ministère éducation nationale

ÉduSCOL [dgesco]

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Formation continue
Publications

Actes de l'université d'été

Expérimentation et démarches d'investigation en mathématiques

Les sciences expérimentales et le traitement de l'image

Saint Four du 20 au 24 août 2007

octobre 2008

eduscol.education.fr/forensactes

LES SCIENCES EXPÉRIMENTALES ET LE TRAITEMENT DE L'IMAGE

par

Yves Meyer

Je remercie les organisateurs de l'Université d'été de Saint-Flour de m'avoir invité à faire cet exposé. Pour parler de l'expérimentation et des démarches d'investigation en mathématiques, j'ai choisi le cas particulier du traitement de l'image. Les mathématiques, la physique statistique, l'informatique et les neuro-sciences interviennent dans le traitement de l'image. Il en est de même pour la philosophie et la peinture. Je vais vous parler de ce riche dialogue entre disciplines scientifiques en me concentrant sur les problèmes posés par la perception et la compression des images.

1. Introduction

Le *traitement de l'image* est une discipline très jeune qui est née avec la révolution numérique. Seules les *images numériques* peuvent être manipulées à l'aide d'algorithmes qui soient *reproductibles*, comme doivent l'être toutes les expériences dans les sciences dures. Remontons dans le temps afin de mieux comprendre cette remarque. Avant l'invention de la photographie, les peintres avaient le monopole de la production des images et la perception du monde extérieur était incluse dans l'acte de peindre. En quelques traits de crayon, un dessinateur doué peut faire surgir, de façon précise et exacte, un visage familier. Le problème de la compression des *images naturelles* était ainsi résolu par le dessin. Mais l'art du dessinateur n'est pas reproductible, car la sensibilité de l'artiste y joue un rôle essentiel. Les manipulations sur les photographies argentiques étaient de l'ordre du savoir-faire et n'étaient donc pas non plus reproductibles. On pense aux retouches effectuées sur les photographies officielles des dignitaires de l'ex-URSS. Aujourd'hui les caméras numériques nous proposent des "reproductions objectives" des images naturelles. Un appareil de photographie numérique ou un camescope numérique utilisent un capteur CCD (charge couple device). Ce capteur CCD se présente comme une énorme matrice $M \times N$ comportant plusieurs millions de photorécepteurs. Chaque photorécepteur génère des charges électriques d'intensités variables en fonction de la quantité de lumière reçue et de la durée d'exposition. Ce dispositif produit les pixels qui constituent l'image numérique. Une image numérique (en noir et blanc) est donc une énorme matrice (au sens des mathématiciens) $M = m(j,k), 1 \leq j \leq M, 1 \leq k \leq N$. Chaque coefficient $m(j,k)$ représente l'éclairement de l'image en ce point (hj, hk) où $h > 0$ est le pas d'échantillonnage. Chaque image contient autant d'information qu'environ mille pages de texte. Cette image sera stockée dans la carte mémoire de l'appareil. L'analogie avec les photorécepteurs (cônes et bâtonnets) de la rétine est frappante. Les "reproductions" numériques de l'image fournies par un capteur CCD ne dépendent pas de notre perception. Elles sont objectives. On peut ensuite les modifier à volonté par des algorithmes mathématiques qui sont reproductibles. Le traitement de l'image intervient alors pour améliorer ou du moins transformer ces images numériques "brutes". *Le traitement de l'image, effectué sur des images numériques, se prête donc à l'expérimentation, condition essentielle pour être une science dure.*

Signalons tout de suite que nous n'étudierons ici que la compression avec perte (dite *lossy*) et que les images concernées sont des *images naturelles* digitalisées. Les images naturelles sont celles des scènes familières (animaux ou personnes immobiles ou en mouvement), des objets de notre environnement ou des paysages qui nous entourent. Au cours de l'évolution, le système visuel des mammifères s'est progressivement adapté à ces images naturelles. Un des résultats de cette adaptation est la capacité du cortex visuel primaire à détecter les figures géométriques simples que l'on trouve dans beaucoup d'images naturelles (alignements, courbes, etc). Cette thèse que l'on retrouve chez David Hubel demande à être développée, ce que nous ferons dans les pages qui suivent.

Après une compression avec perte, l'image n'est plus exactement la même ; elle a été simplifiée. La première relation entre perception et compression est donc évidente. *On ne dispose pas à l'heure actuelle de critères objectifs permettant de décider de la qualité d'un algorithme de compression.* On demande que l'image produite par l'algorithme soit *perceptuellement proche* de l'image originale. Dès lors la perception joue un rôle dans les critères utilisés pour mesurer la qualité des résultats. La seconde relation entre perception et compression est plus subtile, car elle repose sur l'étude de la cohérence interne des images. La compression des images fixes peut se faire à des taux fantastiques. S'il s'agissait d'un texte écrit, cela reviendrait à résumer mille pages en une demi-page. Cela reviendrait à résumer en une quizaine de lignes les Éssais de Montaigne. Ce taux est atteint dans l'algorithme de compression développé par Mallat. Ceci est évidemment impossible dans le cas du texte écrit, mais est possible dans le cas de l'image. La raison est que l'information fournie par une image naturelle est extrêmement redondante. Cette redondance vient de ce qu'une image naturelle est structurée. La *syntaxe* qui est responsable des structures présentes dans les images naturelles peut être vue comme une traduction des contraintes logiques imposées par la rationalité du monde qui nous entoure. Le philosophe Ernst Cassirer considère que la perception d'une image naturelle est basée sur la reconnaissance des formes ou des structures sous-jacentes qui y figurent. On ne peut résumer un texte si ce texte n'a pas de sens. De même on ne peut comprimer une image qui ne soit pas structurée. Comment ces structures sous-jacentes sont-elles définies ? Existe-t-il dans les images des formes simples et universelles qui soient à la base de la compression et de la perception ?

Tout ceci nous conduit aux quatre questions suivantes :

(a) *Les algorithmes de compression des images digitales sont-ils semblables à ceux utilisés par le cortex visuel primaire* ?
(b) *Quel rôle joue la perception dans la compression des images* ?
(c) *Comment la perception s'élabore-t-elle à partir de l'information de "bas niveau" fournie par le cortex visuel primaire ?*
(d) *Peut-on relier cette élaboration de la perception aux algorithmes utilisés en traitement de l'image ?*

Dans la première partie de cet exposé nous essayerons d'y répondre. Cela nous amènera à présenter plusieurs analyses des problèmes posés par la perception. Ces analyses viendront de la philosophie, de la peinture et de la neurophysiologie. Cet exposé s'orientera ensuite vers des objectifs plus modestes : nous présenterons et justifierons deux choix de modèles convenant à la description des images fixes. Nous nous limitons ici aux images numériques. Le premier modèle consiste à modéliser une image par un *sketch* défini par la donnée d'un ensemble fini de contours. Pour le peintre, il s'agira d'une esquisse, d'un dessin préalable. En traitement de l'image, les contours sont les bords des objets, ils les délimitent. Les contours peuvent être tracés ; pour le mathématicien, la somme des longueurs des contours est finie. Cela nous aménera à modéliser une image par une *fonction à variation bornée.* Ce modèle adopté, il conviendra de choisir une représentation adaptée au sens donné par David Marr ou une langue adaptée au sens de Cassirer. C'est là que nous rencontrerons l'analyse par ondelettes. Cet algorithme permet d'effectuer des "zooms arrière", c'est-à-dire de voyager à travers les échelles. Nous décrirons et interpréterons le nouveau standard de compression des images fixes dans le cadre des théories de la perception et des modèles mathématiques qui s'y raccordent. Ce nouveau standard, JPEG2000, est destiné à remplacer l'ancien standard nommé JPEG ; JPEG2000 n'est pas encore complètement achevé, bien que l'essentiel ait déjà été fait. JPEG2000 est un logiciel libre, gratuit, déjà expérimenté sur le web, dont l'amélioration sera un processus infiniment continué. Le lecteur intéressé se reportera à l'excellent site : http ://jj2000.epfl.ch (l'absence de www n'est pas une erreur). Dans les dernières sections de l'exposé écrit, nous présenterons quelques applications médicales des algorithmes de compression des images fixes, puis nous décrirons les nouveaux algorithmes (dus à Emmanuel Candès, David Donoho et Stéphane Mallat) qui font mieux que JPEG2000. Nous terminerons par l'extraordinaire découverte du *compressed sensing* qui est un premier pas vers les algorithmes de la seconde génération. On peut, en effet, envisager une nouvelle génération d'algorithmes de compression basés sur une analyse du contenu sémantique de l'image. Nous n'en sommes pas là et les algorithmes de compression dont nous disposons s'apparentent encore à la partie "bas niveau" du système visuel humain. Cette longue introduction est maintenant terminée.

2. Perception et perceptions

Comment définir le mot *perception*? Le dictionnaire *Robert* nous dit que la perception est la fonction par laquelle l'esprit se représente les objets ; cette définition y est illustrée par deux citations (de J-J. Rousseau et J-P. Sartre) qui nous aideront à organiser notre débat. Jean-Jacques Rousseau écrivait :

> "*Nos sensations sont purement passives, au lieu que toutes nos perceptions ou idées naissent d'un principe actif qui juge.*"

Jean-Paul Sartre va dans la même direction quand il nous dit que "*dans la perception, un savoir se forme lentement.*" En partant des découvertes de la neurophysiologie et des sciences cognitives, Pierre Buser analyse bien plus finement les articulations liant la sensation à la perception ; en fait, pour chacun des problèmes qu'il étudie, Pierre Buser nous propose plusieurs solutions ou théories antagonistes et c'est ce qui fait le charme et la richesse de son ouvrage, *Cerveau de soi, Cerveau de l'autre*, publié chez Odile Jacob. Reprenant l'histoire du concept de perception et remontant à 1785, il écrit :

> *C'est à Reid (1785) que nous devons semble-t-il la distinction (qui fera du chemin pendant cent cinquante années au moins) entre sensation, donnée brute de l'expérience, et perception, opération mentale de prise de conscience de la présence réelle de l'objet externe. Comme Berkeley, Reid vit dans cette opération perceptive de connaissance du monde une origine divine...*

Analysant ensuite l'œuvre de Helmholtz, Pierre Buser écrit :

> *"Helmholtz fut de ceux qui plaidèrent pour un processus en deux étapes, la sensation donnée brute liée à la mise en jeu de nos capteurs sensoriels, et la perception, représentation consciente de la réalité que nous bâtissons à partir de la sensation, grâce à nos inférences et éventuellement nos jugements...Les données récentes maintiennent-elles la dualité hierarchique entre sensation et perception?...Que la perception implique une interprétation de la donnée brute par une opération "psychologique" (qui s'opposerait au "physiologique" pur de la sensation) ne semble pas poser problème, dans la mesure précisément où l'opposition entre les deux domaines, physiologique et mental, s'estompe aux yeux de tant*

