PERCEPTION ET COMPRESSION DES IMAGES FIXES

par

Yves Meyer

La révolution numérique envahit notre civilisation en nous inondant d'images digitales. On peut dire que ce déluge n'est pas à craindre, car la plupart de ces images ne présentent aucun intérêt. Nous n'aborderons pas ici les problèmes posés par l'accès au contenu sémantique d'une image et supposerons qu'il faille archiver ou transmettre certaines de ces images. Les premiers obstacles rencontrés sont le volume des données ou le débit limité des canaux de transmission. On ne peut archiver ou transmettre sans comprimer et nous voilà dans le vif du sujet, c'est-à-dire face au problème de la compression en imagerie numérique. **Les progrès en imagerie numérique sont-ils reliés à une meilleure compréhension du fonctionnement du système visuel humain ?** Cette question peut paraître paradoxale, tant la révolution numérique semble faire partie de la technologie et n'avoir aucun lien avec la neurophysiologie, ni avec les sciences cognitives. C'est pourtant l'inverse qui est vrai. En effet, nous observerons, dès la première section, que l'étude de la *perception* intervient inévitablement dans les problèmes posés par la compression des images fixes. Cela nous amènera à nous poser une autre question : *Quels sont les fondements de notre perception des objets ou des images ?* En fait, ce problème concerne tout autant le langage (dont une des fonctions est de classer et de désigner les choses) que la vision proprement dite. C'est pourquoi nous examinerons, dans la première partie de l'exposé, les points de vue de trois philosophes, George Berkeley, Ernst Cassirer et Merleau-Ponty, et celui d'un peintre, Nicolas Poussin. La neurophysiologie consolidera, en les précisant, leurs analyses et notre discussion sera alors basée sur les travaux de Pierre Buser, de

David Hubel et de Torsten Wiesel. Dans la seconde partie de l'exposé, c'est-à-dire dans les sections 5 à 10, nous étudierons trois algorithmes de compression utilisés en imagerie numérique (JPEG-2000, les curvelets et les bandelets). L'étude du "compressed sensing" ouvre de nouvelles et passionnantes directions de recherche. Enfin nous pourrons conclure en reliant ces trois algorithmes de compression des images fixes aux acquis de la neurophysologie ou aux travaux sur la perception.

1. Perception et compression sont-elles reliées ?

Le traitement de l'image est une discipline jeune, née avec la révolution numérique, parce que seules les images numériques peuvent être manipulées à l'aide d'algorithmes explicites et reproductibles. Mais le traitement de l'image se rattache, en fait, à un savoir-faire très ancien. Voyons plutôt. Avant l'invention de la photographie, les peintres avaient le monopole de la production des images et leurs œuvres explicitaient leur *perception du monde extérieur*. Ceci n'a pas changé et Pascal avait bien tort d'écrire :

> *Quelle vanité que la peinture, qui attire l'admiration par la ressemblance des choses dont on n'admire pas les originaux !*

Le peintre n'imite pas, contrairement à ce que pensait Pascal ; il nous propose une représentation du monde ; il nous livre sa vision intérieure et ses images mentales, car il a aussi peint pour nous. Ce faisant, le peintre effectue une compression des images fixes ! En effet, en quelques traits de crayon, un dessinateur habile fait surgir, de façon précise et exacte, un visage familier. Le problème de la compression des images est ainsi résolu par le dessin, mais le résultat dépend du talent du dessinateur. Les manipulations sur les photographies argentiques étaient de l'ordre du savoir-faire et n'étaient donc pas reproductibles. On pense aux retouches effectuées sur les photographies officielles des dignitaires de l'ex-URSS. [On pourra consulter : *Le commissaire disparaît ; la falsification des photographies et des œuvres dans la Russie de Staline*, David King (1997)]. Aujourd'hui les caméras numériques nous proposent des images "objectives" du monde, vierges de toute perception humaine, et que chacun peut manipuler à volonté. Les algorithmes dont nous disposons pour manipuler ces images ne sont plus condamnables. Il ne s'agit plus de tromper, mais plutôt de rehausser l'image, de la débruiter ou de la

comprimer.

Revenons maintenant à notre question : *La compression de ces images digitales est-elle reliée à leur perception ?* Signalons tout de suite que nous passerons sous silence la "compression en ligne" qui accélère le débit de transmission des messages et bénéficie de la théorie de Shannon. De façon très grossière cette compression imite le fonctionnement du code Morse. Le codage le plus court y est offert aux lettres (ou mots) les plus fréquents. Nous étudierons un second type de compression ; il s'agit de la compression avec perte (dite *lossy*). La compression avec perte est un sujet très ardu et les techniques employées dépendent de la nature des documents à transmettre. Nous ne traiterons pas ici les signaux "audio", comme le signal de parole, et nous nous concentrerons sur le cas particulier des images fixes, en excluant la video. L'image n'est plus exactement la même après avoir subi une compression avec perte, ; elle a été simplifiée. Pouvons-nous accepter cette simplification ? Ne risque-t-on pas de caricaturer l'image de départ ? On ne dispose malheureusement pas, à l'heure actuelle, de critères objectifs permettant de décider de la qualité d'un algorithme de compression. On peut demander que l'écart quadratique moyen entre l'image originale et l'image comprimée soit le plus faible possible. Mais ce point de vue donne la même importance à toutes les parties de l'image, ce qui est absurde du point de vue perceptuel (pensez à un visage et à l'importance des yeux). On demandera donc plutôt que l'image produite par l'algorithme soit *perceptuellement proche* de l'image originale. Mais adopter ce point de vue nous impose d'en savoir un peu plus sur la perception.

La seconde relation entre *perception et compression* est plus subtile, car elle repose sur l'étude de la cohérence interne des images. Je vais préciser ce point en prenant l'exemple du langage écrit et de l'exercice consistant à résumer un texte. Résumer un texte est impossible si ce texte n'a pas de sens. Si le texte est bien écrit, avec un plan et des articulations claires, il est possible de le résumer, à moins qu'il ne soit trop dense. Paul Valéry observe qu'on ne peut résumer un poème. En d'autres termes, un texte ne peut être résumé que si trois conditions sont remplies : le texte doit être bien écrit, il doit être légèrement redondant et la personne qui lira le résumé doit être suffisamment avertie pour savoir "lire entre les lignes" et ainsi retrouver l'information qui manque. Retournons aux images. Si la compression des images fixes peut se faire à des taux

impressionnants, c'est parce que (1) l'information fournie par *une image naturelle est structurée et/ou redondante* et que (2) *notre perception est adaptée* à "lire entre les lignes", c'est-à-dire à compléter une information visuelle incomplète ou altérée. On ne peut percevoir ou comprimer une image qui ne soit pas 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. Cette *syntaxe* est-elle accordée à notre perception ? Le philosophe Ernst Cassirer rèpond à cette question et affirme que *la perception d'une image naturelle est basée sur la reconnaissance des formes ou des structures sous-jacentes qui y figurent* (ce point sera approfondi dans la section suivante). Nous arrivons au cœur de notre débat en posant les questions suivantes :

a) *Comment les images naturelles sont-elles structurées ?*
b) *Existe-t-il des formes simples et universelles qui soient à la base de la perception et de la compression des images naturelles ?*
c) *La perception des images a-t-elle un équivalent algorithmique ?*

Pour répondre aux deux premières questions nous utiliserons diverses approches statistiques dont *l'analyse en composantes indépendantes* (ICA). L'ICA a été appliquée à un certain corpus d'images naturelles (des photographies de la campagne anglaise). Les résultats seront détaillés dans la section 6. Nous présenterons d'importants résultats provenant d'autres études statistiques portant sur des corpus d'images naturelles. Nous décrirons ensuite un modèle, dû à Stanley Osher et Leonid Rudin, où toute image f est décomposée en une somme de trois composantes. La première u tient compte des objets présents dans l'image f. La seconde v décrit les textures (souvent délimitées par les bords des objets) et la dernière w est un bruit additif. La réponse à la troisième question nous amènera à parler des processus de bas niveau en traitement de l'image et en théorie de la perception. La perception serait le produit d'une chaîne d'opérations. Le premier (bas) niveau serait consacré à la détection et au codage des structures les plus élémentaires ou des formes les plus simples qui soient présentes dans l'image. L'information accédant aux niveaux suivants serait "calculée", de proche en proche, grâce à des algorithmes pyramidaux, à partir des résultats obtenus dans les étapes antérieures. L'étude du cortex visuel primaire et les algorithmes pyramidaux illustreront ce propos. Enfin la réponse à la dernière question sera donnée dans la conclusion.

