,y_j) \tau_\alpha^j(x) \frac{\partial^\alpha}{\alpha!} \right] v(x) + \mathbf{R}^r v, \end{aligned} \quad (4.58)$$

où le reste  $\mathbf{R}^r v$  s'écrit sous forme intégrale

$$\mathbf{R}^r v = \sum_{|\alpha|=r+1} \frac{r+1}{\alpha!} \int_{S \times [0,1]} \sigma(x,y) (y-x)^\alpha (1-u)^r \partial^\alpha v^j(x + u(y-x)) dy du,$$

et comme le domaine est borné (disons dans une boule de rayon  $\varepsilon$ ), alors

$$|\mathbf{R}^r v|_{0,\infty} < C \varepsilon^{r-1} |v|_{r+1,\infty},$$

où  $C$  est une constante.

D'autre part, on peut montrer que l'opérateur de diffusion (4.51) se décompose sous la forme suivante

$$D_2v(x) = \operatorname{div}(L(x)\nabla v) = \left[ \sum_k \operatorname{div}_k(L(x))\partial^{e_k} + \sum_{kl} L^{kl}(x)\partial^{e_k+e_l} \right] v(x). \quad (4.59)$$

En comparant (4.58) et (4.59) jusqu'à l'ordre  $r$ , on obtient les relations (4.56). ■

**Remarque** L'approximation intégrale (4.52) avec (4.55) donne donc l'approximation suivante

$$\begin{aligned} Qv(x) &= \int_{\mathcal{S}} \sigma(x, y) (v(y) - v(x)) dy \\ &= \int_{\mathcal{S}} \sum_j \chi_{\mathcal{S}_j} \sigma(x, y_j) \mu_j(y) (v(y) - v(x)) dy \\ &= \sum_j \sigma(x, y_j) \int_{\mathcal{S}_j} \mu_j(y) (v(y) - v(x)) dy \\ &= \sum_j \sigma(x, y_j) \left( \int_{\mathcal{S}_j} \mu_j(y) v(y) dy - v(x) \int_{\mathcal{S}_j} \mu_j(y) dy \right) \\ &= \sum_j \sigma(x, y_j) \left( A[v(y)]_{\mu_j}^y - A[v(x)]_{\mu_j}^x \right). \end{aligned} \quad (4.60)$$

où nous avons utilisé la définition (4.54) dans la dernière égalité. ■

*Dans cette contribution, nous nous sommes essentiellement intéressés à établir la correspondance entre un opérateur de diffusion et son approximation discrète, étant donnée une architecture neuronale discrète. Dans la partie suivante, nous considérons comment un calcul sur une carte peut être spécifié au moyen d'approches variationnelles, dont la minimisation va mener à des termes de diffusion, et donc des échanges dans le réseau.* 👁️

### Spécification de processus discrets par une approche variationnelle générale

*Dans la partie précédente, nous avons montré comment passer d'un opérateur différentiel linéaire, anisotrope du type  $D_2v(x) = \operatorname{div}(L(x)\nabla v)$ , à une approximation discrète étant donné une architecture neuronale (colonnes corticales, champs récepteurs et connectivités). Dans cette contribution nous proposons une formulation variationnelle générale comme un moyen de spécifier le calcul effectué dans une carte corticale, et étudions comment l'interaction entre cartes et en particulier les récurrences peuvent être modélisées de manière stable par des ensembles de formulations variationnelles couplées. C'est un travail en cours.*

La surface du cortex est découpée en aires corticales, et il semble apparaître des séparations physiologiques qui permettent de définir une carte corticale, comme par exemple la classification en aires de Brodmann<sup>10</sup>. Les chercheurs en neurosciences computationnelles ont une vue précise de l'architecture des cartes corticales et de la façon dont l'architecture neuronale (microscopique et macroscopique) permet de faire émerger des fonctions visuelles. On peut citer notamment les travaux de Grossberg<sup>[39]</sup>, Friston<sup>[34]</sup>, Dayan et Abbott<sup>[26]</sup> ou Burnod<sup>[14]</sup>.

Que ce soit pour la vision algorithmique ou biologique, la perception repose sur des calculs effectués et représentés par des "cartes" (corticales dans le cas du cortex). La notion de carte dont nous parlons ici comprend donc la représentation de l'information que l'on peut voir comme fonction définie sur une surface ou un domaine bidimensionnel, mais aussi les processus de calculs qui permettent d'évaluer cette

<sup>10</sup>Toutefois, les limites de ces aires ne sont pas toujours précises, dépendent de leur définition (par le type de cellule ou par la morphologie corticale par exemple) et ne correspondent pas forcément à des séparations fonctionnelles.

[39] Stephen Grossberg. Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1(1):1–97, 1988.

[34] Karl Friston. Functional integration and inference in the brain. *Prog Neurobiol*, 68:113–143, 2002.

[26] P. Dayan and L. F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[14] Y. Burnod. *An adaptive neural network: the cerebral cortex*. Masson, Paris, 1993. 2nd edition.

fonction. Par exemple, on trouve des cartes qui codent de façon rétinotopique (c'est-à-dire topologique, en correspondance avec un repère géométrique dans la rétine) les contrastes d'intensité ou de couleurs, les contrastes d'orientation, la courbure, les informations liées à la disparité, la sélectivité à une certaine réponse, le mouvement, etc.

La spécification d'un processus, i.e., du calcul d'une carte, est souvent basée sur un mécanisme qui corrige de manière itérative la sortie de manière à prédire l'entrée. On peut penser aux méthodes classiques d'*expectation-minimization* (EM), de telle sorte que l'*expectation* produit une sortie à partir de l'entrée, et la *minimization* va prédire l'entrée à partir d'estimés *a priori* de la sortie. Cela peut s'écrire sous la forme d'un problème d'optimisation, ce qui nous a amenés à considérer de manière plus large les formulations variationnelles dans ce contexte.

Soit  $w : \mathbb{R}^n \rightarrow \mathbb{R}^q$  donné, représentant l'entrée d'une carte corticale, définie sur un ensemble  $\Omega$  et caractérisée par une valeur  $v : \mathbb{R}^n \rightarrow \mathbb{R}^p$ . Nous nous intéressons dans cette partie aux processus pouvant être définis par une formulation variationnelle du type

$$\bar{v} = \underset{v \in H/c(v)=0}{\operatorname{argmin}} \mathcal{L}(v), \text{ avec } \mathcal{L}(v) = \int_{\Omega} |Pv - w|_{\Lambda}^2 + \int_{\Omega} \phi(|\nabla v|_L) + \int_{\Omega} \psi(v), \quad (4.61)$$

où  $\nabla$  est l'opérateur gradient,  $P$ ,  $\phi(\cdot)$ ,  $\psi(\cdot)$ ,  $c(\cdot)$  déterminent le calcul à effectuer, et  $\Lambda$  et  $L$  vont correspondre à des métriques qui joueront un rôle particulier par la suite. Plus précisément,  $|u|_M = u^T M u$  où  $M$  est une matrice symétrique définie positive.