> *d'auteurs, ne serait-ce qu'avec l'abandon d'un certain dualisme. Il n'empêche que les perceptions complexes, au nombre desquelles sont, en bonne place, les classiques figures ambiguës (cubes de Necker, etc.) sont là pour nous rappeler que la perception d'une forme implique sans doute davantage que la seule réception des messages visuels..."*

Pierre Buser insiste enfin sur l'existence de différents niveaux de perception ; il y a une perception implicite ou préconsciente et il y a aussi une appréhension consciente. Il écrit à propos de cette première :

> *"Helmholtz (1866), à son tour, discuta des problèmes généraux de la perception. Il introduisit la notion d'inférence inconsciente (unbewusster Schluss) pour signifier que, dans la perception, la référence objective peut trouver sa source dans des repères qui ne sont pas immédiatement accessibles à la conscience. Ainsi, dans une perception de profondeur et de distance relative des objets, créée par la disparité des images rétiniennes, nous savons maintenant que le sujet est totalement inconscient des processus intermédiaires..."*

Un peu plus loin, Pierre Buser écrit, en commentant l'œuvre de J. Fodor :

> *"Lorsque soudain nous percevons l'arrêt d'un bruit dont nous n'étions soi-disant pas conscient, et que nous pouvons parfaitement l'identifier, comment nous souvenons-nous si bien de quelque chose dont nous n'avions pas été conscients précédemment ? Les réflexes sont stupides parce que non inférentiels, c'est-à-dire n'impliquant aucun processus mental, leur voie d'acheminement étant fixée une fois pour toutes ; ils le sont aussi parce qu'ils sont encapsulés, c'est-à-dire qu'ils se déroulent sans s'adapter aux besoins directs et immédiats de l'organisme. Quant à la perception, elle est, contrairement aux réflexes, effectivement inférentielle, c'est-à-dire qu'elle implique, lato sensu, une opération mentale et se rapproche, à cet égard de la cognition ; mais elle reste elle aussi encapsulée, comme les réflexes. Le cas de certaines illusions optico-géométriques est typique : même si le sujet sait qu'il s'agit d'une illusion, il ne pourra pas en faire abstraction. Autrement dit, le cognitivisme ancien confondrait "la complexité inférentielle de la perception avec sa pénétrabilité cognitive". Fodor, tout à son idée de séparer perception*

> *et cognition, en arrive ainsi à définir le module perceptif comme un système computationnel encapsulé, ne permettant qu'un nombre limité d'opérations qu'il réalise d'une manière rapide, mais rigide, et qui sont largement de type ascendant (bottom-up) pour l'information reçue."*

La perception peut donc être basée sur un savoir inconscient. En outre la perception est subjective. Les peintres modifient notre perception à notre insu. En fait, les peintres nous apprennent à voir ; même les peintres classiques, comme Nicolas Poussin, nous éloignent du réél naïf, du simple reflet du monde visible, pour nous introduire dans un monde nouveau qui nous paraît d'abord étrange, mais qui nous deviendra ensuite familier. Entrer dans ce nouveau monde, c'est modifier à jamais notre propre perception. Mais quelles sont les conditions requises à l'élaboration de la perception ?

3. Peintres et philosophes

Nicolas Poussin y répond quand il écrivait : "*Il ne se crée rien de visible sans distance"* Le philosophe Maurice Merleau-Ponty nous parle aussi d'éloignement :

> "*Dans la vie silencieuse de la perception, nous adhérons à quelque chose, nous le faisons nôtre, et cependant nous nous en retirons et le tenons à distance, sans quoi nous n'en saurions rien.*"

La distance vis à vis d'une œuvre, nous la créons en prenant du recul et en sentant alors le tableau se réorganiser au cours de ce "*zoom-arrière*". Cette réorganisation du tableau conduit à la découverte des rapports de structure, tels que les définit le philosophe Ernst Cassirer. Il écrit :

> "*L'imitation ne consiste jamais à redessiner, trait pour trait, un contenu de réalité, mais ...à tracer les contours caractéristiques de sa silhouette...En ce sens, reproduire un objet ne consiste pas simplement à rassembler les caractères singuliers de sa forme, mais à en saisir les rapports de structure...*".

Cassirer incorporait les problèmes posés par la perception à ceux, tout aussi redoutables, posés par le langage. Il écrit :

> *"Le chaos des impressions immédiates ne s'éclaircit et ne s'articule pour nous que parce que nous le "nommons" et le pénétrons ainsi par la fonction de la pensée linguistique et de l'expression linguistique... Le langage devient ainsi un des moyens fondamentaux de l'esprit, grâce auquel s'accomplit le progrès qui nous fait passer du monde des simples sensations à celui de l'intuition et de la représentation...C'est ici l'origine de cette fonction universelle de division et de liaison qui trouve son plus haut degré d'expression consciente dans les analyses et les synthèses de la pensée scientifique."*

Les peintres et les philosophes nous tiennent en fait le même langage. La perception d'une image ne nous fournit pas son reflet, sa copie conforme, pas plus que les mots ne peuvent être identifiés aux objets qu'ils désignent. Au contraire et dans les deux cas, le passage de l'objet à la perception que nous en avons (ou au nom qui le désigne) est le résultat d'une opération intellectuelle (Ernst Cassirer dirait "spirituelle") complexe, dépendant d'une modélisation appropriée du monde extérieur. Cette opération intellectuelle complexe ressemble-t-elle aux algorithmes utilisés en traitement de l'image ? Peut-on en savoir un peu plus sur les modèles implicites du monde qui nous entoure, modèles qui fondent le langage ou la perception ? Pour tenter de répondre à ces questions, nous utiliserons les acquis de la neuro-physiologie et les découvertes qui ont valu le Prix Nobel à David Hubel et Torsten Wiesel.

4. Le cortex visuel primaire

Voici ce que David Hubel écrit sur le fonctionnement du cortex visuel primaire et de la perception :

> *Ramón y Cajal fut le premier à comprendre que les connections dans le cortex sont très courtes... Quel que soit le traitement effectué par le cortex, il reste certainement local : l'information concernant une petite région de l'environnement visuel atteint une petite région du cortex, où elle est transformée, analysée, digérée (utilisez l'expression que vous préférez), puis envoyée dans une autre région corticale où elle subit un autre type de traitement, indépendant du traitement effectué dans la région voisine. L'environnement visuel est ainsi analysé, fragment*

> *par fragment, dans le cortex visuel primaire : par conséquent, celui-ci n'est pas l'endroit du cerveau où les objets entiers (bateaux, chapeaux, visages, etc.) sont reconnus, perçus ou traité ; le cortex visuel primaire n'est pas le centre de la "perception".*

Cette description de l'activité du cortex visuel primaire exclut une modélisation du type "transformée de Fourier", car cette dernière est une transformation globale, opérant sur toute l'image. En juin 1997, à l'Université Autonoma de Madrid, David Hubel nous expliquait que l'un des buts de ses recherches était de comprendre comment l'information visuelle, captée par l'œil, est ensuite traitée par le cortex visuel primaire. Il s'agit ici d'un pré-traitement qui ne met pas encore en jeu les fonctions cognitives du cerveau. David Hubel, Torsten Wiesel et Margaret Livingstone ont montré que certaines cellules du cortex visuel primaire sont affectées à des tâches infiniment modestes, parcellaires, répétitives, un peu comme un travail à la chaîne. Ces neurones ne procèdent pas à un découpage de l'image en blocs 8×8, comme le ferait l'algorithme JPEG, mais détectent des patterns, des structures universelles et rudimentaires qui se retrouvent dans toutes les images. Ces patterns ne se réduisent pas à des petits morceaux de l'image mais en fournissent une sorte de croquis ou de sketch. Nous reviendrons sur cette notion de sketch et sur sa modélisation.

Par exemple, certaines cellules sont responsables de la détection des contours et il est tout à fait étonnant que différentes orientations soient prises en charge par différentes cellules, chacune étant spécialisée dans une orientation particulière. D'autres neurones sont spécialisés dans la détection de motifs périodiques etc. Tous ces neurones font partie des aires primaires du cerveau. Aucun ne fournit la compréhension de l'ensemble de l'image. Cette compréhension fera appel à des processus cognitifs mettant en jeu des populations de neurones. L'information "prétraitée" fournie par les neurones du cortex visuel primaire est, en effet, envoyée vers les aires secondaires ou associatives, responsables des processus cognitifs et y subit un "post-traitement". Comme l'écrit Bernard Mazoyer :

> *"Les fonctions cognitives sont basées sur la mise en jeu d'un réseau distribué d'aires corticales possédant une dynamique temporelle".*

D. Hubel s'est alors demandé si le cortex visuel primaire était déjà câblé à la naissance ou si le câblage s'élaborait dans les premiers mois suivant la naissance, grâce aux stimuli visuels que reçoit le bébé. L'air du temps était en faveur de l'apprentissage et du conditionnement. Le cerveau du bébé était vu comme une page blanche sur laquelle la vie allait écrire et organiser de façon cohérente notre aptitude à comprendre ce que nous voyons. La cohérence de notre cerveau serait un reflet de celle du monde qui nous entoure. Mais D. Hubel a découvert que c'était l'inverse qui a lieu dans certaines fonctions. Cette découverte a bouleversé les préjugés sur l'apprentissage. Certaines parties du cerveau sont précâblées et ce câblage peut s'effacer, faute de simulations.

D. Hubel a travaillé sur des chats et des singes. En étudiant la période critique d'élaboration du fonctionnement de la vision, il a découvert un moyen de guérison d'une forme de cécité qui s'appelle l'amblyopie. Cette forme de cécité vient du fait que, dans certains cas, une partie du cortex visuel primaire n'est pas stimulée, car elle ne reçoit pas l'information de l'œil (à cause d'un fort strabisme, par ailleurs temporaire, qui affecte certains bébés). Les neurones dégénèrent alors de manière irrécupérable, et le bébé devient aveugle. Le problème n'est donc pas celui d'un défaut dans le pré-câblage. C'est le contraire qui est vrai : les neurones, qui existent dès la naissance, ont besoin d'être stimulés pour survivre.

Le travail de David Hubel illustre le lien entre recherche pure et recherche appliquée : D. Hubel dit, en effet, qu'il n'a pas cherché à guérir une forme de cécité et que sa découverte sur la façon de traiter l'amblyopie est un produit inattendu de son étude du fonctionnement du cerveau.

Il ajoute, de façon très ironique, que si son programme de recherche avait porté sur l'étude de cette forme de cécité, il n'aurait jamais obtenu la moindre subvention. On lui aurait enjoint de travailler, comme cela semblait évident, sur l'œil et cela n'aurait conduit à rien puisque le problème se situe au niveau du cerveau ! Citons Hubel :

> *"During the past forty years, in collaboration with Torsten Wiesel and later with Margaret Livingstone, I have attempted to build upon the work of Ramón y Cajal in an attempt to obtain a detailed understanding of the physiology of one small*

part of the cerebral cortex -the striate cortex, or primary visual cortex. Since our work has focused on understanding the human brain, we have confined our studies to higher mammals, especially macaque monkeys. Our experiments have aimed at understanding brain function in normal adult cats and monkeys, and in newborn animals.