2. Perception et petites impressions

Avant d'être enfin abordée à l'aide des outils fournis par les neurosciences, la question de savoir comment la *perception* s'élabore à partir des simples *sensations* avait été posée par les philosophes. Dans "La philosophie des Lumières" Ernst Cassirer nous rappelle que le théologien et philosophe irlandais George Berkeley (1685-1753) avait soigneusement analysé les problèmes posés par la perception :

> *La question avait étét posée pour la première fois dans l'Optique de Molyneux et avait éveillé aussitôt le plus vif intérêt philoso-phique. Les expériences que nous avons faites dans l'un de nos secteurs sensoriels peuvent-elles nous permettre de constituer un secteur de contenu qualitativement différent et d'une autre structure spécifique? Y a-t-il une connection interne nous permettant de passer directement d'un secteur à l'autre, par exemple du monde tactile au monde visible? Un aveugle de naissance qui aurait acquis, grâce à l'expérience du toucher la connaissance excate de certaines formes corporelles et qui saurait faire entre elles à coup sr la différence possédera-t-il encore ce même don de distinction lorsqu'une heureuse opération lui aura rendu le sens de la vue et qu'il devra dès lors juger de ces formes sur la base de données purement optiques ? Va-t-il d'emblée pouvoir distinguer, par le seul moyen de la vue, un cube d'une boule, ou lui faudra-t-il un long et difficile effort de conciliation avant de parvenir à établir la liaison entre les impressions tactlies et la forme visible de l'un et de l'autre ?*

Il est remarquable que la cohérence entre la vision et le toucher soit l'un des sujets d'étude des neurosciences contemporaines. Nous verrons également que le problème posé par Molyneux se retrouve dans les recherches de David Hubel sur lesquelles nous reviendrons. Mais écoutons encore Ernst Cassirer :

> *Berkeley, dans sa* **Nouvelle théorie de la vision et ses principes de la connaissance humaine**, *était parti de ce paradoxe : la seule matière, le seul matériau dont disposons pour édifier notre monde perceptif, ne consiste que dans les simples impressions sensibles ; mais, d'autre part, ces impressions sensibles elles-mêmes ne comportent pas la moindre*

> *indication des "formes" sous lesquelles se présente à nous la réalité perçue. Nous croyons voir cette réalité en face de nous, comme une structure solide, où chaque élément singulier aurait sa place désignée et ses relations avec tous les autres exactement déterminées...C'est en donnant à son concept fondamental de perception une signification plus large que Berkeley surmonte ce dilemme, en y faisant rentrer, outre la simple sensation, l'activité de* **représentation.**...*Cette interaction des impressions sensibles, cette régularité avec laquelle elles s'appellent et se représentent mutuellement devant la conscience, est le fondement dernier de la représentation de l'espace... Et lorsque Chedelsen parvint en 1728 à guérir grâce à une heureuse opération un garçon de quatorze ans, aveugle de naissance, il parut que cette question, posée par Molyneux comme pure hypothèse, avait trouvé sa solution expérimentale. Les observations effectuées sur ce garçon semblèrent en effet confirmer en tous points la thèse empiriste. Les prédictions théoriques de Bekeley étaient entièrement vérifiées : il s'avérait que le malade, en recouvrant la lumière, n'avait nullement gagné d'emblée la faculté de voir, qu'en particulier il lui fallait apprendre progressivement et péniblement à distinguer les formes corporelles qu'on présentait à sa vue.*

Nous verrons avec émerveillement que cette forme de cécité sera expliquée par David Hubel et Torsten Wiesel dans leurs travaux sur l'amblyopie et la période critique d'apprentissage. Mais revenons encore une fois à *La philosophie des formes symboliques.* Ernst Cassirer y écrit :

> *Et c'est ainsi partout que le libre agir de l'esprit dissipe le chaos des impressions sensibles et lui donne pour nous une valeur stable. Ce n'est qu'en opposant à l'impression fuyante un pouvoir constructif, dans une des directions de la symbolisation, que cette impression acquiert pour nous forme et durée. Ce passage à la forme s'accomplit de diverses façons et selon des principes de construction différents dans la science, dans le langage et dans le mythe ; principes et façons qui coïncident cependant dans le fait que le produit final de leur activité, tel qu'il nous apparaît, ne ressemble plus par aucun trait au simple matériau dont il procède originairement.*

Jean-Jacques Rousseau allait dans la même direction que George Berkeley quand il écrivait :

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

Et, de même, Jean-Paul Sartre nous dit que "*dans la perception, un savoir se forme lentement.*"

En s'appuyant sur les avancées de la neurophysiologie et des sciences cognitives, Pierre Buser analyse 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.

Analysant 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 hiérarchique 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 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..."*

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 réel 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 dans *La philosophie des formes symboliques*, Yale Univ. Press (1953) :

> "*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'arti-cule 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.*"

La théorie de la Gestalt illustre les propos d'Ernst Cassirer. Pour Wertheimer, les objets sont perçus directement comme des entités globales parce que l'esprit est "programmé" pour reconnaître instantanément des formes géométriques simples et même pour les créer quand elles n'existent pas. On trouvera une belle discussion de cette théorie dans le livre de David Marr, *Vision, A computational investigation into the human representation and processing of visual information* (W.H. Freeman, 1982). Une des justifications de la théorie de la Gestalt est fournie par les illusions visuelles. Rappelons que dans des expériences aussi reproductibles et assurées que celles des sciences dures, la perception préattentive de certaines images fait apparaître des formes simples, quitte à ce que notre perception trace ou prolonge des lignes qui ne figurent pas dans l'image observée. Mais David Marr va bien plus loin dans la modélisation du fonctionnement du sytème visuel. Il suppose que le traitement bas-niveau nous fournit une traduction de l'information fournie par la rétine en ce qu'il appelle une *représentation.* Il écrit : Il définit ce terme de la façon suivante :

> *A representation, therefore, is not a foreign idea at all–we all use representations 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 dicussed 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 chooses 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.*

David Marr définit *the first complete symbolic representation of an image* à l'aide de sa théorie des "zero-crossings" qui est détaillée dans l'ouvrage déjà mentionné. On pourra trouver une analyse critique de cette théorie dans *Wavelets, Tools for Science & Technomogy* [Stéphane Jaffard, Y.M. and Robert Ryan, SIAM 2001]. Pour conclure, les peintres et les philosophes et Pierre Buser nous tiennent en fait le même langage. Ils nous enseignent que 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 complexe ; Ernst Cassirer parlerait d'une "opération spirituelle". Ces opérations dépendent peut-être d'une modélisation appropriée du monde extérieur. Le système visuel humain aurait pris en compte et incorporé, au cours de l'évolution, les modèles les plus appropriés au traitement de l'information fournie par les images de la nature. Nous verrons réapparaître cette hypothèse dans la section 6. Peut-on en savoir un peu plus sur ces modèles implicites qui fondent le langage ou la perception du monde qui nous entoure ? Nous essayerons de répondre à cette question en nous appuyant sur les données de la neurophysiologie et sur 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 dans *Eye, Brain and Vision* (publié en 1988 par W.H. Freeman), sur le fonctionnement du cortex visuel primaire :

> "*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 traités ; le cortex visuel primaire n'est pas le centre de la "perception"."*

Cette description exclut un algorithme du type "transformée de Fourier", car cette dernière est une transformation globale, opérant sur toute l'image. Bien au contraire, 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 un croquis ou un 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. Ecoutons David Hubel (Madrid, 1995) décrire son travail :

> "*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.*"

D'autres neurones sont affectés à 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, ni même la perception des objets qui y figurent. Cette compréhension fera appel à des processus cognitifs mettant en jeu des populations de neurones. Après de nombreuses étapes, l'information "prétraitée" fournie par les neurones du cortex visuel primaire parvient aux aires secondaires ou associatives, responsables des processus cognitifs. 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 les expériences vécues s'écrivent et organisent nos modes de perceptions. La cohérence que notre cerveau acquiert se construirait ainsi dans l'apprentissage et refléterait celle du monde qui nous entoure. Mais D. Hubel a découvert que c'était souvent l'inverse qui a lieu. Certaines parties du cerveau sont *précâblées* et ce câblage peut s'effacer, faute de simulations. Cette découverte a balayé bien des préjugés sur l'apprentissage et la pédagogie.

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. 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 !

