

HABILITATION À DIRIGER DES RECHERCHES

spécialité : **Mathématiques**

**Modèles continus, numériques et statistiques
pour l'analyse d'images**

Lionel MOISAN

— Jury —

M. Albert COHEN
M. Guy DAVID
M. Olivier FAUGERAS
M. Donald GEMAN
M. Jean-Michel GHIDAGLIA
M. Yves MEYER
M. Jean-Michel MOREL

janvier 2003

à Claire.

Je voudrais tout d'abord remercier Yves Meyer, dont l'enthousiasme si communicatif n'a cessé de me soutenir depuis mes premiers pas dans la recherche, et en particulier pendant la rédaction de ce mémoire. Qu'il sache combien, au-delà de l'estime que je lui porte, je lui suis reconnaissant de présenter aujourd'hui mes travaux.

Don Geman et Olivier Faugeras m'ont fait l'honneur d'être les rapporteurs de ce mémoire. Je les remercie profondément de cette attention portée à mon travail. Toute ma reconnaissance va également à Albert Cohen, Guy David et Jean-Michel Ghidaglia pour avoir accepté de faire partie de ce jury.

Une partie importante des recherches présentées dans ce mémoire a été réalisée avec Jean-Michel Morel et Agnès Desolneux. Qu'il me soit permis de les remercier ici pour toutes les discussions passionnantes que nous avons eues ensemble. Je ne saurais comment exprimer ma gratitude à Jean-Michel Morel pour la confiance qu'il m'a accordée ces dernières années et pour tout ce que j'ai pu apprendre à son contact.

Travailler en collaboration est une aventure stimulante et enrichissante, tant sur le plan scientifique que sur le plan humain. Merci à tous ceux avec qui j'ai eu la chance de partager cette aventure, notamment Frédéric Cao, Georges Koepfler, Saïd Ladjal, José-Luis Lisani, Pascal Monasse, Bérenger Stival, Angélique Videau, Emmanuel Villéger et Dan Zhang.

Je voudrais aussi témoigner ma profonde sympathie à Vicent Caselles et Bernard Rougé, et les remercier pour les nombreuses idées que nous avons pu échanger et pour la passion de la recherche qu'ils ont contribué à me faire partager.

Je ne saurais citer ici tous les autres mathématiciens qui, d'une façon ou d'une autre, ont influencé ce travail. Qu'ils me pardonnent de ne pas les nommer, et qu'ils soient assurés de ma sincère reconnaissance pour leur aide discrète mais précieuse.

L'ambiance d'un laboratoire tient pour une bonne part à ses doctorants. Merci à eux, et à tous les membres du CMLA qui ont su insuffler la bonne humeur dans nos murs. Je voudrais en particulier témoigner mon amitié à Corinne Vachier, Laure Quivy, François Alouges, Laurent Younes, Christophe Labourdette, Pascal Bringas, Véronique Almadovar, Micheline Brunetti et Zahia Zendagui.

Enfin, pour m'avoir invité à participer à la naissance de Digital Optics, merci à Frédéric Guichard et à Jérôme Ménière, ainsi qu'à Laurent Chanas, Jean-Marc Lavest, Bruno Liège et tous les autres membres de Vision IQ avec qui j'ai eu le plaisir de partager cette belle aventure industrielle.

Table des Matières

1	Introduction	7
1.1	L'image physique et l'image numérique	7
1.2	L'image mathématique : une idéalisation nécessaire	8
2	Modèles mathématiques	9
2.1	Les spécificités de l'image	9
2.2	Le modèle de Shannon	10
2.2.1	Du bon échantillonnage des images	10
2.2.2	Extrapolation de spectre	12
2.3	Le modèle variationnel	12
2.3.1	Régularisation	13
2.3.2	Variation totale	13
2.3.3	Variation totale pondérée	13
2.4	Equations aux Dérivées Partielles et scale-space	14
2.4.1	La démarche axiomatique : exemple de la reconstruction du relief	15
2.4.2	Le principe morphologique	18
2.4.3	Mouvements par courbure	19
2.4.4	Application à la comparaison morphologique et affine-invariante d'images	20
3	Modèles numériques	21
3.1	Le modèle de Shannon	22
3.2	Le pixel discret	23
3.3	Le pixel infinitésimal	24
3.4	Discrétisation des EDP morphologiques	25
3.5	L'image géométrique	26
3.5.1	Un algorithme rapide pour le scale-space affine	27
3.5.2	Généralisation à d'autres mouvements par courbure	29
4	Analyse d'images et modélisation statistique	31
4.1	La révolution statistique en marche	31
4.2	Le modèle bayésien	32
4.3	Les conclusions de la théorie de la Gestalt	34
4.4	Le problème des paramètres	35
4.5	Une approche systématique : la détection a contrario	36
4.5.1	Formalisation du principe de Helmholtz	37
4.5.2	L'image est son propre modèle	39
4.5.3	Maximalité	40
4.6	Vision humaine et vision artificielle	42
4.7	Vers des modèles sans paramètres ?	43
5	Modélisation et expérimentation	44
6	Axes de recherche	46
6.1	Détection systématique de structures surdéterminées	46
6.2	Echantillonnage et modèles numériques	46
	Publications et Brevets	49
	Références	53
	Publications jointes	59

Modèles continus, numériques et statistiques pour l'analyse d'images

1 Introduction

Que dire du traitement d'images¹ après plus de trente ans d'existence ? Certes, l'un des objectifs les plus ambitieux des années 70 n'a pas encore été atteint : la vision artificielle, qui devait mener à la création de robots "intelligents", capables notamment de se mouvoir sans difficulté et de reconnaître des objets, reste un problème non résolu à ce jour, malgré les travaux fondateurs de David Marr [70] et les nombreuses recherches qui ont suivi. Cependant, il est indéniable que de nombreuses percées ont été accomplies, et ce dans tous les domaines du traitement d'images : pour n'en donner que quelques exemples, citons la décomposition en ondelettes [16, 23, 73, 68], désormais essentielle en compression (algorithmes EZW [91], JPEG 2000 [57]), la restauration (déconvolution et débruitage) par minimisation de la variation totale [84, 85] ou par seuillage des coefficients d'ondelettes [27], la représentation multiéchelle d'images (*scale-space*) [99], linéaire [81] ou morphologique [2], et, pour donner un dernier exemple, l'exploration systématique de la géométrie épipolaire en stéréovision [33].

Un des aspects les plus frappants de ces théories est la diversité des techniques mathématiques auxquelles elles se rattachent : analyse de Fourier, ondelettes, statistiques, calcul des variations, équations aux dérivées partielles, géométrie projective, etc. De même qu'en physique, où une même entité reçoit couramment plusieurs modélisations différentes (la lumière, par exemple, est représentée tantôt par une onde électromagnétique, tantôt par un flux de photons), on peut facilement se convaincre de la nécessité d'utiliser plusieurs modèles mathématiques différents pour la représentation des images : fonction à bande limitée, fonction à variation bornée, champ de Markov, etc. sont autant de représentations (c'est-à-dire d'interprétations) pertinentes du tableau de nombres que constitue une image numérique. Cette variété, un peu déconcertante à première vue, se révèle en fin de compte d'une grande richesse puisqu'elle permet d'aborder un même problème sous des angles différents, avec les multiples interprétations croisées que cela autorise.

1.1 L'image physique et l'image numérique

D'un point de vue purement physique, une image est une distribution de lumière focalisée sur un plan. Cette distribution peut elle-même être décomposée selon la longueur d'onde (400-800 nm pour le spectre visible), afin de distinguer des couleurs. Bien que cette décomposition requiert *a priori* un espace de dimension infinie, on sait en pratique que la perception humaine

¹Nous employons ici le terme "traitement d'images" dans son acceptation la plus large, recouvrant toutes les techniques visant à transformer ou analyser une image.

se joue dans un espace de dimension 3 : d'une part, la couleur est analysée sur notre rétine grâce à trois types de cellules (les cônes), sensibles à des domaines spectraux différents ; d'autre part, toutes les perceptions colorimétriques de l'œil humain peuvent être reproduites par combinaison de trois sources lumineuses indépendantes dont on fait varier l'intensité² [32].

Une image numérique, quant à elle, est généralement constituée d'un tableau de valeurs qui représentent l'intensité lumineuse en chacun des points d'une grille discrète rectangulaire. Si l'on s'intéresse à des images couleur, il y a alors autant de tableaux que de composantes spectrales, soit en général 3 pour les images usuelles. Le lien entre ces tableaux de valeurs et l'image physique qu'ils reflètent dépend notamment du capteur qui a permis de mesurer ces valeurs, et de l'optique qui a focalisé l'image. Nous reviendrons plus tard sur cette correspondance.

La couleur pose des problèmes très intéressants en traitement d'images, mais nous n'aurons pas le loisir de les aborder ici. Nous nous limiterons donc au cas d'un espace de dimension 1, c'est-à-dire à des images monochromes produites par une certaine pondération des différentes composantes spectrales.

1.2 L'image mathématique : une idéalisation nécessaire

Lorsqu'il s'agit de définir des algorithmes de traitement d'images, on se trouve tiraillé entre les deux définitions précédentes : l'image discrète est exactement l'objet sur lequel travaille un algorithme numérique, alors que l'image physique répond plus à notre intuition de la réalité que transcrit l'image, à savoir la projection d'un univers tridimensionnel continu.

Mais le choix ne se limite pas à cette alternative. D'un point de vue mathématique, une abstraction de l'image est souvent nécessaire, afin de dépasser le cadre assez limité de l'image discrète. Ce n'est qu'au prix d'une étape de modélisation, où l'on choisit de représenter l'image par des objets mathématiques précis mais idéalisés, que l'on parvient à approfondir notre compréhension du traitement d'images, en exploitant la formidable capacité des mathématiques à mettre en relation des modèles et à établir de véritables dictionnaires qui permettent une lecture croisée de nombreux algorithmes. Cette étape de modélisation n'est pas immédiate : alors que l'image numérique est un objet discret relativement limité (un tableau de nombres), les modèles que l'on peut lui associer ne manquent pas et l'étape de modélisation est souvent déterminante. Il s'agit non seulement de sélectionner des modèles pertinents, au sens où ils transposent une certaine réalité physique sous-jacente, mais aussi intéressants, c'est-à-dire permettant des avancées théoriques et algorithmiques importantes. Il y a un compromis permanent à faire entre, d'une part, l'idéalisation et la simplification mathématique du modèle, qui alimente la réflexion théorique, et, d'autre part, l'adéquation du modèle à la réalité physique de l'image.

La modélisation mathématique ne suffit pas, car elle ne permet pas à elle seule de mettre en correspondance les objets idéaux qu'elle manipule avec la réalité discrète des images numériques. Elle doit être complétée par une véritable *modélisation numérique*, qui permet de transcrire les

²Cette combinaison peut cependant parfois nécessiter des intensités "négatives", évidemment impossibles à réaliser directement. Dans le cas d'un modèle relatif, où l'on cherche à reproduire une couleur-test à partir de trois sources, on réalise une intensité négative en ajoutant la source concernée à la couleur-test au lieu de la faire contribuer à la couleur synthétisée.

équations abstraites en algorithmes effectifs. Tout comme la modélisation mathématique, cette modélisation numérique est soumise à un compromis : elle doit non seulement transcrire le plus fidèlement possible le modèle mathématique dont elle est issue, mais aussi conduire à une interprétation réaliste de l'image numérique. Prenons un exemple : si l'on souhaite *effectivement* modéliser une image par une fonction à variation bornée de \mathbb{R}^2 dans \mathbb{R} (qui mesure en chaque point l'intensité lumineuse), il faut définir une correspondance entre ce type de fonction et le tableau de nombres que représente l'image numérique (dans ce cas précis, nous verrons qu'une solution intéressante consiste à utiliser la décomposition de l'image en lignes de niveau). Cette dernière étape, parfois négligée, est dans certains cas aussi importante que la modélisation mathématique elle-même : nous aurons l'occasion d'illustrer ce constat à plusieurs reprises, et nous verrons que certains modèles théoriques doivent être assouplis pour pouvoir être mis en œuvre numériquement, alors que d'autres, au contraire, ne révèlent tout leur intérêt qu'au prix d'une transcription numérique littérale.

Dans la section 2, je décrirai quelques modèles mathématiques utilisés en traitement d'images, afin de présenter mes propres travaux dans leur contexte. Je m'intéresserai ensuite aux modèles numériques, dont j'essaierai de dégager les limites respectives, dans la perspective du compromis que nous venons d'évoquer. J'aborderai ensuite mes recherches les plus récentes en analyse d'images, qui s'appuient sur une modélisation statistique particulière, la détection *a contrario*. Je conclurai enfin par quelques réflexions personnelles sur le rôle de l'expérimentation numérique, avant de présenter dans une dernière partie deux axes de recherche que je souhaite développer dans les années à venir.

2 Modèles mathématiques

2.1 Les spécificités de l'image

Historiquement, le traitement d'images a tout d'abord été fortement influencé par le traitement du signal (signaux électriques, en particulier liés au codage du son, puis de la télévision), et par l'analyse harmonique (décomposition de Fourier) qui lui était alors systématiquement associée. Une grande part de la théorie monodimensionnelle se transpose sans difficulté aux images : les notions d'échantillonnage (théorème de Shannon), de représentation spectrale, de filtre linéaire, de réponse impulsionnelle, etc. sont essentiellement les mêmes. Certaines difficultés apparaissent cependant, notamment lorsque la notion de causalité entre en jeu : la relation d'ordre naturellement présente sur la droite réelle n'a plus d'équivalent dans le plan. Ceci est à l'origine des principales difficultés que l'on rencontre dans l'utilisation de modèles autorégressifs (ARMA) en analyse et synthèse de textures [15, 18, 58] et plus généralement lorsque l'on cherche à généraliser l'extension d'une fonction de type positif (extrapolation d'une covariance par exemple) de la dimension 1 [22, 45, 61, 87] à la dimension 2 [41, 64].

Le passage à la dimension 2 entraîne certes la perte de la relation d'ordre, mais permet en contrepartie l'émergence de la géométrie. Son rôle prépondérant dans notre perception des images explique probablement pourquoi, depuis plus de 20 ans, elle occupe une place de plus en plus importante dans les modèles mathématiques. Parmi les manifestations les plus marquantes

de ceci, on peut citer l'utilisation de l'espace des fonctions à variation bornée, de la variation totale comme critère de régularité des images, et de la décomposition des images en ensembles de niveau, modèles que nous avons déjà évoqués plus haut et que nous développerons plus en détail par la suite. Même la fameuse théorie des ondelettes, pourtant linéaire, a été récemment déclinée dans des variantes géométriques par plusieurs groupes de recherche, sous la forme de *ridgelets* [10, 28, 26], *curvelets* [11], *bandelettes* [63], etc.

2.2 Le modèle de Shannon

2.2.1 Du bon échantillonnage des images

La théorie de l'échantillonnage modélise une image (ou, plus généralement, un signal de dimension n) par une fonction $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ à énergie finie ($u \in L^2$) et à bande limitée, c'est-à-dire dont le spectre est à support compact : si \hat{u} est la transformée de Fourier de u , on impose que

$$\text{supp}(\hat{u}) \subset \left[-\frac{\pi}{\delta}, \frac{\pi}{\delta}\right]^2$$

pour un certain réel $\delta > 0$. Ce modèle est justifié par le théorème de Shannon [90], qui montre l'existence d'une correspondance *exacte* entre l'image physique u , définie sur un domaine continu (\mathbb{R}^2), et sa contrepartie discrète, formée des échantillons $u(k\delta, l\delta)$ pour $(k, l) \in \mathbb{Z}^2$.

Théorème 1 (Shannon) *Soit $u \in L^2(\mathbb{R}^2)$ et $\delta > 0$. Si $\text{supp}(\hat{u}) \subset [-\frac{\pi}{\delta}, \frac{\pi}{\delta}]^2$, alors*

$$\forall (x, y) \in \mathbb{R}^2, \quad u(x, y) = \sum_{(k, l) \in \mathbb{Z}^2} u(k\delta, l\delta) \text{sinc}\left(\frac{x}{\delta} - k\right) \text{sinc}\left(\frac{y}{\delta} - l\right), \quad (1)$$

où sinc est l'extension continue de la fonction $t \mapsto \frac{\sin(\pi t)}{\pi t}$.

Ce théorème justifie l'utilisation des images à énergie finie et à bande limitée par le fait qu'elles sont décrites exactement par des échantillons discrets. Ceci suggère la construction d'un modèle numérique exact, qui justifie *a posteriori* le modèle mathématique initial. Cependant, l'équation (1) ne suffit pas à définir un modèle numérique, car cette formule d'interpolation met en jeu un nombre infini d'échantillons. Il faut donc en pratique adapter cette relation, et renoncer à la fidélité exacte du modèle numérique (de toute façon compromise par la nécessaire quantification des échantillons). Nous discuterons ce point dans la section 3.

Indépendamment de ce lien avec l'échantillonnage, le modèle de Shannon est remarquable par la nature profondément physique des contraintes qu'il impose. Même dans un univers peu régulier où l'on autorise des sources de lumière ponctuelles (Dirac), l'image formée par un système optique classique est toujours, du fait de la diffraction, à bande limitée. Dans la modélisation physique du processus de formation et d'acquisition de l'image, bien décrit par des opérations linéaires (convolutions, échantillonnage), le modèle de Shannon (ou, plus exactement, la représentation de Fourier munie de la contrainte de Nyquist) est précieux pour mettre en relation le système optique (diffraction, flou de mise au point), la géométrie des capteurs (taille et forme des photosites) et la grille d'échantillonnage formée par ces capteurs. Il est tout aussi irremplaçable pour appréhender simplement un phénomène complexe comme l'*aliasing*.

Néanmoins, le modèle de Shannon n'est pas aussi pertinent pour la représentation des images que pour celle des signaux monodimensionnels. Alors que le son, par exemple, est très bien adapté à une décomposition additive (un son est une superposition d'ondes élémentaires), l'image s'y prête moins car elle résulte essentiellement d'un processus non linéaire : l'occlusion. Les images que nous percevons du monde extérieur sont composées d'objets au sens large, qui, parce qu'ils ne sont pas transparents en général, se masquent les uns les autres. Ce masquage produit de nombreuses discontinuités dans les images, et ces contours (*edge* en anglais), en délimitant naturellement les frontières des objets, jouent un rôle capital dans le processus de vision. Le modèle de Shannon, qui représente les images par des fonctions analytiques (car à bande limitée), n'offre pas la possibilité de modéliser ces discontinuités, ce qui limite son champ d'application.

Soyons plus précis, et considérons une image physique $u \in L^2(\mathbb{R}^2)$, que nous souhaitons échantillonner sur une grille de pas δ . Pour que cet échantillonnage soit licite, on doit théoriquement avoir

$$\text{supp}(\hat{u}) \subset R = \left[-\frac{\pi}{\delta}, \frac{\pi}{\delta}\right]^2,$$

ce qui suggère de filtrer u en $u' = \varphi \star u$ avec $\hat{\varphi} = 1_R$ avant de procéder à l'échantillonnage de u' . Indépendamment du fait que ce filtre passe-bas idéal est difficilement réalisable dans le cas d'un système physique, une telle opération n'est même pas souhaitable puisque φ serait alors une fonction fortement oscillante (un sinus cardinal) qui introduirait inévitablement des rebonds (oscillations) au niveau des *edges* de l'image. Ce phénomène, appelé *ringing*, illustre bien les limites du modèle de Shannon pour la représentation des images. En pratique, on cherche donc une pondération plus douce entre l'intérieur et l'extérieur de R , sous la forme d'une fonction $\hat{\varphi}$ qui passe rapidement (mais pas trop brusquement) de 1 à 0 autour de la frontière du carré R . La théorie de Shannon nous dit alors l'effet d'une telle pondération :

1. les fréquences hors de R affectées d'un poids non nul seront aliasées en des basses fréquences, ce qui produira des effets de crénelage des contours et l'apparition de moirés dans certaines textures de l'image ;
2. les fréquences dans R affectées d'un poids inférieur à un seront atténuées, ce qui produira un effet de flou ;
3. enfin, les oscillations de la fonction φ produiront des rebonds parallèles aux contours de l'image.