Nous avons choisi cette formulation la plus générale possible où l'on retrouve les mêmes ingrédients que dans les approches classiques en vision algorithmique. Le premier terme est un terme d'attache aux données, le second est un terme de lissage qui définira la régularité de la solution, enfin le troisième permet de contraindre la forme de la solution. Noter aussi le rôle de l'espace fonctionnel et la possibilité de rajouter des contraintes. La formulation (4.61) est donc une manière de spécifier un calcul, elle en explique l'objectif, i.e., ce qui doit être fait, mais sans préciser comment cela doit être fait. Cette formulation générale comprend bien sûr un grand nombre d'applications de vision algorithmique (voir [LI 2]), mais aussi des calculs issus du monde des neurosciences comme le *winner-take-all* (voir [J 3]).

Etant donnée cette formulation, il devient possible de définir les poids synaptiques d'un réseau de neurones (modélisant l'activité d'une carte corticale), en dérivant le critère (4.61) et en utilisant l'approche exposée précédemment pour les passages continus-discrets. Dans ce cas, on modélise des phénomènes de plasticité puisque l'opérateur différentiel est à présent non-linéaire.

Sur cette base, nous avons souhaité établir des analogies entre la formulation variationnelle (4.61) et le cortex visuel, quand on considère plusieurs cartes corticales en interaction. La communication cortico-corticale est faite au travers de connexions montantes (*forward*) et descendantes (*backward*) qui ont des propriétés anatomiques différentes<sup>[13,34]</sup>. Mais cette distinction entre connexions montantes et descendantes n'est pas seulement anatomique, elle est aussi fonctionnelle : les connexions montantes agissent pour la promulgation et la ségrégation de l'information sensorielle, les connexions descendantes ont une action modulatrice pour la médiation des effets contextuels ou la coordination des calculs. Remarquons que cet effet de modulation n'intervient pas nécessairement après les connexions montantes, mais dès qu'une information est transmise.

Nous nous sommes alors posé la question de la modélisation de tels mécanismes, et en particulier de l'interaction entre plusieurs cartes, *via* le formalisme variationnel. Considérons  $M$  cartes corticales définies par des critères du type (4.61).

$$\underset{v_m \in H_m/c_m(v_m)=0}{\operatorname{argmin}} \mathcal{L}_m(v_m) = \int_{\Omega_m} |P_m v_m - w_m|_{\Lambda_m}^2 + \int_{\Omega_m} \phi_m(|\nabla v_m|_{L_m}) + \int_{\Omega_m} \psi_m(v_m). \quad (4.62)$$

La question est de savoir comment concevoir des couplages entre ces cartes, et comment faire en sorte que ces couplages génèrent une dynamique stable. Pour cela, nous avons proposé les mécanismes de couplages suivants :

- Pour les connexions descendantes, nous avons proposé que ces couplages soient effectués au travers des métriques  $\Lambda_m$  et  $L_m$  qui définissent le comportement d'une carte. Pour les définir, nous avons pris en compte une propriété importante des connexions *backward*, à savoir leur grande divergence spatiale : avant d'être retournées, les valeurs sont d'abord moyennées spatialement (cette même

[13] J. Bullier. Integrated model of visual processing. *Brain Res. Reviews*, 36:96–107, 2001.

[34] Karl Friston. Functional integration and inference in the brain. *Prog Neurobiol*, 68:113–143, 2002.

idée se retrouve aussi dans le contrôle de gain dans la rétine, voir section 4.2). Ainsi, étant donné une carte  $m$  et  $i_1, i_2, \dots, i_m$  les indices des cartes connectées à  $m$  en rétroaction, alors les métriques  $\Lambda_m$  et  $L_m$  dépendent de  $v_{i_1}, v_{i_2}, \dots, v_{i_m}$  de la façon suivante

$$\Lambda_m = f(\rho(\mathcal{S} * (v_{i_1}, v_{i_2}, \dots, v_{i_m}))) \quad \text{et} \quad L_m = g(\rho(\mathcal{S} * (v_{i_1}, v_{i_2}, \dots, v_{i_m}))), \quad (4.63)$$

où  $\mathcal{S}$  est un opérateur de lissage spatial,  $\rho$  est une fonction de rectification (par exemple  $\rho(v) = \max(v, 0)$ ) pour prendre en compte le fait que seules des valeurs positives sont retournées par le cortex, et  $f, g$  sont des fonctions régulières.

- Pour les connexions montantes, elles définissent un treillis (sans bouclage, des connexions point à point). Nous avons donc proposé comme modèle de connexion montante entre la carte  $n$  et la carte  $m$

$$w_m = h(\rho(\mathcal{T} * v_n)), \quad (4.64)$$

où  $\mathcal{T}$  est un filtre temporel qui prend en compte les délais de propagation et  $h$  une fonction régulière.

Etant donnée cette architecture, nous avons effectué des simulations numériques, portant par exemple sur la restauration d'images contrôlée par des *feedbacks* de ce type : le lissage des valeurs de la carte définissant la rétroaction s'avère critique pour la stabilité. Nous avons donc avancé la conjecture suivante :

**Conjecture 4.3.1** *Soient  $M$  cartes corticales définies par des formulations variationnelles du type (4.62), telles que pour tout  $\Lambda_m$  et  $L_m$  constants il existe une unique solution au problème. Soient des connexions entre ces cartes, vérifiant les propriétés (4.63) et (4.64). Alors, il existe  $(\bar{v}_1, \dots, \bar{v}_M)$  tel que*

$$\bar{v}_i = \underset{v_i}{\operatorname{argmin}} \mathcal{L}_i(v_i) \quad \forall i = 1, \dots, M.$$

Cette conjecture n'est pas du tout triviale *a priori* car la mise en place de couplages peut entraîner des comportements complexes tels que oscillations ou chaos. Nous cherchons à la démontrer.

## Approches variationnelles et segmentation par synchronisation

*Ce travail correspond au travail de Master de Léonard Gérard (Université Paris 7), que j'ai co-encadré avec Thierry Viéville. C'est un travail en cours.*

Segmenter une image, c'est identifier ses différentes composantes. Une composante se définit par une région qui possède des caractéristiques homogènes par rapport à son environnement. La caractéristique la plus commune est l'intensité (le niveau de gris)<sup>11</sup> : le problème de segmentation revient donc à séparer l'image originale en différentes régions d'intensités similaires. Segmenter a un intérêt majeur, c'est simplifier. Cette simplification du contenu fournit une représentation abstraite plus compacte de l'information.

Identifier différentes composantes, c'est être capable de délimiter des régions, de trouver où sont les discontinuités à préserver, pour représenter au mieux l'image originale. Il apparaît donc cette dualité entre (i) trouver des régions homogènes et (ii) trouver les discontinuités pertinentes. Cela a été formalisé dans le domaine des approches variationnelles par Mumford et Shah<sup>[57]</sup> (voir [LI 2] pour une étude détaillée de ce modèle) : étant donnée une image  $u_0$  que l'on souhaite segmenter, on cherche une fonction  $u$  régulière par morceaux et  $K$  l'ensemble des discontinuités, comme solutions de la formulation variationnelle

$$\inf_{u, K} F(u, K) = \int_{\Omega} (u - u_0)^2 dx + \alpha \int_{\Omega - K} |\nabla u|^2 dx + \beta \int_K d\sigma, \quad (4.65)$$

où l'on trouve un terme d'attache aux données, un terme de régularisation excluant l'ensemble des sauts, et la mesure des sauts. Les paramètres  $\alpha, \beta$  vont permettre de définir la régularité de la solution obtenue.

