

Le traitement des images

tome 1

I. Bloch, Y. Gousseau, H. Maître, D. Matignon,

B. Pesquet-Popescu, F. Schmitt, M. Sigelle, F. Tupin

Polycopié du cours ANIM
Département TSI - Télécom-Paris

version 5.0
8 décembre 2005

Table des matières

1	Vision et perception	13
1.1	Anatomie du système visuel humain	13
1.1.1	L'œil	13
1.1.2	Les voies optiques	14
1.1.3	Les aires supérieures de la vision	14
1.2	Comment percevons nous les images ?	14
1.2.1	Comportement passif du système visuel	15
1.2.2	Comportement actif du système visuel	17
2	Propriétés statistiques des images	25
2.1	Introduction	25
2.1.1	Pourquoi étudier des propriétés statistiques	25
2.1.2	Quelles images	26
2.2	L'amplitude	26
2.2.1	Propriétés	26
2.2.2	Sensibilité aux modifications d'histogrammes	27
2.3	Les sauts de l'amplitude	28
2.4	La fonction d'autocorrélation	29
2.4.1	À une dimension	29
2.4.2	En multi-dimensions	32
2.4.3	Le spectre de densité de puissance	33
2.5	Entropie	34
2.5.1	Entropie d'ordre 0	34
2.5.2	Entropie des sauts	35
2.5.3	Entropie d'ordre N	35
2.5.4	Entropie conditionnelle, redondance	36
2.5.5	Retour sur un paradoxe	36
2.6	Un modèle d'image	37
2.6.1	Processus de Markov Gauss	37
2.6.2	Mais	37

2.7	Critique de la démarche précédente	38
2.7.1	La stationnarité	38
2.7.2	La causalité	38
2.7.3	La monodimensionnalité	38
2.7.4	Vers un modèle complet d'image ?	39
3	L'échantillonnage des images, la représentation fractale	41
3.1	Les signaux monodimensionnels à bande limitée	41
3.1.1	Échantillonnage du signal	42
3.1.2	Reconstruction du signal échantillonné à la fréquence de Nyquist	43
3.1.3	Reconstruction dans le cas de suréchantillonnage	44
3.2	Signaux réels	44
3.2.1	Les limitations physiques du spectre	44
3.2.2	Filtrage du signal	45
3.3	L'extension en dimension N	46
3.4	Le modèle fractal	47
3.4.1	La dimension fractale	48
3.4.2	L'homothétie interne	50
3.4.3	Que faire de la théorie fractale ?	50
4	Représentations discrètes	55
4.1	Introduction	55
4.2	Pavages et maillages	56
4.2.1	Définitions et contraintes	56
4.2.2	Pavages plans réguliers	57
4.2.3	Pavages plans semi-réguliers	57
4.2.4	Dualité pavage / maillage	58
4.3	Topologie discrète	58
4.3.1	Quelques approches	58
4.3.2	Topologie à partir de la notion de voisinage élémentaire	60
4.3.3	Nombre d'Euler : un exemple de caractéristique topologique d'un objet	65
4.4	Représentations géométriques	67
4.4.1	Discrétisation d'une droite continue	68
4.4.2	Caractérisation d'un segment de droite discret	69
4.4.3	Droites analytiques discrètes	70
4.4.4	Cercles discrets	72
4.4.5	Pavage de Voronoï et triangulation de Delaunay	74
4.5	Exemples de structures aléatoires	77
4.6	Fonction distance	78
4.6.1	Définition de distances discrètes	78
4.6.2	Exemples	78

4.6.3	Algorithmes de calcul	79
5	Restauration des images	81
5.1	Le filtrage inverse	81
5.1.1	Approche analytique	82
5.1.2	Approche algébrique	82
5.1.3	Discussion	83
5.1.4	Conclusions (provisoires)	85
5.2	Défauts singuliers et mal conditionnés	85
5.2.1	Défauts singuliers	85
5.2.2	Système mal conditionné	86
5.2.3	Décomposition en valeurs singulières	88
5.3	Restauration des signaux bidimensionnels	88
5.4	La restauration itérative	89
5.4.1	Méthode de Jacobi	90
5.4.2	Méthode de Gauss-Seidel	91
5.4.3	Méthode de plus grande pente	91
5.4.4	Méthode du gradient conjugué	91
5.4.5	Méthodes de projection (POCS)	92
5.5	L'estimation des défauts	92
5.5.1	Détermination par calibrage	93
5.5.2	Détermination statistique	93
5.6	Réduction des effets de bord	93
6	Morphologie mathématique	95
6.1	Introduction et préliminaires	95
6.2	Les quatre opérations	97
6.2.1	Notion d'élément structurant et cadre ensembliste	97
6.2.2	Erosion et dilatation binaires	98
6.2.3	Erosion et dilatation de fonctions	100
6.2.4	Ouverture et fermeture binaires	104
6.2.5	Ouverture et fermeture numériques	105
6.3	Cadre topologique	106
6.4	Cadre algébrique	108
6.4.1	Treillis	108
6.4.2	Erosion et dilatation algébriques, adjonctions	109
6.4.3	Lien avec les opérateurs morphologiques	109
6.4.4	Ouverture et fermeture algébriques	110
6.5	Cadre probabiliste	110
6.5.1	Ensembles fermés aléatoires	110
6.5.2	Un exemple : le schéma booléen	110

6.6	Applications de l'érosion et de la dilatation	111
6.6.1	Mesures	111
6.6.2	Erodé ultime	111
6.6.3	Rehaussement de contraste	112
6.6.4	Gradient morphologique	112
6.7	Applications de l'ouverture et de la fermeture	113
6.7.1	Filtres alternés séquentiels	113
6.7.2	Filtres auto-duaux	113
6.7.3	Chapeau haut-de-forme	114
6.7.4	Granulométries	115
6.7.5	Ouverture surfacique	115
6.7.6	Ouverture annulaire	116
6.8	Transformation en tout ou rien et opérateurs dérivés	116
6.8.1	Transformation en tout ou rien	116
6.8.2	Amincissement et épaissement	116
6.8.3	Squelette	117
6.8.4	Squelette discret	118
6.8.5	SKIZ	118
6.9	Géodésie	119
6.9.1	Distance géodésique et boules géodésiques	119
6.9.2	Opérations morphologiques géodésiques, reconstruction	120
6.9.3	Squelette géodésique par zones d'influence	121
6.10	Ligne de partage des eaux	121
6.10.1	Définition et propriétés	121
6.10.2	Applications en segmentation	122
6.11	Conclusion	123
7	Définition et simulation d'un champ de Markov	127
7.1	Définition et simulation d'un champ de Markov	127
7.1.1	Description de l'image	127
7.1.2	Modélisation probabiliste de l'image	128
7.1.3	Champs de Markov - Champs de Gibbs	129
7.1.4	Echantillonnage de MRF	131
7.1.5	Le recuit simulé	133
7.1.6	Quelques MRF fondamentaux	136
7.1.7	Applications : restauration et segmentation	137
7.2	Estimateurs dans un cadre markovien	141
7.2.1	Introduction	141
7.2.2	Modélisation bayésienne et fonction de coût	141
7.2.3	Estimateur MAP	143

7.2.4	Estimateur MPM	144
7.2.5	Estimateur TPM	145
7.2.6	Comparaison des estimateurs MAP, MPM, et TPM	146
7.3	Estimation des paramètres	148
7.3.1	Introduction	148
7.3.2	Données complètes	149
7.3.3	Données incomplètes	152
7.4	Processus de bords	160
7.4.1	Processus de bords explicites et implicites	160
7.4.2	Algorithmes de minimisation	161
7.5	Quelques applications des champs markoviens	162
7.5.1	Applications sur le graphe des pixels	162
7.5.2	Segmentation	164
7.6	Applications sur des graphes de primitives	167
7.6.1	Graphes de régions : application à la segmentation d'une image d'IRM cérébrale	167
7.6.2	Graphes de segments : application à la détection du réseau routier	168
7.6.3	Conclusion	170
8	Equations aux dérivées partielles et traitement d'images	175
8.1	L'équation de la chaleur et ses limitations	176
8.1.1	Quelques notations	176
8.1.2	Pourquoi l'équation de la chaleur	176
8.1.3	Equation de la chaleur et noyau gaussien	177
8.1.4	Application aux images	177
8.1.5	L'équation de la chaleur inverse	177
8.1.6	Limitations, invariance par changement de contraste	178
8.2	Equations de diffusion non-linéaires	181
8.2.1	Notations et structure locale des images	181
8.2.2	L'équation de Malik et Perona	182
8.2.3	Mouvement par courbure moyenne et variantes	183
8.2.4	L'équation de Rudin-Osher-Fatemi	185
8.2.5	Le filtre de choc de Rudin	186
8.3	EDP et analyse multi-échelles	186
8.3.1	Définition et propriétés des analyses multi-échelles	187
8.3.2	Pourquoi les analyse multi-échelles sont régies par des E.D.P.	189
8.3.3	Classification des analyses multi-échelles	190
9	Ondelettes et traitement d'images	193
9.1	Principes de l'analyse linéaire des images	193
9.2	Trames	197
9.2.1	Objectif	197

9.2.2	Définition	197
9.2.3	Propriétés	198
9.2.4	Trame duale	198
9.2.5	Algorithme des trames	200
9.2.6	Exemples de trames	201
9.3	Poursuite adaptative	202
9.3.1	Motivation	202
9.3.2	Algorithme de poursuite	203
9.3.3	Améliorations	204
10	Les prétraitements	209
10.1	Les traitements photométriques ou colorimétriques	209
10.1.1	Linéarité	209
10.1.2	Homogénéité	209
10.1.3	Binarisation, seuillage	211
10.1.4	Augmentation de contraste	211
10.2	Suppression des bruits	211
10.2.1	Filtrages linéaires	211
10.2.2	Filtrage de rang	213
10.2.3	Filtrages morphologiques	213
10.2.4	Filtrages par équations de diffusion	213
10.3	Les filtres adaptatifs	216
10.3.1	Les filtres à coefficients adaptatifs	216
10.3.2	Les filtres à fenêtres adaptatives	217
10.4	Le rééchantillonnage des images	218
10.4.1	Interpolation avec des polynômes	218
10.4.2	Interpolation par des B-splines	219
10.4.3	Interpolation adaptative	221
11	La détection des contours dans les images	223
11.1	Un modèle continu de contour.	224
11.2	Les approches classiques	225
11.2.1	Les détecteurs de gradient par filtrage	225
11.2.2	Les détecteurs de gradient par masques	226
11.2.3	Pré- et post traitements	227
11.2.4	Les détecteurs de passage par zéro du laplacien	228
11.2.5	Les détecteurs par masquage adapté	229
11.3	Les approches analytiques	230
11.3.1	Les critères de Canny	230
11.3.2	Les filtres de Deriche et Shen et Castan	231
11.3.3	L'extension à 2D	232

11.3.4	Variantes autour des filtres analytiques	233
11.4	Les contours actifs = les <i>snakes</i>	234
11.4.1	Une approche physique des contours	234
11.4.2	Mise en équation des contours actifs	235
11.4.3	Les ensembles de niveaux (<i>level sets</i>)	237
11.4.4	Les modèles géodésiques actifs	238
11.5	La poursuite et la fermeture des contours	238
11.5.1	Méthodes de recherche dans des graphes	239
11.5.2	Les automates	240
12	La segmentation par régions	243
12.1	Les méthodes sur histogramme	243
12.1.1	Avec apprentissage	243
12.1.2	Seuillage sans apprentissage	244
12.1.3	Méthodes de classification	245
12.1.4	Sélection sur l’histogramme et dans l’image	245
12.1.5	Sélection sur histogramme et régularisation	245
12.2	Les méthodes par transformation de régions	247
12.2.1	La croissance de région	248
12.2.2	Le partage de région	248
12.2.3	La réunion de région	249
12.2.4	Les pyramides	250
12.3	Les graphes d’adjacence	250
12.4	La méthode MDL = Minimum Description Length	251
12.5	L’approche de Mumford et Shah	253
12.5.1	Approche formelle	253
12.5.2	Les variantes autour de la formulation de Mumford-Shah	254
13	Les textures	257
13.1	Qu’est ce qu’une texture ?	257
13.1.1	Distribution aléatoire ou régulière ?	258
13.1.2	Les échelles des textures	259
13.1.3	Analyse ou synthèse	259
13.2	Modèles de texture	259
13.2.1	Un modèle biologiquement plausible	259
13.2.2	Modèles stochastiques	260
13.3	Analyse et reconnaissance de textures	260
13.3.1	Schéma général	260
13.3.2	Approches par mesures statistiques	261
13.3.3	La détection d’énergie dans des canaux fréquentiels	265
13.3.4	Les filtres optimisés	266

13.3.5	Les modélisations autorégressives	267
13.4	Les approches par champs markoviens	267
13.4.1	La méthode de Manjunath et Chellappa	268
13.4.2	La méthode de Kervrann et Heitz	269
13.5	Les méthodes structurales	270
13.6	Textures hétérogènes	270
14	Description de contours et de formes	271
14.1	Fonction caractéristique	271
14.2	Description de formes	273
14.2.1	Représentation par les moments	273
14.2.2	Représentation par les moments invariants	273
14.2.3	Boîtes englobantes et boîtes minimales	274
14.3	Polygones de Guzman	274
14.4	Chaînes de Freeman	275
14.4.1	Définition	276
14.4.2	Les propriétés des chaînes de Freeman	277
14.4.3	Reconnaissance des formes par des chaînes de Freeman	279
14.5	Descripteurs de Fourier	280
14.5.1	Descripteur par tangente	280
14.5.2	Représentation complexe	280
14.6	Approximations polynomiales	281
14.6.1	Approximation d'un nuage de points par une droite unique	282
14.6.2	Approximations polygonales, simplification de contours polygonaux	285
14.6.3	Approximation par des splines	286
14.7	Transformation de Hough	288
14.7.1	Définitions	288
14.7.2	Mise en œuvre de la TH	290
14.7.3	Détections par TH	292
14.8	Conclusion	293
15	Éléments de géométrie tridimensionnelle	295
15.1	Géométrie projective	295
15.1.1	Espace projectif et coordonnées homogènes	295
15.1.2	Espaces projectifs \mathcal{P}^1 , \mathcal{P}^2 et \mathcal{P}^3	296
15.1.3	Expression linéaire de problèmes géométriques	297
15.1.4	Stratification, groupes de transformations et invariants	298
15.2	Représentations des rotations de \mathbf{R}^3	300
15.2.1	Les angles d'Euler	300
15.2.2	Exponentielle d'une matrice anti-symétrique	301
15.2.3	Forme canonique et théorème de Rodriguez	302

15.2.4	Quaternions	302
15.3	Géométrie différentielle	305
15.3.1	Courbes, repère de Frenet	305
15.3.2	Courbes sur une surface, formes fondamentales	306
15.3.3	Courbures sur des surfaces	308
15.3.4	Une notation fréquente	309
15.3.5	Interprétations géométriques	309
15.3.6	Principales méthodes de calcul des courbures	311
15.3.7	Géométrie différentielle dans des volumes numériques	313
16	Formation de l'image et couleur	315
16.1	Formation des images	315
16.1.1	Imagerie cohérente et incohérente	315
16.1.2	Interaction matière rayonnement	317
16.1.3	Les sources de lumière	320
16.2	Définition des différents espaces couleur	320
16.2.1	Trivariance et primaires	320
16.2.2	Choix d'un espace de couleur	321
16.2.3	L'espace RVB de la CIE 1931	322
16.2.4	L'espace XYZ	324
16.2.5	L'espace Lab et le repère TLC	325
16.2.6	L'espace YC1C2	327
16.2.7	L'espace TLS	327
16.2.8	Le système visuel hypothétique	328
16.2.9	L'espace TLC	330
16.2.10	Les espaces de la télévision	330
16.3	Modélisation de la perception colorée des images	331
16.4	Les traitements des images en couleur	333
16.4.1	Un ordre pour l'espace R^3 ?	333
16.4.2	Qu'est ce qu'une couleur représentative d'un nuage ?	335
16.4.3	Comment calculer une palette ?	335
16.4.4	Comment filtrer 2 couleurs ?	336
16.4.5	La détection des contours des images en couleur	336
16.4.6	Représentation par quaternions	337
16.4.7	Couleur et quaternion	337
A	Décomposition en Valeurs Singulières : SVD	339
A.1	Définition	339
A.2	Lien avec le rang et interprétation de la décomposition	339
A.3	Normes	340

B Applications de la SVD à la résolution des systèmes linéaires	341
B.1 Conditionnement d'un système linéaire non singulier	341
B.2 Pseudo-inverse	341
B.3 Solutions robustes d'un système linéaire, moindres carrés	342
C Optimisation sous contraintes	345
C.1 Contraintes égalités	345
C.2 Contraintes inégalités	345

Chapitre 1

La vision et la perception humaines

Chapitre rédigé par Francis SCHMITT

1.1 Anatomie du système visuel humain

1.1.1 L'œil

L'œil est le récepteur de la vision [LeGrand, 1956, Pirenne et Crouzy, 1972]. Il reçoit les rayons lumineux, en forme une image et transmet celle-ci au cerveau par le nerf optique, cf. figure 1.1.

L'image est formée à partir du cristallin qui joue le rôle d'une lentille de focale variable, sur la rétine, couche photosensible qui tapisse le fond de l'œil (surface à peu près sphérique). L'iris contrôle le diamètre de la pupille, comme un diaphragme automatique contrôlant l'énergie arrivant sur la rétine. Les muscles oculomoteurs assurent l'orientation de l'œil et la convergence des deux yeux sur le point d'intérêt. Convergence et focalisation sont généralement couplées dans la vision réflexe.

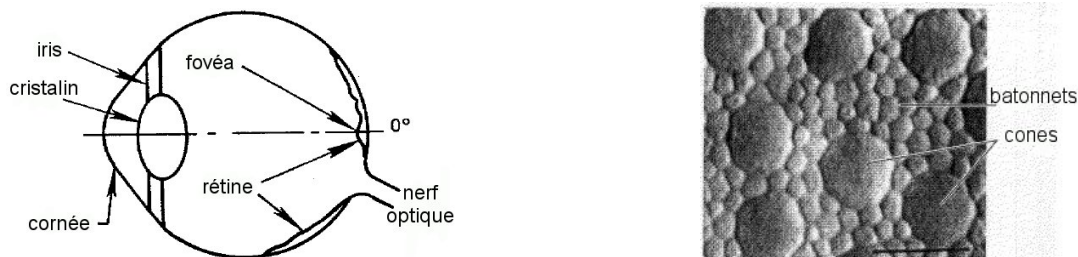


FIG. 1.1 – Schéma de l'œil humain et photographie du fond de la rétine.

Le champ visuel n'est pas circulaire, il est de 90° du côté temporal, de 60° du côté nasal, de 50° et 80° vers le haut et vers le bas.

La rétine est le transducteur qui transforme le signal optique en signaux électriques. La surface de la rétine est très hétérogène, aussi bien en nature de récepteurs qu'en densité, cf. figure 1.1. Cette hétérogénéité permet à des zones différentes d'assumer des fonctions complémentaires.

Le centre de la rétine assure la vision colorée et la vision des détails à fort niveau d'éclairement (vision photopique). Le pourtour assure la vision achromatique à faible niveau de lumière (vision scotopique), la vision des

mouvements rapides, la perception globale des scènes et la vision achromatique à faible niveau de lumière (vision scotopique). L'acuité visuelle est maximale sur la fovéa (zone de la rétine correspondant à l'axe du regard). La zone d'où part le nerf optique ne capte pas la lumière (elle est appelée tache aveugle).

La rétine est constituée de nombreux types de cellules nerveuses : les cellules photo-réceptrices (cônes et bâtonnets), les cellules bipolaires, ganglionnaires et amacrines. Ces cellules constituent 2 voies croisées : la voie directe et la voie indirecte, interconnectées de façon complexe. Dans la fovéa, la rétine adopte un schéma beaucoup plus simple : il n'y a que des cônes (et pas de bâtonnets) et chaque cône n'est connecté qu'à une cellule bipolaire elle-même connectée à une seule cellule ganglionnaire.

L'acuité visuelle baisse régulièrement lorsqu'on s'éloigne de la fovéa.

Les cônes sont principalement concentrés dans la fovéa et leur densité décroît très rapidement vers le pourtour de l'œil. Ils sont au nombre de 6 à 7 millions. Dans la fovéa ils sont très petits (de 1 à 2 microns) et très serrés (entre 2 et 10 microns). Les cônes sont peu sensibles, mais une même fibre nerveuse ne regroupe qu'un tout petit nombre de cônes, assurant une bonne discrimination angulaire. Dans toutes les opérations de traitement d'image, on considère que l'image est observée en vision par les cônes. La vision chromatique est assurée par 3 types de cônes dont la sensibilité diffère selon le type de pigment qu'ils contiennent. Ces cônes sont appelés L, M et S en fonction des longueurs d'ondes auxquels ils ont leur pic de sensibilité : L= *Long* autour du jaune vert (585 nm), M= *Middle* autour du vert (550 nm) et S= *Short* autour du bleu (430 nm).

Les bâtonnets sont au contraire plus densément localisés sur une couronne centrée à 20° de la fovéa environ. Leur densité est nulle à la fovéa mais demeure très forte encore à plus de 60° de celle-ci. Ils sont très nombreux (de 80 à 150 millions), d'un diamètre moyen de 3 microns. Ils ont une très bonne réponse temporelle (c'est en périphérie de l'œil que passent les étoiles filantes), une très bonne sensibilité (ils assurent la vision nocturne - scotopique). Ils ont tous la même réponse spectrale (à peu près indépendante de la longueur d'onde) (« la nuit tous les chats sont gris. »).

1.1.2 Les voies optiques

Elles sont constituées de l'ensemble des cellules nerveuses qui partent de la rétine et aboutissent au cortex visuel primaire.

Le nerf optique tout d'abord part de l'œil et aboutit au chiasma optique, une structure où se croise une partie et se mélange une autre partie des fibres des nerfs optiques issus de chacun des yeux.

Sorties du chiasma optique les fibres se distribuent soit vers le Tubercule Quadrijumeau Antérieur soit vers le Corps Genouillé Latéral puis ensuite vers les aires du Cortex Visuel Primaire situées à l'arrière des lobes occipitaux.

(... à compléter ...)

1.1.3 Les aires supérieures de la vision

(... à compléter ...)

1.2 Comment percevons nous les images ?

Ce n'est pas en général le genre de question que l'on se pose lorsque l'on est face à un écran de télévision ou de cinéma ou lorsque l'on est plongé dans des bandes dessinées. Le système visuel fournit déjà à notre cerveau une quantité d'information suffisamment grande à traiter pour qu'on ne se pose pas encore le problème supplémentaire de savoir comment tout cela fonctionne. Ça marche et ça marche même bien ; merci dame Nature ! Ce type de préoccupation est en fait laissé au soin de diverses espèces de coupeur de cheveux en quatre : psychologues, physiologistes et autres individus comme ces ingénieurs qui traitent des images.

Sachant que l'image est un des supports physiques les plus privilégiés pour transmettre un message à notre cerveau, interrogeons-nous donc sur la manière dont ce message est transmis, décodé et synthétisé par le système nerveux. C'est une question extrêmement complexe qui n'est encore que très partiellement élucidée et qu'on va essayer de débroussailler grossièrement. Les divers types d'approche du phénomène de vision sont comparés dans [Gordon, 1998]. Il n'est pas question de faire ici un large tour d'horizon de l'état de nos connaissances sur le comportement du système visuel face à l'information contenue dans les images. Tout au plus allons-nous essayer de donner une vague vue synthétique qui n'évitera naturellement pas l'écueil d'une schématisation excessive. Le but, en fait, est de sensibiliser le lecteur sur le phénomène de la perception visuelle. Nous espérons que nous aurons ainsi l'occasion de lui procurer quelques sujets de distraction et de réflexion. Il pourra également se référer à [Gregory, 1966], [Cornsweet, 1973], [Murch, 1973] ou [Carterette et Friedman, 1975].

On peut distinguer dans le traitement de l'information « image » par notre système visuel deux procédures distinctes : une qui correspondrait à un comportement passif du système et l'autre à un comportement actif. Nous allons voir successivement à quoi correspondent ces deux aspects, sans cependant insister trop sur le premier qui est un tantinet technique et qui pourrait par là paraître un peu trop rébarbatif.

1.2.1 Comportement passif du système visuel

C'est ce comportement qu'étudient essentiellement les physiologistes et les ingénieurs qui mettent au point de nouveaux systèmes de traitement d'image. Ces derniers personnages ont besoin de bien connaître les propriétés statiques et dynamiques du système visuel pour pouvoir adapter, en fonction de ces contraintes, les caractéristiques physiques de l'image comme sa dimension, sa résolution, sa luminosité, son contraste... L'étude de ces propriétés se fait par des expériences de psychophysique : un individu placé dans des conditions d'observation parfaitement spécifiques regarde un signal donné, appelé stimulus, pouvant prendre des formes très diverses suivant la nature du phénomène étudié. Le problème est de relier les grandeurs physiques de ce stimulus aux grandeurs perceptives de la réponse qu'il engendre chez l'observateur. Dans la pratique on distingue deux types d'expériences : celles de détection et celles de perception. La réponse du système visuel est en effet différente suivant que l'on met en jeu de petites ou de grandes variations du signal. Lorsque celles-ci sont faibles, l'important est de savoir si elles sont au-dessus ou en-dessous du seuil de détection du système visuel. L'observateur aura par conséquent vu ou non la variation du signal et pourra donner une réponse assez objective sans trop se créer d'angoisses métaphysiques. Dans les expériences de perception où les variations du signal sont très grandes, le problème est plus délicat.

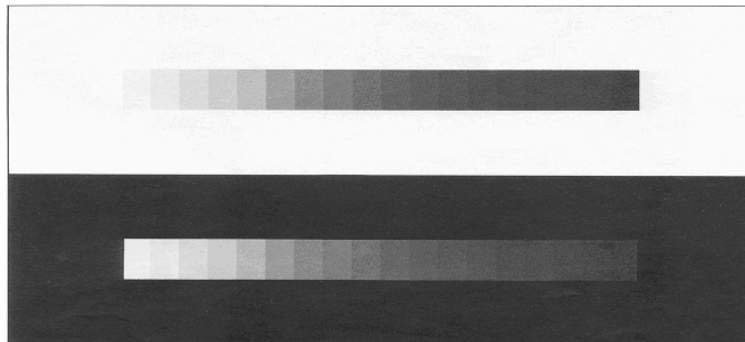


FIG. 1.2 – La perception des différentes gradations d'une échelle de gris dépend du niveau lumineux du fond.

Illustrons cela par un exemple et demandez-vous comment vous procéderiez pour fabriquer une échelle de gris d'une dizaine de tons allant du blanc au noir et dont les écarts de luminosité vous paraîtraient le plus homogène possible, d'un point de vue perceptif bien sûr. Vous vous apercevriez alors que cet exercice vous demande des processus intellectuels beaucoup plus élaborés que celui de fournir une réponse à la simple question « vu ou pas vu ? ». Il

vous faudra entre autres définir un critère de jugement qui vous permettra après entraînement de mieux sélectionner vos gris. Cela souligne l'aspect plus subjectif de telles expériences de perception comparées à celles permettant de mesurer les seuils de détection. La figure 1.2 montre l'importance des conditions selon lesquelles on opère et qu'il est nécessaire de bien contrôler : en changeant la luminosité du fond on modifie sérieusement la sensation que procure une échelle de gris déterminée. Le comportement passif du système visuel s'explique globalement par le fait que c'est essentiellement un système biophysique. Ses propriétés découlent des caractéristiques de sa structure, à commencer par l'œil et la rétine, pour finir avec l'immense complexité de ses connexions nerveuses (cf. section 1.1.2). Ce comportement peut donner lieu à quelques phénomènes particulièrement démonstratifs comme ceux des figures 1.2 à 1.5, dûs aux interactions spatiales entre les régions de l'image ayant des luminosités différentes. Ces exemples indiquent à quel point l'information « image » une fois prise en charge par le système visuel peut être rapidement transformée et déformée. *indexcontraste simultané*

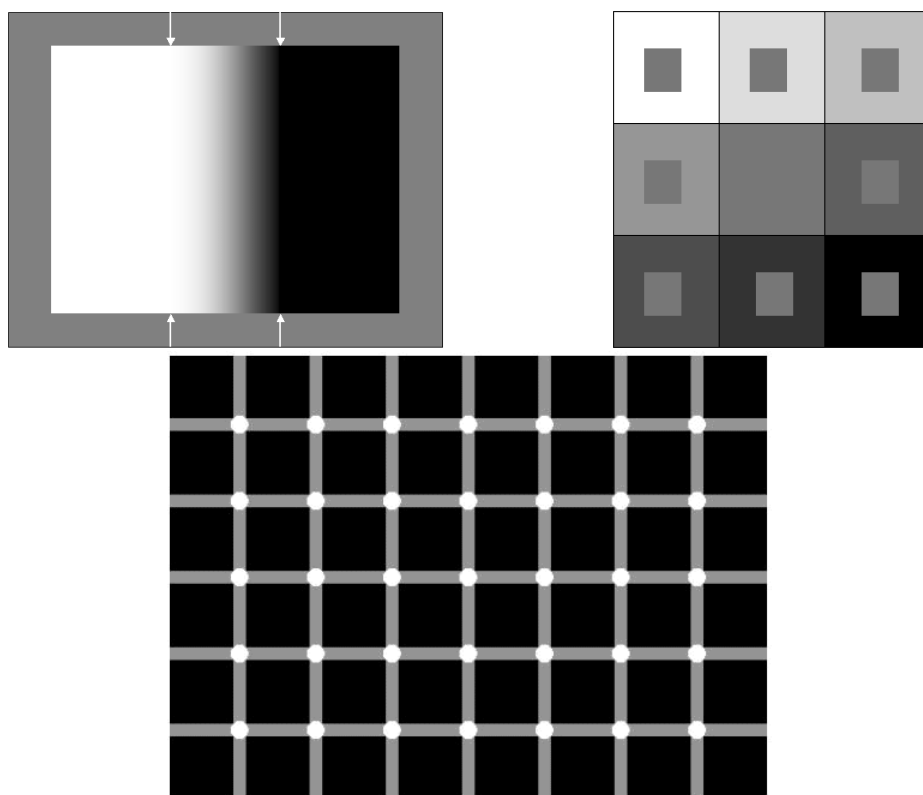


FIG. 1.3 – Phénomènes dus aux interactions spatiales dans l'image.

En haut à gauche une bande de Mach (c'est à dire un renforcement local du contraste) apparaît à la fin d'une variation spatiale continue de la luminance et est plus claire que le fond blanc sur lequel elle se situe.

En haut à droite, le contraste simultané est un exemple classique de l'interaction de la couleur d'un fond sur celle d'un médaillon : les neuf petits carrés ont le même niveau de gris.

En bas, le centre des croisements de la grille blanche d'Hermann paraît plus sombre excepté celui que l'on fixe du regard : la partie centrale de la rétine ne traite pas l'image de la même manière que sa partie périphérique.

A la figure 1.2 les plages dans l'échelle de gris ne sont pas perçues comme uniformes alors qu'elles le sont physiquement. Les trois phénomènes de la figure 1.3, trouvent leur origine aux niveaux des interconnexions nerveuses des cellules rétinienne qui effectuent un véritable filtrage linéaire. Mais dès qu'intervient un contour, des phénomènes non linéaires (figure 1.4) peuvent se produire. Les contours dans certains cas peuvent procurer des

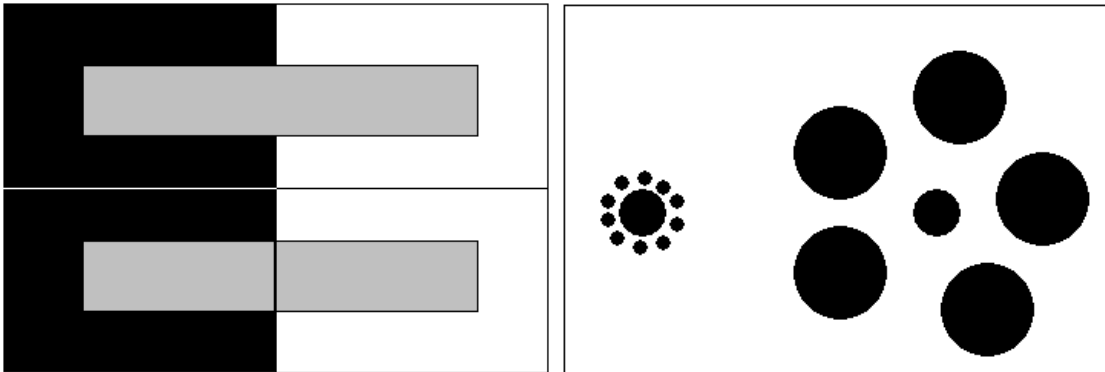


FIG. 1.4 – Phénomènes dus aux interactions spatiales dans l'image. À gauche, la seule adjonction d'un contour au milieu de la barre grise au motif du haut permet d'obtenir un net contraste simultané (motif du bas). À droite, dans l'illusion de Titchener les deux cercles centraux sont de taille égale.

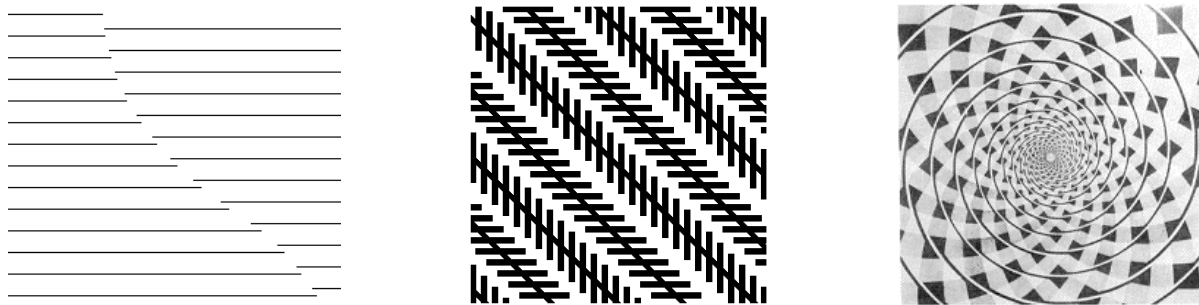


FIG. 1.5 – Phénomènes dus aux interactions spatiales dans l'image. Les contours vrais ou faux (à gauche) jouent par leur orientation un rôle primordial dans certains effets : les diagonales dans l'illusion de Zöllner (au centre) sont parallèles et la spirale de Snearl (à droite) n'est formée... que de cercles concentriques.

résultats assez spectaculaires sur la figure 1.5. Leur explication n'est pas aussi claire que dans le cas de la figure 1.3 à gauche, mais se situerait cependant au niveau de la structure neurophysiologique des voies visuelles où des cellules extrêmement spécialisées dans le traitement des contours ont été découvertes. Mais il existe des illusions notamment sur la dimension des formes (figure 1.4 à droite), pour lesquelles on n'est plus en mesure de donner une explication d'ordre neurophysiologique. Elles correspondraient à des processus mentaux beaucoup plus complexes. Le système visuel ne jouerait déjà plus dans ce cas un rôle aussi passif.

1.2.2 Comportement actif du système visuel

Le rôle actif du système visuel se situe essentiellement au niveau des aires supérieures du cerveau où l'information image arrive après toute une série de transformations et de codages que nous n'avons fait qu'entrevoir. Là, l'information y est rassemblée et synthétisée par des opérations mentales très nombreuses mais encore inconnues. Regardez le trapèze élémentaire de la figure 1.7-A. Suivant le contexte spatial où on le place (B ou C) on en modifie radicalement la perception puisqu'on aura tendance à le voir en position debout ou allongée.

La présence d'éléments de perspective ou de gradients de texture donne en effet de fortes indications sur les distances relatives des objets et permet à l'observateur de mieux structurer l'espace qu'il regarde. Vous constaterez ainsi que les déformations planes dans la figure 1.8 sont finalement interprétées comme un volume caché sous

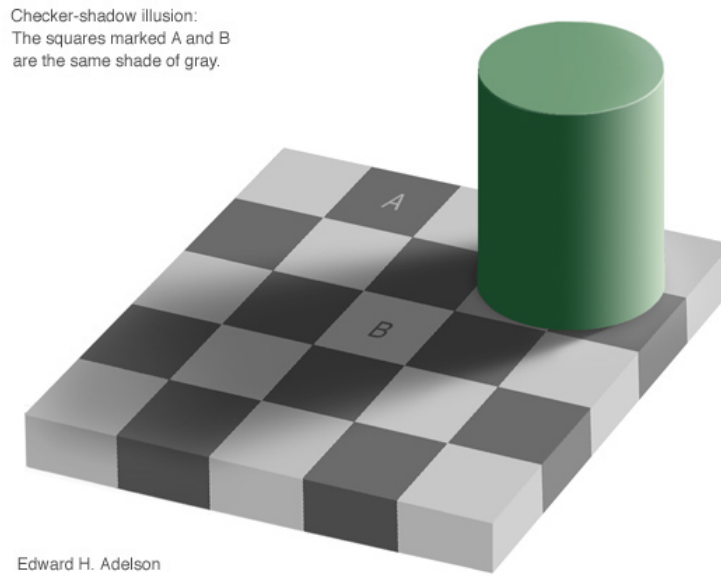


FIG. 1.6 – Dans cet échiquier, les deux carrés marqués A et B ont le même niveau de gris. Plusieurs phénomènes se conjuguent pour renforcer cette illusion. certains relèvent de processus de contraste simultané, d'autres de phénomènes d'interprétation de la scène. © E. H. Adelson.

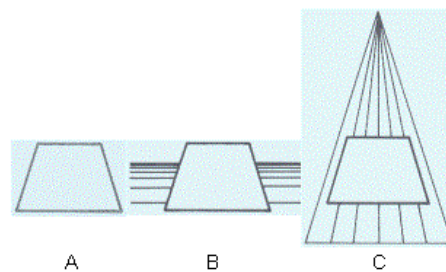


FIG. 1.7 – La perception des formes est modifiée selon le contexte spatial où elle est située.

un damier régulier. Ces indications peuvent aussi agir sur des mécanismes mentaux particuliers comme celui de la constance de la taille et provoquer de fortes illusions géométriques comme la figure 1.9. On y vérifie pourtant que les trois motifs sont absolument identiques et que par conséquent les images rétinienne qu'ils forment ont la même taille.

Ces exemples soulignent le caractère actif avec lequel le cerveau reconstruit l'information image mais n'indiquent pas la manière avec laquelle il procède pour y parvenir. Comment faisons-nous pour reconnaître les formes, analyser les scènes ? Actuellement ces questions n'ont pas encore trouvé de réponses satisfaisantes.

Quelques approches par le biais de la théorie de l'information ont été tentées, mais semblent être restées jusqu'ici stériles. En fait une constatation toute simple permet de montrer combien il est difficile d'attaquer le problème de cette manière. La photographie 1.10 à gauche n'est rien d'autre que le positif de la photo de l'auteur, dont la figure de droite est le négatif. Or, il est extrêmement difficile de reconnaître une personne sur un négatif sans un entraînement préalable. Et cependant la quantité d'information contenue dans un positif ou un négatif est

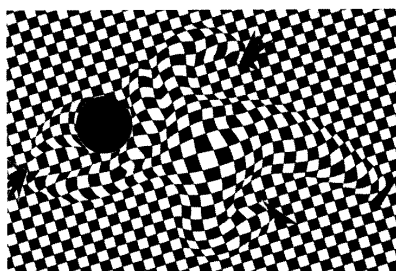


FIG. 1.8 – Rôle des gradients de texture dans l'« Arlequin » de Vasarely (cf. texte).

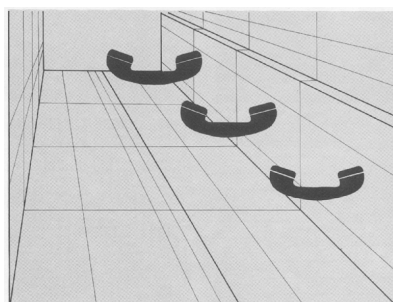


FIG. 1.9 – Illusion géométrique sur la taille provoquée par la perspective.

la même.

L'approche adoptée par les gestaltistes ([Wertheimer, 1944] ou [Gordon, 1998] chapitre 3), très différente, a donné de meilleurs résultats. Pour eux, la perception des signes, des objets ne s'expliquerait pas par un processus de concaténation ou de sommation en série de leurs différentes parties. Ils seraient perçus comme un tout. Regardez les 3 segments de la figure 1.11. Lorsqu'ils finissent par se rapprocher suffisamment pour se rejoindre, ils forment une figure triangulaire que l'on perçoit en soi plutôt que comme le résultat de ce rapprochement. Cette construction suivrait un certain nombre de lois indiquées à la figure 1.12. Ces lois poussent à croire qu'il existerait une sorte de principe de simplicité qui dirigerait le cerveau et lui permettrait de limiter au maximum ses efforts pour reconstruire l'information. C'est une idée un peu naïve, mais prenez par exemple le cas des surfaces subjectives de la figure 1.13. Le triangle blanc à droite n'a pas de contour réel pourtant on le perçoit très vivement. Il paraît même être plus lumineux que le fond. Or, qu'y a-t-il dans cette image ? Trois angles et trois disques tronqués. Ces formes paraissent incomplètes au cerveau. Il lui suffit alors de faire l'hypothèse supplémentaire de la présence d'un triangle blanc recouvrant ces formes pour que l'image s'organise plus simplement.

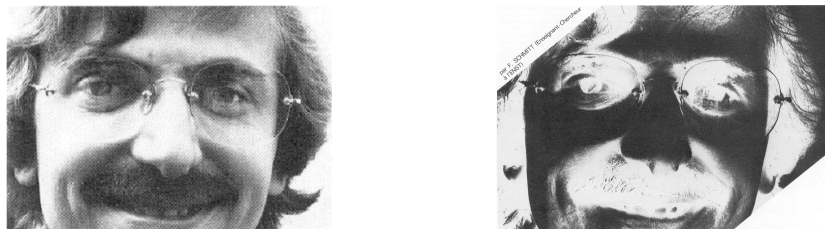


FIG. 1.10 – Positif et négatif de la photo de l'auteur. Vous pouvez vérifier.

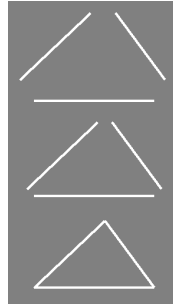


FIG. 1.11 – Le triangle est perçu comme une unité graphique en soi.

Vous constatez que le cerveau ne se contente pas de percevoir ce qui existe dans le signal. Il n'hésite pas à le compléter si cela l'arrange, c'est-à-dire si cela lui permet de faire son travail de synthèse à moindre frais. Ce que le cerveau rajoute au signal pour effectuer ces opérations constructives dépend bien sûr de l'image mais aussi de la mémoire. Plus la mémoire pourra apporter d'éléments significatifs pour comprendre l'image regardée et moins vous aurez besoin de signal pour percevoir l'information. C'est vrai pour la lecture. Lorsque vous appreniez à lire vous étiez obligé de détailler chaque caractère. Maintenant ceux-ci défilent à toute vitesse devant vos yeux.

Il est des cas où le cerveau peut construire pour une image donnée plusieurs solutions distinctes. Ces solutions s'excluent mutuellement car on ne perçoit que l'une des deux et le passage de l'une à l'autre demande un important effort intellectuel. La figure 1.14 est un exemple célèbre de ces figures ambiguës. Certains reconnaîtront une vieille femme et d'autres une jeune femme fort élégante. Pourquoi ? Il est difficile de répondre. Peut-être les amoureux verront plus facilement la jeune femme alors que les maris tourmentés par leur belle-mère verront plutôt la plus âgée. De nombreux facteurs psychologiques interviennent et la mémoire joue encore ici un rôle prépondérant surtout si on la stimule sélectivement, par exemple en vous racontant au préalable des histoires de vieilles rombières ou au contraire des histoires de belles romantiques.

Inversement avec les images en fausse perspective de la figure 1.15 à gauche, le cerveau ne parvient pas à construire de solution cohérente ce qui provoque même une certaine gêne mentale. À la photo 1.15, à droite, cette gêne diminue et même disparaît. Car il faut réfléchir pour s'apercevoir que les escaliers de cette maison sont d'un genre spécial. C'est le caractère réaliste de la photographie avec ses multitudes de petits détails qui permet au cerveau de s'affranchir de cette construction impossible.

Nous terminerons en signalant que le rôle actif du système visuel se manifeste aussi au niveau de l'analyse physique de l'image, c'est-à-dire dans la manière avec laquelle vous la regardez. Peut-être n'en avez-vous pas vraiment conscience, mais vos yeux bougent beaucoup et pas n'importe comment. Yarbus a enregistré leurs mouvements [Kolers, 1972]. La figure 1.16 en haut montre comment dans une première étape l'axe visuel du regard balaye l'image en suivant très approximativement les contours principaux. Puis dans une deuxième phase, le regard se stabilise sur des points particuliers où l'information est très importante comme les yeux et la bouche d'un visage ou tout autre détail inhabituel. Mais ce qui est particulièrement intéressant c'est de savoir que ce type de « lecture » varie énormément suivant les préoccupations de l'observateur.

Sur la figure 1.16 en bas ont été enregistrées différentes lectures de la photographie figurant en haut à gauche et faites par un même individu. On a simplement posé à ce dernier une question précise correspondant à la scène représentée par cette photographie en lui demandant d'y répondre après quelques secondes d'observation. On peut remarquer alors qu'il va chercher l'information là où elle est la plus significative pour résoudre son problème. Enfin, que tout cela ne vous tourmente pas. Dans deux minutes vous aurez complètement oublié que votre système visuel joue un rôle passif ou actif et vous continuerez à regarder autour de vous comme si de rien n'était.

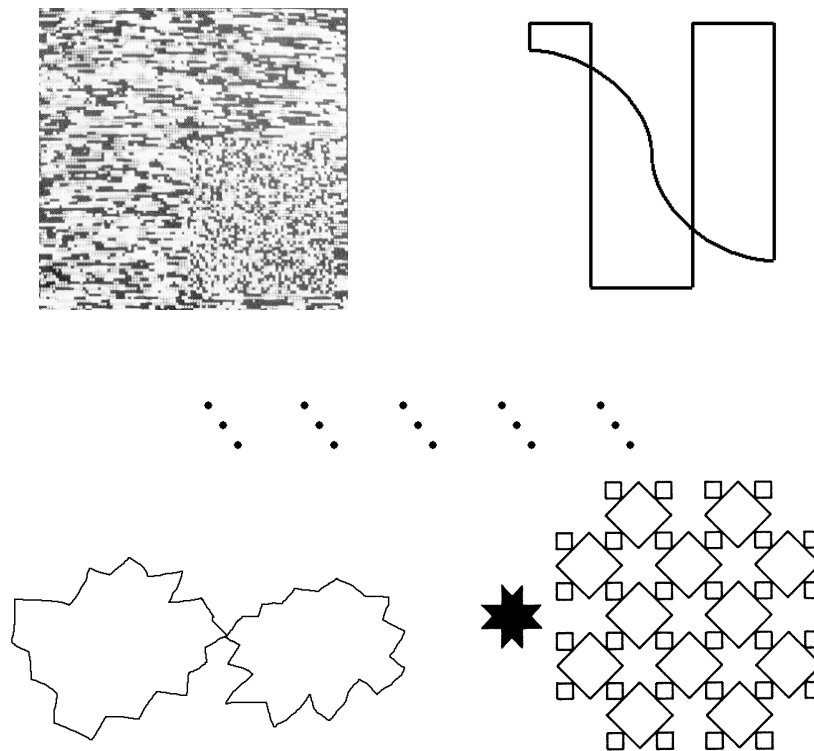


FIG. 1.12 – Lois de la perception (d'après les gestaltistes).

En haut à gauche - lois de similitude : on regroupe les zones ayant des caractéristiques similaires.

En haut à droite - loi de continuité : au lieu de percevoir 3 formes côte à côte, on préfère voir une ligne courbe et une ligne rectangulaire superposée.

Au centre - loi de proximité : on perçoit plutôt 5 triplets de 3 points que 3 lignes de 5 points.

En bas à gauche - loi de fermeture : on perçoit deux formes fermées se touchant en un point plutôt que deux courbes quelconques se croisant.

En bas à droite - le fait qu'on perçoive deux trames de carrés imbriquées, plutôt que la simple juxtaposition de l'élément indiqué en noir est une conséquence de ces lois.

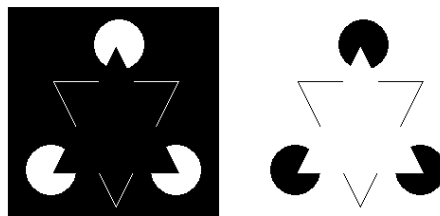


FIG. 1.13 – Deux triangles subjectifs qui n'ont pas de contours (d'après G. Kanizsa, Scientific American, avril 76).



FIG. 1.14 – « Jeune fille ou vieille femme ? » figure ambiguë de R. Leeper.