Les travaux de Hubel et Wiesel ouvrent plusieurs voies de recherche. La première nous conduira à des modèles qui soient basés principalement sur la notion de contour, puisque, selon David Hubel, la détection des contours est le point de départ du processus conduisant à la perception d'une image. D. Hubel va plus loin et interprète la détection des lignes effectuée par le cortex visuel primaire en termes de compression de l'information. Nous y reviendrons dans la conclusion. Ce premier groupe de modèles consiste donc à représenter une image par un ensemble de courbes que l'on puisse dessiner. Cela nous conduira soit aux *level-sets* de Stanley Osher et de Jean-Michel Morel, soit aux modèles basés sur les fonctions à variation bornée (modèle d'Osher, L. Rudin et E. Fatemi). Tous ces modèles mettent l'accent sur la géométrie. Dans le modèle d'Osher, L. Rudin et E. Fatemi, on ne demande aucune régularité aux contours ; on leur impose seulement d'être de longueur finie.

Mais ces mêmes travaux de Hubel et Wiesel nous conduiront aussi aux modèles analytiques qui utilisent des "briques de base" ou "building blocks" pour l'analyse et la synthèse des images. Il s'agit des ondelettes. Les ondelettes de Gabor ou de Morlet-Grossmann seront définies dans les sections suivantes. Elles apparaissent dans l'analyse des images au niveau du cortex visuel primaire. En effet, dans l'ouvrage "Computational Neuroscience of Vision" par Edmund T. Rolls et Gustavo Deco, [Oxford University Press, (2002)], Chapitre 2, puis Chapitre 9, on lit :

The receptive fields of simple V1 neurons have additional lobes of excitation and inhibition, such that these simple neurons are not only sensitive to a specific position and orientation but also to the spatial frequency of the stimuli. In fact, the size of the receptive field is strongly correlated with the spatial frequency at which they preferentially respond. These kinds of profiles can be matched very well with the so-called Gabor functions or wavelets.

The pioneering work of Hubel and Wiesel (1962) about the organization of the primary visual cortex has given impetus to the theoretical and experimental research of the response properties of the cells in this area. The theoretical investigations of Daugman (1988) and Marcelja (1980) proposed that simple cells in the primary visual cortex can be modelled by 2D-Gabor functions. There is a trade off between the resolution that can be achieved in the spatial and the frequency domain (Daugman 1997). The 2D-Gabor function achieves the optimal resolution limit in the conjoint spatial and frequency domain. The Gabor receptive fields have five degrees of freedom given essentially by the product of an elliptical Gaussian and a complex plane wave. The first two degrees of freedom are the 2D-locations of the receptive field's centre ; the third is the size of of the receptive field's centre ; the fourth is the orientation of the boundaries separating excitatory and inhibitory regions and the fifth is the symmetry. This fifth degree of freedom is given in the standard Gabor transformation by the real and imaginary part, i.e. by the phase of the complex function representing it, whereas in a biological context this can be done by combining pairs of neurons with even and odd receptive fields (Daugman 1988). This design is supported by the experimental work of Pollen and Ronner (1981), who found simple cells in quadrature-phase pairs.

Even more, Daugman (1988) proposed that an ensemble of simple cells is best modelled as a family of 2D-Gabor wavelets sampling the frequency domain in a log-polar manner. Experimental neurophysiological evidence constrains the relation between the free parameters that define a 2D-Gabor receptive field (De Valois and De Valois, 1988, Kulikowski and Bishop 1981,

> *Webster and De Valois 1985). There are three constraints fixing the relation between the width, height, orientation, and spatial frequency (Lee 1996)... Further, we assume that the mean is zero in order to have an admissible wavelet basis (Lee 1996). The V1 neuronal pools implemented in the model described in this Chapter consist of an ensemble of simple cells whose receptive fields correspond to a 2D-Gabor function sensitive to a particular orientation, symmetry, and spatial frequency.*

Les modèles géométriques et les modèles analytiques sont-ils conciliables ? Nous savons aujourd'hui, grâce aux travaux d'Albert Cohen, d'Ingrid Daubechies et de leurs collaborateurs que les ondelettes de Morlet-Grossmann, qui seront définies dans la section suivante, constituent la *meilleure base* permettant de décrire les fonctions à variations bornées sous forme de séries. Ces fonctions sont celles qui représentent le niveau de gris dans le modèle d'Osher, L. Rudin et E. Fatemi. Les ondelettes incorporent également le *zoom à travers les échelles* et il semble que l'on ne puisse rêver d'un meilleur choix. Ce n'est peut-être qu'une illusion, car tout autre modèle décrivant, de façon plus précise, les contours conduirait à une autre solution. Nous verrons le bien-fondé de cette remarque lorsque nous étudierons les *ridgelets*, les *curvelets*, les *bandlets* ou les *contourlets*.

5. Les ondelettes

L'analyse par ondelettes est née, à la fin des années 70, d'une étonnante découverte faite par un ingénieur, Jean Morlet. Cette découverte fut comprise et acceptée par un physicien, Alexandre Grossmann, puis par des spécialistes du traitement du signal et par quelques mathématiciens. Vingt ans après, l'analyse par ondelettes débouchait sur le nouveau standard, nommé JPEG2000, de compression des images fixes. 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 (c'est-à-dire des comportements brutaux et inattendus). 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 qui seront définies dans la section suivante). 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.

J'ai souvent discuté avec Jean Morlet. Il ressemble beaucoup à Benoît Mandelbrot. Tout comme Mandelbrot, Morlet a 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 escroc

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, lui aussi situé dans l'autre pièce. 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 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 Alexandre 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 servent à *comprimer et transmettre* les données recueillies dans les campagnes pétrolières. Ce sont des signaux 1-D (représentant les variations ou

fluctuations d'une fonction du temps). L'analyse de ces données est une tout autre histoire. De même, les ondelettes servent à *comprimer et transmettre* les images qui sont bi-dimensionnelles (ou 2-D).

Ni Grossmann, ni Morlet ne sont des numériciens. Ils avaient bien proposé des algorithmes de calcul de la transformée en ondelettes et de la transformée inverse, mais ces algorithmes étaient lourds sous leur forme exacte et imprécis sous leur forme approchée, alors que la révolution numérique repose sur l'utilisation d'algorithmes exacts et rapides. Par exemple, 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 ce qui concerne l'analyse par ondelettes, un long chemin restait à parcourir et il a fallu attendre les travaux d'Ingrid Daubechies, d'Albert Cohen et de Stéphane Mallat pour que la Fast Wavelet Transform se hisse 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 en 1983, dans le cadre du traitement de l'image. Dans le programme de Mallat, l'analyse par ondelettes d'une image se présente comme un cas particulier d'*algorithme pyramidal*. Utiliser un algorithme pyramidal pour analyser une image revient à prendre du recul en reculant pas à pas (la pyramide se construit pendant cette itération). Mais les algorithmes pyramidaux ne fournissent pas encore les bases orthonormées d'ondelettes. Pour cela il faut construire la pyramide à partir de la version orthogonale du *codage en sous-bandes* ou *subband coding.* Ce codage et sa version orthogonale avaient été inventés en 1977 par D. Esteban et C. Galand au

centre d'IBM de La Gaude. Leur motivation était le téléphone digital. Ni l'application à l'image, ni le problème de la stabilité dans le *zoom arrière* n'avaient pas été abordés par D. Esteban et C. Galand. Les algorithmes qu'ils proposaient ne pouvaient être utilisés sans précaution. La pyramide construite à l'aide du codage en sous-bandes peut s'effondrer. Nous devons à Albert Cohen et Ingrid Daubechies l'étude complète de la stabilité de ces pyramides.

La construction par Ingrid Daubechies (1987) des bases orthonormées d'ondelettes à support compact, de régularité r donnée apparaît aujourd'hui, dans une perspective historique, comme la suite logique de ce programme. 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 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 Jean-Christophe Feauveau, Ingrid Daubechies construisit les ondelettes bi-orthogonales. Ce sont elles qui seront utilisées dans JPEG2000. Ces ondelettes ont le triple avantage d'être à support compact, d'être symétriques et d'avoir une régularité que l'on peut choisir arbitrairement.

Les succès des ondelettes en traitement de l'image s'expliquent de deux façons différentes que nous avons déjà évoquées dans la section précédente. La première explication est reliée au modèle de Stanley Osher et Leonid Rudin qui sera détaillé dans l'appendice. Dans ce modèle, on considère que les contours des objets contenus dans une image peuvent être tracés parce que la somme de leurs longueurs est finie. Cela amène à modéliser une image par la somme d'une fonction à variation bornée et d'un terme représentant les textures et le bruit. Si l'on adopte ce modèle, il convient de choisir une représentation adaptée au sens donné par David Marr ou une langue adaptée au sens de Cassirer. On peut adopter un point de vue très différent et privilégier des algorithmes permettant d'effectuer un "zoom arrière", c'est-à-dire

de voyager à travers les échelles. Par exemple, on peut utiliser les "algorithmes pyramidaux" de P. Burt et E.H. Adelson. Nous y reviendrons encore à la fin de cette section.