In adult cats and monkeys, our main effort has been to determine how the visual information coming from the eye is handled and transformed by the brain. In 1950, Stephen Kuffler established that nerve fibers from the eye, the optic nerve fibers, respond maximally to a circular spot of light on the retina or in the outside world. For maximum impulse discharge rates the position of the spot and its size must be precisely specified. A typicall retina cell responds best to a region in the visual field that subtends much less than one degree (termed the receptive field center), and, as Kuffler discovered, a spot larger than optimal produces much weaker responses because of inhibition coming from from the area surrounding the field center.

Cells in the primary visual cortex, to which the optic nerve projects (with one intermediate nucleus interposed) are far more exacting in their stimulus requirements. The commonest type of cell fires most vigorously not to a circular spot, but to a short line segment -to a dark line, a bright line, or to an edge boundary between dark and light. Furthermore each cell is influenced in its firing only by a restricted range of line orientations: a line more than about 15 to 30 degrees from the optimum generally evokes no response. Different cells prefer different orientations, and no one orientation, vertical, horizontal or oblique, is represented more than any other. These observations made in 1958, had not been predicted and came as a complete surprise. Evidentlty cells in this part of the cortex are determining whether there are contours (light-dark or color) in the visual scene, and collectively registrating their orientations.

We went on to show that cells could be subdivided into classes according to a hierarchy of increasing complexity, and could be further categorized in terms of specific responses to movement or to color. For example, a cell in the left hemisphere might respond to a bright red line oriented at 45 degrees to the horizontal, in a small region of the right visual field, but fail to repond to vertical lines, horizontal lines, or white or black lines. One of the most striking examples of cortical specificity is observed in the visual failure of these cells to respond to marked changes in diffuse light intensity, as for example when a bright flashlight is directed into the eyes of the animal. This failure of cells to respond to changes in diffuse light was one reason why investigators in the early 1950s thought that cortical cells were unresponsive to visual stimuli.

It soon became apparent that cells were grouped in the cortex according to orientation preference, in parallel column-like aggregations extending through the full 2mm thickness of cortex. We term these groupings "orientation columns". The arrangement is one of extreme order: a movement parallel to the cortical surface of a distance of 1/20 mm corresponds to a change in orientation preference of cells by about 10 degrees, so that a one-millimeter traverse along the cortex corresponds to a full 180 degrees rotation.

In monkey cortex a typical cell responds to the two eyes, though usually better to one eye than to the other. Cells preferring a given eye, left or right, tend also to be grouped together, in alternating left-eye, right-eye slabs 1/2 mm thick, lying perpendicular to the surface. These eye-dominant columns have little or no relationship to the orientation columns. They have been clearly shown to exist in human primary visual cortex. We indeed surmise that most of our findings apply to humans, since visual cortex of cat and monkey (and all higher mammals so far studied) are anatomically and physiologically similar except for minor details.

These columnar formations are closely analogous to the columns that had perviously been demonstrated by Vernon Mountcastle in somatosensory cortex. Columns may indeed be a universal property of cerebral cortex. They have now been demonstrated in dozen of other cortical areas.

At an early stage of our research Torsten Wiesel and I began to ask whether the visual cortex was fully connected at birth, or whether the organization we had seen in adult animals required postnatal experience for its development. It is fair to say that the prevailing opinion of psychologists in the fifties and sixties was that the wiring of the brain and certainly of the cerebral cortex, was largely a product of post-natal experience.

We began by showing that a newborn animal that had one eye covered for a period as short as a few weeks became blind in that eye. We could show that the blindness was not the result of some abnormality in the eye, but of miswiring in the cortex, since the responses from the retina were normal, whereas stimulating the eye that had been covered evoked no reponse from the cortex. These effects of eye occlusion could be produced only during a critical period of several months subsequent to birth. The period of plasticity in this system was therefore strictly limited. Clinical experience, for example with congenital cataracts, suggests that in humans the critical period extends for several years from birth. Surprisingly, blindfolding both eyes in young monkeys produced far less severe effects, even though the animals became blind in both eyes. Presumably the blindness is due in part at least to effects on higher visual areas. This difference in the effects of monocular and binocular occlusion must have been produced by competition, rather than by simple disuse.

Perhaps the most surprising result of these studies was the demonstration that at birth, this region of cortex responds very much like adult cortex. Orientation selectivity, for example, is present to a striking degree, and the orientation columns have begun to develop. We conclude that the deprivation effects that we had seen were the result of deterioration of connections *present in large part at birth, rather than simply a failure to adquire connections as a result of lack of a learning experience. Specific cortical connections, at least for this region of cortex, are thus partly innate and partly the result of experience, and not, as the previous generation of psychologists had thought, exclusively the result of experience.*

This work in young cats and monkeys, especially the concept of the critical period, led to a means of prevention of one of the commonest forms of blindness, the monocular blindness (amblyopia) that result results from strabismus ("cross eye" or "wall eye"). By operating surgically at an early age to correct the condition it is possible to preserve vision in the two eyes, and even prevent the loss of binocular depth perception that almost always results from strabismus. As often happens in science, this beneficial result of our research was quite unforseen. We did not set out to prevent a specific form of blindness; the discoveries came as a by-product of studying the normal brain. Ironically, had we set at the very beginning to understand amblyopia, we would have begun by working on the eye, not he brain, and would not have gotten anywhere. I suspect no funding agency would have supported a proposal to study amblyopia by making microlectrode recordings from the cortex. The deprivation studies began as a side project and did not require any additional resources.

Tout ceci nous oriente vers un modèle de l'image basé principalement sur la notion de contour, puisque, selon David Hubel, les contours sont essentiels à la vision. Comme nous l'avons déjà indiqué, ce modèle revient à représenter une image par un ensemble de courbes que l'on puisse dessiner. Nous arrivons ici aux *level-sets* de Stanley Osher ou aux modèles basés sur les fonctions à variation bornée (modèle d'Osher, L. Rudin et E. Fatemi). Ces deux premiers modèles mettent l'accent sur la géométrie. Un troisième type de modèles est plus analytique et consistera à rechercher des "briques de base" ou "building blocks" dans une image ou une collection d'images. Ce dernier modèle qui s'inspire du meccano nous conduira aux ondelettes. Nous savons aujourd'hui, grâce aux travaux d'Albert Cohen, d'Ingrid Daubechies et de leurs collaborateurs que les ondelettes qui vont être définies dans la section suivante constituent la *meilleure base* permettant de décrire ou représenter les fonctions à variations bornées. En outre les ondelettes incorporent le *zoom à travers les échelles.*

5. Les ondelettes

Le rôle des ondelettes en traitement de l'image a été fort débattu pendant ces vingt dernières années. Aujourd'hui la pertinence de l'analyse par ondelettes des images est établie grâce à de très nombreux travaux. On peut citer, en particulier, ceux d'Albert Cohen et de son équipe. Nous y reviendrons dans les sections 7, 8 et 9. Mais avant d'aller plus loin, commençons par faire un petit détour et rappeler ce que sont l'analyse et la synthèse. L'analyse a pour finalité de dissoudre la complexité. L'hypothèse de travail est que cette complexité n'est qu'apparente et est due à un mélange. L'analyse recherche les éléments simples qui ont été mélangés. La première question qui se posera concerne le caractère universel des éléments simples que l'on découvre. En chimie et en physique, la recherche des éléments simples a été d'une remarquable fécondité scientifique. L'analyse a successivement fourni les corps purs, les molécules, les atomes et enfin les particules élémentaires. Mais en traitement du signal ou en traitement de l'image, la situation est beaucoup plus confuse. Dans le cas du signal de parole, on aboutit aux phonèmes qui dépendent de la langue utilisée. La musique utilise les notes. Voici quatre questions concernant l'analyse des images :

(a) Comment cherche-t-on les éléments simples dans des stocks d'images ?
(b) Trouve-t-on les mêmes éléments simples dans tous les stocks d'images ?
(c) A-t-on trouvé suffisamment d'éléments simples pour rendre compte de la complexité des images analysées ?
(d) Les "éléments simples" obtenus par une analyse automatisée portant sur un corpus d'images digitales sont-ils en accord avec les fonctions du cortex visuel primaire ?

Nous traiterons ces problèmes dans la section suivante. Nous utiliserons alors une des techniques les plus objectives d'analyse qui est *l'analyse en composantes indépendantes.* Nous nous inspirerons des travaux d'un groupe de chercheurs autour de D. J. Field, B. A. Olshausen et de Naoki Saito (University of California, Davis, CA).

Mais avant de nous engager dans cette voie, nous commencerons par tricher avec notre programme et par imposer une fois pour toutes le choix des briques de base, un peu comme l'on fait pour les lettres de l'alphabet. Ces briques de base sont les ondelettes.

L'analyse par ondelettes est née, à la fin des années 70, comme une alternative à l'analyse de Fourier. L'analyse de Fourier est optimale dans l'analyse des signaux stationnaires, c'est-à-dire des signaux dont les propriétés statistiques ne changent pas au cours du temps. Mais l'analyse de Fourier donne de très mauvais résultats dans l'analyse des signaux transitoires, c'est-à-dire dont le déroulement temporel fait surgir des propriétés totalement imprévues. C'est ce que découvrit Jean Morlet qui proposa les algorithmes dont les variantes sont aujourd'hui utilisées dans le monde entier. Cette découverte faite par un ingénieur de terrain fut comprise et acceptée par un physicien, Alexandre Grossmann, puis par des spécialistes du traitement du signal. Vingt ans après, l'analyse par ondelettes débouchait sur le nouveau standard, nommé JPEG2000, de compression des images fixes. Mais revenons au début. Ancien élève de l'Ecole Polytechnique, Morlet était ingénieur de recherche chez Elf-Aquitaine. Quand il découvrit les ondelettes, Morlet travaillait depuis déjà une vingtaine d'années dans le secteur de la vibrosismique. Morlet créa l'analyse par ondelettes pour surmonter certaines difficultés rencontrées dans l'analyse des signaux acquis lors des campagnes pétrolières.

Autrefois, pour chercher du pétrole, on faisait exploser des charges et les échos recueillis permettaient d'estimer la position, la profondeur et la forme de la cavité contenant l'or noir. Les experts engagés par les compagnies pétrolières, les sourciers, étaient alors des physiciens. Analyser les bruits répercutés par le sous-sol, c'était imiter le savoir-faire du médecin qui, à l'aide du stéthoscope, ausculte le malade en écoutant sa respiration ou les battements de son cœur.