<sup>11</sup>Il existe bien d'autres caractéristiques que l'on peut utiliser pour segmenter une image. Par exemple la texture a été largement utilisée ces dernières années. La même idée s'applique plus généralement à toute quantité : des vecteurs, des tenseurs, etc

[57] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.

Mais résoudre (4.65) d'un point de vue numérique est un problème délicat, notamment à cause de la difficulté de minimiser par rapport à une courbe en discret. Plusieurs solutions ont été proposées, parmi lesquelles l'approximation par une énergie discrète par Chambolle et Dal Maso<sup>[20]</sup> :

$$F_h(u) = 2(u(x) - u_0(x)) + h^N \sum_{\substack{x \in h\mathbb{Z}^N \\ x \in \Omega}} \sum_{\substack{\xi \in \mathbb{Z}^N \\ x + h\xi \in \Omega}} \frac{1}{h} f_\xi \left( \frac{(u(x) - u(x + h\xi))^2}{h} \right) \phi(\xi), \quad (4.66)$$

où  $\phi(\xi) > 0$  représente l'amplitude des connexions locales, et  $f_\xi$  donne la nature de ces connexions : excitatrices ( $f_\xi > 0$ ), symétriques, et dépendant de  $u(x) - u(x + h\xi)$ . Les caractéristiques de la fonction  $f_\xi$  sont indiquées dans la figure 4.25. Les auteurs ont montré que (4.66)  $\Gamma$ -converge vers une énergie qui se ramène à (4.65).

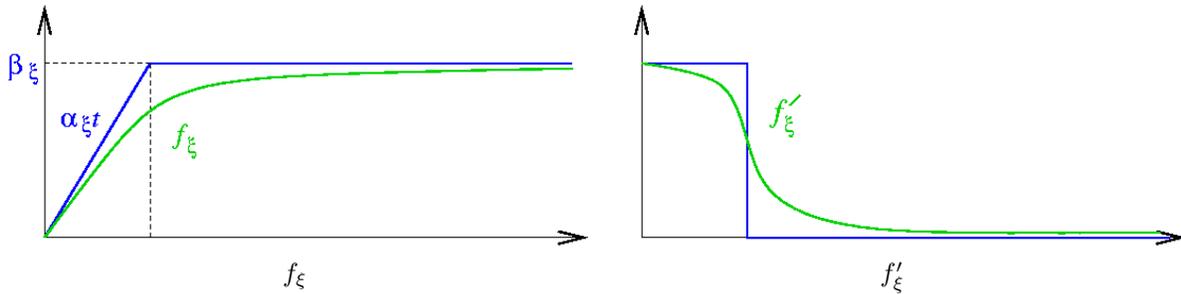


Fig. 4.25: Fonctions intervenant dans l'approximation de Mumford et Shah par Chambolle et Dal Maso (4.66). On a  $\alpha_\xi = f'_\xi(0)$ ,  $\beta_\xi = \lim_{t \rightarrow +\infty} f_\xi(t)$ ,  $f_\xi(0) = 0$ ,  $f_\xi$  est croissante, positive, symétrique en  $\xi$ , bornée par la fonction  $t \mapsto \max(\alpha_\xi t, \beta_\xi)$ .

L'intérêt de considérer (4.66) plutôt que (4.65) est qu'on a déjà discrétisé le problème et qu'il n'y a plus de difficulté pour calculer la dérivée de cette énergie. Si on note  $u_x$  la valeur de  $u$  en  $x$ , on obtient ainsi la dynamique

$$\frac{du_x}{dt} = 2(u_x - u_{0x}) + \sum_{\xi} \frac{1}{h^2} f'_\xi \left( \frac{(u_x - u_{x+h\xi})^2}{h} \right) (u_{x+h\xi} - u_x) \phi(\xi). \quad (4.67)$$

Le deuxième terme du membre de droite est un terme de diffusion contrôlé par  $f'_\xi$ , de sorte que :

- Si  $(u_x - u_{x+h\xi})^2 \approx 0$ , alors  $f'_\xi = \alpha_\xi$ , ce qui va avoir tendance à "égaliser"  $u_x$  et  $u_{x+h\xi}$ .
- Si  $(u_x - u_{x+h\xi})^2$  est grand, alors  $f'_\xi = 0$ , ce qui va supprimer toute interaction.

Un exemple de résultat est donné dans la figure 4.26.

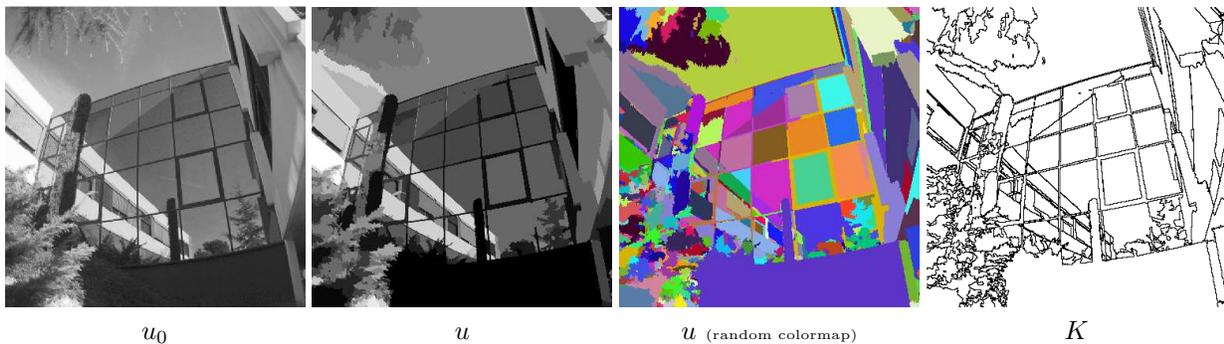


Fig. 4.26: Exemple de résultat de segmentation avec l'approche de Mumford et Shah.

Notre objectif dans cette contribution est d'établir comment les approches variationnelles (et en particulier l'approche de Mumford et Shah) pourraient être utilisées dans le contexte de la segmentation

[20] A. Chambolle and G. Dal Maso. Discrete approximations of the Mumford–Shah functional in dimension two. *M2AN*, 33(4):651–672, 1999. (also available as Technical report 9820 from Université Paris Dauphine, Ceremade).

d'images avec des réseaux de neurones. Ce rapport entre l'approche de Mumford et Shah et les réseaux de neurones a déjà été abordé par Sarti et al. [68] : les auteurs montrent que la solution de l'équation de différence de phase d'un réseau d'oscillateurs faiblement connectés,  $\Gamma$ -converge quand la dimension de la grille tend vers zéro, vers le "gradient" de l'énergie de Mumford et Shah. Dans cet esprit, en considérant également les neurones comme des oscillateurs, nous souhaitons comprendre comment définir les conditions sur les poids synaptiques pour qu'un réseau d'oscillateurs reproduise des comportements d'oscillations synchrones à l'intérieur des régions, produisant ainsi une segmentation [A 14].

Le rôle et l'utilité des synchronisations dans les trains d'impulsions reste un sujet controversé, malgré un certain nombre d'observations.