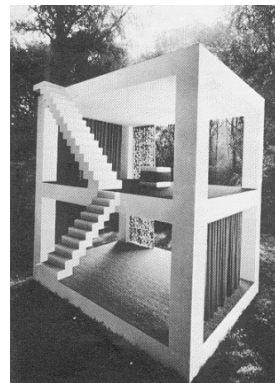
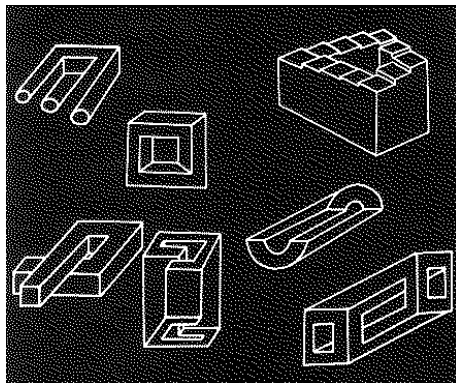


FIG. 1.15 – Quelques constructions impossibles (à gauche), ... qui n'ont pas arrêté cet architecte (photographie A.P. Lamothe), à droite.

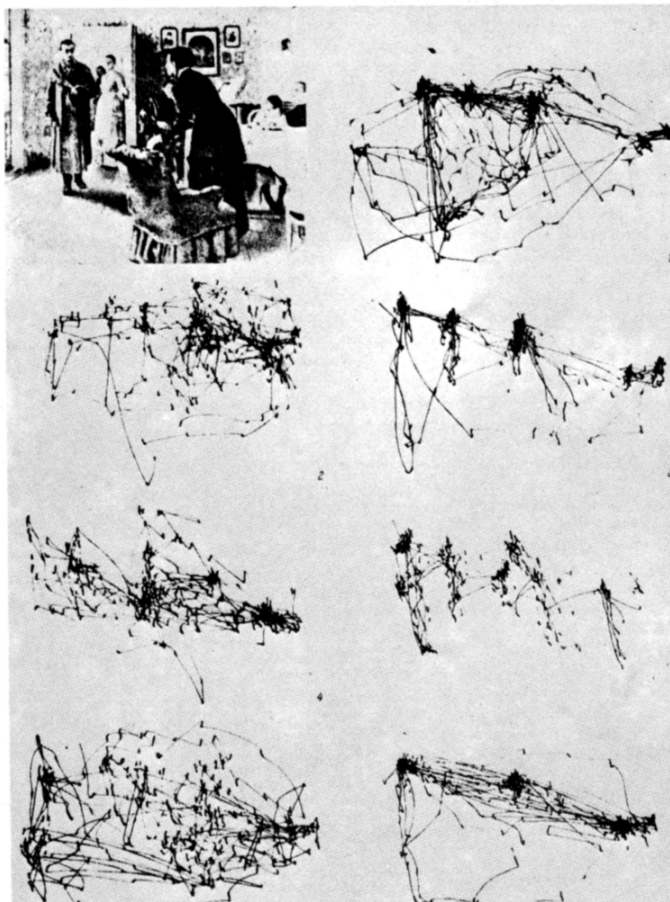
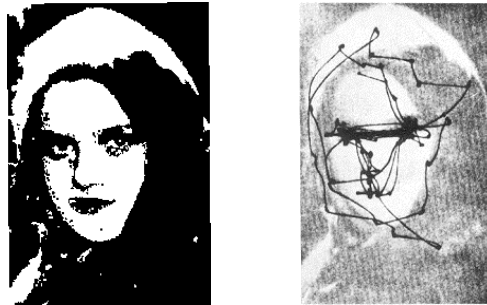


FIG. 1.16 – Chemin que suit l'axe du regard (d'après Yarbus et Kolers [Kolers, 1972]) : en haut, lorsqu'on observe une image ; en bas lorsqu'on demande à l'observateur de se souvenir de diverses choses concernant l'image.

Chapitre 2

Propriétés statistiques des images

Chapitre rédigé par Henri MAÎTRE

2.1 Introduction

2.1.1 Pourquoi étudier des propriétés statistiques

L'analyse des propriétés statistiques des images est dictée par le souci d'adapter des traitements ultérieurs comme le filtrage, la restauration, le codage ou la reconnaissance des formes au signal d'image. Les techniques de base mises en œuvre pour supprimer un bruit ou pour rehausser un signal affaibli reposent toutes sur des hypothèses sur ce qu'est le signal et ce qu'est le bruit, c'est-à-dire sur des modélisations du signal et du bruit. Les techniques statistiques, reprises par les méthodes de traitement de signal, font ainsi abondamment l'hypothèse que les signaux sont gaussiens ou uniformément répartis dans un intervalle par exemple et les bruits blancs, c'est-à-dire à corrélation microscopique. De ces modèles, on peut déduire par des enchaînements rigoureux l'optimalité d'un filtre ou l'estimation de ses performances, en moyenne ou dans le pire des cas. La validité ou la qualité du filtre du codeur ou du détecteur dépendra en pratique de l'adéquation des hypothèses qui sont à la base de son calcul et des propriétés présentées par le signal réellement traité.

On a donc très tôt mesuré les propriétés statistiques des images, et essayé d'en déduire des modèles permettant d'expliquer les images. Ces modélisations statistiques ont cependant très peu d'écho dans notre vie quotidienne car elles ont beaucoup moins d'utilisation que des descriptions « structurelles » ou « syntaxiques » qui donnent de chaque image une représentation à base de langage naturel et de géométrie, comme par exemple : *une femme au sourire énigmatique, les bras croisés sur la poitrine devant un paysage de montagne.*

Dans l'objectif d'un traitement automatique par ordinateur, la représentation statistique est cependant très utile, car elle nourrit immédiatement des algorithmes plus ou moins élaborés manipulant les pixels, composants élémentaires de l'image. L'universalité de codeurs comme JPEG ou MPEG, qui reposent beaucoup sur ce type de propriétés, atteste de l'intérêt technique de ces études.

Notre démarche sera la suivante : dans un premier temps nous passerons en revue les mesures statistiques les plus simples, celles qui sont obtenues par balayage mono-dimensionnel de l'image. Nous en déduirons un modèle, assez universellement utilisé. Nous en montrerons les limites. Nous présenterons quelques propriétés bidimensionnelles mal couvertes par ce modèle. Nous suggérerons des améliorations du modèle de base.

2.1.2 Quelles images

Une étude expérimentale doit distinguer deux types d'images très différents :

- les images **incohérentes** (c'est le cas des images vidéo ainsi que de la plupart des photos satellitaires ou médicales) obtenues avec les sources de lumière naturelle ou des rayonnement incohérents,
- les images **cohérentes**, telles que celles obtenues par holographie, par imagerie radar ou ultrasonore, en imagerie sismique.

Les premières sont obtenues par la sommation **en énergie** des radiations émises par les divers objets associés dans un pixel élémentaire.

Les secondes sont formées de la sommation **en amplitude complexe** des rayonnements issus des objets constituant un même pixel. Ces rayonnements, selon qu'ils sont en phase ou en opposition de phase, créeront un pixel clair ou sombre, ces variations dépendant moins des objets qui constituent le pixel que des conditions de propagation, les termes de phase dépendant notamment de la géométrie d'observation. Ces images, compte tenu de l'état de surface souvent grossier des objets à l'échelle de la longueur d'onde incidente, sont généralement entachées d'un très important bruit appelé chatoiement (en radar), ou tavelure (en astronomie), ou granularité (en optique) et *speckle* en anglais. Les propriétés statistiques des images cohérentes seront toujours très différentes de celles des images incohérentes. Ce chapitre traite des images incohérentes, les plus fréquentes. Les propriétés des images cohérentes seront cependant rappelées lorsqu'elles sont bien connues.

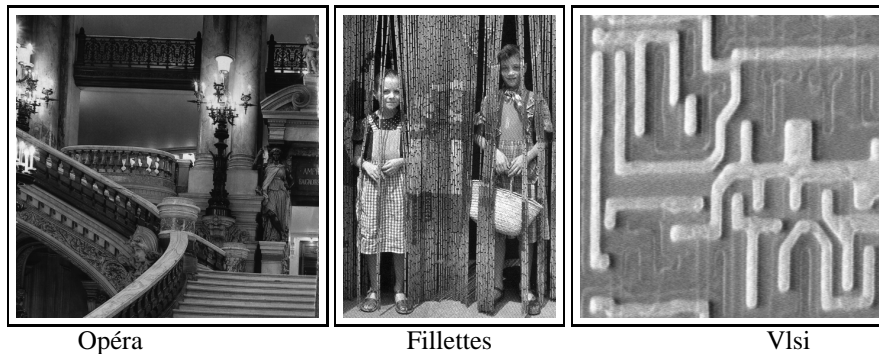


FIG. 2.1 – Trois images « quelconques ».

Notre propos est aussi de cerner les propriétés des *images en général*, les images usuellement rencontrées. Mais ces termes même définissent mal notre objet, car il n'y a pas de normalité en image. Ils excluent cependant clairement des images trop simples : le « carré blanc sur un fond blanc » des peintres, l'échiquier ou la grille, ainsi que les images « pathologiques » représentant des motifs très particuliers (rosaces, mosaïques, etc). Nous précisons lorsque ce sera nécessaire ce qui peut être normal ou anormal en traitement d'image. Retenons cependant que des images normales forment l'essentiel des scènes en vidéo, des images satellitaires, des images médicales, des images de robotique ou de microscopie.

2.2 L'amplitude

2.2.1 Propriétés

Nous représenterons l'image par une fonction soit continue $f(x, y)$, soit discrète $f(i, j)$ des deux variables d'espace x et y ou i et j . Nous utiliserons également fréquemment la restriction monodimensionnelle obtenue par balayage ligne à ligne $f_j(i)$ que nous écrirons $f(i)$ lorsqu'aucune ambiguïté ne sera à craindre. Nous limiterons parfois les variables i et j à N et M pour rendre compte de la dimension finie des images. Mais nous prendrons

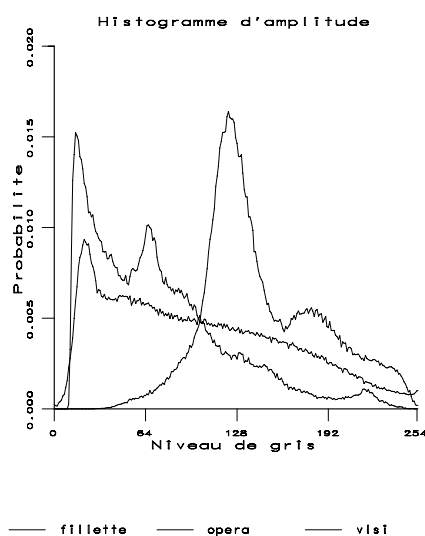


FIG. 2.2 – Les histogrammes des 3 images de la figure 2.1

alors N et M très grands. Les propriétés les plus simples auxquelles on s'intéresse tout d'abord sont liées à l'amplitude de l'image. Dans la plupart des cas, les images sont représentées sur 8 bits, parfois sur 10, 12 pour des images de très bonne qualité ou, exceptionnellement, 16. La première description d'intérêt est l'**histogramme** de l'amplitude qui exprime la fréquence d'apparition des divers niveaux.

Les images étant constituées de très nombreux échantillons (entre 100 000 et plusieurs millions), on invoque souvent le théorème central limite pour justifier l'hypothèse que l'amplitude de l'image serait gaussienne. Cela n'est généralement pas vrai. Les exemples de la figure 2.2 présentent des histogrammes typiques des trois images de la figure 2.1. Ils sont quelconques et toute autre allure serait également possible : l'histogramme d'une image n'est pas plus gaussien qu'uniforme.

2.2.2 Sensibilité aux modifications d'histogrammes

On peut étudier plus en détail la dépendance de l'apparence de l'image à son histogramme. D'intéressants travaux ont ainsi été menés par Estournet [Estournet, 1969] qui a fait réaliser plusieurs copies d'une même photo sur des supports variés : papiers photographiques de marques, de gradations et de duretés différentes. L'analyse de ces documents sur la figure 2.3 montre clairement de grandes différences d'histogrammes, différences bien plus remarquables que celles qui affectent les images elles-mêmes. Estournet en déduit que l'histogramme de l'image, non seulement n'est pas significatif de l'image (puisque de toute évidence beaucoup d'images auront des histogrammes très voisins compte tenu de la faible dimension des histogrammes), mais aussi peut être modifié artificiellement sans que l'aspect extérieur de l'image n'en soit beaucoup affecté, et surtout sans que la signification de l'image ne soit altérée.

Cette démarche peut être poussée plus loin. Est-il possible qu'une image ait un histogramme quelconque ? Pour répondre à cette question, nous allons utiliser un programme qui modifie les niveaux de gris de l'image par un transcodage garantissant que l'image résultat aura un histogramme prédéterminé. Imposons nous une contrainte pour assurer que l'aspect de l'image demeurera assez inchangé : que les relations d'ordre entre niveaux de gris ne soient pas perdues, c'est-à-dire que si n_1 et n_2 sont deux niveaux de gris vérifiant $n_1 < n_2$, alors on les transfor-

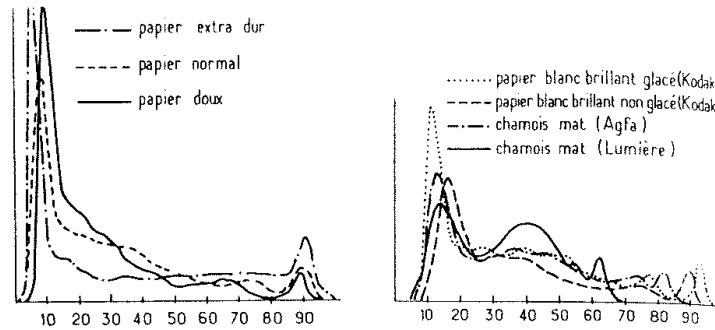


FIG. 2.3 – Histogrammes d’une même image obtenus pour des présentations sur des papiers de natures différentes : à gauche, dureté du papier, à droite, aspect du papier (d’après [Estournet, 1969]).

mera en m_1 et m_2 tels que $m_1 \leq m_2$. Un tel programme est aisé à concevoir. Appliqué à des images il donne les résultats de la figure 2.4, où les histogrammes d’arrivée sont respectivement une loi uniforme, une somme de 2 gaussiennes et une sinusoïde. Nous percevons bien sûr des différences dans les images ainsi créées, mais les images demeurent parfaitement reconnaissables et leur interprétation n’en est pas affectée. Cette expérience nous confirme le rôle très superficiel de l’histogramme dont l’allure n’est liée que de très loin aux éléments significatifs de l’image.

Est ce que ces conclusions sont toujours valables ? Non bien sûr, puisque des images très particulières peuvent avoir des histogrammes très spécifiques : un échiquier sera composé de deux pics et un texte noir sur blanc de deux modes nettement séparés. Si nous prenons un champ de vagues très régulier, ou un mur crépi, il est possible également que l’on obtienne des histogrammes gaussiens.

Dans le cas des images obtenues avec des illuminations cohérentes, on peut mener théoriquement le calcul de la loi de probabilité de l’amplitude réfléchiée par une surface uniforme rugueuse (hypothèse dite de « speckle pleinement développé » [Goodman, 1976]).

On montre alors qu’on obtient, pour probabilité du module de l’amplitude complexe, une loi de Rayleigh :

$$p(f) = \frac{f}{\sigma^2} \exp\left(-\frac{f^2}{2\sigma^2}\right). \quad (2.1)$$

On vérifie sur l’image 2.5 que l’histogramme suit précisément ce modèle.

2.3 Les sauts de l’amplitude

Définissons un « saut d’amplitude » comme la différence entre niveaux de gris de deux pixels adjacents le long d’une ligne : $s_j(i) = s(i) = f_j(i+1) - f_j(i)$. La variable $s(i)$ varie généralement de -255 à +255. Étudions expérimentalement la probabilité de s . La figure 2.6 présente ces probabilités. Nous remarquons que :

- elles ont un maximum très marqué pour $s = 0$, ce qui exprime que le saut nul est le plus probable,
- elles sont symétriques,
- elles ont une décroissance très rapide et régulière.

Au vu de ces courbes il a été proposé de modéliser la probabilité du saut d’amplitude par une gaussienne de moyenne nulle, dont le seul paramètre, l’écart-type σ_s , est une caractéristique de l’image. La probabilité du saut s de l’amplitude pourra alors s’écrire :

$$p(s) = \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma_s^2}\right) \quad (2.2)$$

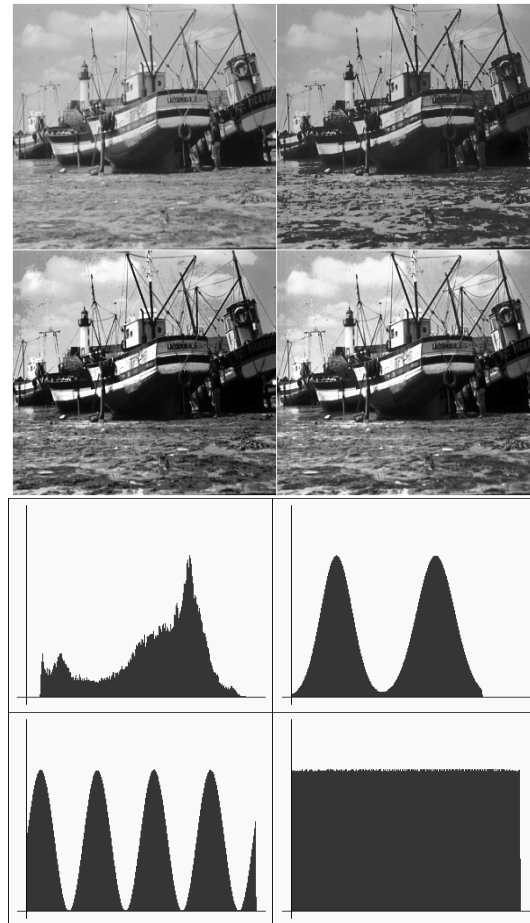


FIG. 2.4 – Modification d'histogramme. En haut à gauche : image originale, en haut à droite somme de 2 gaussiennes, en bas à gauche sinusoïde, en bas à droite répartition uniforme.

Certaines images à transitions douces se caractérisent par une faible valeur de σ_s , d'autres par une forte. Ce domaine de variation de σ_s est typiquement de 3 à 15.

Il est clair que des contre-exemples peuvent être aisément trouvés d'images dont l'histogramme n'a pas les propriétés ci-dessus. Un dégradé régulier du blanc au noir présenterait un histogramme non symétrique avec des valeurs uniformément réparties dans le demi-espace gauche. Un échiquier ne présenterait que trois types de transitions, des transitions nulles, des positives à +255 et des négatives à -255. Ces exemples n'entrent bien sûr pas dans nos hypothèses d'images normales.

2.4 La fonction d'autocorrélation

2.4.1 À une dimension

Une autre grandeur est d'un grand intérêt pour les traiteurs de signaux, c'est la fonction d'autocorrélation. À cela deux raisons : tout d'abord elle exprime d'une façon très complète les dépendances spatiales du signal, d'autre

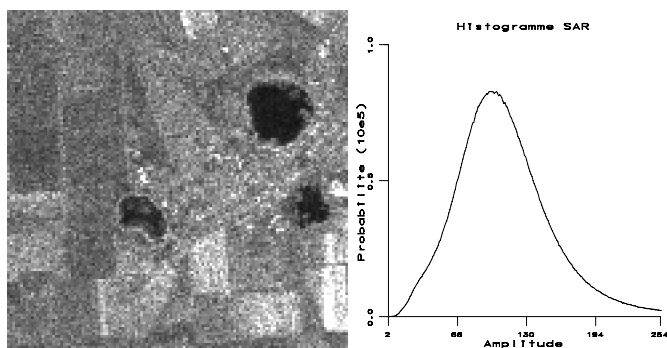


FIG. 2.5 – Agrandissement d’une image de radar à vision latérale (radar ERS-1) sur une région agricole de l’Ukraine (©Cnes) montrant l’importance du bruit de cohérence, et histogramme de la scène complète, très proche d’une loi de Rayleigh.

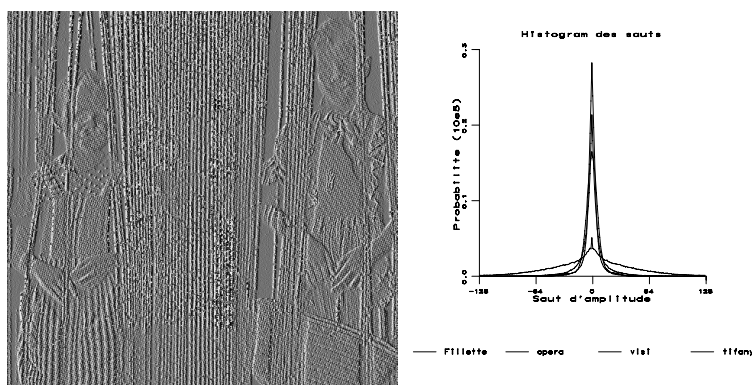


FIG. 2.6 – Une image des gradients horizontaux d’un extrait de l’image « Fillettes » (la valeur 127 a été ajoutée à toute l’image pour l’affichage) et 4 histogrammes de sauts de 4 images différentes.

part elle permet un accès commode au domaine de Fourier par l’intermédiaire du théorème de Wiener-Khinchine¹.

La fonction d’autocorrélation $\gamma_f(k)$ d’une fonction discrète $f(i)$ est définie par :

$$\gamma_f(k) = \langle f(i) \cdot f^*(i+k) \rangle \quad (2.3)$$

où le symbole $\langle x \rangle$ exprime l’espérance mathématique de la variable aléatoire x . La fonction d’autocorrélation se calcule usuellement par :

$$\gamma_f(k) = \sum_{i=-\infty}^{i=+\infty} f(i) \cdot f^*(i+k) \quad (2.4)$$

Les images sont des signaux réels ($f = f^*$). Ce sont des signaux positifs dont on retire la valeur moyenne \bar{f} pour calculer la **fonction d’autocorrélation centrée** :

¹Le théorème de Wiener-Khinchine établit que la fonction d’autocorrélation est transformée de Fourier (TF) du spectre de densité de puissance [Blanc-Lapierre et Picinbono, 1981].

$$\gamma_{f-\bar{f}}(k) = \sum_{i=-\infty}^{i=+\infty} (f(i) - \bar{f}) \cdot (f(i+k) - \bar{f}) \quad (2.5)$$

et on calcule une **fonction d'autocorrélation normée centrée** en rapportant cette fonction à son maximum $\gamma_{f-\bar{f}}(0)$ (qui est également la variance σ_f^2 de f) :

$$C_f(k) = \frac{\gamma_{f-\bar{f}}(k)}{\gamma_{f-\bar{f}}(0)} = \frac{\gamma_{f-\bar{f}}(k)}{\sigma_f^2} \quad (2.6)$$

$$C_f(k) = \frac{\sum_{i=-\infty}^{i=+\infty} (f(i) - \bar{f}) \cdot (f(i+k) - \bar{f})}{\sum_{i=-\infty}^{i=+\infty} (f(i) - \bar{f})^2} \quad (2.7)$$

Cette fonction dispose de nombreuses propriétés intéressantes :

- elle a un maximum à l'origine (égal à 1) ;
- elle est bornée à l'intervalle $[-1, +1]$;
- elle est paire ($C_f(-k) = C_f(k)$) ;
- si elle possède un autre maximum égal à 1 en un autre point p , alors elle possède une infinité de maxima aux positions np , et elle est périodique de période p , ainsi que la fonction f elle-même ;
- il existe des relations simples entre les dérivées de la fonction d'autocorrélation et les fonctions d'autocorrélation des dérivées de f .

On présente sur la figure 2.7 la fonction d'autocorrélation d'une image, pour des décalages de 1 à 27 points. On note un maximum marqué à l'origine (souvent accompagné d'une discontinuité de la dérivée), suivi d'une décroissance rapide et régulière. Si l'on observe la décroissance de C_f sur un plus long domaine, on constate que ses amplitudes deviennent très faibles, exprimant une faible corrélation des pixels de l'image très éloignés les uns des autres.

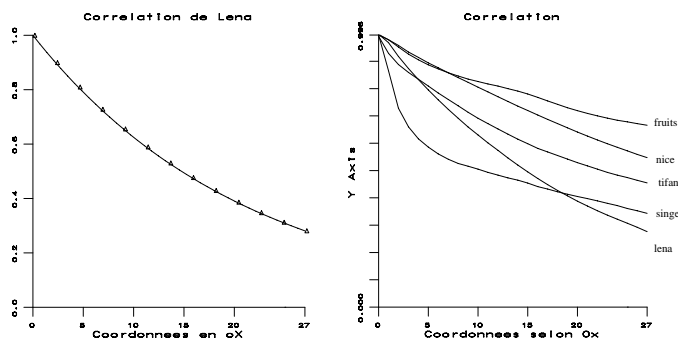


FIG. 2.7 – A gauche sont superposées les valeurs mesurées d'une fonction d'autocorrélation (sur l'image « Lena ») à la loi exponentielle théorique la plus proche, à droite 5 fonctions de corrélation d'images différentes.

Ce comportement très général a permis de modéliser la fonction d'autocorrélation normée centrée par une fonction exponentielle :

$$C_f(k) = \exp(-\alpha|k|) \quad (2.8)$$

Le paramètre α est le second paramètre caractéristique de l'image, on voit sur la figure 2.7 que les images à variation lente ont une corrélation très étalée (et donc un très faible α). Au contraire, les images à variations très rapides ont une autocorrélation pointue et un α grand. Le domaine de variation de α est typiquement de 0,01 à 0,3.



FIG. 2.8 – Les 5 images ayant servis à calculer les corrélations de la figure 2.7 : Nice, Fruits, Tiffany, Lena et Singe.

Les contre-exemples de fonctions dont l'auto-corrélation n'est pas exponentielle sont aisés à trouver, en particulier, parmi les fonctions périodiques. Si les fonctions purement périodiques sont rares dans les images naturelles, les images présentant un motif qui se répète abondent (grilles, murs, vignes, tissus, etc.). D'autres ont des motifs moyens (foule dans des gradins, façades d'immeubles dans une rue, etc.). Ces images auront une autocorrélation composée de deux motifs multipliés : l'un périodique, l'autre exponentiellement décroissant exprimant la perte du motif avec la distance (cf. figure 2.9).

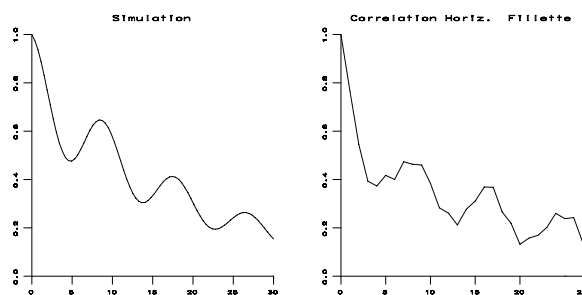


FIG. 2.9 – Allure de l'autocorrélation d'une image avec un fort contenu périodique : théorique à gauche et mesurée (sur *Fillette*), à droite.

2.4.2 En multi-dimensions

Si l'on considère maintenant non plus le signal le long d'une ligne de l'image, mais selon une direction quelconque, on retrouve un comportement semblable de la fonction d'autocorrélation, mais avec peut-être des coefficients α différents selon les directions. Ainsi, il ne sera pas surprenant de trouver une forte corrélation verticale

dans un paysage de sous-bois de sapins alors qu'elle sera faible horizontalement. On voit sur la figure 2.10 des mesures sur une image réelle reflétant cette variété. La fonction d'auto-corrélation bidimensionnelle $C_f(x, y)$ exprime bien les propriétés d'isotropie de la scène.

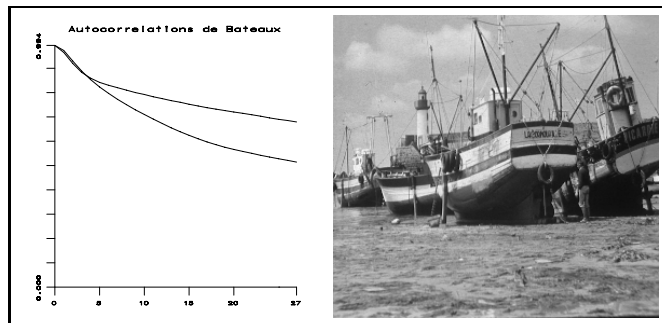


FIG. 2.10 – Corrélation obtenue dans les directions horizontales (en haut) et verticales (en bas) pour l'image « Bateaux ».

Si l'on veut modéliser ce comportement, on est amené à choisir l'un des deux modèles bidimensionnels suivants :

$$C_f(x, y) = \exp(-\sqrt{Q(x, y)}) \tag{2.9}$$

où $Q(x, y)$ est une forme quadratique en x et y ,

$$C_f(x, y) = \exp(-\alpha|x| - \beta|y|) \tag{2.10}$$

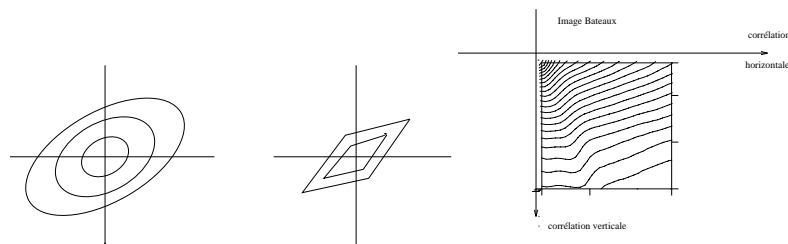


FIG. 2.11 – Courbes d'isocorrélation pour le modèle non-séparable (à gauche) et pour le modèle séparable (au centre) et courbes d'isocorrélation mesurée sur l'image « Bateaux » (le quart inférieur du plan est représenté).

Cette deuxième forme est moins satisfaisante pour la continuité de la fonction d'autocorrélation, mais elle est cependant très souvent retenue car sa forme séparable la rend très commode d'emploi (séparation en deux processus indépendants en ligne et en colonne), particulièrement dans le plan de Fourier (cf. figure 2.11).

2.4.3 Le spectre de densité de puissance

Déoulant du théorème de Wiener-Khinchine, le spectre de densité de puissance possédera donc une forme déterminée [Bracewell, 1978]. Soit F la TF de f , soit u la variable associée à la variable d'espace (*fréquence spatiale*) et notons par P_f le spectre de densité de puissance de f :

$$P_f(u) = \langle |F(u)|^2 \rangle \tag{2.11}$$

$$P_f(u) = \frac{2\alpha}{\alpha^2 + u^2} \tag{2.12}$$

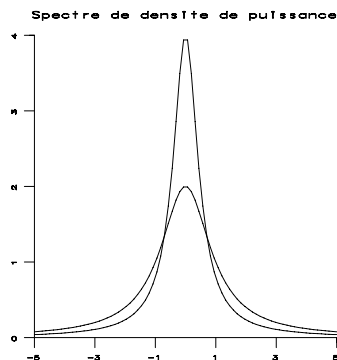


FIG. 2.12 – Spectre de densité de puissance d’images à autocorrélation exponentielle ($\alpha = 0,5$ et $\alpha = 1$)

On voit sur cette fonction (distribution de Lorentz, figure 2.12) des propriétés bien connues des opticiens : l’énergie est maximale à l’origine ($u = 0$), c’est-à-dire, en théorie de la diffraction [Françon, 1970], sur l’axe optique, elle va en diminuant régulièrement vers les hautes fréquences spatiales (positives ou négatives), d’autant plus vite que la corrélation est plus forte (α faible) [Nicolas, 1995].

2.5 Entropie

Suivant les idées de Shannon, l’entropie est une mesure statistique du **contenu d’information** d’un message. Elle est donc adaptée à caractériser les images, chacune d’entre-elles étant prise comme un message spécifique.

L’idée d’entropie ² s’appuie, bien sûr, sur l’analogie de la physique : thermodynamique de Clausius ou statistique de Boltzman et attribuée à des configurations particulièrement improbables un plus grand contenu d’information qu’aux configurations très fréquentes.

2.5.1 Entropie d’ordre 0

Suivant l’idée précédente, Hartley en 1925 attribue à un symbole i émis par une source stationnaire avec une probabilité p_i , la quantité d’information :

$$I_i = -\log p_i \quad (2.13)$$

exprimé en *bit par symbole* si le logarithme est à base 2. Shannon [Shannon, 1948] définit alors l’**information moyenne par symbole** ou *entropie* par :

$$S = -\sum_{i=1}^{i=n} p_i \log p_i \quad (2.14)$$

Considérant une image comme une source stationnaire de pixels, l’entropie se mesure donc à partir du seul histogramme de l’amplitude.

Utilisant les propriétés de la fonction $\log x$, on montre que l’entropie est maximale pour une répartition uniforme de probabilité $S_{max} = \log n$. Pour les images, cela conduit le plus souvent à une entropie S inférieure à

²c’est une des approches possibles de la définition de l’information. Une autre approche repose sur la théorie de la complexité [Delahaye, 1994] et mesure l’information apportée par un message par la longueur du plus petit programme capable de décrire cette information. C’est l’approche de Kolmogorov [Kolmogorov, 1965], elle est complétée par la notion de profondeur logique de Benett qui distingue les séquences complexes mais calculables, de celles complexes mais aléatoires, au moyen du nombre de pas élémentaires accomplis par le programme pour retrouver l’information. Le lecteur curieux trouvera dans [Maître, 1996] une application de ces principes aux images.

8 bits [Gallager, 1966] Les chiffres ordinairement rencontrés vont de 4 à 7,5 bits pour des images normalement contrastées, utilisant l'essentiel de la dynamique des niveaux de gris.

L'effet d'une quantification uniforme de l'image (réduction des niveaux de gris) conduit à une réduction régulière de l'entropie avec une pente presque égale à -1 (cf. figure 2.13).

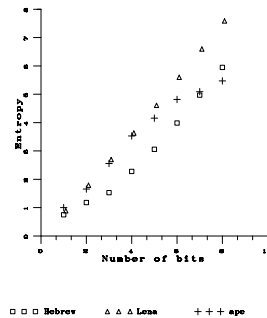


FIG. 2.13 – Décroissance de l'entropie pour 3 images en fonction de la quantification des niveaux de gris.

2.5.2 Entropie des sauts

En utilisant les propriétés soulignées ci-dessus, on peut en particulier calculer l'entropie des sauts d'intensité s dont la probabilité est gaussienne :

$$S_s(x) = \frac{1}{2} \log(2\pi e \sigma_s)$$

Cette entropie a le mauvais goût d'être négative pour de très faibles valeurs de σ_s , mais il faut alors prendre en compte l'aspect discret des niveaux de quantification de l'image et remplacer le résultat obtenu en continu ci-dessus par sa forme discrète.

2.5.3 Entropie d'ordre N

Dans l'approche précédente on considérait chaque pixel indépendant des autres. Cela reflète bien mal ce que nous connaissons maintenant sur la corrélation entre pixels. Il n'est pas difficile d'étendre les idées précédentes à une source qui émettrait non plus 1 pixel indépendamment des autres, mais 2 (ou 3, 4, etc.). On prend alors le couple (désigné par i et j) comme symbole, il a une probabilité $p(i, j)$ et l'on définit par analogie l'entropie de la source d'ordre 1 par :

$$S_1 = - \sum_{i=1, j=1}^{i=n, j=n} p(i, j) \log p(i, j) \quad (2.15)$$

On démontre sans peine que :

$$S \leq S_1 \leq 2S$$

l'une des égalités étant vérifiée si les pixels sont indépendants, l'autre si leur dépendance est linéaire.

Cette extension se reconduit naturellement pour des sources d'ordre N , c'est-à-dire composées de symboles indépendants constitués de $N + 1$ pixels, à partir de la connaissance des probabilités jointes de ces pixels. Notons cependant que cette probabilité étant un tableau de dimension G^{N+1} , (où G est le nombre de niveaux de gris dans l'image), elle devient de plus en plus difficile à calculer.

2.5.4 Entropie conditionnelle, redondance

Si l'on considère 2 images exactement superposables (par exemple deux canaux d'une image satellitaire ou le canal rouge et le canal vert d'un signal vidéo), que l'on dénote par x_i et y_i les valeurs en un même pixel i , on peut définir bien sûr l'entropie S_x et l'entropie S_y de chaque image :

$$S_x = - \sum_{i=1}^{i=n} p(x_i) \log p(x_i)$$

$$S_y = - \sum_{i=1}^{i=n} p(y_i) \log p(y_i)$$

on peut également définir l'entropie jointe de x et y :

$$S_{x,y} = - \sum_{i=1}^{i=n} p(x_i, y_i) \log p(x_i, y_i) \quad (2.16)$$

mais on peut aussi définir l'**entropie conditionnelle** de y par rapport à x , comme l'information moyenne, pour toutes les valeurs de y , de l'information conditionnelle apportée par y sachant x :

$$I_{y|x} = -\log p(y|x)$$

par moyennage sur tous les couples (x, y) :

$$S_{y|x} = - \sum_{i=1}^{i=n} p(x_i, y_i) \log p(y_i|x_i)$$

L'information apportée par x sur y est alors exprimée par :

$$S_{x:y} = S_y - S_{y|x} \quad (2.17)$$

On vérifie que :

$$S_{x:y} = S_{y:x}$$

Cette grandeur est appelée la **redondance** de x et y . On peut vérifier les propriétés suivantes :

$$S_{x,y} = S_x + S_{y|x}$$

$$S_{y|x} \leq S_y$$

$$S_{x:y} = S_x + S_y - S_{x,y}$$

Le codage prédictif en vigueur dans les années 80 (codage MICD : Modulation par Impulsions Codées Différentielle ou DPCM [Guillois, 1996b, Barlaud et Labit, 2002]) s'est beaucoup appuyé sur la recherche des meilleures configurations de N pixels permettant de minimiser l'entropie conditionnelle du $N + 1$ ^{ème} : $S_{x_{N+1}|x_1, x_2, \dots, x_N}$.

2.5.5 Retour sur un paradoxe

Reprenant le petit exercice de modification d'histogramme présenté dans la section 2.2.2, on peut s'interroger sur l'évolution de l'entropie au cours des opérations effectuées. Reprenons par exemple l'image des *Bateaux* dont l'entropie originale était de 6,1. Après la troisième modification d'histogramme (figure 2.4), son entropie est passée

à 8 bits par pixel. Comment explique-t-on ce gain en information moyenne ? Où est pratiquement localisée cette information ? Que peut en faire l'utilisateur ?

Pour répondre à ces questions, il faut bien sûr analyser en détail les schémas de transformation, et en en imaginant d'astucieux, il est possible de se convaincre que l'on peut effectivement ajouter une information utile, par exemple en donnant une image moins bruitée (ou au contraire plus contrastée), ou en incorporant quelque bit de parité ou autre marqueur permettant d'assurer des vérifications sur l'intégrité de la transmission, l'identité de l'auteur ou autre.

2.6 Un modèle d'image

2.6.1 Processus de Markov Gauss

Partant des deux résultats précédemment exposés (probabilité des sauts gaussienne et autocorrélation à décroissance exponentielle), il est venu assez naturellement à l'idée de proposer le modèle générateur de signal d'image suivant : **l'image est un processus markovien, à temps discret, et à accroissements gaussiens.**

Rappelons [Cullmann, 1975] que les processus markoviens à temps discret sont des processus qui se développent le long d'une suite ordonnée d'instant (ici les pixels le long de la ligne) et dont la probabilité de se trouver à l'instant i dans l'état f_i est fonction du seul état à l'instant $i - 1$: f_{i-1} et de la probabilité de transition $P(f_i|f_{i-1})$. Si le processus est stationnaire, cette probabilité ne dépend pas de i , si ses accroissements sont gaussiens, alors cette probabilité est de la forme :

$$\begin{aligned} P(f_i|f_{i-1}) &= P(f_i - f_{i-1}) \\ &= \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma_s^2}\right) \quad \text{avec} \quad s = f_i - f_{i-1} \end{aligned}$$

La probabilité conditionnelle $P(f_i|f_{i-1})$ se représente, dans le langage des chaînes de Markov, par une matrice M de taille $G \times G$, où G représente le nombre de niveaux de gris. Cela permet d'écrire :

$$P(f_i) = M.P(f_{i-1})$$

M est une matrice **stochastique** (somme des colonnes égale à 1 pour une ligne donnée), mais non bi-stochastique. Ce modèle très simple est très utile pour déterminer de nombreuses propriétés des images. Il a permis en particulier de nombreux travaux sur le codage prédictif (MICD ou DPCM) [Guillois, 1996b, Barlaud et Labit, 2002].

2.6.2 Mais ...

Mais ce modèle simple laisse plusieurs problèmes mal résolus :

- la stationnarité est difficile à obtenir avec des processus markoviens. En effet, le processus de Markov est stationnaire si P a une limite quand i tend vers l'infini. Or on peut écrire :

$$P(f_i) = M^i.P(f_0)$$

cela ne tendra vers une limite indépendante du premier point de la ligne f_0 que si M est une matrice dont la seule valeur propre de module 1 est 1, c'est-à-dire si M est régulière (toutes les colonnes de M sont égales) ;

- en l'absence de stationnarité, on a affaire à un processus de diffusion (de Wiener-Levy), et l'on montre que la variance du point i est liée à celle du point 0 par :

$$\sigma_i^2 = \sigma_0^2 + i\sigma_s^2$$

et sa fonction d'autocorrélation dépend du point i où elle est calculée [Fleuret, 1977].

Tout cela montre que l'on ne peut pousser trop loin ce modèle gaussien markovien. Il rend néanmoins bien compte de quelques propriétés très générales et demeure à ce titre souvent utilisé.

2.7 Critique de la démarche précédente

Plutôt que de torturer les équations pour les plier aux données expérimentales, il est préférable de questionner les hypothèses sous-jacentes au modèles précédent pour trouver les faiblesses de la démarche.

Nous en noterons 3 grandes familles, tournant autour de trois notions importantes : la stationnarité, la causalité et la dimensionnalité.

2.7.1 La stationnarité

Nous n'avons pas porté attention jusqu'ici à l'hypothèse de stationnarité qui est pourtant sous-jacente à toutes les estimations que nous avons faites jusqu'à présent, qu'il s'agisse d'amplitude, de saut ou de corrélation. Qu'exprime-t-elle ? Que les estimations que nous faisons sont indépendantes du point de l'image où elles sont faites, c'est-à-dire que les symboles $\langle \rangle$ expriment des estimations représentatives de l'image globalement et non d'un champ particulier de celle-ci.

Cela écarte de nos modèles des images qui s'éloignent trop clairement de cette hypothèse : par exemple un paysage dont la partie haute (le ciel) est très clair et uniforme tandis que le sol est sombre et riche en détails. Dans ce cas, ni les propriétés du premier ordre (moyenne, variance, entropie), ni les propriétés d'ordre supérieur comme la corrélation ne peuvent s'évaluer globalement pour l'image entière. Ecartera-t-on ainsi toutes les images ou existe-t-il des documents qui vérifient cette hypothèse de stationnarité ? C'est une question délicate à laquelle nous pouvons répondre par deux voies :

- par une approche théorique si l'on connaît la source d'image, son contenu et les raisons de ses variations ;
- par une approche expérimentale où l'on confronte à l'image originale une version modifiée de celle-ci faite en sorte qu'elle vérifie l'hypothèse de stationnarité.

La première approche permet d'imaginer des configurations particulières, homogènes, par exemple un champ de houle, un tissu biologique, ou un portion de ciel, certaines images satellitaires ... Mais pour les scènes ordinaires, il apparaît évidemment que les images sont composées d'objets indépendants : (une table, un meuble, un arbre) dont les propriétés sont indépendantes et possiblement très différentes.

La seconde expérience montre qu'il est pratiquement très difficile de concevoir des images ordinaires qui soient stationnaires.

2.7.2 La causalité

Elle est sous-jacente à la plupart des signaux temporels, mais perd son sens pour les images pour lesquelles il n'existe pas d'écoulement irréversible du temps, à moins qu'elles n'aient été balayées comme par exemple en télévision. La causalité est l'une des propriétés fortes des processus markoviens qui n'a pas de signification pour l'image. Recréer une causalité en 2D impose aussi des arbitrages et l'on définit ainsi des « passés » qui peuvent être quart de plan, ou demi-plan, selon les besoins (cf. figure 2.14). Ces réflexions sont à la base des travaux conduits sur les filtres de Kalman bidimensionnels [Wood, 1981].

2.7.3 La monodimensionnalité

La structure bidimensionnelle de l'image rend sa modélisation plus complexe. Les schémas à 1D se combinent difficilement à 2D sans perdre certaines de leurs propriétés, ainsi en est-il des chaînes de Markov. L'extension peut prendre plusieurs formes : une formulation selon 2 équations couplées en lignes et en colonnes, ou une formulation directement planaire.

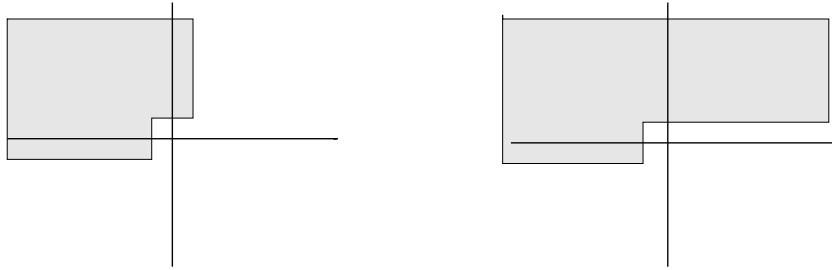


FIG. 2.14 – Prédiction d'un point par « passé quart de plan », à gauche, et « passé demi-plan », à droite. Les zones grisées constituent le passé du point au centre du repère, au sens de la causalité markovienne.

Ainsi Roesser a proposé un processus du type :

$$\begin{aligned} r(i+1, j) &= a_1 r(i, j) + a_2 s(i, j) + b_1 u(i, j) \\ s(i, j+1) &= a_3 r(i, j) + a_4 s(i, j) + b_2 u(i, j) \end{aligned}$$

où u est la partie aléatoire du processus (typiquement gaussien), l'image finale étant donnée par :

$$f(i, j) = c_1 r(i, j) + c_2 s(i, j)$$

et Attasi [Attasi, 1975] :

$$f(i+1, j+1) = a_1 f(i, j+1) + a_2 f(i+1, j) - a_1 a_2 f(i, j) + b u(i, j)$$

D'autres modèles peuvent être trouvés dans les références [Huang, 1976] et [Huang, 1981].

2.7.4 Vers un modèle complet d'image ?

Reprenant attentivement les propriétés du modèle présenté en 2.6.1 à la lumière des remarques précédentes, on constate que les écarts majeurs aux lois ci-dessus apparaissent pour les sauts de transition élevés (les queues des courbes de la figure 2.6) que la loi gaussienne sous-estime considérablement. Ces points sont ceux des transitions importantes, donc des contours de l'image. Le modèle 2.6.1 suppose un champ uniformément couvert d'une seule texture et non la présence d'objets variés séparés par des contours.

Peut-on alors construire un modèle global d'image prenant en compte ces propriétés ? L'objectif est de rendre compte simultanément des trois limites soulignées ci-dessus.

De nombreuses propriétés bidimensionnelles ont été mesurées dans les images pour compléter les analyses à 1D et les relier à des propriétés mono-dimensionnelles [Maître, 1977]. Ainsi diverses études ont porté sur la mesure des probabilités les plus variées :

- la probabilité d'apparition de deux contours successifs à une distance d le long d'une droite quelconque,
- la probabilité de la longueur de plages constantes à ϵ près le long d'une ligne,
- la probabilité de l'orientation des contours,
- la probabilité de la surface d'une aire constante à ϵ près, etc.

La conclusion la plus utile de ces études concerne les longueurs de plages, ou distances entre contours qui suivent assez bien une répartition poissonnienne :

$$p(l) = \frac{1}{\bar{l} - 1} \left[\frac{\bar{l} - 1}{\bar{l}} \right]^l$$

où \bar{l} dénote la longueur moyenne de la longueur de plage, valeur spécifique d'une image donnée. Cette propriété a souvent été utilisée dans des schémas de codage ou de détection, elle est cependant très difficile à inclure dans des modèles et aucun modèle aujourd'hui disponible n'est capable de l'intégrer à notre connaissance.

La prise en compte des dépendances non-causales entre pixels a abouti à l'abandon presque général de la piste des **chaînes de Markov** au profit des **champs de Markov** (cf. chapitre 7). Dans les champs de Markov, un pixel ne dépend plus simplement des 2 ou 3 pixels qui l'ont précédé lors du balayage régulier de l'image, mais d'un voisinage qui l'entoure et qui fait porter les dépendances de façon symétrique entre les points (voisinage de 4 points, ou de 8 points dans la plupart des cas). L'inconvénient des champs markoviens est de ne plus se prêter à des techniques de filtrage directes, mais de faire appel à des techniques itératives généralement plus coûteuses en temps machine, et plus difficiles à optimiser.