En pratique, il y a donc un compromis à réaliser entre ces trois effets indésirables, et la technologie grand public actuelle a plutôt tendance à éviter le flou, immédiatement perceptible, au détriment de l'*aliasing*, auquel de nombreuses personnes ne sont pas très sensibles. Que ce soit la conséquence d'un choix délibéré ou d'une méconnaissance de la théorie de Shannon, force est de constater qu'un nombre important de DVD actuels souffrent cruellement d'*aliasing* [P31]. Ces effets sont bien mieux contrôlés dans le domaine de l'imagerie satellite, où la maîtrise de l'échantillonnage des images est un enjeu majeur [82].

J'encadre actuellement Angélique Videau pour un stage de DEA sur un sujet assez proche ("Mesure de résolution et fusion d'images"), qui nous conduit à considérer de très près ces questions liés à l'échantillonnage des images.

2.2.2 Extrapolation de spectre

Je voudrais maintenant essayer d'illustrer ce qui précède avec le problème de l'extrapolation de spectre [83], auquel j'ai eu le loisir de m'intéresser dans le cadre d'une étude réalisée pour le CNES en 2000-2001. Considérons une image $u \in L^2$, dont le spectre est à support compact ($\text{supp}(\hat{u}) \in R$), mais dont la transformée de Fourier n'est connue que dans un sous-domaine D de R . Le problème de l'extrapolation de spectre consiste à essayer de "compléter" \hat{u} dans $R \setminus D$ de façon à produire une image u plausible. Un cas particulier assez représentatif est donné par le "zoom $\times 2$ ", où

$$D = \left[-\frac{\pi}{2\delta}, \frac{\pi}{2\delta}\right]^2 \subset \left[-\frac{\pi}{\delta}, \frac{\pi}{\delta}\right]^2 = R.$$

Si l'on synthétise l'image haute résolution (*i.e.* avec un pas d'échantillonnage de δ) en utilisant les seules valeurs du spectre connu, ce qui revient à supposer $\hat{u} = 0$ dans $R \setminus D$ (c'est la méthode bien connue du *zero-padding*), on assiste à la formation d'oscillations, phénomène de *ringing* que nous avons évoqué plus haut. Ces oscillations sont d'autant plus accentuées que la coupure fréquentielle ne se produit pas au bord de R , mais à l'intérieur de R , sur la frontière de D .

Pour essayer de supprimer ces oscillations et produire une image acceptable, deux voies s'ouvrent à nous : soit atténuer les fréquences connues au bord de D , de façon à "lisser" la discontinuité spectrale (par exemple en convolant u avec une fonction prolate [93]), soit tenter d'extrapoler le spectre sur $R \setminus D$ en maintenant les valeurs connues dans D . C'est bien sûr cette deuxième solution qui nous intéresse ici, et elle va nous donner l'occasion d'introduire le modèle variationnel.

2.3 Le modèle variationnel

Un problème récurrent en traitement d'images (et plus généralement en traitement du signal) est la déconvolution. Il est notamment au cœur du processus de restauration des images [25] : on cherche, à partir d'une image floue, à reconstruire une image nette, au moyen d'un modèle (plus ou moins précis) de la dégradation produite. Typiquement, on observe une image $u_0 : \Omega \rightarrow \mathbb{R}$ (image en niveaux de gris, définie sur un domaine $\Omega \subset \mathbb{R}^2$), et on cherche à retrouver l'image "originale" u dont u_0 est une version dégradée par

$$u_0 = Ku + n, \tag{2}$$

où $Ku = k \star u$ représente la convolution de u par un noyau k (que nous supposerons connu ici), et n un terme de bruit, modélisant par exemple l'erreur de mesure, la quantification des niveaux de gris, l'échantillonnage, etc. Même si K est inversible (*i.e.* si \hat{k} ne s'annule pas), ce qui est faux en général, la restauration de u par déconvolution brutale ($K^{-1}u_0$) est inefficace à cause du terme de bruit. En effet, dans le domaine de Fourier on a

$$\widehat{K^{-1}u_0} = \hat{u} + \frac{\hat{n}}{\hat{k}},$$

et le deuxième terme "explose" aux points où \hat{k} devient très petit, ce qui produit en général des oscillations démesurées qui noient le reste de l'image.

2.3.1 Régularisation

La solution classique pour résoudre ce type de problème mal posé consiste à utiliser une régularisation, comme l'a suggéré Tikhonov en 1963 [95]. Parmi toutes les fonctions u vérifiant

$$\|u_0 - Ku\|_2 \leq \sigma \quad (3)$$

(où σ est la variance *a priori* du bruit), on va sélectionner la plus plausible en un certain sens, en l'occurrence la plus régulière. Typiquement, u pourra être définie comme l'unique minimum de

$$\int_{\Omega} |Du|^2 \quad (4)$$

sous la contrainte (3) [53]. On peut montrer, en introduisant un multiplicateur de Lagrange, que ceci revient à minimiser la fonctionnelle

$$E(u) = \int_{\Omega} |Du|^2 + \lambda \int_{\Omega} (u_0 - Ku)^2$$

pour un certain λ . Dans ce cas précis, on peut expliciter la solution dans le domaine de Fourier, et retrouver un cas particulier du Filtre de Wiener (voir [13] par exemple), donné par

$$\hat{u}(\xi) = \frac{\hat{k}^*(\xi)}{|\hat{k}(\xi)|^2 + \frac{|\xi|^2}{\lambda}} \cdot \hat{u}_0(\xi),$$

où z^* désigne le complexe conjugué de z .

2.3.2 Variation totale

Le terme de régularité (4) ne modélise pas bien la plausibilité d'une image, dans le sens où il favorise fortement les images très lisses. En particulier, ce terme n'autorise pas les discontinuités, qui, comme nous l'avons signalé plus haut, sont pourtant inhérentes au processus même de formation des images (l'occlusion). En 1992, Rudin, Osher et Fatemi [84] ont proposé d'utiliser à la place la variation totale de l'image, soit

$$\int_{\Omega} |Du|. \quad (5)$$

Le bon cadre théorique est alors l'espace des fonctions à variation bornée, c'est-à-dire des fonctions $u \in L^1$ pour lesquelles $|Du|$ (au sens des distributions) est une mesure de masse totale finie [42, 101]. En l'absence de solution explicite, on minimise alors

$$E(u) = \int_{\Omega} |Du| + \lambda \int_{\Omega} (u_0 - Ku)^2$$

par un algorithme itératif de descente de gradient.

2.3.3 Variation totale pondérée

Revenons maintenant au problème de l'extrapolation de spectre. Si s représente les valeurs connues de la transformée de Fourier de u , nous devons sélectionner, parmi toutes les fonctions u telles que

$$\forall \xi \in D, \quad \hat{u}(\xi) = s(\xi), \quad (6)$$

la plus acceptable en un certain sens. Pour cela, nous pouvons utiliser une formulation variationnelle, en minimisant une certaine énergie $E(u)$ sous la contrainte (6). L'utilisation de la variation totale pour $E(u)$ semble *a priori* intéressante, d'autant que Guichard et Malgouyres [47] ont montré en 1998 qu'elle donnait de bons résultats lorsque l'on suppose que l'image basse résolution contient déjà (sous la forme d'*aliasing*) les hautes fréquences de l'image recherchée. Dans le cas de l'extrapolation de spectre, malheureusement, la variation totale ne donne pas des résultats très satisfaisants, et certaines oscillations résiduelles restent visuellement décelables. On peut en donner l'explication suivante : perceptuellement, nous sommes beaucoup plus sensibles à l'ajout d'oscillations dans les régions uniformes que dans les régions texturées (car ces dernières sont déjà par nature fortement oscillantes). Or, le nivellement de régions fortement texturées permet un gain de variation totale beaucoup plus important que l'élimination de petites oscillations dans des zones uniformes de l'image. Ceci explique pourquoi l'extrapolation de spectre par minimisation de la variation totale a plus tendance à atténuer le contraste des régions texturées qu'à supprimer efficacement le *ringing*.

Dans [P15], j'ai proposé une fonctionnelle permettant d'améliorer les résultats obtenus avec la variation totale, en pondérant celle-ci (voir [7] pour une justification théorique) par un terme dépendant de "l'uniformité locale" de l'image. J'ai ainsi utilisé une *variation totale pondérée*, définie par

$$E(u) = \int_{\Omega} w_{\lambda}(x) |Du(x)| dx, \quad (7)$$

où λ est un paramètre fixé (en gros le niveau de bruit moyen) et

$$w_{\lambda}(x) = \sup \left\{ r; \lambda \geq \inf_{|y-x| \leq r} \sup_{|z-y| \leq r} |u_0(z) - u_0(x)| \right\}. \quad (8)$$

Cette définition de w_{λ} correspond à une estimation non locale de l'inverse du module du gradient de u_0 . En effet, on vérifie sans difficulté que si $Du_0(x) \neq 0$, alors

$$w_{\lambda}(x) \sim \frac{\lambda}{|Du_0(x)|} \quad \text{quand } \lambda \rightarrow 0.$$

En un sens, (7) est donc une sorte de définition locale de la variation totale, où l'on paye en chaque point la norme du gradient relativement au gradient ambiant, ce qui, à en juger par les résultats obtenus, semble modéliser plus précisément notre perception du *ringing* que la variation totale classique.

2.4 Equations aux Dérivées Partielles et scale-space

Revenons à la restauration d'images, et intéressons-nous au cas particulier du débruitage, c'est-à-dire au cas où le flou est absent ($Ku = u$). Si l'on utilise la variation totale comme terme de régularité, on cherche donc à minimiser (5) sous la contrainte

$$\int_{\Omega} (u - u_0)^2 \leq \sigma^2. \quad (9)$$

Plutôt que de considérer la minimisation de (5) sous la contrainte (9), on peut, comme le suggèrent Bellettini, Caselles et Novaga [8], étudier l'équation d'Euler (descente de gradient)

associée à la minimisation (sans contrainte) de (5), qui conduit au flot $u(\mathbf{x}, t)$ défini par $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ et

$$\frac{\partial u}{\partial t} = \operatorname{div} \left(\frac{Du}{|Du|} \right) \quad (10)$$

(Du représente ici le gradient de u par rapport à \mathbf{x} seulement).

Qu'a-t-on gagné en remplaçant un problème variationnel par un problème d'évolution ? Du point de vue des paramètres, rien, puisqu'on a remplacé le paramètre σ (généralement inconnu) par une variable d'échelle³ t . Du point de vue de la solution, on passe d'une interprétation globale (minimisation d'énergie) à une interprétation locale (évolution infinitésimale), ce qui a autant d'avantages que d'inconvénients. Certes, on comprend souvent mieux la fonctionnelle que l'on cherche à minimiser en étudiant l'effet d'une descente de gradient infinitésimale qu'en analysant les propriétés des minima de cette fonctionnelle. Mais d'un autre côté, la formulation variationnelle offre généralement l'attrait d'une interprétation énergétique intuitive. En fait, le principal intérêt de la formulation par EDP est qu'elle ouvre de nouvelles perspectives de modélisation, en autorisant le cas d'évolutions qui ne seraient pas issues de problèmes variationnels. Le degré de généralité offert par les EDP est énorme, puisqu'il contient en germe l'expression de tout algorithme itératif local. C'est ainsi que l'on peut par exemple classifier les opérateurs locaux du traitement d'images (moyenne, médian, etc.) en termes d'EDP limites associées, en considérant le générateur infinitésimal donné par le comportement de l'opérateur au premier ordre lorsque le pas de la grille tend vers 0, comme l'ont montré Frédéric Guichard et Jean-Michel Morel [48]. Nous aurons l'occasion de détailler ce point dans la section 3.3.

Alors que les modèles variationnels sont généralement envisagés de façon assez intuitive, à partir de la définition d'une énergie combinant 2 termes (voire plus) entre lesquels on cherche un compromis, comment s'orienter dans la variété des EDP envisageables ? Une solution, utilisée historiquement à de nombreuses reprises (par exemple pour justifier les équations de la mécanique classique), consiste à déduire l'EDP cherchée de principes : c'est la démarche axiomatique.

2.4.1 La démarche axiomatique : exemple de la reconstruction du relief

L'idée de la démarche axiomatique est d'imposer un certain nombre de propriétés "naturelles" au modèle, jusqu'à obtenir l'unicité d'une EDP qui le décrit. Ceci permet souvent une justification esthétique et une classification efficace des modèles étudiés. Cette idée a été concrétisée pour l'analyse multiéchelle d'images en 1994 par Alvarez, Guichard, Lions et Morel [2], et a abouti à la découverte du *scale-space* affine, décrit par l'EDP⁴,

$$\frac{\partial u}{\partial t} = |Du| \operatorname{curv}(u)^{1/3}, \quad (11)$$

où

$$\operatorname{curv}(u) = \operatorname{div} \left(\frac{Du}{|Du|} \right)$$

³On peut aussi voir t comme une variable de temps, ce qui est bien compatible avec l'idée d'une équation d'évolution; cependant, nous garderons le terme d'*échelle* pour éviter toute confusion lorsque nous ferons évoluer des films (séquences d'images), qui possèdent déjà une variable temporelle.

⁴On peut donner un sens à (11) lorsque u est seulement continue et bornée, grâce à la théorie des *solutions de viscosité* [20, 21].

représente la courbure de la ligne de niveau de u passant en un point donné où Du est non nul. Nous aurons l'occasion de revenir sur cette équation un peu plus tard.

Pendant ma thèse, je me suis inspiré de cette idée pour modéliser le problème de la reconstruction du relief à partir d'une séquence d'images. L'objectif de ce travail était de pouvoir calculer le relief d'une scène tridimensionnelle à partir d'un film de survol. Si l'on suppose que la scène observée est lambertienne⁵, le relief cherché (la profondeur de chaque point) est inversement proportionnel à la vitesse apparente sur le film $u(x, y, \theta)$, qui, avec des choix d'axes appropriés, s'écrit

$$v = -\frac{u_\theta}{u_x} \quad (12)$$

(θ représente le temps, x la direction du mouvement de la caméra, et les indices doivent être compris comme des dérivées partielles). L'estimation de v à partir de (12) est plus complexe qu'il n'y paraît, d'une part à cause de la présence de bruit, qui rend instable le calcul des dérivées partielles, et d'autre part à cause de l'échantillonnage en temps, généralement insuffisant, qui rend illusoire l'estimation de u_θ par simples différences finies. Il y a donc ici un écart important entre la modélisation mathématique directe du problème, dans laquelle u_θ a un sens évident, et sa réalisation numérique, beaucoup plus délicate. Pour calculer v , une solution est de filtrer préalablement u , ce qui est particulièrement bienvenu ici à cause de la forte redondance de la donnée initiale (les images du film, prises deux à deux, sont autant de paires stéréo qui permettent en théorie de retrouver le relief). On se convainc ainsi facilement de la nécessité d'une représentation multiéchelle, selon le principe du *scale-space* introduit par Witkin [99]. Cette représentation consiste en l'addition d'une variable d'échelle $t \geq 0$, de sorte que le film "à l'échelle t " est décrit par la fonction $u(x, y, \theta, t)$.

Les principes qui définissent le *scale-space* recherché est une véritable étape fondamentale en physique, d'une part, puisqu'il faut rejeter l'utilisation d'une modélisation basée sur des critères d'invariance (au sens large) ; modélisation mathématique afin de pouvoir restreindre les solutions à des objets mathématiques bien définis et étudiés. Idéalement, les axiomes retenus doivent définir un modèle unique, non trivial, ce qui en pratique nécessite certains principes. Dans le cadre de la reconstruction du relief, j'ai abouti à la construc-

tion d'une unique analyse multiéchelle locale, causale, invariante par les transformations [v-Compatibility], [Galilean Invariance], et [Zoom Invariance]. Cette modélisation est réalisée près de la variable t , par l'EDP

$$\frac{\partial u}{\partial t} = u_{\theta\theta} - 2\frac{u_\theta}{u_x}u_{\theta x} + \left(\frac{u_\theta}{u_x}\right)^2 u_{xx}, \quad (13)$$

soumise à la condition initiale $u(x, y, \theta, t) = u_0(x, y, \theta)$.

⁵Une surface est lambertienne si elle diffuse la même intensité lumineuse dans toutes les directions. Elle est au contraire spéculaire si elle a tendance à se comporter comme un miroir.

Ces axiomes, dont la formulation détaillée pourra être trouvée dans [P18], ne sont pas tous de même nature. Les trois derniers axiomes relèvent spécifiquement d'une modélisation du problème de la reconstruction du relief. Ils traduisent essentiellement le fait que le calcul du relief (c'est-à-dire de v) ne doit pas changer lorsque l'on applique certaines transformations physiques au film. Inversement, les axiomes de localité et de causalité relèvent plutôt de la modélisation mathématique, au sens où leur rôle principal est de limiter les modèles possibles à des EDP paraboliques du second ordre. Certes, on peut justifier leur présence par des arguments de nature plus physique [2], mais y renoncer est très difficile, à moins de les remplacer par d'autres axiomes susceptibles de conduire à des modèles mathématiques intéressants.

En effet, il ne suffit pas de pouvoir formellement déduire une EDP des axiomes retenus ; encore faut-il pouvoir donner un statut mathématique précis à cette EDP, c'est-à-dire définir un concept de solution (en général des solutions faibles) qui permet de montrer l'existence, l'unicité et la stabilité de la solution au problème d'évolution avec condition initiale. Pour des EDP non-linéaires, la marge théorique est souvent étroite et l'on ne peut pas toujours atteindre le degré de généralité souhaité. Dans [P11], j'ai montré l'existence et l'unicité des solutions de (13) lorsque u est limite uniforme de fonctions régulières telles que $u_x \neq 0$. La très forte singularité de l'équation (13) en $u_x = 0$ est en quelque sorte une conséquence de l'irréversibilité du temps (une trajectoire n'est pas une courbe plane quelconque)⁶. Ce cadre théorique a son importance, car en n'autorisant pas les fonctions discontinues, on renonce à la possibilité de prendre en compte le phénomène d'occlusion, ce qui limite le champ d'application du modèle proposé.

Même si la plupart des EDP ne sont pas des équations d'Euler associées à une énergie, on peut tout de même parfois obtenir des interprétations globales. On pourrait penser aux états asymptotiques de l'équation d'évolution — obtenus en remplaçant $\frac{\partial u}{\partial t}$ par 0 dans (13) —, mais ils sont en général trop simples pour être réellement utiles. En fait, l'intérêt du *scale-space* n'est pas tant dans les états asymptotiques ($t = \infty$) que dans la succession des premiers états intermédiaires ($t \in [0, T]$). Les interprétations globales s'obtiennent plutôt en essayant de montrer la conservation ou la monotonie de certaines intégrales. Dans le cas du modèle (13), j'ai montré dans [P9] que l'énergie lumineuse totale $\int u^2$ est conservée, et que la quantité totale d'accélération $\int \Gamma^2$ (où $\Gamma = v_\theta + vv_x$ est l'accélération apparente sur le film) décroît avec le temps.