Comme Stéphane Mallat l'avait prévu, ces points de vue ne sont pas antagonistes : les bases orthonormées d'ondelettes qui sont adaptées aux algorithmes pyramidaux constituent aussi les bases optimales pour la compression des fonctions à variation bornée. Ceci a été prouvé en 1997 par Albert Cohen, Ingrid Daubechies et leurs collaborateurs. Les deux options (contours de longueurs finies ou "zoom arrière") conduisent donc au même algorithme et cet algorithme est utilisé dans le nouveau standard JPEG2000 de compression des images fixes. La beauté et l'intérêt de ces résultats ne doivent cependant pas faire illusion. En effet, toute cette discussion repose sur le modèle particulier utilisant l'espace BV. D'autres choix de modèles conduisent à d'autres algorithmes

Pour conclure cette section consacrée à l'analyse par ondelettes, je voudrais relier cette analyse à une technique plus ancienne, connue sous le nom d'analyse de Littlewood-Paley chez les mathématiciens et de "banc de filtres à surtension constante" en traitement du signal. L'intérêt de ce rapprochement est de nous permettre de revenir, une fois encore, aux algorithmes réalisant un "zoom arrière".

Avant de présenter l'analyse de Littlewood-Paley, revenons un instant au point de vue des philosophes : ils nous rappellent qu'il convient de prendre du recul pour mieux percevoir une image ou un tableau dans un musée. On devient alors sensible à la façon dont notre perception se réorganise dans ce "zoom arrière". En termes mathématiques, on a le choix entre plusieurs options pour modéliser ce recul. Une première solution consiste à lisser l'image donnée en la rendant floue et imprécise et à la remplacer par une suite de convolutions $f_j = f \star \phi_j,\ j \in \mathbf{N}$, où ϕ est une fonction suffisamment régulière, suffisamment localisée (au sens de sa décroissance à l'infini) et vérifiant $\int \phi(x)\,dx = 1$, tous les autres moments étant nuls et où $\phi_j(x) = 2^{nj}\phi(2^j x)$. Le modèle de la perception consiste alors à examiner ce qui change lorsqu'on passe de f_j à f_{j+1}, 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 fournit l'analyse par ondelettes du signal ou de l'image. Dans cette seconde option, les f_j sont des versions simplifiées de l'image, c'est-à-dire des sketchs. Plaçons-nous en dimension n. Alors les f_j ne dépendent (localement) que d'un "petit nombre" de paramètres et sont définis par $f_j(x) = \sum \alpha(j,k) 2^{nj/2} \phi(2^j x - k)$; la "fonction d'échelle" ϕ et ses translatées entières $\phi(x-k)$, $k \in \mathbf{Z}^n$, forment une suite orthonormée et la somme définissant f_j porte sur $k \in \mathbf{Z}^n$. De même $g_j(x) = f_{j+1}(x) - f_j(x) = \sum \beta(j,k) 2^{nj/2} \psi(2^j x - k)$ et finalement la suite double $2^{nj/2}\psi(2^j x - k)$, $j \in \mathbf{Z}$, $k \in \mathbf{Z}^n$, forme une base orthonormée de $L^2(\mathbf{R}^n)$. En fait $2^n - 1$ ondelettes 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 dernier résultat est dû à Ingrid Daubechies. Nous obtenons finalement l'idendité remarquable suivante :

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

En d'autres termes la fonction f que l'on analyse est découplée dans des canaux dyadiques Γ_j ; chacun de ces canaux couvre approximativement une octave. Les différents canaux se déduisent de l'un d'eux par simple changement d'échelle. A l'intérieur de Γ_j, tout se passe donc à l'échelle 2^{-j}. Les "building blocks" $2^{nj/2}\psi(2^j x - k) \in \Gamma_j$ sont les ondelettes engendrées par les $2^n - 1$ "ondelettes mères" ψ. Elles sont aussi localisées en variable d'espace et se déduisent des $2^n - 1$ ondelettes ψ par translations entières, puis dilatations dyadiques. En dimension deux, il faut donc utiliser trois ondelettes ψ. On obtient alors la décomposition multi-échelle des images. La série fournit une analyse qui se situe à l'opposé de celle que donnerait une série de Fourier, car les morceaux sont de plus en plus localisés 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. Signalons que, dès 1972, l'existence de telles bases orthonormées avait été prévue par Kenneth Wilson dans ses travaux sur la renormalisation.

6. La décomposition en composantes indépendantes ou ICA

Peut-on mieux comprendre le succès des ondelettes en traitement de l'image ? Plus généralement existe-t-il une méthodologie objective pour construire la base orthonormée convenant le mieux à un ensemble donné de signaux ? Cette adéquation peut se définir de deux façons différentes : la première concerne la concision de la décomposition et la seconde la pertinence de l'analyse. La pertinence de l'analyse est reliée à l'intérêt présenté par les "mesures" qui sont effectuées. Ici une mesure est un coefficient dans la base choisie. Ces mesures sont d'autant plus importantes qu'elles donnent des points de vue différents sur les signaux ou les images étudiées. Cette seconde option conduit à *l'analyse en composantes indépendantes* alors que la première nous oriente vers les algorithmes du type "best basis" développés par R. Coifman et son groupe. Voici une illustration de ces idées.

Les neurophysiologistes se sont intéressés aux représentations efficaces des images et en particulier des images de scènes naturelles. Une image contient une information gigantesque et il est exclu qu'elle soit intégralement perçue par le système visuel animal ou humain ; on ne perçoit qu'une partie infime d'une image que l'on regarde. L'hypothèse de travail de certains neurophysiologistes est la suivante : si les cellules du cerveau 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 l'os est optimal 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.

C'est pourquoi D. J. Field et de B. A. Olshausen ont essayé d'expliquer l'analyse des images effectuée par le cortex visuel primaire à l'aide des résultats de l'ICA, *independant component analysis* ou analyse en composantes indépendantes. L'ICA est appliquée à une collection d'images de la campagne anglaise. On a pensé reproduire ainsi l'environnement des premiers hominidés (le lecteur a le droit de sourire). Cela nous amène à poser la question suivante : *Si l'on optimisait l'analyse des images des scènes naturelles, c'est-à-dire si on leur appliquait l'ICA, on*

retrouverait précisément les images élémentaires qui correspondent aux diverses spécialisations des cellules du cerveau?

Ecoutons B. A. Olshausen et E. Simoncelli [Natural Image statistics and neural representation, Annu. Rev. Neurosci. 24 (2001) 1193-1216] :

> *Understanding the function of neurons and neural systems is a primary goal of systems neuroscience. The evolution and development of such systems is driven by three fundamental components : (a) the tasks that the organism much perform, (b) the computational capabilities and limitations of neurons (this would include metabolic and wiring constraints), and (c) the environment in which the organism lives...The use of such ecological constraints is most clearly evident in sensory systems, where it has long been assumed that neurons are adapted, at evolutionary, developmental, and behavioral timescales, to the signals to which they are exposed...The establishment of a precise quantitative relationship between environmental statistics and neural statistics is important for a number of reasons... More thant 40 years ago, motivated by developments in information theory, F. Ayttneave (1954) suggested that the goal of visual perception is to produce an efficient representation of the incoming signal. In a neurobiological context, H.B. Barlow (1961) hypothesized that the role of early sensory neurons is to remove statitical redundancy in the sensory input. Variants of this "efficient coding" hypothesis has been formulated by numerous other authors.*

L'analyse en composantes indépendantes répond à ce dernier défi d'éliminer la redondance statistique. L'ICA est basée sur le concept intuitif de "contraste". Comme nous l'avons déjà dit, 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.

L'analyse en composantes indépendantes est un programme scientifique excitant pour les uns, décevant pour les autres. Nous verrons pourquoi en donnant des exemples et des contre-exemples. 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 le traitement des signaux captés par la tour de contrôle d'un aéroport. 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". Ce mélange a créé la complexité apparente et a brouillé les sources. Voici une modélisation mathématique.