- La synchronie serait un mécanisme utilisé pour le problème du *biding*<sup>[70]</sup>, c'est-à-dire que les neurones répondraient de manière synchrone dans le système visuel, s'agissant des différents attributs d'un même objet, ce qui nous permettrait d'avoir un percept unifié combinant tous ces attributs.
- La synchronie serait aussi un mécanisme utilisé pour coder les différentes régions dans le champ visuel, c'est-à-dire faire de la segmentation. Parmi les expériences allant dans ce sens, Neuenschwander et al. [58] montrent que pour un stimulus représentant un rectangle uniforme, on observe une synchronisation des impulsions pour les cellules ganglionnaires dont les champs récepteurs correspondent au rectangle. Par contre, si on sépare le rectangle en deux parties, on observe alors deux assemblées synchrones<sup>12</sup>.

Des réseaux de neurones qui se synchronisent quand les entrées correspondent à une région homogène? Cela existe. Par exemple, l'algorithme LEGION<sup>[74]</sup> (*Locally Excitatory Globally Inhibitory Oscillator Networks*), qui avec un inhibiteur global, parvient à différencier et rendre synchrones les différentes régions d'une image. Mais une telle approche n'a pas grand sens biologique. . .

L'approche que nous proposons rentre dans le cadre de la théorie de la réduction de dimensionnalité (voir un récapitulatif dans le manuscrit de Brown<sup>[12]</sup>). On considère un réseau de neurones où chaque neurone est un oscillateur. On définit par exemple le neurone intègre-et-tire (*integrate-and-fire*, noté IF) par

$$\begin{cases} \frac{dv_i}{dt} = I + D(v), & \text{tant que } v_i < \text{Seuil}, \\ \text{Emission d'une impulsion et } v_i = 0 & \text{dès que } v_i = \text{Seuil}, \end{cases} \quad (4.68)$$

où  $I$  correspond à la dynamique propre du neurone et  $D(v)$  les entrées externes. Si on suppose que  $D(v)$  reste faible devant  $I$ , alors on peut associer à la dynamique de  $v$  une variable scalaire  $\theta \in [0, 2\pi]$ , telle que  $\theta$  croît avec le temps et passe par zéro à chaque impulsion, en suivant la dynamique

$$\frac{d\theta_i}{dt} = \omega + cD(\theta/c), \quad (4.69)$$

où  $\omega$  est la vitesse constante associée à (4.68) quand  $D(v) = 0$ .

La première contribution que nous proposons est de définir, au niveau de la phase, la dynamique qui va se traduire en potentiel grâce à l'équivalence (4.68)–(4.69) dans le cas d'un neurone de type IF. En effet, reprenons la dynamique (4.67) issue de l'approximation de Chambolle et Dal Maso (1999), et définissons

$$D(u_x - u_{x+\xi}) = \sum_{\xi} \frac{1}{h^2} f'_{\xi} \left( \frac{(u_x - u_{x+h\xi})^2}{h} \right) (u_{x+h\xi} - u_x) \phi(\xi). \quad (4.70)$$

Alors, si on définit l'encodage comme une phase relative, i.e.,

$$u_x = \theta_x - \text{clock},$$

---

<sup>12</sup>Le lien entre synchronisation et segmentation reste controversé et d'autres mécanismes pourraient être à l'origine de la synchronisation (Roelfsema, Lamme et al., 2004).

[68] A. Sarti, G. Citti, and M. Manfredini. From neural oscillations to variational problems in the visual cortex. *Journal of Physiology - Paris*, 97:379–395, 2003.

[70] W. Singer. Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24(1):49–65, 1999.

[58] S. Neuenschwander, M. Castelo-Branco, and W. Singer. Synchronous oscillations in the cat retina. *Vision Research*, 39(15):2485–2497, 1999.

[74] D. L. Wang and D. Terman. Locally excitatory globally inhibitory oscillator networks. *IEEE Trans. Neural Net.*, 6:283–286, 1995.

[12] E. T. Brown. *Neural oscillators and integrators in the dynamics of decision tasks*. PhD thesis, Faculty of Princeton University, jun 2004.

la phase  $\theta_x$  vérifie donc

$$\frac{d\theta_x}{dt} = \omega + D(\theta_x - \theta_{x+\xi}), \quad (4.71)$$

où  $D$  est défini par (4.70), et grâce à la correspondance simple (4.68)–(4.69) dans le cas de neurones de type IF, on obtient l'équation sur le potentiel

$$\frac{dv_x}{dt} = I + \sum_{\xi} \frac{1}{h^2} f'_{\xi} \left( \frac{(v_x - v_{x+h\xi})^2}{h} \right) (v_{x+h\xi} - v_x) \phi(\xi), \quad (4.72)$$

tant que  $v_x$  est inférieur au seuil. Cette équation a été implémentée et donne le résultat attendu (voir figure 4.27) : partant d'une image originale bruitée, on observe une intégration progressive de l'information et des émissions d'impulsions qui tendent à se synchroniser dans les régions homogènes.

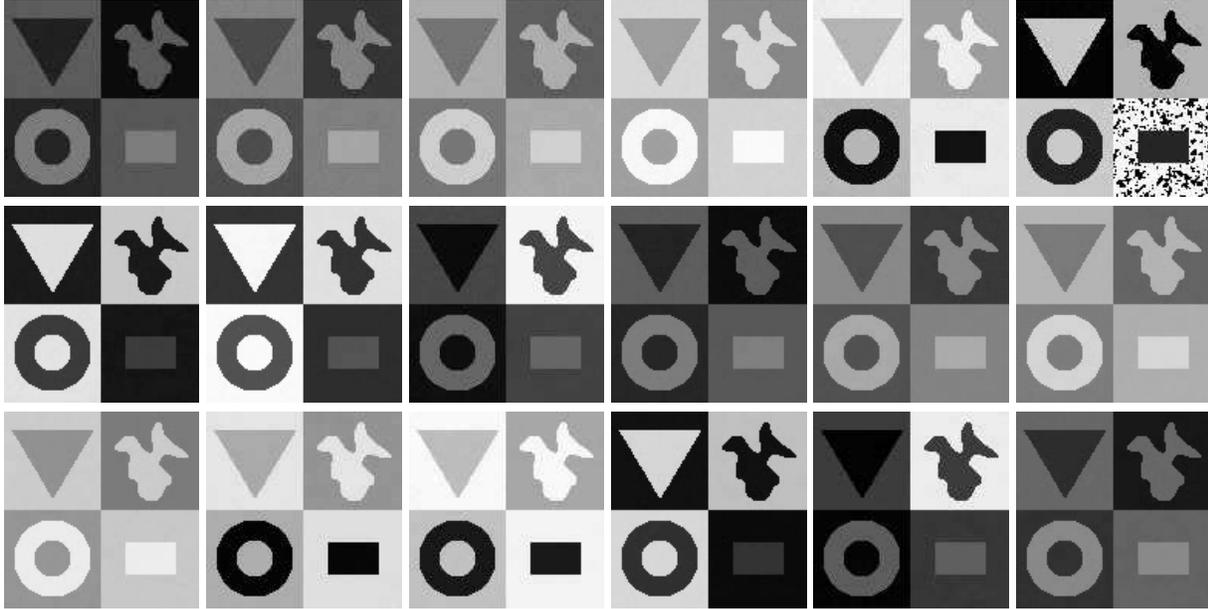


Fig. 4.27: Exemple de simulation sur une image synthétique avec l'équation (4.71), modélisant un neurone intègre-et-tire avec des connexions latérales excitatrices. Le temps évolue de haut en bas et de gauche à droite, montrant l'intégration et les émissions d'impulsions quand les potentiels des neurones sont réinitialisés (passages discontinus du blanc au noir, deux cycles sont représentés).