La prise en compte des propriétés de non-stationnarité des images a incité les auteurs à abandonner l'idée d'un modèle global au profit d'un modèle **en mosaïque**, dans lequel l'image est constituée d'une partition de plages, chacune relevant d'un modèle semblable à celui que nous avons développé en 2.6.1, c'est-à-dire chaque plage suivant une loi de Markov-Gauss définie par la valeur moyenne \bar{f} , le paramètre de probabilité de transition σ_s , le paramètre de décroissance de la corrélation α . Ce modèle définit deux grandeurs qui seront fondamentales pour tout le traitement des images : **les contours** qui représentent les transitions d'une plage à une autre, et les **textures** qui décrivent les propriétés statistiques de chaque plage. C'est le modèle qui sera le plus utilisé pour les opérations de reconnaissance des formes, de filtrage et de détection. C'est aussi un modèle qui est particulièrement exploré pour les techniques de codage avancé (MPEG-4 : codage par objets [Barlaud et Labit, 2002]), ainsi que pour la restauration des images préservant les contours.

Chapitre 3

L'échantillonnage des images, la représentation fractale

Chapitre rédigé par Henri MAÎTRE

Chronologiquement, l'une des toutes premières étapes du traitement numérique des images est la tâche d'échantillonnage qui réduit l'ensemble continu du monde observable en une série de valeurs discrètes. Nous n'insisterons pas sur l'importance de cette étape en traitement des images, tant il est évident que c'est par elle que l'on contrôle la finesse des détails enregistrés, et par là même, la nature de l'information retenue dans l'image numérique. Cette étape s'accompagne d'une quantification des niveaux de gris et très souvent débouche ensuite soit sur la compression, soit sur le traitement.

Mais en traitement des images, l'échantillonnage apparaît en de nombreuses autres occasions que lors de l'acquisition. En effet, les images sont très souvent appelées à être ré-échantillonnées, par exemple pour des conversions de format (passage d'un format SECAM à un format NTSC¹ en télévision numérique), ou pour en transformer la géométrie (corriger une perspective, orienter une forme dans un repère convenable ou recalculer l'image sur une référence). Toutes ces étapes posent à nouveau le problème de l'échantillonnage et nécessitent les mêmes précautions que l'acquisition elle-même.

Dans tout le champ du traitement du signal, l'échantillonnage est abordé à l'aide de la théorie classique proposée par Nyquist et Shannon. Cette théorie s'applique bien sûr aux images et nous la développerons donc dans un premier temps. Nous verrons cependant qu'elle est parfois mal adaptée. Nous verrons également que certains objets peuvent être très mal traités par cette approche. C'est le cas des objets fractales que nous présenterons en fin de chapitre.

3.1 Les signaux monodimensionnels à bande limitée

La théorie de l'échantillonnage s'applique à des signaux f , fonctions d'une variable continue x , dont la bande passante est limitée, c'est-à-dire dont la transformée de Fourier (TF) F , fonction de u , variable associée² à x ,

¹Les formats SECAM et NTSC sont les normes de représentation des signaux de télévision utilisés en France et aux États-Unis [Guillois, 1996a].

²À une variable x temporelle correspond une variable u fréquentielle, exprimée en Hertz ; à une variable x spatiale, correspond une fréquence spatiale u , exprimée en m^{-1} .

possède un support fini $[-U, U]$:

$$F(u) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi iux} dx \quad : \quad F(u) \equiv 0 \quad \forall |u| \geq U$$

3.1.1 Échantillonnage du signal

Le théorème d'échantillonnage établit l'équivalence, sous certaines conditions, entre un signal continu $f(x)$, connu pour tout x , et un ensemble d'échantillons discrets g_i . Cette équivalence se traduit en particulier par la possibilité de calculer f à partir des valeurs g_i et réciproquement la possibilité de calculer g_i à partir de f . Les conditions sont les suivantes ;

- les g_i sont les valeurs du signal $f(x)$ prises à des positions x régulières ;
- ces positions sont séparées d'un intervalle p au plus égal à $\frac{1}{2U}$.

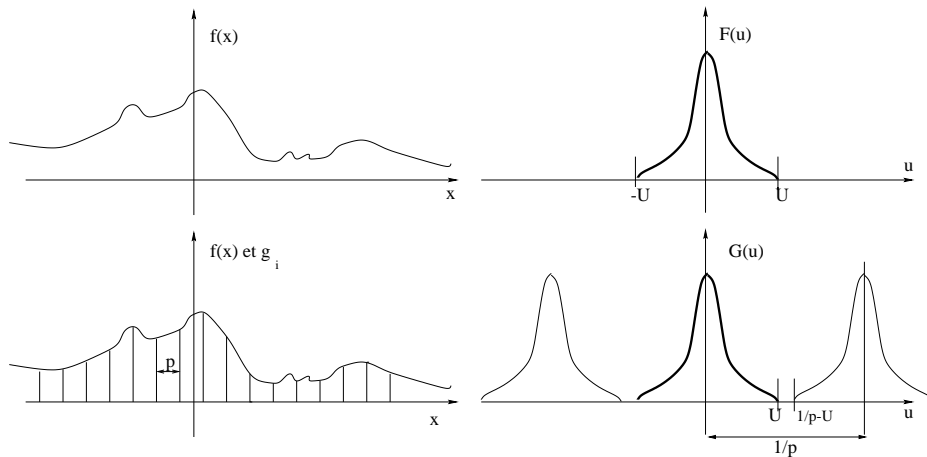


FIG. 3.1 – Échantillonnage d'un signal continu $f(x)$, à spectre $F(u)$ à support borné, avec un pas d'échantillonnage p . Le signal résultant discret a un spectre $|G(u)|^2$.

Ces échantillons sont donc en nombre infini et ont pour valeur :

$$g_i = f(x = ip) = f(x)\delta(x - ip)$$

Ce que nous écrirons aussi :

$$g(x) = f(x) \sum_{i=-\infty}^{\infty} \delta(x - ip) = f(x)\mathbb{I}(x/p)$$

la fonction $\mathbb{I}(x/p)$ étant le peigne de Dirac de période p .

La TF $G(u)$ se déduit de F par répétition de période $\frac{1}{p}$ (cf. figure 3.1) et atténuation de $\frac{1}{p}$. Comme $\frac{1}{p} \geq 2U$, sur l'intervalle $]-U, U[$, $G(u)$ et $F(u)$ coïncident exactement (au facteur $\frac{1}{p}$ près).

Pour reconstruire le signal, il est intéressant de distinguer le cas où le signal a été échantillonné à la fréquence la plus basse, $p = \frac{1}{2U}$, appelée fréquence de Nyquist, du cas où l'on choisit un nombre plus élevé d'échantillons $p < \frac{1}{2U}$.

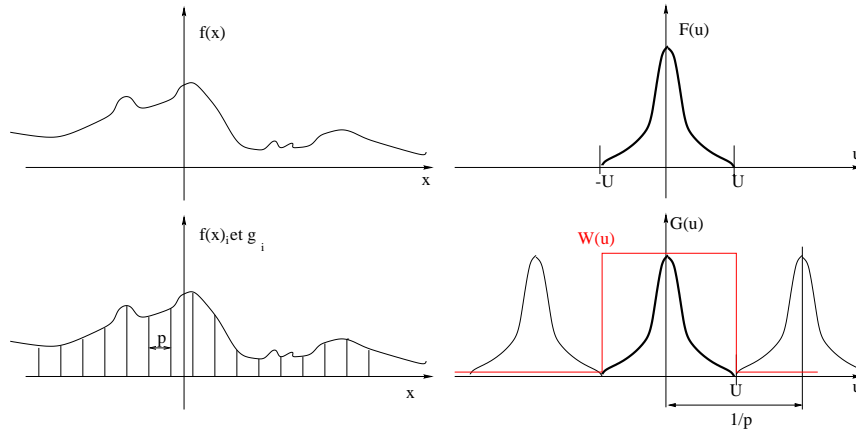


FIG. 3.2 – Échantillonnage d'un signal continu $f(x)$ à la fréquence de Nyquist. Le filtre de reconstruction $W(u)$ est le filtre canonique $W_N(u)$, c'est-à-dire la fonction porte.

3.1.2 Reconstruction du signal échantillonné à la fréquence de Nyquist

Dans le cas où l'on choisit juste le nombre minimal d'échantillons, on se trouve dans la situation particulière décrite sur la figure 3.2 où les ordres parasites dus à l'échantillonnage affectent toutes les fréquences immédiatement supérieures à U . La reconstruction est alors totalement contrainte. Elle se fait par le filtre unique $W_N(u)$ tel que :

$$W_N(u) = \begin{cases} 1 & \forall |u| < U \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

La reconstruction de f à partir de g s'écrit alors :

$$f(x) = pg(x) * w_N(x)$$

ou de façon équivalente :

$$F(u) = pG(u)W_N(u)$$

La fonction w_N qui vérifie les relations 3.1 s'appelle la fonction canonique d'échantillonnage. Nous en verrons l'usage plus loin. Dans le cas monodimensionnel qui nous intéresse ici, nous avons :

$$w_N(x) = \text{sinc}(xp)$$

et la reconstruction, en dehors des valeurs connues $x = kp$, prend la forme classique d'une sommation de sinus cardinaux pondérés par les valeurs des échantillons :

$$f(x) = \sum_i pg_i \text{sinc}(x - ip)$$

Pour les positions des échantillons $x = kp$, on vérifie que les contributions de tous les autres échantillons s'annulent (cela exprime l'indépendance des échantillons à la fréquence de Nyquist). On voit donc que les échantillons pris à la distance $\frac{1}{2U}$ forment une base pour la reconstruction des signaux dont la bande est limitée à l'intervalle $[-U, U]$.

3.1.3 Reconstruction dans le cas de suréchantillonnage

Si l'on choisit de prendre plus d'échantillons qu'il n'est strictement nécessaire ($p < \frac{1}{2B}$), on se trouve dans la situation décrite sur la figure 3.1, où certaines fréquences, comprises entre U et $\frac{1}{p} - U$, ne contiennent pas d'énergie. Lors de la reconstruction, tout filtre vérifiant :

$$W_N(u) = \begin{cases} 1 & \forall |u| < U \\ 0 & \forall |u| > \frac{1}{p} - U \end{cases} \quad (3.2)$$

est satisfaisant pour la reconstruction. Il existera donc de nombreuses formules de reconstruction du type :

$$f(x) = pg(x) * w_k(x)$$

où les w_k seront autant de fonctions de reconstruction (cf. figure 3.3). Il n'y a donc en général plus d'indépendance entre les échantillons, mais au contraire **redondance**. Ces échantillons ne forment plus une base mais un système générateur. La redondance peut être exploitée de diverses façons, par exemple pour corriger des distorsions, atténuer l'effet d'un bruit ou combler des pertes. Elle permet également de choisir les filtres w_k les plus adaptés à nos contraintes. Elle se paie par un signal plus volumineux.

Parmi toutes les fonctions de reconstruction possibles, la fonction canonique vue en éq.3.1 se distingue car elle garantit que la reconstruction qu'elle procure, même en présence de bruit, sera la plus fidèle au signal original selon un critère d'écart quadratique moyen.

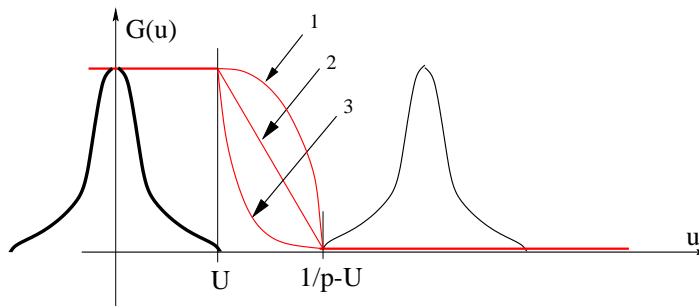


FIG. 3.3 – Trois filtres possibles pour reconstruire exactement un signal suréchantillonné (les filtres sont symétriques pour les u négatifs). Ils diffèrent par leurs valeurs dans l'intervalle compris entre U et $1/p - U$.

3.2 Signaux réels

À quelque échelle que nous l'observions, l'univers nous apparaît comme une fonction des deux variables d'espace, définie en tout point et continue en presque tout point et dotée de détails aussi fins que nous le souhaitons. Il en est ainsi des vues aériennes, des scènes de la rue quotidienne et mêmes des microphotographies, tant les structures observées sont usuellement grandes devant les discontinuités de la matière : moléculaires ou atomiques. L'univers ne se présente donc pas comme une fonction à spectre borné comme nous le souhaiterions, mais comme une fonction à spectre à support quasi-infini.

3.2.1 Les limitations physiques du spectre

Mais l'univers s'observe à travers des instruments (microscope, lentille, caméra) dont la résolution est limitée en particulier par la diffraction (cf. [Pérez, 1991]). Ce sont ces instruments qui introduisent généralement la limitation en bande passante exigée par l'échantillonnage. Au lieu de mesurer l'image exacte $f(x)$ de l'univers, on en

mesure une image dégradée f' par la réponse impulsionnelle $h(x)$ du système d'observation :

$$f'(x) = f(x) * h(x)$$

La limite fréquentielle U est alors imposée par le support de H , fonction de transfert du système. C'est donc à ce signal f' que l'on appliquera les recommandations issues de la théorie de Shannon et exposées dans les lignes précédentes (Section 3.1).

Les limites des optiques

Pour tous les systèmes dont l'optique est parfaite et limitée par le seul diamètre D de la pupille d'entrée, et dans le cas d'une imagerie incohérente, l'optique introduit un filtrage bidimensionnel, fonction de la seule variable $\rho = \sqrt{u^2 + v^2}$:

$$H(\rho) = \frac{J_1^2(\pi\rho D/\phi)}{(\pi\rho D/\phi)^2} \quad (3.3)$$

Cette fonction est connue sous le nom de tache d'Airy [Pérez, 1991], J_1 est la fonction de Bessel d'ordre 1 et ϕ la distance focale de l'optique utilisée. Un tel filtre n'est malheureusement pas un filtre passe-bas parfait mais il est généralement considéré comme approximativement nul au-delà des fréquences $\rho = 1,22D/\phi$.

Les capteurs solides

Une autre limitation du spectre du signal est introduite lorsqu'on utilise des détecteurs solides (matrices ou lignes de photodétecteurs). Ces détecteurs sont constitués d'un assemblage de détecteurs élémentaires placés dans le plan image du système optique, chacun étant chargé de mesurer un échantillon. L'échantillon ainsi mesuré n'est donc pas ponctuel mais il résulte de l'intégration de l'énergie sur la surface du capteur élémentaire. Le filtrage a pour réponse impulsionnelle une fonction porte dont, pour des raisons de construction, la largeur est nécessairement inférieure au pas d'échantillonnage (donné par la distance entre capteurs). Un tel système est lui aussi un mauvais filtre passe-bas puisque sa fonction de transfert (la fonction *sinc*) est à support large d'une part et ne s'annule qu'à l'extérieur de la bande passante compatible avec le pas d'échantillonnage d'autre part. Lui aussi laissera place à du repliement de spectre si l'on ne prend pas soin de filtrer optiquement le signal.

3.2.2 Filtrage du signal

Pour de nombreuses applications, le pas d'échantillonnage p , dont découle le débit de l'échantillonneur, est la grandeur limitante d'un système d'imagerie. C'est par exemple le cas pour certaines applications de télésurveillance ou de télédétection qui utilisent un canal de transmission à débit réduit. Dans ces cas, le problème de l'échantillonnage est pris à l'envers et c'est la bande passante de l'optique qui est ajustée pour assurer la relation $p = \frac{1}{2U}$. On adapte donc la dégradation du signal au taux d'échantillonnage que l'on peut effectuer. Comment choisir alors h si p est fixé ?

Une idée tentante est de filtrer parfaitement le signal original f de façon que son spectre ne déborde pas de la bande de fréquence $[-U, U]$. Cela conduit à choisir $H(u)$ égale à la fonction porte $W_U(u)$. Cette solution est aisée si l'on dispose de la TF du signal $f(x)$, comme en optique cohérente [Cozannet et al., 1981], ou en radar [Maître, 2001]. Il suffit alors de supprimer les fréquences au-delà de U . Lorsque l'on ne peut accéder directement au domaine spectral, un tel filtrage passe-bas idéal est difficile à réaliser puisqu'il nécessite que l'on convole le signal par un sinus cardinal, fonction à support infini. On choisit alors d'approcher cette opération en convolant $f(x)$ par une fonction $h'(x)$ à support "presque-compact" et dont le spectre $H'(u)$ est proche de la fonction porte. Cette approximation entraîne deux types d'erreurs (cf. figure 3.4) :

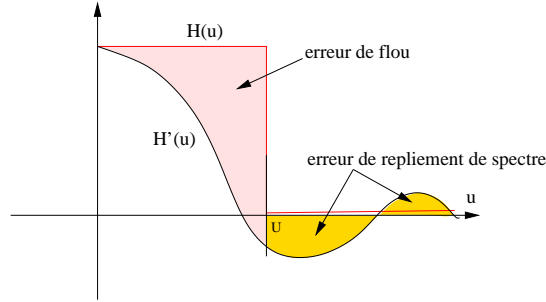


FIG. 3.4 – Les deux types d’erreurs introduites par le filtrage d’un signal par un filtre $h'(x)$ différent du passe-bas idéal $h(x)$ lors de son échantillonnage. Ils sont présentés ici dans l’espace de Fourier. À l’intérieur de la bande, l’erreur est une distorsion des fréquences appelée flou en traitement d’images, à l’extérieur de la bande, l’énergie non annulée viendra se replier lors de l’échantillonnage et créera le défaut de repliement de spectre.

1. tout d’abord, à l’intérieur de la bande passante $[-U, U]$, des fréquences seront atténuées, introduisant un flou sur l’image. Son énergie est donnée par :

$$E_i = \int_{-U}^U F^2(u) [H'(u) - H(u)]^2 du$$

2. d’autre part, à l’extérieur, des fréquences ne seront pas annulées. Lors de l’échantillonnage, ces fréquences répliquées dans les ordres parasites viendront s’insérer dans la bande conservée, introduisant une distorsion du signal appelée repliement de spectre ou *aliasing*. L’énergie concernée est :

$$E_o = 2 \int_U^\infty F^2(u) H'^2(u) du$$

Parmi les fonctions h' les plus fréquemment utilisées, nous pouvons citer les suivantes :

- la fonction porte : $h'_1(x) = W_{p/2}(x)$ qui conduit à $H'_1(u) = p \operatorname{sinc}(up)$;
- la fonction triangle (convolution de 2 portes) : $h'_2(x) = \operatorname{Tri}(x/p)$ qui conduit à $H'_2(u) = p^2 \operatorname{sinc}^2(up)$;
- la spline cubique (convolution d’une porte et d’un triangle) :

$$h'_3(x) = (x)_+^3 - 4(x-1)_+^3 + 6(x-2)_+^3 - 4(x-3)_+^3$$

où : $z_+ = \max(z, 0)$, qui conduit à $H'_3(u) = p^3 \operatorname{sinc}^3(up)$;

- la fonction gaussienne $h'_4(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ qui a pour fonction de transfert la gaussienne de variance $1/\sigma$.

3.3 L’extension en dimension N

L’extension de la théorie de l’échantillonnage à des espaces de dimension N ($N = 2$ pour l’image, $N = 3$ pour les séquences d’images ou pour les volumes en imagerie sismique et médicale, $N = 4$ pour une séquence de volumes temporels, etc.) est simplifiée par la linéarité des opérations dans le domaine spectral qui permettent de décomposer la démarche en autant de problèmes d’échantillonnage que de dimensions. Mais il révèle cependant des différences notables dès que $N > 1$ et entraîne une complexité croissante du choix de la base d’échantillonnage avec N .

Nous noterons par \mathbf{x} et \mathbf{u} les variables vectorielles conjuguées d’espace et de fréquence. A une image continue $f(\mathbf{x})$ est associée sa transformée de Fourier $F(\mathbf{u})$.

$$F(\mathbf{u}) = \iint \dots \int f(\mathbf{x}) \exp(-2i\pi\mathbf{u}\cdot\mathbf{x}) d\mathbf{x}$$

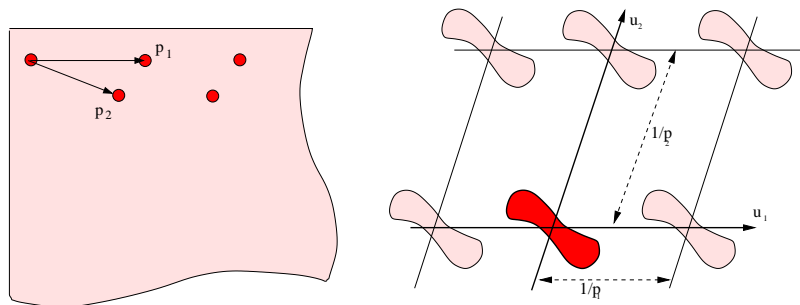


FIG. 3.5 – Échantillonnage d'un signal bidimensionnel selon 2 pas d'échantillonnage p_1 et p_2 (à gauche). Le spectre de l'image est reproduit aux nœuds du réseau réciproque (à droite). Le domaine \mathcal{S} est le domaine où $F(u)$ n'est pas nul ; le domaine \mathcal{V} est la maille de répétition du motif.

Un signal de dimension N est à spectre borné s'il existe U tel que $\forall \mathbf{u}, |\mathbf{u}| > U$. Se donner un échantillonnage de l'espace, c'est choisir un ensemble de N vecteurs d'échantillonnage \mathbf{p} formant une base de l'espace image. À ces vecteurs est associé par transformée de Fourier le réseau réciproque de vecteurs de base \mathbf{k} qui pave l'espace des fréquences (cf. figure 3.5). La reconstruction du signal continu se fera exactement si les ordres parasites ne se superposent pas au spectre du signal.

On voit qu'il est simple d'échantillonner un signal de dimension N de façon à vérifier l'inégalité stricte de Shannon. Il suffit de choisir tous les pas d'échantillonnage p_i de façon que $p_i < 1/2U$. Cette condition suffisante pour faire un échantillonnage parfait n'est cependant pas nécessaire et conduit généralement à un très fort suréchantillonnage et donc à une grande redondance dans les échantillons ainsi mesurés.

On définit l'efficacité de l'échantillonnage η par le rapport du volume \mathcal{S} où le spectre n'est pas nul, sur le volume élémentaire \mathcal{V} de répétition du spectre par les \mathbf{p} :

$$\eta = \frac{\mathcal{S}}{\mathcal{V}} \leq 1$$

On vérifie que l'échantillonnage au critère de Nyquist (cf. Section 3.1.2) correspond au cas $\eta = 1$, sans redondance. S'il est facile d'adapter le pas d'échantillonnage d'un signal monodimensionnel de façon à éliminer la redondance, c'est beaucoup plus difficile pour $N > 1$, et très souvent impossible. En effet, on ne pourra assurer $\eta = 1$ que si le domaine \mathcal{S} a la même géométrie que \mathcal{V} , c'est-à-dire si l'on peut paver l'espace à partir de translats de \mathcal{S} . À 2D, les seules formes vérifiant cette propriété sont les parallélogrammes ou des hexagones, mais il est exceptionnel de trouver des scènes dont le spectre est borné par de telles figures. En conséquence, pour la dimension 2 et les dimensions supérieures, on doit généralement accepter des échantillonnages redondants. On peut alors chercher à maximiser η . Dans le cas général d'un spectre quelconque, cette optimisation se fait de façon empirique. Un cas très fréquemment rencontré est celui de spectres isotropes (\mathcal{S} est une sphère de R^N). Les échantillonnages qui s'imposent alors sont bâtis sur le maillage hexagonal (cf. figure 3.6) :

- à 2D c'est le maillage hexagonal lui-même ;
- à 3D c'est le maillage hexagonal compact ou le maillage cubique centré qui conduit à des maillages appelés en quinconce en vidéo (quinconce ligne et quinconce trame).

Les échantillonnages bâtis sur le maillage carré sont plus redondants mais leur simplicité de mise en œuvre les a rendus populaires malgré leur efficacité médiocre et leurs faibles propriétés topologiques (cf. Chapitre 4).

3.4 Le modèle fractal

Nous allons voir maintenant un modèle d'univers qui s'oppose en tous points à l'approche précédemment développée de l'échantillonnage. C'est le modèle fractal.

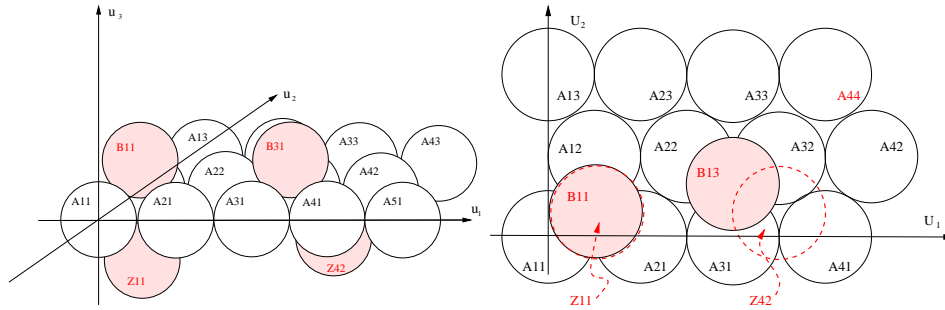


FIG. 3.6 – Échantillonnage de signaux tridimensionnels à spectre isotrope. Les spectres sont représentés par des sphères. la vue de gauche représente une vue perspective du domaine de Fourier, celle de droite une vue de dessus depuis la direction des u_3 positifs. On a représenté un échantillonnage hexagonal dans le plan $u_1 u_2$ (sphères A_{ij}) et 2 schémas différents : l'hexagonal compact et le cubique centré. Dans les deux cas, la couche supérieure (représentée par les sphères B_{ij}) est placée dans les trous de la couche A de façon que chaque sphère B soit tangente à 3 sphères A . Mais dans le schéma hexagonal compact (B_{11} et Z_{11}), la couche inférieure Z est à l'aplomb de la couche B tandis que dans le cas cubique centré (B_{13} et Z_{42}) elle occupe les sites inoccupés de la couche B .

La notion de dimension fractale a été introduite par Mandelbrot [Mandelbrot, 1977] pour rendre compte du comportement surprenant de certains objets soumis à des mesures de plus en plus fines. On s'aperçoit alors que le résultat de la mesure (longueur, périmètre ou surface) dépend fortement de la précision de l'instrument utilisé, et que ce résultat, invariablement, croît sans limite apparente pour des précisions de plus en plus grandes. Cette propriété est bien illustrée par l'exemple d'une côte rocailleuse, tiré de [Richardson, 1961], dont la longueur ne cesse de croître, selon qu'on la mesure sur une carte, sur le terrain en voiture, en vélo, à pied, ou avec un double décimètre. Chaque nouvelle méthode nous permet de contourner des rochers nouveaux, de s'introduire dans des anfractuosités plus petites, ce qui rallonge d'autant l'estimation faite auparavant. Mandelbrot a souligné que ce comportement, qui s'oppose résolument à toutes les notions de l'analyse mathématique classique (fondée sur la régularité et la continuité des fonctions, au moins dans des petits domaines), est en fait très commun et peut être aisément étendu à des objets les plus variés : suspension d'eau savonneuse et ses infinies cascades de particules en émulsion, éponges aux formes torturées, étagement de végétations dans les sous-bois, ramifications sans fin des connexions nerveuses, emboîtement des cratères à la surface de la lune, et bien d'autres encore.

Sur la figure 3.7, nous avons représenté certains de ces objets, tels que la nature nous les propose, véritables casse-têtes pour le traicteur d'image. Une éponge, ou un poumon, présentent, par exemple, une surface continue, mais si déchiquetée, si percée d'alvéoles, si ponctuée de protubérances de toutes tailles qu'elle apparaît finalement remplir tout l'espace : tout le poumon n'est que surface intérieure vers le système sanguin, extérieure vers les bronches et l'atmosphère (cf. figure 3.9).

Cette propriété d'une surface de remplir tout un volume n'est pas sans rappeler les courbes " pathologiques " de Peano ou de Von Koch (fig. 3.8). C'est cette analogie avec ces courbes célèbres que Mandelbrot utilise pour définir deux qualités qui seront pour nous précieuses : la dimension fractale et l'homothétie interne.

De la même façon un phénomène, apparemment ponctuel, pourra avoir une dimension plus grande que 0, et une surface dans l'espace pourra avoir une dimension supérieure à 2.

3.4.1 La dimension fractale

Elle exprime la propriété (soulignée plus haut) d'une courbe qui passe plus ou moins par tous les points d'un plan. On connaît de longue date les dimensions 1, 2, ou 3 attachées respectivement aux courbes, aux surfaces, ou

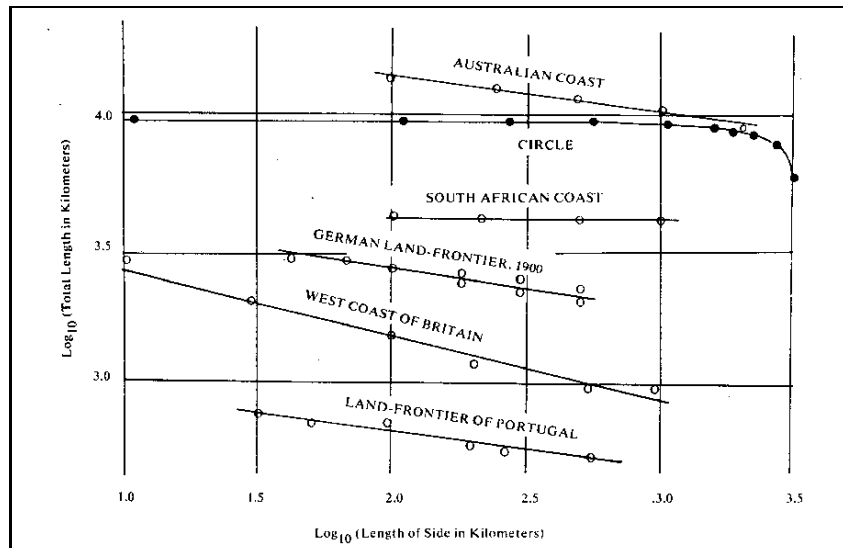


FIG. 3.7 – Évolution de la longueur de quelques frontières et côtes en fonction de la mesure utilisée pour calculer cette longueur (d'après [Richardson, 1961] citée dans [Mandelbrot, 1977]) - échelles logarithmiques.

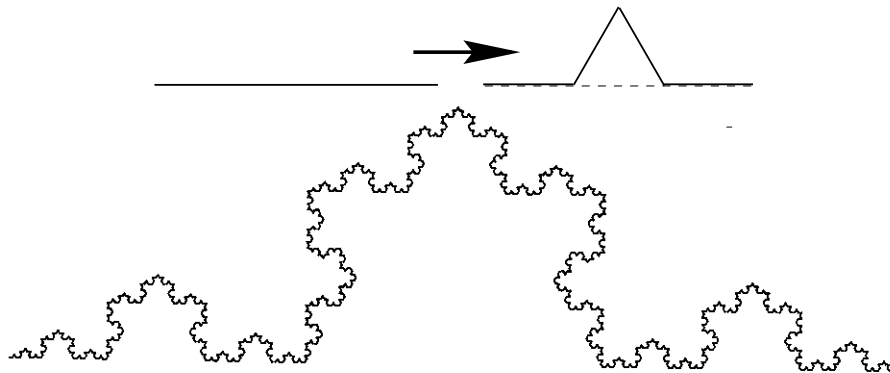


FIG. 3.8 – La courbe de von Koch est un cas d'école de courbe fractale. Elle se construit à partir d'un segment de droite qui est divisé en 3 et remplacé par 4 segments de taille $1/3$. La dimension fractale vaut alors $D = \log 4 / \log 3$.

aux volumes. Mandelbrot postule qu'il existe des courbes de dimension intermédiaire entre 1 et 2, des surfaces de dimension supérieure à 2, et que ces objets possèdent justement la propriété de n'avoir pas de longueur ou de surface précise, pas plus qu'un volume n'a de surface, ou un carré de longueur. Cette dimension, intermédiaire entre les valeurs entières, a été baptisée du néologisme " fractale " afin qu'aucune confusion ne soit faite entre une surface classique (de dimension $D = 2$). L'une et l'autre passent, par exemple, par tous les points d'un plan, mais si l'une est une vraie fonction de deux variables, l'autre n'est que fonction d'une seule variable (ensemble totalement ordonné).

Comment mesure-t-on une dimension, classique ou fractale? Tout d'abord, conformément à la démarche mathématique, il nous faut nous munir d'une mesure. C'est une opération en apparence délicate, mais en pratique

très simple. Elle est discutée dans [Mandelbrot, 1977], et nous n'en reprendrons pas les termes ici. Disons simplement que si nous disposons, par exemple, d'une famille de confettis de rayons décroissants, alors nous sommes à même d'évaluer aussi bien des longueurs ou des surfaces classiques, que des courbes fractales. Pour cela, ayant confronté à un étalon universel la taille du confetti (nous le dénoterons η), il ne nous reste plus qu'à recouvrir la courbe de ces confettis, de façon qu'aucun point de la courbe ne soit plus visible. Le produit du nombre de confettis $N(\eta)$ par la taille élémentaire η : $L(\eta) = \eta \cdot N(\eta)$ est une bonne mesure de la grandeur qui nous intéresse lorsque η tend vers 0 : longueur ou surface par exemple. Mesurons ainsi un segment de droite ou une courbe classique. Si l'on divise η par 10, $N(\eta)$ est multiplié par 10 et la longueur totale est constante. Cette variation, linéaire, de η et $N(\eta)$ est caractéristique des dimensions 1. Si l'on mesure l'aire d'un carré, chaque fois que η est divisé par 10, il faut 100 fois plus de confettis pour recouvrir la surface. La dépendance quadratique indique une dimension 2. Répétée sur une courbe fractale, l'expérience mettra en évidence une dimension intermédiaire entre 1 et 2 : par exemple, $\log 4 / \log 3 = 1,26$ pour la courbe de Von Koch de la figure 3.8, ou bien 1,25 pour la courbe de l'Angleterre conformément à la figure 3.7.

Ainsi, la dimension fractale est déterminée par le graphe de $\log(L(\eta)) = f(\log(\eta))$. Si ce graphe est linéaire la dimension D est donnée par :

$$\log L(\eta) \propto (1 - D) \log \eta \quad (3.4)$$

sinon par son asymptote :

$$\lim_{\eta \rightarrow 0} D = 1 - \frac{\log L(\eta)}{\log \eta} = \frac{-\log N(\eta)}{\log \eta}. \quad (3.5)$$

3.4.2 L'homothétie interne

L'homothétie interne est une idée assez intuitive que certaines formes se retrouvent identiques à elles-mêmes lorsqu'on les observe à des échelles différentes :

- 1 m^3 de nuage est identique à 1 km^3 ,
- la côte de Bretagne est un massif granitique de quelques centaines de km de long, mais la Pointe du Finistère qu'elle contient semble posséder les mêmes propriétés, de même qu'un rocher isolé de cette Pointe ou un morceau de granit de ce rocher,
- la courbe de von Koch (figure 3.8) est exactement identique à elle-même lors d'une division d'échelle de 3.
- un mouvement brownien observé pendant 1 s ou pendant 10 s semble le même.

L'homothétie interne peut être exacte (géométriquement comme pour la courbe de von Koch, ou en loi statistique comme pour le mouvement brownien), ou approchée, comme pour beaucoup de phénomènes physiques. Lorsque plusieurs homothéties s'appliquent à une même structure, on emploie le terme de cascade d'homothéties (comme par exemple pour la courbe de von Koch).

La dimension (fractale) de l'objet D et le nombre N de structures nécessaires pour reconstruire un objet d'une taille donnée à partir d'une réplique k fois plus petite sont liés par la relation :

$$N = k^D. \quad (3.6)$$

Par exemple, un segment de droite est identique à k segments k fois plus petits mis bout à bout, un carré est semblable à k^2 carrés de côté k fois plus petit judicieusement assemblés, et il faut $N = 4$ segments de taille $k = 3$ fois plus petite pour reconstruire un flocon de von Koch.

3.4.3 Que faire de la théorie fractale ?

Voyons maintenant de quelle façon ces surprenantes propriétés interviennent lors de l'étape d'échantillonnage d'une image.

Tout d'abord, l'existence de formes fractales conforte notre opinion que l'approche de Nyquist-Shannon est parfois impuissante à rendre compte de la structure d'un objet, par le simple fait qu'il n'y a pas de fréquence limite qui le caractérise de façon certaine. À quelle résolution faudra-t-il en effet analyser une coupe de poumon pour évaluer la surface des échanges d'oxygène ? Les seules réponses rigoureuses seraient, là-encore, des majeurs bien incompatibles avec nos moyens de traitement : par exemple les plus grosses structures moléculaires, ou bien les plus fines structures cellulaires. C'est pourquoi, dans de nombreuses études (en biologie, métallurgie, ou même télédétection des ressources terrestres), des travaux identiques, menés sur des échantillonnages différents conduisent à des résultats discordants. Le dépouillement automatique des données entraîne souvent une confiance trop grande dans les résultats pour qu'une remise en question des hypothèses fondamentales soit consentie.

En second point, lorsqu'elle est clairement perçue, la notion de forme fractale permet de réaliser une économie de temps d'analyse et de traitement.

Pour cela, il nous faut connaître d'une part la dimension fractale de l'objet, d'autre part l'existence et la limite de son homothétie interne. Cette dernière connaissance est un des points-clefs de l'analyse et sa connaissance nous est habituellement enseignée par le spécialiste : biologiste, métallurgiste, etc. On sait que l'architecture intime de l'univers est hétérogène : les constructions nucléaires précèdent les structures atomiques, qui elles-mêmes engendrent les édifices moléculaires, et ainsi de suite. Clairement, de cette architecture découlent les discontinuités dans les diverses structures, et donc dans les dimensions fractales. La limite d'homothétie est issue de cette double notion de discontinuité ; elle nous informe que, jusqu'à une certaine échelle de résolution, la fonction des structures est bien celle qui nous concerne (échanges gazeux dans le cas des poumons). Nous choisissons alors un pas grossier d'échantillonnage δx , très supérieur à la limite $\delta'x$ qui serait nécessaire pour rendre compte de la totalité du phénomène. De la mesure m faite à cette échelle (exprimée en unité δx), et de la dimension fractale D , nous déduisons alors la valeur vraie m' (exprimée en unités $\delta'x$) par la relation :

$$m' = m \left[\frac{\delta x}{\delta'x} \right]^D \quad (3.7)$$

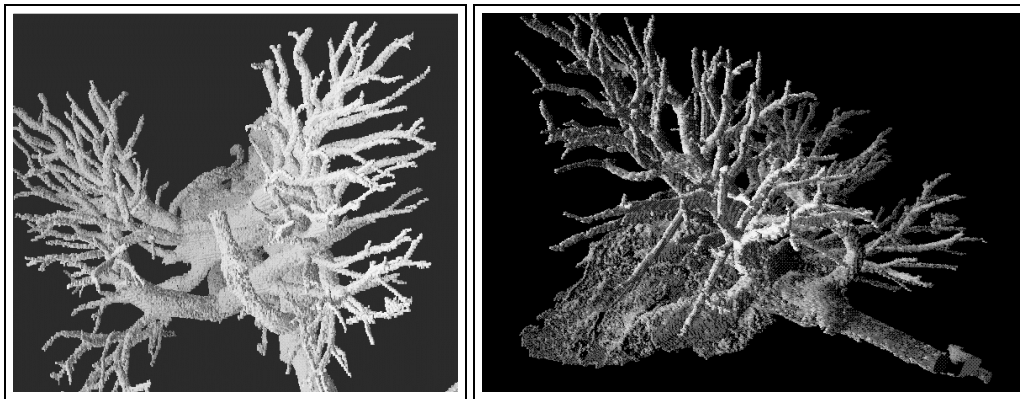


FIG. 3.9 – Quelques exemples de fractales : arbre bronchique de l'homme recalculé à partir du corps du Human Visual Project par Jorge Marquez.

Soulignons à ce point la propriété remarquable du pixel (point élémentaire d'image) de fournir une mesure universelle (au sens où nous l'avons entendu au paragraphe 3.4.1). Contrairement au double décimètre, le pixel permet de mesurer non seulement des longueurs ($D = 1$), mais aussi des surfaces ($D = 2$) ou des volumes ($D = 3$). Et l'équation 3.5 demeure valable, tant pour des formes classiques que fractales pour D variant de 0 à 3. Ainsi cette notion de forme fractale mérite-t-elle une attention particulière ; elle bouleverse nos schémas traditionnels de l'univers peuplé de "bonnes" fonctions, continues, dérivables, mesurables, etc. Elle se montre

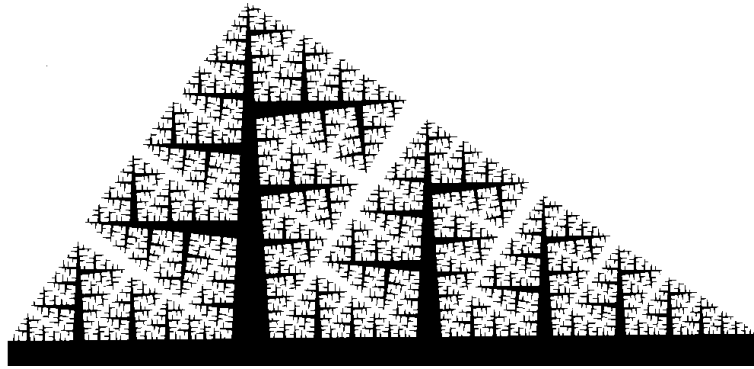


FIG. 3.10 – La forêt étagée de Peano, une forêt fractale artificielle.

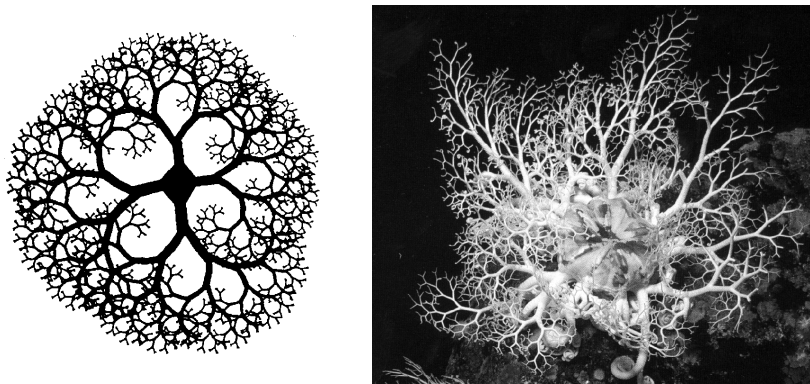


FIG. 3.11 – Corail synthétique et naturel

néanmoins un outil remarquable pour représenter beaucoup des objets rencontrés chaque jour. Elle s'est montrée particulièrement précieuse pour décrire les objets de la géomorphologie : rivières, lignes de crêtes, etc.

Enfin les approches fractales ont prêté leur nom à des outils de compression et de codage d'images (le codage fractal). Ces techniques utilisent les propriétés du point fixe de certaines transformations contractantes du plan (généralement des combinaisons d'homothéties, de rotations et de changements de contraste) [Fisher, 1995]. Cela permet de ne transmettre, pour chaque image issue de la scène à transmettre, que la seule transformation qui permet d'obtenir cette image à partir d'une autre plus grande issue d'une autre partie de l'image (cascade interne).

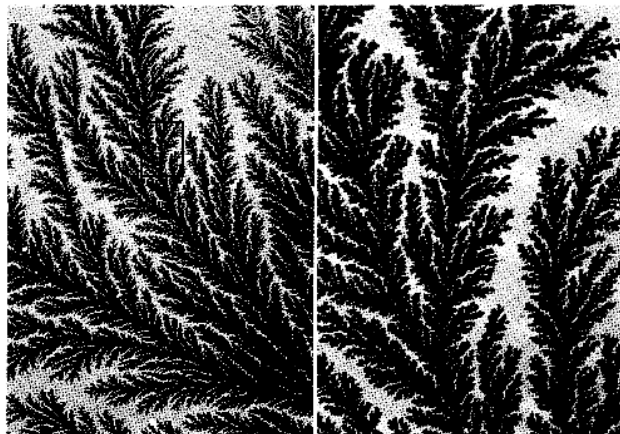


FIG. 3.12 – Electrodeposition de cuivre, c'est un exemple de percolation fractale. La cascade d'homothétie est illustrée par la petite fenêtre de l'image gauche, agrandie sur l'image droite (d'après [Bunde et Havlin, 1994]).

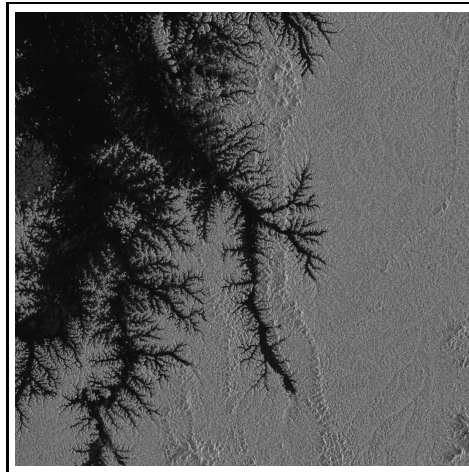


FIG. 3.13 – Un lac en forêt guyanaise vu par le satellite radar ERS-1 ©ESA.

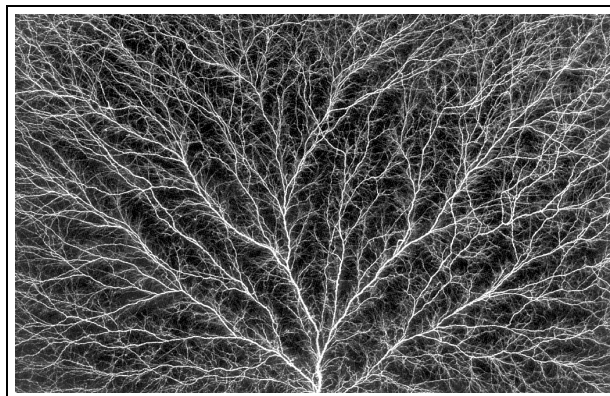


FIG. 3.14 – Décharge électrique claquant un diélectrique (d'après [Bunde et Havlin, 1991]).

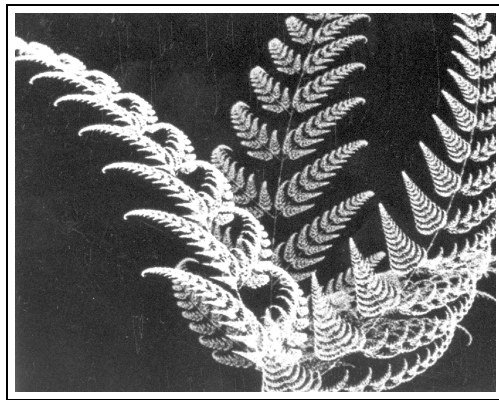


FIG. 3.15 – Fougère simulée par 5 attracteurs superposés (d'après [Peitgen et Saupe, 1988]).

Chapitre 4

Représentations discrètes

Chapitre rédigé par Isabelle BLOCH

4.1 Introduction

Le traitement des images par ordinateur nécessite de disposer de représentations discrètes des scènes observées. On dispose alors d'un échantillonnage de \mathbb{R}^2 ou \mathbb{R}^3 dans \mathbb{Z}^2 ou \mathbb{Z}^3 , qui est de plus fini, ce qui justifie de représenter les images sous forme matricielle. Pour traiter de tels ensembles de points, deux solutions sont possibles :

- soit on plonge les points discrets de \mathbb{Z}^n dans \mathbb{R}^n , et on les traite comme s'ils appartenaient à un espace continu ; ce sont alors souvent des procédés de géométrie algorithmique qui sont mis en œuvre ;
- soit on définit directement les traitements dans l'espace discret, en essayant d'en préserver les effets et les propriétés.

Le choix de l'une ou l'autre de ces solutions dépend beaucoup du type de traitement à effectuer, du type de représentation sur lequel s'appuie ce traitement, des propriétés souhaitées, de la complexité des calculs, etc. Par exemple, on garantit de meilleures propriétés de connexité avec des rotations continues que des rotations discrètes. Au contraire, les opérations morphologiques de base sont beaucoup plus simples à effectuer dans un espace discret (voir le chapitre 6 sur la morphologie mathématique). Dans les cas où les deux approches sont possibles, elles ne sont pas toujours équivalentes, et il est souvent difficile de garantir que la transformation appliquée à une version discrète d'un objet fournira le même résultat que la discrétisation du résultat de la transformation continue.

Dans ce chapitre, nous nous intéressons surtout aux problèmes de représentations géométriques discrètes, qui serviront de support aux traitements et transformations effectués directement dans \mathbb{Z}^n . Des exemples de tels traitements seront abordés en particulier dans le chapitre de morphologie mathématique (chapitre 6). Nous aborderons ici succinctement les pavages et maillages (section 4.2), sur lesquels sont définies les images discrètes, puis les questions de topologie (section 4.3), les problèmes de représentations de quelques structures géométriques simples, déterministes (section 4.4) ou aléatoires (section 4.5), et enfin la définition et le calcul de distances discrètes (section 4.6). Un exposé plus détaillé sur tous ces points ainsi que sur d'autres aspects de la géométrie discrète en traitement d'images peut être trouvé dans [Chassery et Montanvert, 1991, Klette et Rosenfeld, 2004].