Nous partons d'une famille Ω de signaux $x_\omega,\ \omega \in \Omega$. Ici Ω est un ensemble. La loi de probabilité $dP(\omega)$ sur cet ensemble Ω de signaux interviendra dans un instant. Pour simplifier la discussion, nous supposerons que les signaux x_ω soient échantillonnés sur $\{1, 2, ..., N\}$. Ils seront alors notés $x(j, \omega),\ 1 \leq j \leq N,\ \omega \in \Omega$. On utilisera par la suite une écriture vectorielle plus concise où l'ensemble des signaux donnés, enregistrés est noté $X(\omega),\ \omega \in \Omega$ et $X(\omega) \in \mathbb{R}^N$. On cherche à savoir si ces signaux peuvent s'écrire comme N combinaisons linéaires

$$x(j, \omega) = \sum_1^m \alpha(j, k)\, s_k(\omega) \quad (1 \leq j \leq N) \tag{1}$$

de m sources indépendantes, notées $s_1(\omega), ..., s_m(\omega),\ \omega \in \Omega$. Dans le cas d'un processus $X(t, \omega)$, $t \in [a, b]$, cette décomposition se généralisera en

$$X(t, \omega) = \sum_0^\infty \alpha_j(\omega) f_j(t) \tag{2}$$

où les $\alpha_j(\omega)$ sont des variables aléatoires indépendantes.
Pour définir l'indépendance des sources, il est maintenant nécessaire que l'ensemble Ω soit muni d'une loi de probalité $dP(\omega)$. Une telle loi modélise l'ensemble des connaissances a priori que l'on possède sur ces

signaux. L'indépendance des signaux sources peut alors être définie et se réfère à cet espace de probabilité Ω, muni de la loi de probabilité $dP(\omega)$.

Sous une forme matricielle condensée, on écrit $X(\omega) = AS(\omega),\ \omega \in \Omega$, où $X(\omega),\ \omega \in \Omega$, est l'ensemble des signaux donnés, enregistrés, mais où ni la matrice A (de type $N \times m$), ni les sources $S(\omega)$ ne sont connues. Les signaux $X(\omega)$, $\omega \in \Omega$, forment maintenant un nuage de points dans $\mathbb{R}^N$ et l'on cherche à découvrir un modèle permettant de rendre compte de la formation de ce nuage. Le but du jeu est d'estimer des valeurs probables de m et de A à partir de la donnée de X. Cette recherche est basée sur l'hypothèse très forte de l'indépendance statistique des sources. En supposant $m = N$ et que la décomposition en composantes indépendantes existe, une solution consiste à construire une matrice $B = A^{-1}$ qui minimise la dépendance statistique entre les coordonnées de BX. Nous allons définir cette dépendance statistique en introduisant la "distance de Kullback".

Dans le cas gaussien, l'indépendance entre les sources $s_k(\omega),\ 1 \leq k \leq N$, et la décorrélation sont la même propriété. La décorrélation signifie que $\int_\Omega s_k(\omega) s_l(\omega) dP(\omega) = 0,\ k \neq l$, en supposant que les sources soient centrées, c'est à dire aient une espérance nulle. Ceci entraîne que, dans le cas gaussien, la décomposition en composantes indépendantes coïncide avec la décomposition en composantes principales ou décomposition de Karhunen-Loève (KL). Avant de retourner à l'analyse en composantes indépendantes, disons quelques mots de l'analyse en composantes principales.

En toute généralité la décomposition de Karhunen-Loève s'écrit

$$X(\omega) = \lambda_1 g_1(\omega) Y_1 + ... + \lambda_N g_N(\omega) Y_N \tag{3}$$

où les $g_1(\omega), ..., g_N(\omega)$ sont indépendantes dans le cas gaussien et décorrélées dans le cas général; les λ_j sont des nombres réels positifs ou nuls et les $Y_1, ..., Y_N$ forment une base orthonormée de $\mathbb{R}^N$. Cette décomposition s'obtient de la façon suivante. On part d'un nuage de points $X(\omega) \in \mathbb{R}^N,\ \omega \in \Omega$, et l'on suppose que les coordonnées $x_k(\omega),\ 1 \leq k \leq N$, soient centrées, c'est à dire que la moyenne en ω (ou espérance) de chaque coordonnée soit nulle. On calcule alors les

covariances

$$\gamma_{k,l} = \int_\Omega x_k(\omega) x_l(\omega) dP(\omega) \quad (1 \leq k, l \leq N) \tag{4}$$

puis la matrice des covariances $\Gamma = \left(\gamma_{k,l}\right)_{1 \leq k,l \leq N}$. Les vecteurs propres orthogonaux de cette matrice définie positive sont les composantes principales cherchées. En termes plus géométriques, l'analyse en composantes principales revient à déterminer l'axe d'inertie du "nuage de points" $X(\omega),\ \omega \in \Omega$, c'est à dire la droite affine D telle que la somme des carrés des distances à D soit minimale. Une fois D déterminée, on cherche le second axe d'inertie etc.

Dans le cas général de processus non gaussiens, la propriété d'indépendance est beaucoup plus forte que la décorrélation et la décomposition en composantes indépendantes est unique, à une permutation près des indices (à condition qu'au plus une des composantes de $\mathbf{s}$ soit gaussienne).

La dépendance statistique est mesurée à l'aide de la "distance de Kullback". Soient f et g les densités de répartition de deux vecteurs aléatoires $\mathbf{y}(\omega) \in \mathbb{R}^n$ et $\tilde{\mathbf{y}}(\omega) \in \mathbb{R}^n,\ \omega \in \Omega$. Ces densités sont définies sur $\mathbb{R}^n$. La "distance de Kullback" entre ces densités est

$$K(f, g) = \int_{\mathbb{R}^n} f(x) \log \frac{f(x)}{g(x)} dx \tag{5}$$

Cette distance (appelée divergence) est positive ou nulle et n'est nulle que si $f = g$. Le minimum des divergences de Kullback entre la distribution d'un vecteur aléatoire donné $\mathbf{y}(\omega)$ et des vecteurs (arbitraires) $\tilde{\mathbf{y}}(\omega)$ dont les composantes sont indépendantes s'appelle l'information mutuelle entre les composantes de $\mathbf{y}(\omega)$. Cette information mutuelle est nulle si et seulement si les coordonnées de $\mathbf{y}(\omega)$ sont elles-mêmes indépendantes.

L'algorithme utilisé pour retrouver A et S à partir de F consiste, dans un premier temps, à "blanchir" les données F, en appliquant, par exemple, l'algorithme de Karhunen-Loève, puis à transformer ces données blanchies $\tilde{F}$ par une nouvelle matrice orthogonale U de façon à minimiser l'information mutuelle entre les composantes du vecteur transfomé $U\tilde{F}$. Lorsque les données ont été blanchies, minimiser l'information mutuelle

revient à minimiser la somme des entropies (au sens de Shannon) des composantes du vecteur transformé.

Citons Jean-François Cardoso:

> *Mixing the entries of* $\mathbf{s}$ *'tends' to increase their entropies; it seems natural to find separated source signals as those with minimum marginal entropies. It is also interesting to notice that (up to a negative constant) the entropy of each component* y_k *is the Kullback divergence between the distribution of* y_k *and the zero mean unit-variance normal distribution. Therefore, minimizing the sum of the marginal entropies is also equivalent to driving the marginal distributions of* $\mathbf{y}$ *as far away as possible from normality. Again, the interpretation is that mixing tends to gaussianize the marginal distributions so that a separating technique should go in the opposite direction.*

Cette dernière formulation conduit à maximiser les kurtosis des coordonnées du vecteur $U\tilde{F}$ et une approximation de la solution s'obtient alors à l'aide des cumulants d'ordre supérieur chers à Jean-Louis Lacoume.

La décomposition de Karhunen-Loève fournit une vision en moyenne du processus. En revanche l'algorithme ICA a tendance, grâce à son caractère non-linéaire, à détecter des évènements transitoires.

Voici un exemple géométrique très simple. On part du cube unité $\Omega \subset \mathbb{R}^n$. Ce cube est défini par

$$\omega = \omega_1 \mathbf{e_1} + ... + \omega_n \mathbf{e_n} \tag{6}$$

où les coordonnées $\omega_1, ..., \omega_n$ sont regardées comme des variables aléatoires indépendantes sur l'espace de probabilité $\Omega = [0,1]^n$. En d'autres termes, le cube Ω (vu comme un nuage de points) a une décomposition triviale en composantes indépendantes. Cette décomposition est donnée par (6).

Si maintenant on examinait ce cube en vision cavalière, c'est à dire dans la direction d'une diagonale, il ne ressemblerait plus à un cube; en dimension 3, on obtiendrait un hexagone. Supposons que l'on dispose de n telles vues cavalières prises à partir de n points de vue différents. Alors

l'analyse en composantes indépendantes a pour but de retrouver le cube d'origine à partir de ces prises de vues modifiées par les changements de perspective et de calculer la rotation qui a été effectuée. On observera que ceci ne serait pas possible si le cube était remplacé par une sphère. Cette sphère a la même forme quelque soit le point de vue envisagé. On peut effectivement reconnaître un cube à partir de vues cavalières et même deviner la rotation effectuée; ceci est une illustration du cas non gaussien dans le problème de séparation de sources. En revanche on ne peut détecter la rotation dans le cas de la sphère (qui modélise le cas gaussien).