L'équation (13) a aussi la propriété intéressante de s'interpréter comme une diffusion anisotropique, à l'instar du mouvement par courbure moyenne par exemple. En effet, elle s'écrit encore

$$\frac{\partial u}{\partial t} = u_{\xi\xi} \quad \text{avec } \xi = \left(-\frac{u_\theta}{u_x}, 0, 1 \right) \quad \text{et } u_{\xi\xi} = [D^2u](\xi, \xi). \quad (14)$$

On peut même montrer que toutes les composantes du mouvement (v et ses dérivées totales successives) vérifient la même équation de diffusion [P9]. Ceci est assez naturel lorsque l'on réalise que (14) n'est autre que l'équation de la chaleur monodimensionnelle (non-linéaire) sur chaque trajectoire issue du mouvement apparent sur le film u .

La démarche axiomatique est extrêmement puissante, car elle donne immédiatement accès

⁶Evans a étudié en détail ce problème en s'intéressant à l'équation (13) en tant que réalisation scalaire de l'équation de la chaleur monodimensionnelle multivaluée [31].

à certaines propriétés du modèle (puisqu'elles sont à la base de sa construction), et induit de ce fait une classification naturelle des différents modèles. Dans [46], Frédéric Guichard a ainsi procédé à une classification des opérateurs de lissage locaux sur les images. Alors qu'un modèle variationnel se limite à la définition d'une énergie, la démarche axiomatique autorise de multiples caractérisations. J'ai ainsi montré que (13) peut se déduire d'une axiomatique plus physique, comme l'illustre le théorème suivant.

Nous dirons qu'un film $u(x, y, \theta)$ est *idéal* si l'on peut trouver des fonctions $C(\theta)$, $Z(X, Y)$ et $U(X, Y)$ telles qu'en tout point, on ait

$$u\left(\frac{X - C(\theta)}{Z(X, Y)}, \frac{Y}{Z(X, Y)}, \theta\right) = U(X, Y),$$

ce qui signifie que u n'est autre que l'observation de la scène $Z(X, Y), U(X, Y)$ (relief et luminosité lambertienne) par une caméra de trajectoire $C(\theta)$.

Théorème 3 ([P18, P19]) *L'équation (13) produit, à partir d'un film idéal u_0 , une séquence de films idéaux $u(t)$, dont l'interprétation du relief est identique à celle de u_0 , et pour lesquels le mouvement sous-jacent de la caméra évolue linéairement. C'est l'unique analyse multiéchelle de films v -compatible qui possède ces propriétés.*

Une des conséquences de l'axiome [v -Compatibility] est de rendre l'analyse du film indépendante du contraste : si g est une fonction continue de \mathbb{R} dans \mathbb{R} et $u'_0 = g \circ u_0$, alors les évolutions de u_0 et u'_0 par (13) vérifieront $u' = g \circ u$. Dans le contexte de l'analyse du mouvement, cette propriété découle naturellement de l'hypothèse lambertienne : les trajectoires physiques sont repérées par la conservation de l'intensité ($u = cte$), mais la valeur précise de u ne joue aucun rôle en elle-même.

2.4.2 Le principe morphologique

Le principe que nous venons d'évoquer, à savoir l'invariance par rapport à l'intensité absolue de l'éclairage, a aussi un sens pour l'analyse d'une seule image. Comme le souligne Frédéric Guichard dans sa thèse [46], *“le port de lunettes de soleil, le changement de contraste d'un téléviseur à un autre, ne semblent pas perturber notre compréhension des images perçues.”* Ce principe d'invariance par changement de contraste a été très largement utilisé en morphologie mathématique [71, 88, 69, 49], et est pour cette raison parfois appelé “invariance morphologique”. Ainsi, un opérateur T agissant sur les images est dit morphologique s'il vérifie $g \circ T = T \circ g$ pour tout homéomorphisme croissant de \mathbb{R} . Il est équivalent de dire que T est un *stack filter* [97], c'est-à-dire un opérateur qui commute avec les seuillages.

Précisons ce dernier point. Soit une image en niveaux de gris représentée par une fonction $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ semi-continue supérieurement. En seuillant u , on obtient ses ensembles de niveau, définis par

$$\chi_\lambda(u) = \{\mathbf{x} \in \mathbb{R}^2, u(\mathbf{x}) \geq \lambda\}.$$

Les bords topologiques de ces ensembles sont appelés lignes de niveau de u . Ces lignes de niveau sont globalement invariantes lorsque l'on change u en $g \circ u$ (g étant un homéomorphisme croissant

de \mathbb{R}), car on vérifie facilement que

$$\chi_{g(\lambda)}(g \circ u) = \chi_\lambda(u).$$

De plus, toute image peut être reconstruite à partir de ces ensembles de niveau, par

$$u(\mathbf{x}) = \sup\{\lambda, \mathbf{x} \in \chi_\lambda(u)\}. \quad (15)$$

Par conséquent, l'invariance morphologique force l'analyse d'une image à ne dépendre que des lignes (ou ensembles) de niveau de l'image : un *stack filter* peut toujours s'écrire

$$T(u) = \sup\{\lambda, \mathbf{x} \in T(1_\lambda(u))\},$$

où $1_\lambda(u)$ est la fonction indicatrice de $\chi_\lambda(u)$.

2.4.3 Mouvements par courbure

Comme on le voit, l'invariance morphologique est un axiome fort dans la construction d'un *scale-space* d'images, puisqu'il réduit théoriquement (et pratiquement, comme nous le verrons plus tard) le filtrage d'une image au filtrage d'un ensemble de courbes (les lignes de niveau de l'image). En ajoutant les axiomes de localité et de causalité, et un principe d'invariance euclidienne (invariance par translation et rotation), Christian Lopez et Jean-Michel Morel [65, 48] ont montré que chaque courbe doit évoluer selon un mouvement par courbure,

$$\frac{\partial \mathbf{C}}{\partial t} = F(\kappa, t) \mathbf{N}, \quad (16)$$

où κ et \mathbf{N} représentent respectivement la courbure et la normale de la courbe \mathbf{C} , et F est une fonction croissante. L'équivalent scalaire de (16) n'est autre que

$$\frac{\partial u}{\partial t} = |Du| F(\text{curv}(u)),$$

équation que l'on retrouve avec une axiomatique similaire dans le cas scalaire [2]. Parmi les évolutions de courbes (16), deux cas particuliers se distinguent : le mouvement par courbure moyenne⁷, défini par

$$\frac{\partial \mathbf{C}}{\partial t} = \kappa \mathbf{N}, \quad (17)$$

et le *scale-space* affine [2, 86], donné par

$$\frac{\partial \mathbf{C}}{\partial t} = \kappa^{1/3} \mathbf{N}, \quad (18)$$

avec la convention $\kappa^{1/3} = -|\kappa|^{1/3}$ lorsque $\kappa < 0$. Cette dernière équation n'est autre que la transposition géométrique (mouvement des lignes de niveau) de l'évolution (11).

L'étude théorique du mouvement par courbure moyenne dans la plan fut réalisée en 1986 par Gage et Hamilton [36] dans le cas convexe, puis généralisée en 1987 par Grayson [44]. Leurs travaux révèlent notamment que (17) est l'équation d'Euler (descente de gradient) associée à

⁷Cette appellation provient du fait que la généralisation de cette évolution en dimension $n \geq 3$ (évolution d'une hypersurface) se fait en remplaçant κ par la courbure moyenne de la surface.

la minimisation du périmètre de la courbe, et que le flot associé fait aussi décroître l'aire de la région bornée délimitée par la courbe et tendre le rapport isopérimétrique de la courbe vers celui d'un cercle : il y a disparition de la courbe en temps fini, et la "forme limite" avant disparition est un cercle.

L'étude théorique du cas affine fut réalisée en 1994 par Sapiro et Tannenbaum [86] pour les courbes convexes, puis généralisée en 1998 avec l'aide d'Angenent [4]. Les propriétés sont essentiellement les mêmes, mais transposées à la géométrie affine : l'équation (18) est ainsi l'équation d'Euler associée à la minimisation du périmètre affine, l'aire décroît, et il y a disparition de la courbe en temps fini, la "forme limite" étant une ellipse.

Le *scale-space* affine, découvert simultanément en 1994 par Alvarez, Guichard, Lions et Morel [2] dans sa version scalaire (11), et par Sapiro et Tannenbaum [86] dans sa version géométrique (18), est particulièrement intéressant en traitement d'images. Par rapport au mouvement par courbure moyenne, il présente en effet l'avantage de commuter avec les transformations spéciales affines du plan, soit $\mathbf{x} \mapsto A\mathbf{x}$, où A est une matrice 2×2 de déterminant 1. Etant donné le processus de formation des images (projection perspective de l'univers tridimensionnel), les transformations affines jouent un rôle particulier puisqu'elles correspondent à la projection d'une image plane située à l'infini, mais non nécessairement parallèle au plan focal de la caméra. Pouvoir analyser une image indépendamment de toute transformation affine est, comme nous allons le voir maintenant, un premier pas vers la reconnaissance de formes indépendante du point de vue.

2.4.4 Application à la comparaison morphologique et affine-invariante d'images

Par forme, nous entendons toute région plane dont la frontière est donnée par un nombre fini de courbes de Jordan. Etant donnée une base de formes F_1, \dots, F_n , on se pose le problème de reconnaître une nouvelle forme F , c'est-à-dire par exemple d'identifier l'élément F_i de la base le plus proche de F en un certain sens. On exige généralement que cette comparaison soit réalisée modulo un certain groupe \mathcal{G} de transformations, c'est-à-dire que F et $g.F$ soient considérées comme une même forme pour tout élément g de \mathcal{G} . Dans le cas de la reconnaissance affine-invariante, \mathcal{G} est engendré par les translations et les applications linéaires de déterminant non nul (ce qui inclut les rotations et les homothéties).

L'idée naturelle pour s'affranchir de \mathcal{G} est de normaliser toute forme F en une forme particulière $n(F)$ telle que

$$n(F) = n(F') \quad \Leftrightarrow \quad \exists g \in \mathcal{G}, F' = g.F$$

Dans le cas simplissime où \mathcal{G} est le groupe des translations du plan, on peut par exemple construire $n(F)$ en translatant F de façon à ce que son barycentre coïncide avec un point fixe. Dans le cas du groupe affine évoqué plus haut, une normalisation similaire, basée sur les moments, est possible, comme l'a montré Thierry Cohignac dans sa thèse [19] en 1994.

Si l'on accepte le principe morphologique, la comparaison de deux images doit reposer principalement sur la comparaison de leurs lignes de niveau. Ceci est cohérent avec le modèle physique de production des images, puisque, comme nous l'avons déjà remarqué, les occlusions entre ob-

jets provoquent des discontinuités le long de certains contours de ces objets, et ces discontinuités sont (au moins localement) des lignes de niveau. Pascal Monasse a ainsi pu montrer dans sa thèse que l'on pouvait recaler (*i.e.* mettre en correspondance) deux images proches en comparant tous leurs ensembles de niveau [76]. Mais si l'on veut autoriser de fortes variations entre les deux images, ce n'est plus les ensembles de niveau qu'il faut comparer mais *tous les morceaux de lignes de niveau*, puisque, par exemple, un objet peut être partiellement masqué dans l'une des images mais pas dans l'autre.

C'est à ce stade que la nécessité du *scale-space* se fait sentir. En effet la normalisation de chaque morceau de ligne de niveau requiert l'estimation de certaines dérivées de la courbe (afin d'identifier ses tangentes, ses points d'inflexion, etc.), et, comme dans le cas de l'analyse de films — estimation de la vitesse (12) —, un filtrage préalable est nécessaire. En calculant le *scale-space* affine de chaque ligne de niveau, on garde ainsi la possibilité d'effectuer la comparaison des deux images de façon affine-invariante, c'est-à-dire dans une large mesure indépendamment du point de vue (non seulement les applications affines sont, comme nous l'avons vu, des cas limites de transformations projectives, mais elles suffisent aussi à approximer localement au premier ordre toute déformation régulière).

En 2000, José-Luis Lisani, Pascal Monasse, Jean-Michel Morel et moi-même avons ainsi montré dans [P13] que la comparaison affine-invariante de tous les morceaux de lignes de niveau de deux images était possible, tant sur le plan mathématique que technologique (temps de calcul). Notre méthode tient en trois étapes :

1. Le calcul de l'arbre des lignes de niveau de chaque image, au moyen de la *Fast Level-Set Transform* développée par Frédéric Guichard et Pascal Monasse [75] ;
2. Le calcul rapide du *scale-space* affine de toutes ces lignes de niveau, grâce à un algorithme géométrique très simple [P12] sur lequel je reviendrai dans la section 3 ;
3. Le codage (normalisation) de tous ces morceaux de lignes de niveau et leur comparaison au moyen d'un dictionnaire.

Le transfert industriel de ces algorithmes fait actuellement l'objet d'une collaboration entre le CMLA et la société Poseidon Technologies, sous la forme d'un projet baptisé ISII [55].

3 Modèles numériques

In fine, c'est toujours l'image numérique qui est traitée ou analysée par des programmes informatiques. L'abstraction mathématique qui, en s'affranchissant de cette réalité discrète, permet une meilleure compréhension des algorithmes, a donc un prix : tous les modèles mathématiques que nous venons de voir doivent être finalement transposés numériquement pour devenir effectifs. Cette tâche est parfois délicate, car il peut être très difficile de rester fidèle au modèle mathématique en gardant une représentation réaliste de l'image physique.

3.1 Le modèle de Shannon

Comme nous l'avons vu dans la section 2.2, le modèle de Shannon est très utile pour analyser le processus physique d'acquisition des images. Examinons maintenant son intérêt sur le plan numérique. La formule d'interpolation (1) n'est pas applicable telle quelle car elle met en jeu un nombre infini d'échantillons. On la contourne en pratique en symétrisant puis en périodisant l'image. Si u est une image de taille $M \times N$ que l'on suppose périodique, l'interpolation (1) devient (en prenant $\delta = 1$ pour simplifier)

$$u(x, y) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} u(k, l) \operatorname{sincd}_M(x - k) \operatorname{sincd}_N(y - l), \quad (19)$$

où le sinus cardinal discret d'ordre N vaut

$$\forall t \in \mathbb{R} \setminus N\mathbb{Z}, \quad \operatorname{sincd}_N(t) = \begin{cases} \frac{\sin \pi t}{N \sin \frac{\pi}{N} t} & \text{si } N \text{ est impair,} \\ \frac{\sin \pi t}{N \tan \frac{\pi}{N} t} & \text{si } N \text{ est pair,} \end{cases}$$

et $\operatorname{sincd}_N(kN) = 1$ pour $k \in \mathbb{Z}$. Ceci est une autre formulation de la représentation de Fourier discrète donnée par⁸

$$u(x, y) = \frac{1}{MN} \sum_{p=-\lfloor \frac{M}{2} \rfloor}^{\lfloor \frac{M-1}{2} \rfloor} \sum_{q=-\lfloor \frac{N}{2} \rfloor}^{\lfloor \frac{N-1}{2} \rfloor} e^{2i\pi(\frac{px}{M} + \frac{qy}{N})} \tilde{u}(p, q), \quad (20)$$

où $\tilde{u}(p, q)$ (transformée de Fourier discrète de u) est définie par

$$\tilde{u}(p, q) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{-2i\pi(\frac{pk}{M} + \frac{ql}{N})} u(k, l).$$

A noter que cette représentation n'est cohérente⁹ que si u vérifie la condition de Shannon $\tilde{u}(p, q) = 0$ pour tous couples d'entiers (p, q) tels que $p = M/2$ ou $q = N/2$ (il n'y a donc pas de contrainte si M et N sont impairs). Pour des images naturelles, cette condition n'a pas d'impact réel : on peut forcer cette relation sans affecter visuellement l'image.

L'interpolation d'une image numérique par (19) est-elle pertinente en général ? On pourrait être tenté de répondre "oui, si la condition de Shannon a été respectée lors du processus de numérisation". Mais indépendamment de cette information (qui n'est pas nécessairement connue), on peut avancer que tout dépend du comportement de $|\tilde{u}(\xi)|$ lorsque ξ tend vers le bord du domaine fréquentiel discret

$$F = \left[-\frac{N}{2}, \frac{N}{2} \right]^2 \cap \mathbb{Z}^2$$

(on suppose pour simplifier que u est une image de taille $N \times N$ avec N pair). L'interpolation de u à un domaine continu (qui conduit à poser $\tilde{u}(\xi) = 0$ pour toute fréquence $\xi \notin F$) sera

⁸ $[x]$ représente ici la partie entière de x .

⁹au sens où (20) définit une fonction à valeurs réelles.

acceptable si $\tilde{u}(\xi) \simeq 0$ au bord de F . Si cette condition n'est pas remplie, la discontinuité au bord de F produira des rebonds (*ringing*) dans les zones de discontinuité de l'image. Pour éviter ce phénomène, on peut soit atténuer \tilde{u} au bord de F , soit réaliser une extrapolation de spectre, problème que nous avons déjà évoqué dans la section 2.2.

En un sens, donc, l'interpolation de Shannon n'est intéressante que pour des images raisonnablement floues. En l'appliquant à des images contenant des hautes fréquences trop marquées (et donc vraisemblablement de l'*aliasing*), on s'expose à l'apparition de rebonds gênants. Le cas le plus extrême est probablement celui des images binaires, pour lesquelles l'interpolation de Shannon produit des rebonds inacceptables. Ceci explique le nombre important de modèles numériques plus locaux qui ont été proposés pour approcher le modèle de Shannon. Parmi eux, on peut citer l'interpolation par splines [96], et bien sûr la représentation en ondelettes, qui permet de limiter les effets de bords [17].

Si le modèle de Shannon propose une interprétation parfois discutable des images numériques, il a toutefois un avantage de taille : il permet d'appliquer des transformations euclidiennes à une image (translation non entière, rotation¹⁰, homothétie de facteur $\lambda \geq 1$) sans perte d'information, contrairement aux méthodes d'interpolation plus classiques. La translation de vecteur (α, β) , par exemple, se traduit dans le domaine de Fourier par l'opération

$$\tilde{u}'(p, q) = \tilde{u}(p, q) \cdot e^{-2i\pi\left(\frac{\alpha p}{M} + \frac{\beta q}{N}\right)}, \quad (21)$$

opération parfaitement réversible.

Ce procédé trouve une application inattendue à la déquantification d'images. Imaginons une image discrète $u(k, l)$ fortement quantifiée, au sens où u est à valeurs dans un ensemble de cardinal faible. Même si l'œil humain peut s'accommoder parfaitement d'une image en 16 ou 32 niveaux de gris, une telle quantification pose des problèmes numériques, notamment lorsque l'on essaie de calculer des quantités géométriques locales. La direction du gradient, estimée par différences finies à partir de 4 pixels, est alors fortement biaisée (des orientations particulières sont privilégiées), et ceci d'autant plus que les gradients faibles (pour lesquels l'estimation de la direction est particulièrement sensible) sont majoritaires dans une image naturelle. Agnès Desolneux, Saïd Ladjal, Jean-Michel Morel et moi-même avons montré dans [P7] que l'on pouvait déquantifier une image en lui appliquant une translation de vecteur $(\frac{1}{2}, \frac{1}{2})$ selon (21). Cette opération a, en plus d'être inversible, l'avantage de transformer le bruit de quantification en un bruit quasi-gaussien et d'uniformiser ainsi la distribution des directions du gradient.

3.2 Le pixel discret

A l'opposé du modèle de Shannon, qui propose une interpolation globale de l'image à partir des échantillons mesurés, se trouve l'interpolation au plus proche voisin. Le passage du domaine discret au domaine continu se fait alors en supposant l'image constante sur chaque pixel discret, compris comme un carré de côté 1. Ce modèle a l'avantage d'être très simple, et de décrire

¹⁰La très surprenante (et trop peu connue) méthode de Yaroslavskly [100] permet d'appliquer une rotation à une image dans le domaine de Fourier. Elle repose sur une factorisation astucieuse de la rotation en trois glissements, qui se réalisent comme des translations selon chaque ligne ou chaque colonne de l'image.

exactement la représentation de l'image numérique sur un écran (spécialement sur un écran LCD, constitué de pixels carrés jointifs). En revanche, il n'offre pas une très grande souplesse mathématique, et ne facilite pas les définitions géométriques.