Les principes décrits peuvent être appliqués aussi bien à la synthèse qu'à l'analyse d'images. La géométrie discrète constitue toujours un axe de recherche important dans ces deux domaines.

4.2 Pavages et maillages

4.2.1 Définitions et contraintes

Un pavage est une partition de l'espace continu (\mathbb{R}^n) en cellules élémentaires (tout l'espace est recouvert par les cellules, et les cellules n'ont pas d'intersection deux à deux).

Cette définition, la plus générale, peut donner lieu à une infinité de solutions si l'on n'impose pas de contraintes supplémentaires sur la forme et la disposition des cellules. En pratique, des contraintes sont imposées :

- d'une part par les capteurs, dont la surface sensible a généralement une structure régulière,
- et d'autre part par l'utilisation que l'on souhaite faire de la représentation, qui impose une certaine régularité et simplicité, ainsi que certaines propriétés sur le pavage résultant sur lesquelles nous reviendrons.

Deux méthodes peuvent être envisagées pour construire un tel pavage :

1. La première consiste à se donner une distribution de points dans l'espace. Puis à chaque point P est associée une cellule V_P , de telle sorte qu'elle n'intersecte aucune autre cellule et qu'elle ne laisse pas de « vide » dans l'espace. La solution la plus simple consiste à choisir pour V_P l'ensemble des points de l'espace qui sont les plus proches de P que de tout autre point de la distribution. Dans le cas où la distribution des points est régulière, on aboutit à un pavage classique, tel que ceux que nous verrons dans la section suivante. Dans le cas où la distribution est irrégulière, on aboutit à un pavage de Voronoï, que nous décrirons plus en détails dans la section 4.4. Les points de la distribution peuvent être interprétés comme des germes, à partir desquels on fait croître des cellules à vitesse constante, jusqu'à ce que tout l'espace soit rempli (voir figure 4.1).

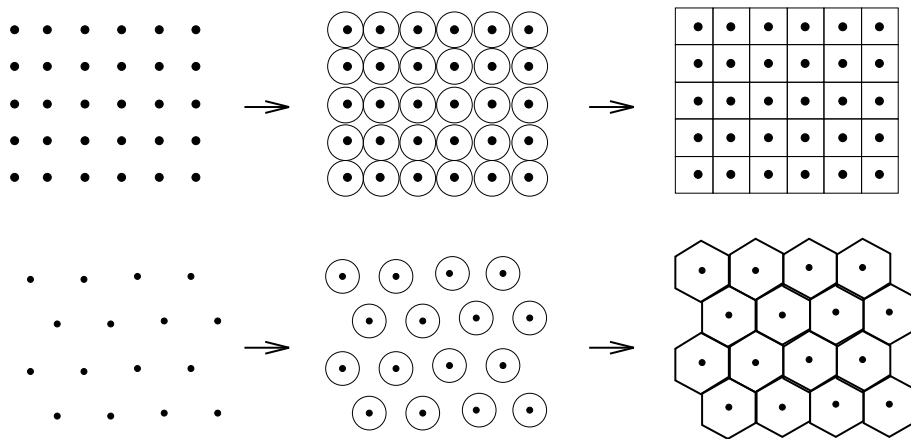


FIG. 4.1 – Deux exemples de construction d'un pavage par propagation à partir d'une distribution de points régulière.

2. La deuxième solution consiste à se donner au contraire un modèle a priori de cellule élémentaire, et à reproduire ce modèle en juxtaposant les cellules pour construire une partition de l'espace [Grunbaum et Shepard, 1989]. Cela n'est bien sûr pas possible avec n'importe quel modèle de cellule. On impose généralement que celui-ci soit un polygone (ou polyèdre) convexe, régulier, et que dans la juxtaposition les sommets soient en contact avec d'autres sommets mais pas avec des arêtes (typiquement, la configuration de la figure 4.2 ne satisfait pas cette dernière condition).

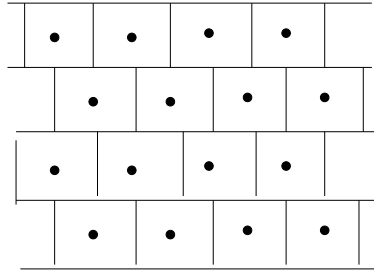


FIG. 4.2 – Exemple d’une configuration où les sommets du pavage ne sont pas en coïncidence avec d’autres sommets. Notons qu’en déformant ce maillage, on peut obtenir un pavage hexagonal qui est le pavage régulier associé à cette distribution de points, comme on le verra par la suite.

4.2.2 Pavages plans réguliers

Les contraintes les plus strictes que l’on puisse imposer aux cellules élémentaires est qu’elles soient toutes identiques et « régulières », c’est-à-dire qu’elles soient des polygones dont tous les côtés et tous les angles soient égaux, et que chaque sommet soit en contact avec un nombre fixe de sommets d’autres cellules. Avec de telles cellules, on obtient des pavages dits « réguliers ».

On montre que, dans le plan, il n’existe que trois types de pavages réguliers, correspondant respectivement à des cellules triangulaires¹, carrées et hexagonales. Ils sont illustrés sur la figure 4.3.

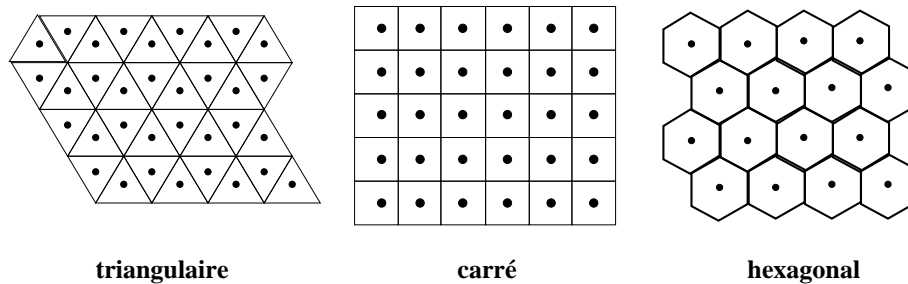


FIG. 4.3 – Les trois pavages réguliers du plan.

4.2.3 Pavages plans semi-réguliers

Si l’on relâche un peu les contraintes précédentes, en permettant aux cellules d’être de types différents (avec des nombres de côtés pouvant varier d’une cellule à l’autre), tout en conservant le nombre constant de cellules adjacentes, on obtient des pavages « semi-réguliers » [Grunbaum et Shepard, 1989].

Il existe 21 solutions respectant ces contraintes. Seules 11 définissent effectivement une partition (dont les 3 régulières vues précédemment). Deux de ces solutions sont illustrées sur la figure 4.4. L’ensemble de ces solutions est décrit dans [Chassery et Montanvert, 1991].

Si l’on relâche plus de contraintes, on peut obtenir des pavages non uniformes, ou même complètement irréguliers (voir par exemple les dessins d’Escher [Locher et al., 1972]).

¹Notons que le pavage triangulaire est régulier au sens où nous l’avons défini, mais qu’il n’est pas uniforme car les triangles n’ont pas tous la même orientation.

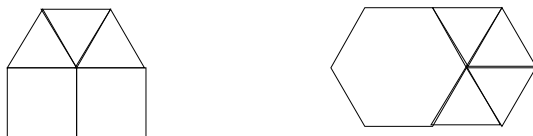


FIG. 4.4 – Deux exemples de configurations semi-régulières permettant de paver le plan.

4.2.4 Dualité pavage / maillage

À partir d'un pavage, on associe à chaque cellule V_P constituant le pavage un point P situé à l'intérieur de cette cellule. Par exemple, P peut être le centre de gravité de V_P (il sera bien à l'intérieur de V_P si l'on se restreint comme ci-dessus à des cellules convexes). Dans ce cas, on retrouve bien sûr des P et des V_P qui correspondent à la construction du pavage à partir d'une distribution de points. Ce sont ces points P qui sont appelés pixels en traitement d'images.

Un maillage est alors défini de la manière suivante : à tout point P on associe les points Q tels que V_P et V_Q ont une arête commune (dans le plan). Le maillage est constitué de tous les segments $[P, Q]$ que l'on peut construire de cette manière.

Ce maillage définit dans l'espace un nouveau pavage, dont les cellules sont centrées sur les sommets du pavage initial. Il y a donc dualité entre pavage et maillage.

Pour les pavages réguliers, qui sont les plus utilisés, cette dualité se traduit de la manière suivante (voir figure 4.5) :

- le maillage associé au pavage triangulaire est hexagonal,
- le maillage associé au pavage carré est carré,
- le maillage associé au pavage hexagonal est triangulaire.

Le maillage carré est certainement le plus utilisé. Le maillage hexagonal n'est cependant pas sans intérêt car il possède de meilleures propriétés d'isotropie et de topologie que le maillage carré. Le maillage triangulaire n'est en revanche presque jamais utilisé.

Dans l'espace discret à 3 dimensions, le maillage cubique est le plus utilisé. C'est aussi le seul qui soit régulier. On trouve aussi quelques applications de maillages cubiques à faces centrées ou rhombododécaédriques.

4.3 Topologie discrète

Dans cette section, nous abordons les problèmes de topologie sur les maillages discrets. Nous verrons en particulier que la topologie discrète classique n'est pas adaptée et qu'une approche fondée sur la définition de voisinages est préférable.

4.3.1 Quelques approches

Considérons tout d'abord la topologie discrète classique. Pour cette topologie, tout point de l'ensemble discret considéré est un ouvert. Comme un ensemble connexe est un ensemble qui ne peut pas se décomposer en réunion de deux ouverts non vides disjoints, les seuls ensembles connexes que l'on puisse construire avec cette topologie sont réduits à des singletons.

Cependant du point de vue de l'analyse d'images et de la reconnaissance des formes, on a besoin de définir des objets, que l'on considère comme des entités connexes généralement constituées de plusieurs points. La topologie discrète classique ne permet pas de répondre à ces exigences, et elle n'est donc pas adaptée aux besoins du traitement d'images.

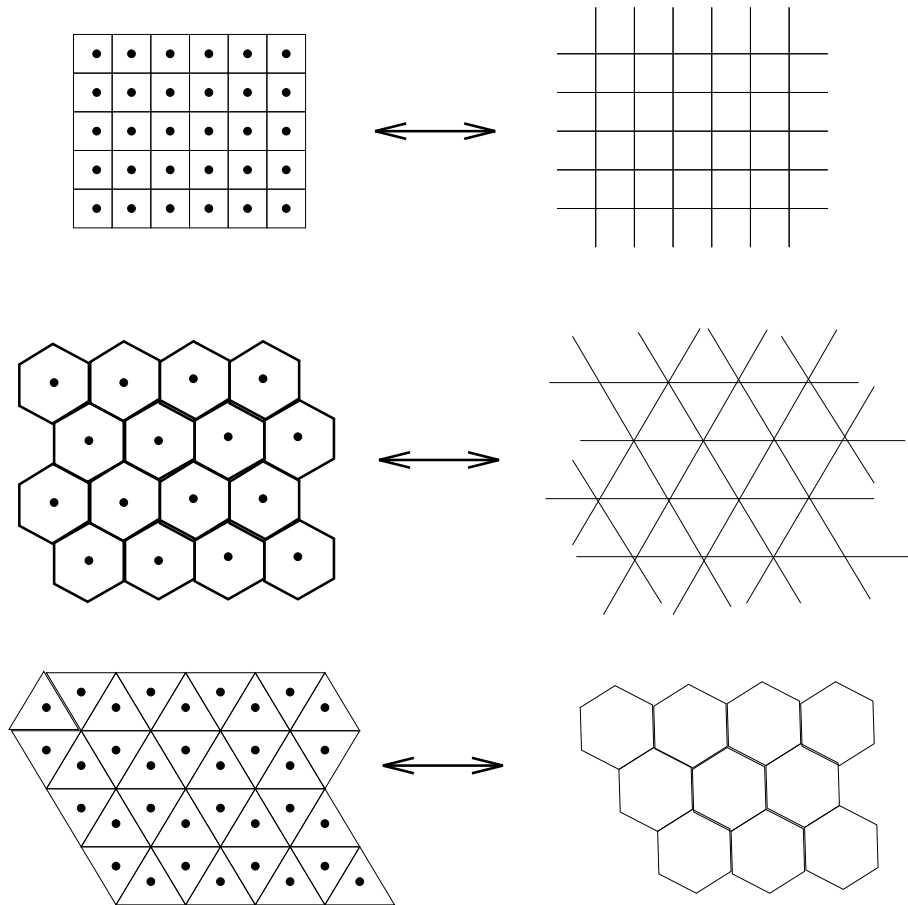


FIG. 4.5 – Dualité entre les maillages et les pavages réguliers du plan.

Plusieurs approches ont été tentées afin de définir des topologies plus satisfaisantes. Nous décrivons ici succinctement l'une d'elles [Chassery et Chenin, 1980, Khalimsky et al., 1990]. Elle consiste (dans le plan) à construire une base topologique de \mathbb{Z}^2 . Une topologie est alors une classe \mathcal{T} d'ensembles de \mathbb{Z}^2 telle que chaque élément de \mathcal{T} soit réunion dénombrable d'éléments de la base topologique, et telle que :

1. $\mathbb{Z}^2 \in \mathcal{T}, \emptyset \in \mathcal{T}$,
2. $\forall (A, B) \in \mathcal{T}^2, A \cap B \in \mathcal{T}$.

Sur une trame carrée, on peut par exemple construire la base topologique de la manière suivante : pour tout point P de \mathbb{Z}^2 , de coordonnées (i, j) , on associe à P l'ensemble $\mathcal{U}(P)$ défini par (voir figure 4.6) :

$$\begin{aligned} \mathcal{U}(P) &= \{P, (i-1, j), (i+1, j), (i, j-1), (i, j+1)\} \text{ si } i+j \text{ pair,} \\ &= \{P\} \text{ sinon.} \end{aligned}$$

Les ensembles $\mathcal{U}(P)$ ainsi construits vérifient bien les conditions suivantes :

1. $\mathbb{Z}^2 = \cup_{P \in \mathbb{Z}^2} \mathcal{U}(P)$,
2. si $\mathcal{U}(P)$ et $\mathcal{U}(Q)$ sont deux éléments de la base topologique, alors $\mathcal{U}(P) \cap \mathcal{U}(Q)$ est une réunion d'éléments de cette base.

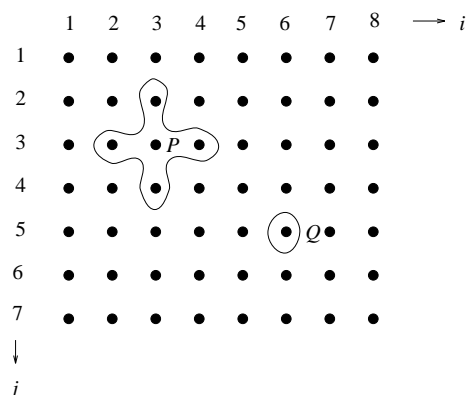


FIG. 4.6 – Définition d’une base topologique dans le plan discret sur une trame carrée.

Ce type de construction a un inconvénient majeur : la forme du voisinage induit par les $\mathcal{U}(P)$ autour de chaque point dépend de la position du point dans l’espace. En particulier, la topologie ainsi construite n’est pas invariante par translation. De plus, ce type de construction ne garantit pas toutes les propriétés souhaitées, en particulier le théorème de Jordan. Enfin, si cette construction reste possible sur une trame triangulaire (voir figure 4.7), elle ne l’est pas sur une trame hexagonale.

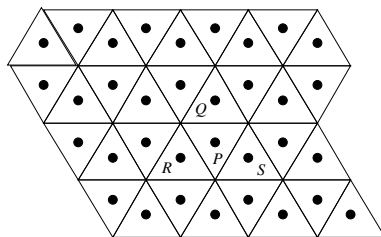


FIG. 4.7 – Définition d’une base topologique dans le plan discret sur une trame triangulaire. Sur cette figure, $\mathcal{U}(P) = \{P, Q, R, S\}$, $\mathcal{U}(Q) = \{Q\}$, $\mathcal{U}(R) = \{R\}$, $\mathcal{U}(S) = \{S\}$.

4.3.2 Topologie à partir de la notion de voisinage élémentaire

Pour le traitement d’images, il est donc difficile de définir des notions topologiques satisfaisantes et opérationnelles à partir de la construction d’ouverts. Il est préférable de définir d’abord les notions de voisinage élémentaire, puis de connexité, qui est une des notions topologiques essentielles pour l’interprétation des images et la reconnaissance des formes [Rosenfeld, 1970, Rosenfeld, 1979].

Pour cela, on considère l’image discrète comme un graphe. Les nœuds du graphe sont les points discrets (les pixels de l’image). Les arcs sont définis par les relations de voisinage entre les points. Par exemple, sur une trame carrée, on peut considérer que deux points sont voisins si une et une seule de leurs coordonnées diffère d’une unité. Chaque point (encadré sur la figure 4.8 a) a ainsi 4 voisins, et on parle de 4-connexité. Le graphe correspondant est illustré sur la figure 4.9.

Si l’on ajoute dans le graphe les arcs reliant les points aussi en diagonale (dont les deux coordonnées diffèrent d’une unité), chaque point a alors 8 voisins, et on parle de 8-connexité (figure 4.8 b). En 4-connexité, très peu de directions de l’espace sont représentées, en 8-connexité, tous les voisins ne sont pas à la même distance du point central.

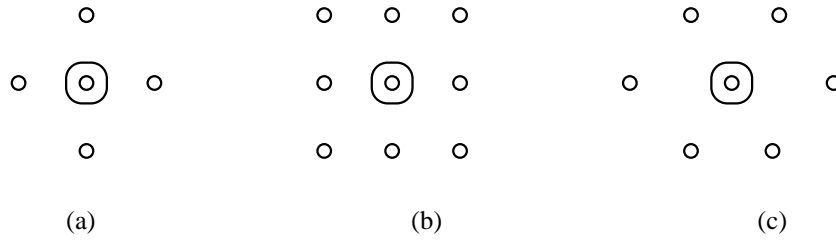


FIG. 4.8 – Connexité sur une trame discrète carrée (a : 4-connexité, b : 8-connexité), ou hexagonale (c : 6-connexité).

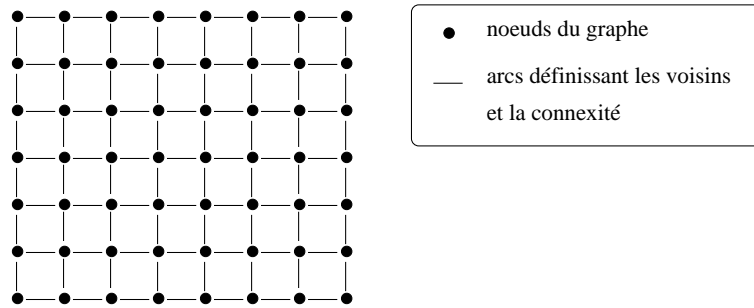


FIG. 4.9 – Graphe définissant la 4-connexité sur une trame carrée.

Sur une trame hexagonale, la connexité la plus naturelle est la 6-connexité (figure 4.8 c). Chaque point a 6 voisins qui sont tous à même distance du point central et les 6 directions représentées sont équiréparties, ce qui constitue un des avantages de la trame hexagonale.

Sur une trame triangulaire, trois types de voisinages élémentaires peuvent être définis, correspondant respectivement à la 3-, 9- et 12-connexité. Ces définitions sont illustrées sur la figure 4.10.

Dans une représentation matricielle des images, l'accès aux voisins d'un point de coordonnées (i, j) se fait très aisément sur une trame carrée. Ce sont :

- en 4-connexité : les points de coordonnées $(i - 1, j)$, $(i + 1, j)$, $(i, j - 1)$, $(i, j + 1)$,
- en 8-connexité : les points de coordonnées $(i - 1, j)$, $(i + 1, j)$, $(i, j - 1)$, $(i, j + 1)$, $(i - 1, j - 1)$, $(i - 1, j + 1)$, $(i + 1, j - 1)$, $(i + 1, j + 1)$.

Sur une trame hexagonale, l'accès aux voisins est un peu plus délicat. En effet, s'il est toujours possible de représenter une image de manière matricielle, il faut considérer que, géométriquement, les lignes sont décalées les unes par rapport aux autres. Les indices des voisins d'un point dépendent donc de la parité de la ligne sur laquelle se situe le point. Si les lignes sont numérotées comme l'indique la figure 4.11, alors les voisins d'un point de coordonnées (i, j) sont :

- si j est pair : $(i - 1, j - 1)$, $(i, j - 1)$, $(i - 1, j)$, $(i + 1, j)$, $(i - 1, j + 1)$, $(i, j + 1)$,
- si j est impair : $(i, j - 1)$, $(i + 1, j - 1)$, $(i - 1, j)$, $(i + 1, j)$, $(i, j + 1)$, $(i + 1, j + 1)$.

Sur une trame triangulaire, de la même manière, les indices des voisins dépendent de la parité de la ligne sur laquelle se trouve le point considéré (et bien sûr de la connexité choisie).

Si l'on se place maintenant en dimension trois sur une trame cubique, trois types de voisinages élémentaires sont classiquement définis, correspondant respectivement à la 6-, 18- et 26-connexité. Si l'on représente chaque point P de l'image 3D comme un cube élémentaire, les 6-voisins correspondent aux cubes adjacents à P par une face, les 18-voisins comprennent les 6 précédents et ceux qui sont adjacents à P par une arête, et les 26-voisins

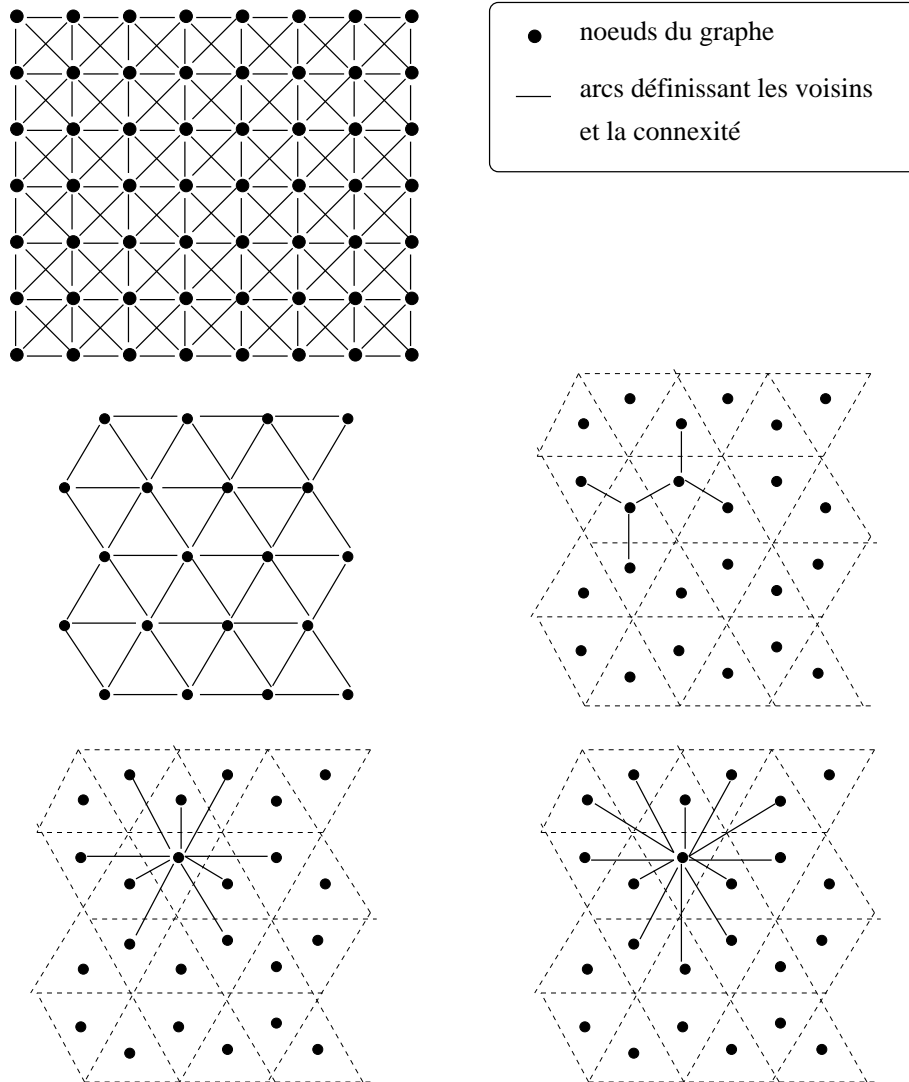


FIG. 4.10 – Graphes définissant la 8-connexité sur une trame carrée, la 6-connexité sur une trame hexagonale, et les 3-, 9- et 12-connexités sur une trame triangulaire (seuls quelques arcs sont représentés pour la trame triangulaire).

s'obtiennent en ajoutant les cubes adjacents par un sommet (voir figure 4.12).

Une fois que sont définis les voisinages élémentaires, les notions de connexité s'en déduisent aisément, grâce à l'interprétation en termes de théorie des graphes [Berge, 1958].

Un chemin est ainsi défini comme une suite de sommets du graphe telle que deux points consécutifs de la suite soient joints par un arc du graphe, où les arcs sont définis en fonction de la connexité choisie. Par exemple, sur une trame carrée, on définit :

- un 4-chemin comme une suite de points $(i_k, j_k)_{1 \leq k \leq n}$ tels que :

$$\forall k, 1 \leq k < n, |i_k - i_{k+1}| + |j_k - j_{k+1}| \leq 1,$$

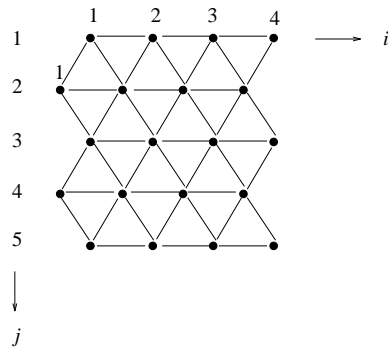


FIG. 4.11 – Les indices des voisins d’un point sur une trame hexagonale dépendent de la parité de la ligne sur laquelle se trouve le point.

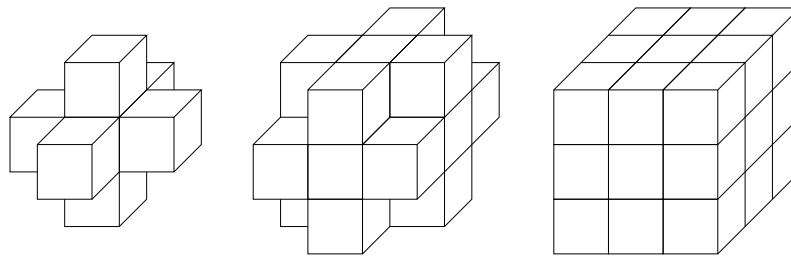


FIG. 4.12 – Les trois types de voisinages élémentaires pour une image 3D définie sur une trame cubique.

- un 8-chemin comme une suite de points $(i_k, j_k)_{1 \leq k \leq n}$ tels que :

$$\forall k, 1 \leq k < n, \max(|i_k - i_{k+1}|, |j_k - j_{k+1}|) \leq 1.$$

Les composantes connexes s’en déduisent alors en reprenant les définitions classiques des graphes. On obtient ainsi sur une trame carrée :

- une composante 4-connexe est un ensemble de points \mathcal{S} tel que pour tout couple de points (P, Q) de \mathcal{S} , il existe un 4-chemin d’extrémités P et Q et dont tous les points soient dans \mathcal{S} ;
- une composante 8-connexe est un ensemble de points \mathcal{S} tel que pour tout couple de points (P, Q) de \mathcal{S} , il existe un 8-chemin d’extrémités P et Q et dont tous les points soient dans \mathcal{S} .

Ces définitions sont illustrées sur la figure 4.13.

De manière analogue, on définit sur une trame hexagonale les notions de 6-chemin et de composante 6-connexe. Sur une trame triangulaire, les chemins peuvent être 3-, 9- ou 12-connexes, de même que les composantes connexes.

Sur une trame carrée, des paradoxes peuvent apparaître pour certaines configurations locales de points de l’objet et de son complémentaire. Sur la figure 4.14, en 8-connexité, les points a et d de l’objet sont voisins ainsi que les points b et c du complémentaire. Cela signifie que les composantes connexes de l’objet et du complémentaire se « croisent ». Si on se place au contraire en 4-connexité, ni les points a et d, ni les points b et c ne sont voisins,

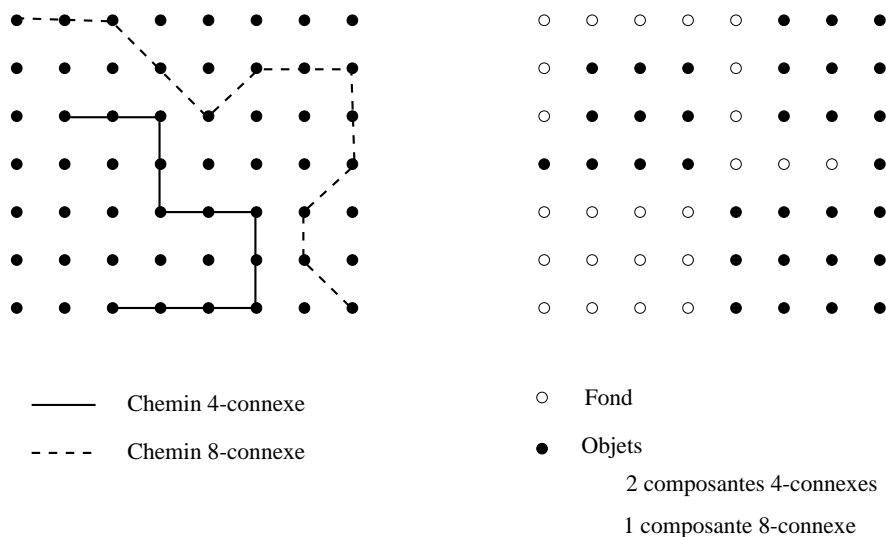


FIG. 4.13 – Illustration des définitions des chemins et composantes connexes sur une trame carrée, en 4- et 8-connexités.

laissant ainsi une zone qui n'est ni l'objet ni son complémentaire. Ce paradoxe peut être évité si l'on considère des connexités différentes pour l'objet et pour son complémentaire.

Sur une trame hexagonale, ce type de phénomène ne se produit pas et on travaille en 6-connexité aussi bien pour la forme que pour le fond.



FIG. 4.14 – Paradoxe sur une trame discrète carrée en 8- et 4-connexité.

Un des résultats importants qui permet de garantir de bonnes propriétés topologiques aux objets (et d'éviter ce type de paradoxe) est le théorème de Jordan. Dans le cas continu, il s'exprime sous la forme suivante : toute courbe simple fermée sépare l'espace en deux composantes connexes, l'intérieur et l'extérieur de la courbe. Avec les définitions de la connexité discrète données ci-dessous, on montre les deux résultats suivants :

- sur une trame carrée, tout 4-chemin (respectivement 8-chemin) simple fermé² sépare l'espace en deux composantes 8-connexes (respectivement 4-connexes), l'intérieur et l'extérieur (voir l'illustration de la figure 4.15) ;
- sur une trame hexagonale, tout 6-chemin simple fermé sépare l'espace en deux composantes 6-connexes.

²Un 4-chemin simple fermé est un 4-chemin (A_0, \dots, A_n) tel que $n \geq 4$, $A_i = A_j$ si et seulement si $i = j$, et A_i est un 4-voisin de A_j si et seulement si $i = j \pm 1$ ou $i = j \pm 2$.

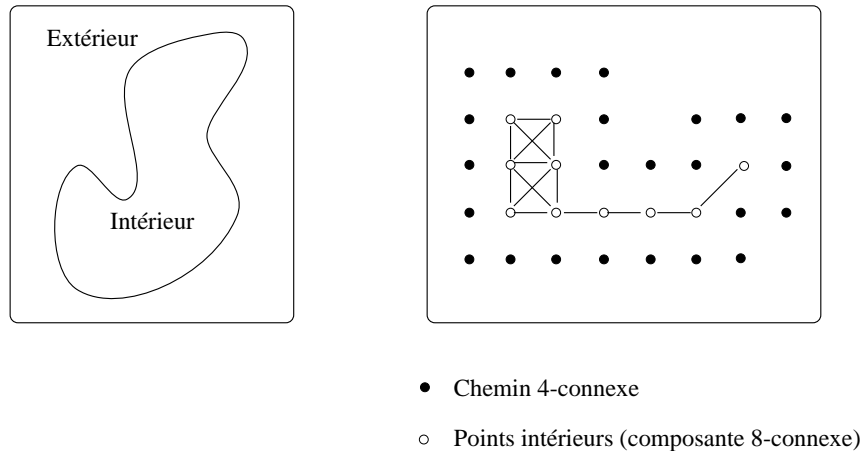


FIG. 4.15 – Illustration du théorème de Jordan dans le cas continu et dans le cas discret sur une trame carrée (les points extérieurs ne sont pas représentés).

Le premier de ces résultats fait apparaître une dualité entre la 4- et la 8-connexité [Kong et Rosenfeld, 1989]. Ainsi, pour garantir que le théorème de Jordan soit respecté, il faut changer de connexité quand on passe d'un objet à son complémentaire. On considérera donc par exemple un objet en 4-connexité et son complémentaire en 8-connexité, ou le contraire. Sur l'exemple de la figure 4.15, les points de l'intérieur du chemin 4-connexe forment bien une composante 8-connexe, mais deux composantes 4-connexes. Le résultat n'est donc pas vrai si on prend la même connexité pour les points du chemin et pour le complémentaire du chemin.

Le deuxième résultat montre qu'au contraire ce type de problème ne se pose pas en 6-connexité. La trame hexagonale présente donc de meilleures propriétés topologiques que la trame carrée.

Sur une trame triangulaire, il existe de manière analogue une dualité entre la 3- et la 12-connexité, mais, comme nous l'avons souligné, cette trame est peu usitée.

En dimension 3 sur une trame cubique, il existe une dualité entre la 6- et la 26-connexité, et on a un analogue du théorème de Jordan en considérant cette fois une surface simple fermée.

La difficulté rencontrée en 4- et 8-connexités vient de ce que l'appartenance des frontières d'un pixel à un objet ou à son complémentaire n'est pas clairement définie. Elle peut être levée dans le cadre des complexes cellulaires [Kovalesky, 1989]. Chaque pixel ou voxel est décomposé en éléments de dimension locale variant de 0 (les sommets) à 2 (l'intérieur du carré représentant un pixel) ou 3 (l'intérieur du cube représentant un voxel), et l'appartenance de chacun de ses éléments aux objets et à leur complémentaire est définie. Ces structures ne posent pas de problèmes topologiques et le théorème de Jordan est vérifié, ainsi que l'illustre la figure 4.16. Ces structures peuvent également être généralisées à trois dimensions [Cointepas et al., 2001]. Il faut cependant noter que ces structures sont un peu plus lourdes à manipuler que les pixels ou les voxels car elles sont plus volumineuses et non homogènes (plusieurs types d'éléments, dont les voisinages sont également différents).

4.3.3 Nombre d'Euler : un exemple de caractéristique topologique d'un objet

Un objet ou un ensemble de points peut être caractérisé par plusieurs types de mesures topologiques. Les plus simples sont le nombre de composantes connexes dont il est constitué et le nombre de trous qu'il contient. La différence entre ces deux nombres est appelé nombre d'Euler, et est utilisé dans des problèmes de reconnaissance des formes, comme un des attributs qui caractérisent un objet.

Si l'on appelle N_{cc} le nombre de composantes connexes d'un objet et N_t le nombre de trous, le nombre d'Euler

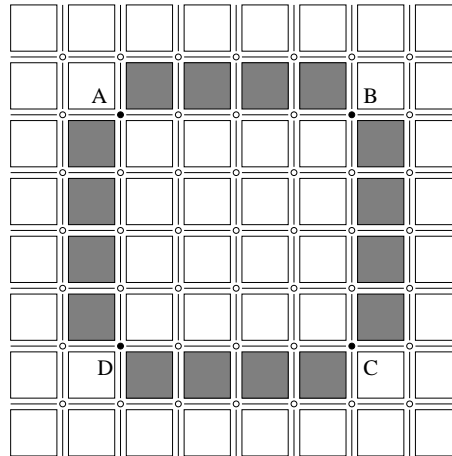


FIG. 4.16 – Le théorème de Jordan est toujours vérifié dans le modèle des complexes cellulaires. Chaque pixel est décomposé en carré, arête et sommet. La couleur des points A , B , C et D détermine la connexité de la courbe noire et celle du fond.

vaut $E = N_{cc} - N_t$. La connexité utilisée pour calculer le nombre de trous et le nombre de composantes connexes doit respecter les contraintes du théorème de Jordan discret. Sur l'exemple de la figure 4.17, le nombre d'Euler vaut -1 si on considère les objets en 8-connexité et les trous en 4-connexité, et 0 si on choisit les conventions inverses.

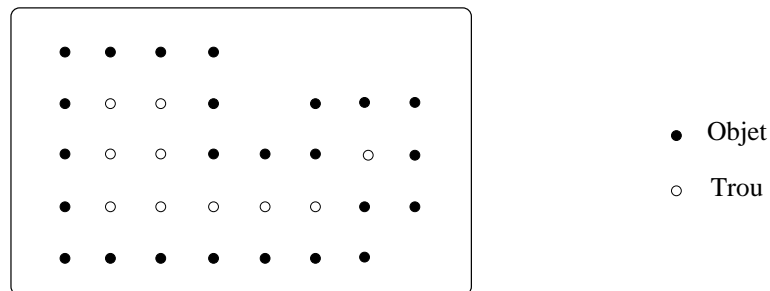


FIG. 4.17 – Nombre d'Euler sur une trame carrée : si on considère les objets en 8-connexité et les trous en 4-connexité, $N_{cc} = 1$ et $N_t = 2$, donc $E = -1$. Avec les conventions inverses, $N_{cc} = 1$ et $N_t = 1$, donc $E = 0$.

Le nombre d'Euler peut être calculé par simple dénombrement de configurations locales, décrites dans la figure 4.18. Avec les notations de cette figure, on montre que :

- si on considère les objets en 8-connexité et les trous en 4-connexité, alors :

$$E = v - e - d + t - q,$$

- si on considère les objets en 4-connexité et les trous en 8-connexité, alors :

$$E = v - e + q,$$

Le lecteur pourra vérifier les résultats de ces formules sur l'exemple de la figure 4.17.

Il est remarquable que ces caractéristiques ne fassent intervenir que des points de l'objet, alors que le nombre d'Euler implique à la fois l'objet et son complémentaire. Cet exemple est également particulièrement intéressant

v	•
e	• • • •
d	• •
t	• • • • • • • • • • • •
q	• • • •

FIG. 4.18 – Configurations locales pour le calcul du nombre d’Euler.

car il montre qu’il est possible de calculer le nombre d’Euler uniquement avec des caractéristiques locales, bien qu’il s’agisse d’une mesure globale sur un objet. Le calcul est donc beaucoup plus facile, et la mise en œuvre informatique avec ces caractéristiques locales est immédiate, alors que le calcul direct des nombres de composantes connexes et de trous serait plus difficile.

Cet exemple illustre ainsi une idée qu’on retrouvera souvent en traitement d’images, qui consiste à ramener des notions globales à des calculs qui peuvent se faire simplement localement. On voit apparaître ici un intérêt supplémentaire de la notion de voisinages élémentaires, puisque ce sont les points définis par ces voisinages qui interviennent dans les calculs locaux.

4.4 Représentations géométriques

Si la plupart des entités géométriques simples (droites, cercles, courbes) sont parfaitement maîtrisées dans le cas continu, leurs représentations discrètes soulèvent deux types de problèmes :

- comment représenter sur une trame discrète une entité géométrique continue, et quelles sont les propriétés de cette représentation par rapport aux propriétés vérifiées dans le cas continu ?
- à partir d’une entité géométrique discrète, quelles sont les représentations continues qui lui correspondent ?

Si l’on prend l’exemple de la droite, il est possible de se donner des règles (comme on le verra plus loin) qui permettent de déterminer les points de la trame représentant une droite continue. Inversement, étant donné un ensemble de points de la trame, on peut vérifier s’ils correspondent ou non à la discrétisation d’une droite continue suivant ces règles. Si c’est le cas, on trouvera en général qu’il n’y a pas unicité de la droite continue dont l’ensemble de points est la discrétisation, mais que les droites possibles constituent tout un faisceau.

On pourra se reporter à la référence [Chassery et Montanvert, 1991] pour une étude de ces problèmes dans le cas de courbes quelconques. Ici, nous nous restreignons à l’étude des droites et des cercles. Puis nous abordons les pavages de Voronoï et les triangulations de Delaunay, qui permettent de représenter les images de manière structurée, non pas systématique comme avec les pavages réguliers, mais en fonction du contenu des images.

Dans toute la suite, on se placera dans le cas 2D, sur une trame carrée. Les résultats peuvent en général être étendus à d’autres types de trames et à des espaces de dimension supérieure.

4.4.1 Discrétisation d'une droite continue

Une première méthode de discrétisation des droites est la méthode dite du « pavé semi-ouvert ». Le pavé semi-ouvert associé à un point discret P de coordonnées (i, j) est l'ensemble des points de \mathbb{R}^2 dont les coordonnées (x, y) vérifient :

$$i - \frac{1}{2} < x \leq i + \frac{1}{2} \text{ et } j - \frac{1}{2} < y \leq j + \frac{1}{2}.$$

La méthode de discrétisation par pavé semi-ouvert consiste alors à garder dans la représentation discrète tout point P de la trame tel que le pavé semi-ouvert qui lui est associé ait une intersection non vide avec l'entité à discrétiser. La figure 4.19 illustre le résultat dans le cas d'un segment de droite.

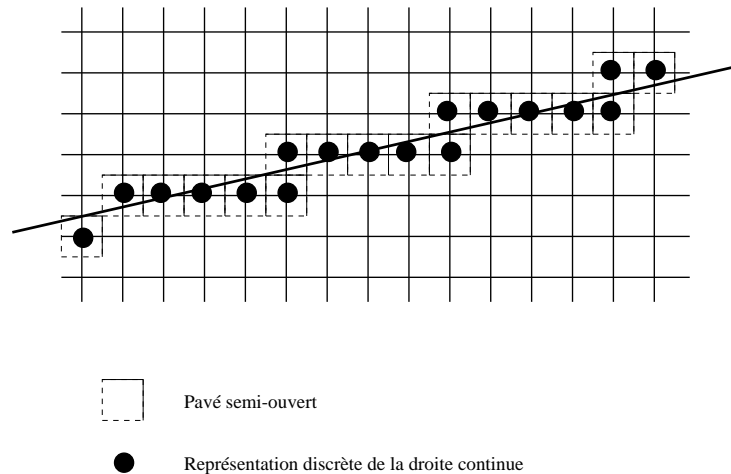


FIG. 4.19 – Discrétisation d'un segment de droite par la méthode du pavé semi-ouvert.

L'ensemble discret ainsi obtenu ne constitue pas nécessairement un chemin simple au sens de la 4- ou de la 8-connexité. Sur l'exemple de la figure 4.19, l'ensemble de points obtenu n'est pas 4-connexe, et au sens de la 8-connexité, certains points ont plus de 2 voisins.

Si l'on cherche maintenant à discrétiser la frontière du demi-plan fermé limité par la droite (plutôt que la droite elle-même), en imposant une contrainte d'unilatéralité, on obtient cette fois un chemin 8-connexe [Dorst et Smeulders, 1984]. La méthode consiste à garder les points discrets s'ils sont situés d'un même côté de la droite et tels que le segment vertical du maillage qui en part coupe la droite. Ce procédé est illustré sur la figure 4.20.

On remarquera que le résultat obtenu est différent de celui obtenu avec la méthode du pavé semi-ouvert, partant de la même droite continue.

Il est possible de combiner ces deux méthodes, en associant à chaque point du maillage un segment semi-ouvert, centré en ce point et vertical. Un point discret est alors conservé si le segment qui en est issu intersecte la droite continue. Le résultat de ce procédé est illustré sur la figure 4.21.

On obtient toujours un chemin 8-connexe, mais où les points discrets sont répartis des deux côtés de la droite continue³.

³Le résultat est équivalent à celui qu'on obtiendrait avec l'algorithme de Bresenham, fréquemment utilisé en synthèse d'images, qui minimise l'erreur locale en chacun des points du tracé [Hégron, 1985].

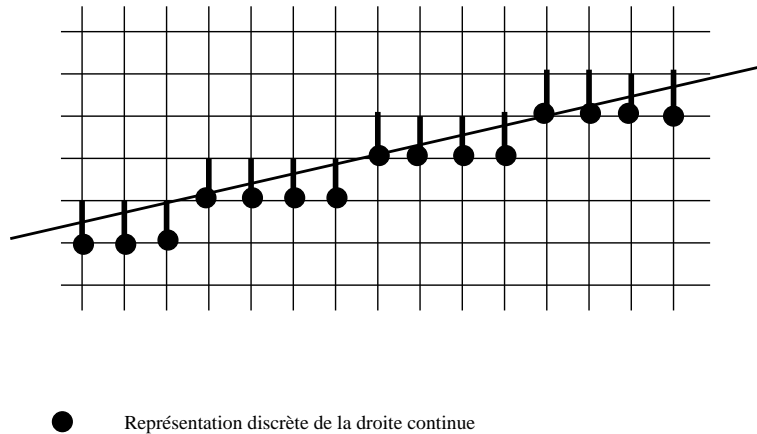


FIG. 4.20 – Discretisation d'un segment de droite par la contrainte d'unilatéralité.

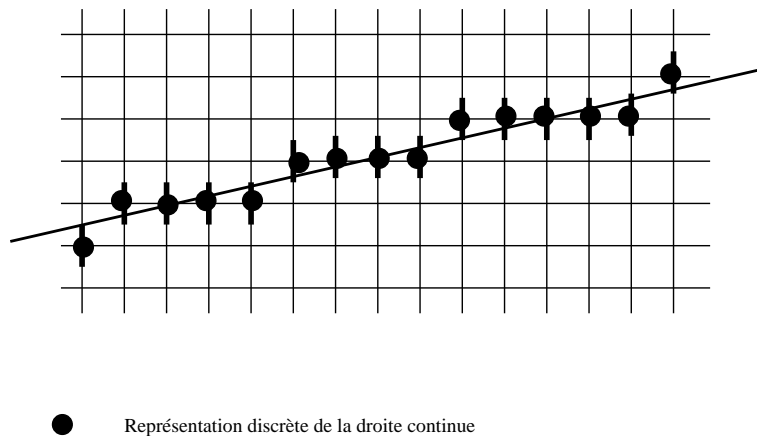


FIG. 4.21 – Discretisation d'un segment de droite par la méthode de segment vertical semi-ouvert.

4.4.2 Caractérisation d'un segment de droite discret

Considérons maintenant le problème inverse : étant donné un ensemble de points discrets, est-il la discrétisation d'un segment de droite continue ? En utilisant les caractérisations précédentes, cela revient à vérifier qu'il existe un segment de droite dont la discrétisation selon les règles choisies donne exactement l'ensemble des points discrets. Cette démarche n'est pas toujours très opérationnelle, et on lui préfère deux autres méthodes, l'une reposant sur la propriété de la corde, et l'autre sur une description syntaxique d'un segment de droite discret.

Soit S un ensemble de points discrets (toujours en dimension deux et sur une trame carrée). On dit que S vérifie la propriété de la corde si et seulement si :

$$\forall (P, Q) \in S, \forall R \in [P, Q], \exists T \in S, d_{\infty}(T, R) < 1$$

où $[P, Q]$ désigne le segment de \mathbb{R}^2 (continu) joignant P à Q , et d_{∞} désigne la distance obtenue à partir de la norme L_{∞} dans \mathbb{R}^2 ($d_{\infty}((x, y), (x', y')) = \max(|x - x'|, |y - y'|)$).

La figure 4.22 présente un cas où la propriété de la corde est vérifiée et un cas où elle ne l'est pas.

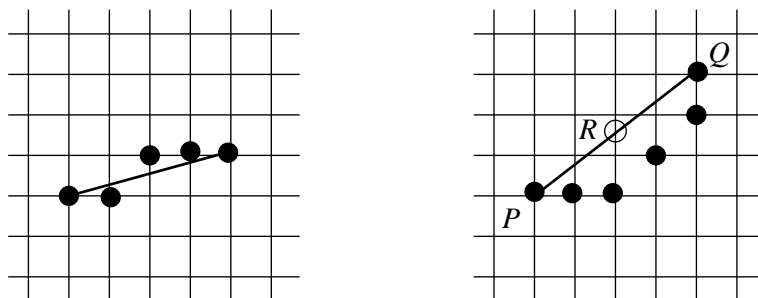


FIG. 4.22 – L'ensemble de points discrets de gauche vérifie la propriété de la corde, celui de droite ne la vérifie pas (le point R du segment $[P, Q]$ est à une distance de plus de 1 de tout point discret de S).