Pierre Goupillaud, collègue et ami de Jean Morlet, était au départ un ingénieur français. Il s'expatria aux USA et travailla pour la compagnie pétrolière Conoco, (aujourd'hui ConocoPhillips) dans le secteur de la géophysique. Goupillaud suggéra d'envoyer dans le sous-sol une vibration, courte et modulée en fréquence, au lieu de faire exploser des charges. L'énergie dépensée et les dégâts occasionnés sont alors réduits. Ce même principe est utilisé par le sonar de la chauve-souris. La vibrosismique était née. Mais les échos recueillis sont bien plus complexes à analyser que dans le cas des explosions de charges. Les physiciens durent céder la place à des spécialistes du traitement du signal. Ces derniers élaborèrent des logiciels informatiques qui, en un sens, imitent le fonctionnement du cerveau de la chauve-souris. Grâce à la vibrosismique, Elf-Aquitaine a pu mener une campagne pétrolière à Paris même. Les camions-vibrateurs ont sillonné les artères parisiennes pendant une quinzaine de jours, au milieu de nuits d'hiver de l'année 1986. L'exploitation des résultats a demandé une année entière. Cela donne une idée des difficultés rencontrées dans la vibrosismique.

Jean Morlet analysait donc les signaux provenant de la vibrosismique. Ces signaux sont des courbes graphiques assez irrégulières qui présentent de fortes parties transitoires (ce terme sera défini plus tard). Jean Morlet étudiait ces courbes à l'aide d'une technique éprouvée, l'analyse de Fourier à fenêtre (en fait à l'aide des ondelettes de Gabor, nous y reviendrons). Mais un jour, Morlet, lassé des artefacts (erreurs systématiques) entraînés par cette technique, découvrit une nouvelle façon de représenter ce type de signaux. C'est ainsi que Morlet créa l'analyse par ondelettes. L'idée de base de Morlet est d'incorporer l'aspect transitoire dans la définition même des ondelettes. On part d'une fonction $\psi(x)$ de la variable réelle x (le temps en traitement du signal) qui est localisée, régulière et d'intégrale nulle, par exemple

$$\psi(x) = (1 - x^2)\exp(-x^2/2). \tag{1}$$

Les ondelettes alors sont les fonctions $\psi_{a,b}$ qui sont déduites de l'ondelette mère ψ par translation (par b) ou dilatation (par a). On a donc

$$\psi_{a,b}(x) = a^{-1/2}\psi(\frac{x-b}{a}),\ a > 0,\ b \in \mathbb{R}. \tag{2}$$

Ces fonctions $\psi_{a,b}$ sont donc indexées par deux paramètres réels a et b. L'analyse s'effectue en calculant la transformée en ondelettes définie par :

$$W(a,b) = \int f(x)\overline{\psi_{a,b}(x)}\,dx. \tag{3}$$

La synthèse s'opère comme si les ondelettes formaient une base orthonormée. On a donc

$$f(x) = c \int_0^\infty \int_{\mathbb{R}} W(a,b)\psi_{a,b}(x)\, db\, \frac{da}{a^2} \tag{4}$$

où c est une constante positive. Les ondelettes $\psi_{a,b}(x)$ sont souvent appelées les ondelettes "temps-échelle" tandis que les ondelettes de Gabor sont appelées "temps-fréquence". L'analyse par ondelettes "temps-échelle" permet de détecter les changements brutaux dans le déroulement de $f(x)$. En effet les petites valeurs du paramètre d'échelle a permettent de cerner (avec une erreur ne dépassant pas Ca) la position d'une discontinuité, car les coefficients d'ondelette prennent de grandes valeurs aux petites échelles. Marie Farge parle d'un microscope mathématique. Bien entendu, pour que l'analyse par ondelettes conduise à un algorithme rapide, il conviendra de pourvoir remplacer les valeurs continues des paramètres a et b par des valeurs discrètes. Morlet suggérait de prendre $a = \alpha 2^j, b = k\beta 2^j$ où j et k sont des entiers relatifs, les pas d'échantillonnage α et β étant positifs. On peut faire beaucoup mieux. Comme je l'ai établi, on peut construire une fonction ψ appartenant à la classe de Schwartz de sorte que les fonctions

$$2^{j/2}\psi(2^j x - k),\ \ j,k \in \mathbb{Z} \tag{5}$$

consituent une base orthonormée de $L^2(\mathbb{R})$ composée d'ondelettes.

J'ai souvent discuté avec Jean Morlet. Il ressemblait beaucoup à Benoît Mandelbrot. Tout comme Mandelbrot, Morlet avait une extraordinaire intuition et une réelle vision scientifique. Il a tout de suite compris la portée de sa découverte et a essayé d'alerter Elf-Aquitaine. Mais Elf-Aquitaine venait d'être la victime consentante d'une énorme escroquerie; un malfaiteur belge était arrivé à persuader les têtes pensantes de l'entreprise que l'on pouvait "flairer le pétrole" à l'aide des trop célèbres "avions renifleurs". Pour bluffer les "têtes pensantes" et autres "décideurs" d'ELF, l'escroc présentait, dans un certain ordre, des objets dans une pièce. Ces objets étaient "reniflés" par un miraculeux "gadget", situé dans une autre pièce. Ce gadget reconstruisait, en temps réel, les images des objets sur un écran d'ordinateur. Les décideurs d'ELF étaient médusés. L'escroquerie fut révélée par Jules Horowitz, membre de l'Institut, qui eut l'idée d'inverser l'ordre de passage de deux des objets présentés au "nez" du gadget. Comme tout était pré-enregistré, les images défilèrent évidemment dans l'ordre ancien. Mais c'était trop tard et l'argent d'ELF avait disparu.

Passant de l'extrême crédulité à une extrême méfiance, Elf-Aquitaine répondit à la découverte de Jean Morlet en lui octroyant une retraite anticipée. Plus de dix ans après cette mise à la retraite, Morlet obtint le prix Reginald Fessenden de la Société Américaine de Géophysique. Lors de la cérémonie, Pierre Goupillaud présenta l'œuvre de Morlet et dit :

> *A product of the renowned Ecole Polytechnique, Morlet performed the exceptional feat of discovering a novel mathematical tool which has made the Fourier transform obsolete after 200 years of uses and abuses, particularly in its fast version... Until now, his only reward for years of perseverance and creativity in producing this extraordinary tool was an early retirement from ELF.*

Roger Balian qui enseignait la physique à l'Ecole Polytechnique orienta Jean Morlet vers Alex Grossmann. Alex Grossmann, directeur de recherches au CNRS, travaillait à Marseille-Luminy, au centre de physique théorique. Alex Grossmann fut patient, subtil et comprit ce que Morlet avait dans l'esprit. Grâce à la clairvoyance de Grossmann, les résultats de Morlet ont pu être publiés en 1984. Ecouter Morlet n'était certainement pas une tâche aisée, tant ses idées étaient originales, allusives, approximatives et souvent exagérément optimistes. J'en parle d'expérience. Morlet pensait, par exemple, que l'analyse par ondelettes allait tout de suite révolutionner la vibrosismique et la prospection pétrolière. Quelque chose d'autre s'est produit : les ondelettes ne servent qu'à *comprimer et transmettre* les données recueillies dans les campagnes pétrolières. L'analyse de ces données est une tout autre histoire. En outre les ondelettes servent à *comprimer et transmettre* les images qui sont bi-dimensionnelles. Voici pourquoi. L'éclairement d'une image présente de très fortes discontinuités ; il s'agit des bords des objets où l'éclairement change très brutalement. Cela explique que l'analyse par ondelettes, adaptée aux signaux transitoires, convienne si bien aux images.

La révolution numérique repose sur l'utilisation d'algorithmes rapides et efficaces. Grossmann et Morlet ont donc proposé un algorithme approché permettant de calculer la transformée en ondelettes. Cet algorithme consistait à remplacer par des approximations les intégrales doubles (4) dont le coût de calcul est élevé. Mais ces approximations sont imprécises et engendrèrent des erreurs incontrôlables.

En revanche la transformation de Fourier peut se calculer par un algorithme rapide, dénommé *Fast Fourier Transform* ou FFT. C'est un algorithme exact. Il a été découvert, en 1965, aux Etats-Unis, par James W. Cooley et John W. Tukey. Sans la FFT le calcul d'une transformation de Fourier serait prohibitif. Il faudrait N^2 opérations pour un signal de longueur N. Avec la FFT ceci se réduit à $2N \log_2 N$. Plus concrètement cela revient à comparer un temps de calcul qui ne prend qu'une seconde à un temps de calcul qui prendrait plusieurs semaines. Sans la possibilité qu'offre la FFT de calculer en temps réel, l'imagerie médicale ou la biologie moléculaire n'auraient pas vu le jour.

En revenant à l'analyse par ondelettes, un long chemin restait à parcourir et c'est grâce aux travaux d'Ingrid Daubechies et de Stéphane Mallat que la Fast Wavelet Transform a finalement pu de hisser au niveau atteint par la FFT. Le calcul de la FWT d'un signal de longueur N est exact et ne nécessite que CN opérations (C est une constante que nous retrouverons dans ce qui suit).

La construction de la FWT bénéficiait de deux découvertes antérieures : les *algorithmes pyramidaux* et le *codage en sous-bandes.* Les *algorithmes pyramidaux* furent découverts par P. Burt et E. H. Adelson (1983). L'analyse par ondelettes d'une image devient, dans les travaux de Mallat, un cas particulier d'*algorithmes pyramidaux.* Dans cette perspective, analyser une image par ondelettes revient à étudier cette image en prenant du recul. Le recul permet de comparer mentalement différentes vues à différentes échelles. La différence entre ces vues est précisément ce que fournit l'analyse par ondelettes. La seconde découverte utilisée par Mallat est le *codage en sous-bandes* ou *subband coding.* Ce codage avait été inventé en 1977, au centre d'IBM, à La Gaude par D. Esteban et C. Galand. La motivation était le téléphone digital. L'importance de la découverte du codage en sous-bande n'a pas été perçue par IBM. La thèse de Mallat fut le point de départ du standard JPEG2000 de compression des images fixes.

Puis vint la construction, par Ingrid Daubechies (1987), des bases orthonormées d'ondelettes à support compact, de régularité r donnée. On se reportera à (5). Cette régularité peut être aussi élevée que l'on veut, mais la base choisie dépend alors de r. La longueur du support de l'ondelette est la constante C intervenant dans les algorithmes pyramidaux. Le seul cas connu était celui du système de Haar (1909). Les ondelettes à support compact conduisent à des algorithmes qui travaillent en temps réel, alors même que le signal défile. Le calcul se fait sur une *fenêtre mobile*, ne mettant en jeu qu'un nombre fixe, limité de valeurs du signal. Dans le cas du système de Haar, l'algorithme calcule la demi-somme et la demi-différence entre deux valeurs consécutives. Mais le manque de régularité ($r = 0$) du système de Haar excluait toute application à la compression des images fixes. L'année suivante, en collaboration avec Albert Cohen et Feauveau, I. Daubechies construisit les ondelettes bi-orthogonales. Ce sont elles qui seront utilisées dans JPEG2000.