Les lignes de niveau, par exemple, sont pour ce modèle constituées uniquement de segments horizontaux et verticaux, et la mise en défaut du théorème de Jordan (toute courbe fermée simple partage le plan en deux composantes connexes) oblige à introduire une dissymétrie entre les ensembles de niveau supérieurs et inférieurs, en distinguant deux notions de connexité (la 4-connexité et la 8-connexité). Le seul avantage que l'on pourrait trouver à ce modèle est de respecter à la lettre l'invariance morphologique, au sens où les lignes de niveau ainsi calculées sont exactement les mêmes pour l'image u et l'image $g(u)$ (avec g strictement croissante), ce qui n'est plus tout à fait le cas dès lors qu'on utilise une interpolation. Mais cette caractéristique ne suffit pas à compenser la pauvreté géométrique du modèle, et une simple interpolation bilinéaire améliore considérablement le réalisme des lignes de niveau.

3.3 Le pixel infinitésimal

Le modèle de Shannon, qui interprète une image numérique comme une fonction analytique, permet de calculer des dérivées à tout ordre. Cependant, même pour des dérivées du premier ordre (gradient), ce calcul a deux inconvénients : en plus d'être global (et donc coûteux en temps de calcul), il est très sensible au bruit et au *ringing*. Pour cette raison, on lui préfère souvent une estimation par différences finies, qui consiste à calculer les dérivées de l'image par des combinaisons de pixels voisins sous l'hypothèse que le pas de la grille (*i.e.* la taille du pixel) est infiniment petit. Cette construction, bien sûr très classique en analyse numérique, a des avantages indéniables : elle permet d'interpréter tous les opérateurs discrets locaux en termes de dérivées partielles. Les correspondances suivantes (l'équivalence est vraie à une constante près qui ne dépend que de r) illustrent l'intérêt de cette approche.

$$\begin{aligned} \max_{|y| \leq r} \{u(x+y)\} - u(x) &\longleftrightarrow |Du| \\ \text{moyenne}_{|y| \leq r} \{u(x+y)\} - u(x) &\longleftrightarrow \Delta u \\ \text{median}_{|y| \leq r} \{u(x+y)\} - u(x) &\longleftrightarrow |Du| \text{curv}(u) \\ \max_{|y| \leq r} \{u(x+y)\} + \min_{|y| \leq r} \{u(x+y)\} - 2u(x) &\longleftrightarrow D^2u(Du, Du) \end{aligned}$$

Sous réserve de passer correctement à la limite, on peut alors montrer que l'itération de ces opérateurs discrets s'interprète comme la résolution numérique d'équations d'évolution. Les opérateurs ci-dessus permettent ainsi de calculer les solutions de la dilatation, de l'équation de la chaleur, du mouvement par courbure moyenne (17) et, à condition de considérer l'état asymptotique, de calculer l'extension Lipschitz absolument minimisante (AMLE) d'une image [5, 6, 56, 12].

Cette approche des opérateurs discrets a cependant ses limites, puisque le pas de la grille (que l'on suppose infiniment petit) est en pratique imposé par l'image elle-même, contrairement à ce qui se passe dans les simulations numériques classiques. Si l'image présente de brusques

variations, le comportement asymptotique peut très bien ne pas être révélateur du comportement effectif de l'opérateur. Il faut donc toujours être prudent lors de l'utilisation de schémas numériques basés sur des différences finies, et tout particulièrement, comme chacun sait, pour des schémas explicites. Dans le cas d'une formulation variationnelle notamment, il est plus sage de dériver l'équation d'Euler de la fonctionnelle discrétisée que de discrétiser l'équation d'Euler de la fonctionnelle continue. Dans le premier cas, on minimise numériquement une énergie discrète, avec la possibilité de contrôler effectivement la décroissance de cette énergie, alors que dans le second, on simule numériquement une EDP sans réel contrôle de la convergence. En un sens, ceci traduit la supériorité de la méthode des éléments finis, par nature variationnelle, sur celle des différences finies, différentielles : *“numériquement, il est toujours plus sage de travailler sur une fonction que sur sa dérivée”*¹¹.

3.4 Discrétisation des EDP morphologiques

Pour transposer numériquement une EDP déduite d'une formulation axiomatique, il est naturel de chercher à reproduire numériquement les axiomes imposés au modèle continu. Le principe de comparaison, par exemple, est crucial d'un point de vue numérique, puisqu'il garantit la stabilité du schéma. Les invariances géométriques complexes (invariance par rotation, ou par transformation affine) sont en revanche impossibles à réaliser sur une image numérique, du fait de la non-invariance de la grille d'échantillonnage elle-même. Pour ce qui est de l'invariance morphologique, la situation est plus complexe : comme nous allons le voir, il est possible mais généralement peu souhaitable de conserver cet axiome lors de la transposition numérique du modèle.

Rappelons tout d'abord le théorème de représentation de Matheron [71, 48], qui en un sens est à la morphologie mathématique ce que le théorème de Riesz est à l'algèbre linéaire.

Théorème 4 (Matheron) *Pour tout opérateur T morphologique, monotone, et invariant par translation, il existe une famille d'ensembles \mathcal{B} telle que pour toute image u ,*

$$Tu(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in B} u(\mathbf{x} + \mathbf{y})$$

pour presque tout \mathbf{x} .

L'intérêt de ce théorème est qu'il décrit un opérateur morphologique de façon géométrique (par l'intermédiaire de la famille d'éléments structurants \mathcal{B}), ce qui donne un moyen commode de le transposer numériquement, en projetant simplement chaque élément de \mathcal{B} sur la grille discrète.

L'invariance morphologique, néanmoins, ne fait pas bon ménage avec le monde numérique, et ceci spécialement pour les équations d'évolution. La raison est simple : puisqu'un opérateur morphologique agit indépendamment sur les ensembles de niveau d'une image, on peut se contenter d'examiner son action sur une image binaire. Or, la vitesse à laquelle peut évoluer un front binaire est quantifiée, au sens où pour un front rectiligne par exemple, il n'y a pas d'intermédiaire entre une vitesse de 1 pixel et une vitesse nulle. En d'autres termes, on ne peut coder des vitesses

¹¹Ce principe est couramment attribué à Stanley Osher.

subpixelles qu'en renonçant à l'invariance morphologique. Illustrons ceci dans le cadre du mouvement par courbure,

$$\frac{\partial u}{\partial t} = |Du| \text{curv}(u), \quad (22)$$

qui n'est autre que la formulation scalaire de (17). Comme nous l'avons vu dans la section précédente, l'opérateur médian est consistant avec (22) et, avec une renormalisation adéquate, on peut montrer la convergence du médian itéré vers les solutions de l'EDP. Cette propriété de convergence n'est malheureusement qu'un théorème asymptotique, qui ne traduit pas très fidèlement l'action du médian itéré. Pour les raisons que nous avons évoquées plus haut, les vitesses faibles (autrement dit les faibles courbures) sont numériquement tronquées à 0, ce qui conduit à une résolution incorrecte de (22) lorsque le nombre d'itérations augmente. En particulier, l'opérateur médian possède de nombreux points fixes qui ne sont pas des états stationnaires de (22). En un sens, ceci témoigne encore de la défaillance du modèle du "pixel discret", seul modèle compatible avec l'invariance morphologique.

Paradoxalement, il faut donc renoncer à transcrire aussi fidèlement que possible les propriétés du modèle continu pour résoudre numériquement une EDP morphologique, tout au moins en ce qui concerne l'invariance morphologique elle-même. Pour produire des simulations numériques acceptables, il est nécessaire d'accepter un effet diffusif artificiel, soit en discrétisant "classiquement" l'opérateur différentiel, soit, dans une moindre mesure, en combinant linéairement plusieurs opérateurs morphologiques. C'est dans cette optique que, à l'instar de Francine Catté, Françoise Dibos et Georges Koepfler [14] pour le mouvement par courbure moyenne, j'ai proposé de résoudre numériquement (13) à l'aide d'un schéma réalisant la moyenne de deux filtres inf-sup. L'utilisation de filtres inf-sup pour (13) était d'autant plus nécessaire que l'échantillonnage des films, généralement insuffisant par rapport à la variable de temps (θ), rend déraisonnable l'utilisation de différences finies pour estimer les dérivées temporelles.

Théorème 5 ([P19]) *Etant donné un film $u : \mathbb{R}^3 \rightarrow \mathbb{R}$, on pose*

$$\begin{aligned} IS_h u(x_0, y_0, \theta_0) &= \inf_{v \in \mathbb{R}} \sup_{-h \leq \theta \leq h} u(x_0 + v\theta, y_0, \theta_0 + \theta), \\ SI_h u(x_0, y_0, \theta_0) &= \sup_{v \in \mathbb{R}} \inf_{-h \leq \theta \leq h} u(x_0 + v\theta, y_0, \theta_0 + \theta), \\ \text{et } T_h u &= \frac{1}{2} (IS_h u + SI_h u). \end{aligned}$$

Alors,

$$T_h u = \frac{h^2}{2} \left(u_{\theta\theta} - 2 \frac{u_\theta}{u_x} u_{\theta x} + \left(\frac{u_\theta}{u_x} \right)^2 u_{xx} \right) + O(h^3)$$

en tout point où u est C^3 et $u_x \neq 0$.

3.5 L'image géométrique

Tous les modèles numériques que nous avons évoqués jusqu'à présent sont des modèles scalaires, au sens où leur rôle est d'interpréter l'image numérique, définie sur une grille, en une fonction définie sur un domaine continu. Compte tenu de l'importance de la géométrie dans

le contenu des images et notre manière de les percevoir, il paraît naturel d'essayer de définir des modèles purement géométriques. Ce modèle est en germe dans le principe d'invariance morphologique, qui réduit une image à ses lignes de niveau. En représentant une image par l'ensemble de ses lignes de niveau, considérées comme des courbes (et codées numériquement comme telles), on dote l'analyse d'images d'un vrai modèle géométrique, capable de s'affranchir de toute grille.

En 1999, Georges Koepfler et moi-même avons proposé un modèle de ce type, et nous avons montré qu'il était numériquement réalisable [P12]. La mise en œuvre pratique de notre modèle nécessite trois étapes :

1. l'extraction de toutes les lignes de niveau de l'image, selon le modèle du pixel discret évoqué dans la section 3.2 ;
2. le lissage rapide de toutes ces courbes, grâce à une implémentation rapide du *scale-space* affine (18), que nous allons évoquer dans la section 3.5.1 ;
3. la reconstruction d'une image discrète à partir d'un ensemble de lignes de niveau, selon (15).

L'étape de lissage (étape 2) permet de s'affranchir de la pixellisation des lignes de niveau, et fait office d'interpolation morphologique de l'image. Malgré sa relative lourdeur, ce modèle a des avantages indéniables : totalement géométrique, il rend extrêmement simple l'application de transformations géométriques sur l'image, puisqu'il suffit de réaliser ces transformations entre l'étape 2 et l'étape 3.

Le principal défaut de ce modèle est probablement la perte de résolution due à l'utilisation de l'interpolation au plus proche voisin lors de l'extraction des lignes de niveau. Par la suite, il a été amélioré lorsque j'ai travaillé avec José-Luis Lisani, Pascal Monasse et Jean-Michel Morel sur la comparaison morphologique d'images [P13]. En utilisant une extraction des lignes de niveau par interpolation bilinéaire, au moyen de la *Fast Level Set Transform* (FLST) [75], nous avons pu obtenir une précision géométrique bien supérieure, au prix d'une légère relaxation de l'invariance morphologique (les lignes de niveau *bilinéaires* de u et de $g(u)$ ne sont pas exactement les mêmes en général).

3.5.1 Un algorithme rapide pour le *scale-space* affine

L'intérêt du modèle géométrique que nous venons d'évoquer est de fournir un cadre numérique idéal à la résolution des EDP morphologiques. Ceci est particulièrement vrai dans le cas du *scale-space* morphologique affine (11), dont nous avons déjà illustré l'intérêt dans la section 2.4.4. En effet, la résolution numérique de (11) est non seulement problématique à cause du conflit entre invariance morphologique et discrétisation de la grille, mais aussi à cause de l'invariance affine, peu compatible avec l'utilisation d'une grille carrée fixe. L'invariance affine de (11) a d'ailleurs quelque chose d'un peu miraculeux pour une EDP locale, étant donné qu'une application affine peut introduire des distorsions d'amplitude relative quelconque sur les deux axes.

A l'inverse, la représentation géométrique, qui transforme l'équation (11) en un problème d'évolution de courbes (équation 18), est très bien adaptée au respect des invariances géométriques. L'avantage que l'on pourrait voir à la représentation scalaire est qu'elle gère automatiquement les changements de topologie (c'est l'argument classique en faveur des *Level Set Methods* [78, 89]), mais dans le cas du *scale-space* affine cet avantage est nul puisqu'il a été démontré qu'aucun changement de topologie ne peut se produire [4] (une courbe de Jordan ne peut se briser en plusieurs morceaux : elle reste une courbe de Jordan avant de disparaître).

Pour autant, la résolution numérique de (18) n'est pas immédiate. Idéalement, on souhaite un schéma à la fois très rapide (pour pouvoir lisser les quelques milliers de lignes de niveau d'une image en un temps raisonnable), invariant affine (dans la perspective d'une analyse d'images indépendante du point de vue), et monotone (afin de préserver la structure d'inclusion des lignes de niveau et permettre une reconstruction de l'image après lissage). Je vais essayer d'expliquer brièvement comment je suis arrivé à la construction d'un tel schéma.

Tout commence avec le théorème suivant, que j'ai découvert pendant ma thèse.

Théorème 6 ([P3]) *Soit $s \mapsto \mathbf{C}(s)$ une courbe fermée convexe de classe C^2 , paramétrée par abscisse curviligne, et σ un réel positif. A tout point $\mathbf{C}(s)$, on associe le point $\mathbf{C}_\sigma(s)$, défini comme le milieu de la corde $(\mathbf{C}(s - \delta), \mathbf{C}(s + \delta))$, où $\delta > 0$ est déterminé de façon à ce que l'aire de la région délimitée par cette corde et l'arc $\mathbf{C}|_{[s-\delta, s+\delta]}$ soit égale à σ (cf. figure 1). Alors on a le développement asymptotique*

$$\mathbf{C}_\sigma(s) = \mathbf{C}(s) + \omega \sigma^{\frac{2}{3}} \cdot \kappa(s)^{\frac{1}{3}} \mathbf{N}(s) + o_{\sigma \rightarrow 0}(\sigma^{\frac{2}{3}}), \quad (23)$$

où $\mathbf{N}(s)$ est la normale intérieure à la courbe, et $\omega = \frac{1}{2} \left(\frac{3}{2}\right)^{\frac{2}{3}}$.

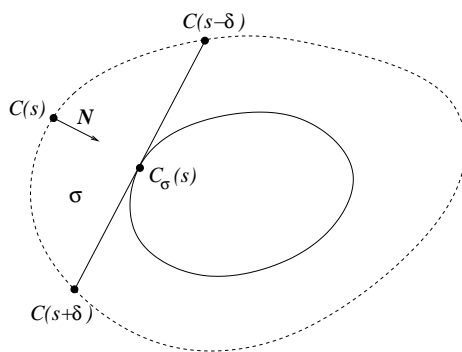


Figure 1: L'érosion affine d'une courbe convexe est donnée par l'enveloppe (ou encore la succession des milieux) des cordes délimitant une région d'aire fixée (σ) avec la courbe initiale.

Ce théorème est un résultat de consistance entre l'opérateur E_σ , qui fait passer de la courbe \mathbf{C} à la courbe \mathbf{C}_σ , et l'EDP (18). J'ai aussi montré que \mathbf{C}_σ est encore l'enveloppe de toutes les σ -cordes de \mathbf{C} (i.e. les cordes délimitant une aire σ avec la courbe). Hormis dans certains cas particuliers où cette enveloppe présente des points de rebroussement, \mathbf{C}_σ est encore la frontière de la forme obtenue en privant la forme initiale (l'intérieur de \mathbf{C}) de toutes les régions corde-arc d'aire σ . Cette caractérisation m'a conduit à donner le nom d'*érosion affine* à E_σ .

L'érosion affine est non seulement consistante avec (18), mais elle vérifie de plus les deux propriétés que nous avons souhaitées plus haut : l'invariance affine et la monotonie (deux courbes disjointes restent disjointes, ce qui garantit la stabilité du schéma puisque l'on peut toujours cerner une courbe très irrégulière entre deux courbes régulières). En généralisant la définition de l'érosion affine à des courbes non convexes, j'ai pu finalement montrer la convergence¹² d'un schéma basé sur l'itération alternée de E_σ et de son opérateur dual vers les solutions de (18). Il restait alors à choisir une discrétisation numérique de l'érosion affine. J'ai tout d'abord recherché le schéma le plus précis possible, afin de ne pas perdre lors de la discrétisation les précieuses propriétés d'invariance et de monotonie de l'érosion affine. J'ai ainsi défini un algorithme capable de calculer *exactement* l'érosion affine d'un polygone (constituée d'une concaténation de segments et d'arcs d'hyperboles) puis de discrétiser la courbe obtenue de façon invariante affine [P3].

Par la suite, je me suis peu à peu rendu compte que l'on pouvait fortement simplifier l'algorithme en raisonnant de manière purement discrète, et en se ramenant systématiquement à des courbes convexes. Outre la réduction importante du code, cette simplification avait l'avantage de réduire la complexité quadratique de l'algorithme initial en une complexité linéaire, assurant ainsi des gains énormes en temps de calcul. Cette accélération était nécessaire pour parvenir à traiter rapidement toutes les lignes de niveau d'une image, et montrer ainsi la faisabilité du modèle géométrique proposé dans [P12].

J'ai ainsi défini l'*érosion affine discrète* d'une courbe polygonale convexe en considérant uniquement les σ -cordes s'appuyant sur au moins un sommet du polygone initial, puis généralisé cette définition au cas non convexe en découpant la courbe initiale en "composantes convexes", en appliquant l'érosion discrète à chaque composante, puis en concaténant les nouvelles composantes ainsi produites. Moyennant la résolution de quelques difficultés (prise en compte de la précision numérique dans l'identification des composantes convexes, définition d'une aire d'érosion maximale afin de ne pas ralentir le mouvement des points d'inflexions), j'ai finalement obtenu un algorithme tout aussi précis mais bien plus rapide, et reposant uniquement sur des calculs d'aires et de milieux. Cet algorithme permet de lisser avec une bonne précision toutes les lignes de niveau d'une image (soit typiquement quelques milliers de courbes) en quelques secondes.

3.5.2 Généralisation à d'autres mouvements par courbure

Contrairement à l'algorithme que nous venons de décrire, le schéma numérique d'évolution de courbes décrit par Mikula et Sevcovic [74] est assez lent (car implicite), et sa construction non-géométrique ne peut garantir ni l'invariance affine, ni la monotonie. En revanche, il n'est pas limité au seul *scale-space* affine, et permet de simuler numériquement tout mouvement par courbure de type

$$\frac{\partial \mathbf{C}}{\partial t} = F(\kappa) \mathbf{N} \quad (24)$$

¹²J'ai montré cette convergence en replongeant l'érosion affine dans un cadre scalaire, car l'existence et l'unicité des solutions du *scale-space* affine géométrique [4] n'avaient pas encore été montrées à l'époque.

avec F régulière. Même si le *scale-space* affine est, du point de vue des invariances géométriques, le mouvement par courbure optimal, il était tentant d'essayer d'étendre le principe de l'érosion affine à des évolutions plus générales.