*L'équation d'évolution (4.72) est intéressante car elle met en jeu des calculs simples, des différences entre potentiels, ce que l'on retrouve dans de nombreux modèles de réseaux de neurones et qui peut s'expliquer du point de vue biologique. Par contre, si l'on considère des modèles de neurones plus compliqués qu'un IF, alors l'équivalence (4.68)–(4.69) devient plus complexe. On pourra toujours définir la dynamique de  $\theta$  comme on le souhaite par (4.71), mais l'équation correspondante en potentiel va faire apparaître des termes plus complexes, n'ayant plus d'interprétation biologique simple. Pour élargir ce travail, nous cherchons à mieux comprendre les propriétés du réseau qui permettent ce type de synchronisation. Pour cela, nous étudions plus précisément les dynamiques possibles entre des couples de neurones, comme dans les travaux de Lewis et Rinzel<sup>[50]</sup>.*



[50] T. J. Lewis and J. Rinzel. Dynamics of spiking neurons connected by both inhibitory and electrical coupling. *Journal of Computational Neuroscience*, 14(3):283–309, 2003.



## CHAPITRE 5

# Liste des publications après la thèse

### Livre

- LI 1 G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, January 2002.
- LI 2 G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations (Second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.

 <http://www-sop.inria.fr/books/imath>



### Chapitres de livres

- CH 1 G. Aubert and P. Kornprobst. Traitement des images numériques. In J. Akoka and I. Comyn-Wattiau, editors, *Encyclopédie de l'informatique et des systèmes d'information*, number 6, chapter 18, pages 861–879. Vuibert, November 2006.

 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/aubert-kornprobst:06b.pdf>

- CH 2 G. Aubert and P. Kornprobst. Mathematics of image processing. In J. P. Francoise, G.L. Naber, and S.T. Tsou, editors, *Encyclopedia of Mathematical Physics*, volume 3, pages 1–9. Elsevier, Oxford, nov 2006.

 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/aubert-kornprobst:06.pdf>

- CH 3 T. Brox, R. van den Boomgaard, F. Lauze, J. van de Weijer, J. Weickert, P. Mràzek, and P. Kornprobst. Adaptive structure tensors and their applications. In J. Weickert and H. Hagen, editors, *Visualization and Processing of Tensor Fields*, pages 19–47. Springer, 2006.

### Publications dans des journaux

- J 1 O. Faugeras, G. Adde, G. Charpiat, C. Chefd'Hotel, M. Clerc, T. Deneux, R. Deriche, G. Hermosillo, R. Keriven, P. Kornprobst, J. Kybic, C. Lenglet, L. Lopez-Perez, T. Papadopoulos, J.-P. Pons, F. Ségonne, B. Thirion, D. Tschumperlé, T. Viéville, and N. Wotawa. Variational, geometric, and statistical methods for modeling brain anatomy and function. *NeuroImage*, 23S1:S46–S55, 2004. Special issue: Mathematics in Brain Imaging - Edited by P.M. Thompson, M.I. Miller, T. Ratnanather, R.A. Poldrack and T.E. Nichols.

- J 2 R.R. Peeters, P. Kornprobst, M. Nikolova, S. Sunaert, T. Viéville, G. Malandain, R. Deriche, O. Faugeras, M. Ng, and P. Van Hecke. The use of superresolution techniques to reduce slice thickness in functional MRI. *International Journal of Imaging Systems and Technology (IJIST)*, *Special issue on High Resolution Image Reconstruction*, 14:131–138, 2004.

- J 3 T. Viéville, S. Chemla, and P. Kornprobst. How do high-level specifications of the brain relate to variational approaches? *Journal of Physiology - Paris*, 101(1-3):118–135, 2007.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2007/vieville-chemla-et-al:07.pdf> 
- Publications dans des conférences
- C 1 P. Kornprobst and G. Médioni. Tracking segmented objects using tensor voting. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 118–125, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2000/kornprobst-medioni:00.pdf> 
- C 2 P. Kornprobst and G. Médioni. A 2D+t tensor voting based approach for tracking. In *Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 1104–1107, Barcelona, Spain, September 2000. Computer Society Press.
- C 3 C. Lacombe, P. Kornprobst, G. Aubert, and L. Blanc-Feraud. A variational approach to one dimensional phase unwrapping. In *Proceedings of the International Conference on Pattern Recognition*, Québec City, Canada, August 2002. Computer Society Press.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2002/lacombe-kornprobst-et-al:02.ps.gz>
- C 4 P. Kornprobst, R. Peeters, M. Nikolova, R. Deriche, M. Ng, and P. Van Hecke. A superresolution framework for fmri sequences and its impact on resulting activation maps. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI2003*, volume 2 of *Lecture Notes in Computer Science*, pages 117–125. Springer-Verlag, 2003.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/kornprobst-peeters-et-al:03.ps.gz> 
- C 5 R. Deriche, P. Kornprobst, M. Nikolova, and M. Ng. Half-quadratic regularization for mri image restoration. In *ICASSP'03*, pages 585–588, 2003.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/deriche-kornprobst-et-al:03.ps.gz>
- C 6 C. Lacombe, G. Aubert, L. Blanc-Féraud, and P. Kornprobst. Filtering interferometric phase images by anisotropic diffusion. In *Proceedings of the International Conference on Image Processing*. IEEE Signal Processing Society, 2003.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/lacombe-aubert-et-al:03b.ps.gz> 
- C 7 C. Lacombe, G. Aubert, L. Blanc-Féraud, and P. Kornprobst. Filtrage adaptatif des interférogrammes par diffusion anisotrope. In *Journées Orasis 2003*, 2003.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/lacombe-aubert-et-al:03.ps.gz>
- C 8 F. Lauze, P. Kornprobst, and E. Mémin. A coarse to fine multiscale approach for linear least squares optical flow estimation. In *A Coarse To Fine Multiscale Approach For Linear Least Squares Optical Flow Estimation*, volume 2, pages 767–776, 2004.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2004/lauze-kornprobst-et-al:04b.pdf> 
- C 9 F.B. Lauze, P. Kornprobst, C. Lenglet, R. Deriche, and M. Nielsen. Sur quelques méthodes de calcul de flot optique à partir du tenseur de structure : Synthèse et contribution. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, 2004.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2004/lauze-kornprobst-et-al:04.pdf> 
- C 10 P. Kornprobst, T. Viéville, and I. Dimov. Could early visual processes label and segment motions? In *17th IMACS World Congress, Scientific Computation, Applied Mathematics and Simulation*, 2005.
- C 11 P. Kornprobst, T. Viéville, and I. Dimov. Could early visual processes be sufficient to label motions? In *International Joint Conference on Neural Networks*, 2005.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2005/kornprobst-vieville-et-al:05.pdf> 