6. La décomposition en composantes indépendantes, ou ICA

Certains neurophysiologistes se sont intéressés aux représentations efficaces des images et en particulier des images de scènes naturelles. Une image brute provenant du monde extérieur contient une information gigantesque et il est exclu qu'elle soit intégralement perçue par le système visuel animal ou humain. L'hypothèse de travail de ces neurophysiologistes est la suivante : si les cellules du cortex visuel primaire sont affectées à des tâches spécifiques de reconnaissance de certaines structures géométriques, c'est peut-être parce que cette solution biologique au problème de la lecture d'une image est optimale en terme de d'analyse et de compression des données. De même que la géométrie de l'os est optimale en terme de poids (faible) et de solidité (forte) grâce à sa texture singulière. On peut penser qu'à la suite d'un lent processus d'évolution, la sélection naturelle ait conduit à cette spécialisation croissante des cellules rétiniennes.

Peut-on établir scientifiquement l'assertion suivante : *en optimisant l'analyse et la compression des images des scènes naturelles, on retrouve précisément les algorithmes élémentaires qui correspondent aux diverses spécialisations des neurones du cortex visuel primaire ?*

On peut aujourd'hui étudier cette question fondamentale en faisant appel à *l'analyse en composantes indépendantes.* Cette technique d'analyse est basée sur le concept intuitif de "contraste". L'hypothèse de travail est que, pour extraire une information pertinente d'un ensemble riche de données complexes et non structurées, il faut optimiser le contraste, c'est à dire disposer de différents points de vue, à partir de directions les plus éloignées les unes des autres. Cela afin d'obtenir les informations les plus contrastées possibles les unes par rapport aux autres.

Voici un exemple. Supposons que l'on ne dispose que de cinq photographies pour comprendre la structure d'un objet tridimensionnel. Alors il serait maladroit de prendre cinq vues trop voisines, en ne se déplaçant que légèrement autour de l'objet. Il vaut mieux tourner carrément autour de l'objet et choisir des angles de vue les plus différents possible.

Un niveau plus profond de compréhension des images pourrait-il être obtenu en suivant cette métaphore ? Peut-on, comme le firent D.J. Field, B.A. Olshausen et Naoki Saito, relier l'analyse des images effectuée par le cortex visuel primaire à l'analyse en composantes indépendantes ?

L'analyse en composantes indépendantes (ICA ou *independant component analysis*) est un programme scientifique excitant pour les uns, décevant pour les autres. Nous dirons pourquoi. L'analyse en composantes indépendantes est née des problèmes de la séparation de sources et de la "déconvolution aveugle" qui se rencontrent, par exemple, dans l'utilisation des radars. On peut citer (en vrac) les noms de Bernard Picinbono, Christian Jutten, Odile Macchi, Jean-François Cardoso et Jean-Louis Lacoume, sans oublier David Donoho. Voici le point de départ : un signal observé nous semble inintelligible, mais, en fait, cette complexité provient de ce que ce signal est un mélange, une combinaison linéaire entre plusieurs signaux "sources", supposés statistiquement indépendants. Le mélange a créé la complexité apparente et a brouillé les sources. Peut-on extraire ces signaux sources à partir du signal donné ? Voici deux exemples. Dans le premier, on dispose d'un enregistrement numérique d'une sonate pour piano et violon. Peut-on algorithmiquement extraire la partie jouée par le violon ? Cet exemple ne convient pas dans la mesure où le piano et le violon ne jouent pas de façon indépendante. Le second exemple est plus proche des conditions requises pour faire une ICA. C'est celui de la "party", élément de la convivialité aux USA où une vingtaine de personnes, debout, parlent en buvant des vins ou des apéritifs. Les conversations se mêlent sans que cela vous empêche de comprendre la personne à laquelle vous vous adressez. Le cerveau effectue une analyse en composantes indépendantes. L'ICA repose sur une hypothèse très forte, à savoir que le signal observé est une combinaison linéaire de signaux statistiquement indépendants (au sens fort, c'est-à-dire au sens des moments d'ordres supérieurs). Ceci est très rare. C'est pourquoi l'ICA est décevante pour certains. La beauté de cette analyse non linéaire est d'être effective et de donner une réponse essentiellement unique avec si peu d'information. Je n'en dirai pas plus. La meilleure référence que je connaisse sur l'ICA est le site web de Jean-François Cardoso. Une première initiation à ces méthodes m'a été donnée par Jean-Louis Lacoume.

Nous nous proposons maintenant, en suivant D. J. Field, B. A. Olshausen et Naoki Saito, d'utiliser ce concept pour essayer de comprendre la façon dont fonctionne la vision humaine.

Ces chercheurs ont appliqué un algorithme d'analyse en composantes indépendantes à une base de données d'images naturelles. A leur plus grande surprise, les fonctions de base obtenues sont des ondelettes ! Nous ne faisons pas ici de distinction entre ondelettes de Gabor (des gaussiennes d'écart type σ, translatées en position, puis modulées arbitrairement en fréquence) et ondelettes de Grossmann-Morlet (une fonction $\psi(x)$ régulière et localisée, d'intégrale nulle et ayant une certain nombre de moments nuls, que l'on dilate arbitrairement, puis translate arbitrairement). Les ondelettes de Gabor sont indexées par deux paramètres ω et x_0 appartenant à $\mathbb{R}^2$ et sont définies par

$$w_{(\omega,x_0)}(x) = \exp(i\omega \cdot x) g_\sigma(x - x_0), \quad \omega \in \mathbb{R}^2, x_0 \in \mathbb{R}^2 \tag{6}$$

où $g_\sigma(x)$ est une gaussienne d'écart-type σ.

En revanche, les ondelettes de Grossmann-Morlet bidimensionnelles (dites aussi "temps-échelle") sont indexées par un paramètre positif a et un vecteur $x_0 \in \mathbb{R}^2$ et sont définies par

$$\psi_{(a,x_0)}(x) = a^{-1}\psi(\frac{x - x_0}{a}), \quad a \in (0, \infty), x_0 \in \mathbb{R}^2 \tag{7}$$

et la fonction ψ (l'ondelette-mère) doit tout d'abord satisfaire les trois propriétés essentielles qui ont déjà été évoquées : être localisée (un support compact est idéal), être oscillante (l'intégrale de ψ doit être nulle, mais on a souvent besoin de plusieurs autres moments nuls) et enfin être la plus régulière possible, compte tenu des deux conditions précédentes. Mais en outre l'ondelette mère doit être isotrope au sens suivant : pour tout vecteur unitaire $\nu \in \mathbb{R}^2$, la transformée de Fourier $\hat{\psi}$ de ψ doit vérifier

$$\int_0^\infty |\hat{\psi}(t\nu)|^2 \frac{dt}{t} = 1 \tag{8}$$

Cette condition d'isotropie est le point faible de l'analyse par ondelettes "temps-échelle" bidimensionnelles. Elle est masquée en dimension 1, car il suffit que l'ondelette mère soit réelle et d'intégrale nulle pour que cette condition soit satisfaite.

Or ces "ondelettes" sont précisément les motifs (ou "primitives" selon la terminologie de David Marr) qui sont détectés de façon privilégiée par les neurones du cortex visuel primaire. L'analyse en composantes indépendantes rendrait donc compte de l'analyse si particulière de l'image effectuée par les neurones du cortex visuel primaire. Rappelons les titres "Emergence of simple-cell receptive field properties by learning a sparse code for natural images" ou bien "Sparse coding : A strategy employed by V1 ?" des travaux de D.J. Field et B.A. Olshausen. Ces titres sont en eux-mêmes des programmes scientifiques.

7. Le modèle d'Osher, Rudin et Fatemi

Une image en noir et blanc est vue comme une fonction $f(x)$ définie dans le plan et dont les valeurs sont comprises entre 0 et 1. Ici 0 correspond au noir tandis que 1 est un blanc fortement éclairé. Par ailleurs cette fonction $f(x)$ peut (et doit) présenter de fortes discontinuités qui correspondent aux bords des objets. Pour nous autres mathématiciens, une telle fonction est mesurable et bornée. Bien entendu la réciproque n'est pas vraie : toute fonction mesurable et bornée ne correspond pas à une image. Une quantité vraiment minuscule de telles fonctions provient d'images et le but de la modélisation consiste à en savoir un peu plus. A cet effet, on utilise l'espace des fonctions à variation bornée dont la définition a été trouvée, en 1925, par le mathématicien italien Leonida

Tonneli. Il s'agissait de résoudre un problème posé par Lebesgue : savoir quand la surface du graphe Γ d'une fonction continue f définie sur le carré unité est finie. Un problème contigu est la définition de cette surface. Mais ce dernier problème avait été déjà résolu par Lebesgue. Tonneli découvrit que le graphe de f a une surface finie si et seulement si f est à variation bornée. Encore faut-il définir cette dernière propriété. Oublions le carré et considérons le cas, conceptuellement plus simple, d'une fonction définie dans le plan. Elle est à variation bornée si et seulement s'il existe une contante C telle que pour tout vecteur y on ait $\int\int |f(x+y) - f(x)|\, dx_1 dx_2 \leq C|y|$. La norme de f dans BV est, par définition, $||f||_{BV} = \int\int |\nabla f(x)|\, dx_1 dx_2$ et est équivalente à la borne inférieure des constantes C figurant dans la première définition. Ici $\nabla f(x)$ est le gradient de f. Si f est la fonction indicatrice d'un ensemble E, alors la norme de f dans BV est la longueur du bord de E, aussi notée le périmètre de E. Ceci est vrai si E est régulier et demande à être précisé dans le cas d'un ensemble E arbitraire. Le bord de E doit être remplacé par le bord distingué de E dont la définition est due à Ennio de Giorgi. Le modèle d'Osher, Rudin et Fatemi consiste à définir une image f comme une somme $f = u + v$ entre une composante $u \in BV$ qui modélisera les objets contenus dans l'image et une composante v qui modélisera la texture et le bruit. La composante v est tout simplement de carré intégrable, car on ne peut plus trouver de structure géométrique dans v. Bien entendu, il y a une infinité de telles décompositions et l'on décide de choisir celle qui minimise la fonctionnelle $||u||_{BV} + \lambda||v||_2^2$. Le paramètre positif λ sert à décider à partir de quelle taille les "petits" objets sont regardés comme des éléments de textures. Le modèle d'Osher, Rudin et Fatemi est la traduction en langage mathématique de l'art du dessinateur. L'essentiel dans une image peut être dessiné. Ceci fournit la composante u dans le modèle. Ce qui ne peut être dessiné (les cheveux, par exemple) constitue v et est beaucoup plus irrégulier que u.