En 1999, Frédéric Cao et moi avons montré [P6] que l'on pouvait effectivement réaliser un schéma géométrique et monotone pour toute évolution de type (24) avec soit F croissante, nulle en 0 et Hölderienne d'exposant $1/3$ (premier cas), soit $F(t) = t^\gamma$ avec $\gamma > 1/3$ (second cas). Cette généralisation, d'un point de vue théorique, était suggérée dans les travaux d'Ishii [54].

Donnons un aperçu de cette généralisation dans le premier cas (le second cas est une adaptation du premier tirant parti de l'homogénéité de la fonction puissance). Posons tout d'abord $G(t) = \alpha F(t^3)$, où $\alpha > 0$ est choisi de telle sorte que G soit exactement 1-Lipschitz. Soit maintenant $(\mathbf{C}(s), \mathbf{C}(t))$ une σ -corde de \mathbf{C} , et $C_{s,t}$ la région corde-arc associée. On note h la hauteur de cette région corde arc (*i.e.* la distance maximale d'un point de $C_{s,t}$ à la corde), et l'on définit $C'_{s,t}$ comme l'ensemble des points de $C_{s,t}$ dont la distance à la corde est supérieure à

$$\delta = h - \omega \sigma^{2/3} G\left(\frac{h}{\omega \sigma^{2/3}}\right).$$

La F -érosion de \mathbf{C} est alors obtenue en privant la forme initiale de tous les ensembles $C'_{s,t}$ définis précédemment. Par rapport à l'érosion affine, ceci revient à translater parallèlement chaque σ -corde vers l'extérieur de la forme avant de considérer l'enveloppe de toutes ces nouvelles cordes (voir figure 2). Comme dans le cas affine, une telle enveloppe peut avoir des points de rebroussement, ce qui engendre des morceaux de courbes qui doivent être supprimés à chaque itération de l'opérateur.

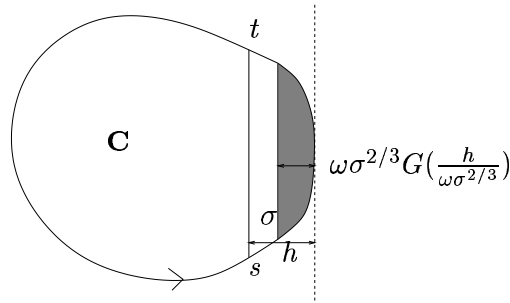


Figure 2: *Généralisation de l'érosion affine à des mouvements par courbure de type (24). Les σ -cordes sont translattées vers l'extérieur de la forme avant que leur enveloppe ne soit calculée.*

Si la consistance de la F -érosion avec (24) découle assez directement du théorème 6, en revanche la monotonie est moins intuitive et repose principalement sur le caractère 1-Lipschitz de G . L'algorithme d'érosion affine discrète peut être adapté à la F -érosion, mais sa rapidité est moindre en raison de calculs plus complexes et d'un raffinement nécessaire de l'échantillonnage de la courbe initiale et du pas de temps. Il nous a cependant permis de simuler des mouvements par courbure présentant une forte dynamique, comme par exemple le cas $F(t) = t^{10}$, très délicat numériquement (une variation de courbure d'un facteur 10 entraîne une variation de vitesse d'un facteur 10^{10}). Les grandes puissances ont l'avantage de lisser les détails de manière très séquentielle (les petits détails disparaissent sans que les détails plus gros soient affectés), ce qui

pourrait avoir des applications intéressantes en traitement d'images. L'influence de la fonction F est discutée plus en détail dans [P6].

4 Analyse d'images et modélisation statistique

Les modèles que nous avons évoqués jusqu'à présent se définissent dans une logique de *traitement* d'images (*Image Processing*), au sens où ils décrivent des opérations qui prennent une (ou plusieurs) image en entrée et produisent une (ou plusieurs) images en sortie. Même le *scale-space*, qui réalise une simplification progressive de l'image, donc en un sens une sélection de l'information pertinente qu'elle contient, ne vas pas jusqu'à décrire le moyen d'extraire cette information.

La vision artificielle, quant à elle, demande des modèles qui calculent des objets du monde réel : il s'agit bien d'*analyser* des images par rapport à l'univers physique qu'elles représentent, afin de guider des processus de décision, comme, par exemple, la navigation d'un robot. Cet objectif ambitieux peut sembler irréaliste si l'on pense que les images naturelles sont souvent très ambiguës au niveau local, et que ces ambiguïtés ne disparaissent que lors d'une interprétation globale. Le *scale-space*, tel que nous l'avons présenté, ne saurait donc être un modèle générique pour la vision artificielle : en effet, une structure globale résultant de coïncidences de ministructures locales (par exemple une ligne droite faite de pointillés) ne peut être détectée par une analyse multiéchelle classique, puisque les ministructures seront condamnées à disparaître dès les premières échelles de l'analyse, sans que la structure globale puisse émerger.

4.1 La révolution statistique en marche

L'inférence statistique n'est pas une discipline nouvelle, mais à l'instar de David Mumford [77], je pense qu'elle va connaître dans les années à venir un essor sans précédent (à vrai dire déjà commencé), tout particulièrement dans le domaine de l'analyse automatique d'images.

Tout d'abord, les images numériques sont partout, et il n'a jamais été aussi facile d'en produire. A elle seule, la toile est une gigantesque base de données d'images : le manque d'échantillons, problème souvent rencontré en statistiques, n'est donc pas de mise ici. Chaque image est d'ailleurs déjà en elle-même un objet de taille intéressante du point de vue statistique, et ceci d'autant plus que la résolution des photoscopes courants augmente régulièrement. Avec une image de plusieurs millions de pixel (typiquement $N = 3 \cdot 10^6$ de nos jours), le terme d'erreur relative ($1/\sqrt{N}$) dans le théorème de la limite centrale (pour donner un exemple) est tout à fait acceptable. L'étude des statistiques des images naturelles est, du fait de ses implications en restauration d'images notamment, un domaine actuellement en pleine effervescence, comme en témoignent les travaux récents de David Mumford et ses élèves [50, 51], de Donald Geman et Alexey Koloydenko [37], de Yann Gousseau [1, 43], et de Simoncelli et Olshausen [92].

La situation est d'autant plus favorable que, en même temps que cette formidable disponibilité d'images, on dispose maintenant d'une respectable capacité de calcul sur le moindre micro-ordinateur récent. Ce fort potentiel numérique nous offre en particulier la possibilité de réaliser

des modèles moins “opportunistes”, plus simples et plus objectifs. Un exemple assez frappant est donné par la méthode de synthèse de textures développée par Efros et Leung [29], qui surpasse les modèles paramétriques complexes à base d'ondelettes ou de champs de Markov. Le plus surprenant est la simplicité de leur algorithme : plutôt que de proposer un modèle paramétrique de la texture, d'identifier ces paramètres sur un échantillon pour pouvoir la reproduire à volonté ensuite, ils partent du principe que l'échantillon lui-même est le modèle de la texture. Ce type de modélisation, ou les données — associées à une modélisation stochastique très simple — remplacent les paramètres d'un modèle complexe, offre à mon avis une voie royale pour réaliser des modélisations statistiques complexes. Nous aurons l'occasion de revenir sur ce point dans la section 4.5.2.

Pour espérer tirer parti de cette conjonction favorable entre l'abondance des données et la capacité informatique de les traiter, il est plus que jamais nécessaire de s'appuyer sur des principes, afin de pouvoir maîtriser l'extrême diversité des algorithmes potentiellement utilisables. La définition de ces principes, est, là encore, une étape de modélisation, mais en un sens plus abstrait que ce que nous avons évoqué dans la section 1.2, puisqu'il ne s'agit plus de définir ce qu'est une image, mais de contraindre *les possibilités d'analyse* d'une image. On rejoint ici le cadre axiomatique utilisé pour la classification des analyses multiéchelles (cf. section 2.4.1).

Même si l'outil statistique semble être particulièrement adapté pour appréhender la complexité de l'image, il ne s'agit pas pour autant de rejeter en bloc les modèles déterministes, d'autant qu'ils peuvent souvent être réinterprétés dans un cadre probabiliste. Ceci concerne tout particulièrement les modèles variationnels, qui présentent un lien étroit avec le modèle bayésien (*maximum a posteriori*), que nous allons maintenant évoquer.

4.2 Le modèle bayésien

L'utilisation de la loi de Bayes (combinée à un modèle de champs de Markov) a été introduite en traitement d'images par Stuart et Donald Geman [38]. Elle se formalise de la manière suivante. Etant donné un modèle M des structures potentiellement présentes dans une image (ce modèle doit prendre en compte la variabilité en forme et en localisation des structures cherchées, ce qui se fait généralement par l'introduction de paramètres), on cherche, pour une image u donnée, l'instance $m \in M$ du modèle qui maximise la probabilité *a posteriori*

$$P(m|u) = \frac{P(u|m)P(m)}{P(u)}. \quad (25)$$

Ceci suppose une modélisation préalable conduisant au calcul des termes présents dans le membre de droite de (25).

En quoi consistent ces termes ? Le terme $P(u|m)$ est un terme d'instanciation, qui traduit la probabilité (vraisemblance) qu'un modèle m se réalise en une image u . Ce terme quantifie donc le passage du modèle mathématique idéal à l'image numérique, avec la part d'aléa qu'il implique. Le terme $P(m)$ est la probabilité *a priori* du modèle, c'est-à-dire la propension de m à se réaliser dans l'absolu, indépendamment de l'observation de u . Il s'agit ici de favoriser certains modèles soit parce qu'ils sont objectivement plus souvent rencontrés, soit par exemple

parce que l'on considère qu'ils donnent en général des interprétations plus réalistes dans le cadre d'une application donnée. Le principe du Rasoir d'Occam (l'explication la plus simple est la meilleure), qui peut se concrétiser mathématiquement au moyen du paradigme *Minimum Description Length* (MDL) [79, 80, 62], peut nous inciter à utiliser un argument de complexité pour modéliser $P(m)$. Le terme $P(u)$, quant à lui, est censé représenter la probabilité *a priori* d'observer une image donnée. Heureusement, ce terme joue seulement le rôle d'une constante de normalisation lorsque l'on cherche à maximiser $P(m|u)$ par rapport à m , de sorte que sa délicate estimation n'est jamais réalisée en pratique.

La méthode du maximum *a posteriori* que nous venons de décrire a des liens étroits avec nombre de modèles variationnels. En effet, la maximisation de $P(m|u)$ (ou, ce qui revient au même, de $P(m|u)P(u)$) se réécrit comme la minimisation de l'énergie

$$E(m) = -\log P(m) - \log P(u|m).$$

Comme dans bien des modèles variationnels, l'énergie est la combinaison d'un terme de régularité ($-\log P(m)$) et d'un terme de fidélité ($-\log P(u|m)$). Par exemple, le modèle de débruitage de Rudin-Osher évoqué précédemment, qui s'écrit

$$E(m) = \int_{\Omega} |Dm| + \lambda \int_{\Omega} (m - u)^2$$

(dans le cadre de la restauration d'images qui nous intéresse ici, le "modèle" m est l'image que l'on reconstruit), s'interprète dans le cadre bayésien comme un maximum *a posteriori* associé à la loi *a priori*

$$P(m) = \frac{1}{Z} \exp\left(-\int_{\Omega} |Dm|\right)$$

et à un modèle de bruit Gaussien

$$P(u|m) = \frac{1}{Z'} \exp\left(-\lambda \int_{\Omega} (m - u)^2\right),$$

Z et Z' étant ici des constantes de normalisation.

Cette possibilité de transposer des formulations variationnelles dans le cadre bayésien permet une interprétation différente des termes de l'énergie, et suggère aussi l'utilisation de lois *a priori* plus générales, où la probabilité $P(m)$ ne serait pas donnée par une expression analytique simple mais apprise sur un échantillon d'images par exemple. En revanche, la méthode du maximum *a posteriori* n'est pas réellement un outil de *détection* : elle décrit la meilleure interprétation (dans le cadre d'un modèle donné) de la donnée observée mais ne décide pas si cette interprétation est licite (*i.e.* très probable) ou pas. En effet, en calculant par commodité le maximum de $P(m|u)P(u)$ (et non de $P(m|u)$), on perd la possibilité d'interpréter le résultat en termes de vraisemblance du modèle associé à l'image. Le maximum *a posteriori* s'apparente donc en pratique à une formulation variationnelle, qui cherche non pas à affirmer la présence ou l'absence d'une structure dans une image, mais à trouver sa localisation la plus probable sous l'hypothèse qu'elle est présente.

Une autre objection classique à l'encontre du modèle bayésien (qui s'applique aussi aux modèles variationnels qui s'interprètent dans ce cadre) est le choix difficile de la loi *a priori*. Le

choix le plus simple, qui consisterait à utiliser une loi uniforme pour m , simplifie la méthode du maximum *a posteriori* en celle du maximum de vraisemblance (on maximise alors $P(u|m)$). Malheureusement, ce choix n'est pas adapté car dans le cas d'objets aussi complexes que les images, il existe généralement une infinité de modèles pour lesquels $P(u|m)$ est arbitrairement proche de 1. La définition de la loi *a priori* est donc essentielle, au même titre que le terme de régularisation proposé par Tikhonov dans le cadre variationnel. Elle reste malheureusement souvent difficile, parce que la méthode du maximum *a posteriori* est d'une trop grande généralité pour assurer à elle seule une maîtrise de la modélisation. La découverte et l'utilisation — maintenant très répandue — de la variation totale comme loi *a priori* pour les images (ou plutôt comme terme de régularité dans les formulations variationnelles) n'a vu le jour qu'après quinze ans de recherches en traitement d'images !

Ainsi, derrière la difficulté de sélection de la loi *a priori* dans le modèle de Bayes, se cache un véritable enjeu de modélisation. La solution n'est probablement pas à chercher dans une définition d'une loi *a priori* idéale, sorte de Graal mathématique inaccessible, mais dans la modélisation des principes qui régissent l'analyse des images. Pour faire face à l'inquiétante nécessité de modèles non locaux évoquée précédemment, il faut utiliser des principes qui permettent d'organiser l'analyse globale des images en une combinaison d'analyses élémentaires. Parmi les stratégies possibles, l'approche *coarse-to-fine* (littéralement, du plus grossier au plus fin), très séduisante, consiste à rejeter des hypothèses en analysant l'image à une échelle très grossière, puis à raffiner peu à peu la résolution de l'analyse en limitant la recherche aux hypothèses restantes. On minimise ainsi le coût algorithmique total, en n'explorant, à une échelle donnée, que des hypothèses qui n'ont pas été invalidées aux échelles supérieures. Cette approche est notamment étudiée intensivement par Donald Geman et ses collaborateurs [39, 3, 35].

4.3 Les conclusions de la théorie de la Gestalt

Le principe d'analyse *coarse-to-fine* permet certes d'organiser la *stratégie* de détection, mais ne spécifie pas la nature des objets à chercher, et une importante étape de modélisation reste nécessaire pour appréhender chaque application. Pour définir un système de vision complètement automatique, il faut donc définir précisément les primitives élémentaires sur lesquelles l'analyse de l'image doit reposer. Comme nous allons le voir, l'analogie avec la vision humaine est instructive à ce stade.

Le programme de recherche sur la perception visuelle lancé par Wertheimer [98] en 1921, et repris ensuite par Metzger [72], Kanizsa [59, 60] et leurs écoles, a abouti à ce qu'il est convenu d'appeler la théorie de la Gestalt. Cette théorie nous dit que la vision humaine s'articule autour d'un mécanisme de *groupement* récursif. Des points d'une image (ou des objets précédemment formés) sont groupés quand ils provoquent l'émergence d'une qualité géométrique (connexité, convexité, parallélisme, etc.) ou quand ils partagent eux-mêmes une qualité (couleur, forme, taille, convexité, etc.). Ces qualités sont en nombre restreint (une vingtaine) et ont été étudiées en détail par les gestaltistes. Pour prendre un exemple, un carré dessiné sur une feuille de papier est vu comme la collaboration de plusieurs qualités : connexité (un seul trait continu), fermeture, convexité, constance de largeur (épaisseur du trait), parallélisme (des côtés opposés),

orthogonalité, même longueur pour chaque côté, symétrie verticale et horizontale, etc.

Ceci nous apprend que même les tâches de vision que l'on pourrait qualifier d'évoluées passent par de multiples étapes intermédiaires qui construisent des représentations partielles de l'information perçue par la rétine. Un algorithme d'analyse d'images qui voudrait s'inspirer de la vision humaine doit donc s'attaquer tout d'abord à la détection des structures géométriques simples mises en évidence par la théorie de la Gestalt. C'est le but que j'ai poursuivi depuis 4 ans avec Agnès Desolneux et Jean-Michel Morel : en quelque sorte, nous avons essayé de formaliser mathématiquement certains mécanismes de groupement décrits par les gestalites (voir [P17] pour une discussion de nos travaux dans ce cadre).

Cette formalisation, que je développe dans la section 4.5, nécessitait la résolution du problème des paramètres, que je vais maintenant évoquer.

4.4 Le problème des paramètres

Même pour réaliser la détection automatique de structures simples, une difficulté survient immédiatement sous la forme d'une prolifération de paramètres, à savoir les paramètres libres (constantes du modèle), et les seuils de détection liés à la nécessaire multiplication des mesures. Dans le formalisme bayésien, la prise en compte de ces paramètres est intimement liée à la difficulté de choix de la loi *a priori*. En l'absence d'une méthode systématique pour fixer ces paramètres, on peut alors difficilement parler d'une méthode de détection automatique.

Une méthode à un paramètre est tout à fait acceptable, voire souhaitable, ne serait-ce que parce que l'on peut vouloir contrôler la fiabilité d'un système en fonction d'un contexte externe (certaines applications nécessitent une détection très sensible, pour laquelle on autorise un grand nombre de fausses alarmes mais le moins possible de détections manquées, alors que d'autres requièrent de prendre en compte seulement les détections les plus fiables). Une méthode à deux paramètres est déjà plus problématique, et son contrôle nécessite un apprentissage préalable (contrôler deux paramètres en temps réel, c'est typiquement conduire une voiture). A partir de trois paramètres, on doit parler de pilotage (on peut faire l'analogie avec la navigation aérienne) et non d'analyse automatique.

Un cas extrême est donné par les contours actifs (*snakes*), qui visent à délimiter les bords d'un objet dans une image ou dans un film. La formulation classique des *snakes* est variationnelle : on cherche un contour (disons une courbe plane pour se limiter au cas bidimensionnel) qui minimise une énergie comprenant essentiellement deux termes : un terme de fidélité, qui oblige la courbe à épouser les contours de l'image (en général, on incite la courbe à passer par des points de l'image à fort gradient), et un terme qui contrôle la régularité de la courbe (longueur, courbure, etc.). Les *snakes* sont un outil interactif très intéressant, qui permettent d'optimiser un contour initial approché. Mais, en tant que méthode de détection automatique, leur utilisation est plus discutable, tant la prolifération des paramètres est flagrante : coefficients intervenant dans la définition de l'énergie (un ou deux), degré de lissage appliqué à l'image avant le calcul du gradient (un), et surtout contour initial, qui représente à lui seul un grand nombre de paramètres : en effet, les énergies qui définissent les *snakes* ont toujours beaucoup de minima locaux, ce qui

rend leur minimisation itérative très sensible aux conditions initiales. Dans [P2], j'ai montré avec Agnès Desolneux et Jean-Michel Morel que pour réaliser une *détection* de contours, on avait généralement intérêt à employer une méthode statistique directe plutôt qu'une méthode de *snakes*.