Un segment discret vérifiant la propriété de la corde satisfait aux caractéristiques utilisées dans la méthode de discrétisation du segment semi-ouvert [Rosenfeld, 1974, Ronse, 1985, Santalo, 1940].

Une deuxième caractérisation opérationnelle de segments de droite discrets repose sur leur description syntaxique [Pham, 1986]. En effet, un segment de droite discret est constitué d'une suite de points, que l'on peut suivre en observant les changements de direction. La caractérisation est alors la suivante :

- on appelle section une sous-suite de points maximale sans changement de direction (les 8 directions possibles sont celles définies par la 8-connexité sur la trame carrée, comme l'indique la figure 4.23),
- dans un segment de droite discret, les sections ne peuvent avoir que deux directions distinctes, qui sont consécutives (selon le schéma de la figure 4.23),
- pour une de ces directions, les sections sont toutes de longueur 1, et pour l'autre direction, les sections sont de longueur n ou $n + 1$, où la valeur de n dépend de la pente de la droite.

On peut donc utiliser cette caractérisation, illustrée sur la figure 4.24, pour vérifier si l'ensemble de points considéré est bien un segment de droite discret.

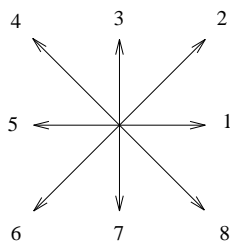


FIG. 4.23 – Les 8 directions possibles sur une trame carrée. Deux directions sont consécutives si leurs numéros diffèrent de 1 (modulo 8).

4.4.3 Droites analytiques discrètes

On considère ici le problème des droites discrètes d'un point de vue différent. Au lieu d'essayer de discrétiser une droite par une suite de points connexes, on cherche au contraire à savoir quels sont les points d'intersection avec la trame d'une droite continue quelconque.

Une droite d'équation $y = ax + b$ sera alors représentée par les points d'intersection avec les points du maillage.

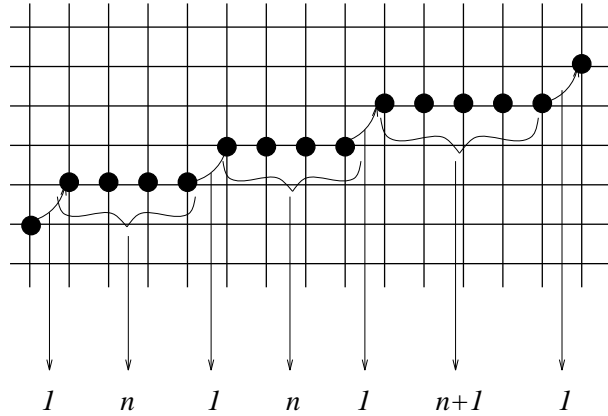


FIG. 4.24 – Caractérisation syntaxique d'un segment de droite discret.

Pour que cette intersection soit non vide, il faut que la pente de la droite soit de la forme :

$$a = \frac{p}{q}$$

avec p et q entiers, premiers entre eux, et vérifiant :

$$p \leq q \leq N,$$

si l'image est de taille $N \times N$, et pour une pente de droite inférieure à 1 (les autres cas s'obtiennent par symétrie).

Les pentes de droites possibles forment donc une suite de Farey d'ordre N , notée $F(N)$ [Frael, 1924]. Pour une image de taille 4×4 , les droites possibles telles que $a \leq 1$ sont représentées sur la figure 4.25 ; on a dans ce cas :

$$F(4) = \{0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}.$$

Le nombre de droites possibles augmente bien sûr avec la taille de l'image. Par exemple pour $N = 6$, on a :

$$F(6) = \{0, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1\}.$$

Le cardinal de $F(N)$ est de l'ordre de $3N^2/\pi^2$. Des propriétés des suites de Farey permettent de les construire. En particulier, si les fractions $\frac{p}{q}, \frac{p'}{q'}, \frac{p''}{q''}$ sont des termes consécutifs de $F(N)$, on a les relations suivantes :

$$p'q - pq' = 1,$$

$$\frac{p'}{q'} = \frac{p + p''}{q + q''}.$$

On peut ainsi calculer le nombre de droites possibles suivant la taille de l'image, mais aussi les nombres d'occurrences de ces droites en fonction de la pente a . En effet, si l'on essaye de retrouver une droite dans une image à partir de deux points discrets, on trouvera plus souvent certaines droites que d'autres. Par exemple, les droites de pente 0 et 1 sur la figure 4.25 seront obtenues pour plus de couples de points que les autres droites.

Pour certaines transformations, ce phénomène constitue un véritable biais lié à la discrétisation dont il va falloir tenir compte. Par exemple, si l'on essaye de détecter des contours rectilignes en comptant les couples de points

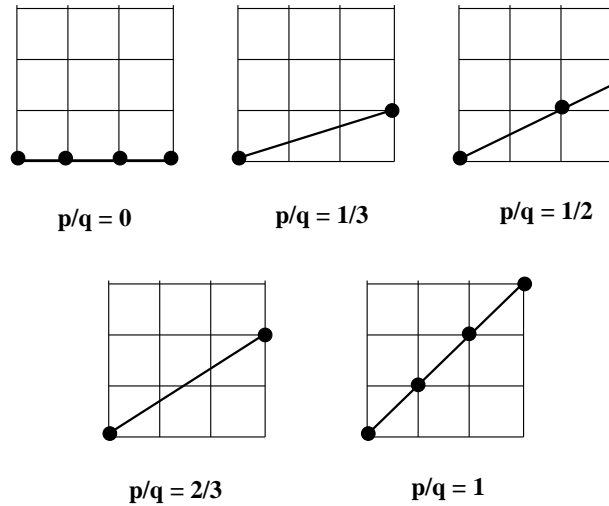


FIG. 4.25 – Les droites possibles (d’intersection non vide avec l’ensemble des points de la trame) sur une image 4×4 .

qui contribuent à un contour d’orientation donnée (transformation de Hough [Maître, 1985a, Maître, 1986], voir chapitre 14), on détectera beaucoup plus facilement les contours de pente 0 ou 1 que des contours de pente très faible par exemple.

Si l’on cherche maintenant quelles sont les longueurs possibles de segments discrets, on arrive à des conclusions de même type. Soit L le carré d’une longueur de segment. L est obtenu comme la distance quadratique entre deux points discrets, donc de coordonnées entières. L est donc solution d’une équation diophantienne de la forme :

$$a^2 + b^2 = L$$

avec a et b entiers. On retrouvera pour les cercles le même type d’équation, qui sera étudié plus en détails à cette occasion.

Cette équation n’a pas toujours de solution. En effet, on constate que, suivant la parité de a et b , $a^2 + b^2$ est congru à 0, 1 ou 2 modulo 4. Les valeurs de L qui sont congrues à 3 modulo 4 ne peuvent donc jamais être obtenues. La figure 4.26 illustre l’irrégularité des nombres d’occurrences de L sur une image 16×16 et sur une image 50×50 . La décroissance générale des courbes obtenues est liée au fait que les images sont bornées.

Pour plus de détails sur les droites et courbes discrètes, on pourra consulter [Debled-Rennesson et Reveillès, 1995].

4.4.4 Cercles discrets

De la même manière que pour les droites, le nombre de points discrets qui sont situés sur un cercle continu peut varier beaucoup, et de manière irrégulière suivant le rayon. Pour certaines valeurs du rayon, il n’y a pas de solution. Ici encore, l’équation à résoudre est :

$$a^2 + b^2 = n$$

où a , b et n sont entiers, n représente le carré du rayon et $a^2 + b^2$ la distance quadratique d’un point discret au centre du cercle (on suppose que le centre coïncide avec un des points de la trame). On trouve ainsi 4 points sur un cercle de rayon 1, 4 points pour $n = 2$, aucun point pour $n = 3$, etc., comme l’illustre la figure 4.27.

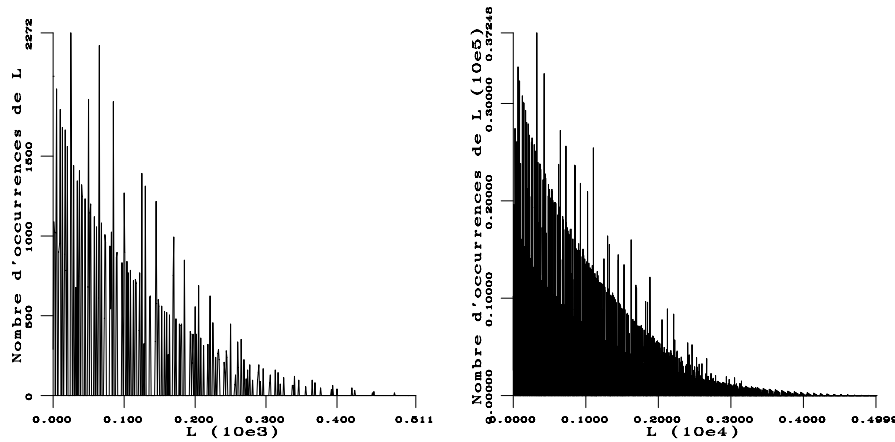


FIG. 4.26 – Distribution des carrés des longueurs des segments discrets sur une image 16 × 16 et sur une image 50 × 50.

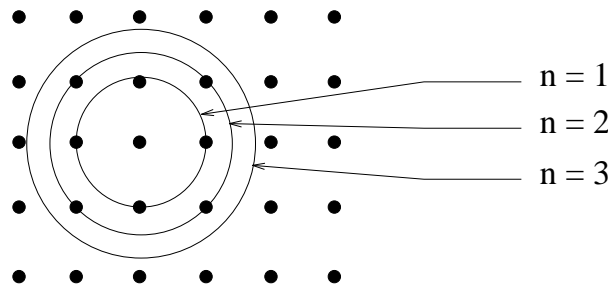


FIG. 4.27 – Intersection de cercles continus avec des points de la trame carrée.

Le nombre de solutions de cette équation est donné par $r(n)$, calculé à partir de la décomposition de n en nombres premiers :

$$r(n) = 4 \prod_t (\tau_p + 1) \prod_q \left(\frac{1 + (-1)^{\sigma_q}}{2} \right).$$

où n est décomposé en facteurs premiers sous la forme :

$$n = 2^\alpha \prod_p p^{\tau_p} \prod_q q^{\sigma_q}$$

où les p représentent les facteurs premiers congrus à 1 modulo 4, et les q ceux qui sont congrus à 3 modulo 4.

Par exemple, pour $n = 5$ on obtient 8 points, pour $n = 25$ on obtient 12 points. En effet, $25 = 5^2$ et 5 est congru à 1 modulo 4. Le seul terme qui intervient dans $r(25)$ correspond donc à $p = 5$, avec $\tau_p = 2$, et $r(25) = 4 \times (2 + 1) = 12$. On remarque dans ces formules que, dès qu'intervient dans la décomposition de n un terme congru à 3 modulo 4 avec un exposant impair, alors $r(n) = 0$.

En pratique, la définition d'un cercle discret limitée aux points d'intersection avec le maillage est très restrictive. On préfère donc souvent considérer que le cercle a une certaine « épaisseur », et donc regarder plutôt l'intersection d'une couronne circulaire avec les points du maillage⁴.

⁴On peut bien sûr faire la même chose avec des droites et des segments, en considérant des bandes.

Si k est l'épaisseur de la couronne, on cherche alors le nombre de points satisfaisant l'équation :

$$n \leq a^2 + b^2 < n + k$$

toujours avec a et b entiers. Le nombre de solutions est simplement donné par :

$$r(n, k) = \sum_{i=n}^{i=n+k-1} r(i).$$

Par exemple, pour $n = 54$ et $k = 4$, on trouve $r(n, k) = 0$. Ici encore, la grande variabilité de la fonction $r(n, k)$ entraîne des biais dans les méthodes d'analyse ou de synthèse d'images, ainsi que de reconnaissance des formes, qu'il ne faut pas négliger. Il existe des méthodes de correction, prenant en compte la forme des distributions obtenues.

Pour plus de détails sur les cercles et sphères discrets, on pourra consulter [Andres, 1994].

4.4.5 Pavage de Voronoï et triangulation de Delaunay

Nous nous attachons maintenant à des représentations structurées adaptées au contenu des images. Nous présentons ici les pavages de Voronoï, qui constituent un exemple de système de représentation structurée pour l'interprétation de formes. Nous en verrons un autre exemple très différent avec le squelette dans le chapitre 6 sur la morphologie mathématique.

On se donne un ensemble de points $\{P_1, P_2, \dots, P_n\}$, appelés germes. À chacun de ces germes, on associe un domaine du plan \mathbb{R}^2 défini par :

$$V(P_i) = \{P \in \mathbb{R}^2 / \forall j, 1 \leq j \leq n, d(P, P_i) \leq d(P, P_j)\},$$

où d désigne une distance du plan.

Si l'on choisit pour d la distance euclidienne, alors les $V(P_i)$ sont des polygones convexes. En effet, si l'on prend tout d'abord $n = 2$, alors la séparation entre $V(P_1)$ et $V(P_2)$ est la médiatrice des deux points. Les $V(P_i)$ pour n quelconque sont donc des intersections de demi-plans, donc des polygones convexes (pas forcément bornés).

La figure 4.28 illustre cette définition (pour la distance euclidienne), ainsi que la convexité des pavés obtenus.

Notons que le pavage obtenu correspond à la première des méthodes décrites pour construire des pavages dans la section 4.2. Les pavés sont appelés polygones de Voronoï, et sont formés d'arêtes et de sommets de Voronoï.

La définition du pavage de Voronoï se généralise directement à des germes quelconques (plus nécessairement ponctuels) mais, dans ce cas, les pavés obtenus ne sont pas nécessairement polygonaux.

Dans le cas discret, la définition du pavage de Voronoï demeure inchangée. Cependant, au lieu de la distance euclidienne, on en utilise généralement des approximations : les distances discrètes, telles qu'elles seront vues dans la section 4.6, qui permettent des calculs rapides tout en fournissant une bonne approximation de la distance euclidienne.

Le pavage de Voronoï a les propriétés suivantes [Klein, 1989] :

1. s'il n'existe pas de quadruplets de germes cocirculaires, tout sommet de Voronoï est équidistant de 3 germes exactement,
2. tout sommet du diagramme de Voronoï est centre d'un cercle passant par 3 germes et ne contenant aucun autre germe, ce cercle est appelé cercle de Delaunay,
3. $V(P_i)$ est non borné si et seulement si P_i appartient à la frontière de l'enveloppe convexe des P_j .

Si l'on joint par des segments les 3 germes équidistants d'un sommet de Voronoï, on obtient une triangulation de l'ensemble de P_i , appelée triangulation de Delaunay. Le résultat obtenu pour les points de la figure 4.28 est illustré sur la figure 4.29.

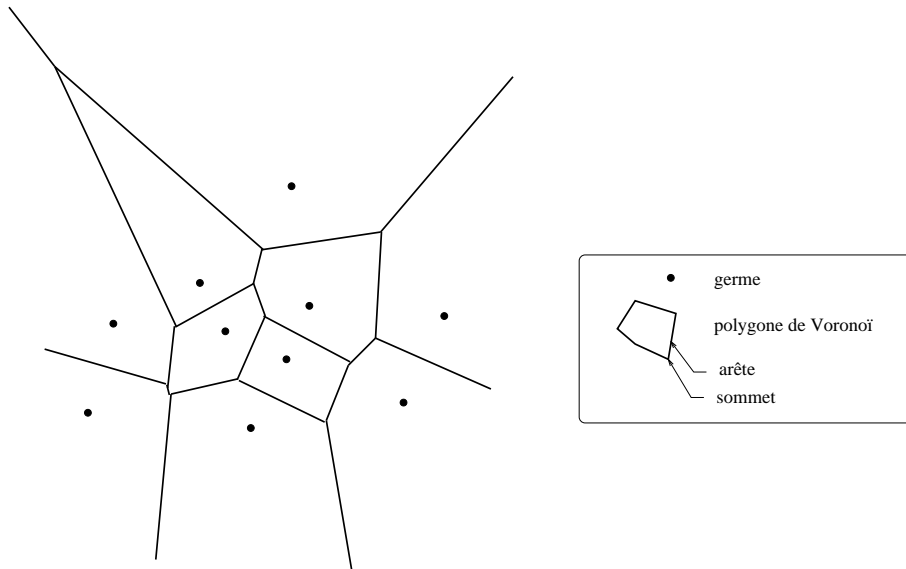


FIG. 4.28 – Pavage de Voronoï à partir d’un ensemble de points.

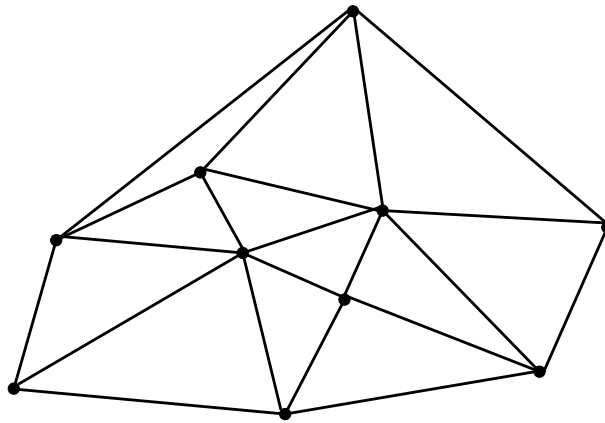


FIG. 4.29 – Triangulation de Delaunay d’un ensemble de points.

La triangulation de Delaunay et le pavage de Voronoï sont deux structures de l’espace duales l’une de l’autre, comme l’illustre la figure 4.30. En effet, à chaque arête de Voronoï correspond une arête de la triangulation, qui lui est orthogonale ; chaque germe est un sommet de la triangulation, et chaque sommet de Voronoï est le centre d’un triangle de Delaunay.

La deuxième propriété est illustrée sur la figure 4.31. La triangulation de Delaunay est celle qui vérifie que le cercle circonscrit à chaque triangle ne contienne aucun sommet de la triangulation. On a donc ici une première application géométrique du pavage de Voronoï et de la triangulation de Delaunay.

La troisième propriété permet de déduire directement l’enveloppe convexe d’un ensemble de points, ce qui nous donne une deuxième application géométrique de ces deux notions.

Il existe beaucoup d’autres applications. Nous n’en mentionnerons plus qu’une ici, qui concerne le calcul de la

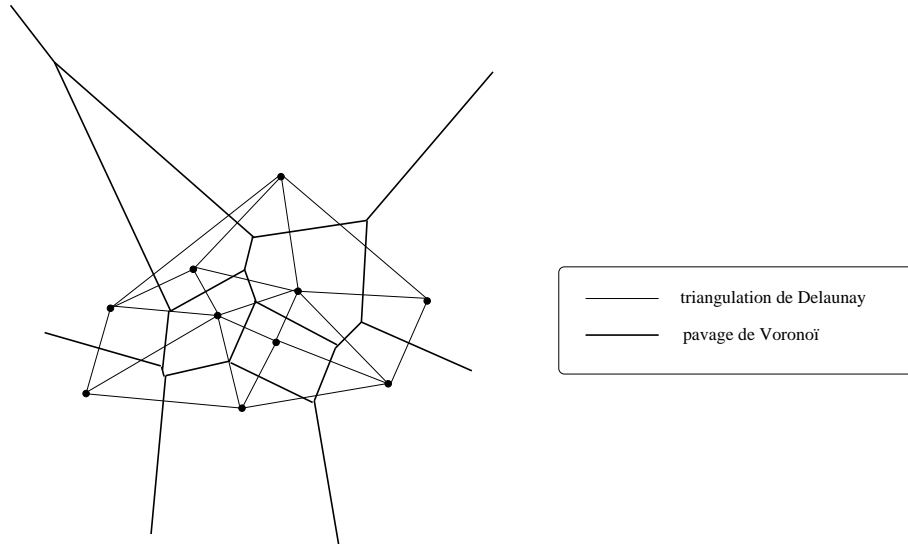


FIG. 4.30 – Dualité entre triangulation de Delaunay et pavage de Voronoï.

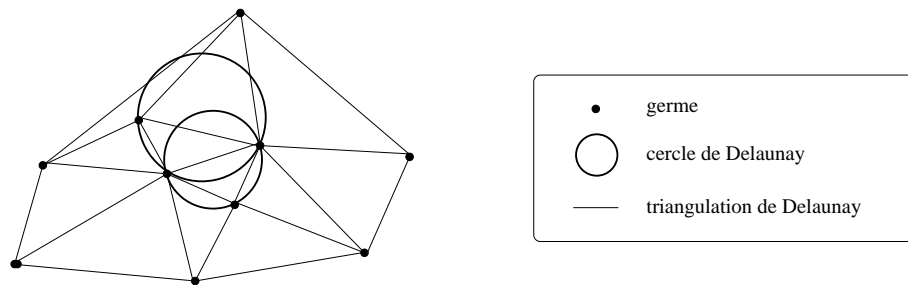


FIG. 4.31 – Propriétés géométriques de la triangulation de Delaunay.

distance minimum entre deux ensembles de points A et B . Cette distance est définie par :

$$d_{\min}(A, B) = \min_{a \in A, b \in B} d(a, b),$$

où $d(a, b)$ désigne la distance euclidienne entre les points a et b . Pour calculer cette distance minimum, il suffit d'effectuer la triangulation de Delaunay de la réunion de tous les points $A \cup B$, et de déterminer l'arête la plus courte ayant une extrémité dans A et l'autre dans B .

Enfin, mentionnons quelques algorithmes de calcul du pavage de Voronoï [Toussaint, 1985, Preparata et Shamos, 1988]. Il existe des algorithmes récursifs qui sont optimaux (en termes de complexité algorithmique), mais dont la mise en œuvre n'est pas très facile. Il existe également des algorithmes incrémentaux, dont la complexité dans le cas le pire n'est pas optimale, mais qui sont plus simples à mettre en œuvre [Green et Sibson, 1978]. Le principe consiste à rajouter les points un à un et à modifier le pavage précédemment obtenu. L'intérêt est que l'ajout d'un point n'entraîne que des modifications locales du pavage. Supposons que l'on connaisse le pavage de $\{P_1, \dots, P_{n-1}\}$. Lorsqu'on ajoute P_n , on commence par rechercher le germe P_i ($1 \leq i \leq n-1$) le plus proche de P_n . La médiatrice de $[P_i, P_n]$ est le support d'une nouvelle arête du pavage. En se donnant un sens de parcours de cette médiatrice, on recherche l'intersection avec une arête du pavage précédent. Le germe correspondant, voisin de P_i , est le nouveau point P_i et on itère la procédure jusqu'à ce qu'on revienne au point de départ. On « tourne » ainsi

autour de P_n pour construire le pavé qui lui est associé. Les parties d'arêtes du pavage précédent qui tombent dans ce nouveau pavé sont alors supprimées. La figure 4.32 illustre le principe de cette construction.

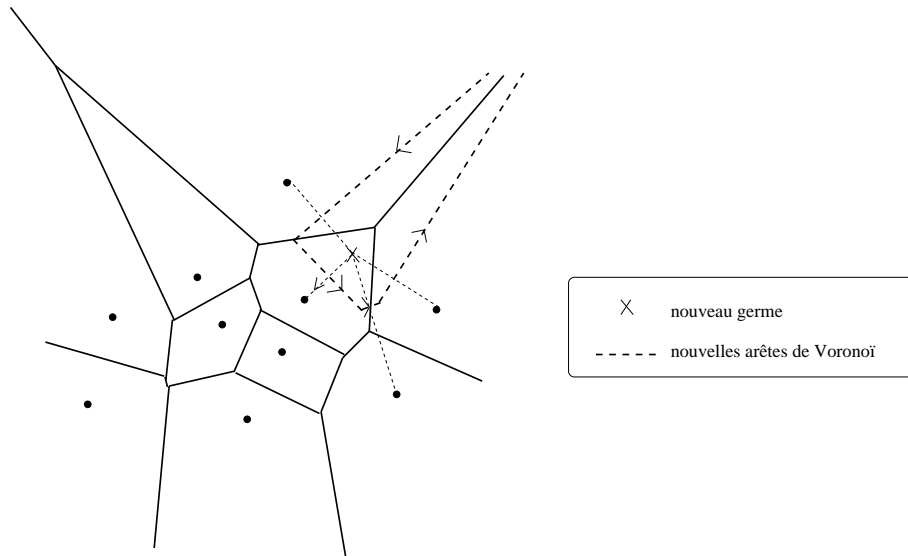


FIG. 4.32 – Construction incrémentale du pavage de Voronoï.

4.5 Exemples de structures aléatoires

Dans nombre d'applications, en particulier pour l'analyse de textures, qu'elles soient binaires ou à niveaux de gris, il est nécessaire de générer des germes aléatoires dans l'image.

Une des structures aléatoires ponctuelles les plus utilisées repose sur des distributions de Poisson⁵. Une distribution de points de Poisson vérifie les propriétés suivantes [Serra, 1982a] :

- si X est un compact, et $N(X)$ la variable aléatoire qui compte les points de Poisson tombés dans X , alors $N(X)$ est une variable aléatoire d'espérance $\lambda V(X)$, où λ est la densité du processus de Poisson et $V(X)$ le volume du compact (sa surface dans le cas du plan) ;
- si X et X' sont deux compacts tels que $X \cap X' = \emptyset$, alors $N(X)$ et $N(X')$ sont deux variables aléatoires indépendantes ;
- la variable aléatoire $N(X)$, pour X compact vérifie :

$$P(N(X) = n) = \frac{\Phi^n}{n!} \exp(-\Phi)$$

avec : $\Phi = \int_X \lambda dv$, (dans le cas d'une densité constante, $\Phi = \lambda V(X)$) ;

- étant donné n points x_i dans X , x_i suit une loi uniforme sur X .

Si l'on essaye de généraliser la notion de points de Poisson à des structures plus complexes, on est confronté à des difficultés de génération de ces structures. Par exemple, la génération de droites de Poisson n'est pas immédiate. L'idée la plus simple serait de se donner une distribution de points de Poisson dans le plan, et en chaque point, de tracer une droite dont la direction est choisie de manière uniforme. Malheureusement la distribution de droites ainsi obtenue n'est pas de Poisson. Une autre solution, plus compliquée mais fournissant effectivement une distribution de Poisson, consiste à faire tourner un axe Oz autour d'un point O . Quand Oz balaye un secteur angulaire ($\alpha, \alpha +$

⁵Elles constituent par exemple une des composantes des schémas booléens que nous aborderons au chapitre 6.

$d\alpha$), on tire sur la droite des points de Poisson de densité $\lambda d\alpha$, et on trace les droites orthogonales à Oz passant par ces points. On obtient ainsi un ensemble de droites de Poisson. Elles peuvent être utilisées par exemple pour définir des pavages aléatoires isotropes.

4.6 Fonction distance

La fonction distance est une transformation qui permet de passer d'une image binaire à une image numérique où la valeur de chaque point représente la distance à l'objet le plus proche. Bien que le concept soit global, la fonction distance peut être calculée de manière locale en propageant des distances locales (entre points voisins) définies dans un masque. Ce type d'algorithme évite de calculer de manière exhaustive les distances d'un point à tous les points de l'objet pour en trouver le minimum, ce qui serait prohibitivement long. La contre-partie est qu'on ne calcule alors qu'une approximation de la distance euclidienne. Les exigences sont alors d'obtenir, par des algorithmes rapides, une bonne approximation de la distance euclidienne. Ces méthodes, appelées aussi distances du chanfrein [Borgefors, 1996], sont un exemple où un procédé algorithmique permet de lever les difficultés liées à la discrétisation de l'espace.

Nous verrons dans le chapitre 6 que la fonction distance a de nombreuses applications en morphologie mathématique (outre toutes les applications nécessitant une estimation de la distance entre objets comme le recalage).

4.6.1 Définition de distances discrètes

Les distances discrètes sont définies comme des distances sur un graphe, dans lequel les nœuds sont les pixels ou voxels, et les arcs sont définis par un ensemble de vecteurs de base. Ils peuvent représenter les relations de connexité élémentaires (par exemple 4 ou 8 connexité), mais également relier des points plus éloignés. Ils représentent les directions selon lesquelles on peut se déplacer pour aller d'un point à un autre.

Soit $\mathcal{P} = \{\vec{p}_1, \dots, \vec{p}_m\}$ l'ensemble de ces vecteurs de base, qui engendrent le graphe, et auxquels on associe des longueurs d_i . On impose les conditions suivantes :

- $\vec{p}_i \in \mathcal{P} \Rightarrow -\vec{p}_i \in \mathcal{P}$,
- $\vec{p}_i \in \mathcal{P}, \lambda \vec{p}_i \in \mathcal{P} \Rightarrow \lambda = \pm 1$,
- $\|\vec{p}_i\| = \|\vec{p}_j\| \Rightarrow d_i = d_j$.

On définit alors la distance entre deux nœuds x et y (deux points) comme :

$$d(x, y) = \frac{1}{s} \min \left\{ \sum_{i=1}^m n_i d_i \mid n_i \in \mathbb{N}, \sum_{i=1}^m n_i \vec{p}_i = \vec{xy} \right\}$$

où s est un facteur d'échelle (typiquement $s = d_1$ si d_1 est la longueur associée au vecteur élémentaire selon un des axes de coordonnées). Cette distance est exactement la longueur du plus court chemin entre x et y sur le graphe engendré par \mathcal{P} . On peut vérifier aisément que toutes les propriétés d'une distance sont bien satisfaites.

En pratique, les \vec{p}_i et d_i sont représentés par des masques. Chaque point du masque est l'extrémité d'un \vec{p}_i et a pour coefficient le d_i correspondant, qui représente la distance locale entre ce point et le point central du masque (celui-ci est affecté du coefficient 0).

4.6.2 Exemples

Pour la qualité de l'approximation, il est possible de jouer sur deux paramètres : la taille du masque (donc le nombre de directions représentées, ou encore le nombre de \vec{p}_i) et les coefficients du masque (donc les d_i). Plusieurs critères ont été proposés pour estimer cette qualité, le plus fréquemment utilisé étant de minimiser le maximum de la différence avec la distance euclidienne. Les masques les plus employés sont illustrés sur la figure 4.33. Les deux

premiers correspondent respectivement à la 4- et à la 8-connexité sur une trame carrée. Le troisième (figure 4.33 c) permet d'avoir une erreur faible ($\leq 8\%$) par rapport à la distance euclidienne en modifiant les coefficients (toutes les distances sont alors multipliées par 3). Le dernier (figure 4.33 d) donne une erreur inférieure à 2% en jouant à la fois sur les coefficients et sur la taille du masque (toutes les distances sont alors multipliées par 5).

							11	11					
	1		1	1	1	4	3	4	11	7	5	7	11
1	0	1		1	0	1	3	0	3	5	0	5	
	1		1	1	1	4	3	4	11	7	5	7	11
									11		11		
a			b			c							d

FIG. 4.33 – Exemples de masques pour la distance du chanfrein.

4.6.3 Algorithmes de calcul

Une première famille d'algorithmes opère sur une représentation classique de l'image, sous forme de tableau. Les algorithmes de cette famille peuvent être de deux types : parallèle ou séquentiel. Nous les illustrons sur le problème du calcul de la fonction distance (distance de chaque point de l'image au point des objets le plus proche, au sens de la distance discrète définie plus haut).

Si f^k représente l'image à l'itération k et g le masque choisi, l'algorithme parallèle s'exprime de la manière suivante : les points de l'objet sont mis à 0 et les points du complémentaire à une valeur infinie (en pratique, la plus grande valeur stockable), pour former l'image f^0 ; puis la formule suivante est itérée jusqu'à convergence :

$$f^k(x) = \min\{f^{k-1}(y - x) + g(y), y \in \text{support}(g)\}.$$

L'inconvénient majeur de cet algorithme est que le nombre d'itérations dépend de la taille de l'image, de la taille de l'objet et de sa forme. De plus deux images doivent être gardées en mémoire. Cet algorithme s'applique à n'importe quelle trame et n'importe quel masque.

L'algorithme séquentiel ne nécessite que deux itérations. Il procède en deux balayages en sens opposés de l'image et le masque est divisé en deux parties g_1 et g_2 contenant les points déjà examinés dans le sens courant de balayage. Pour $k = 1, 2$, on effectue l'opération suivante (f^0 étant calculée comme précédemment) :

$$f^k(x) = \min\{f^{k-1}(x), f^k(y - x) + g_k(y), y \in \text{support}(g_k)\}.$$

Cet algorithme vérifie l'exigence de la rapidité. Il ne nécessite qu'une image en mémoire et s'applique également à n'importe quelle trame et n'importe quel masque.

Une deuxième famille d'algorithmes repose sur les contours des objets. En effet, les points situés à l'intérieur des objets ont toujours une distance nulle à l'objet, et les points les plus proches des points extérieurs aux objets sont toujours des points situés sur les contours des objets. Ainsi, seuls ceux-ci interviennent réellement dans le calcul.

Un premier type d'algorithme dans cette famille repose sur une structure de chaîne pour représenter les contours des objets. Le principe du calcul de la fonction distance consiste à déplacer ces contours par des règles de réécriture, pour déterminer successivement les points à une distance 1 des objets, puis les points à une distance

2, etc. [Vincent, 1992]. Une étape d'ajustement est parfois nécessaire, par exemple quand le déplacement d'un contour conduit à connecter deux objets. Ce type d'algorithme ne nécessite qu'une image en mémoire. L'accès aux voisins doit être facile car il est fréquemment utilisé dans les règles de ré-écriture. Il est très rapide, mais ne peut pas être généralisé à des images 3D car il n'y a pas d'ordre naturel sur les points de la surface d'un objet 3D permettant de les représenter sous forme de chaîne. Les règles de ré-écriture et d'ajustement comportent beaucoup de cas particuliers et la procédure complète est donc un peu délicate.

Un deuxième type d'algorithme repose sur une représentation des points de contours sous la forme d'une simple file d'attente FIFO (*first in, first out*). Cette file est initialisée par les points de contours des objets (l'ordre n'a pas d'importance). L'algorithme extrait alors le premier point p de la file, et recherche ses voisins. Pour chaque voisin q , si celui-ci est à l'extérieur des objets, il est ajouté à la file, et on lui affecte une valeur de distance qui vaut celle de p plus la distance locale entre p et q (d_i si $\vec{pq} = \vec{p}_i$). Cet algorithme est également très rapide, généralisable à 3D, et nécessite également un accès rapide aux voisins.

Notons que ces différentes classes d'algorithmes sont utilisées pour de nombreuses autres transformations.

Chapitre 5

Restauration des images

Chapitre rédigé par Henri MAÎTRE

La restauration des images est l'opération qui corrige des images dégradées et reconstruit un signal de bonne qualité à partir d'un signal de médiocre qualité. On distingue parfois la restauration d'image de l'amélioration d'images en ce que la seconde s'attache à corriger des défauts généralement inconnus par des techniques heuristiques tandis que la seconde se consacre à corriger des défauts connus par inversion du défaut. Néanmoins, les progrès récents de la restauration permettent aujourd'hui d'envisager des " restaurations aveugles " dans lesquelles le défaut est inconnu. Mais la restauration aveugle partage avec la restauration l'objectif d'inverser le défaut, défaut qui est donc appris au cours du processus d'inversion. La préoccupation de connaître le défaut n'existe pas en amélioration et les traitements qui y sont proposés sont souvent moins formalisés. Ils sont présentés au chapitre 10. L'équation fondamentale de la restauration relie une image originale f à une image dégradée g par l'intermédiaire d'un opérateur \mathcal{D} :

$$g = \mathcal{D}(f) \quad (5.1)$$

Une équation aussi générale ne possède pas de solution universelle et l'on est amené à réduire le problème au seul cas où :

- \mathcal{D} est un opérateur linéaire ;
- \mathcal{D} est un opérateur spatialement invariant.

Dans ces conditions, on montre que l'équation 5.1 prend alors la forme d'une convolution :

$$g = d * f \quad (5.2)$$

où d est la réponse impulsionnelle du défaut. C'est cette seule opération que nous traiterons dans ce chapitre. Pour cette raison, la restauration sera réduite pour nous à une opération de déconvolution ou de filtrage. De plus nous ne sommes concernés ici que par le cas de défauts parfaitement connus, c'est-à-dire pour lesquels nous disposons de la réponse impulsionnelle d .

Nous allons, dans ce chapitre, examiner tout d'abord l'inversion de l'équation précédente. Nous étudierons ensuite les conditions pour que la restauration d'un signal se passe bien. Enfin nous proposerons des techniques pour évaluer le défaut entachant l'image, défaut que nous aurons jusqu'alors supposé connu.

5.1 Le filtrage inverse

Nous considérerons tout d'abord la restauration des signaux monodimensionnels et repousserons l'aspect bi- et tri-dimensionnel à la section 5.3. Nous allons examiner deux façons différentes d'aborder ce problème : d'une part de façon analytique, d'autre part de façon algébrique et nous soulignerons les correspondances entre elles.

5.1.1 Approche analytique

L'équation 5.2 se réécrit de façon développée sous la forme :

$$g(x) = \int_{-\infty}^{+\infty} f(t)d(x-t)dt \quad (5.3)$$

Par TF, cette équation nous donne :

$$G(u) = F(u)D(u)$$

qui, si toutes les conditions sont remplies, (nous y reviendrons plus loin), permet d'obtenir une estimée $\hat{F}(u)$ de la TF de f par :

$$\hat{F}(u) = G(u)/D(u) = G(u)W_I(u)$$

puis, par TF inverse, une estimée \hat{f} de f :

$$\hat{f}(x) = g(x) * w_I(x) \quad (5.4)$$

Nous voyons que, si toutes les conditions sont remplies (existence du filtre W_I , existence de la TF inverse), le signal restauré $\hat{f}(x)$ et le signal idéal f sont égaux. Il y a restauration parfaite. Cette restauration est obtenue par filtrage linéaire de l'image défectueuse. Logiquement, le filtre w_I est appelé filtre inverse, sa fonction de transfert W_I est en effet l'inverse de la fonction de transfert du défaut D . Nous reviendrons plus loin sur les problèmes associés aux conditions d'existence du filtre inverse :

$$W_I(u) = \frac{1}{D(u)} \quad (5.5)$$

5.1.2 Approche algébrique

Dans cette approche, nous supposons les signaux échantillonnés conformément aux recommandations établies dans le chapitre 3 et représentés par des vecteurs¹. Nous reviendrons également plus tard sur les conséquences de cet échantillonnage. L'équation 5.3 peut se réécrire de façon discrète sous la forme :

$$g_j = \sum_{j'} d_{j-j'} f_{j'} \quad (5.6)$$

soit, en définissant par Δ la matrice de terme général $\delta_{jk} = d_{j-k}$:

$$\Delta = \{\delta_{jk} = d_{j-k}\} \quad (5.7)$$

sous une forme vectorielle :

$$\mathbf{g} = \Delta \mathbf{f} \quad (5.8)$$

Cette équation fournit une solution immédiate au problème de restauration, par inversion d'un système linéaire (nous supposons là-aussi que les conditions de cette inversion existent) :

$$\hat{\mathbf{f}} = \Delta^{-1} \mathbf{g} \quad (5.9)$$

¹Comme aux chapitres précédents, nous représenterons les vecteurs par des caractères gras lorsqu'il y aura un risque d'ambiguïté

5.1.3 Discussion

On voit ainsi que selon que l'on choisit une approche discrète (algébrique), ou continue (analytique), on est conduit à des solutions en apparence différentes : l'équation 5.4 ou l'équation 5.9. Notons tout d'abord que la nature continue du signal n'est pas la cause des outils différents employés puisque, à partir de l'équation 5.6, il est également possible d'appliquer une TF aux deux membres de l'équation pour aboutir à une solution, discrète, mais mettant en œuvre des transformées de Fourier et non des inverses de matrices. Expliquons en quoi ces deux approches sont bien les deux faces d'un même problème [Andrews et Hunt, 1977].

L'équation 5.7 exprime la propriété d'invariance par translation du défaut ; la matrice Δ est une matrice de Töplitz (nous la noterons pour cela Δ_T) : elle s'écrit sous la forme remarquable :

$$\Delta_T = \begin{bmatrix} d_0 & d_1 & d_2 & d_3 & \dots & 0 \\ d_{-1} & d_0 & d_1 & d_2 & d_3 & \dots \\ \vdots & d_{-1} & d_0 & \ddots & \ddots & \\ \vdots & & & \ddots & & \\ & & & & \ddots & \vdots \\ 0 & & & \dots & d_{-1} & d_0 \end{bmatrix} \quad (5.10)$$

La forme des matrices Töplitz rappelle celle des matrices circulantes. Elles n'en diffèrent que par les termes repliés aux coins supérieur droit et inférieur gauche. Une matrice circulante construite sur d_i , au lieu d'utiliser la définition 5.7, a pour définition :

$$\Delta_c = \{d_{(j-k) \bmod N}\}$$

où N est la dimension de la matrice, ce qui conduit à :

$$\Delta_c = \begin{bmatrix} d_0 & d_1 & d_2 & \dots & \dots & d_{-2} & d_{-1} \\ d_{-1} & d_0 & d_1 & d_2 & \dots & \dots & d_{-1} \\ \vdots & d_{-1} & d_0 & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & & & \\ & & & & \ddots & & \vdots \\ d_2 & & & & & & d_1 \\ d_1 & d_2 & & \dots & d_{-1} & & d_0 \end{bmatrix} \quad (5.11)$$

Il est connu que les matrices circulantes ont toutes pour vecteurs propres les racines de l'unité [Lelong-Ferrand et Arnaudès, 1971], ce qui permet de les écrire : $\Delta_c = P^t \Lambda P$, avec $P = [\mathbf{p}_k]$ et :

$$\mathbf{p}_k = \{e^{-2i\pi k/N} \quad e^{-2i\pi 2k/N} \quad \dots \quad e^{-2i\pi jk/N} \quad \dots\}$$

Leurs valeurs propres valent :

$$\lambda_k = \sum_j d_{jk} e^{-2i\pi jk/N}$$

On montre de plus que les matrices Δ_T et Δ_c sont asymptotiquement équivalentes, sous les conditions suivantes, que nous discuterons plus loin :

- que la dimension du système soit très grande ($N \rightarrow \infty$),
- que le défaut soit à support étroit.

Remplaçons donc Δ (qui était égal à Δ_T) par Δ_c dans l'équation 5.9, nous obtenons :

$$\hat{\mathbf{f}} = \Delta_c^{-1} \mathbf{g} = [P^t \Lambda P]^{-1} \mathbf{g}$$

que nous pouvons récrire :

$$\hat{\mathbf{f}} = P^t \Lambda^{-1} P \mathbf{g}$$

Nous reconnaissons maintenant, en les regroupant à partir de la droite, les divers termes qui constituent l'approche analytique :

- le produit $P\mathbf{g}$, qui représente la TF discrète de \mathbf{g} , dénotée G ;
- le produit par les termes diagonaux $1/\lambda_k$ qui constituent les termes courants de la matrice inverse. Or $1/\lambda_k = 1/\sum_j d_{jk} e^{-2i\pi jk/N}$, c'est donc l'inverse de la fonction de transfert du défaut ;
- puis la transformée de Fourier inverse, accomplie par le produit matriciel par P^t .

On voit donc qu'il y a bien équivalence des deux approches dans le cas où le défaut est correctement représenté par une matrice circulante : l'approche par inversion de matrice est l'inversion directe du défaut, tandis que l'approche par TF est l'inversion par projection du signal sur la base des vecteurs propres de l'opérateur \mathbf{d} .

Mais nous avons vu également que lorsque l'on représente le signal par un ensemble fini d'échantillons, la matrice qui découle des hypothèses de linéarité et d'invariance spatiale n'est pas circulante mais Töplitz. À quelles conditions y a-t-il donc équivalence entre Δ_T et Δ_c ? Nous avons vu que la différence entre les deux matrices réside dans les termes des coins supérieur droit et inférieur gauche qui apparaissent lors du passage de Töplitz à circulante. Ce sont les termes de la demi-réponse impulsionnelle gauche qui ont disparu sur la gauche du signal, et de même à droite. Ces différences seront négligeables si la taille du défaut est très petite devant la dimension du signal. Pourquoi l'approche continue (équation 5.3) conduit-elle à une solution par TF quelle que soit la taille du défaut ? Parce que dans cette approche, la TF est faite de $-\infty$ à $+\infty$, éliminant de fait les effets de bord. Si l'on souhaite discrétiser les divers signaux, sous l'hypothèse qu'ils vérifient la contrainte de spectre borné, il nous faudra également prendre un nombre infini d'échantillons (cf. chapitre 3) et l'équivalence $\Delta_T \equiv \Delta_c$ découle alors de la convergence asymptotique quelle que soit la taille, finie, du défaut².

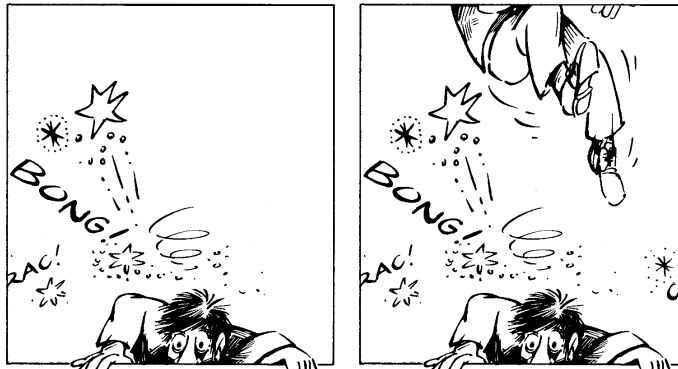


FIG. 5.1 – Dans une matrice Töplitz (à gauche), le signal est tronqué aux bords. Dans une matrice circulante (à droite), les effets de bord sont repliés : on considère que le signal est périodique au détriment d'une bonne lisibilité du signal (merci Fred).

²Dans un cas tout à fait exceptionnel, si le signal est périodique, les 2 approches sont équivalentes avec un nombre fini d'échantillons. En effet sa TF peut être obtenue par une série de Fourier sur une période.

5.1.4 Conclusions (provisoires)

Nous disposons donc, en fait, de 3 approches différentes pour restaurer le signal si celui-ci est représenté sous forme échantillonnée :

- l’inversion la plus ordinaire de la matrice Δ , sans tenir compte de ses propriétés spécifiques, (méthode du pivot de Gauss, ou méthode de Gauss Seidel par exemple) [Durant, 1961, Porte et Vignes, 1974, Golub et van Loan, 1996]. Elle permet d’inverser tous les défauts linéaires, même s’ils ne sont pas spatialement invariants (ce qui dépasse les intentions affichées en début de ce chapitre mais peut être très précieux dans de nombreuses applications de traitement d’images). Sa complexité est en $\mathcal{O}(N^3)$.
- l’inversion de la matrice Töplitz Δ_T , qui ne permet de corriger que les défauts spatialement invariants, mais en tenant compte de la dimension finie des signaux et des effets de bord. Sa complexité est en $\mathcal{O}(N^2)$ (algorithmes de Trench ou de Zohar, exploitant la répétition des termes le long des diagonales pour résoudre des sous-systèmes d’ordre croissant) [Zohar, 1974].
- l’inversion de la matrice circulante Δ_c , qui négligera les effets de bord mais permettra des résolutions rapides (la TF Rapide de Cooley et Tuckey a une complexité en $\mathcal{O}(N \log N)$, en utilisant les propriétés de symétrie des racines de l’unité pour regrouper les termes des opérations du changement de base) [Cooley et Tukey, 1965].

	complexité	domaine de validité
matrice quelconque	$\mathcal{O}(N^3)$	tous défauts linéaires
matrice Töplitz	$\mathcal{O}(N^2)$	défauts linéaires spatialement invariants
matrice circulante	$\mathcal{O}(N \log N)$	défauts linéaires spatialement invariants à réponse impulsionnelle étroite

TAB. 5.1 – Les diverses représentations des défauts, leur complexité et leur domaine d’application.

Dès que les signaux sont un peu longs (quelques centaines d’échantillons), on voit donc l’intérêt d’adopter une représentation par matrice circulante, même si l’on risque de dégrader la solution trouvée par des effets de bord non désirés. Nous verrons plus loin 5.6 comment on peut atténuer les effets de bord lors de la restauration d’images.

5.2 Défauts singuliers et mal conditionnés

Nous avons négligé de nombreux aspects des signaux réels dans les lignes précédentes. Reprenons donc les conditions d’application de la restauration dans des cas plus généraux.