8. L'analyse multi-échelle des images

Les philosophes et les peintres nous disent que pour analyser une image ou un tableau dans un musée, il faut prendre du recul, en étant alors sensible à la façon dont notre perception se réorganise dans ce "zoom arrière". En termes mathématiques, on a le choix entre deux options. La première consiste à imiter la simplification progressive fournie par le recul. Cela revient à lisser progressivement l'image donnée f. Pour cela, on remplace f par une suite de convolutions $f_j = f \star \phi_j$, $j \in \mathbb{N}$, où ϕ est une fonction suffisamment régulière, suffisamment localisée (au sens de sa décroissance à l'infini) et vérifiant $\int \phi(x)\, dx = 1$, les moments suivants étant nuls. En outre $\phi_j(x) = 2^{2j}\phi(2^j x)$. La perception basée sur le recul revient à examiner ce qui change lorsqu'on passe de f_{j+1} à f_j, ce qui revient à analyser f à l'aide de la suite des différences $g_j = f_{j+1} - f_j$. On retombe sur l'analyse de Littlewood-Paley et les g_j sont appelés des blocs dyadiques. On a évidemment $f = f_0 + \sum_0^\infty g_j$. La seconde option est plus subtile et sera la seule retenue en traitement du signal ou de l'image. Nous présenterons d'abord cette seconde option en dimension 1. Il s'agit d'une version améliorée de l'analyse en ondelettes. Outre l'ondelette mère ψ, intervient un autre personnage ϕ, appelé le “père des ondelettes” ou “fonction d'échelle”. La fonction ϕ est localisée, régulière et d'intégrale égale à 1. Les translatées entières $\phi(x-k)$, $k \in \mathbb{Z}$, de la fonction d'échelle forment une suite orthonormée. On pose ensuite $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$, $j, k \in \mathbb{Z}$. Alors les f_j et les g_j ne dépendent (localement) que d'un “petit nombre” de paramètres. Commençons par f_j. On pose $\alpha(j,k) = \int f(x)\phi_{j,k}(x)\, dx$ et $f_j(x) = \sum \alpha(j,k)\phi_{j,k}(x)$. Alors il vient, grâce à un choix approprié de la fonction d'échelle ϕ, $g_j(x) = f_{j+1}(x) - f_j(x) = \sum \beta(j,k) 2^{j/2}\psi(2^j x - k)$ et finalement la suite double $2^{j/2}\psi(2^j x - k)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}$, est, en dimension un, une base orthonormée de L^2. En dimension n, la situation est semblable à ceci près que $2^n - 1$ fonctions ψ sont nécessaires. Enfin on peut choisir ϕ et ψ dans la classe de Schwartz, mais on peut aussi, pour tout indice de régularité r choisir ces deux fonctions de sorte qu'elles soient de classe C^r, à support compact. Ce denier résultat est dû à Ingrid Daubechies. Nous avons finalement, en dimension n,

$$f(x) = \sum_j \sum_k \beta(j,k) 2^{nj/2}\psi(2^j x - k) \tag{9}$$

où les coefficients d'ondelette sont définis par

$$\beta(j,k) = \int f(x) 2^{nj/2}\psi(2^j x - k)\, dx \tag{10}$$

En d'autres termes la fonction que l'on analyse est découplée dans des canaux dyadiques couvrant approximativement une octave. A l'intérieur de chaque canal, tout se passe à l'échelle 2^{-j}. Les différents canaux se déduisent de l'un d'eux par simple changement d'échelle. Enfin les "building blocks" $2^{nj/2}\psi(2^j x - k)$ sont localisés et se déduisent de la "fonction mère" ψ par translation entière, puis dilatation dyadique. En deux dimensions il faut utiliser trois ondelettes ψ. On obtient alors la décomposition multi-échelle des images. La série d'ondelettes ainsi obtenue se situe à l'opposé d'une série de Fourier, car les ondelettes sont de plus en plus localisées lorsque l'on va vers les hautes fréquences. Ceci convient admirablement aux images, car les bords des objets génèrent des hautes fréquences et demandent à être localisés avec le plus grand soin. Dans la section suivante, nous verrons que ces bases d'ondelettes étaient implicites dans les algorithmes utilisés sous le nom de sub-band coding en traitement du signal.

9. Le standard JPEG 2000 et la compression des images fixes

Venons-en enfin à ce qu'on appelle la technologie et, dans le cadre de la révolution numérique, au problème de la compression des images fixes.

Représenter une image par une suite finie de 0 et de 1 pose des problèmes qui touchent à un ensemble de disciplines incluant les mathématiques, mais aussi la physique statistique ou les sciences cognitives.

La *chaîne de transmission* des images se compose de trois tronçons. Le premier est le codage de l'image : on vous donne une image et vous devez la représenter par une suite finie de 0 et de 1, ce qui vous oblige à négliger une partie importante de l'information contenue dans l'image en question. Vous voulez transmettre cette image et apparaissent alors tous les problèmes de codage en ligne qui font appel à la théorie des nombres, aux codes correcteurs d'erreurs, etc. Enfin vous devez décoder la suite de 0 ou de 1 reçue afin de reconstruire une image qui soit perçue comme la plus proche possible de l'image de départ.

On touche, à nouveau, un point qui se relie aux neurosciences : "perçue comme la plus proche possible" signifie que c'est l'œil qui décide si l'image reçue est de bonne qualité, si elle lui plaît ou ne lui plaît pas. On profite alors des capacités de masquage liées à la vision humaine, c'est-à-dire du fait que l'œil est plus sensible à certains défauts qu'à d'autres. L'algorithme doit être élaboré en tenant compte de la sensibilité de l'œil et, dans la littérature portant sur le traitement d'images et les problèmes de compression, vous avez toujours le jugement de l'expert qui vous dit si l'erreur est admissible ou pas ! Le jugement se fait donc toujours par un groupe d'experts : JPEG signifie d'ailleurs Joint Photographic Expert Group, groupe d'experts commun à l'ISO (organisation internationale de normalisation) et à la CEI (commission électronique internationale) chargé d'établir les normes de codage numérique de compression pour les images fixes.

L'algorithme de compression des images fixes qui a été utilisé jusqu'aujourd'hui est le célèbre JPEG. L'algorithme JPEG consiste à découper l'image en blocs 8×8 et à utiliser une analyse de Fourier discrète sur chaque bloc. JPEG fournit une bonne qualité pour des taux de compression de l'ordre de 10. L'objectif de JPEG 2000 est de passer de 10 à 100 !

Le point de départ de JPEG 2000 a été le "codage en sous-bandes" ou "subband coding", découvert à la fin des années 70 par D. Esteban et C. Galand, au centre de recherche IBM de La Gaude (près de Nice). Ce travail était motivé par le téléphone digital et jouera un rôle crucial dans la construction des ondelettes à support compact d'Ingrid Daubechies. En effet, comme Stéphane Mallat l'a montré, toute base orthonormée d'ondelettes peut s'interpréter comme un algorithme de codage en sous-bande. La réciproque est génériquement vraie : sauf si l'on est très malchanceux, une paire de filtres associés à un algorithme de codage en sous-bande conduit à une base orthonormée d'ondelettes. On doit à Albert Cohen d'avoir précisé ce point. Le codage en sous-bande va remplacer, dans JPEG 2000, la DCT qui était utilisée dans JPEG. En gros, pour comprimer une image, on commence par l'analyser en utilisant l'algorithme rapide de calcul des coefficients d'ondelettes (cet algorithme utilise le codage en sous-bande). Ensuite on quantifie les coefficients obtenus (qui sont des nombres réels) en les remplaçant par des approximations digitales appropriées. On hérite ainsi d'un ensemble fini de 0 et 1 que l'on transmet.

J'ai relu les travaux de C. Galand et il est clair qu'il ne s'est pas aperçu que les algorithmes qu'il découvrait convenaient plus à l'image qu'au téléphone. Il présentait une technique de traitement du signal où on descendait (ou montait) à travers les échelles et où l'aspect dynamique du suivi temporel paraissait mal pris en compte. Pour le signal de parole c'était peut-être inadéquat, mais sa découverte, le "codage en sous-bande", allait avoir un avenir prodigieux. On retrouve ici ce que François Jacob appelle le bricolage biologique : un petit mécanisme utilisé par une bactérie pour une certaine fonction est employé par l'homme pour une fonction complètement différente.

Faire une découverte sans s'en apercevoir est un phénomène assez courant. A l'époque où D. Esteban et C. Galand ont découvert le subband coding, IBM ne s'intéressait pas à l'image mais au téléphone numérique. Les programmes lancés par Esteban et Galand ont été abandonnés.

Je voudrais encore insister sur cette liaison profonde, organique, entre le traitement d'images du point de vue de l'informaticien et les problèmes qui touchent la neurobiologie et les neurosciences en citant (en traduction) un texte écrit par Peter Burt et son équipe. Il s'agit de scientifiques américains qui furent des pionniers de ce domaine. Ils décrivent l'analyse de l'image dans une base orthonormée d'ondelettes. Le lien entre le codage en sous-bande et les bases orthormées d'ondelettes a été découvert par Stéphane Mallat en 1986. Ceci est repris dans le texte qui suit que je traduis :

> *Pour comprimer les images, on va utiliser des transformations pyramidales. On décompose donc l'image en un ensemble de fonctions de base qui correspondent à une orientation en fréquences spatiales et qui sont localisées et auto-similaires. Pour des raisons de facilité de calcul, on veut aussi que cet ensemble soit orthogonal et soit fourni par un algorithme rapide...*
>
> *De telles transformations sont très utiles pour de nombreux aspects des images, car elles donnent des informations sur les changements d'intensité lumineuse à différentes échelles et les endroits où ces changements sont en train d'apparaître. Il y a aussi de fortes présomptions pour que le système visuel humain opère une décomposition de l'image similaire à celles-ci dans son traitement de bas niveau, c'est à dire avant que les fonctions cognitives n'entrent en jeu.*

Il importe de commenter le texte de Peter Burt. Il nous dit que chaque ondelette a une orientation privilégiée dans le domaine fréquenciel. Or nous avons toujours insisté sur le fait contraire : les ondelettes sont isotropes. Dans l'analyse continue décrite par (3) et (4), l'ondelette mère n'a aucune selectivité directionnelle. Elle peut parfaitement être une fonction radiale. En revanche les ondelettes orthogonales utilisées dans le traitement de l'image ont une faible selectivité fréquencielle. Les ondelettes mères sont au nombre de trois. La première oscille dans la direction verticale, la seconde dans la direction horizontale et la troisième dans les directions définies par les deux bissectrices des axes de coordonnées. C'est en ce sens que l'on peut parler d'une orientation. Mais ces orientations sont infiniment moins précises que celles définies par les neurones du cortex visuel primaire. Ces remarques motivent les travaux de Candès et Donoho qui seront décrits en section 11.

Les personnes qui travaillent sur le traitement d'images ont toujours à l'esprit l'idée d'améliorer les réalisations informatiques à l'aide des découvertes des neurosciences. La raison profonde de ce point de vue, c'est que finalement nous sommes les juges de la qualité de l'image et que, si elle ne nous plaît pas, elle sera rejetée quel que soit le rapport signal sur bruit (c'est à dire la mesure "objective" de la qualité de l'algorithme utilisé).

Il importe d'avertir le lecteur que JPEG-2000 est aujourd'hui largement dépassé et qu'un groupe de jeunes et audacieux chercheurs, réunis autour de Stéphane Mallat, dans une start-up dénommée "let it wave" et dont le site web est `www.letitwave.fr` a réussi l'exploit de comprimer les photos d'identité à 500 octets (500 bytes). Nous y reviendrons dans l'avant-dernière section.