La meilleure référence que je connaisse sur l'ICA est le site web de Jean-François Cardoso qui est http://sig.enst.fr/cardoso/stuff.html et une première initiation à ces méthodes m'a été donnée par Jean-Louis Lacoume.

Nous nous proposons maintenant, en suivant D. J. Field et B. A. Olshausen, 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 la plus grande surprise des chercheurs, 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). Dans le premier cas, les ondelettes sont indexées par deux paramètres ω et x_0 appartenant à $\mathbb{R}^2$ et sont définies par:

$$w_{(\omega,x_0)}(x) = \exp(i\omega \cdot x) g_\sigma(x - x_0), \quad \omega \in \mathbb{R}^2, x_0 \in \mathbb{R}^2 \tag{7}$$

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

Dans le second cas, les ondelettes sont indexées par un paramètre positif a et un vecteur $x_0 \in \mathbb{R}^2$ et sont définies par

$$\psi_{(a,x_0)}(x) = a^{-1}\psi(\frac{x - x_0}{a}), \quad a \in (0, \infty), x_0 \in \mathbb{R}^2 \tag{8}$$

Or ces "ondelettes" sont précisément les motifs (ou "primitives" selon la terminologie de David Marr) qui sont détectés de façon privilégiée par les cellules rétiniennes. L'analyse en composantes indépendantes modéliserait donc le travail effectué par les cellules rétiniennes. 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.

Il importe de préciser que l'analyse de Karhunen-Loève appliquée à cette même base de données fournirait la base de Fourier. Mais les cellules du cortex visuel primaire ne calculent évidemment pas les coefficients de Fourier (cette évidence intellectuelle est confirmée par l'expérimentation).

Deux exemples nous permettront de mieux comprendre les forces et les faiblesses de l'analyse en composantes indépendantes appliquée à des signaux non-stationnaires et non-gaussiens, où apparaissent des évènements transitoires brutaux. Le premier exemple est celui de la rampe. Le second sera celui des processus de Lévy. La rampe est l'exemple le plus simple, en une dimension, d'une fonction régulière par morceaux et ayant des discontinuités de saut (jump discontinuities). En deux dimensions, on considérera des fonctions qui sont régulières à l'intérieur de domaines ayant eux-mêmes des bords réguliers, mais qui présentent des discontinuités de saut à travers les bords de ces domaines. De telles fonctions modélisent grossièrement des images géométriques.

Mais revenons à la rampe. Il s'agit d'un processus $X(t,\omega)$ indexé par $\omega \in [0,1]$ et défini par $X(t,\omega) = t$ si $0 \leq t < \omega$ et $X(t,\omega) = t-1$ si $\omega \leq t \leq 1$. L'analyse en composantes principales de ce processus est décevante, car elle revient à décomposer la rampe dans le système trigonométrique. La convergence, mesurée en écart quadratique moyen, est très lente. En outre, l'analyse en composantes principales ne fournit aucun renseignement direct sur la particularité la plus importante de ce signal: la discontinuité en $t = \omega$.

Du point de vue de l'ICA, l'exemple de la rampe semble un contre-exemple. En effet on ne peut, stricto sensu, décomposer ce processus en composantes indépendantes. Plus précisément, il n'est pas possible

d'écrire

$$X(t,\omega) = \sum_0^\infty \alpha_k(t)\phi_k(\omega) \tag{9}$$

où les $\phi_k(\omega)$ constituent une suite de variables aléatoires indépendantes. On peut seulement appliquer à ce processus l'algorithme qui fournirait les composantes indépendantes si elles existaient. Alors la surprise est que "l'algorithme ICA" fournit approximativement la décomposition de la rampe dans une base d'ondelettes "temps-échelle". La convergence de chaque réalisation a acquis une vitesse exponentielle! Mais l'ordre des termes de la série dépend de la réalisation. Ceci nous amène à penser que l'ICA est adaptée à des classes de signaux présentant de fortes transitoires, comme c'est le cas pour les images.

Un exemple qui est plus satisfaisant pour le mathématicien est celui "vols de Lévy" décrits par Benoît Mandelbrot dans "Les objets fractals" (Champs, Flammarion). Voici leur construction. On part d'un exposant $D \in (0,2)$ et d'une suite $X_k, k \in \mathbb{Z}$, de variables aléatoires indépendantes, centrées et identiquement distribuées. La loi de répartition de X_k est donnée par

$$\begin{cases} \text{Prob}\{X_k \in [-1,1]\} = 0 \\ \text{Prob}\{X_k > \lambda\} = Prob\{X_k < -\lambda\} = (1/2)\lambda^{-D} \quad (\lambda > 1) \end{cases} \tag{10}$$

Alors on considère la marche aléatoire $Y_n = \sum_0^n X_k$ que l'on peut encore écrire $Y(t,\omega) = \sum_0^\infty X_k(\omega)H(t-k),\ t \in \mathbb{N},$ où $H(t)$ est la fonction indicatrice de $[0,\infty]$ (fonction de Heaviside). Nous avons bien ici une décomposition en composantes indépendantes du processus $Y(t,\omega)$. Pour relier cette marche aléatoire aux processus de Lévy, on commence par lisser $H(t)$ en la remplaçant par $\tilde{H}(x) = 0$ si $x \leq 0, \tilde{H}(x) = x$ si $x \in [0,1]$ et $\tilde{H}(x) = 1$ si $x \geq 1$. Ensuite on utilise le pas de temps $h > 0$ au lieu de $h = 1$, ce qui conduit à choisir l'échantillonnage $t = kh, k \in \mathbb{Z}$. Finalement on change d'échelle dans les déplacements et l'on obtient

$$Y_h(t,\omega) = h^{1/D} \sum_0^{+\infty} X_k(\omega)\tilde{H}(\frac{t}{h} - k) \tag{11}$$

On vérifie alors que, si h tend vers 0, Y_h converge en loi vers un processus de Lévy stable. Les processus de Lévy ont des sauts, présentent des transitoires brutales et constituent, à ce titre, une version stochastique et autosimilaire de la rampe. L'analyse en composantes indépendantes

est adaptée à ce type de comportement.

Voici une alternative intéressante à l'analyse en composantes indépendantes. La version "analyse fonctionnelle" de cet algorithme est la recherche de *bases inconditionnelles dans un espace de Banach*. On part d'un espace fonctionnel E modélisant la classe des signaux ou des images Γ que l'on étudie. Cela signifie que les propriétés ou la connaissance a priori dont on dispose sur Γ peuvent être décrites à l'aide d'une norme, dans un espace fonctionnel E adapté. Par exemple les images géométriques sont souvent modélisées par des fonctions à variations bornées. Cela se traduit par $\Gamma \subset \lambda B$ pour un certaine constante positive notée λ, , B désignat la boule unité de E. Ensuite on applique une variante de l'analyse en composantes indépendantes à la boule unité B de E. Cela signifie que l'on cherche une suite e_n de vecteurs de E telle que les fonctions (ou signaux) $f \in B$ s'écrivent, de façon unique,

$$f = \sum_{0}^{\infty} \alpha_n \, e_n \tag{12}$$

où les coefficients α_n sont indépendants les uns des autres; c'est à dire que l'on peut arbitrairement les modifier en les rempaçant par d'autres coefficients β_n dont les modules sont du même ordre de grandeur ($|\beta_n| \leq |\alpha_n|$) sans changer l'appartenance de f à B. On dira alors que $e_n,\ n \in \mathbf{N}$, est une *base inconditionnelle de* E.

Pour conclure cette section, signalons l'existence d'une abondante littérature sur les propriétés statistiques des images [J-P. Nadal (ENS-Ulm), A. Turiel (Dto de Física Teórica, Universidad Autónoma de Madrid)]. Les auteurs écrivent :

> *The description of the early stages of the visual pathway in mammalians and other animals must be addressed from the knowledge of the properties of the signals that this system is intended to encode : natural images. These are very complex objects, and truly random from the point of view of the observer. However, natural images are structured, highly redundant objects, a fact that becomes clear for instance in that the luminosity changes smoothly over the reflecting surfaces. This redundancy, which should be used as a priory knowledge about*

the signal, is useful to develop optimal coding strategies, which are learnt by sensory system.

Les auteurs continuent leur discussion en utilisant le concept de multifractalité, dû à Uriel Frisch and Giorgio Parisi, pour modéliser les images naturelles. Nous n'en dirons pas plus et revoyons le lecteur intéressé aux travaux de Jean-Pierre Nadal. On se reportera au site web de Jean-Pierre Nadal qui est www.lps.ens.fr/Jean-Pierre.Nadal. On pourra aussi consulter William E. Vinje and Jack L. Gallant, *Sparse coding and decorrelation in primary visual cortex during natural vision,* Science, 287, (18 Februray 2000).