- C 12 A. Wohrer, P. Kornprobst, and T. Vieville. From light to spikes: a large-scale retina simulator. In *International Joint Conference on Neural Networks*, Vancouver, 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/wohrer-kornprobst-etal:06.pdf> 
- C 13 M.-J. Escobar, A. Wohrer, P. Kornprobst, and T. Vieville. Biological motion recognition using an mt-like model. In *Proceedings of 3rd Latin American Robotic Symposium*, 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/escobar-wohrer-etal:06b.pdf> 
- C 14 P. Kornprobst, S. Chemla, O. Rochel, and T. Vieville. A 1st step towards an abstract view of computation in spiking neural-networks. In *1ère conférence francophone de Neurosciences Computationnelles*, October 2006.
- C 15 T. Viéville and P. Kornprobst. Modeling cortical maps with feed-backs. In *International Joint Conference on Neural Networks*, Vancouver, 2006.
- C 16 A. Wohrer, P. Kornprobst, and T. Vieville. Contrast gain control through a feedback in the retina. In *1ère conférence francophone de Neurosciences Computationnelles*, October 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/wohrer-kornprobst-etal:06b.pdf> 
- C 17 A. Ramirez-Manzanares, M. Rivera, P. Kornprobst, and F. Lauze. A variational approach for multi-valued velocity field estimation in transparent sequences. In *Proceedings of the Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCS*, pages 227–238, 2007.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2007/ramirez-manzanares-rivera-etal:07.pdf> 
- C 18 M.-J. Escobar, A. Wohrer, P. Kornprobst, and T. Vieville. Biological motion recognition using an mt-like model. In *1ère conférence francophone de Neurosciences Computationnelles*, October 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/escobar-wohrer-etal:06c.pdf>
- Communications avec soumission d'abstract
- A 1 C. Barthou, D. Bouvier, O. Faugeras, P. Kornprobst, R. Keriven, and T. Papadopoulo. A level set method for the inverse EEG/MEG problem. In *Applied Inverse Problems: Theoretical and Computational Aspects*, page 127, Montecatini Terme, Italy, June 2001.
- A 2 R.R. Peeters, P. Kornprobst, S. Sunaert, T. Viéville, O. Faugeras, and P. Van Hecke. The use of superresolution reconstruction algorithms to enhance spatial resolution in fMRI. In *ISMRM 11th Scientific Meeting and Exhibition*. International Society for Magnetic Resonance in Medicine, 2003.
- A 3 A. Wohrer, P. Kornprobst, and T. Viéville. A biologically-inspired spiking retina model for the encoding of visual sequences. In *European Conference on Visual Perception*, 2005.
- A 4 T. Viéville and P. Kornprobst. How fast-brain object categorization allows top-down processes of segmentation. In *European Conference on Visual Perception*, 2005.
- A 5 M.-J. Escobar, A. Wohrer, P. Kornprobst, and T. Vieville. Biological motion recognition using an mt-like model. In *Proceedings of 3rd Latin American Robotic Symposium*, 2006.
- A 6 P. Kornprobst, F. Chavane, S. Chemla, A. Reynaud, and T. Vieville. Reverse-engineering of the visual brain cortical maps computation using optical-imaging. In *29th European Conference on Visual Perception*, page 54, aug 2006.
- A 7 P. Kornprobst, T. Vieville, S. Chemla, and O. Rochel. Modeling cortical maps with feed-backs. In *29th European Conference on Visual Perception*, page 53, aug 2006.
- A 8 A. Wohrer, G. Masson, L. Perrinet, P. Kornprobst, and T. Vieville. Contrast sensitivity adaptation in a virtual spiking retina and its adequation with mammals retinas. In *29th European Conference on Visual Perception*, page 67, aug 2006.
- A 9 T. Viéville and P. Kornprobst. Modeling cortical maps with feed-backs. In *International Conf. on Cognitive and Neural Systems*, 2006.

- A 10 Maria-Jose Escobar, Pierre Kornprobst, and Thierry Vieville. Spike to spike mt model and applications. In *Sixteenth Annual Computational Neuroscience Meeting (CNS)*, jul 2007.
- A 11 S. Chemla, F. Chavane, T. Vieville, and P. Kornprobst. Biophysical cortical column model for optical signal analysis. In *Sixteenth Annual Computational Neuroscience Meeting (CNS)*, jul 2007.
- A 12 M-J Escobar, G S Masson, Vieville T, and P Kornprobst. Spiking mt model: Dynamics and motion patterns. In *Perception 36 ECVF Abstract Supplement*, 2007.
- A 13 É. Tlapale, G. Masson, T. Viéville, and P. Kornprobst. Model of motion field diffusion controlled by form cues. In *Perception 36 ECVF Abstract Supplement*, sep 2007.
- A 14 L. Gérard, P. Kornprobst, and T. Viéville. From variational to spiking network image-segmentation techniques. In *Perception 36 ECVF Abstract Supplement*, 2007.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2007/gerard-kornprobst-et-al:07.pdf>

#### Rapports de recherche

- R 1 O. Faugeras, F. Clément, R. Deriche, R. Keriven, T. Papadopoulo, J. Roberts, T. Viéville, F. Devernay, J. Gomes, G. Hermosillo, P. Kornprobst, and D. Lingrand. The inverse EEG and MEG problems: The adjoint space approach I: The continuous case. Technical Report 3673, INRIA, May 1999.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-3673.html>
- R 2 P. Kornprobst, R. Peeters, T. Viéville, G. Malandain, S. Mierisova, S. Sunaert, O. Faugeras, and P. Van Hecke. Superresolution in MRI and its influence in statistical analysis. Technical Report 4513, INRIA, July 2002.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-4513.html>
- R 3 J. Bullier, R. Deriche, O. Faugeras, D. Fieze, P. Girard, R. Guyonneau, P. Kornprobst, T. Papadopoulo, S. Thorpe, and T. Viéville. Rivage feedback during visual integration : towards a generic architecture. Technical Report 5451, INRIA, 2004.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-5451.html>
- R 4 I. Dimov, P. Kornprobst, and T. Viéville. Could early visual processes be sufficient to label motions? Technical Report 5240, INRIA, jun 2004.
- 📄 <http://www-sop.inria.fr/rapports/sophia/RR-5240.html>
- R 5 É. Tlapale, J.-B. Bernard, E. Castet, and P. Kornprobst. Le projet solaire: Un système d'optimisation de la lecture par asservissement de l'image au regard. Technical Report RT-0326, INRIA, oct 2006.
- 📄 <http://hal.inria.fr/inria-00102542>
- R 6 P. Kornprobst and G. Aubert. Explicit reconstruction for image inpainting. Research Report 5905, INRIA, April 2006.
- 📄 <http://www.inria.fr/rrrt/rr-5905.html>
- R 7 A. Wohrer, P. Kornprobst, and T. Viéville. A biologically-inspired model for a spiking retina. Technical Report 5848, INRIA, February 2006.
- 📄 <http://www.inria.fr/rrrt/rr-5848.html>
- R 8 Adrien Wohrer, Pierre Kornprobst, and Thierry Viéville. Virtual retina: a biological retina model and simulator, with contrast gain control. Research Report 6243, INRIA, jul 2007.
- 📄 <https://hal.inria.fr/inria-00160716> 
- R 9 G. Aubert and P. Kornprobst. New algorithm for solving variational problems in  $W^{(1,p)}$  and BV: Application to image restoration. Technical Report RR-6245, INRIA, jul 2007.