10. La radiologie numérique

Je vais maintenant dire un mot sur quelques applications médicales des algorithmes de traitement de l'image. Ces progrès techniques peuvent "alléger la souffrance des hommes".

La radiologie numérique a été plébiscitée lors des "Journées françaises de la radiologie" qui se sont tenues à Paris du 23 au 27 octobre 2000. Mon information provient du compte-rendu fait par Michel Alberganti dans "Le Monde", daté du 27 octobre 2000. Une des raisons de ce succès de la radiologie numérique vient de ce que ces images numériques peuvent être téléchargées. Ainsi, s'il vous arrive quelque chose, peut-être un accident, à Nice alors que vous vivez normalement en région parisienne, le mdecin pourra avoir accès en temps réel à votre dossier médical ainsi qu'à toutes vos radios qui sont stockées dans l'ordinateur de votre hôpital de référence. Pour le moment, il s'agit d'un rêve.

Toute la télé-médecine, très importante en particulier pour les zones rurales, va être basée sur l'imagerie numérique. La France s'est mise avec un certain retard à cette technique et, selon M. Alberganti, à l'hôpital de Villejuif, par exemple, les images ne seraient toujours pas acquises numériquement et il faudrait effectuer une conversion analogique-numérique, alors qu'aux États-Unis ou en Allemagne les images médicales sont acquises directement en numérique ce qui rend plus souple et commode leur manipulation.

Le système de numérisation et transmission des images médicales s'appelle PACS et donne entièrement satisfaction aux praticiens. Un des avantages de l'image numérique est, pour ne retenir qu'un exemple, l'aide à la détection précoce de cancers. Il existe un algorithme de détection automatique des micro-calcifications qui peuvent être le signe d'un cancer du sein. Le médecin peut ainsi être aidé, accompagné, par un algorithme qui détecte les endroits où il y a une micro-calcification et donc un cancer potentiel.

Dans le cadre de la médecine du travail, des tests ont été effectués sur les médecins qui ont la responsabilité d'examiner des mammographies et on s'est aperçu que dans de nombreux cas, le médecin, à cause de la fatigue, laissait passer un cancer potentiel !

En imagerie médicale numérique, d'autres fonctionnalités de post-traitement et d'aides à la décision peuvent être incorporées. À mon avis la télé-médecine est vraiment un enjeu capital et elle ne peut exister sans la télé-radiologie et sans le pouvoir de transmettre des images de haute qualité sur des lignes téléphoniques ordinaires ou sur des lignes à haut débit, aussi bien à l'intérieur de l'hôpital qu'à l'extérieur.

JPEG-2000 est le seul algorithme de compression des images médicales qui ait été approuvé par la Food and Drug Administration et il a été développé par une corporation dont le nom est “Pegasus Imaging”. Le site web de l’organisme officiel qui développe JPEG-2000 est www.jpeg.org et celui de Pegasus, qui est très bien fait, très amusant et d’où vous pouvez télécharger des algorithmes est www.jpg.com.

Que peuvent faire ces nouvelles méthodes de traitement d’images que ne faisaient pas les anciennes ? Tout d’abord, elles sont plus performantes, mais encore elles offrent la capacité de zoomer sur un détail, c’est-à-dire de répartir comme vous le souhaitez “votre budget bits”. Au lieu de le répartir uniformément sur l’image : vous pouvez ignorer quatre-vingt-dix pour cent de l’image et zoomer sur un détail sur lequel vous concentrez votre allocation de bits. En terme de téléchargement, cela signifie que le médecin peut immédiatement regarder la zone qui l’intéresse et rendre flou tout le reste de l’image. Comme il est capital pour la médecine de respecter les budgets, vous donnez ainsi la possibilité d’obtenir un contraste plus grand et une qualité plus grande pour une certaine partie de l’image. En transformant l’image en une suite de 0 et de 1, vous faites des sacrifices et il convient de les faire le plus intelligemment possible.

11. Les ondelettes ne sont pas optimales et 500 octets suffisent.

Pendant les années où JPEG2000 a été élaboré, la ‘quête du graal’ a continué. Il s’agissait de faire mieux que les ondelettes dans la course à la compression. Plusieurs chercheurs ont ainsi cherché à mieux tirer parti des particularités géométriques des images. Un des modèles utilisés consiste à décrire les images en imitant le peintre Ingres. Une image est alors un ensemble de lignes assez régulières délimitant des zones où l’éclairement varie peu. Ces lignes seront appelées des bords. Si l’on utilise l’analyse par ondelettes pour décrire une telle image et si l’on désire atteindre une précision de l’ordre de $1/N$, il faut utiliser au moins N ondelettes distinctes, ce qui revient à coder les positions d’au moins N points (nous ne tenons pas compte de constantes multiplicatives et N signifie, en fait, CN). Il est clair que ce codage conduit à un gaspillage et que la même précision peut être obtenue en utilisant seulement $2\sqrt{N}$ données. Il suffit pour le voir de fournir les positions de $\sqrt{N}$ points des bords et des $\sqrt{N}$ tangentes à ces points. En utilisant l'approximation des bords par les cercles osculateurs, on fait encore mieux etc. Les ondelettes, quant à elles, sont sensibles à la présence de bords, mais n'arrivent pas à suivre la direction de ce bord, car elles sont 'isotropes'. Elles ne comportent pas un paramètre indiquant une direction. Seules la position et l'échelle sont prises en compte. Cette remarque doit être nuancée. D'une part les ondelettes utilisées dans les bases orthogonales ont une légère selectivité directionnelle. D'autre part cette selectivité peut être renforcée, comme m'a montré Sylvain Durand.

La première percée dans la direction de l'anisotropie a été réalisée par Emmanuel Candès et David Donoho. Ces deux chercheurs ont créé les *ridgelets* qui sont des *ondelettes de seconde génération.* Les *ridgelets* et leurs "cousines" les *curvelets* s'orientent et s'allongent en épousant automatiquement la géométrie d'un bord éventuel. La construction des *ridgelets* repose sur celle des ondelettes.

Stéphane Mallat a relevé le défi lancé par Cands et Donoho et a fourni une solution plus réaliste en inventant les *bandelettes.* Elles aussi épousent le plus longtemps possible le tracé des bords des images. La nouvelle technologie proposée par Mallat est développée par la sart-up *Let It Wave* et s'appelle *Let It Wave Codec. Let It Wave* a obtenu le premier prix de l'innovation technologique 2002, décerné par le Ministre de la Recherche et de la Technologie.

Quelques mots sur ce dernier algorithme. Le point de départ théorique est la thèse d'Erwann Le Pennec. Dans ce travail, une image est modélisée comme il vient d'être indiqué. Rappelons que les objets contenus dans l'image sont délimités par des bords nets, que l'on peut dessiner comme le ferait Ingres. L'éclairement à l'intérieur de chaque objet est supposé très régulier. A cette image brute, assez schématique, est ajouté un bruit aléatoire. Le problème posé est la description la plus concise, la plus économique, et la plus précise de ce type d'images. L'algorithme utilisé par Le Pennec est hybride, car il mêle les *level sets* de Jean-Michel Morel et Stanley Osher aux ondelettes. Le point de départ est la recherche des bords (supposés doux) des objets. En fait, on cherche seulement une information directionnelle décrivant les lignes parallèles aux bords. C'est ici qu'apparaissent les lignes de niveau. On tire ensuite parti de la régularité locale de l'éclairement pour l'analyser à l'aide d'ondelettes anisotropes dont les paramètres sont réglés à l'aide de l'information directionnelle déjà obtenue. Les *bandelettes* de Mallat et Le Pennec ne sont pas fixées une fois pour toutes, comme le sont les ridgelets, mais leur construction s'adapte (automatiquement) à l'image analysée. Le passage d'un travail de recherche à l'application industrielle développée par *Let It Wave* n'était pas une mince affaire et le résultat final obtenu par l'équipe de Stéphane Mallat est une prouesse.

12. Le "compressed sensing" d'Emmanuel Candès

Comme nous allons le voir, le "compressed sensing" est né d'un problème d'imagerie médicale, plus précisément de tomographie. Mais plaçons-nous d'abord dans un cadre très général. Les découvertes faites par Emmanuel Candès et Terence Tao s'inscrivent dans la problématique des problèmes inverses mal posés. Il s'agit de reconstruire une image inconnue alors qu'on ne dispose à son sujet que de quelques informations fragmentaires. Fournir dans ces conditions un résultat exact, comme si l'on disposait d'une information complète, est impossible et l'on doit pallier l'information manquante par une hypothèse a priori. Pour que la démarche proposée soit intéressante, il convient que cette connaissance a priori soit suffisamment souple et n'exclue aucun des signaux ou des images que l'on cherche à reconstruire.

Désignons par $X \in \mathbb{R}^n$ le signal (ou l'image) inconnu et supposons que l'on n'ait réalisé que quelques mesures $y_j =< A_j, X >,\ j = 1, ...p$, sur ce vecteur inconnu X. Les vecteurs $A_j \in \mathbb{R}^n$ sont donnés. Dans ce qui suit n est extrêmement grand (supérieur à 10^5) et p est beaucoup plus petit que n. La connaissance a priori sur la solution X est modélisée par un certain ensemble Γ et s'écrit $X \in \Gamma$. Dans ce qui suit Γ n'est pas convexe. Alors on cherche le vecteur inconnu X en essayant de résoudre

$$y_j =< A_j, X >,\ j = 1, ...p,\ \ X \in \Gamma \tag{11}$$

L'objet des travaux de d'Emmanuel Candès et de Terence Tao est double. Pour une classe importante d'ensembles Γ, Candès et Tao démontrent que (11) a une solution unique (ou bien n'a pas de solution) et, dans un second temps, ils proposent un algorithme pour la déterminer.

La définition de la classe Γ est dictée par les travaux sur les *représentations creuses*. Le point de départ concerne l'analyse par ondelettes des images. Les images que l'on peut modéliser par des fonctions appartenant à l'espace BV des fonctions à variation bornée ont une représentation creuse dans une base d'ondelettes (Albert Cohen) et cette découverte mathématique est en plein accord avec les acquis de la neurophysiologie (travaux de David Hubel et Torsten Wiesel sur la détection des contours par les neurones de l'aire V_1 du cortex visuel primaire). La base dans laquelle la représentation d'un signal ou d'une classe de signaux est creuse sera désormais notre base de référence. Dans cette nouvelle base, l'ensemble T des indices des coefficients significatifs a une petite cardinalité $|T|$, compte tenu de la taille du signal. Mais cette propriété ne fournit aucune information sur la localisation de cet ensemble T. Dans le cas du traitement de l'image, cette localisation est déterminée par les positions des bords de l'image. C'est ce qui amène Candès et Tao à considérer pour tout entier $q \leq n$, l'ensemble Γ_q des signaux $X = (x_1, ..., x_n)$ tels que $|T| \leq q$ où $|T|$ est le cardinal du support $T = \{j\,;\ x_j \neq 0\}$ de X. Cet ensemble Γ_q n'est évidemment pas un espace vectoriel.