5.2.1 Défauts singuliers

Nous avons évité jusqu’à présent d’évoquer le problème principal du filtre inverse : le défaut peut être non inversible. Cela se traduit, soit par la présence de valeurs nulles dans la fonction de transfert $D(u)$ (représentation continue), soit par la singularité de la matrice Δ (représentation discrète). On dit dans ce cas que le défaut est singulier [Porte et Vignes, 1974, Forsythe et Moler, 1967]. Le filtre inverse, s’il était calculé sans précaution, donnerait alors une divergence inacceptable. Dans l’approche analytique, il est clair que si $D(u) = 0$ pour certaines fréquences, alors $G(u)$ ne porte plus aucune information sur $F(u)$ à ces fréquences. Sans autre information sur le signal à restaurer, nous devons nous résoudre à ne rien savoir sur les fréquences de $\hat{F}(u)$; cette constatation nous conduit à proposer une solution pour tenir compte de la singularité du défaut, souvent appelée ” solution principale au sens de Bracewell ” [Bracewell, 1995], version améliorée du filtre inverse. Nous la représentons par sa fonction de transfert, notée $W_B(u)$ qui vérifie :

$$W_B(u) \begin{cases} = 0 & \text{si } D(u) = 0 \\ = \frac{1}{D(u)} & \text{sinon} \end{cases} \quad (5.12)$$

Ce filtre est identique au filtre inverse si le défaut n'est pas singulier, mais ne diverge plus en cas de singularité.

Dans une représentation vectorielle (équation 5.8), deux types de problèmes se présentent :

1. d'une part la matrice Δ peut être rectangulaire,
2. d'autre part, elle peut ne pas être de rang plein.

Ces deux problèmes trouvent une solution par la matrice pseudo inverse Δ^- [Albert, 1976, Hestenes, 1975, Decreusefond et Maruani, 2005] qui est solution de l'équation (issue de l'équation 5.6) :

$$\begin{aligned} \Delta^t \mathbf{g} &= \Delta^t \Delta \mathbf{f} \\ \Delta^- &= (\Delta^t \Delta)^{-1} \Delta^t \end{aligned}$$

Cette matrice a la propriété suivante : son rang est égal à celui de la plus grande matrice carrée régulière issue de Δ . Si Δ est régulière, Δ^- est égale à Δ^{-1} . Comme l'indique le dernier facteur Δ^t , si Δ est singulière, Δ^- a une valeur propre nulle pour tout vecteur propre appartenant au noyau de $\Delta^t \Delta$. La restauration consiste donc bien à inverser le plus grand système régulier issu de Δ , et d'annuler toutes les composantes de $\hat{\mathbf{f}}$ appartenant au noyau de Δ (ou de $\Delta^t \Delta$, qui est le même).

5.2.2 Système mal conditionné

Conditionnement

Mais la matrice Δ est, non seulement parfois singulière, mais également souvent mal conditionnée [DUR], c'est-à-dire que certaines de ses valeurs propres non nulles sont très faibles, qu'elles ne pourront être inversées exactement, et risquent, selon la méthode mise en œuvre, de donner des résultats différents. On définit le degré de conditionnement d'un système linéaire comme le rapport de sa plus grande à sa plus petite valeur propre :

$$\kappa(d) = \frac{\max_j \lambda_{jj}}{\min_j \lambda_{jj}}$$

Un système est mal conditionné si κ est typiquement de l'ordre de l'inverse de la précision de la machine sur laquelle on travaille. Il doit alors être régularisé. De nombreuses techniques ont été proposées pour régulariser l'équation 5.2 [Twomey, 1965, Phillips, 1962, Tikonov et Arsenine, 1974]. Nous présentons ici l'approche analytique proposée par Wiener [Hellstrom, 1967].

Le filtre de Wiener

Wiener suppose que l'équation 5.2 est perturbée par un bruit $b(x)$ additif, aléatoire, de moyenne nulle, non corrélé au signal. L'équation 5.2 devient alors :

$$g(x) = f(x) * d(x) + b(x)$$

Wiener recherche une solution linéaire (donc par convolution de g avec un filtre dénoté w_W). Il veut une solution qui minimise ϵ l'espérance de l'erreur quadratique intégrale entre le signal et son estimée, espérance estimée sur toutes les réalisations possibles du bruit b :

$$\epsilon = \left\langle \int_{-\infty}^{+\infty} (f(x) - \hat{f}(x))^2 dx \right\rangle$$

Appliquant le théorème de Parseval, on obtient :

$$\epsilon = \left\langle \int_{-\infty}^{+\infty} |(F(u) - \hat{F}(u))^2 du \right\rangle$$

remplaçant \hat{F} par GW_W , permutant intégrale et espérance, et imposant à chaque composante u de minimiser sa contribution à ϵ , on obtient :

$$\frac{\partial \epsilon}{\partial W_W(u)} = 0 \quad \Rightarrow \quad W_W(u) = \frac{D^*}{DD^* + \Phi_b/\Phi_f} \quad (5.13)$$

où Φ_b et Φ_f sont respectivement les densités de puissance du bruit et du signal :

$$\Phi_b(u) = \langle B(u)B^*(u) \rangle \quad (5.14)$$

$$\Phi_f(u) = \langle F(u)F^*(u) \rangle \quad (5.15)$$

Nous voyons le rôle du filtre de Wiener : dans le cas d'un défaut singulier, et en l'absence de bruit, il conduit à une solution équivalente à celle du filtre inverse. Lorsque le défaut est singulier, il est égal à la solution de Bracewell. Mais si D , sans être nul, devient faible, le rapport du dénominateur l'emporte sur le terme quadratique DD^* , et fait tendre le filtre vers 0. Le filtre ne corrige plus, mais en ces valeurs où le signal est faible devant le bruit, le bruit n'est pas amplifié par le filtre.

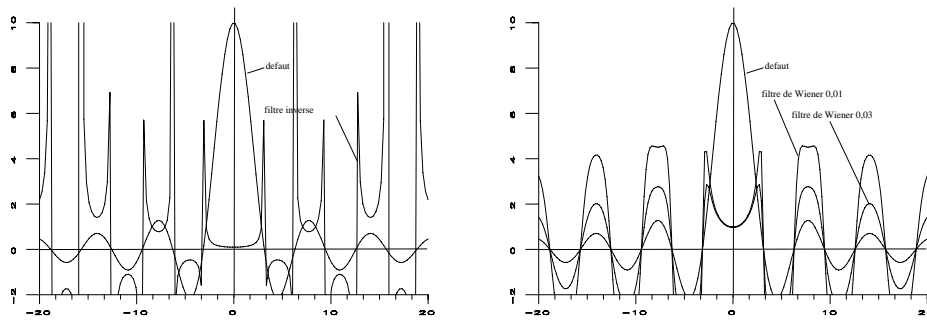


FIG. 5.2 – A gauche : fonction de transfert ($\sin u/u$) d'un défaut de bougé à vitesse constante et le filtre inverse permettant de le corriger. A droite, deux filtres de Wiener de ce même défaut obtenus à partir de deux valeurs différentes du rapport Φ_b/Φ_f supposés constants.

Des filtres de Wiener identiques existent pour l'approche algébrique. En règle générale, une solution algébrique régularisée de l'équation 5.6 utilise une matrice inverse approchée de Δ^{-1} sous la forme :

$$\Delta^+ = (\Delta^t \Delta + \Phi)^{-1} \Delta^t$$

où Φ est une matrice diagonale à termes ϕ_j strictement positifs et majorés par 1. Pour les composantes j où ϕ_j est très faible, la correction sera celle du filtre inverse, pour les composantes où ϕ_j est égal à 1, il n'y aura aucune correction si $\Delta^t \Delta$ est très faible. De la même façon, on déduit le filtre de Wiener matriciel :

$$\Delta^W = (\Delta^t \Delta + \Phi_f^{-1} \Phi_b I)^{-1} \Delta^t \quad (5.16)$$

à partir des matrices de covariance de signal Φ_f et de bruit Φ_b et de la matrice identité I .

Estimation des spectres de densité de puissance

Les deux termes de densité spectrale Φ_b et Φ_f revêtent une importance particulière dans la régularisation du filtrage dans l'approche de Wiener. Ils sont malheureusement le plus souvent inconnus lors de la résolution d'un problème particulier. Une façon très répandue de circonvenir cette ignorance consiste à considérer le rapport signal à bruit $R = \Phi_b/\Phi_f$ comme une constante qui est, soit estimée à partir des connaissances acquises sur le capteur, soit recherchée par des essais successifs de restauration par des valeurs de R variables.

Une méthode un peu plus rigoureuse consiste à exploiter les propriétés statistiques des images naturelles telles que nous les avons vues au chapitre 2. Considérant le bruit blanc de variance σ_b , et l'image comme un processus markovien à accroissements gaussiens, caractérisé par la décroissance α de sa corrélation (cf. équation 2.12), le terme de régularisation prend la forme :

$$\frac{\Phi_b(u)}{\Phi_f(u)} = \frac{\sigma_b^2(\alpha^2 + u^2)}{2\alpha}$$

Il reste alors à estimer les seules inconnues σ_b et α .

5.2.3 Décomposition en valeurs singulières

La représentation en valeurs singulières vient généraliser aux systèmes non carrés la décomposition en valeurs propres et fournit un cadre théorique unificateur pour présenter les résultats acquis à ce point. Partant de la formulation vectorielle de l'équation 5.8, la décomposition en valeurs singulières (*Singular Value Decomposition* ou SVD [Golub et van Loan, 1996]) propose d'écrire la matrice rectangulaire Δ sous la forme :

$$\Delta = U^t \Omega V$$

où V est une matrice unitaire, matrice des vecteurs propres de $\Delta^t \Delta$, U^t est la matrice transposée (égale à la matrice inverse) de la matrice des vecteurs propres de $\Delta \Delta^t$ et Ω une matrice diagonale à valeurs réelles positives ordonnées par ordre décroissant.

$$\Delta^t \Delta = V^t \Omega^2 V \quad (5.17)$$

$$\Delta \Delta^t = U^t \Omega^2 U \quad (5.18)$$

et, avec les notations déjà employées :

$$\Omega^2 = \Lambda \quad (5.19)$$

La restauration de l'équation 5.8 se ramène alors à la succession d'opérations suivantes :

1. pour toutes les valeurs propres de Ω très supérieures au bruit affectant le signal et engendré par les calculs, inverser le système, c'est-à-dire, multiplier par l'inverse de la valeur singulière correspondante chaque composante du signal ;
2. pour toute valeur singulière nulle ou très inférieure au bruit, et en l'absence d'une information a priori sur le signal, annuler les composantes du signal ;
3. pour les valeurs singulières très proches du bruit, mettre en œuvre toute technique de régularisation qui apparaîtra appropriée, en tirant parti au mieux des connaissances existant sur le signal.

Cette démarche serait à conseiller systématiquement si sa complexité en $\mathcal{O}(N^3)$ n'en rendait la mise en œuvre pratiquement impossible.

5.3 Restauration des signaux bidimensionnels

L'approche analytique de la restauration s'accommode bien sûr de signaux de toutes dimensions, puisque la transformée de Fourier s'étend naturellement aux espaces de dimensions supérieures à 1. L'extension des filtres inverses ou de Wiener est triviale.

passent par le calcul d'une série de fonctions f^k , convergent vers l'estimée recherchée \hat{f} et qui permettent donc de s'approcher par étapes de la restauration. Ces méthodes reposent généralement sur une réécriture de l'équation 5.2 (respectivement 5.6), en décomposant la fonction d (resp. la matrice de défaut Δ) en 2 termes :

$$d = d_1 - d_2 \quad \text{ou} \quad \Delta = \Delta_1 - \Delta_2$$

et en résolvant :

$$d_1 * f = d_2 * f + g \quad \text{ou} \quad \Delta_1 \mathbf{f} = \Delta_2 \mathbf{f} + \mathbf{g}$$

en approximant f par la série f^k qui vérifie :

$$d_1 * f^{k+1} = d_2 * f^k + g \quad \text{ou} \quad \Delta_1 \mathbf{f}^{k+1} = \Delta_2 \mathbf{f}^k + \mathbf{g} \quad (5.22)$$

On choisit généralement d_1 et Δ_1 de façon qu'ils soient aisément inversibles afin d'obtenir :

$$\mathbf{f}^{k+1} = \Delta_1^{-1} \Delta_2 \mathbf{f}^k + \Delta_1^{-1} \mathbf{g}$$

et la solution initiale \mathbf{f}^0 peut être choisie, soit identiquement nulle, soit aléatoire, soit enfin, le plus souvent, égale à l'image défectueuse \mathbf{g} .

5.4.1 Méthode de Jacobi

C'est la plus simple des méthodes, on prend d_1 égal à l'unité, ce qui conduit au schéma :

$$f^{k+1} = f^k + g - d * f^k \quad (5.23)$$

qui, s'il converge, vérifie bien : $g - d * f^k = r^k = 0$. La quantité r^k est appelé $k^{\text{ème}}$ résidu de l'itération.

Les conditions de cette convergence sont aisément discutées à partir de la TF de l'équation 5.23 :

$$F^{k+1} = F^k + G - DF^k$$

qui, si l'on choisit $F^0 = G$, donne par récurrence : $F^{k+1} = \sum_{i=0}^k G(1 - D)^i$
équation qui converge vers :

$$\hat{F} = \frac{G}{D}$$

si et seulement si $0 \leq |D| \leq 1$. On reconnaît donc une opération équivalente au filtrage inverse (équation 5.2), avec ses qualités (restauration exacte dans les cas réguliers et bien conditionnés), et ses défauts (divergence dans le cas de système singulier, instabilité dans le cas de mauvais conditionnement). On peut cependant y ajouter des qualités propres aux techniques itératives :

1. il n'est pas nécessaire d'inverser le défaut, mais il suffit de l'appliquer n fois aux solutions successives de l'itération (cela permet en particulier, pour des systèmes d'imagerie disponibles à l'utilisateur, de ne pas déterminer le défaut, mais d'utiliser le système défectueux lui-même pour se corriger [Maître, 1981]) ;
2. si le défaut est très compact (support de d très étroit), l'opération de convolution $d * f$ peut être réduite à un nombre très petit d'opérations, et la solution itérative est alors beaucoup moins coûteuse que le filtrage inverse ;
3. l'opérateur peut voir les résultats des itérations successives de la restauration et intervenir sur elle en interrompant par exemple la correction en cas de divergence du système (l'expérience montre que ce sont généralement les valeurs propres les mieux conditionnées qui sont l'objet des corrections des premières itérations, la divergence n'intervenant qu'après une première phase de restauration [Maître, 1981]) ;
4. la correction itérative se prête bien à l'introduction, dans la recherche de la solution optimale, de connaissances a priori ou de contraintes non linéaires (par exemple contraindre l'image solution à être positive, ou d'amplitude bornée, ou à support étroit, etc.).

En cas de convergence de la méthode de Jacobi, il est possible de modifier le schéma de l'équation 5.23 en introduisant un facteur de relaxation α choisi entre 0 et 2 et permettant, soit d'accélérer la convergence ($\alpha > 1$) on parle alors de sur-relaxation, soit d'améliorer son conditionnement ($\alpha \leq 1$) :

$$f^{k+1} = f^k + \alpha(g - d * f^k)$$

5.4.2 Méthode de Gauss-Seidel

Elle ne prend de forme intéressante qu'en écriture algébrique (équation 5.8). Elle repose sur l'utilisation de la matrice unité pour Δ_1 , et d'une décomposition de Δ_2 en deux matrices, l'une Δ_2' n'ayant que des zéros au-dessus de la diagonale, l'autre Δ_2'' sur et en-dessous de la diagonale. Cette forme permet d'utiliser à l'itération k les valeurs déjà corrigées à cette itération pour corriger les suivantes, selon le schéma :

$$f^{k+1} = g + \Delta_2' f^k + \Delta_2'' f^{k+1}$$

5.4.3 Méthode de plus grande pente

Utilisant le résidu r^k à la $k^{\text{ème}}$ itération, la méthode de la plus grande pente s'écrit, algébriquement :

$$f^{k+1} = f^k + \frac{|r^k|^2}{[r^k]^t \Delta r^k} r^k$$

Ces formules sont des expressions classiques en analyse numérique ; en traitement des images, on leur préfère des techniques dont la convergence est généralement assurée vers la solution principale au sens de Bracewell (ou vers la matrice pseudo-inverse de norme minimale) comme la méthode du gradient conjugué ou les méthodes de projection (méthodes POCS).

5.4.4 Méthode du gradient conjugué

Elle se propose, non pas de résoudre l'équation linéaire 5.4, mais de minimiser la forme quadratique associée :

$$\Lambda = f^t \Delta f - 2 f^t g$$

dans l'hypothèse où Δ est une matrice définie non-négative par la récurrence :

$$\begin{aligned} f^{k+1} &= f^k + a^k \rho^k \\ a^k &= \frac{\rho^k r^k}{\rho^k \Delta \rho^k} \\ r^{k+1} &= r^k - a^k \Delta \rho^k \\ \rho^{k+1} &= r^{k+1} - \frac{r^k \Delta \rho^k}{\rho^k \Delta \rho^k} \rho^k \end{aligned}$$

Si la matrice n'est pas définie non négative (ou même, si elle n'est pas carrée), on résout, comme on l'a vu en section 5.2.1, le système $\Delta^t g = \Delta^t \Delta f$ au lieu du système $g = \Delta f$. Les avantages de la méthode du gradient conjugué sont tout d'abord sa convergence inconditionnelle (vers la solution pseudo-inverse), et ensuite sa très grande vitesse de convergence [Hestenes, 1975].

5.4.5 Méthodes de projection (POCS)

Elles forment une famille de méthodes extrêmement populaires depuis quelques années, qui procèdent en calculant itérativement les projections de l'équation 5.4 sur des variétés linéaires choisies de l'espace des solutions (droites, plans, voire surfaces convexes). Elles prennent la forme générale [Householder et Bauer, 1960] :

$$F^{k+1} = f^k + \frac{V^t r^k}{|\Delta^t V|^2} \Delta^t V$$

où V est un vecteur choisi par l'utilisateur.

5.5 L'estimation des défauts

Nous avons supposé jusque là que le défaut d était parfaitement connu. Cette hypothèse n'est que rarement vraie. Il s'agit principalement des cas où la physique de l'acquisition est parfaitement maîtrisée. Par exemple dans les applications d'imagerie de très haute qualité, la limite de résolution est souvent imposée par la diffraction qui introduit un défaut déterminé par le plus petit diaphragme du système. Dans le cas d'ouvertures circulaires de diamètre D et de focale ϕ , on retrouve pour réponse impulsionnelle la fonction de Bessel déjà rencontrée à l'équation 3.3 :

$$d(x, y) = \frac{J_1^2(\pi \rho D \sqrt{x^2 + y^2} / \phi)}{(\pi \sqrt{x^2 + y^2} \rho D / \phi)^2} \quad (5.24)$$

On est également dans ces conditions favorables lorsque le système d'imagerie est constitué autour de l'équation de convolution 5.2 afin de permettre de mesurer indirectement une grandeur f inaccessible. C'est le cas de la tomographie par exemple, ou de l'imagerie sismique ou encore dans les systèmes à *synthèse d'ouverture* comme on en rencontre en radar [Maître, 2001], en acoustique ou en astronomie [Roddier, 1999].

Ainsi, en tomographie médicale [Grangeat, 2001], le signal mesuré g est l'intégrale, le long de la ligne de propagation des rayons X , du signal intéressant f , la densité des tissus biologiques observés. Lorsque la source de rayons X et le capteur se déplacent en translation et en rotation autour du patient, les signaux f et g sont liés par une réponse impulsionnelle $d(x, y) = 1/\sqrt{x^2 + y^2}$ [Grangeat, 2001].

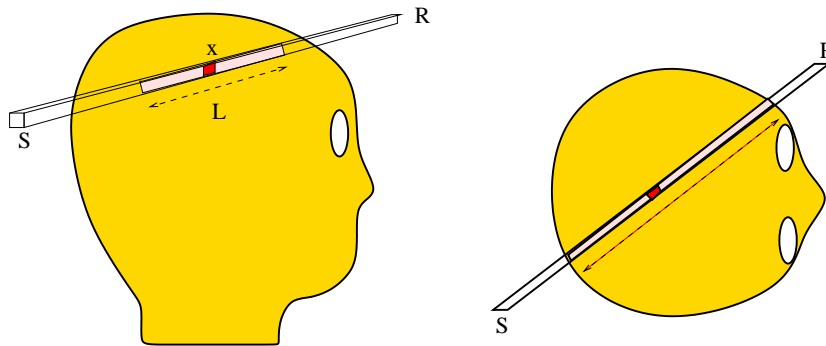


FIG. 5.3 – Construction de l'image en tomographie médicale. S est la source de rayons X , R est le récepteur. Une mesure g est l'intégrale le long de L de l'atténuation des rayons X due aux tissus.

5.5.1 Détermination par calibrage

Il s'agit surtout des cas où il est possible d'étalonner le système optique, c'est à dire de placer en entrée du système un signal f parfaitement connu. On détermine alors le défaut d par inversion de l'équation 5.2 mais cette fois en inversant f , dans la mesure du possible.

C'est ainsi que l'on procède en astronomie ainsi qu'en télédétection, en utilisant pour f soit des étoiles isolées (alors $f = \delta$) soit des mires naturelles (ponts, routes), ou artificielles tracées au sol.

5.5.2 Détermination statistique

Mais dans de nombreux cas on ne dispose que de la seule image défectueuse g . Il est alors très difficile de remonter au défaut sans aucune information a priori sur celui-ci. Heureusement on sait souvent quelle classe de défaut peut avoir affecté l'image. C'est le plus souvent un défaut de mise au point, ou un défaut de bougé, l'objet ou la caméra étant en mouvement pendant la prise de vue. Il s'agit alors de confirmer le défaut, puis de déterminer son ampleur. Une bonne façon de procéder consiste à étudier le spectre P_g de densité de puissance de l'image défectueuse. D'après l'équation 5.4, il s'exprime par :

$$P_g = \langle |G|^2 \rangle = \langle |FD|^2 \rangle = \langle |F|^2 \rangle |D|^2 = P_f |D|^2 \quad (5.25)$$

Or la plupart des défauts présentent des zéros dans leur fonction de transfert D . C'est ainsi vrai pour :

1. les bougés à vitesse constante dont la réponse impulsionnelle est une fonction porte dans la direction du déplacement et de largeur fixée par la vitesse et le temps de pose : D est alors un sinus cardinal ;
2. les défauts de mise au point dont la réponse impulsionnelle est un cercle et la fonction de transfert une fonction de Bessel de première espèce semblable à celle de l'équation 5.24, mais dans le plan des fréquences.

On voit sur l'équation 5.25 que ces zéros se retrouvent dans le spectre P_g , alors qu'ils sont très improbables dans le spectre d'une image (cf. chapitre 2). En détectant ces zéros, il est alors possible de confirmer la présence du défaut et de déterminer les paramètres qui le caractérisent.

5.6 Réduction des effets de bord

Nous avons vu qu'il était très souvent préférable, pour des questions d'efficacité de calcul, d'utiliser la forme circulante du défaut plutôt que sa forme Töplitz. On est alors conduit à commettre une erreur de reconstruction puisque l'on fait implicitement intervenir par périodisation de l'image, des signaux à droite et à gauche, ainsi qu'en haut et en bas, qui ne sont pas les signaux qui ont contribué au défaut. Comment réduire l'effet de ces signaux parasites ?

Une première solution, de médiocre qualité, consiste à prolonger l'image de marges uniformes (par exemple de niveau égal à la valeur moyenne de l'image), puis de procéder à la déconvolution de l'image ainsi agrandie. On peut également répéter la première et la dernière ligne ou la première et la dernière colonnes à cet effet.

Préférable est l'approche qui consiste à reproduire l'image par une symétrie miroir sur tous ses bords. Les images répliques ainsi créées possédant un spectre de densité de puissance très proche de celui du signal d'origine, perturbent assez peu la restauration et donnent des résultats souvent de bonne qualité.

Ces deux approches se traduisent l'une et l'autre par un accroissement des calculs puisque tout se passe comme si l'on traitait en pratique une image plus grande.

Chapitre 6

Morphologie mathématique

Chapitre rédigé par Isabelle BLOCH

6.1 Introduction et préliminaires

La morphologie mathématique est une théorie essentiellement non linéaire, utilisée en particulier en analyse d'images, dont le but est l'étude des objets en fonction de leur forme, de leur taille, des relations avec leur voisinage (en particulier topologiques), de leur texture, et de leurs niveaux de gris ou de leur couleur. Par les transformations qu'elle propose, elle se situe à différents niveaux du traitement d'images (filtrage, segmentation, mesures, analyse de texture) et fournit ainsi des outils pour la reconnaissance des formes. La morphologie mathématique, développée à l'origine pour l'étude des matériaux poreux, trouve maintenant ses applications dans de nombreux domaines du traitement d'images, aussi bien 2D que 3D, en biologie et cytologie quantitative, en imagerie médicale, en imagerie aérienne et satellitaire, en robotique et vision par ordinateur, en contrôle industriel non destructif, dans les études sur les documents et les œuvres d'art. Hors du domaine du traitement des images, on trouve des applications par exemple en analyse de données, ou encore en théorie des jeux.

La morphologie mathématique a été développée à l'origine à l'École des Mines de Paris. Elle repose essentiellement sur les travaux de G. Matheron effectués dans les années 60-70, puis sur ceux de J. Serra et de son équipe. Depuis ces premiers développements, elle a pris une ampleur internationale et plusieurs équipes s'y consacrent.

Elle s'appuie sur la théorie des ensembles, des treillis, de la topologie des fermés et des probabilités. Elle s'applique ainsi à des structures algébriques variées (ensembles, fonctions, mais également ensembles flous ou propositions logiques), pouvant avoir un caractère aléatoire (pour l'analyse de textures par exemple).

Ce chapitre n'est qu'une introduction à la morphologie mathématique. Il en donne succinctement les bases mathématiques, décrit les opérations de base et leurs principales propriétés, et présente des exemples d'outils de filtrage et de segmentation qui en sont déduits. On se placera de manière générale dans l'espace \mathbb{R}^n , puis on définira les notions discrètes associées, dans \mathbb{Z}^n ou \mathbb{N}^n . Une présentation détaillée peut être trouvée dans les ouvrages [Serra, 1982a, Serra (Ed.), 1988, Dougherty (Ed.), 1992, Schmitt et Mattioli, 1994a, Soille, 1999].

On trouve à la base des transformations de morphologie mathématique quatre principes fondamentaux qui guident leur construction et qui sont vérifiés pour la plupart des opérations. Ils sont exprimés ci-dessous pour une opération Ψ quelconque, agissant sur un ensemble ou une fonction f définie sur \mathbb{R}^n .

Compatibilité avec les translations : Ce principe exprime l'indépendance des transformations par rapport à

l'origine de l'espace¹ :

$$\forall t \in \mathbb{R}^n, \Psi(f + t) = \Psi(f) + t.$$

Compatibilité avec les homothéties : Ce principe assure l'indépendance des transformations par rapport à un paramètre d'échelle :

$$\forall \lambda \in \mathbb{R}, \Psi(\lambda f) = \lambda \Psi(f).$$

Connaissance locale : Pour connaître $\Psi(f)$ dans un domaine Z' de \mathbb{R}^n , il suffit de connaître f dans un domaine Z de \mathbb{R}^n :

$$\forall Z' \subset \mathbb{R}^n, Z' \text{ borné}, \exists Z \subset \mathbb{R}^n, Z \text{ borné}, [\Psi(f \cap Z)] \cap Z' = \Psi(f) \cap Z',$$

où $f \cap Z$ désigne la restriction de f au domaine Z .

Semi-continuité : ce principe (voir définitions 18 et 22) assure la robustesse des transformations.

Outre ces propriétés fondamentales, les opérateurs de morphologie mathématique peuvent avoir des propriétés algébriques dont les principales sont définies ci-dessous.

Définition 1. Croissance : Une transformation Ψ sur des ensembles de \mathbb{R}^n ou des fonctions de \mathbb{R}^n dans $\overline{\mathbb{R}}$ est croissante si :

$$\forall X, Y, X \subset Y \Rightarrow \Psi(X) \subset \Psi(Y),$$

$$\forall f, g, f \leq g \Rightarrow \Psi(f) \leq \Psi(g).$$

Définition 2. Extensivité et anti-extensivité : Une transformation Ψ sur des ensembles de \mathbb{R}^n ou des fonctions de \mathbb{R}^n dans $\overline{\mathbb{R}}$ est extensive si :

$$\forall X, X \subset \Psi(X),$$

$$\forall f, f \leq \Psi(f).$$

Ψ est anti-extensive si :

$$\forall X, \Psi(X) \subset X,$$

$$\forall f, \Psi(f) \leq f.$$

Définition 3. Idempotence : Une transformation Ψ sur des ensembles de \mathbb{R}^n ou des fonctions de \mathbb{R}^n dans $\overline{\mathbb{R}}$ est idempotente si :

$$\forall X, \Psi[\Psi(X)] = \Psi(X),$$

$$\forall f, \Psi[\Psi(f)] = \Psi(f).$$

Définition 4. Dualité : Deux transformations Ψ et Φ sur des ensembles de \mathbb{R}^n ou des fonctions de \mathbb{R}^n dans $\overline{\mathbb{R}}$ sont duales par rapport à la complémentation si :

$$\forall X, \Psi(X^C) = [\Phi(X)]^C,$$

où X^C désigne le complémentaire de X dans \mathbb{R}^n (c'est-à-dire $\mathbb{R}^n \setminus X$),

$$\forall f, \Psi(-f) = -\Phi(f).$$

¹Cette propriété est appelée aussi invariance spatiale dans d'autres théories du traitement d'images.

Notons que pour des images numériques, représentées par des fonctions à valeurs positives et bornées par une valeur maximale M , la dualité s'exprime par :

$$\Psi(M - f) = M - \Phi(f).$$

Nous verrons dans la suite que ce principe de dualité exprime le fait que deux opérations, qui ne sont pas l'inverse l'une de l'autre, ont des effets contraires (les opérations de morphologie mathématique étant généralement non inversibles).

Pour décrire de manière très synthétique la boîte à outils de la morphologie mathématique, il faut retenir les points suivants :

- les transformations sont non linéaires, elles sont fondées sur des opérations de type \sup et \inf ;
- les transformations sont généralement non inversibles, et elles perdent donc de l'information ; le travail du morphologue consiste alors à déterminer les transformations adaptées à son problème, c'est-à-dire qui vont simplifier les images en retenant l'information pertinente ;
- des propriétés analytiques et algébriques sont attachées aux opérations, ce qui permet d'assurer des propriétés précises sur les objets ou images issues des transformations ; c'est sur ces propriétés que l'on s'appuie pour enchaîner les transformations afin de résoudre un problème particulier ;
- aux transformations sont également associés des algorithmes, qui permettent de les appliquer de manière efficace.

Les parties suivantes s'attachent à décrire plus précisément ces concepts pour les principales opérations morphologiques.

6.2 Les quatre opérations

6.2.1 Notion d'élément structurant et cadre ensembliste

Des premières applications à l'étude des milieux poreux est née l'approche ensembliste de la morphologie mathématique. Elle s'applique à des images ou objets binaires et les étudie sous l'angle de leurs relations avec un ensemble fixé. Cet ensemble, dont on choisit la forme et la taille, est appelé élément structurant. Les relations sont de type ensembliste (réunion, intersection, etc.). Étant donné un élément structurant et une relation, l'image (ou l'objet) de départ est transformé en translatant l'élément structurant en tout point et en examinant si la relation entre l'objet et l'élément structurant translaté est vérifiée. Les propriétés des opérations ainsi définies et des images transformées découlent de la théorie des ensembles. C'est selon ce principe que seront par exemple définies la dilatation et l'érosion binaires. L'élément structurant définit un voisinage autour de chaque point de l'image et ce sont donc des propriétés locales des objets qui sont ainsi mises en évidence.

En pratique, les images sont définies sur des trames discrètes. Le problème consiste alors à étudier dans quelle mesure les définitions et les propriétés des transformations décrites pour l'espace continu sont transposables à un espace discret.

Les trames les plus utilisées en dimension 2 sont la trame carrée (ou rectangulaire) et la trame hexagonale où les points sont disposés en quinconce. Plus de détails peuvent être trouvés dans la partie sur la géométrie discrète (chapitre 4) ou dans les références qui y sont citées.

Les voisinages élémentaires associés aux différentes trames constituent les éléments structurants élémentaires (de taille 1) utilisés en morphologie mathématique. Effectuer une dilatation de taille n signifie alors effectuer n fois une itération de taille 1, par un de ces éléments structurants.

6.2.2 Erosion et dilatation binaires

Les définitions qui suivent reposent sur la définition classique de l'addition de Minkowski de deux ensembles, qui correspond à l'addition vectorielle :

$$X \oplus Y = \{x + y \mid x \in X, y \in Y\}. \quad (6.1)$$

Dans toute cette partie, on considérera des ensembles de \mathbb{R}^n . Un élément structurant est également un ensemble dont on se donne la forme et la taille, et dont on particularise un point (l'origine) qu'on appelle le centre (qui n'est pas forcément le centre géométrique).

Définition 5. Dilatation binaire : La dilatation binaire d'un ensemble X par un élément structurant B est définie comme l'ensemble obtenu par addition de Minkowski de X par le symétrique \check{B} de B par rapport à son centre :

$$D(X, B) = X \oplus \check{B} = \{x + y \mid x \in X, y \in \check{B}\}. \quad (6.2)$$

Il existe des définitions légèrement différentes, pour lesquelles la dilatation est égale à l'addition de Minkowski. Cela implique des modifications des propriétés, en particulier dans les relations de dualité, mais permet une meilleure conformité avec le cadre algébrique que nous verrons plus loin.

L'équation 6.2 est équivalente aux expressions suivantes :

$$\begin{aligned} D(X, B) &= \bigcup_{x \in X} \check{B}_x \\ &= \{x \in \mathbb{R}^n \mid B_x \cap X \neq \emptyset\}, \end{aligned} \quad (6.3)$$

où B_x désigne le translaté de l'élément structurant au point x (c'est-à-dire le centre de l'élément structurant coïncide avec x).

Dans cette formulation (équation 6.3) apparaissent clairement les concepts de l'approche ensembliste : ici la relation imposée entre la forme étudiée et l'élément structurant est l'intersection. Ainsi, un point appartient à l'objet résultat si l'élément structurant centré en ce point intersecte l'objet initial.

Propriétés de la dilatation. La dilatation a les propriétés suivantes (les démonstrations sont immédiates en utilisant les formules 6.2 ou 6.3) :

- elle est extensive ($X \subset D(X, B)$) si le centre de B appartient à B ,
- elle est croissante ($X \subset Y \Rightarrow D(X, B) \subset D(Y, B)$),
- $B \subset B' \Rightarrow D(X, B) \subset D(X, B')$,
- elle commute avec la réunion mais pas avec l'intersection :

$$D(X, B \cup B') = D(X, B) \cup D(X, B'),$$

$$D(X, B \cap B') \subset D(X, B) \cap D(X, B'),$$

- elle vérifie une relation d'itération (associativité) :

$$D[D(X, B), B'] = D(X, B \oplus B').$$

Ces propriétés algébriques ont des conséquences importantes sur les applications de cette transformation, à la fois du point de vue algorithmique et du point de vue des propriétés des objets obtenus. Par exemple, la relation d'itération permet de calculer une dilatation par un disque de rayon 2 soit directement, soit comme une suite de deux dilatations par un disque de rayon 1.

De plus, la dilatation vérifie les 4 principes fondamentaux donnés en introduction.

Exemples. La dilatation par un segment centré à une de ses extrémités a pour effet ;; d'étendre ;; l'objet dans la direction opposée à celle du segment (alors que l'addition de Minkowski étend l'objet dans la direction du

segment). Notons que la distinction entre addition de Minkowski et dilatation tombe dès qu'on utilise des éléments structurants symétriques, ce qui est souvent le cas.

La dilatation par un disque a pour effet d'augmenter la taille des objets selon la taille du disque, de relier entre elles les composantes proches et de boucher les petits trous (plus petits que l'élément structurant).

La figure 6.1 illustre ces effets.

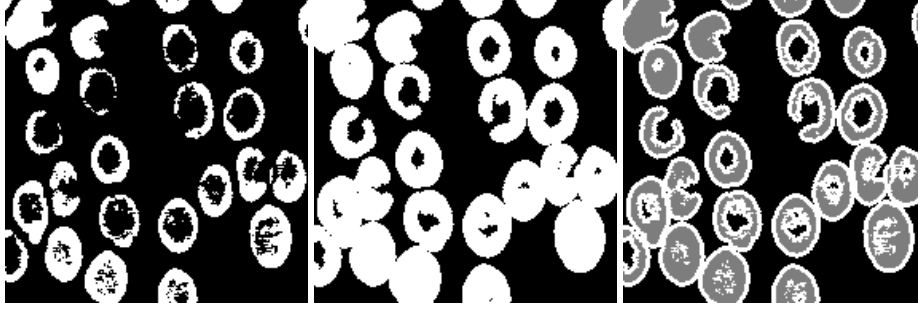


FIG. 6.1 – Exemple de dilatation binaire (de gauche à droite : image initiale, dilatation par un disque de taille 3, différence : en blanc, les parties rajoutées par la dilatation).

Définition 6. Érosion binaire : L'érosion d'un ensemble X par un élément structurant B est définie par :

$$E(X, B) = \{x \in \mathbb{R}^n / B_x \subset X\} \quad (6.4)$$

$$= \{x / \forall y \in B, x + y \in X\} = X \ominus \check{B}. \quad (6.5)$$

L'équation 6.4 correspond à l'approche ensembliste, où cette fois la relation imposée entre l'élément structurant et la forme est l'inclusion. L'équation 6.5 fait référence à la soustraction de Minkowski notée \ominus .

Propriétés de l'érosion. La propriété essentielle de l'érosion est qu'elle est la transformation duale de la dilatation par rapport à la complémentation :

$$E(X, B) = [D(X^C, B)]^C. \quad (6.6)$$

Ainsi, il est équivalent d'éroder un objet ou de dilater son complémentaire. Cette propriété peut également être présentée comme définition, dont on déduit alors les expressions 6.4 et 6.5.

L'érosion a les propriétés algébriques suivantes (elles peuvent être démontrées soit directement à partir des formules 6.4 et 6.5, soit déduites de celles de la dilatation grâce à la formule de dualité 6.6) :

- elle est anti-extensive si le centre de B appartient à B ,
- elle est croissante,
- $B \subset B' \Rightarrow E(X, B') \subset E(X, B)$,
- elle satisfait les relations suivantes par rapport à la réunion et à l'intersection (en particulier, elle commute avec l'intersection) :

$$E[(X \cap Y), B] = E(X, B) \cap E(Y, B),$$

$$E[(X \cup Y), B] \supset E(X, B) \cup E(Y, B),$$

$$E[X, (B \cup B')] = E(X, B) \cap E(X, B'),$$

$$E[X, (B \cap B')] \supset E(X, B) \cup E(X, B'),$$

- elle satisfait la relation d'itération suivante :

$$E[E(X, B), B'] = E(X, B \oplus B'),$$

- la succession d’une érosion et d’une dilatation vérifie l’inclusion suivante :

$$D[E(X, B), B'] \subset E[D(X, B'), B]. \quad (6.7)$$

De plus, l’érosion vérifie les 4 principes énoncés en introduction.

Exemples. L’érosion par un disque a les effets suivants : l’objet est diminué selon la taille de l’élément structurant, les composantes connexes de l’objet plus petites que l’élément structurant sont supprimées, les parties des objets reliées par des isthmes plus fins que l’élément structurant sont déconnectées.

La figure 6.2 illustre ces effets.

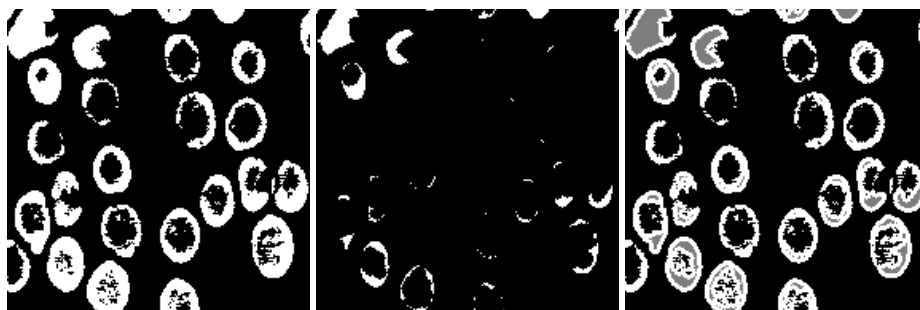


FIG. 6.2 – Exemple d’érosion binaire (de gauche à droite : image initiale, érosion par un disque de taille 3, différence : en blanc, les parties supprimées par l’érosion).

Liens avec la fonction distance. La fonction distance (voir chapitre 4) a de nombreuses applications en morphologie mathématique (outre toutes les applications nécessitant une estimation de la distance entre objets comme le recalage). En voici quelques exemples :

- elle permet d’effectuer très rapidement des dilations et érosions de taille quelconque : en effet, la fonction distance calculée avec les masques de la figure 4.33 a et b (chapitre 4) fournit des courbes de niveaux autour de l’objet qui correspondent exactement aux points qui seraient rajoutés à l’objet par des dilations successives par les éléments structurants élémentaires ; il suffit alors de seuiller la fonction distance à la taille de dilatation souhaitée pour obtenir, en deux passes seulement, le dilaté (pour l’érosion, on procède de même, en calculant cette fois la distance des points de l’objet au complémentaire) ;
- elle permet de réaliser des dilations et érosions par des éléments structurants proches d’un disque (par exemple avec le masque de la figure 4.33 d, chapitre 4) ;
- elle permet d’obtenir très rapidement les érodés ultimes d’un ensemble d’objets X (c’est-à-dire les composantes connexes de $E(X, B_r)$ qui disparaissent par une érosion de taille $r + 1$), comme maxima régionaux de la fonction distance ;
- elle peut être utilisée pour le calcul de squelettes.

La figure 6.3 montre que les courbes d’iso-distances à l’intérieur des objets sont les érodés successifs de ces objets.

6.2.3 Erosion et dilatation de fonctions

La généralisation des transformations binaires à des transformations numériques (sur des fonctions) peut s’effectuer de deux manières :

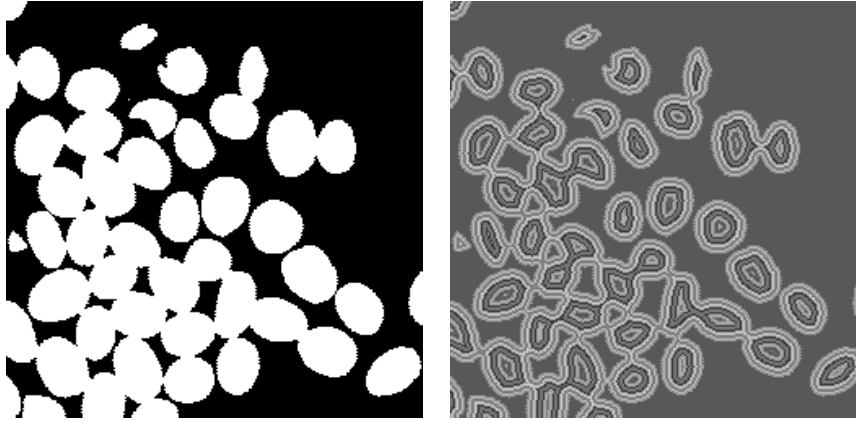


FIG. 6.3 – Les courbes d’iso-distances à l’intérieur des objets sont les érodés successifs.

1. soit en remplaçant dans les définitions tous les concepts ensemblistes par leurs équivalents fonctionnels :

$$\begin{aligned}
 \cup &\rightarrow \sup / \vee \\
 \cap &\rightarrow \inf / \wedge \\
 \subset &\rightarrow \leq \\
 \supset &\rightarrow \geq
 \end{aligned}
 \tag{6.8}$$

2. soit en considérant le sous-graphe de la fonction (voir définition 23) comme un ensemble binaire (dans un espace de dimension $n + 1$) et en lui appliquant les opérations binaires.

Ces deux approches sont équivalentes (pour des fonctions semi-continues supérieurement à valeurs dans $\mathbb{R} \cup \{-\infty\}$ et des éléments structurants bornés) mais la première est plus opérationnelle. Les équivalences des expressions 6.8 se retrouveront dans les définitions et les propriétés qui suivent.

Cette généralisation est indispensable pour pouvoir traiter des images à niveaux de gris. Nous considérerons tout d’abord des éléments structurants binaires, puis nous donnerons les définitions générales des transformations sur des fonctions avec des éléments structurants fonctionnels.

Définition 7. Dilatation fonctionnelle (*élément structurant binaire*) : La dilatation d’une fonction f par un élément structurant B est la fonction définie par :

$$\forall x \in \mathbb{R}^n, \quad D(f, B)(x) = \sup\{f(y) \mid y \in B_x\}. \tag{6.9}$$

L’équation 6.9 fait apparaître explicitement que les valeurs de f intervenant dans le résultat de l’opération en un point sont celles prises dans un voisinage de ce point, ce voisinage étant défini par l’élément structurant².

Propriétés. La dilatation a les propriétés suivantes (analogues des propriétés de la dilatation binaire) :

- elle est extensive si le centre de B appartient à B ,
- elle est croissante,
- $D(f \vee g, B) = D(f, B) \vee D(g, B)$,
- $D(f \wedge g, B) \leq D(f, B) \wedge D(g, B)$,
- elle vérifie les quatre principes énoncés en introduction.

Exemples. Sur une image à niveaux de gris, la dilatation par un disque augmente les niveaux de gris, propage les maxima locaux des niveaux de gris (dans une région correspondant à la taille et à la forme de l’élément structurant).

La figure 6.4 illustre ces effets.

²On a le même résultat pour la dilatation binaire.



FIG. 6.4 – Exemple de dilatation numérique (de gauche à droite : image initiale, considérée comme une fonction sur l'espace à deux dimensions, dilatation par un disque de taille 3).

Définition 8. Érosion fonctionnelle (élément structurant binaire) : L'érosion d'une fonction f par un élément structurant B est la fonction définie par :

$$\forall x \in \mathbb{R}^n, \quad E(f, B)(x) = \inf\{f(y) \mid y \in B_x\}. \quad (6.10)$$

Ici encore, la valeur prise en un point dépend uniquement des valeurs de f dans un voisinage, défini par B , de ce point.

Propriétés. L'érosion a les propriétés suivantes (similaires à celles de l'érosion binaire) :

- elle est duale de la dilatation,
- elle est anti-extensive si le centre de B appartient à B ,
- elle est croissante,
- $E(f \vee g, B) \geq E(f, B) \vee E(g, B)$,
- $E(f \wedge g, B) = E(f, B) \wedge E(g, B)$,
- elle vérifie les quatre principes de l'introduction.

Exemples. L'érosion d'une image à niveaux de gris par un disque a pour effet de diminuer les niveaux, et de propager les minima dans une région définie par l'élément structurant.

La figure 6.5 illustre ces effets.

Notons que la dilatation et l'érosion sont des cas particuliers de filtres de rang (voir chapitre 10). Dans un tel filtre, les niveaux de gris des voisins d'un point sont classés par ordre croissant et la valeur de ce point est remplacée par celle d'un rang donné. Le min utilisé dans l'érosion correspond donc à un filtre de rang 1 et le max de la dilatation à un filtre de rang p (si le voisinage comporte p points). Cependant ce ne sont pas des filtres au sens morphologique du terme.

Cas discret. Une fois les notions de topologie définies sur la trame discrète, la transposition des définitions morphologiques de base ne pose pas de problème. Elles s'expriment de la même manière que dans le cas continu, ou plus simplement puisque les \sup et les \inf deviennent des \max et des \min lorsque l'élément structurant B contient un nombre fini de points. Par exemple, la dilatation par un élément structurant de taille 1 en 4-connexité, s'exprime en tout point M de coordonnées (x, y) de la trame par :

$$D(f, B)(M) = \max\{f(x, y - 1), f(x - 1, y), f(x, y), f(x + 1, y), f(x, y + 1)\},$$

que f soit une fonction ou un ensemble. Cette formule fournit en même temps un moyen opérationnel simple pour calculer la dilatation.



FIG. 6.5 – Exemple d'érosion numérique (de gauche à droite : image initiale et érosion par un disque de taille 3).

Il est remarquable de noter que toutes les propriétés de la dilatation, de l'érosion, de l'ouverture et de la fermeture (qui seront vues plus loin) sont encore valables sur une trame discrète (si l'on exclut celles liées à la semi-continuité qui n'ont pas de sens dans ce cas)³.

Si la transposition de ces opérations de base est si aisée, c'est que celles-ci impliquent essentiellement des notions ensemblistes (éventuellement généralisées à des fonctions comme nous l'avons vu plus haut) et des concepts topologiques très simples, de voisinage entre points uniquement.

Des transformations plus complexes impliquant des notions topologiques plus globales (préservation de la connexité et de l'homotopie d'un objet par exemple) sont beaucoup plus délicates à transposer. Ce sera le cas par exemple du squelette (voir parties 6.8.3 et 6.8.4) : les définitions continues ont de très bonnes propriétés mais leur transposition directe au cas discret en fait perdre la plupart. On est alors conduit à redéfinir le squelette directement dans l'espace discret, par des procédés qui garantissent les propriétés que l'on veut conserver.

Sans aller chercher des transformations aussi complexes, une simple rotation peut dégrader les propriétés topologiques d'un objet (par exemple lui faire perdre sa connexité, ou faire apparaître des trous là où il n'y en avait pas).