 <https://hal.inria.fr/inria-00161706>



R 10 M.-J. Escobar, G. Masson, T. Vieville, and P. Kornprobst. Spike to spike model and applications: A biological plausible approach for the motion processing. Research Report 6280, INRIA, 09 2007.

 <https://hal.inria.fr/inria-00170153>



R 11 É. Tlapale, G. Masson, and P. Kornprobst. Biological model of motion integration and segmentation based on form cues. Technical Report RR-6293, INRIA, sep 2007.

 <http://hal.inria.fr/inria-00172412>



#### Articles soumis dans des revues

- S 1 Travail correspondant à [R 9], soumis à SIAM Journal of Numerical Analysis
- S 2 Travail correspondant à [C 13], soumis à International Journal of Computer Vision
- S 3 Travail correspondant au tutorial donné à SIGGRAPH 2007, présentant une revue du filtre bilateral, soumis à Foundations and Trends in Computer Graphics and Vision
- S 4 Travail correspondant à [C 12], [R 8], soumis à Journal of Computational Neuroscience
- S 5 Travail correspondant à [C 17], soumis à Journal of the Optical Society of America A



## APPENDICE A

# Principaux développements logiciels

Les développements de logiciels ont été décrits dans le curriculum-vitæ. Dans ce chapitre, il s'agit de donner plus de détails sur les deux principaux.

### **Systèmes d'aide à la vision pour les patients atteints de scotomes**

*Collaboration avec Eric Castet et Jean-Baptiste Bernard (INCM), réalisé avec Emilien Tlapale (Master, UNICE)*

Les causes habituelles de l'apparition de scotomes sont la maladie de démyélinisation, les substances toxiques, les carences nutritives et des blocages vasculaires dans la rétine ou dans le nerf optique. Mais la cause la plus importante est sans doute la dégénérescence maculaire, avec 13 millions de cas aux USA et 1 million en France. Ces nombres sont en augmentation et, selon les spécialistes, ils devraient être triplés d'ici 25 ans.

Un problème majeur rencontré par les déficients visuels qui sont aveugles dans la portion centrale de leur champ visuel est la lecture de textes. Les systèmes actuels d'aide à la lecture sont pour l'essentiel des systèmes de télé-agrandisseurs, et ils ne tirent pas partie de l'essor récent des recherches et technologies concernant la vision, telles que la vision par ordinateur ou la réalité augmentée. Notre objectif était donc de proposer des alternatives aux systèmes existants. Mais développer de tels systèmes n'est possible que grâce à d'étroites collaborations avec psychophysiciens et orthoptistes, qui connaissent les réelles difficultés des patients et ont la possibilité de tester avec eux des solutions pour les aider. C'est dans cet esprit que le travail avec Eric Castet (INCM) a été entrepris.

Nous travaillons sur deux systèmes.

- Le premier système s'inscrit dans le cadre du projet SOLAIRE (Système d'Optimisation de la Lecture par Asservissement de l'Image au Regard) porté par Eric Castet (INCM), nous développons un système de "vision augmentée" dédié à la lecture en combinant les compétences d'ophtalmologistes, de chercheurs en neurosciences visuelles et de spécialistes de vision artificielle. L'idée maîtresse de ce projet est d'asservir l'image du texte lu à la position instantanée du regard en adaptant le traitement de l'image aux caractéristiques de la pathologie propre à chaque patient. Le rapport [R 5] en décrit la mise en œuvre et la figure A.1 en illustre le principe. Nous avons réalisé un prototype, mais la mise en œuvre d'un tel système dans des conditions réelles d'utilisation reste à faire.
- Le second système est un logiciel dont le but est de traiter des documents PDF, partant du principe que de plus en plus d'hebdomadaires ou documents sont désormais disponibles en PDF. Contrairement au précédent, il n'est pas asservi au regard. Ce logiciel permet de naviguer dans un document PDF et de rentrer dans un mode de texte rehaussé (analogue au système précédent), ce qui permet aux patients de lire avec plus de confort des paragraphes de leur choix. Le logiciel a été mis à disposition de l'hôpital de la Timone à Marseille pour évaluation, permettant aux patients de le comparer à des systèmes existants.

*Ce travail est en cours. Il s'agit côté INRIA d'un travail de développement, dans lequel nous intégrons les* 

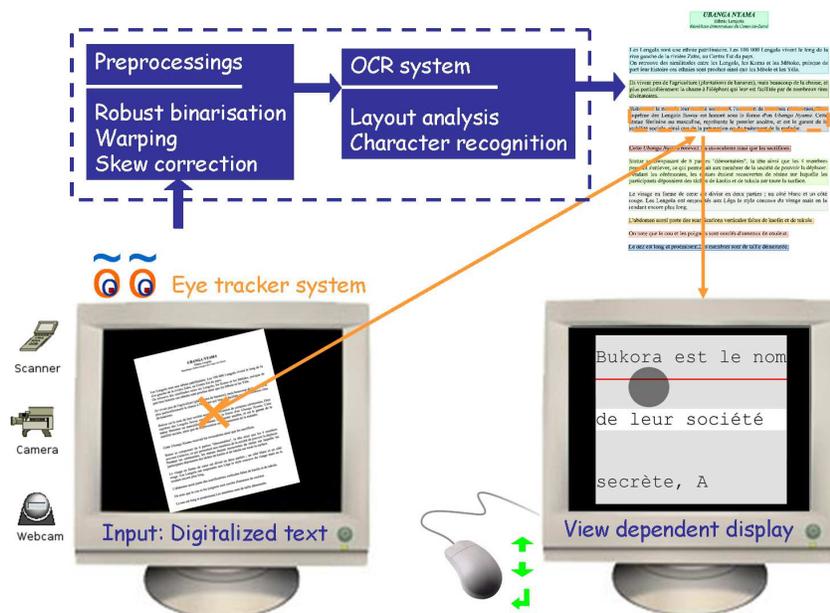


Fig. A.1: Systèmes d'aide à la vision pour les patients atteints de scotomes. Ce système décrit dans [R 5] permet de traiter un document observé via un capteur (type caméra ou scanner) et de proposer une vision rehaussée aux patients. Le système comprend plusieurs modules de traitement d'images dont la reconnaissance de caractères.

résultats des études psychophysiques menées par les collègues de Marseille. Les retours des patients seront aussi déterminants pour l'évolution de ce système. Par la mise en application d'algorithmes efficaces de vision par ordinateur, on espère donc offrir un outil d'aide à la lecture pour ces patients atteints de scotome. Cela nous permet aussi d'appréhender le monde de la psychophysique, et des difficultés rencontrées pour mesurer et améliorer la vitesse de lecture et comprendre quelles seraient des stratégies de lecture efficace.

## Virtual Retina

Cadre contractuel : projet Européen FACETS ("Fast Analog Computing with Emergent Transient States", EC IP project FP6-015879, 2005–2009). FACETS est un grand consortium entre plusieurs communautés scientifiques visant développer les connaissances sur le système nerveux. Le travail décrit dans cette section correspond au travail de thèse d'Adrien Wohrer (2004–2007) que j'encadre avec Thierry Vieville.

"Virtual Retina" est un simulateur de rétine qui transforme une vidéo en trains d'impulsions (*spikes*). Le modèle sous-jacent est bio-inspiré : chaque élément correspond à une réalité physiologique. Le modèle comprend une partie de filtrage linéaire (*Outer Plexiform Layer*), un mécanisme de contrôle de gain, et la génération de trains d'impulsions correspondant aux cellules ganglionnaires. Les détails du modèle sont présentés dans la section 4.2, correspondant aux publications [C 16], [R 7].