Comment éliminer un paramètre indésirable dans un modèle de détection ? Si ce paramètre est une constante du modèle (et non un seuil de détection), une solution consiste à balayer les valeurs de ce paramètre, c'est-à-dire à lancer la procédure de détection pour toutes les valeurs possibles du paramètre (en fait, un échantillonnage des valeurs possibles). Il faut alors adapter la réponse du système en conséquence, car en augmentant le nombre de tests, on augmente la probabilité d'une fausse détection. En ce qui concerne les seuils de détection, la situation est plus délicate : si la détection repose sur au moins deux mesures, il faut plutôt parler de *domaine de détection* (dans l'espace des mesures), et si l'on ne souhaite pas choisir un domaine arbitraire (par exemple un hypercube, ce qui revient à choisir un seuil pour chaque mesure), une solution est de concevoir un modèle qui met en relation ces mesures. Dans ce qui suit, je vais présenter une approche systématique, qui utilise les deux solutions que nous venons d'évoquer pour produire des algorithmes à un seul paramètre (la fiabilité de la détection).

4.5 Une approche systématique : la détection a contrario

Le principe de groupement mis en évidence par les gestaltistes est de nature qualitative, et n'est donc pas utilisable tel quel pour construire un système automatique de vision. Comment le rendre quantitatif ? Une des possibilités est de suivre un principe simple dû semble-t-il à Helmholtz. Ce principe stipule que notre perception repose sur la détection de structures qui s'opposent au hasard : en l'absence d'information, notre système visuel suppose que les caractéristiques des objets perçus (position, orientation, couleur, etc.) sont réparties aléatoirement et relativement uniformément, et c'est la violation de cette hypothèse qui conduit au groupement des objets.

Ce principe suggère de réaliser une détection *a contrario*, c'est-à-dire de chercher des structures en tant que négation d'un modèle banal (répartition aléatoire et uniforme des qualités), qui décrit ce que ne sont *pas* ces structures. Ce point de vue est séduisant, car il est beaucoup plus facile (et certainement beaucoup moins arbitraire) de construire des modèles de ce type que des modèles très précis des objets que l'on désire détecter.

L'idée d'utiliser une détection *a contrario* avait été plusieurs fois proposée en analyse d'images, notamment par Lowe [66] et Stewart [94]. Le travail que j'ai réalisé avec Agnès Desolneux et Jean-Michel Morel depuis près de quatre ans a consisté notamment à donner un statut mathématique simple et surtout systématique à la détection *a contrario*, puis à l'appliquer à la détection d'alignements de directions [P4] ou de points [P8], de bords contrastés [P5], de modes d'histogrammes [P8] et d'amas de points [P8]. Comme nous allons le voir, la simplification que nous avons introduite en mesurant la fiabilité d'une détection par une *espérance de nombre de fausses alarmes* plutôt que par la *probabilité d'une fausse alarme* permet, moyennant un nécessaire échantillonnage des détections possibles, de déterminer explicitement un domaine de détection absolu combinant automatiquement de multiples mesures.

4.5.1 Formalisation du principe de Helmholtz

Les outils de détection *a contrario* que nous avons proposés et appliqués à l'analyse d'images reposent essentiellement sur deux idées : la détection par réfutation d'un modèle naïf, ce qui est à la base même du principe de Helmholtz, et l'emploi systématique d'espérances mathématiques (plutôt que de probabilités) qui permettent de s'affranchir du problème de la non-indépendance de détections multiples.

Notre définition des événements ε -significatifs dans une image peut se résumer à cette propriété : l'espérance du nombre d'événements significatifs dans une image aléatoire est inférieure à ε . Cette définition très simple a l'avantage de garantir une chose : dans une image aléatoire, en moyenne on observe moins de ε fausses détections. Plus formellement, ceci donne la

Définition 1 Soit E l'ensemble des événements potentiellement détectables. A tout élément $e \in E$ on associe son nombre de fausses alarmes, défini par

$$NFA(e) = |E| \cdot P(e) \quad (26)$$

et pour tout réel $\varepsilon > 0$, e est dit ε -détectable ssi

$$NFA(e) \leq \varepsilon. \quad (27)$$

Cette définition suppose d'avoir donné un sens préalable à $P(e)$, la probabilité qu'un événement donné se produise. Ceci se fait par le principe de Helmholtz, en supposant que l'image observée est le résultat d'un tirage aléatoire uniforme¹³.

Détection d'alignements. Prenons un exemple, en considérant une image numérique u de taille $N \times N$. On peut calculer une orientation en chaque point de l'image en appliquant une rotation de $+\pi/2$ à la direction du gradient (quand elle est définie). Cette orientation mesure la direction de la tangente à la ligne de niveau qui passe en chaque point, et donc coïncide aussi avec celle d'un bord éventuel. Soit maintenant S un segment de l'image défini par deux extrémités entières, soit, après approximation entière, par une suite de l pixels de l'image. Si les orientations de ces points coïncident avec l'orientation du segment lui-même, alors ce segment est probablement un "bord rectiligne" de l'image. Dans [P4], nous avons utilisé la définition 1 pour quantifier ceci, et déduire ainsi des seuils précis pour la détection d'alignements dans une image. L'événement à considérer est "parmi les l points de S , au moins k ont une direction qui coïncide avec celle de S à $\pm p\pi$ radians près", et le nombre de fausses alarmes associé est

$$NFA(k, l) = N^4 B(p, l, k), \quad (28)$$

où B désigne la queue de distribution de la loi binomiale, définie par

$$B(p, l, k) = \sum_{i=k}^l \binom{l}{i} p^i (1-p)^{l-i}. \quad (29)$$

¹³Cette notion de tirage "uniforme" est à prendre au sens large, et nous verrons qu'elle peut revêtir des formes très diverses.

L'intérêt de la relation $NFA(k, l) \leq \varepsilon$ est qu'elle combine les paramètres k , l , N , p et ε pour définir un seul critère de détection. En simplifiant, on pourrait dire que nous avons simplement rendu explicites les constantes de détection dans la transformée de Hough [67]. Le seuil de détection ε est un paramètre de fiabilité, qui peut typiquement être fixé à 1 ou à 10^{-10} selon les besoins. En utilisant des estimations statistiques plus ou moins subtiles (théorème de la limite centrale, grandes déviations), on peut en déduire des estimations du seuil k_0 (dépendant de l , N , p et ε) à partir duquel un segment est ε -significatif, ainsi que des propriétés qualitatives du modèle de détection donné par (28). Sans entrer dans une description trop précise (voir pour plus de détails [P4], [P25]), citons le résultat suivant.

Théorème 7 ([P4]) *Dans une image $N \times N$, un segment de longueur l contenant au moins k points alignés à la précision p ($p < \frac{1}{2}$) est ε -significatif dès lors que*

$$k \geq pl + \sqrt{l \cdot \frac{(1-2p)(\ln N^4 - \ln \varepsilon)}{\ln(1-p) - \ln p}}. \quad (30)$$

On remarque au passage la faible dépendance du seuil par rapport à ε .

Dans le cas que nous venons de voir, le modèle probabiliste “naïf” que nous avons utilisé est facile à construire d'après le principe de Helmholtz : on choisit naturellement une distribution uniforme et indépendante des directions mesurées en chaque point. Ce modèle est complètement faux (les directions ne sont *pas* indépendantes) pour la plupart des images, et c'est justement là le point essentiel, puisque nous avons défini la significativité des segments comme négation de ce modèle, c'est-à-dire *a contrario*. En fait, chaque segment ε -significatif est précisément une preuve que le modèle est faux, mais ce n'est bien sûr pas ce résultat en tant que tel qui nous intéresse ici mais les preuves que nous en construisons.

On pourrait objecter qu'observer un champ de direction (*i.e.* une fonction à valeurs dans S^1) est un cas particulièrement favorable à l'application du principe de Helmholtz, puisque le modèle uniforme est difficilement discutable sur S^1 (à moins de renoncer à l'invariance par rotation de l'analyse). Néanmoins, notre approche reste applicable dans des situations très différentes, comme nous allons le voir dans l'exemple suivant.

Modes d'histogrammes. Revenons à notre projet de formalisation des lois de la Gestalt, et tout particulièrement à celui de quantifier le groupement d'un nombre fini (N) d'objets en fonction d'une qualité scalaire (un nombre réel q , qui peut être une intensité lumineuse, une longueur, une aire, etc.). On commence par quantifier les valeurs de q , de façon à obtenir une qualité discrète \tilde{q} prenant L valeurs possibles ($0 \leq \tilde{q} \leq L-1$). Dans [P8], nous avons considéré deux cas distincts : le cas d'une qualité que l'on peut *a priori* supposer uniforme (par exemple une intensité, une orientation, une position sur une droite, etc.) et le cas d'une qualité que l'on peut *a priori* supposer décroissante (par exemple une longueur ou une aire).

Dans le premier cas, l'application du principe de Helmholtz est immédiate (on suppose que les qualités des objets sont indépendantes et uniformément distribuées dans $\{0..L-1\}$), et les événements à considérer sont du type “*au moins k objets (parmi N) ont une qualité $q \in \{a..b\}$ ”.*

On obtient ainsi la relation

$$NFA(a, b, k) = \frac{L(L+1)}{2} B\left(\frac{b-a+1}{L}, N, k\right). \quad (31)$$

Le cas d'une distribution décroissante est assez différent et mérite qu'on s'y attarde quelque peu. Supposer que la distribution des aires d'objets indépendants est *a priori* décroissante peut être justifié notamment par le fait que cette décroissance est effectivement observée dans les images naturelles [1, 43], où les aires des objets suivent en moyenne la seule loi invariante par changement d'échelle, où $P(\text{aire} = t)$ est proportionnel à $1/t^2$. Plutôt que de prendre cette loi particulière, somme toute assez restrictive, comme hypothèse de Helmholtz, nous avons préféré n'en retenir qu'un aspect qualitatif : la décroissance. Dans ce cadre, le nombre de fausses alarmes associé au même événement "*au moins k objets (parmi N) ont une qualité $q \in \{a..b\}$* " est

$$NFA(a, b, k) = \frac{L(L+1)}{2} \max_{n \in \mathcal{D}} B\left(\frac{1}{N} \sum_{i=a}^b n(i), N, k\right), \quad (32)$$

où \mathcal{D} est l'ensemble des fonctions $n : \{1..L-1\} \rightarrow \mathbb{N}$ décroissantes et de masse totale N . Ainsi, le modèle que l'on réfute pour caractériser les événements significatifs peut avoir des formes assez différentes selon le type de propriété recherché, et n'est pas limité à des distributions uniformes de variables indépendantes.

4.5.2 L'image est son propre modèle

Pour pouvoir effectuer la détection *a contrario* en toute généralité, il faut toutefois aller plus loin, car certaines situations se prêtent mal à des modèles prédéfinis. Un exemple est donné par la détection de contours significatifs, telle que nous l'avons formalisée dans [P5]. L'idée est de considérer toutes les lignes de niveau d'une image $u = u_0$ (associées à une certaine quantification des niveaux de gris) et de calculer pour chacune son contraste μ , défini comme le minimum de $|Du(\mathbf{x})|$ quand \mathbf{x} décrit la ligne de niveau. Pour déterminer le nombre de fausses alarmes associé à l'événement "*une ligne de niveau de longueur l a un contraste au moins égal à μ* ", il faut choisir quel modèle de $|Du|$ on va réfuter. Exiger l'indépendance des variables aléatoires $|Du(\mathbf{x})|$ paraît naturel (on remarque un bord parce que *tous* ses points sont contrastés), et à bien y réfléchir c'est même la seule propriété que l'on veut réellement mettre en défaut. Choisir une distribution prédéfinie pour $|Du|$, uniforme ou non, serait illogique car on risquerait alors de détecter des bords non comme coïncidence de points très contrastés mais simplement comme une conséquence de la différence entre la distribution de $|Du|$ observée sur l'image u_0 et celle imposée *a priori*. Ainsi, il est naturel de choisir comme modèle *a contrario* des variables aléatoires $|Du(\mathbf{x})|$ indépendantes, mais dont la distribution est donnée par une estimation empirique sur l'image, à savoir

$$P[|Du(\mathbf{x})| \geq \mu] = H_{u_0}(\mu), \quad (33)$$

où

$$H_{u_0}(\mu) = \frac{1}{|\Omega|} \left| \{ \mathbf{x} \in \Omega, |Du_0(\mathbf{x})| \geq \mu \} \right| \quad (34)$$

et Ω est le domaine de définition (fini) de u_0 . Le nombre de fausses alarmes associé à l'événement décrit précédemment est alors simplement

$$NFA(\mu, l) = N_l \cdot \left(H_{u_0}(\mu) \right)^l, \quad (35)$$

où N_l désigne le nombre de lignes de niveau de u_0 .

Cette possibilité d'utiliser la répartition empirique d'une variable comme distribution *a priori*, puis de réfuter simplement l'*indépendance* spatiale des échantillons, illustre l'étendue des applications possibles du principe de Helmholtz. On peut donc l'utiliser non seulement dans le cas de distributions naturellement uniformes (direction) ou décroissantes (aire), mais aussi dans le cas de distributions quelconques pour peu que nier l'*indépendance* suffise à exhiber des objets intéressants.

Utiliser l'image comme "son propre modèle", c'est en réalité se donner les moyens de l'analyser récursivement par une succession d'étapes simples. On pourra être "surpris" d'observer une image composée uniquement de carrés, mais on le sera encore davantage si cette image contient en plus un cercle. Ce principe d'une analyse hiérarchique et contextuelle (la fameuse distinction fond/forme) est au cœur même de la théorie de la Gestalt, et les lois de la Gestalt visuelle existent précisément parce que ce principe est inhérent à la vision humaine.

Le modèle précédent (équation 33) est déjà lié à cette récursivité, puisqu'il peut être vu comme la deuxième étape d'une analyse en deux temps, où l'on commencerait par "s'étonner" de la distribution empirique de $|Du_0|$ (en analysant H_{u_0}), puis du fait que les valeurs de $|Du_0|$ ne sont pas réparties uniformément dans le plan image. D'un point de vue statistique, ceci peut se traduire sous la forme d'un conditionnement,

$$P \left[\left\{ |Du(\mathbf{x})|, \mathbf{x} \in \Omega \right\} \right] = \sum_H P \left[\left\{ |Du(\mathbf{x})|, \mathbf{x} \in \Omega \right\} \mid H_u = H \right] \cdot P[H_u = H].$$

Le lien avec (33) est justifié par le fait que si l'on n'analyse simultanément que de faibles proportions de Ω , alors, compte tenu de la taille importante des images numériques, un modèle probabiliste basé sur la répartition uniforme de l'ensemble des valeurs de $|Du|$ imposées par la contrainte $H_u = H$ a les mêmes effets qu'un modèle où chaque $|Du(\mathbf{x})|$ est tiré indépendamment selon la loi dérivée de H .

4.5.3 Maximalité

La méthode de détection *a contrario* que nous avons introduite permet d'associer à tout événement e potentiellement détectable dans une image son nombre de fausses alarmes $NFA(e)$. Les événements e ε -significatifs sont alors ceux pour lesquels $NFA(e) \leq \varepsilon$. Lorsque l'on balaye systématiquement l'ensemble des événements possibles, on constate généralement un phénomène d'adhérence : du fait de la corrélation entre les mesures effectuées au voisinage d'un point de l'image, un événement ε -significatif a peu de chances d'être isolé : on doit s'attendre à ce que la déviation du hasard qu'il manifeste se répercute au moins partiellement sur les événements spatialement proches, et rendent certains d'entre eux ε -significatifs.

Prenons le cas de la détection de bords non fermés (*edges*), traité dans [P5]. Etant donné un morceau A (connexe) de ligne de niveau, on mesure son contraste $\mu = \min_{\mathbf{x} \in A} |Du(\mathbf{x})|$ comme dans la section précédente, et son nombre de fausses alarmes par (35), excepté que N_{ll} doit être remplacé par le nombre de *morceaux* de lignes de niveau dans l'image (en nombre fini après une quantification appropriée des points sur les lignes de niveau). Si A est ε' -significatif avec $\varepsilon' < \varepsilon$, alors certains sous-ensembles de A sont probablement ε -significatifs, car leur contraste est au moins égal à celui de A et leur longueur est, pour certains d'entre eux, à peine inférieure. De même, des sur-ensembles de A seront aussi probablement détectés. On voit ici à l'œuvre le phénomène d'adhérence que nous venons d'évoquer, et qui conduit à une multiplication indésirable des détections.

Puisque l'on obtient une multiplicité de “réponses” pour chaque événement très détectable, il est tentant de choisir la meilleure réponse, c'est-à-dire le meilleur représentant au sein d'un “amas” spatialement localisé d'événements ε -significatifs. Si l'on associe à chaque événement e son support S_e , *i.e.* l'ensemble des points de l'image impliqués dans la définition de e , on peut définir une notion de maximalité locale par rapport à l'inclusion des supports :

Définition 2 ([P4, P5, P25]) *Un événement e est dit maximal ssi*

1. $\forall e', \quad S_{e'} \subset S_e \Rightarrow NFA(e') \geq NFA(e)$,
2. $\forall e', \quad S_{e'} \supset S_e \Rightarrow NFA(e') > NFA(e)$,

Cette notion de maximalité se révèle essentielle, car elle permet de se débarrasser de bon nombre de détections indirectes. D'un point de vue théorique, elle pose un problème structurel intéressant : est-ce que les événements maximaux sont par nature disjoints ?

Dans le cas que nous venons de voir (*edges*), cette propriété est immédiate. Soient A et A' deux morceaux de lignes de niveau de contrastes respectif μ et μ' (on peut supposer $\mu \leq \mu'$ par symétrie) tels que $A \cap A' \neq \emptyset$. Alors $A \cup A'$ est au moins aussi détectable que A (car plus long et aussi contrasté), donc A ne peut être maximal.

Pour d'autres *gestalts*¹⁴, cette propriété est parfois plus délicate à démontrer. Dans le cas de modes d'histogramme, elle repose sur une propriété de convexité de la distance de Kullback [P25]. Dans le cas des alignements, le caractère disjoint des événements maximaux significatifs est une conjecture [P4, P27], qui peut se formuler de la manière suivante :

Conjecture 1 *Si pour $i = 1, 2, 3$, k_i et l_i sont des entiers positifs tels que $l_i \neq 0$ et $k_i \leq l_i$, alors*

$$\min \left(p, B(p, l_1, k_1), B(p, l_1 + l_2 + l_3, k_1 + k_2 + k_3) \right) < \max_{i=2,3} B(p, l_1 + l_i, k_1 + k_i). \quad (36)$$

Cette conjecture a été vérifiée numériquement pour toutes les valeurs atteignables en pratique, lorsque $p = 1/16$. Elle peut aussi être vue comme une conséquence de la conjecture suivante, asymptotiquement prouvée [P27] :

¹⁴Par commodité, nous utiliserons, improprement, le pluriel “gestalts” au lieu du pluriel germanique “Gestalten”.

Conjecture 2 Soit, pour $p \in]0, 1[$,

$$P(k, l) = \frac{\int_0^p x^{k-1} (1-x)^{l-k} dx}{\int_0^1 x^{k-1} (1-x)^{l-k} dx}. \quad (37)$$

Pour tous $(k, l) \in \mathbb{R}^+ \times \mathbb{R}^+$ tels que $p(l-1) + 1 \leq k \leq l$, on a $\text{curv}(P)(k, l) < 0$.