Considérons maintenant des éléments structurants fonctionnels, c'est-à-dire des fonctions g de \mathbb{R}^n dans $\mathbb{R} \cup \{-\infty\}$ telles que $\{x \in \mathbb{R}^n / g(x) \neq -\infty\}$ est borné.

Définition 9. Dilatation par une fonction : La dilatation d'une fonction f par une fonction g est la fonction définie par :

$$\forall x \in \mathbb{R}^n, \quad D(f, g)(x) = \sup\{f(y) + g(y - x) / y \in \mathbb{R}^n\}. \quad (6.11)$$

Définition 10. Érosion par une fonction : L'érosion d'une fonction f par une fonction g est la fonction définie par :

$$\forall x \in \mathbb{R}^n, \quad E(f, g)(x) = \inf\{f(y) - g(y - x) / y \in \mathbb{R}^n\}. \quad (6.12)$$

L'intérêt de ces éléments structurants est qu'ils permettent de réaliser des opérations plus fines qu'avec des éléments structurants binaires. Par exemple, il est ainsi possible de modifier un relief (l'écarter par exemple) dans n'importe quelle direction, et pas seulement parallèlement au plan de l'image comme avec des éléments structurants binaires (plans).

³Notons que dans l'expression de la dualité entre dilatation et érosion, et entre ouverture et fermeture, l'élément structurant reste le même des deux côtés de l'égalité.

6.2.4 Ouverture et fermeture binaires

L'équation 6.7 montre, si on l'exprime pour $B = B'$, que la dilatation n'est pas l'inverse de l'érosion. Par exemple, si l'on effectue d'abord une érosion puis une dilatation, les composantes connexes de l'objet qui ont été supprimées par l'érosion (à cause de leur petite taille) ne peuvent plus être recouvertes par la dilatation et sont donc définitivement perdues. On construit donc ainsi une nouvelle transformation par composition d'une érosion et d'une dilatation, appelée ouverture.

Définition 11. Ouverture binaire : L'ouverture de l'ensemble X par l'élément structurant B est définie par :

$$X_B = D[E(X, B), \tilde{B}]. \quad (6.13)$$

Propriétés. L'ouverture a les propriétés algébriques suivantes :

- elle est anti-extensive ($X \supset X_B$)⁴,
- elle est croissante ($X \subset Y \Rightarrow X_B \subset Y_B$),
- elle est idempotente ($(X_B)_B = X_B$).

Ces propriétés sont fondamentales puisqu'elles font de l'ouverture un filtre morphologique.

On a de plus $B \subset B' \Rightarrow X_{B'} \subset X_B$, et si X_n désigne l'ouvert de X par un élément structurant de taille n , $(X_n)_{n'} = (X_{n'})_n = X_{\max(n, n')}$.

L'ouverture vérifie les 4 principes fondamentaux (comme composée de deux opérations les satisfaisant).

Exemples. L'ouverture a pour effet de supprimer les parties des objets plus petites que l'élément structurant, et de lisser les contours en supprimant les petites excroissances (trop fines pour pouvoir contenir l'élément structurant). C'est l'effet de filtrage décrit algébriquement ci-dessus. Elle ne réduit pas systématiquement toutes les structures comme le fait l'érosion.

La figure 6.6 illustre ces effets.

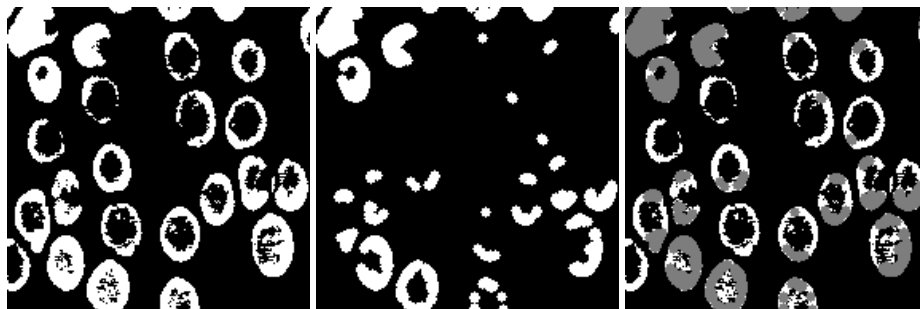


FIG. 6.6 – Exemple d'ouverture binaire (de gauche à droite : image initiale, ouverture par un disque de taille 3, différence : en blanc, les parties supprimées par l'ouverture).

Définition 12. Fermeture binaire : La fermeture de l'ensemble X par l'élément structurant B est définie par :

$$X^B = E[D(X, B), \tilde{B}]. \quad (6.14)$$

Propriétés. La fermeture a les propriétés algébriques suivantes :

- elle est extensive ($X \subset X^B$),
- elle est croissante,

⁴Contrairement à la dilatation et à l'érosion, cette propriété est vraie sans restriction sur l'élément structurant.

– elle est idempotente.

De même que pour l'ouverture, ces propriétés font de l'ouverture un filtre morphologique.

On a de plus $B \subset B' \Rightarrow X^B \subset X^{B'}$, et si X^n désigne le fermé de X par un élément structurant de taille n , $(X^n)^{n'} = (X^{n'})^n = X^{\max(n, n')}$.

La fermeture vérifie les 4 principes fondamentaux (comme composée de deux opérations les satisfaisant).

Enfin, la fermeture est l'opération duale de l'ouverture par rapport à la complémentation :

$$X^B = [(X^C)_B]^C, \quad (6.15)$$

ce qui permet de déduire les propriétés et l'action d'une des opérations de celles de l'autre.

Exemples. La fermeture a pour effet de boucher les trous des objets qui sont plus petits que l'élément structurant. Elle lisse les contours des objets en rajoutant des points dans les concavités étroites (dans lesquelles ne peut pas se glisser l'élément structurant). On retrouve l'effet de filtrage, dual de celui de l'ouverture, décrit par les trois propriétés algébriques ci-dessus.

La figure 6.7 illustre ces effets.

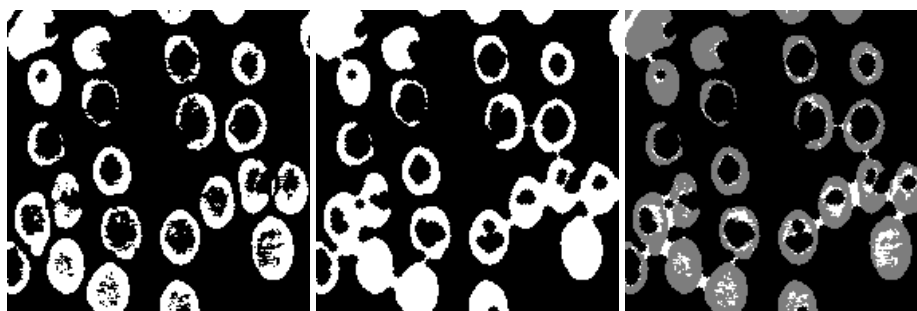


FIG. 6.7 – Exemple de fermeture binaire (de gauche à droite : image initiale, fermeture par un disque de taille 3, différence : en blanc, les parties rajoutées par la fermeture).

6.2.5 Ouverture et fermeture numériques

Définition 13. Ouverture numérique : (par un élément structurant binaire) L'ouverture d'une fonction f par un élément structurant B est définie comme dans le cas binaire par :

$$f_B = D[E(f, B), \check{B}]. \quad (6.16)$$

Propriétés. L'ouverture est croissante, anti-extensive et idempotente, ce qui lui confère une nature de filtre morphologique.

Exemples. L'ouverture sur une image à niveaux de gris a pour effet d'écarter les pics (si on considère le sous-graphe de la fonction comme un relief) qui sont plus étroits que l'élément structurant.

La figure 6.8 illustre ces effets.

Définition 14. Fermeture numérique : (par un élément structurant binaire) La fermeture d'une fonction f par un élément structurant B est définie comme dans le cas binaire par :

$$f^B = E[D(f, B), \check{B}]. \quad (6.17)$$



FIG. 6.8 – Exemple d’ouverture numérique (de gauche à droite : image initiale et ouverture par un disque de taille 3).

Propriétés. La fermeture est duale de l’ouverture. De plus, comme dans le cas binaire, elle est croissante, extensive et idempotente, ce qui lui confère une nature de filtre morphologique.

Exemples. La fermeture a l’effet dual de l’ouverture : elle comble les vallées qui sont plus étroites que l’élément structurant.

La figure 6.9 illustre ces effets.



FIG. 6.9 – Exemple de fermeture numérique (de gauche à droite : image initiale et fermeture par un disque de taille 3).

6.3 Cadre topologique

Nous avons vu plus haut les fondements ensemblistes de la morphologie mathématique. Celle-ci repose ensuite sur des concepts de topologie. Dans cette approche, on cherche à définir le voisinage non plus d’un point mais d’un ensemble. Deux types de topologies sont utilisées en morphologie mathématique : la topologie en tout ou rien sur les fermés et la topologie myope définie sur les compacts. La topologie myope a l’avantage d’être reliée à la distance de Hausdorff, métrique bien adaptée à la comparaison de formes dans de nombreuses applications.

Dans la suite, on note \mathcal{F} l'ensemble des fermés de \mathbb{R}^n , \mathcal{G} l'ensemble de ses ouverts, et \mathcal{K} l'ensemble de ses compacts, pour la topologie usuelle de \mathbb{R}^n .

Définition 15. Topologie en tout ou rien : La topologie en tout ou rien sur \mathcal{F} est engendrée par la famille :

$$\mathcal{F}_{G_1 \dots G_p}^K = \{F \in \mathcal{F}, F \cap K = \emptyset, \forall i, 1 \leq i \leq p, F \cap G_i \neq \emptyset\},$$

où $p \in \mathbb{N}$, $K \in \mathcal{K}$ et $G_i \in \mathcal{G}$.

La topologie en tout ou rien est donc engendrée par les fermés qui, étant donnés un compact et une famille d'ouverts, n'intersectent pas le compact et intersectent chacun des ouverts.

La convergence pour la topologie en tout ou rien est définie de la manière suivante :

Définition 16. Convergence dans \mathcal{F} : Une suite de fermés $(F_n)_{n \in \mathbb{N}}$ converge vers $F \in \mathcal{F}$ si pour tout ouvert G qui intersecte F et tout compact K qui n'intersecte pas F , on a :

$$\exists N, \forall n \geq N, G \cap F_n \neq \emptyset, K \cap F_n = \emptyset.$$

On montre que la réunion de $\mathcal{F} \times \mathcal{F}$ dans \mathcal{F} est continue pour cette définition mais que l'intersection ne l'est pas. Cela nous conduit à introduire les notions de semi-continuité.

Définition 17. Limites inférieure et supérieure : La limite inférieure d'une suite de fermés $(F_n)_{n \in \mathbb{N}}$, notée $\underline{\lim}(F_n)$ est l'intersection des points d'adhérence de la suite, et sa limite supérieure, notée $\overline{\lim}(F_n)$ est la réunion des points d'adhérence.

Définition 18. Semi-continuité : Une application f d'un ensemble Ω dans \mathcal{F} est semi-continue supérieurement (s.c.s.) si pour tout ω de Ω et toute suite (ω_n) de Ω convergeant vers ω :

$$\overline{\lim} f(\omega_n) \subset f(\omega).$$

De même, f est semi-continue inférieurement (s.c.i.) si :

$$\underline{\lim} f(\omega_n) \supset f(\omega).$$

Notons qu'une application est continue ssi elle est à la fois s.c.s. et s.c.i.

L'intersection de $\mathcal{F} \times \mathcal{F}$ dans \mathcal{F} est s.c.s. Ces propriétés de l'intersection et de la réunion seront directement utilisées pour en déduire des propriétés des transformations de morphologie mathématique construites selon l'approche ensembliste. On a ainsi les résultats suivants :

- la dilatation d'un fermé par un compact est continue (le résultat est un fermé), c'est-à-dire à la fois s.c.s. et s.c.i.,
- la dilatation d'un compact par un compact est continue (le résultat est un compact),
- l'application de $\mathcal{F} \times \mathcal{K}$ dans \mathcal{F} qui à (F, K) associe $E(F, K)$ est s.c.s.,
- de même, l'application de $\mathcal{K} \times \mathcal{K}$ dans \mathcal{K} qui à (K', K) associe $E(K', K)$ est s.c.s.,
- l'application de $\mathcal{F} \times \mathcal{K}$ dans \mathcal{F} qui à (F, K) associe F_K est s.c.s.,
- de même, l'application de $\mathcal{K} \times \mathcal{K}$ dans \mathcal{K} qui à (K', K) associe K'_K est s.c.s.,
- l'application de $\mathcal{F} \times \mathcal{K}$ dans \mathcal{F} qui à (F, K) associe F^K est s.c.s.,
- de même, l'application de $\mathcal{K} \times \mathcal{K}$ dans \mathcal{K} qui à (K', K) associe K'^K est s.c.s.

Ces propriétés fortes de la dilatation et de l'érosion en termes de continuité lorsque l'élément structurant est compact incitent à choisir pour les applications des éléments structurants compacts.

Définition 19. Topologie myope : La topologie myope sur \mathcal{K} est engendrée par la famille :

$$\mathcal{K}_{G_1 \dots G_p}^F = \{K \in \mathcal{K}, K \cap F = \emptyset, \forall i, 1 \leq i \leq p, K \cap G_i \neq \emptyset\},$$

où $p \in \mathbb{N}$, $F \in \mathcal{F}$ et $G_i \in \mathcal{G}$.

La topologie myope est plus fine que la topologie induite sur \mathcal{K} par la topologie en tout ou rien sur \mathcal{F} , et équivalente sur $\mathcal{K} \setminus \emptyset$ à la topologie induite par la distance de Hausdorff.

Définition 20. Distance de Hausdorff : La distance de Hausdorff entre deux compacts non vides K et K' est définie par :

$$\delta(K, K') = \max\left\{\sup_{x \in K} d(x, K'), \sup_{x' \in K'} d(x', K)\right\},$$

où $d(x, K)$ est la distance classique d'un point à un compact ($d(x, K) = \inf_{y \in K} d(x, y)$ où $d(x, y)$ est la distance euclidienne classique sur \mathbb{R}^n).

Cette distance est souvent utilisée pour comparer deux ensembles et estimer leur proximité. Notons que cette distance est très sensible aux déformations qui peuvent se produire à la surface des objets⁵.

6.4 Cadre algébrique

Le cadre algébrique général de la morphologie mathématique est celui des treillis et de la notion fondamentale d'adjonction [Heijmans et Ronse, 1990, Ronse et Heijmans, 1991, Heijmans, 1991].

6.4.1 Treillis

Définition 21. Treillis : Un treillis est un ensemble \mathcal{T} muni d'une relation d'ordre \leq , tel que toute paire d'éléments (ou toute famille finie) possède une borne supérieure (notée $x \vee y$) et une borne inférieure (notée $x \wedge y$). Un treillis est complet si toute famille d'éléments (finie ou non) possède un plus petit majorant et un plus grand minorant.

Un treillis complet contient en particulier un plus petit élément et un plus grand élément.

L'ensemble des fonctions de \mathbb{R}^n dans $\overline{\mathbb{R}}$ est un treillis complet, pour la relation d'ordre définie par :

$$f \leq g \Leftrightarrow \forall x \in \mathbb{R}^n, f(x) \leq g(x).$$

La semi-continuité pour des fonctions est définie de la manière suivante.

Définition 22. Semi-continuité d'une fonction : Une fonction f est semi-continue supérieurement en x si :

$$\forall t > f(x), \exists V(x), \forall y \in V(x), t > f(y),$$

où $V(x)$ est un voisinage de x dans \mathbb{R}^n . De même, f est semi-continue inférieurement si :

$$\forall t < f(x), \exists V(x), \forall y \in V(x), t < f(y).$$

Définition 23. Sous-graphe : À toute fonction f de \mathbb{R}^n dans $\overline{\mathbb{R}}$, on associe son sous-graphe

$$SG(f) = \{(x, t) \in \mathbb{R}^n \times \overline{\mathbb{R}}, f(x) \geq t\}.$$

Remarquons qu'une fonction est s.c.s. ssi son sous-graphe est fermé.

L'ensemble des fonctions s.c.s. de \mathbb{R}^n dans $\overline{\mathbb{R}}$ est un treillis complet, pour la relation d'ordre définie par :

$$f \leq g \Leftrightarrow SG(f) \subset SG(g).$$

⁵Le problème de la définition d'une distance entre objets est réputé difficile et la plupart des distances utilisées, comme la distance de Hausdorff, ne donnent qu'une information partielle sur la ressemblance des objets. La distance de Hausdorff est cependant bien adaptée à une grande classe de problèmes, en particulier grâce à ses liens étroits avec la morphologie mathématique.

Cette propriété montre l'intérêt du cadre fourni par les treillis pour la morphologie mathématique. Un autre intérêt vient du fait que les opérations de morphologie mathématique sont fondées sur des \vee et des \wedge comme nous l'avons vu dans la partie 6.2.

La notion de sous-graphe permet de faire un lien direct avec l'approche ensembliste d'une part puisque les sous-graphes sont des ensembles de $\mathbb{R}^n \times \overline{\mathbb{R}}$, et avec l'approche topologique d'autre part puisque la topologie utilisée sur l'espace des fonctions s.c.s. est celle induite par la topologie en tout ou rien sur $\mathcal{F}(\mathbb{R}^n \times \overline{\mathbb{R}})$.

6.4.2 Erosion et dilatation algébriques, adjonctions

Définition 24. Dans un treillis complet (\mathcal{T}, \leq) , une **dilatation algébrique** est définie de manière générale comme une fonction δ sur \mathcal{T} qui commute avec le sup, c'est-à-dire telle que :

$$\forall (x_i) \in \mathcal{T}, \delta(\vee_i x_i) = \vee_i \delta(x_i), \quad (6.18)$$

où (x_i) est une famille (éventuellement vide) d'éléments de \mathcal{T} .

De manière similaire, une **érosion algébrique** est une fonction ε sur \mathcal{T} qui commute avec l'inf, c'est-à-dire telle que :

$$\forall (x_i) \in \mathcal{T}, \varepsilon(\wedge_i x_i) = \wedge_i \varepsilon(x_i). \quad (6.19)$$

Notons que cette définition suffit à garantir que les dilatations et érosions algébriques sont croissantes.

Une propriété fondamentale dans le cadre des treillis est celle d'adjonction.

Définition 25. Une paire d'opérateurs (ε, δ) sur \mathcal{T} est une **adjonction** si et seulement si :

$$\forall (x, y) \in \mathcal{T}^2, \delta(x) \leq y \Leftrightarrow x \leq \varepsilon(y). \quad (6.20)$$

Propriétés. On a les relations suivantes entre opérations algébriques et adjonctions :

- si (ε, δ) est une adjonction, alors ε est une érosion algébrique et δ est une dilatation algébrique ;
- un opérateur croissant δ est une dilatation algébrique si et seulement si il existe un opérateur ε tel que (ε, δ) soit une adjonction ; ε est alors une érosion algébrique et vaut :

$$\varepsilon(x) = \vee \{y \in \mathcal{T}, \delta(y) \leq x\};$$

- un opérateur croissant ε est une érosion algébrique si et seulement si il existe un opérateur δ tel que (ε, δ) soit une adjonction ; δ est alors une dilatation algébrique et vaut :

$$\delta(x) = \wedge \{y \in \mathcal{T}, \varepsilon(y) \geq x\};$$

6.4.3 Lien avec les opérateurs morphologiques

Sur le treillis booléen des parties de \mathbb{R}^n ou \mathbb{Z}^n (la relation d'ordre est alors l'inclusion), le dilaté de n 'importe quelle partie peut être exprimé comme la réunion des dilatés des singletons qui la composent :

$$\delta(X) = \cup_{x \in X} \delta(\{x\}).$$

De plus, toute dilatation invariante par translation peut s'exprimer comme une dilatation morphologique, définie à partir d'un élément structurant comme dans la section 6.2.

Des résultats similaires sont valables pour les érosions.

6.4.4 Ouverture et fermeture algébriques

Définition 26. Une opération croissante est une **ouverture algébrique** si et seulement si elle est anti-extensive et idempotente, et une **fermeture algébrique** si et seulement si elle est extensive et idempotente.

Si I désigne l'identité sur le treillis, on a pour toute adjonction (ε, δ) :

$$\delta\varepsilon \leq I \leq \varepsilon\delta. \quad (6.21)$$

De plus, $\delta\varepsilon$ et $\varepsilon\delta$ sont toujours idempotentes, ce sont donc des exemples d'ouverture et fermeture algébriques respectivement, appelées ouverture et fermeture morphologiques.

6.5 Cadre probabiliste

6.5.1 Ensembles fermés aléatoires

Enfin, une partie de la morphologie mathématique repose sur une approche probabiliste, en particulier sur le concept d'ensemble fermé aléatoire. Le principe consiste à considérer les équations ensemblistes utilisées dans la première approche comme des événements et à leur affecter une probabilité. Ces événements seront donc du type $\{\{B \subset A\}, \{\{B \cap A \neq \emptyset\}\}$. L'ensemble des valeurs (ou réalisations) considéré ici est l'ensemble des fermés \mathcal{F} . L'ensemble des événements est défini par la tribu borélienne sur \mathcal{F} engendrée par les événements particuliers qui nous intéressent, appelée encore tribu morphologique.

Définition 27. Tribu morphologique : La tribu morphologique est la tribu σ_f sur \mathcal{F} engendrée par les ouverts de la topologie en tout ou rien sur \mathcal{F} , c'est-à-dire par les \mathcal{F}^K et les \mathcal{F}_G pour tous les compacts K de \mathcal{K} et tous les ouverts G de \mathcal{G} :

$$\mathcal{F}^K = \{F \in \mathcal{F}, K \cap F = \emptyset\}, \quad \mathcal{F}_G = \{F \in \mathcal{F}, G \cap F \neq \emptyset\}.$$

La donnée d'une probabilité sur l'espace mesurable (\mathcal{F}, σ_f) conduit alors à la notion d'ensemble fermé aléatoire.

Définition 28. Ensemble fermé aléatoire : Un ensemble fermé aléatoire F , noté $F = (\mathcal{F}, \sigma_f, P)$ est déterminé par la donnée d'une probabilité P sur (\mathcal{F}, σ_f) .

Définition 29. Capacité de Choquet : La capacité de Choquet d'un ensemble fermé aléatoire $F = (\mathcal{F}, \sigma_f, P)$ est la fonctionnelle $T : \mathcal{K} \rightarrow [0, 1]$ définie par :

$$\forall K \in \mathcal{K}, \quad T(K) = P(F \cap K \neq \emptyset).$$

La capacité de Choquet joue un rôle analogue à celui de la fonction de répartition d'une probabilité sur \mathbb{R} . On montre qu'un ensemble fermé aléatoire est entièrement caractérisé par sa capacité de Choquet. En pratique, cela fournit un moyen opératoire de caractérisation d'un phénomène aléatoire, puisque la connaissance des $P(F \cap K \neq \emptyset)$ est suffisante, ce qui limite les mesures à effectuer.

Les applications de ces concepts concernent des mesures de phénomènes aléatoires, comme par exemple des schémas booléens, ou l'analyse de textures.

6.5.2 Un exemple : le schéma booléen

Un schéma booléen stationnaire est défini comme la réunion de compacts aléatoires de même loi translatés aux points d'un processus de Poisson [Schmitt et Mattioli, 1994a]. Sa capacité de Choquet est une fonction de la densité du processus de Poisson sous-jacent et de la moyenne du volume du dilaté du compact aléatoire.

Soit un processus de Poisson \mathcal{P}_θ d'intensité θ ($\theta > 0$). Pour tout compact K , la probabilité pour que n points du processus tombent dans K vaut :

$$\frac{\lambda^n}{n!} e^{-\lambda} \quad (6.22)$$

avec $\lambda = \theta V(K)$, où $V(K)$ désigne le volume de K . De plus, si deux compacts K_1 et K_2 sont disjoints, alors les variables aléatoires qui comptent les points tombant dans K_1 et dans K_2 sont indépendantes (voir chapitre 4, section 4.5).

On appelle grain primaire un ensemble fermé aléatoire $F' = (\mathcal{F}, \sigma_f, P')$. Le schéma booléen $F = (\theta, F')$ d'intensité θ et de grain primaire F' est alors défini par :

$$F = \cup\{F'_x, x \in \mathcal{P}_\theta\} \quad (6.23)$$

où les F'_x sont des compacts aléatoires indépendants de même loi P' et translatés aux points du processus de Poisson. On garantit que cette réunion est un fermé en imposant que le volume moyen de $(F' \oplus B)$ est fini, pour B égal à la boule unité. Un tel schéma booléen est stable par réunion.

On a alors la caractérisation suivante en fonction de la capacité de Choquet :

$$1 - T(K) = P(K \cap F = \emptyset) = e^{-\theta \overline{V(F' \oplus \tilde{K})}} \quad (6.24)$$

où \overline{V} désigne le volume moyen. Cette caractérisation permet de définir des procédures d'inférence des paramètres du modèle, à une translation près [Schmitt et Mattioli, 1994a].

6.6 Applications de l'érosion et de la dilatation

6.6.1 Mesures

Nous donnons dans cette partie succinctement quelques exemples de transformations qui associent à un objet une mesure, que nous répartissons dans deux classes.

La première classe est issue de la géométrie intégrale et vise à calculer des longueurs, des surfaces, des volumes, des intégrales de courbure moyenne, par l'intermédiaire des fonctionnelles de Minkowski. Le lien avec la morphologie mathématique est fourni par les formules de Steiner, qui permettent de calculer les fonctionnelles de Minkowski du dilaté d'un compact d'après celles du compact [Hadwiger, 1957, Santalo, 1976].

La deuxième classe vise des applications de reconnaissance des formes et étudie le comportement d'objets vis-à-vis de transformations morphologiques appropriées aux caractéristiques à mettre en évidence [Schmitt et Mattioli, 1994a]. Par exemple, les courbes granulométriques (donnant la surface des objets sélectionnés par la transformation ϕ_λ quand λ varie) permettent d'étudier la distribution de tailles des objets. Autre exemple, le covariogramme géométrique, donnant la surface de $E(X, B_r)$ (où B_r est un segment de longueur r dans une direction α) quand r varie, permet d'observer le comportement des objets quand on l'érode dans une direction. Il peut en particulier discriminer des ensembles d'objets de même surface globale mais dont la répartition des composantes connexes en taille et dans l'espace varie.

6.6.2 Erodé ultime

L'érodé ultime est la réunion de toutes les composantes d'un objet binaire qui disparaissent d'une érosion à l'autre dans une séquence d'érosions par un élément structurant élémentaire B . Plus formellement, l'érodé ultime d'un ensemble X est défini par :

$$EU(X) = \cup_n \{E(X, B_n) \setminus R[E(X, B_{n+1}); E(X, B_n)]\} \quad (6.25)$$

où $E(X, B_n)$ désigne l'érodé de X de taille n et $R[Y; Z]$ désigne les composantes connexes de Z qui ont une intersection non vide avec Y .

L'érodé ultime d'un ensemble X est exactement l'ensemble des maxima régionaux de la fonction distance à l'intérieur de X (distance des points de X à X^C).

6.6.3 Rehaussement de contraste

Le rehaussement de contraste morphologique d'une fonction f est défini à partir d'une fonction minorante \underline{f} et d'une fonction majorante \overline{f} et de deux paramètres α et β tels que $\alpha \geq 0$, $\beta \geq 0$ et $\alpha + \beta < 1$. Le résultat g de cette transformation est obtenu en faisant basculer les points vers la fonction minorante ou vers la fonction majorante suivant la règle suivante :

$$\begin{aligned} g(x) &= \underline{f}(x) & \text{si } \underline{f}(x) \leq f(x) \leq \underline{f}(x) + \alpha \Delta f(x) \\ g(x) &= f(x) & \text{si } \underline{f}(x) + \alpha \Delta f(x) \leq f(x) \leq \overline{f}(x) - \beta \Delta f(x) \\ g(x) &= \overline{f}(x) & \text{si } \overline{f}(x) - \beta \Delta f(x) \leq f(x) \leq \overline{f}(x) \end{aligned}$$

avec $\Delta f(x) = \overline{f}(x) - \underline{f}(x)$.

La morphologie mathématique fournit naturellement des fonctions minorantes (resp. majorantes) à partir de transformations anti-extensives (resp. extensives) comme l'érosion de f par un élément structurant centré, ou encore l'ouverture de f (resp. dilatation ou fermeture).

La figure 6.10 donne un exemple de rehaussement de contraste à partir des dilatation et érosion des figures 6.4 et 6.5.



FIG. 6.10 – Exemple de rehaussement de contraste morphologique.

6.6.4 Gradient morphologique

Soit B le disque fermé de rayon unité. Le gradient morphologique d'une fonction f est défini dans le cas continu par la fonction g suivante :

$$g(x) = \lim_{\lambda \rightarrow 0} \frac{D(f, B_\lambda)(x) - E(f, B_\lambda)(x)}{2\lambda},$$

et dans le cas discret par :

$$g(x) = D(f, B)(x) - E(f, B)(x).$$

Cette transformation trouve ses applications dans la détection de contours.

La figure 6.11 donne un exemple de gradient obtenu par différence des dilatation et érosion de taille 1.



FIG. 6.11 – Exemple de gradient morphologique.

6.7 Applications de l'ouverture et de la fermeture

6.7.1 Filtres alternés séquentiels

Nous avons vu que les ouvertures et fermetures étaient des filtres morphologiques. Sans entrer dans la théorie des filtres morphologiques, nous ne décrivons ici que les filtres alternés séquentiels, qui sont beaucoup utilisés en pratique, et qui sont construits à partir de suites d'ouvertures et de fermetures de tailles croissantes. Dans le cas discret, un tel filtre appliqué à une fonction f s'exprime comme :

$$(\dots(((f_{B_1})^{B_1})_{B_2})^{B_2})\dots_{B_n})^{B_n}.$$

Notons qu'on obtient bien ainsi des opérations croissantes et idempotentes (donc des filtres morphologiques).

Ils sont utilisés en pratique pour filtrer progressivement le bruit positif (pics étroits) et le bruit négatif (vallées étroites). Le dernier élément structurant utilisé (de taille n) est déterminé en fonction de la taille minimale des objets de l'image que l'on veut conserver après le filtrage.

La figure 6.12 donne des exemples de filtres alternés séquentiels.

6.7.2 Filtres auto-duaux

Les filtres auto-duaux sont des transformations qui agissent sur les fonctions indépendamment du contraste local et traitent donc les parties claires et les parties sombres de la même manière. Le filtre médian est un exemple de tel filtre.

Un exemple est le centre morphologique, défini à partir d'une famille d'opérateurs $\{\psi_1, \psi_2, \dots, \psi_n\}$ par :

$$(I \vee \wedge_i \psi_i) \wedge \vee_i \psi_i \quad (6.26)$$



FIG. 6.12 – Filtres alternés séquentiels : image initiale, tailles maximales de 1, 2 et 3.

où I désigne l'identité.

Les travaux de H. Heijmans se sont attachés à formaliser cette notion de filtres auto-duaux dans un cadre algébrique.

6.7.3 Chapeau haut-de-forme

La transformation du $\hat{\hat{}}$ chapeau haut-de-forme $\hat{\hat{}}$ d'une fonction f est définie, aussi bien en continu qu'en discret, comme la fonction :

$$f - f_B$$

pour un élément structurant B donné. L'opération duale est la fonction :

$$f^B - f.$$

Cette transformation extrait les pics étroits (plus étroits que l'élément structurant) quelle que soit leur hauteur absolue. Elle permet par exemple d'extraire des lignes fines de niveau intense par rapport à leur voisinage (telles que des routes dans une image satellitaire). L'opération duale extrait au contraire des vallées étroites.

La figure 6.13 donne un exemple de chapeau haut de forme obtenu pour une ouverture de taille 3.

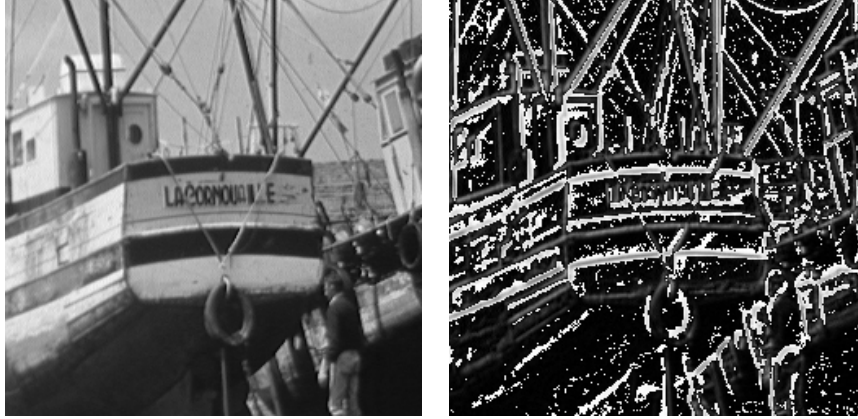


FIG. 6.13 – Exemple de chapeau haut de forme obtenu pour une ouverture de taille 3.

Si l'on applique cette transformation avec des éléments structurants fonctionnels, par exemple des sphères de \mathbb{R}^{n+1} , on sélectionne alors les contours abrupts dans l'image, ou encore les pics de courbure plus forte que l'élément structurant.

6.7.4 Granulométries

Les granulométries sont des opérations de \downarrow tamisage \downarrow_λ qui servent à sélectionner successivement des particules de tailles données croissantes.

Définition 30. Granulométrie : Une granulométrie sur un ensemble \mathcal{A} de parties de \mathbb{R}^n est une famille de fonctions paramétrées ϕ_λ (avec $\lambda > 0$) définies sur \mathcal{A} telle que :

1. $\forall X \in \mathcal{A}, \forall \lambda > 0, \phi_\lambda(X) \subset X$ (ϕ_λ anti-extensive),
2. $\forall (X, Y) \in \mathcal{A}^2, \forall \lambda > 0, X \subset Y \Rightarrow \phi_\lambda(X) \subset \phi_\lambda(Y)$ (ϕ_λ croissante),
3. $\forall X \in \mathcal{A}, \forall \lambda > 0, \forall \mu > 0, \lambda \geq \mu \Rightarrow \phi_\lambda(X) \subset \phi_\mu(X)$ (ϕ_λ décroissante par rapport au paramètre),
4. $\forall \lambda > 0, \forall \mu > 0, \phi_\lambda \circ \phi_\mu = \phi_\mu \circ \phi_\lambda = \phi_{\max(\lambda, \mu)}$.

Il est clair, d'après cette définition, que la famille d'ouvertures par des boules de rayon r définit une granulométrie. On montre même que ϕ_λ est une granulométrie si et seulement si ϕ_λ est une ouverture pour tout λ et la classe des ensembles de \mathcal{A} invariants par ϕ_λ est incluse dans celle des invariants par ϕ_μ pour $\lambda \geq \mu$.

Ainsi, si on applique à un ensemble une suite d'ouvertures de tailles croissantes (par des boules), on sélectionnera d'abord les plus petites parties de l'ensemble (celles qui sont supprimées par l'ouverture), puis des parties de plus en plus grosses.

La figure 6.14 montre une courbe de granulométrie obtenue sur une image binaire par ouvertures de différentes tailles. Lorsque la taille de l'ouverture correspond à la taille caractéristique de la plupart des objets, un saut apparaît dans la courbe.

6.7.5 Ouverture surfacique

Les ouvertures peuvent combiner un critère sur un attribut, par exemple la taille des objets, permettant ainsi de filtrer des petits objets ou des petites parties claires. Une ouverture surfacique de paramètre λ est ainsi définie par :

$$\gamma_\lambda = \bigvee_i \{ \gamma_{B_i}, B_i \text{ est connexe et } S(B_i) = \lambda \} \quad (6.27)$$

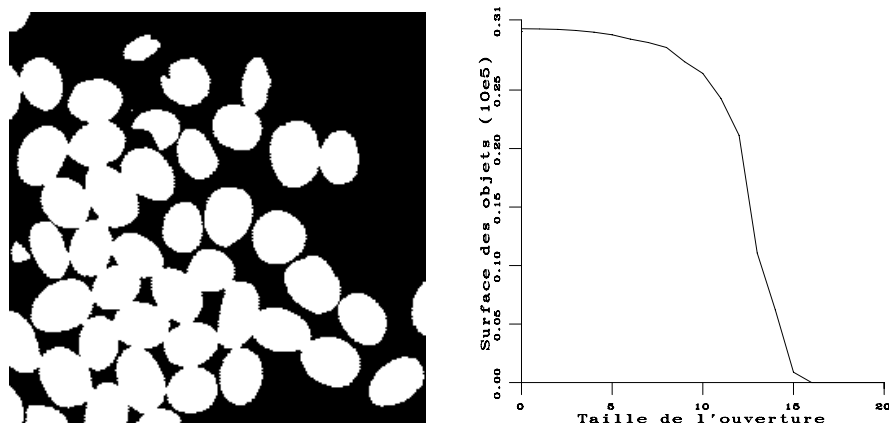


FIG. 6.14 – Image binaire et sa courbe de granulométrie.

où γ_{B_i} désigne une ouverture par B_i et $S(B_i)$ est la surface de B_i (ou son volume en 3D).

On définit bien sûr de manière duale des fermetures surfaciques.

6.7.6 Ouverture annulaire

L'ouverture annulaire est définie comme le min entre l'image originale et son dilaté par un élément structurant en forme d'anneau. Puisque l'élément structurant ne contient alors pas l'origine, la propriété d'extensivité de la dilatation n'est pas vérifiée. Cette opération sert par exemple à récupérer des groupes serrés d'objets dans une image en éliminant des parties isolées (la dilatation de telles parties ne les contient pas).

6.8 Transformation en tout ou rien et opérateurs dérivés

6.8.1 Transformation en tout ou rien

Les transformations vues jusqu'à présent sont toutes fondées sur le même principe : elles examinent si une certaine configuration de points (définie par l'élément structurant) vérifie une relation avec l'objet étudié. La transformation en tout ou rien propose d'examiner des configurations où certains points vérifient une relation avec l'objet et d'autres vérifient une relation avec le complémentaire de l'objet. Ainsi les éléments structurants T considérés dans cette transformation sont décomposés en deux parties T_1 et T_2 (de même origine).

Définition 31. Transformation en tout ou rien : La transformation en tout ou rien de X par l'élément structurant $T = (T_1, T_2)$ est définie par :

$$X \otimes T = E(X, T_1) \cap E(X^C, T_2).$$

Notons que puisque $T_1 \cap T_2 = \emptyset$, l'origine ne peut pas appartenir à la fois à T_1 et à T_2 , donc on ne peut pas avoir simultanément $E(X, T_1) \subset X$ et $E(X^C, T_2) \subset X^C$ (ce qui est heureux...).

6.8.2 Amincissement et épaissement

Deux nouvelles transformations sont déduites de la transformation en tout ou rien :

Définition 32. Amincissement et épaissement : L'aminçissement de X par l'élément structurant $T = (T_1, T_2)$ est défini par :

$$X \circ T = X \setminus X \otimes T, \quad (6.28)$$

et l'épaississement par :

$$X \odot T = X \cup X \otimes T. \quad (6.29)$$

L'aminçissement, opération anti-extensive, supprime donc de X des points qui sont dans une configuration particulière, sélectionnés par la transformation en tout ou rien. Les éléments structurants appropriés sont ceux pour lesquels l'origine appartient à T_1 . L'épaississement rajoute au contraire à X les points sélectionnés par la transformation en tout ou rien et est une opération extensive. Les éléments structurants appropriés sont ceux pour lesquels l'origine appartient à T_2 . Ces deux transformations sont duales l'une de l'autre au sens suivant :

$$X \circ T = (X^C \odot T')^C$$

avec $T' = (T_2, T_1)$ si $T = (T_1, T_2)$.

Il existe tout un jargon d'éléments structurants, appelé alphabet de Golay en trame hexagonale, qui permet d'effectuer une grande variété d'opérations [Serra, 1982a]. Par exemple :

- si T_1 est l'élément structurant élémentaire sur la trame considérée et T_2 est vide, la transformation en tout ou rien est une érosion et l'aminçissement correspondant extrait la frontière des objets ;
- si les configurations recherchées par la transformation en tout ou rien correspondent à un point du complémentaire entouré de points de l'objet formant une configuration localement non convexe, l'épaississement correspondant (appliqué itérativement jusqu'à convergence) permet de calculer le plus petit convexe discret contenant l'objet de départ ;
- si T_1 est le point central de l'élément structurant élémentaire et T_2 contient les autres, la transformation en tout ou rien sélectionne les points isolés de l'objet ;
- etc.

Nous verrons dans la suite que des configurations particulières permettent d'extraire le squelette d'un objet par aminçissement.

6.8.3 Squelette

Le squelette est une notion très utile en reconnaissance de formes : en effet, il s'agit d'une représentation des objets, qu'on souhaite intuitivement compacte (lignes fines centrées dans les objets), homotopique à l'objet de départ (même nombre de composantes connexes et même nombre de trous), et inversible. En morphologie mathématique, on arrive à une définition du squelette dans le cas continu qui a de bonnes propriétés par rapport à ces exigences. En revanche, le passage au discret est très délicat pour cette transformation.

Dans le cas continu, le squelette est défini comme l'ensemble des centres des boules maximales incluses dans l'objet (une boule étant maximale si elle ne peut être incluse dans aucune autre boule incluse dans l'objet). On travaille généralement sur les fermés en morphologie mathématique. Cependant, dans le cas du squelette, cela conduit à de mauvaises propriétés : le squelette d'un ensemble connexe n'est pas nécessairement connexe, il n'est pas nécessairement fermé, et l'application qui à un fermé associe l'adhérence de son squelette n'est ni s.c.s., ni s.c.i. On travaillera donc, pour le cas du squelette, sur des ouverts.

Définition 33. Squelette : Le squelette d'un ouvert A est l'ensemble des centres des boules ouvertes maximales incluses dans A .

En appelant $s_\rho(A)$ l'ensemble des centres des boules ouvertes maximales de rayon donné $\rho > 0$, on obtient le squelette $r(A)$ comme réunion de tels ensembles :

$$r(A) = \bigcup_{\rho > 0} s_\rho(A).$$

On peut montrer que :

$$s_\rho(A) = \bigcap_{\mu > 0} [E(A, B_\rho) \setminus [E(A, B_\rho)]_{\bar{B}_\mu}]$$

où B_ρ (resp. \bar{B}_ρ) désigne la boule ouverte (resp. fermée) de rayon ρ . Cela fournit une caractérisation du squelette de A (et un moyen opératoire pour le calculer) à partir de ses érodés et des ouverts de ses érodés :

$$r(A) = \bigcup_{\rho > 0} \bigcap_{\mu > 0} [E(A, B_\rho) \setminus [E(A, B_\rho)]_{\bar{B}_\mu}]. \quad (6.30)$$

L'objet de départ peut être reconstruit à partir de son squelette (et donc la propriété d'inversibilité est satisfaite) :

$$A = \bigcup_{\rho > 0} D(s_\rho, B_\rho).$$

Pour la définition 33, on montre de plus que l'application qui à A associe l'adhérence de son squelette $\bar{r}(A)$ est s.c.i. de \mathcal{G} dans \mathcal{F} , que si A est connexe alors $\bar{r}(A)$ est connexe, et que le squelette est \mathcal{I} fin \mathcal{I} au sens suivant : l'intérieur du squelette est vide. Le squelette ainsi défini vérifie donc bien les propriétés qu'on en attend intuitivement.

6.8.4 Squelette discret

Si l'on transpose maintenant directement ces notions au cas discret, le squelette s'exprime (en discrétisant l'équation 6.30) comme :

$$S(X) = \bigcup_{n \in \mathbb{N}^*} [E(X, B_n) \setminus E(X, B_n)_B], \quad (6.31)$$

où B est l'élément structurant élémentaire sur la trame. Cette définition, que l'on trouve aussi sous le nom d'axe médian, correspond aux centres des boules maximales (discrètes) incluses dans l'objet.

Si la propriété d'inversibilité est conservée avec l'équation 6.31, les autres sont perdues, en particulier la préservation de la connexité. Deux classes de méthodes ont alors été proposées pour résoudre ce problème : la première tente de relier entre eux les points détectés par l'équation 6.31 pour en faire une \mathcal{I} ligne médiane \mathcal{I} qui satisfasse les propriétés de connexité ; la deuxième tente de trouver d'autres définitions, directement discrètes, en supprimant itérativement des points à la surface de l'objet.

La méthode que propose la morphologie mathématique relève de cette deuxième approche. Elle s'appuie sur la notion d'amincissement (équation 6.28), en utilisant des éléments structurants qui suppriment des points de l'objet sans changer sa topologie (voir figure 6.15). Il s'agit donc d'un amincissement homotopique. On arrive ainsi à un squelette qui vérifie les propriétés de connexité, qui donne une bonne représentation des formes, mais qui n'est plus inversible (il ne contient pas nécessairement les centres des boules maximales) et qui n'est pas nécessairement d'épaisseur unité.

La figure 6.16 illustre la différence entre les centres des boules maximales et le squelette par amincissement dans le cas discret.

6.8.5 SKIZ

Considérons maintenant le complémentaire de l'objet X . On note X_i les composantes connexes de l'objet ($X = \bigcup_i X_i$). Dans X^C , on appelle zone d'influence de X_i l'ensemble des points $ZI(X_i)$ plus proches de X_i que des X_j pour $j \neq i$:

$$ZI(X_i) = \{x \in X^C / d(x, X_i) < d(x, X \setminus X_i)\}.$$

Cette notion permet de définir celle de squelette par zones d'influence (SKIZ) :

$$\begin{array}{ccc} & 1 & 1 \\ & \cdot & 1 & \cdot \\ 0 & & 0 & \end{array}$$

FIG. 6.15 – Élément structurant utilisé en 6-connexité pour obtenir le squelette d'un objet par amincissement : cet élément structurant, ainsi que les 5 qui en sont déduits par rotations, sont appliqués itérativement jusqu'à convergence.

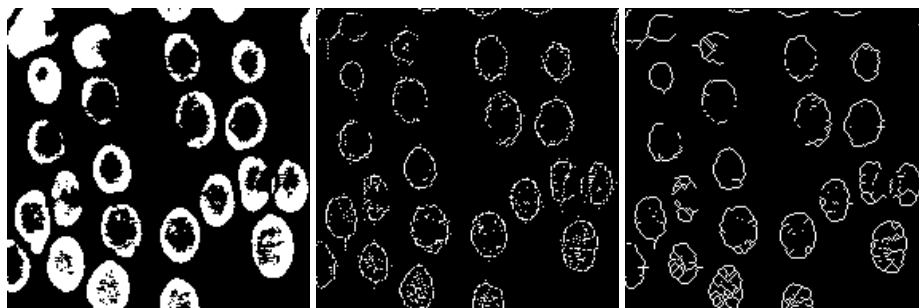


FIG. 6.16 – Centres des boules maximales (au centre) et squelette par amincissement (à droite).

Définition 34. Squelette par zones d'influence : *Le squelette par zone d'influence de X est l'ensemble des points qui n'appartiennent à aucune des zones d'influence :*

$$Skiz(X) = \left(\bigcup_i ZI(X_i) \right)^C.$$

Le squelette par zones d'influence est un sous-ensemble du squelette de X^C , il n'est pas forcément connexe et contient en général moins de barbules que le squelette de X^C (ce qui est souvent exploité dans les applications).

6.9 Géodésie

6.9.1 Distance géodésique et boules géodésiques

Les transformations géodésiques sont celles qui sont contraintes par un ensemble donné (on les appelle aussi conditionnelles). Elles sont fortement liées à la notion de distance géodésique : la distance géodésique d_X conditionnellement à un ensemble X entre deux points de X est la longueur du plus court chemin joignant ces deux points en restant dans l'ensemble.

On considère alors des éléments structurants qui sont des boules de la distance géodésique.

Définition 35. Boule géodésique : *La boule géodésique de centre x et de rayon r conditionnellement au sous-ensemble X de \mathbb{R}^n est définie par :*

$$B_X(x, r) = \{y \in X / d_X(x, y) \leq r\}. \quad (6.32)$$

6.9.2 Opérations morphologiques géodésiques, reconstruction

Définition 36. Dilatation géodésique ou conditionnelle : La dilatation géodésique de Y dans X (appelée aussi dilatation de Y conditionnellement à X) par une boule de rayon r est définie par :

$$D_X(Y, B_r) = \{x \in \mathbb{R}^n / B_X(x, r) \cap Y \neq \emptyset\} = \{x \in \mathbb{R}^n / d_X(x, Y) \leq r\}, \quad (6.33)$$

où $B_X(x, r)$ désigne la boule géodésique de centre x et de rayon r .

En pratique dans un espace discret, l'expression 6.33 se calcule très simplement à partir de la dilatation euclidienne (équation 6.3) par l'élément structurant élémentaire B_1 puisqu'elle est alors équivalente à :

$$D_X(Y, B_r) = [D(Y, B_1) \cap X]^r \quad (6.34)$$

où l'exposant signifie que l'opération est itérée r fois.