Les principales caractéristiques du simulateur "Virtual Retina" sont :

- Possibilité de réaliser des simulations grande échelle, pouvant aller jusqu'à 100.000 cellules ganglionnaires en sortie.
- Facilité de configuration des différentes composantes de la rétine par des fichiers XML (voir la Figure A.2).
- Modélisation des cellules Magno et Parvo par ajustement des paramètres du modèle.
- Reproduction de la géométrie de la rétine, avec des caractéristiques de cellules respectant les propriétés de la fovéa et sa périphérie.

- Modèle dynamique de contrôle de gain, réalisé par une inhibition d'un type de cellules amacrines sur les cellules bipolaires.
- Possibilité d'ajouter des microsaccades dans le signal d'entrée.

```

<retina-description-file>

  <retina
    temporal-step__sec ="0.005"
    input-luminosity-range="255"
    pixels-per-degree="10.0">

    <log-polar-scheme
      fovea-radius__deg="50"
      scaling-factor-outside-fovea__inv-deg="0.02"/>

    <outer-plexiform-layer>
      <linear-version
        sigma-center__deg="0.10"
        additional-sigma-surround__deg="0.30"
        tau-center__sec="0.01"
        additional-tau-surround__sec="0.01"
        opl-amplification="5"
        relative-weight-surround-center="1"
        leaky-heat-equation="1" />
      </outer-plexiform-layer>

    <contrast-gain-control
      input-amplification-from-opl-signal="250"
      bipolar-to-amacrine-threshold="0.5"
      sigma-amacrine-cells__deg="0.5"
      additional-tau-amacrine-cells__sec="0.03"
      amacrine-feedback-amplification="500"
      inert-leak-in-bipolar-cells="10"/>

    <parvocellular-ganglion-layer
      additional-tau-transient__sec="0.03"
      relative-transient-weight="0.75"
      linear-threshold-from-bipolar-signal="0"
      value-at-linear-threshold="300"
      input-amplification-from-bipolar-signal="100">
      <spiking-channel>
        <circular-spiking-channel
          diameter="300" fovea-density="1" g-leak="50"
          noise-V="0.0" noise-refr__sec="0"
          tau-refr__sec="0.003" random-init="1"/>
        </spiking-channel>
      </parvocellular-ganglion-layer>

  </retina>
</retina-description-file>

```

Fig. A.2: Exemple de fichier XML pour la définition des paramètres de rétine.

 <http://www-sop.inria.fr/odyssee/software/virtualretina>  
 APP logiciel Virtual Retina: IDDN.FR.001.210034.000.S.P.2007.000.31235



## APPENDICE B

# Documents joints

Livre

- LI 2 G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations (Second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.

📄 <http://www-sop.inria.fr/books/imath>



Publications dans des journaux

- J 3 T. Viéville, S. Chemla, and P. Kornprobst. How do high-level specifications of the brain relate to variational approaches? *Journal of Physiology - Paris*, 101(1-3):118–135, 2007.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2007/vieville-chemla-et-al:07.ps>



Publications dans des conférences

- C 1 P. Kornprobst and G. Médioni. Tracking segmented objects using tensor voting. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 118–125, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2000/kornprobst-medioni:00.pdf>



- C 4 P. Kornprobst, R. Peeters, M. Nikolova, R. Deriche, M. Ng, and P. Van Hecke. A superresolution framework for fmri sequences and its impact on resulting activation maps. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI2003*, volume 2 of *Lecture Notes in Computer Science*, pages 117–125. Springer-Verlag, 2003.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/kornprobst-peeters-et-al:03.ps.gz>



- C 6 C. Lacombe, G. Aubert, L. Blanc-Féraud, and P. Kornprobst. Filtering interferometric phase images by anisotropic diffusion. In *Proceedings of the International Conference on Image Processing*. IEEE Signal Processing Society, 2003.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2003/lacombe-aubert-et-al:03b.ps.gz>



- C 8 F. Lauze, P. Kornprobst, and E. Mémin. A coarse to fine multiscale approach for linear least squares optical flow estimation. In *A Coarse To Fine Multiscale Approach For Linear Least Squares Optical Flow Estimation*, volume 2, pages 767–776, 2004.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2004/lauze-kornprobst-et-al:04b.pdf>



- C 9 F.B. Lauze, P. Kornprobst, C. Lenglet, R. Deriche, and M. Nielsen. Sur quelques méthodes de calcul de flot optique à partir du tenseur de structure : Synthèse et contribution. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, 2004.

📄 <ftp://ftp-sop.inria.fr/odyssee/Publications/2004/lauze-kornprobst-et-al:04.pdf>



- C 11 P. Kornprobst, T. Viéville, and I. Dimov. Could early visual processes be sufficient to label motions? In *International Joint Conference on Neural Networks*, 2005.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2005/kornprobst-vieville-et-al:05.pdf> 
- C 12 A. Wohrer, P. Kornprobst, and T. Vieville. From light to spikes: a large-scale retina simulator. In *International Joint Conference on Neural Networks*, Vancouver, 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/wohrer-kornprobst-et-al:06.pdf> 
- C 13 M.-J. Escobar, A. Wohrer, P. Kornprobst, and T. Vieville. Biological motion recognition using an mt-like model. In *Proceedings of 3rd Latin American Robotic Symposium*, 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/escobar-wohrer-et-al:06b.pdf> 
- C 16 A. Wohrer, P. Kornprobst, and T. Vieville. Contrast gain control through a feedback in the retina. In *1ère conférence francophone de Neurosciences Computationnelles*, October 2006.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2006/wohrer-kornprobst-et-al:06b.pdf> 
- C 17 A. Ramirez-Manzanares, M. Rivera, P. Kornprobst, and F. Lauze. A variational approach for multi-valued velocity field estimation in transparent sequences. In *Proceedings of the Scale Space and Variational Methods in Computer Vision*, volume 4485 of *LNCS*, pages 227–238, 2007.  
 <ftp://ftp-sop.inria.fr/odyssee/Publications/2007/ramirez-manzanares-rivera-et-al:07.pdf> 
- Rapports de recherche
- R 8 Adrien Wohrer, Pierre Kornprobst, and Thierry Viéville. Virtual retina: a biological retina model and simulator, with contrast gain control. Research Report 6243, INRIA, jul 2007.  
 <https://hal.inria.fr/inria-00160716> 
- R 9 G. Aubert and P. Kornprobst. New algorithm for solving variational problems in  $W^{(1,p)}$  and BV: Application to image restoration. Technical Report RR-6245, INRIA, jul 2007.  
 <https://hal.inria.fr/inria-00161706> 
- R 10 M.-J. Escobar, G. Masson, T. Vieville, and P. Kornprobst. Spike to spike model and applications: A biological plausible approach for the motion processing. Research Report 6280, INRIA, 09 2007.  
 <https://hal.inria.fr/inria-00170153> 
- R 11 É. Tlapale, G. Masson, and P. Kornprobst. Biological model of motion integration and segmentation based on form cues. Technical Report RR-6293, INRIA, sep 2007.  
 <http://hal.inria.fr/inria-00172412> 