A noter que cette propriété est facile à montrer lorsque l'on remplace la fonction $P(k, l)$ — qui interpole $B(p, l, k)$ de manière exacte — par son approximation de type “grandes déviations”,

$$Q(p, l, k) = \exp \left(-l \ln \frac{k}{lp} - (l-k) \ln \frac{l-k}{l(1-p)} \right). \quad (38)$$

4.6 Vision humaine et vision artificielle

Le travail que je viens de décrire dans les sections précédentes s'inscrit dans une volonté de formaliser mathématiquement les lois de la perception visuelle énoncées par les gestaltistes. Ceci amène une réflexion sur la comparaison entre vision humaine et vision artificielle, dont je voudrais discuter certains aspects dans cette section.

Du point de vue purement physique, on ne peut nier une certaine analogie : à l'instar de l'image numérique, l'image rétinienne est échantillonnée par un nombre fini de capteurs. Cependant, la comparaison tourne court car contrairement aux matrices CCD, les capteurs humains sont de deux natures différentes (les cônes et les bâtonnets), et répartis de manière fortement non uniforme (il y a en particulier une très forte concentration de cônes dans une petite zone de la rétine, la fovéa).

Au-delà de cette comparaison, la vision humaine est indéniablement une source d'inspiration en vision artificielle. L'axiome d'architecture pyramidale nécessaire à la construction du *scale-space* [2] a été historiquement justifié par l'absence de boucle de retour possible (*feedback*) en vision préattentive. Les primitives utilisées en analyse d'images (*edges* notamment) sont aussi clairement inspirées par notre propre perception des images. L'utilisation d'ondelettes de Gabor directionnelles en vision artificielle, est une modélisation directe des fameux “détecteurs d'orientation” de l'aire visuelle primaires découverts par Hubel et Wiesel [52].

Jusqu'à quel point faut-il pousser l'analogie entre vision artificielle et vision humaine ? Probablement pas au point de vouloir “copier” cette dernière dans son fonctionnement précis. L'architecture du cerveau (parallèle) et d'un ordinateur (séquentiel) n'ont pas grand chose en commun, et l'on ne voit guère pourquoi le fonctionnement précis du système visuel humain serait particulièrement adapté à la vision artificielle. En tout état de cause, “les avions ne battent pas des ailes”, et la facilité que nous avons à analyser les images n'est qu'apparente, puisqu'un tiers du cortex cérébral est dédié à cette tâche ! Il me semble donc plus raisonnable de se contenter de voir dans la vision humaine une preuve de faisabilité, une borne de complexité, et une source d'inspiration dans laquelle puiser des principes organisateurs pour la vision artificielle.

Cette inspiration est en effet légitime si l'on voit la vision humaine (ou plus généralement animale) comme un modèle de système de vision, pas nécessairement optimal mais indubitablement

performant. Sans aller jusqu'à ériger les lois de la Gestalt en principe nécessaire à tout système de vision artificielle, il semble fructueux de s'en inspirer et d'essayer de les mettre en œuvre dans un système de vision artificielle. L'adéquation étonnante que nous avons montrée dans [P22] entre des seuils psychovisuels mesurés et ceux prédits par notre formalisation du principe de Helmholtz, nous conforte dans cette voie. Si l'on se place dans cette perspective gestaltiste, force est de constater que les travaux menés avec Agnès Desolneux et Jean-Michel Morel ne font que construire certains atomes (les gestalts partielles) d'un système de vision pyramidal qui devra mettre en œuvre la détection récursive des gestalts, la coopération et l'induction [P8, P16]. Il y a donc encore plusieurs principes à formaliser avant de pouvoir parler de "pyramide visuelle".

4.7 Vers des modèles sans paramètres ?

Au-delà de la formalisation de la perception visuelle, le principe de Helmholtz nous apporte un outil précieux pour la détection de structures dans les images. Tout d'abord, la détection *a contrario*, telle que nous l'avons formalisée, peut revendiquer une certaine objectivité, dans la mesure où il est beaucoup plus simple de définir un modèle à réfuter (typiquement un modèle uniforme) qu'un modèle précis des objets que l'on souhaite détecter. Un tel principe d'économie, souvent rencontré en mathématiques, limite sérieusement la part d'arbitraire (choix de constantes notamment) qui intervient dans la définition du procédé de détection, et facilite par conséquent son évaluation et sa justification *a posteriori*. Ensuite, la réduction de toutes les mesures nécessaires à la détection à un seul critère — l'espérance du nombre de fausses alarmes — est extrêmement utile, tant du point de vue pratique (contrôle de la détection à l'aide d'un seul curseur avec, qui plus est, un étalonnage absolu de ce curseur), que théorique (possibilité d'étudier les propriétés qualitatives, quantitatives, et structurelles de l'espace des objets détectables). Cette dernière propriété est si forte qu'elle peut parfois justifier à elle seule le choix du modèle *a contrario* : même si, pour une application précise, la détection *a contrario* n'est pas valide en tant que modèle, elle peut se justifier en tant que moyen objectif de combiner deux ou plusieurs mesures. Si, pour détecter des structures linéaires par exemple, on mesure un paramètre de longueur l et un paramètre de précision α , la combinaison $p = l \log \alpha$ que produira vraisemblablement le principe de Helmholtz est déjà intéressante en soi, ne serait-ce que par le modèle d'iso-détection $l \cdot \log \alpha = \text{constante}$ qu'elle propose.

Voici un premier exemple d'utilisation du modèle *a contrario* hors du cadre de la détection de gestalts. Il y a un an, j'ai entrepris d'étudier avec Bérenger Stival, actuellement en thèse sous ma direction, le problème de la détection de rigidité entre deux nuages de points [P26]. Il s'agissait de pouvoir déterminer automatiquement la validité d'appariements de points entre deux images d'une paire stéréo, c'est-à-dire de prouver la rigidité de la scène observée. Il est bien connu que la projection perspective de n points de \mathbb{R}^3 dans le plan focal de deux caméras non calibrées est contrainte dès que $n \geq 8$, mais nous avons en quelque sorte étendu et quantifié ce résultat en donnant un critère simple permettant de décider si le fait de détecter que, parmi n appariements de points, k d'entre eux sont la projection à α près de k points de \mathbb{R}^3 , est significatif (c'est-à-dire non accidentel) ou non. L'espérance du nombre de fausses alarmes que

nous avons obtenu est

$$NFA(\alpha, n, k) = 3(n-7) \cdot \binom{n}{k} \cdot \binom{k}{7} \cdot \alpha^{k-7}. \quad (39)$$

Asymptotiquement ($n \rightarrow +\infty$, $1-p = k/n$ fixé), on peut en déduire que la rigidité n'est significative que si

$$\log \alpha \leq \log(1-p) + \frac{p}{1-p} \log p - \frac{15 \log k}{2k} + O_{n \rightarrow +\infty} \left(\frac{1}{n} \right), \quad (40)$$

et donc pour k assez grand, aucune détection de rigidité n'est possible lorsque

$$\log \alpha \geq \log(1-p) + \frac{p}{1-p} \log p.$$

Le critère (39) s'est révélé très utile pour la détection effective de rigidité dans une paire stéréo, et pour l'identification des mauvais appariements (*outliers*). L'algorithme stochastique que nous avons proposé permet ainsi de repérer des bons appariements quand ceux-ci ne représentent que 10% du nombre total d'appariements (*i.e.* en présence de $p = 90\%$ de détracteurs). Nous sommes ainsi parvenus, grâce à la combinaison de α , k et n réalisée par (39), à améliorer sensiblement les meilleurs algorithmes actuels (M-estimateurs, LMedS, RANSAC), dont les performances chutent très vite au-delà de 50% de détracteurs. Nous avons aussi étudié le cas de points non appariés, où il apparaît que la détection de rigidité reste possible mais que des seuils absolus existent (par exemple, $\alpha < 1/32$), et que les conditions optimales de détection ne sont pas asymptotiques : à α et p fixés, la détection la meilleure est obtenue pour une valeur finie de n .

Pour citer un autre exemple, j'ai aussi étudié avec Zhang Dan [P28] le moyen de prouver la similarité de deux images modulo n'importe quel changement de contraste. Là encore, nous avons utilisé une modélisation *a contrario*, particulièrement adaptée puisque l'invariance morphologique permet de normaliser le contraste des images et donc de définir une image aléatoire sans référence à son histogramme.

5 Modélisation et expérimentation

J'ai essayé de montrer dans ce mémoire combien la modélisation est au cœur du traitement d'images. Les théories que nous avons pu évoquer en restauration d'images, en vision artificielle, etc. reposent essentiellement sur une modélisation de ce que sont (ou ne sont pas) les images. Même si, comme tout mathématicien, je suis sensible à la beauté intrinsèque d'une belle théorie mathématique, je trouve cette interaction entre la formalisation mathématique et les objectifs concrets que l'on cherche à atteindre réellement fascinante. Quoi de plus excitant, pour un chercheur, que de découvrir une coïncidence inattendue entre le monde réel et le modèle qu'il étudie ?

Puisque la modélisation joue un rôle primordial en traitement d'images, il faut pouvoir étudier, outre les implications théoriques d'un modèle, sa pertinence en tant que modèle. Ceci suppose de réelles possibilités d'expérimentation, et il faut bien reconnaître que le traitement d'images est une discipline particulièrement propice à satisfaire cette exigence, en particulier

depuis l'avènement massif de l'informatique. Néanmoins, le développement systématique de l'expérimentation numérique nécessite des outils adaptés.

Le logiciel *MegaWave*, devenu par la suite *MegaWave2*, est l'exemple d'un tel outil. Il s'agit d'un logiciel libre, créé en 1990 par Jacques Froment et Sylvain Parino, dont les principaux contributeurs font ou ont fait partie de l'équipe de Jean-Michel Morel. Destiné en premier lieu à des chercheurs, il met à la disposition de l'utilisateur à la fois une bibliothèque d'algorithmes issus de la recherche, et une plate-forme de développement qui lui permet de se concentrer uniquement sur les aspects purement algorithmiques des programmes qu'il écrit (les entrées-sorties sur disque, les paramètres optionnels, la syntaxe en ligne de commande, etc. sont automatiquement pris en charge par le préprocesseur). Ce logiciel a permis la construction progressive de systèmes complexes mettant en jeu plusieurs "gros" algorithmes écrits par des chercheurs différents. L'algorithme de reconnaissance de formes évoqué dans la section 2.4.4 en est un exemple.

En 1998, j'ai rejoint Jacques Froment dans l'aventure *MegaWave2*, en prenant la responsabilité du développement de la bibliothèque d'algorithmes, ce qui permettait à Jacques Froment de se concentrer sur l'amélioration du système (préprocesseur, etc.). J'ai entrepris la refonte de l'architecture de cette bibliothèque, et ajouté de nombreuses fonctionnalités, (environ 100 modules, sur un total de 260), notamment à des fins d'enseignement.

Dès que l'on cherche à comprendre l'influence pratique d'au moins deux paramètres d'un système, il faut pouvoir réaliser des expériences réellement interactives, sans quoi les manipulations sont trop fastidieuses. Les outils de visualisation que j'ai introduits dans *MegaWave2* ont, je crois, contribué à une meilleure compréhension de certains modèles. Je pense notamment aux outils liés à la visualisation multi-échelle (zoom) et multi-modèle (interpolation) des lignes de niveau, et à leur calcul en temps réel (modules `llview`, `fkview`). Le développement d'une version Linux, la création d'un site internet fourni (<http://www.cmla.ens-cachan.fr/~megawave>), ont achevé de contribuer à la diffusion importante de *MegaWave2* depuis plus de deux ans. Nous enregistrons actuellement environ 900 connexions par mois, dont 450 liées au téléchargement du logiciel¹⁵.

Ces dernières années, j'ai eu l'occasion d'enseigner le traitement d'images, tantôt d'un point de vue théorique, tantôt sous la forme de travaux pratiques, dans plusieurs écoles d'été ou écoles doctorales. Le cours que je donne actuellement dans le DEA "Mathématiques, Vision et Apprentissage" est centré sur la modélisation : j'essaie de développer au maximum l'interaction entre les connaissances théoriques et l'expérimentation numérique. Il s'agit aussi bien d'illustrer des théorèmes mathématiques par des expériences que de comprendre des phénomènes numériques à l'aide de résultats théoriques. L'intérêt pédagogique d'un logiciel comme *MegaWave* apparaît alors clairement : la possibilité de visualiser à tout moment le programme source d'un module que l'on expérimente est une exigence légitime, au même titre que la démonstration d'un théorème. Un logiciel commercial comme *Matlab*, même s'il offre des possibilités très intéressantes, ne donne pas accès aux sources, et conduit à des situations de type "boîte noire" peu compatibles avec la curiosité justifiée d'un chercheur ou futur chercheur.

¹⁵Ces statistiques, établies par un site indépendant, sont consultables sur le site *MegaWave*.

Je travaille depuis environ un an à la réalisation d'un *journal logiciel*, dans lequel l'exigence de reproductibilité de la recherche prônée par *MegaWave2* serait pleinement satisfaite. Il s'agit de donner à toute personne le moyen de tester facilement (idéalement, en ligne) les algorithmes publiés par les chercheurs, afin de garantir une objectivité maximale de la recherche appliquée et de permettre ainsi une interaction plus importante entre les groupes de recherche en traitement d'images. Ce principe de rendre l'expérimentation numérique interactive et facile à mettre en œuvre est tout à fait dans l'esprit du projet "La Forge" (<http://www.cmla.ens-cachan.fr/forge/>), auquel j'ai l'honneur d'être associé, projet développé au CMLA et subventionné par le RIAM (réseau Recherche et Innovation en Audiovisuel et Multimédia).

6 Axes de recherche

En conclusion, je voudrais présenter deux axes de recherche que je souhaite développer dans les années à venir.

6.1 Détection systématique de structures surdéterminées

Tout d'abord, je souhaite approfondir la formalisation du principe de Helmholtz, que j'ai utilisé avec Agnès Desolneux et Jean-Michel Morel dans le cadre de la détection de gestalts, avant de l'appliquer à la détection de rigidité [P26] et de similarité entre deux images [P28]. Je souhaite maintenant étendre ce modèle de détection *a contrario* plus systématiquement, en lui donnant les moyens de construire automatiquement (par détection pyramidale) les structures les plus saillantes, c'est-à-dire les plus opposées au hasard. A l'instar de Stuart Geman et ses collaborateurs [9, 40], j'ai choisi pour cela de m'intéresser à l'analyse de textes, plus simple par certains aspects : contrairement à l'image, les atomes de base — les lettres — sont bien définis pour un texte, et la donnée est une fonction définie sur un espace de dimension 1. J'ai établi en ce sens un programme de recherche sur trois ans intitulé "Détection de structures surdéterminées dans les images et les textes", inspiré notamment des travaux de Finch et Chater [34] et d'Elman [30]. Ce projet, qu'il serait trop long de décrire ici en détail, consiste en un certain sens à rendre effective (*i.e.* calculable) la notion de statistique suffisante au sens de Kolmogorov. J'ai eu l'honneur de voir ce programme de recherche accepté (et donc financé pour trois ans) dans le cadre de l'Action Concertée Incitative "Jeunes Chercheurs" menée en 2001 par le Ministère de la Recherche.

6.2 Échantillonnage et modèles numériques

Le second axe de recherche que je souhaite développer est lié à la notion de résolution en traitement d'images. La théorie de Shannon, sur laquelle s'appuie toujours l'échantillonnage des images numériques — comme en attestent les travaux de Bernard Rougé qui ont conduit à la définition du système d'acquisition du satellite SPOT 5 lancé en mai 2002 —, pose des problèmes spécifiques pour l'échantillonnage des images. A cause des fréquentes discontinuités que présentent les images, les phénomènes de *ringing* (atténuation fréquentielle trop brutale),

d'*aliasing* (atténuation fréquentielle trop lente) et de flou (atténuation fréquentielle trop rapide) sont souvent inévitables dans le cadre d'un filtrage linéaire. J'étudie actuellement des alternatives non-linéaires à l'échantillonnage avec Angélique Videau, qui effectue un stage de DEA sous ma direction depuis mars 2002. L'application que nous visons est essentiellement la fusion d'images, mais la suppression de l'*aliasing* dans les films [30], par exemple, est l'une des nombreuses autres applications potentielles.

Publications et Brevets

Les références marquées en gras sont jointes en annexe.

Livres et chapitres de livres

- [P1] *Petits problèmes de mathématiques appliquées et de modélisation* (avec Brigitte Bidégaray), Springer, 2000, 145 pages.
- [P2] “Variational Snakes Theory” (avec Agnès Desolneux et Jean-Michel Morel), in *Geometric Level Set Methods in Imaging, Vision and Graphics*, S. Osher and N. Paragios éditeurs, Springer Verlag, à paraître en janvier 2003.

Articles dans des revues avec comité de lecture

- [P3] “Affine Plane Curve Evolution: a Fully Consistent Scheme”, *IEEE Transactions On Image Processing*, vol. 7:3, pp. 411-420, 1998.
- [P4] “Meaningful Alignments” (avec Agnès Desolneux et Jean-Michel Morel), *International Journal of Computer Vision*, vol. 40:1, pp. 7-23, 2000.
- [P5] “Edge Detection by Helmholtz Principle” (avec Agnès Desolneux et Jean-Michel Morel), *Journal of Mathematical Imaging and Vision*, vol 14:3, pp. 271-284, 2001.
- [P6] “Geometric Computation of Curvature Driven Plane Curve Evolutions” (avec Frédéric Cao), *SIAM Journal on Numerical Analysis*, vol. 39:2, pp. 624-646, 2001.
- [P7] “Dequantizing image orientation” (avec Agnès Desolneux, Saïd Ladjal et Jean-Michel Morel), *IEEE Transactions on Image Processing*, vol. 11:10, pp. 1129-1140, 2002.
- [P8] “A grouping principle and four applications” (avec Agnès Desolneux et Jean-Michel Morel), preprint CMLA no 2001-22, à paraître dans *IEEE Transactions on Pattern Analysis and Machine Intelligence* (numéro spécial).

Actes de conférences

- [P9] “Perspective invariant multiscale analysis of movies”, *International Symposium on Optical Science Engineering*, San Diego, vol. 2567, pp. 84-94, 1995.

- [P10] “Multiscale Analysis of Movies for Depth Recovery”, Proceedings of the *International Conference on Image Processing*, Washington, vol. 3, pp. 25-28, 1995.
- [P11] “PDE’s, motion analysis, and 3D reconstruction from movies”, in *Partial Differential Equations - Theory and numerical solution*, Chapman & Hall/CRC Research Notes in Mathematics, vol. 406, pp. 273-282, 1998.
- [P12] “Geometric Multiscale Representation of Numerical Images” (avec Georges Koepfler), in *Springer Lecture Notes in Computer Science*, vol. 1682, pp. 339-350, 1999.
- [P13] “Affine Invariant Mathematical Morphology Applied to a Generic Shape Recognition Algorithm” (avec José-Luis Lisani, Pascal Monasse et Jean-Michel Morel), Proceedings of the *International Symposium on Mathematical Morphology*, Palo Alto, California, 2000.
- [P14] “A stable numerical algorithm for generalized curve shortening flow” (avec Frédéric Cao), in *Free Boundary Problems: Theory and Applications*, Gatuko International Series, Mathematical Sciences and Applications 14, pp. 61-76, 2000.
- [P15] “Extrapolation de Spectre et Variation Totale Pondérée”, actes du GRETSI, Toulouse, 2001.
- [P16] “Automatic image analysis: a challenge for computer vision” (avec Agnès Desolneux et Jean-Michel Morel), actes du GRETSI, Toulouse, 2001.
- [P17] “Gestalt theory and Computer Vision” (avec Agnès Desolneux et Jean-Michel Morel), prépublication CMLA N° 2002-06, à paraître.