Le théorème suivant de Candès et Tao éclairera notre discussion. Soit p un nombre premier et soit $\mathbb{F}_p$ le corps fini à p éléments. La transformée de Fourier d'une fonction f à valeurs réelles ou complexes, définie sur $\mathbb{F}_p$ est, elle-même, définie sur $\mathbb{F}_p$ et est notée $\hat{f}$.

Théorème 1. — *Soit Ω une partie de $\mathbb{F}_p$. Alors toute fonction f définie sur $\mathbb{F}_p$ est déterminée de façon unique par la restriction de sa transformée de Fourier à Ω lorsque le cardinal du support T de f vérifie la condition $|T| \leq \frac{1}{2}|\Omega|$.*

Cet énoncé est faux lorsque $\mathbb{F}_p$ est remplacé par $\mathbb{Z}/n\mathbb{Z}$ et que n n'est pas un nombre premier. Néanmoins il redevient vrai si Ω est un ensemble ayant des propriétés arithmétiques particulières, comme nous allons le voir dans un instant. Le théorème 1 apparaît aujourd'hui comme le point de départ des travaux ultérieurs de Candès et Tao. Mais l'histoire est différente. Le vrai point de départ fut un problème posé par la tomographie. Il s'agissait de reconstruire une image simple f à partir d'une information partielle sur sa transformée de Fourier. La seule donnée dont on dispose est la connaissance de $c(\omega) = \hat{f}(\omega),\ \omega \in \Omega$. L'ensemble Ω se compose d'une centaine de lignes passant par l'origine. L'hypothèse a priori est que f est constante par morceaux, c'est-à-dire que f est égale à une constante c_j à l'intérieur d'un domaine D_j et que ces domaines sont en nombre fini. Les D_j ont des bords réguliers et f est discontinue sur ces bords. Alors Candès et Justin Romberg découvrirent expérimentalement que f est la solution (unique) du problème

$$\inf \|g\|_{BV};\ \hat{g}(\omega) = c(\omega),\ \omega \in \Omega \tag{12}$$

Le gradient ∇f de f est porté par un ensemble fini de lignes. Il a donc un petit support et nous nous retrouvons dans un cadre qui ressemble à celui du théorème 1. On observera que $\|g\|_{BV} = \|\nabla g\|_1$ et l'on pourra alors considérer (12) comme un cas particulier du théorème 2 ci-dessous. Mais le théorème 1 ne disait rien sur l'algorithme permettant de calculer f et n'explique donc pas l'algorithme (12). En outre l'hypothèse que p est premier n'est pas satisfaite. Voici un théorème plus général obtenu par Candès et Tao. Commençons par une définition. Considérons $q \le p$ vecteurs Z_j appartenant à l'espace euclidien $\mathbb{R}^p$. Pour un $0 < \epsilon < 1$, on dira que ces vecteurs $Z_j \in \mathbb{R}^p,\ 1 \le j \le q$ sont $\epsilon-$orthogonaux si, pour tout choix des coefficients c_j, on a

$$(1-\epsilon)\sum_1^q |c_j|^2 \le \|\sum_1^q c_j Z_j\|^2 \le (1+\epsilon)\sum_1^q |c_j|^2 \tag{13}$$

Une matrice $(p \times n)$ A est dite (q, ϵ) incohérente si tout ensemble de q vecteurs colonnes extrait des vecteurs colonnes de A vérifie (13). Pour A et q fixés, ϵ_q désigne la borne inférieure de ces valeurs de ϵ. On revient alors à (11) où $\Gamma = \Gamma_q$ est défini comme ci-dessus (le cardinal du support de X ne dépasse pas q) et où A est la matrice dont les vecteurs lignes sont les A_j, $1 \le j \le p$.

Théorème 2. — *Avec les notations précédentes, on suppose que*

$$\epsilon_q + \epsilon_{2q} + \epsilon_{3q} < 1 \tag{14}$$

Alors le système (11) *où* $\Gamma = \Gamma_q$ *a une solution au plus, qui est définie (si elle existe) en résolvant le problème d'optimisation convexe*

$$\inf \|X\|_1;\ y_j =< A_j, X >,\ j = 1, ...p \tag{15}$$

où $\|X\|_1 = |x_1| + \cdots + |x_n|$.

On notera que la solution unique de (15) appartient automatiquement à Γ_q. On pourrait penser que l'algorithme fourni par le théorème est instable. Il n'en est rien. Candès et Tao démontrent un remarquable théorème de robustesse en prouvant que le théorème 2 s'applique à des signaux bruités et fournit alors un algorithme de débruitage.

La principale difficulté est de construire, pour tout n et tout p des matrices d'usage commode vérifiant (14). Candès et Tao, d'une part, Alain Pajor d'autre part, montrent que des matrices aléatoires conviennent. Il est indispensable d'en savoir plus et ce domaine de recherche touche à la physique statistique et à la théorie des nombres. David Donoho a relié les travaux de Candès et Tao à la géométrie des ensembles convexes en grande dimension. Les site web de Candès, de Donoho et de Pajor sont indiqués à la fin de cet article.

13. Conclusion

Revenons aux questions que nous nous sommes posées. Les recherches portant sur la compression des images peuvent-elles bénéficier des progrès de la neurophysiologie ? Ce que nous avons appris nous apporte des réponses nuancées. Tout d'abord le traitement de l'image et l'étude de la vision des mammifères se sont développés de façon autonome. Il n'y a pas eu d'influence directe entre ces deux secteurs de l'activité scientifique. Les convergences n'en sont que plus troublantes. Le scientifique qui a été le plus influencé par ces convergences fut David Marr que nous citerons dans l'appendice. David Marr, spécialiste de la vision des mammifères, avait été invité par le MIT pour résoudre des problèmes de robotique. Tout son travail, avant qu'il ne soit emporté par la leucémie, porta sur la convergence entre les "algorithmes biologiques" qui sont à la base de la vision humaine et les "algorithmes mathématiques" nécessités par la vision artificielle. Par exemple, David Marr comprit que les fonctions de certains neurones du cortex visuel primaire s'apparentent à une analyse par ondelettes. Les travaux récents d'Emmanuel Candès et de Stéphane Mallat renforcent encore les découvertes de David Marr. En effet, les ridgelets et les bandelettes sont anisotropes et constituent une meilleure approximation des fonctions des neurones du cortex visuel primaire que ne le sont les ondelettes qui sont isotropes. Mais aujourd'hui il me semble que les modèles à base d'ondelettes sont trop simplistes pour expliquer le fonctionnement de la vision des mammifères. Nous nous éloignons du programme de Marr. On pourra à ce propos consulter les remarquables travaux de Guy Orban, Full Professor, Medical School, Katholieke Universiteit te Leuven.

14. Appendice

Nous regroupons ici quelques notes de lecture. Commençons par Ramón y Cajal. Il obtint le Prix Nobel de médecine en 1906. Il a voulu encourager les jeunes espagnols à faire de la recherche scientifique. Il écrit :

> *De los dóciles y humildes pueden salir los santos, pocas veces los sabios...La veneración excesiva, como todos los estados pasionales, excluye el sentido crítico...Todo hombre puede ser, si se lo propone, escultor de su proprio cerebro, y aun en el peor dotado es susceptible, al modo de las tierras pobres, pero bien cultivadas y abonadas, de rendir copiosa mes... Cuando se reflexiona sobre la curiosa propriedad que el hombre posee de cambiar y perfeccionar su actividad mental con relación a un objeto o problema profundamente meditado, no puede menos de sospecharse que el cerebro, merced a su plasticidad, evoluciona anatómica y dinámicamente, adaptandose progresivamente al tema...*

Oui ! Tout homme peut, s'il en prend la décision, sculpter son propre cerveau !

Dans un autre registre, voici comment David Marr analyse le rôle de ce qu'il appelle une représentation. David Marr était un théoricien des problèmes posés par la vision humaine. Il pensait qu'il existait une science ou une grammaire de la vision s'appliquant aussi bien à la vision humaine qu'à la vision artificielle. Il écrit :

> *"A representation, therefore, is not a foreign idea at all ; we all use reprentations all the time. However, the notion that one can capture some aspect of reality by making a description of it using a symbol and that to do so can be useful seems to me a fascinating and powerful idea. But even the simple examples we have discussed introduce some rather general and important issues that arise whenever one chooses to use one particular representation. For example, if one chooses the Arabic numeral representation, it is easy to discover whether a number*

is a power of 10 but difficult to discover whether it is a power of 2. If one choses the binary representation, the situation is reversed. Thus there is a trade-off; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover...But even though one is not restricted to using just one reprentation system for a given type of information, the choice of which to use is important and cannot be taken lightly. It determines what information is made explicit and hence what is pushed further into the background, and it has a far reaching effect on the ease and difficulty with which operations may subsequently be carried out on that information."

Terminons avec David Hubel :

Understanding the human brain is one of the greatest challenges that scientists face. It is surely the most complex machine in the known universe. It has commonly been thought that understanding the brain is a hopeless quest, since the main tool we have to work with is our own brain. I have never seen any merit in this logic, and the work I have described here, together with research in hundred of laboratories, has surely already led to remarkable progress in understanding the brain, even though it is just a beginning.

15. Références

Voici des sites web d'où vous pouvez décharger les travaux de recherche qui m'ont permis de préparer cet exposé.

En ce qui concerne le standard JPEG2000, le meilleur site est :

`http:// jj2000.epfl.ch`

mais on pourra aussi consulter :

`http:// www.jpg.com` (qui est le site de la firme Pegasus) ou

`http:// www.jpeg.org` (qui est le site officiel du comité JPEG2000)

Le site des travaux de Candès est :

`http:// www.acm.caltech.edu/ ˜ emmanuel`

et celui d'Alain Pajor est

`http://perso-math.univ-mlv.fr/users/pajor.alain/`

En ce qui concerne l'analyse en composantes indépendantes, on consultera :

`http://sig.enst.fr/ ˜ cardoso/stuff.html`

Le travail de Saito est disponible sur :

`http://www.math.ucdavis.edu/ ˜ saito`

Le travail de Donoho se trouve à l'adresse suivante :

`http://www-stat.stanford.edu/ ˜ donoho`

On apprend tout sur les travaux de Mallat et sur les ondelettes en consultant le site :

`http://www.cmapx.polytechnique.fr/ ˜ mallat`

ou bien, si l'on s'intéresse aux liens entre la théorie des ondelettes et l'approximation non-linéaire :

`http://www.math.sc.edu/ ˜ devore`

Enfin les mordus d'images astronomiques se régaleront en consultant le site :

`http://oposite.stsci.edu`

qui fournit, en direct, les images du télescope spatial Hubble, ainsi qu'une plongée vertigineuse vers les confins de l'Univers !