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

Venons en à ce qu'on appelle la technologie et, dans le cadre de la revolution numerique, 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 un ensemble d'images et vous devez les représenter par des suites finies de 0 et de 1, ce qui vous oblige à négliger une partie importante de l'information contenue dans les images en question. Vous voulez transmettre ces 0 et 1. 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. Les algorithmes de compression et de décodage doivent être conçus 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) chargée d'établir les normes de codage numérique de compression pour les images fixes.

L'algorithme de compression des images fixes qui était 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 JPEG2000 est de passer de 10 à 100 !

Comme nous l'avons déjà dit, l'un des points de départ de JPEG2000 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). 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'ondelette (cet "algorithme de Mallat" utilise le codage en sous-bande). Ensuite on quantifie les coefficients réels obtenus en les remplaçant par des approximations digitales appropriées. On hérite ainsi d'un ensemble fini de 0 et 1 que l'on transmet. On reconstruit alors l'image en utilisant l'identité remarquable décrite dans la section 5. Comme les coefficients ont été quantifiés, on ne retombe pas exactement sur l'image de départ. L'erreur commise s'analyse en faisant appel à une très belle théorie mathématique, celle des bases inconditionnelles. Cette théorie permet de prévoir les effets du "bruit de quantification". Les bases orthonormées d'ondelettes sont des bases inconditionnelles universelles ; c'est une nouvelle explication de leur supériorité sur les séries de Fourier qui constituent, de ce point de vue, le pire choix possible. C'est encore une des raisons de la supériorité de JPEG2000 sur JPEG.

Le 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

indéfiniment continué. Le lecteur intéressé se reportera à l'excellent site : http://jj2000.epfl.ch (l'absence de www n'est pas une erreur).

Je voudrais encore insister sur cette liaison profonde, organique, entre le traitement d'images, vu du point de vue de l'informaticien, et les problèmes qui touchent la neurophysiologie et les neurosciences en citant (en traduction) un texte écrit par Peter Burt et son équipe :

> "*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, 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 opère une décomposition de l'image similaire à celle-ci dans son traitement de bas niveau, c'est à dire avant que les fonctions cognitives n'entrent en jeu.*"

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

Dans le modèle d'Osher, Rudin et Fatemi, les images sont modélisées par $f = u + v + w$ où $u \in BV$, $v \in L^2$, et où w est une bruit blanc gaussien. La définition de la décomposition optimale est donnée dans l'appendice. Alors la recherche de la composante u, c'est-à-dire des objets contenus dans l'image f, se fait grâce au "*wavelet shrinkage*" de David Donoho. A. Cohen et ses collaborateurs ont éclairé ce travail de Donoho en cherchant la meilleure approximation d'une fonction à variation bornée par une combinaison linéaire f_N de N ondelettes (les positions et les largeurs de ces ondelttes étant, par ailleurs, arbitraires). A. Cohen a démontré qu'il existe une constante absolue C telle que l'on ait

$$\|f - f_N\|_2^2 \leq CN^{-1}\|f\|_{BV}^2 \tag{13}$$

La meilleur approximation par le système trigonométrique ne donnerait aucun résultat intéressant ; en dimension deux, l'espace BV

est inclus dans l'espace de Lorentz $L^{2,1}$ et cette inclusion est optimale. Plus précisément, il n'existe pas de base orthonormée qui fasse mieux que les ondelettes pour comprimer les fonctions à variation bornée.

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. Cela n'est possible qu'en utilisant des modèles plus précis que le modèle BV d'Osher, Rudin et Fatemi. Plusieurs chercheurs ont donc proposé des modèles plus contraignants que nous décrivons maintenant. Il s'agit de modéliser 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. Dans le modèle BV d'Osher, Rudin et Fatemi, la contrainte imposée aux lignes était d'avoir des longueurs finies (elles peuvent alors être assez irrégulières). Dans ce qui suit, nous imposerons, au contaire, une certaine régularité aux bords des ondelettes. 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). En d'autres termes, l'analyse par ondelettes traite une telle image à bords réguliers comme s'il s'agissait de la pire des fonctions à variation bornée. En théorie de l'approximation, ce phénomène est étudié sous le nom de *saturation*. 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.

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* consituent une base orthonormée tandis que leurs "cousines" les *curvelets* forment un "*frame*" c'est-à-dire un ensemble redondant, mais dont la

redondance est limitée par une constante C. Plus précisément un frame est une collection e_j, $j \in J$, de vecteurs d'un espace de Hilbert H telle qu'il existe deux constantes $C_1 > C_0 > 0$ de sorte que l'on ait pour tout $x \in H$, on ait

$$C_0\|x\|^2 \leq \sum_{j\in J} | < x, e_j > |^2 \leq C_1\|x\|^2 \tag{14}$$

Cette condition assure une reconstruction stable. Les curvelets s'orientent et s'allongent en épousant automatiquement la géométrie d'un bord éventuel. Si f est une image de type "cartoon", c-à-d. régulière à l'intérieur de courbes de classe C^2, alors la meilleure approximation f_N de f par une combinaison linéaire de N curvelets vérifie

$$\|f - f_n\|^2 \leq C(\log N)^3 N^{-2} \tag{15}$$

Il convient de noter l'amélioration d'un ordre de grandeur par rapport à ce que l'obtiendrait avec des ondelettes traditionnelles. La construction des ridgelets et des curvelets est un exploit scientifique.

Stéphane Mallat a relevé le défi lancé par Donoho et a fourni, en inventant les *bandelettes*, une solution qui est conceptuellement moins belle, mais bien plus réaliste. Les bandelettes 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, à l'aide de contours réguliers délimitant les objets contenus dans l'image. 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. En fait Le Pennec considére un modéle plus complexe où l'image est donnée comme une somme $T(u)+v$ où T est un opérateur qui modélise l'optique ayant servi à acquérir les images. En général T dégrade l'image qui est devenue légérement floue et où v repésente les textures et le bruit. Ce modéle est cependant trop complexe, parce que ni l'opérateur T, ni le

bruit ne peuvent recevoir de traitement suffisamment général. Chaque cas particulier nécessiterait un traitement approprié. C'est pourquoi dans le travail de Le Pennec, T est la convolution avec une approximation de l'idendité. On a $T(u) = u \star h_s$ où $h_s = s^{-2}h(x/s)$ et où h est une fonction réguliére, à support compact et d'intégrale égale à 1. Nous ignorons les textures, u est supposée réguliére (de régularité Hölderienne α) en dehors d'une collection finie de courbes fermées de même régularité Hölderirenne. Ces courbes représentent les bords des objets que nous voulons trouver dans l'iimage. Enfin v est simplement un bruit blanc gaussien. Ce modéle présente des difficultés intéressantes, car il n'y a plus de bords dans l'image dégradée f. L'algorithme utilisé par Le Pennec est hybride, car il mêle l'utilisation des *level sets* de Jean-Michel Morel et Stanley Osher aux ondelettes. Le point de départ est la recherche des bords des objets. Ces bords n'existent plus ; ils ont été effacés. En fait, on cherche seulement une information directionnelle décrivant les lignes parallèles aux bords. C'est ici qu'apparaissent les lignes de niveau de l'éclairement. Plus précisément on cherche les lignes de niveau de l'image lissée (la présence du bruit additif ne permet pas l'utilisation des lignes de niveau stricto sensu). Ensuite Le Pennec construit un difféomorphisme Φ redressant localement les lignes de niveau de l'image lissée. enfin il s'agit de trouver la représentation optimale de fonctions réguliéres, à l'exception d'une discontinuité à travers l'axe horizontal. Quelles ondelettes doit-on choisir pour que ces fonctions aient une série creuse? Le choix adopté par Le Pennec est celui des ondeletets bi-dimensionnelles anisotropes. Ces ondelettes sont les produits tensoriels $\psi_{j,k}(x_1)\psi j', k'(x_2)$ d'ondelettes unidimensionnelles. On tire ainsi parti de la régularité locale de cet 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. Faute de temps, je ne peux parler des *contourlets* de Minh N. Do et Martin Vetterli ; le but à atteindre est le même que pour les bandlets. Je renvoie le lecteur intéressé au site web http://www.ifp.uiuc.edu/˜mindho/publications.