Autres

- [P18] “Analyse Multiéchelle de films pour la reconstruction du relief”, note au *Compte rendu de l’Académie des Sciences*, Paris, tome 320, série I, pp. 279-284, 1995.
- [P19] *Traitement numérique d’images et de films : équations aux dérivées partielles préservant forme et relief*, thèse de doctorat, Université Paris-Dauphine, 1997.
- [P20] “A Depth-Compatible Multiscale Analysis of Movies”, preprint CMLA no 9803.
- [P21] “A review of P.D.E. models in image processing and image analysis”, (avec Frédéric Guichard et Jean-Michel Morel), *Journal de Physique IV*, vol. 12, pp. 137-154, 2002.
- [P22] “Computational Gestalts and Perception Thresholds” (avec Agnès Desolneux et Jean-Michel Morel), à paraître dans *Journal of Physiology* (numéro spécial).

Brevets

- [P23] “Procédé et système pour modifier une image numérique en prenant en compte son bruit”, demande PCT N°PCT/FR02/01908 déposée le 5 juin 2002 (inventeurs : Laurent Chanas, Frédéric Guichard, Lionel Moisan, Bruno Liège).
- [P24] “Procédé et système pour produire des informations formatées liées aux défauts d’au moins un appareil d’une chaîne, notamment au flou”, demande PCT N°PCT/FR02/01915 déposée le 5 juin 2002 (inventeurs : Laurent Chanas, Frédéric Guichard, Lionel Moisan, Bruno Liège).

Travaux soumis ou en cours

- [P25] “Maximal Meaningful Events and Applications to Image Analysis” (avec Agnès Desolneux et Jean-Michel Morel), preprint CMLA no 2000-22, soumis à *Annals of Statistics*.
- [P26] “A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix” (avec Bérenger Stival), preprint CMLA no 2002-11, soumis à *International Journal of Computer Vision*.
- [P27] “Asymptotic Estimates and Inequalities for the Tail of the Binomial Distribution”, en cours de rédaction.
- [P28] “Illumination Invariant Image Comparison” (avec Zhang Dan), en cours de rédaction.
- [P29] “Detection of constant width in images” (avec Jean-Michel Morel et Emmanuel Villéger), en cours de rédaction.

Articles de vulgarisation

- [P30] “Traitement d’images”, *Pour La Science*, février 1995.
- [P31] “Les défauts du DVD sous l’œil de l’équation” (avec Jean-Michel Morel), *La Recherche*, mai 2001.

Références

- [1] L. Alvarez, Y. Gousseau, and J.-M. Morel, “The size of objects in natural and artificial images”, *Advances in Imaging and Electron Physics* 111, pp. 167-242, Academic Press, 1999.
- [2] L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel, “Axioms and fundamental equations of image processing”, *Archives for Rational Mechanics* 123, pp. 199-257, 1993.
- [3] Y. Amit, D. Geman, “A computational model for visual selection”, *Neural Computation* 11:7, pp. 1691-1715, 1999.
- [4] S. Angenent, G. Sapiro, A. Tannenbaum, “On the affine heat equation for nonconvex curves”, *J. of the Am. Math. Soc.* 11:3, pp. 601-634, 1998.
- [5] G. Aronsson, “Extensions of functions satisfying lipschitz conditions”, *Arkiv för Matematik* 6, 1967.
- [6] G. Aronsson, “On the partial differential equation $u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy} = 0$ ”, *Arkiv för Matematik* 7, 1968.
- [7] G. Bellettini, G. Bouchitté, I. Fragalà, “BV Functions with Respect to a Measure and Relaxation of Metric Integral Functionals”, *Journal of Convex Analysis* 6:2, pp. 349-366, 2000.
- [8] G. Bellettini, V. Caselles, M. Novaga, “The total variation flow in \mathbb{R}^n ”, preprint, 2001, to appear in *J. of Differential Equations*.
- [9] E. Bienenstock, S. Geman, D. Potter, “Compositionality, MDL Priors, and Object Recognition”, in *Advances in Neural Information Processing Systems* 9, M.C. Mozer, M.I. Jordan, and T. Petsche eds, MIT Press, pp. 838-844, 1997.
- [10] E.J. Candes, *Ridgelets : theory and applications*, PhD Dissertation, Stanford University, 1998.
- [11] E.J. Candes, D.L. Donoho, “Curvelets”, Tech. report, Dept. of Statistics, Stanford University, 1999.
- [12] V. Caselles, J-M. Morel, C. Sbert, “An Axiomatic Approach to Image Interpolation”, *IEEE Transactions On Image Processing* 7:3, pp. 376-379, 1998.
- [13] K.R. Castleman, *Digital Image Processing*, Prentice Hall, 1996.
- [14] F. Catté, F. Dibos, G. Koepfler, “A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets”, *SIAM J. of Num. Analysis* 32:6, pp.1895-1909, 1995.

- [15] R. Chellappa and R. Kashyap, "Texture synthesis using 2-D non-causal autoregressive models", *IEEE Trans. Acoust. Speech Signal Process.* 33, pp. 194-203, 1985.
- [16] A. Cohen, "Wavelets and multiscale signal processing", *Appl. Math. & Math. Comp.* 11, Chapman & Hall ed., 1995.
- [17] A. Cohen, J. Froment, "Image compression and multiscale approximation", in *Wavelets and Applications*, ed. Y. Meyer, Masson, pp. 181-206, 1992.
- [18] F.S. Cohen, M.A.S. Patel, "Modeling and Synthesis of Images of 3D Textured Surfaces", *Computer Vision, Graphics and Image Processing* 53:6, pp. 501-510, 1991.
- [19] T. Cohignac, *Reconnaissance de formes planes*, thèse de doctorat, Université Paris-Dauphine, 1994.
- [20] M.G. Crandall, H. Ishii, P.-L. Lions, "User's guide to viscosity solutions of second order partial differential equations", *Bulletin of Amer. Math. Society* 27, pp. 1-67, 1992.
- [21] M.G. Crandall, H. Ishii, P.-L. Lions, "Viscosity solutions : a primer", in *Viscosity Solutions and Applications*, Springer Lecture Notes in Mathematics 1660, 1997.
- [22] D. Dacunha-Castelle, F. Gamboa, "Maximum d'entropie et problème des moments", *Ann. Inst. Henri Poincaré* 26:4, pp. 567-596, 1990.
- [23] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.
- [24] I. Daubechies, "Orthonormal bases of compactly supported wavelets", *Commun. Pure Appl. Math.* 41, pp. 909-996, 1988.
- [25] G. Demoment, "Image reconstruction and restoration : overview of common estimation structure and problems", *IEEE Transactions on Acoustics, Speech and Signal Processing* 37:12, pp. 2024-2036, 1989.
- [26] M. Do, M. Vetterli, "Orthonormal finite ridgelet transform for image compression", *Proc. IEEE Int. Conf. on Image Proc.*, Vancouver, 2000.
- [27] D. Donoho, I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika* 81, pp. 425-455, 1994.
- [28] D.L. Donoho, "Ridge Functions and Orthonormal Ridgelets", Tech. report, Dept of Statistics, Stanford University, 1998.
- [29] A.A. Efros, T.K. Leung, "Texture Synthesis by Non-parametric Sampling", *IEEE Int. Conf. on Computer Vision*, Corfu, Greece, 1999.
- [30] J.L. Elman, "Finding Structure in Time", *Cognitive Science* 14, pp. 179-211, 1990.
- [31] L.C. Evans, "A geometric Interpretation of the Heat Equation with Multivalued Initial Data", *SIAM Journal of Mathematical Analysis* 27:2, pp. 932-958, 1996.

- [32] M.D. Fairchild, *Color Appearance Models*, Addison Wesley, 1998.
- [33] O. Faugeras, Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, 2001.
- [34] S. Finch, N. Chater, “A hybrid approach to the automatic learning of linguistic categories”, in Igor Aleksander and John Taylor eds, *Artificial Neural Networks 2*, Amsterdam, ICANN, Elsevier, 1992.
- [35] F. Fleuret, D. Geman, “Coarse-to-fine face detection”, *Int. J. of Computer Vision* 41:1, pp. 85-107, 2001.
- [36] M. Gage, R.S. Hamilton, “The heat equation shrinking convex plane curves”, *J. of Diff. Geometry* 23, pp. 69-96, 1986.
- [37] D. Geman, A. Koloydenko, “Invariant statistics and coding of natural microimages”, *Proc. IEEE Workshop on Statistical and Computational Theories of Vision*, Fort Collins, CO, 1999.
- [38] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Machine Intell.* 6 , pp. 721-741, 1984.
- [39] D. Geman, B. Jedynek, “An active testing model for tracking roads from satellite images”, *IEEE Trans. Pattern Anal. Mach. Intell.* 18, pp. 1-14, 1996.
- [40] S. Geman, D.F. Potter, Z. Chi, “Composition Systems”, Technical Report, Division of Applied Mathematics, Brown University, 1998.
- [41] S. Geronimo, H. Woerdeman, “Positive Extensions and Riesz-Fejer Factorization for Two-Variable Trigonometric Polynomials”, preprint, accessible en ligne.
- [42] E. Giusti, *Minimal Surfaces and functions of bounded variation*, Birkhäuser, 1984.
- [43] Y. Gousseau, *The distribution of shapes in natural images*, thèse de doctorat, Université Paris-Dauphine, 2000.
- [44] M. Grayson, “The heat equation shrinks embedded plane curves to round points”, *J. of Diff. Geometry* 26, pp. 285-314, 1987.
- [45] U. Grenander, G. Szegö, *Toeplitz orms and Their Applications*, University of California Press, Berkeley, 1958.
- [46] F. Guichard, *Axiomatisation des analyses multi-échelles d’images et de films*, thèse de doctorat, Université Paris-Dauphine, 1994.
- [47] F. Guichard, F. Malgouyres, “Total variation based interpolation”, *Proc. Europ. Signal Proc. Conf.* 3, pp. 1741-1744, 1998.
- [48] F. Guichard, J.-M. Morel, *Image Analysis and PDE’s*, à paraître.
- [49] R.M. Haralick, S.R. Sternberg, X. Zhuang, “Grayscale Morphology”, *Comp. Vision, Graphics, Image Proc.* 35, 1986.

- [50] G. Huang, D. Mumford, "The Statistics of Natural Images and Models", *Proc. IEEE Comp. Vision and Pattern Recognition*, pp. 541-547, 1999.
- [51] G. Huang, A. Lee, D. Mumford, "Statistics of range images", *Proc. IEEE Comp. Vision and Pattern Recognition*, pp. 324-331, 2000.
- [52] D.H. Hubel, T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *Journal of Physiology* 160, pp. 106-154, 1962.
- [53] B.R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer", *IEEE Trans. Comput.* 22, 1973.
- [54] H. Ishii, "An Approximation Scheme for Gauss Curvature Flow and its Convergence", preprint.
- [55] ISII: Indexation sémantique d'images sur internet, projet soutenu par le Réseau National de Recherche en Télécommunications (RNRT), http://www.telecom.gouv.fr/rnrt/projets/res_30_ap00.htm.
- [56] R. Jensen, "Uniqueness of Lipschitz extension : minimizing the sup norm of the gradient", *Arch. Rat. Mech. Anal.* 123, pp. 51-74, 1993.
- [57] JPEG: The Joint Photographic Experts Group, ISO/IEC JTC1 SC29, <http://www.jpeg.org>.
- [58] R.L. Kashyap, "Characterization and estimation of two-dimensional ARMA models", *IEEE Trans. on Information Theory* 30, pp. 736-745, 1984.
- [59] G.Kanizsa, *Grammatica del Vedere*, Il Mulino, Bologna, 1980. Traduction : *La grammaire du voir*, Diderot Editeur, Arts et Sciences, 1996.
- [60] G. Kanizsa, *Organization in Vision*, Holt, Rinehart & Winston, 1979.
- [61] H.J. Landau, "Maximum entropy and the moment problem", *Bull. Amer. Math. Soc.* 16:1, 1987.
- [62] Y. Leclerc, "Constructing Simple Stable Descriptions for Image Partitioning", *Int. J. of Computer Vision* 3, pp. 73-102, 1989.
- [63] E. Le Pennec, S. Mallat, "Image Compression with Geometrical Wavelets", *Proc. IEEE Int. Conf. on Image Proc.*, Vancouver, 2000.
- [64] H. Lev-Ari, S.P. Parker, T. Kailath, "Multidimensional maximum entropy covariance extension", *IEEE Trans. Inf. Theory* 35, pp. 497-598, 1989.
- [65] C. Lopez, J.M. Morel, "Axiomatization of shape analysis and application to texture hyperdiscrimination", *Proceedings of the Trento Conference on Surface Tension and Movement by Mean Curvature*, De Gruyter, Berlin, 1992.

- [66] D. Lowe, *Perceptual Organisation and Visual Recognition*, Kluwer Academic Publishers, 1985.
- [67] H. Maitre, "Un panorama de la transformation de Hough", *Traitement du signal* 2:4, 1985.
- [68] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1988.
- [69] P. Maragos, "Tutorial on advances in morphological image processing and analysis", *Optical Engineering* 26:7, pp. 623-632, 1987.
- [70] D. Marr, *Vision*, Freeman and Co., 1982.
- [71] G. Matheron, *Random Sets and Integral Geometry*, John Wiley, New York, 1975.
- [72] W. Metzger, *Gesetze des Sehens*, Waldemar Kramer, 1975.
- [73] Y. Meyer, *Wavelets and operators*, vol. 1, 2, 3, Cambridge University Press, 1992.
- [74] K. Mikula, D. Sevcovic, "Solution of nonlinearly curvature driven evolution of plane curves", *Applied Numerical Mathematics* 31:2, pp.191-207, 1999.
- [75] P. Monasse, F. Guichard, "Fast Computation of a Contrast Invariant Image Representation" *IEEE Transactions on Image Processing* 9:5, pp. 860-872, 2000.
- [76] P. Monasse, *Représentation morphologique d'images numériques et application au recalage d'images*, thèse de doctorat, Université Paris-Dauphine, 2000.
- [77] D. Mumford, "The Dawning of the Age of Stochasticity", allocution à l'Accademia Nazionale dei Lincei, accessible en ligne, 1999.
- [78] S. Osher, J.A. Sethian, "Fronts Propagating with Curvature-Dependent Speed : Algorithms Based on Hamilton-Jacobi Formulations", *Journal of Computational Physics* 79:1, pp. 12-49, 1988.
- [79] Rissanen, J. "A universal prior for integers and estimation by Minimum Description Length", *Annals of Statistics* 11:2, pp. 416-431, 1983.
- [80] Rissanen, J. *Stochastic Complexity in Statistical Inquiry*, World Scientific Press, 1989.
- [81] B. Romeny, L. Florack, J. Koenderink, M. Viergever, *Scale-Space Theory in Computer Vision*, Springer Verlag, 1997.
- [82] B. Rougé, C. Latry, "Procédé d'acquisition d'une image par balayage pousse-balai", brevet international PCT/FR96/01188, 1996.
- [83] B. Rougé, A. Seghier, "Nonlinear spectral extrapolation: new results and their application to spatial and medical imaging", *SPIE Proceedings* 2568, pp.279-289, 1995.
- [84] L. Rudin, S. Osher, C. Fatemi, "Nonlinear total variation based noise removal algorithm", *Physica D* 60, pp. 259-268, 1992.

- [85] L. Rudin, S. Osher, "Total Variation based image restoration with free local constraints", *Proc. IEEE Int. Conf. on Image Processing 1*, Austin, pp. 31-35, 1994.
- [86] G. Sapiro, A. Tannenbaum, "On affine plane curve evolution", *J. of Functional Analysis* 119, pp. 79-120, 1994.
- [87] A. Seghier, "Extension de fonctions de type positif, II", preprint no 89-47, Université Paris-Sud, 1989.
- [88] J. Serra, *Image analysis and mathematical morphology*, Academic Press, 1982.
- [89] J.A. Sethian, *Level Set Methods*, Cambridge University Press, 1996.
- [90] C.E. Shannon, "A Mathematical Theory of Communication", *Bell System Technical Journal* XXVII:3, pp. 379-423, 1948.
- [91] J. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Signal Processing* 41:12, pp. 3445-3462, 1993.
- [92] E.P. Simoncelli, B. Olshausen, "Natural image statistics and neural representation", *Annual Review of Neuroscience* 24, pp. 1193-1216, 2001.
- [93] D. Slepian, "Some Comments on Fourier Analysis, Uncertainty and Modeling", *SIAM Review* 25:3, pp. 379-393, 1983.
- [94] C.V. Stewart, "MINPRAN: A New Robust Estimator for Computer Vision", *IEEE Trans. Pattern anal. Mach. Intell.* 17, pp. 925-938, 1995.
- [95] A.N. Tikhonov, "Regularization of incorrectly posed problems", *Soviet Mathematics Doklady* 4, pp. 1624-1627, 1963.
- [96] M. Unser, A. Aldroubi, M. Eden, "Fast B-Spline Transforms for Continuous Image Representation and Interpolation", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13:3, pp. 277-285, 1991.
- [97] P.D. Wendt, E.J. Coyle, N.C. Gallagher, "Stack filters", *IEEE Trans. Acoust., Speech, Signal Processing* 34, pp. 898-911, 1986.
- [98] M. Wertheimer, "Untersuchungen zur Lehre der Gestalt", II, *Psychologische Forschung* 4, pp. 301-350, 1923.
- [99] A.P. Witkin, "Scale-Space filtering", *Proceedings of IJCAI*, Karlsruhe, pp. 1019-1021, 1983.
- [100] L. Yaroslavsky, M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, 1996.
- [101] W.P. Ziemer, *Weakly differentiable functions*, Springer-Verlag, New-York, 1989.

Publications jointes

- [P3] L. Moisan, “Affine Plane Curve Evolution: a Fully Consistent Scheme”, *IEEE Transactions On Image Processing*, vol. 7:3, pp. 411-420, 1998.
- [P6] F. Cao, L. Moisan, “Geometric Computation of Curvature Driven Plane Curve Evolutions”, *SIAM Journal on Numerical Analysis*, vol. 39:2, pp. 624-646, 2001.
- [P12] G. Koepfler, L. Moisan, “Geometric Multiscale Representation of Numerical Images” in *Springer Lecture Notes in Computer Science*, vol. 1682, pp. 339-350, 1999.
- [P13] J.-L. Lisani, L. Moisan, P. Monasse, J.-M. Morel, “Affine Invariant Mathematical Morphology Applied to a Generic Shape Recognition Algorithm”, Proceedings of the *International Symposium on Mathematical Morphology*, Palo Alto, California, 2000.
- [P11] L. Moisan, “PDE’s, motion analysis, and 3D reconstruction from movies”, in *Partial Differential Equations - Theory and numerical solution*, Chapman & Hall/CRC Research Notes in Mathematics, vol. 406, pp. 273-282, 1998.
- [P4] A. Desolneux, L. Moisan, J.-M. Morel, “Meaningful Alignments”, *International Journal of Computer Vision*, vol. 40:1, pp. 7-23, 2000.
- [P5] A. Desolneux, L. Moisan, J.-M. Morel, “Edge Detection by Helmholtz Principle”, *Journal of Mathematical Imaging and Vision*, vol 14:3, pp. 271-284, 2001.
- [P17] A. Desolneux, L. Moisan, J.-M. Morel, “Gestalt theory and Computer Vision”, prépublication CMLA N° 2002-06, à paraître.
- [P2] A. Desolneux, L. Moisan, J.-M. Morel, “Variational Snakes Theory”, in *Geometric Level Set Methods in Imaging, Vision and Graphics*, S. Osher and N. Paragios éditeurs, Springer Verlag, à paraître en janvier 2003.
- [P26] L. Moisan, B. Stival, “A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix”, preprint CMLA n° 2002-11, soumis à *International Journal of Computer Vision*.