9. Le compressed sensing

Le compressed sensing est le résultat d'une analyse critique des algorithmes qui permettent de coder et de transmettre un signal ou une image. On supposera, une fois pour toutes qu'il existe une base orthonormée **B**, donnée et disponible telle que le développement en série de ce signal dans cette base **B** ne comporte que peu de termes significatifs. Ceci ne préjuge pas de l'emplacement de ces termes dans la série. Ils peuvent par exemple être le 1492-ième, puis le 2727-ième et enfin le 54792-ième. On parle alors d'une représentation creuse et l'intérêt du jeu qui suit disparaîtrait si l'on savait où sont situés ces "grands coefficients". . Pour coder ce signal ou cette image, il convient alors d'effectuer les trois opérations suivantes: (a) calculer TOUS les coefficients de la décomposition dans la base **B**, (b) jeter à la poubelle tous les coefficients inférieurs à un certain seuil, (c) quantifier, coder et transmettre les coefficients restants. Si nous savions à l'avance où sont localisés les coefficients significatifs, le problème de les calculer tous n'existerait pas et il suffirait de faire les mesures là où il faut. Or nous ne savons pas où sont ces coefficients ; la question posée par Candès et Tao est donc la suivante : pourquoi doit-on calculer tant de coefficients pour ensuite les jeter presque tous à la poubelle ? En fait Candès et Tao prouvent qu'il suffit de faire un nombre de mesures du même ordre de grandeur que le nombre de coefficients significatifs, ceci bien-entendu sans savoir où sont disposés ces coefficients. Voici un des énoncés. On considère un signal périodique f de longueur N qui est porté par un ensemble inconnu T de cardinalité $|T|$ donnée. On veut construire f à l'aide des coefficients de Fourier d'indices $k \in \Omega$ où Ω est un ensemble de cardinalité M. Les auteurs montrent que, pour tout entier q, et pour la plupart des choix de l'ensemble aléatoire Ω de fréquences de cardinalité $|\Omega| = M \geq \alpha(q)|T| \log N$, on peut reconstruire exactement f à l'aide d'un programme d'optimisation convexe tournant en temps réel. Ici "la plupart des choix" signifie : avec une probabilité dépassant $1 - O(N^{-q})$. Un énoncé analogue vaut en dimension deux et a d'importantes applications en imagerie médicale. On suppose qu'une image a une forme géométrique telle qu'elle puisse être bien comprimée dans une base adaptée. On ne connaît qu'un petit nombre de coefficients de Fourier de cette image. Est-il possible d'obtenir une reconstruction exacte ? La réponse est oui, grâce aux techniques du compressed sensing.

10. Conclusion

Revenons à la question posée : *Peut-on relier la compression des images digitales à ce qui fonde la perception ?*

Repassons en revue les modèles que nous avons étudiés. Indiquons tout d'abord que nous pourrions chercher à construire des modèles à partir d'un apprentissage sur un corpus d'images formant un *learning set.* De telles études statistiques sont importantes et cette démarche a été évoquée dans la section 6. Le lecteur qui voudrait en savoir plus se reportera aux travaux sur l'ICA mentionnés ci-dessous ou aux travaux de Robert Azencott, de Donald et Stuart Geman, ou d'Alain Trouvé sur l'apprentissage stochastique. Les modèles que nous avons utilisés sont définis par les propriétés que nous imposons aux contours. Les contours sont les bords des objets ; ils les délimitent. Mais ceci n'est qu'un point de vue naïf, car la définition des bords des objets inclus dans une image est souvent un problème en soi. Les contours ne peuvent être de "vrais contours" qui n'existent que dans des exemples académiques. Il suffit, pour s'en convaincre, de penser au "clair-obscur" de Leonard de Vinci. Les contours ne peuvent donc apparaître que dans un sketch, dans une esquisse de l'image. Ceci étant, les modèles que nous avons étudiés diffèrent (a) par les conditions portant sur la régularité des contours utilisés et (b) par les algorithmes qui en assurent le tracé. C'est là que les points de vue divergent, comme nous l'avons vu en étudiant l'algorithme JPEG2000, les "ridgelets" de David Donoho et finalement les "bandlets" de Stéphane Mallat ou les "contourlets" de Do et Vetterli.

On peut rêver à une nouvelle génération d'algorithmes de compression basés sur une analyse directe du contenu sémantique de l'image. On ne passerait pas par l'étape intermédiaire d'extraction des contours. 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 parenté est explicitée chez David Hubel. Dans *Eye, Brain and Vision*, il écrit :

> *Les contours qui séparent les régions sombres et les régions claires sont les paramètres majeurs de notre perception et de notre compréhension du monde visuel...Le cerveau a logiquement évolué de façon à minimiser les nombre de cellules nécessaires au traitement de l'information visuelle. Or la*

> *seule information dont nous ayons besoin concerne les bords d'une forme; ce qui se passe à l'intérieur ne compte pas...*

David Hubel reprend donc à son compte l'hypothèse suivante : *l'évolution du système visuel des mammifères aurait privilégié la compression de l'information.*

Si l'on accepte cette hypothèse, notre perception serait élaborée à partir d'un prétraitement de l'information visuelle et ce prétraitement reviendrait à comprimer cette information. En outre, David Hubel, David Donoho et Stéphane Mallat s'accordent sur la nature de ce prétraitement, puisque, dans les deux cas, il s'agit de construire ou de calculer des "croquis" à l'aide d'un algorithme. Ces croquis ou sketchs seraient la première étape d'un processus conduisant à la perception. Mais, pour l'instant, il est difficile d'aller plus loin et de choisir l'un des modèles de contours en traitement de l'image en se basant seulement sur les découvertes faites par David Hubel.

11. Appendice : Le modèle d'Osher, Rudin et Fatemi

Une image en noir et blanc y est vue comme une fonction 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 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 naturelle. Une quantité vraiment minuscule de telles fonctions provient d'images naturelles 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 1926, par le mathématicien italien Leonida Tonelli (Note aux CRAS, présentée par Jacques Hadamard, le 10 mai 1926). 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 résolu par Lebesgue. Alors le graphe de f a une surface finie si et seulement si f est à variation bornée. On écrit $f \in BV$. La norme de f dans BV est, par définition

$||f||_{BV} = \int |\nabla f(x)|dx$ 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. Le modèle d'Osher, Rudin et Fatemi consiste à définir une image f comme une somme $f = u + v + w$ entre une composante $u \in BV$ qui modélisera les objets contenus dans l'image, une composante v qui modélisera la texture et enfin une composante w rendant compte du bruit. Les termes v et w sont souvent regroupés. La composante v sera, 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 doivent être regardés comme des textures. Les propriétés de cet algorithme sont étudiées en détail dans la thèse d'Ali Haddad. On trouvera ce texte et d'autres sur des problèmes reliés en consultant le dossier "Prépublications" dans le site: `www.cmla.ens-cachan.fr/Cmla/index.html`.

12. Références

Voici des sites web d'où vous pouvez décharger les travaux de recherche liés à cet exposé.

On apprend tout sur les **ondelettes** en consultant le site :

`http://www.cmapx.polytechnique.fr/ ~mallat` ou le beau livre *A wavelet tour of signal processing* par Stéphane Mallat, publié chez Academic Press (1997).

Les travaux d'**Albert Cohen** sont consultables sur le site:

`http://www.ann.jussieu.fr/ ~cohen/`.

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

http:// jj2000.epfl.ch (l'absence de www n'est pas une erreur)

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 **Let It Wave Codec** se trouve sur le site :

http://www.letitwave.fr/.

L'étude des **propriétés statistiques des images naturelles** nous renvoie au site :

http:// www.lps.ens.fr/ ~Jean-Pierre.Nadal.

En ce qui concerne l'**analyse en composantes indépendantes**, on consultera le livre *Independent Component Analysis,* par Aapo Hyvärinen, Juha Karhunen et Erkki Oja, John Wiley & Sons, 2001, ou les sites :

http://sig.enst.fr/ ~cardoso/stuff.html

http://www.math.ucdavis.edu/ ~saito.

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

http://www-stat.stanford.edu/ ~donoho.

Les travaux de **Martin Vetterli** et de **Minh Do** se trouvent dans :

http://www.ifp.uiuc.edu/ ~mindho/publications.

Si l'on s'intéresse aux liens entre la théorie des ondelettes et l'**approximation non-linéaire**, on consultera :

http://www.math.sc.edu/ ~devore.

Les travaux sur le **compressed sensing** se trouvent sur le site web d'Emmanuel Candès :

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

mais on trouvera aussi des documents sur le "compressed sensing" en consultant le site web de David Donoho.

Enfin mes travaux en collaboration avec **Ali Haddad** se trouvent dans le dossier "Prépublications", sur le site :

`http://www.cmla.ens-cachan.fr/Cmla/index.html`.

---

1 novembre 2005

Yves Meyer