Définition 37. Érosion géodésique : L'érosion géodésique de Y dans X par une boule de taille r est définie par :

$$E_X(Y, B_r) = \{x \in \mathbb{R}^n / B_X(x, r) \subset Y\} = X \setminus D_X(X \setminus Y, B_r) \quad (6.35)$$

et est ainsi reliée à la dilatation par dualité (faisant intervenir le complémentaire $X \setminus Y$ de Y dans X).

L'ouverture et la fermeture géodésiques sont définies à partir de l'érosion et de la dilatation géodésiques, de manière analogue au cas euclidien (équations 6.13 et 6.14).

De la dilatation géodésique, on déduit la notion de reconstruction de X à partir de \mathbb{I} marqueurs \mathbb{I} définis par un ensemble Y .

Définition 38. Reconstruction binaire : La reconstruction de Y dans X est la dilatation géodésique de Y dans X par une boule de rayon infini, c'est-à-dire, les composantes connexes de X qui ont une intersection non vide avec Y .

Les notions de transformations géodésiques ou conditionnelles s'étendent au cas numérique par l'intermédiaire des \mathbb{I} sections \mathbb{I} f_λ de la fonction f considérée :

$$f_\lambda = \{x \in \mathbb{R}^n / f(x) \geq \lambda\}.$$

Définition 39. Dilatation géodésique numérique : La dilatation géodésique de f dans g , pour $f \leq g$ est définie par ses sections à partir de la dilatation géodésique binaire :

$$[D_g(f, B_r)]_\lambda = D_{g_\lambda}(f_\lambda, B_r),$$

et l'érosion géodésique est définie par dualité.

Dans le cas discret, on retrouve des formules analogues à celles obtenues dans le cas binaire, pour la dilatation :

$$D_g(f, B_r) = [D(f, B) \wedge g]^r,$$

où B est l'élément structurant élémentaire et l'exposant r signifie que l'opération est itérée r fois, et pour l'érosion :

$$E_g(f, B_r) = [E(f, B) \vee g]^r.$$

Comme dans le cas binaire, la **reconstruction numérique** de f dans g est définie par $D_g(f, B_\infty)$. La fonction f est appelée fonction de marquage.

6.9.3 Squelette géodésique par zones d'influence

Soient X un ensemble et Y un ensemble de composantes connexes disjointes deux à deux Y_i ($Y = \cup_i Y_i$). La zone d'influence géodésique $ZI_X(Y_i)$ de Y_i conditionnellement à X est l'ensemble des points de X plus proches de Y_i que des autres composantes au sens de la distance géodésique dans X :

$$ZI_X(Y_i) = \{x \in X, d_X(x, Y_i) < d_X(x, Y \setminus Y_i)\}. \quad (6.36)$$

Le squelette géodésique par zone d'influence est alors défini par :

$$Skiz_X(Y) = X \setminus \cup_i ZI_X(Y_i). \quad (6.37)$$

6.10 Ligne de partage des eaux

6.10.1 Définition et propriétés

La ligne de partage des eaux est une notion très importante pour les problèmes de segmentation. Intuitivement, elle est définie par analogie géographique comme le complémentaire des bassins versants, un bassin versant étant la zone associée à un minimum régional telle qu'une goutte d'eau tombant dans cette zone et suivant la ligne de plus grande pente s'arrêtera dans ce minimum (figure 6.17).

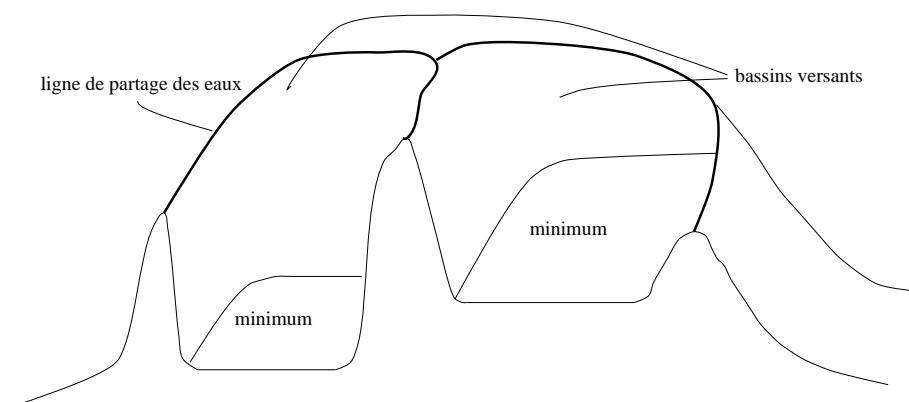


FIG. 6.17 – Ligne de partage des eaux.

La transposition de cette définition intuitive en termes mathématiques ne va pas sans poser de nombreux problèmes (définition d'une ligne de plus grande pente, problème des plateaux, etc.). La plupart des définitions sont algorithmiques, c'est-à-dire que la ligne de partage des eaux est définie d'après le moyen qui permet de la construire. Un des algorithmes les plus populaires est l'algorithme d'immersion qui consiste à remplir progressivement les bassins versants (à partir des minima régionaux) pour déterminer leurs limites [Vincent et Soille, 1991, Vincent, 1992, Soille, 1999]. Ce n'est que récemment que des approches mathématiques rigoureuses ont été proposées dans le cas continu [Schmitt et Mattioli, 1994a].

L'avantage de la ligne de partage des eaux pour la segmentation est qu'elle fournit des régions délimitées par des contours fermés formant une partition de l'image.

La figure 6.18 donne un exemple de ligne de partage des eaux sur une image de gradient.

6.10.2 Applications en segmentation

Une technique puissante de segmentation par morphologie mathématique, qui connaît des développements importants et de plus en plus d'applications, consiste à effectuer un marquage puis une reconstruction, éventuellement précédés de pré-traitements (rehaussement de contraste, filtrage, en particulier par des filtres alternés séquentiels). Nous donnons ici deux exemples d'application de ces techniques, pour séparer des objets binaires se recouvrant partiellement d'une part, et pour segmenter des images numériques à l'aide de la ligne de partage des eaux d'autre part.

Supposons qu'on ait un ensemble d'objets binaires de forme relativement circulaire, et se recouvrant partiellement. La séparation de tels objets peut s'effectuer par morphologie mathématique si le recouvrement n'est pas trop important : les composantes connexes de l'image apparaissent alors avec un rétrécissement qui correspond au lieu où la séparation doit être effectuée. Les marqueurs généralement utilisés pour ce type d'objets sont les érodés ultimes, c'est-à-dire les maxima régionaux de la fonction distance à l'intérieur de l'objet. En effet, une particule isolée donnera lieu à un seul érodé ultime connexe, deux particules se recouvrant partiellement (donc ne formant qu'une seule composante connexe dans l'image) auront un érodé ultime composé de deux composantes connexes qui marquent donc bien chacune des particules, etc.

L'étape de reconstruction permet ensuite de retrouver chaque particule à partir de son marqueur. Elle doit satisfaire deux exigences : préserver le nombre de composantes connexes de l'érodé ultime (puisque chacune d'elles est associée à une particule) et donc ne pas les connecter, et rester dans l'ensemble X de départ (qui représente toutes les particules). La première exigence conduit à utiliser des transformations homotopiques, et la seconde à conditionner les transformations par X donc à se placer dans un cadre géodésique.

Une première solution consiste à calculer le squelette par zones d'influence de l'ensemble des marqueurs conditionnellement à X . L'inconvénient de cette méthode est que les lignes de séparation peuvent être mal positionnées : elles sont situées à mi-distance des marqueurs, et cela n'est pas adapté à la séparation de deux particules de tailles différentes.

Une deuxième solution consiste à reconstruire les particules par dilatation géodésique des marqueurs (conditionnellement à X), en prenant en compte la taille de l'érosion nécessaire pour arriver à l'érodé ultime de chaque particule (c'est-à-dire la valeur du maximum régional correspondant de la fonction distance).

Une troisième solution consiste à calculer la ligne de partage des eaux de la fonction distance inversée (où les maxima deviennent des minima). À cause des irrégularités sur les bords des objets, on peut aboutir à une sur-segmentation. Celle-ci peut être évitée en reconstruisant la distance diminuée d'une faible valeur sous la fonction distance avant de l'inverser pour appliquer la ligne de partage des eaux.

Cette procédure est illustrée sur une image binaire sur la figure 6.19.

Prenons maintenant l'exemple d'une image à niveaux de gris. La ligne de partage des eaux fournit un bon outil de segmentation. Elle peut être appliquée directement sur l'image initiale si l'on cherche à détecter les lignes de crête de l'image pour faire la segmentation, sur une image de gradient (représentant une image de contours numériques), sur le résultat d'un chapeau haut-de-forme si l'on souhaite mettre en évidence des structures fines linéaires (sombres ou claires), etc. Le choix de l'image sur laquelle la ligne de partage des eaux est appliquée dépend bien sûr du problème posé, et doit être effectué dans une première étape. L'inconvénient de la ligne de partage des eaux est qu'elle fournit souvent une image sur-segmentée. La technique de marquage permet alors d'imposer des minima à la ligne de partage des eaux pour éviter ce problème.

Les marqueurs doivent être déterminés à l'intérieur et à l'extérieur des objets que l'on souhaite segmenter. Pour cela, différentes techniques peuvent être envisagées (filtrage par filtre alterné séquentiel de l'image de gradient, recherche de minima ou maxima régionaux, etc.).

Les minima sont imposés à partir de ces marqueurs par la technique suivante : soit f l'image sur laquelle on souhaite appliquer la ligne de partage des eaux, et g la fonction de marquage prenant la valeur 0 aux points des marqueurs détectés et $+\infty$ (en pratique la valeur maximale) aux points du complémentaire des marqueurs. En effectuant l'érosion géodésique $E_{f \wedge g}(g, B_\infty)$ de g par un élément structurant de taille infinie conditionnellement

à $f \wedge g$ (c'est-à-dire une reconstruction numérique), on obtient une version modifiée de f où les minima sont uniquement ceux qui ont été définis par les marqueurs. C'est sur cette image que la ligne de partage des eaux est alors calculée, assurant ainsi que les bassins versants correspondent soit à l'objet à segmenter soit au fond, et qu'on obtient un contour fermé autour de l'objet.

La figure 6.20 donne un exemple de ligne de partage des eaux sur une image de gradient après reconstruction.

6.11 Conclusion

De nombreuses extensions peuvent être trouvées dans la littérature sur la morphologie mathématique, ainsi que des liens avec d'autres domaines. Par exemple, l'élément structurant peut être interprété comme une relation binaire et les opérations de morphologie mathématique sont donc applicables à des structures telles que des graphes, où les éléments structurants sont définis en fonction des arcs du graphe [Vincent, 1989].

Parmi les nombreux liens, citons ceux qui existent avec les équations de diffusion (voir chapitre 8) et l'équation eikonale [Schmitt et Mattioli, 1994a], ainsi qu'avec les approches multi-échelles (pyramides [Goutsias et Heijmans, 2000], ondelettes [Heijmans et Goutsias, 2000]).

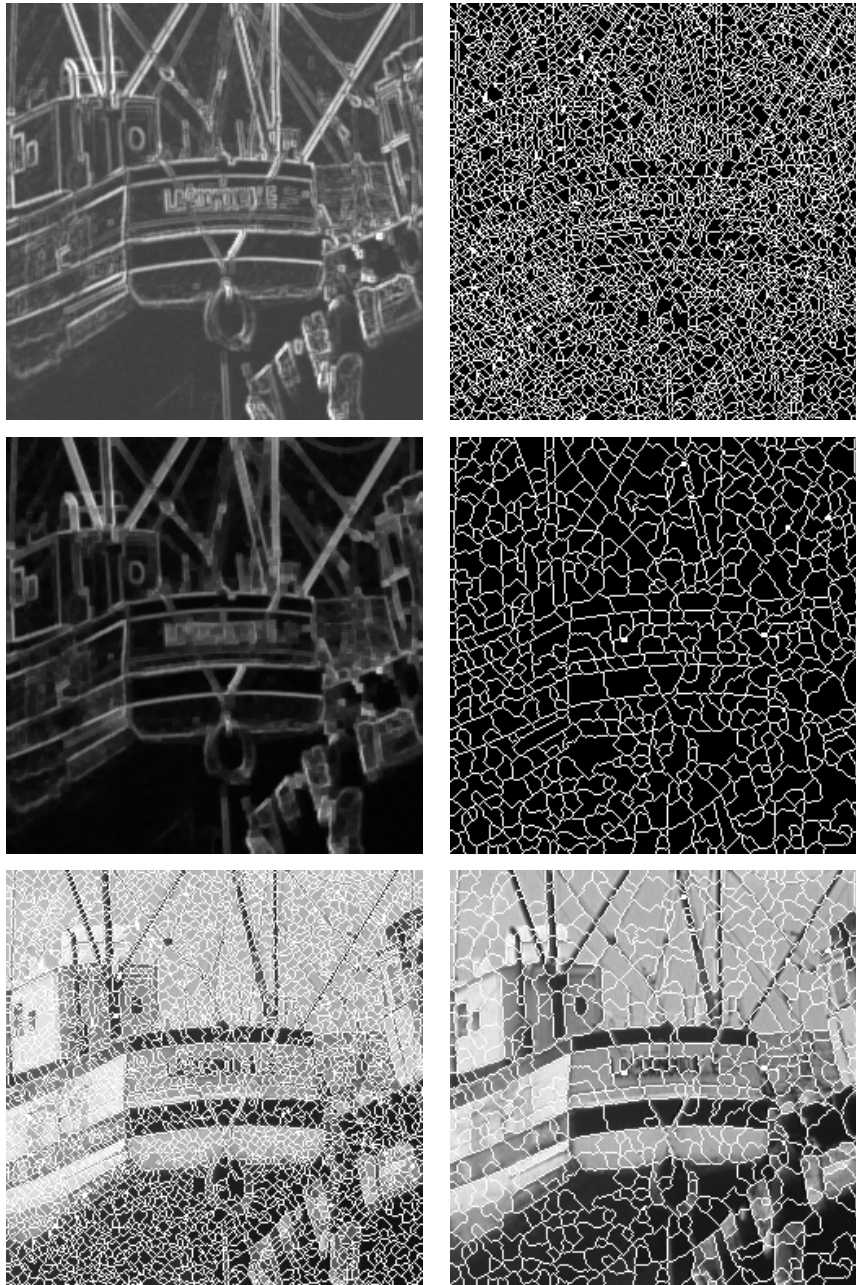


FIG. 6.18 – Ligne de partage des eaux appliquée sur une image de gradient (en haut), et sur une image de gradient fermé (au milieu). En bas : superposition de la ligne de partage des eaux sur l'image originale, sans et avec fermeture du gradient.

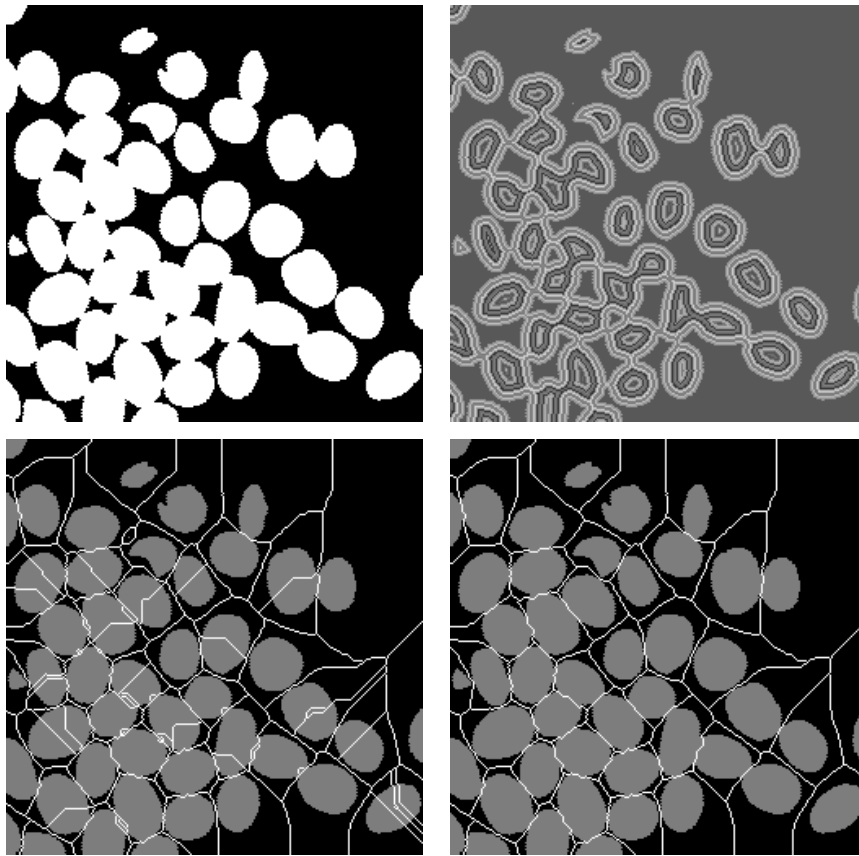


FIG. 6.19 – Image initiale, fonction distance, ligne de partage des eaux appliquée sur la distance inversée, ligne de partage des eaux appliquée à partir de la reconstruction de la distance moins 2 sous la distance.

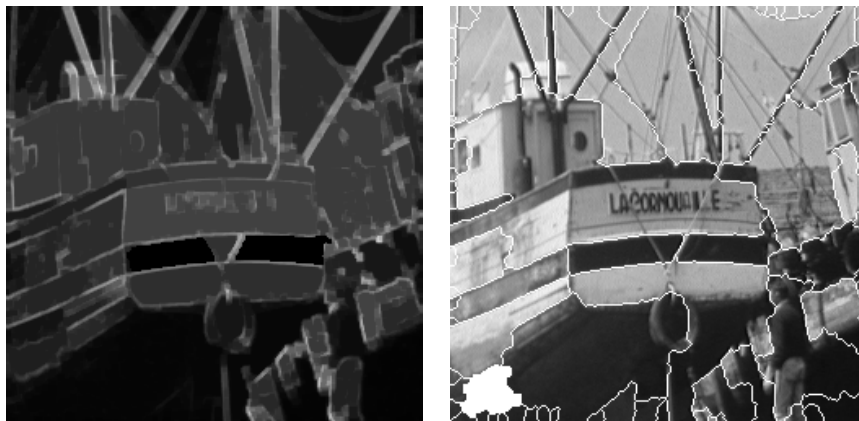


FIG. 6.20 – Image de gradient après reconstruction en imposant des marqueurs (à gauche) et ligne de partage des eaux appliquée sur cette image (à droite), limitant la sursegmentation.

Chapitre 7

Définition et simulation d'un champ de Markov

Chapitre rédigé par Florence TUPIN et Marc SIGELLE

L'information véhiculée par une image va bien au-delà de la seule donnée des niveaux de gris en chaque site (pixel), et la description se fait en termes de zones, contours, structures définies par les contrastes, textures, etc. qui peuvent être présents dans l'image. Le niveau de gris en un site n'est donc souvent pas significatif en lui-même, mais dans ses relations et interactions avec les pixels voisins.

Cette propriété des images, à savoir les interactions locales entre niveaux de gris voisins pour définir les différentes régions de l'image, va nous permettre d'utiliser un formalisme markovien dans de nombreux traitements, qu'il s'agisse de restauration, de segmentation ou plus tard d'analyse complète des images. Le principe est de définir des énergies locales entre groupes de sites reflétant les interactions entre niveaux de gris. L'énergie globale est alors reliée à la probabilité d'apparition de l'image dans le cadre des champs de Gibbs.

Dans ce chapitre, nous introduisons tout d'abord de façon intuitive la notion d'énergie locale avant de définir plus formellement un champ de Markov et d'énoncer le théorème d'équivalence entre champs de Markov et champs de Gibbs. Les algorithmes d'échantillonnage d'un champ de Markov (échantillonneur de Gibbs et algorithme de Métropolis) sont ensuite présentés, ainsi que les modèles markoviens les plus courants. L'utilisation des champs markoviens en traitement d'images dans un cadre bayésien montre la nécessité de pouvoir accéder aux configurations les plus probables d'un champ markovien et nous amène à la présentation du recuit simulé. Dans les parties suivantes, nous abordons les différents estimateurs (MAP, MPM, TPM), le problème de l'estimation des paramètres du champ, les processus de bords, avant de mentionner l'application de la modélisation markovienne à des graphes de primitives de plus haut niveau que les pixels.

7.1 Définition et simulation d'un champ de Markov

7.1.1 Description de l'image

L'image est formée d'un ensemble fini S de sites s_i correspondant aux pixels. S est donc essentiellement un réseau discret fini, partie de \mathbb{Z}^d , si on note d la dimension de l'espace (2 le plus classiquement, 3 pour les volumes, etc.). À chaque site est associé un descripteur, représentant l'état du site et qui peut être son niveau de gris, une étiquette, ou une information plus complexe, et prenant ses valeurs dans E .

La notion d'interactions locales nécessite de structurer les relations spatiales entre les différents sites du réseau.

Pour ce faire, on munit S d'un système de voisinage \mathcal{V} défini de la façon suivante :

$$\mathcal{V}_s = \{t\} \text{ tels que } \begin{cases} s \notin \mathcal{V}_s \\ t \in \mathcal{V}_s \Rightarrow s \in \mathcal{V}_t \end{cases}$$

À partir d'un système de voisinage, un système de cliques peut être déduit : une clique est soit un singleton de S , soit un ensemble de sites tous voisins les uns des autres. En fonction du système de voisinage utilisé, le système de cliques sera différent et fera intervenir plus ou moins de sites comme illustré sur la figure 7.1. On notera \mathcal{C} l'ensemble des cliques relatif à \mathcal{V} , et \mathcal{C}_k l'ensemble des cliques de cardinal k .

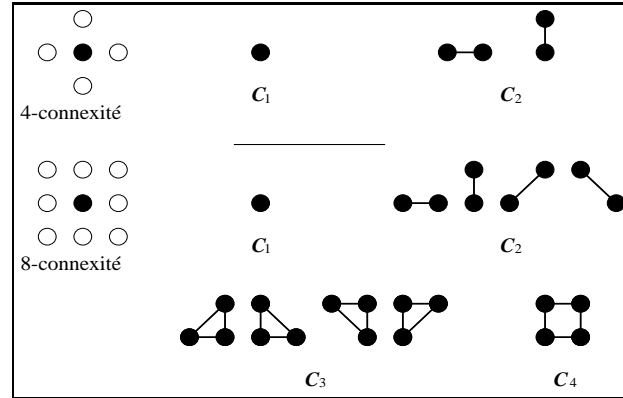


FIG. 7.1 – Les cliques associées à deux systèmes de voisinage en dimension 2.

Les interactions locales entre niveaux de gris (ou descripteurs) de sites voisins peuvent alors s'exprimer comme un potentiel de clique. Soit c une clique, on lui associe le potentiel U_c dont la valeur dépend des niveaux de gris (ou descripteurs) des pixels constituant la clique. En poursuivant ce raisonnement, on peut définir l'énergie globale de l'image comme la somme des potentiels de toutes les cliques :

$$U = \sum_{c \in \mathcal{C}} U_c$$

et l'énergie locale en un site comme la somme des potentiels de toutes les cliques auxquelles il appartient :

$$U_s = \sum_{c \in \mathcal{C} / s \in c} U_c$$

Nous avons jusqu'ici considéré le cas d'une image pour illustrer les notions de voisinage, de clique et de potentiel, mais le formalisme markovien se définit très généralement sur tout graphe. Soit un ensemble de sites S dénombrable (sommets du graphe), et une relation de voisinage, les cliques sont alors définies comme les sous-graphes complets du graphe. C'est l'utilisation de graphes plus généraux que ceux définis sur la grille de l'image qui permet des traitements de plus haut niveau.

7.1.2 Modélisation probabiliste de l'image

La définition des champs de Markov qui sera donnée dans la section suivante nécessite une modélisation probabiliste de l'image. Ainsi, l'image dont nous disposons va être considérée comme une réalisation d'un champ aléatoire. Soit s un site de l'image, on peut en effet lui associer une variable aléatoire (v.a) X_s prenant ses valeurs

dans E . Le niveau de gris x_s en s n'est ainsi qu'une réalisation¹ de la v.a X_s . On définit alors le champ aléatoire $X = (X_s, X_t, \dots)$ prenant ses valeurs dans $\Omega = E^{|S|}$. On trouvera aussi le terme de processus aléatoire pour X ; en toute rigueur, " processus " devrait être réservé au cas d'un ensemble d'indexation continu, et champ au cas discret.

Dans ce cadre probabiliste, l'image considérée est simplement une réalisation x du champ. La probabilité globale de x , $P(X = x)$, permet d'accéder en quelque sorte à la vraisemblance de l'image, et les probabilités conditionnelles locales d'une valeur en un site permettent de mesurer le lien statistique entre un niveau de gris et le reste de l'image. L'hypothèse markovienne permet d'évaluer ces quantités.

Notons que nous nous plaçons dans le cas où E , l'espace de valeurs des descripteurs, est quantifié, ce qui nous permet de manipuler des probabilités. Dans le cas où cet espace est continu, il faut remplacer P par une densité de probabilité, mais dans ce cas le théorème que nous allons voir ci-dessous n'est plus valable.

7.1.3 Champs de Markov - Champs de Gibbs

Définition d'un champ de Markov

Considérons x_s la valeur du descripteur prise au site s et $x^s = (x_t)_{t \neq s}$ la configuration de l'image excepté le site s . La définition d'un champ de Markov est alors la suivante :

X est un champ de Markov ssi la probabilité conditionnelle locale en un site n'est fonction que de la configuration du voisinage du site considéré

ce qui s'exprime de façon formelle par :

$$P(X_s = x_s / x^s) = P(X_s = x_s / x_t, t \in \mathcal{V}_s)$$

Ainsi, le niveau de gris en un site ne dépend que des niveaux de gris des pixels voisins de ce site. Cette hypothèse markovienne se justifie bien dans le cas de la plupart des images naturelles constituées de zones homogènes ou texturées ainsi que pour une large gamme d'images de synthèse. Plus généralement, une connaissance locale de l'image suffit souvent à réaliser son interprétation partielle et donc cette hypothèse markovienne sera souvent justifiée sur des graphes plus globaux que le graphe des pixels.

Notons qu'en l'absence de contrainte sur le système de voisinage, tous les champs aléatoires peuvent être considérés comme markoviens à condition de prendre un voisinage suffisamment grand. L'intérêt de cette modélisation réside bien sûr dans le cas où la propriété markovienne est vérifiée pour des voisinages restreints permettant des calculs rapides.

Equivalence entre champs de Markov et champs de Gibbs

La modélisation markovienne prend toute sa puissance grâce au théorème que nous allons voir maintenant. En effet, celui-ci permettra d'accéder aux expressions des probabilités conditionnelles locales. Il nous faut au préalable définir un certain nombre de notions relatives aux mesures et champs de Gibbs.

Définition d'une mesure de Gibbs : La mesure de Gibbs de fonction d'énergie (ou d'hamiltonien) $U : \Omega \rightarrow \mathbb{R}$ est la probabilité P définie sur Ω par :

$$P(X = x) = \frac{1}{Z} \exp(-U(x))$$

avec

$$U(x) = \sum_{c \in \mathcal{C}} U_c(x)$$

¹On notera généralement en lettres majuscules les variables aléatoires et en minuscules leurs réalisations.

où \mathcal{C} est le système de cliques associé au système de voisinage \mathcal{V} de U^2 .

$Z = \sum_{x \in \Omega} \exp(-U(x))$ est une constante de normalisation appelée fonction de partition de Gibbs. En pratique, il est quasi impossible de calculer cette constante à cause du très grand nombre de configurations possibles. Ne serait-ce que dans le cas d'une image binaire ($\text{Card}(E) = 2$) et de taille $= 512 \times 512$, on a $\text{card}(\Omega) = 2^{262144}$ configurations possibles !

La notation couramment utilisée pour $U(x)$ est abusive car $U_c(x)$ ne dépend pas de l'ensemble de la configuration x mais seulement de x restreinte à la clique c ($U_c(x) = U_c(x_t, t \in c)$).

Nous pouvons maintenant définir le champ de Gibbs de potentiel associé au système de voisinage \mathcal{V} : c'est le champ aléatoire X dont la probabilité est une mesure de Gibbs associée au système de voisinage \mathcal{V} , ce qui implique :

$$P(X = x) = \frac{1}{Z} \exp(-U(x)) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} U_c(x)\right)$$

L'énergie globale d'un champ de Gibbs possède donc la propriété de se décomposer sous forme d'une somme d'énergies locales, qui comme on le verra par la suite permettront d'accéder aux probabilités conditionnelles locales. Notons ici que plus une configuration d'un champ de Gibbs a une énergie faible, plus elle est probable.

Le théorème de Hammersley-Clifford [Besag, 1974] établit alors le résultat fondamental suivant sous les hypothèses :

- S fini ou dénombrable,
- le système de voisinage \mathcal{V} est borné,
- l'espace des états E est discret

X est un champ de Markov relativement à \mathcal{V} et $P(X = x) > 0 \forall x \in \Omega$ si et seulement si X est un champ de Gibbs de potentiel associé à \mathcal{V} .

Par exemple, si nous considérons un champ de Markov de voisinage 4-connexe, nous pouvons écrire l'énergie de la configuration x sous la forme :

$$U(x) = \sum_{c=(s) \in \mathcal{C}_1} U_c(x_s) + \sum_{c=(s,t) \in \mathcal{C}_2} U_c(x_s, x_t)$$

Notons que rien n'interdit la non-stationnarité du champ, c'est-à-dire la variation des potentiels U_c en fonction de la localisation de la clique c dans l'image [Descombes, 1993]. D'autre part, rien n'impose la symétrie des potentiels et on peut avoir $U_{c=(r,s)}(0, 1) \neq U_c(1, 0)$.

Le théorème de Hammersley-Clifford, et la forme bien spécifique de probabilité de X qui en résulte, va permettre de lier les probabilités globales et locales comme nous allons le voir maintenant. En effet si nous cherchons à écrire la probabilité conditionnelle locale $P(x_s / X^s = x^s)$, nous avons grâce au résultat précédent :

$$P(X_s = x_s / X^s = x^s) = \frac{P(X = x)}{P(X^s = x^s)} = \frac{\exp(-U(x_s, x^s))}{\sum_{\xi \in E} \exp(-U(\xi, x^s))}$$

Définissons l'énergie locale U_s par :

$$U_s(x_s / x_t, t \in \mathcal{V}_s) = \sum_{c \in \mathcal{C} / s \in c} U_c(x_s, x_t, t \in \mathcal{V}_s) = \sum_{c \in \mathcal{C} / s \in c} U_c(x_s, V_s)$$

en notant $V_s = (x_t, t \in \mathcal{V}_s)$. Cette énergie locale ne fait donc intervenir que les voisins de s . On peut alors écrire l'énergie globale $U(x)$ sous la forme :

²Il est toujours possible de trouver un système de voisinage \mathcal{V} permettant de décomposer U ; le cas extrême correspondant à des sites tous voisins les uns des autres.

$$U(x) = \sum_{c \in \mathcal{C} / s \notin c} U_c(x) + \sum_{c \in \mathcal{C} / s \in c} U_c(x) = \sum_{c \in \mathcal{C} / s \notin c} U_c(x) + U_s(x_s / V_s)$$

En simplifiant l'expression de la probabilité conditionnelle locale en supprimant les termes communs qui font intervenir les cliques ne contenant pas le site s au numérateur et au dénominateur, on a :

$$\begin{aligned} P(X_s = x_s \mid X^s = x^s) &= \frac{\exp\left(-\sum_{c \in \mathcal{C} / s \notin c} U_c(x) - U_s(x_s / V_s)\right)}{\sum_{x_s \in E} \exp\left(-\sum_{c \in \mathcal{C} / s \notin c} U_c(x) - U_s(x_s / V_s)\right)} \\ &= \frac{\exp(-U_s(x_s / V_s))}{\sum_{x_s \in E} \exp(-U_s(x_s / V_s))} \end{aligned} \quad (7.1)$$

L'expression obtenue, qui ne fait intervenir que les potentiels des cliques contenant le site s (ce qui nous permet de retrouver au passage l'hypothèse markovienne), est très importante. En effet, autant il n'est pas possible partant d'une configuration x d'accéder à sa probabilité à cause de la constante de normalisation, autant il est possible de calculer en chaque site la probabilité conditionnelle locale. Cette expression sera à la base de tous les algorithmes de simulation de champs markoviens que nous verrons dans la section suivante.

Remarque : Le théorème de Hammersley-Clifford n'est valable que lorsqu'aucune configuration n'est interdite (condition $P(X = x) > 0 \forall x$). Des solutions ont été proposées, lorsqu'il est souhaitable de supprimer certaines configurations irréalistes (par exemple avoir une route à l'intérieur d'une zone de mer, ou de l'os dans la matière blanche du cerveau).

7.1.4 Echantillonnage de MRF

Si nous résumons les résultats précédents, la définition d'un champ de Markov passe par la définition de sa fonction d'énergie U . Celle-ci nécessite la définition d'un système de voisinage, qui définit alors le système de cliques, et de fonctions de potentiel associées aux cliques. Ces fonctions de potentiel permettent d'accéder à la probabilité globale d'une configuration, et aux probabilités conditionnelles locales.

Le problème qui se pose alors est, étant défini un champ de Markov, comment pouvons-nous réaliser le tirage d'une configuration (une image ici) en suivant la loi de probabilité de Gibbs caractéristique de ce champ ? Deux algorithmes ont été proposés pour synthétiser des réalisations d'un champ de Markov qui sont :

- l'échantillonneur de Gibbs,
- l'algorithme de Métropolis

que nous allons décrire maintenant.

L'échantillonneur de Gibbs

Cet algorithme, proposé par Geman et Geman [Geman et Geman, 1984a], repose sur la construction itérative d'une suite d'images. A la convergence, i.e après un nombre d'itérations suffisant, les images construites sont des réalisations tirées selon la loi de Gibbs globale.

La méthode de construction de l'image à l'itération n , partant de l'image à l'itération $n - 1$ se fait par mises à jour successives des sites de l'image. À l'étape n :

- choix d'un site s ;

- au site s , selon la configuration des voisins V_s pour l'image $x^{(n-1)}$, calcul de la probabilité conditionnelle locale :

$$P(X_s = x_s | V_s) = \frac{\exp(-U_s(x_s | V_s))}{\sum_{\xi \in E} (\exp(-U_s(\xi | V_s)))}$$

- mise à jour du site s par tirage aléatoire selon la loi $P(X_s = x_s | V_s)$.

On considère que l'algorithme a convergé après un grand nombre d'itérations ou lorsque le nombre de changements est faible. Le choix du site s considéré à l'étape n peut se faire de n'importe quelle façon à condition de balayer tous les sites un très grand nombre de fois (théoriquement un nombre infini de fois). Les méthodes usuelles consistent à tirer un site selon une loi uniforme, ou effectuer un balayage classique, ligne par ligne, de l'image.

Cet algorithme construit en réalité une suite d'images $x^{(n)}$ qui sont les observations d'une suite $X^{(n)}$ de champs aléatoires constituant une chaîne de Markov pour un certain noyau de transition. On peut montrer le théorème suivant, lorsque la séquence balaye chaque site une infinité de fois :

$$\forall x^{(0)} \forall x \in \Omega \lim_{n \rightarrow \infty} P(X^{(n)} = x | X^{(0)} = x^{(0)}) = P(x)$$

où P est la mesure de Gibbs associée au champ de Markov considéré. Ainsi, après un grand nombre d'itérations, les images $x^{(n)}$ générées sont des réalisations de la loi globale $P(x)$, et cela indépendamment de la configuration initiale $x^{(0)}$.

La preuve de ce théorème, et donc de la convergence de l'algorithme, est donnée dans [Winkler, 1995].

On parle de l'échantillonneur de Gibbs comme d'un algorithme de relaxation, car il procède par mises à jour successives des sites, et probabiliste car celle-ci est fondée sur un tirage aléatoire.

L'algorithme de Metropolis

L'échantillonneur de Gibbs est un algorithme très utilisé en traitement d'images pour la synthèse de champs de Markov. Néanmoins, un algorithme antérieur et issu de la physique statistique avait été mis au point dans les années 50 par Metropolis [Metropolis et al., 1953].

Cet algorithme repose sur un principe similaire à l'échantillonneur de Gibbs, et il s'agit également d'un algorithme de relaxation probabiliste. Le principe est là encore de construire une suite d'images qui seront des tirages selon la loi du champ de Markov après un nombre suffisamment grand d'itérations. Mais la mise à jour en un site s'effectue de façon différente. Ainsi à l'étape n :

- choix d'un site s
- tirage aléatoire d'un descripteur λ dans E selon une loi uniforme ;
- calcul de la variation d'énergie pour le passage du label du site s de $x_s^{(n-1)}$ à λ :

$$\Delta U = U_s(\lambda | V_s^{(n-1)}) - U_s(x_s^{(n-1)} | V_s^{(n-1)})$$

- deux cas sont alors possibles :

1. $\Delta U < 0$, le changement est accepté : $x_s^{(n)} = \lambda$;
2. $\Delta U \geq 0$, le changement est accepté ou refusé par tirage selon les probabilités $p = \exp(-\Delta U)$ et $1 - p$.

Le système de balayage des sites et le critère d'arrêt sont similaires à ceux de l'échantillonneur de Gibbs. La différence avec l'échantillonneur de Gibbs réside dans le tirage au sort du nouveau niveau de gris (ou descripteur), au lieu de considérer la loi définie par tous les descripteurs. Comme on ne considère que la variation énergétique entre les 2 configurations, l'algorithme de Metropolis est plus rapide à chaque étape que l'échantillonneur de Gibbs, qui lui nécessite le calcul de la fonction de partition locale. Mais la convergence peut être plus lente car le taux

d'acceptation est strictement inférieur à 1 (les transitions ne sont pas toujours acceptées, contrairement au cas de l'échantillonneur de Gibbs).

Là encore, le principe est de construire une chaîne de Markov selon un certain noyau de transition (différent de celui intervenant dans l'échantillonneur de Gibbs). Le théorème précédent est alors encore vérifié pour l'algorithme de Metropolis.

7.1.5 Le recuit simulé

Nous avons vu dans les paragraphes précédents comment échantillonner selon la loi de probabilité de Gibbs associée au champ de Markov. À chaque application des précédents algorithmes, une nouvelle réalisation est obtenue. Il peut être utile également de pouvoir calculer la ou les configurations les plus probables qui correspondent aux états d'énergie minimale. C'est l'algorithme du recuit simulé qui permet de trouver ces réalisations.

Avant de présenter cet algorithme, nous avons besoin de quelques résultats sur les distributions de Gibbs avec paramètre de température que nous présentons maintenant.

Distribution de Gibbs avec température

Une distribution de Gibbs avec paramètre de température est une probabilité qui s'écrit :

$$P_T(X = x) = \frac{1}{Z(T)} \exp\left(-\frac{U(x)}{T}\right)$$

avec $Z(T) = \sum_x \exp\left(-\frac{U(x)}{T}\right)$ et $T > 0$. Le terme de température provient de l'analogie avec la physique statistique.

Il est intéressant d'étudier le comportement de cette distribution pour des valeurs extrêmes du paramètre de température.

- $T \rightarrow \infty$:

On a $\exp\left(-\frac{U(x)}{T}\right) \rightarrow 1$ et comme $\sum_x P_T(X = x) = 1$, on obtient

$$P_T(X = x) \rightarrow \frac{1}{\text{Card } \Omega}$$

Donc P_T converge vers la probabilité uniforme sur Ω , i.e pour une température infinie tous les états sont équiprobables.

- $T \rightarrow 0$:

Notons U^* l'énergie minimale et Ω^* l'ensemble des configurations atteignant l'énergie minimale $\Omega^* = \{x_1, \dots, x_k\}$ (x_1, \dots, x_k sont les minima globaux de l'énergie). On peut écrire :

$$\begin{aligned} P_T(X = x) &= \frac{\exp\left(-\frac{U(x)}{T}\right)}{\sum_y \exp\left(-\frac{U(y)}{T}\right)} = \frac{\exp\left(-\frac{U(x) - U^*}{T}\right)}{\sum_y \exp\left(-\frac{U(y) - U^*}{T}\right)} \\ &= \frac{\exp\left(-\frac{U(x) - U^*}{T}\right)}{\sum_{y \notin \Omega^*} \exp\left(-\frac{U(y) - U^*}{T}\right) + \sum_{y \in \Omega^*} 1} \end{aligned}$$

◊ Si $x \notin \Omega^*$, on a $U(x) - U^* > 0$ et $\exp(\frac{U(x)-U^*}{T}) \rightarrow 0$ pour $T \rightarrow 0$. Donc $P_T(x) \rightarrow 0$ si x n'est pas un minimum global de l'énergie.

◊ Si $x \in \Omega^*$, on a : $P_T(x_1) = P_T(x_2) = \dots = P_T(x_k) = \frac{1}{k}$ (il y a une somme finie de termes qui tendent vers 0 au dénominateur).

Ce qui signifie que lorsque la température est nulle P_T est uniformément distribuée sur les minima globaux de l'énergie, i.e sur les configurations les plus probables. C'est ce résultat qui est à la base de l'algorithme de recuit simulé.

Algorithme du recuit simulé

Cet algorithme est dédié à la recherche d'une configuration d'énergie minimale d'un champ de Gibbs (on ne cherche plus ici à échantillonner contrairement à précédemment). L'idée d'intégrer un paramètre de température et de simuler un recuit a été initialement proposée par Kirkpatrick [Kirkpatrick et al., 1982] et reprise par Geman et Geman [Geman et Geman, 1984a] qui ont proposé l'algorithme suivant.

Comme les algorithmes de simulation, c'est un algorithme itératif qui construit la solution au fur et à mesure. Le déroulement de l'algorithme est le suivant (en notant n le numéro de l'itération) :

- choix d'une température initiale $T^{(0)}$ suffisamment élevée
- choix d'une configuration initiale quelconque $x^{(0)}$
- à l'étape n
 - simulation d'une configuration $x^{(n)}$ pour la loi de Gibbs d'énergie $\frac{U(x)}{T^{(n)}}$ à partir de la configuration $x^{(n-1)}$; la simulation peut se faire par l'échantillonneur de Gibbs ou l'algorithme de Métropolis ; on réalise en général un balayage complet de l'image à la température $T^{(n)}$;
 - diminution lente de la température : $T^{(n)} > \frac{c}{\log(1+n)}$
- arrêt lorsque le taux de changement est faible.

La décroissance logarithmique de la température est un rythme très lent ; en pratique des décroissances géométriques sont utilisées, souvent sans dégradation notable des résultats obtenus. La constante c intervenant dans la décroissance dépend de la variation énergétique globale maximale sur l'espace des configurations.

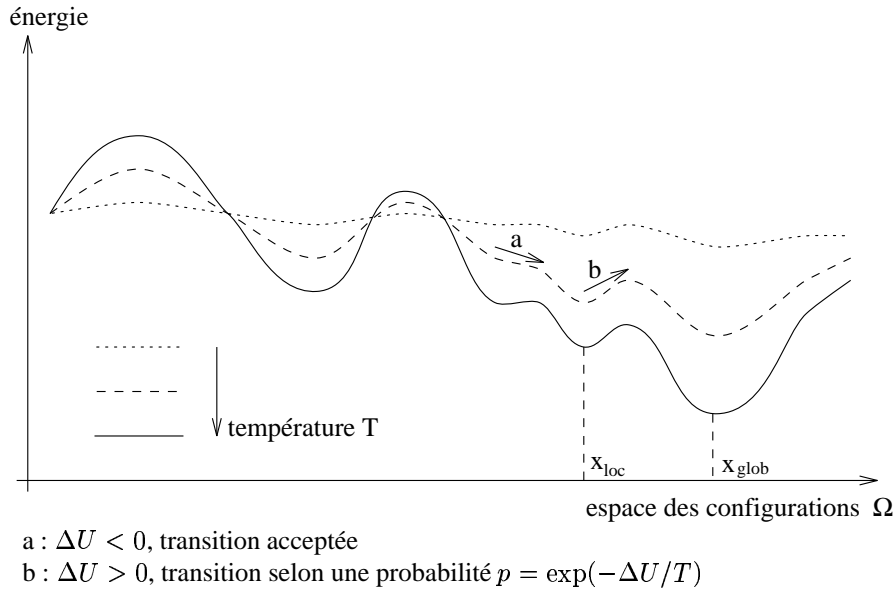
La figure 7.1.5 montre l'évolution du paysage énergétique représenté en 1 dimension au fur et à mesure de la décroissance en température. Au départ, toutes les configurations sont équiprobables puis les minima énergétiques apparaissent et s'accroissent.

Notons que contrairement aux algorithmes de l'échantillonneur de Gibbs et de Métropolis qui échantillonnent selon la loi de Gibbs et qui sont en mesure de donner toutes les configurations possibles, les images obtenues par recuit simulé sont uniques et doivent en théorie correspondre aux minima globaux de l'énergie.

Il existe une preuve de convergence de cet algorithme, qui repose à nouveau sur la construction d'une chaîne de Markov, mais qui est hétérogène cette fois-ci à cause de la variation du paramètre de température [Geman et Geman, 1984a]. Intuitivement, le recuit simulé permet d'atteindre un optimum global, car il accepte des remontées en énergie. Avec la décroissance de la température, ces sauts énergétiques sont progressivement supprimés au fur et à mesure qu'on se rapproche de l'optimum global. La descente en température doit donc se faire suffisamment lentement pour que l'algorithme ne reste pas piégé dans un minimum local de l'énergie.

Algorithme des modes conditionnels itérés (ICM)

Malheureusement, l'algorithme du recuit simulé est très lourd en temps de calcul puisqu'il demande la génération d'un grand nombre de configurations au fur et à mesure que la température décroît. Des algorithmes sous-optimaux sont donc souvent utilisés en pratique. Besag [Besag, 1986] a ainsi proposé un autre algorithme, beaucoup plus



rapide, mais pour lequel nous n'avons pas de preuve de convergence vers un minimum global. Il s'agit de l'ICM, *Iterated Conditional Mode*, que nous allons présenter ici.

Cet algorithme est un algorithme itératif modifiant à chaque étape les valeurs x_s de l'ensemble des sites de l'image. Mais à la différence de ces algorithmes qui étaient stochastiques par essence, la modification d'une valeur se fait ici de façon déterministe.

On construit donc, partant d'une configuration initiale $x(0)$, une suite d'images $x(n)$, convergeant vers une approximation du MAP \hat{x} recherché. Soit un tour la visite de tous les sites de l'image, on parlera dans la suite d'itérations à chaque mise à jour d'un site et d'étape à chaque mise à jour de toute l'image (i.e accomplissement d'un tour).

Le déroulement de l'étape n s'effectue de la façon suivante : on parcourt tous les sites et en chaque site, on effectue les deux opérations suivantes :

1. calcul des probabilités conditionnelles locales, pour toutes les valeurs possibles de λ dans E du site :

$$P(X_s = \lambda / \hat{x}_r(k), r \in \mathcal{V}_s)$$

(en pratique, calcul plus simplement des énergies conditionnelles locales)

2. mise à jour de la valeur par le λ maximisant la probabilité conditionnelle locale :

$$\hat{x}_s(k+1) = \text{Argmax}_{\lambda} P(X_s = \lambda / \hat{x}_r(k), r \in \mathcal{V}_s)$$

(ou de façon équivalente, minimisant l'énergie conditionnelle locale).

Le processus s'arrête lorsque le nombre de changements d'une étape à l'autre devient suffisamment faible.

On peut montrer que l'énergie globale de la configuration \hat{x} diminue à chaque itération. Cet algorithme, contrairement au recuit simulé, est très rapide (une dizaine de balayages permettent d'arriver à convergence) et peu coûteux en temps de calcul puisqu'il ne nécessite que le calcul des énergies conditionnelles locales. En contrepartie, ses performances dépendent très fortement de l'initialisation (par rapport à la forme du paysage énergétique) puisqu'il converge vers un minimum local. L'ICM s'apparente à une descente en gradient (on fait baisser l'énergie à chaque itération) ou à un recuit simulé gelé à température nulle, et peut donc rester bloqué dans le minimum

énergétique local le plus proche de l'initialisation. Le recuit simulé, au contraire, grâce au paramètre de température et aux remontées en énergie qu'il autorise permet d'accéder au minimum global.

Notons qu'il a également été proposé d'utiliser la programmation dynamique pour estimer le MAP [Derin et Elliott, 1987a]. Mais il est alors nécessaire d'être dans une configuration simple de segmentation (peu d'étiquettes, dimensions raisonnables) et seule une approximation peut être obtenue.

7.1.6 Quelques MRF fondamentaux

Nous présentons ici quelques uns des champs de Markov les plus utilisés. Comme indiqué précédemment, ces champs sont définis par leur voisinage et leurs fonctions de potentiel. Ils sont illustrés par le tirage de réalisations selon l'échantillonneur de Gibbs.

◇ Modèle d'Ising :

Ce modèle est le plus ancien (1925 [Ising, 1925]) et a été développé lors de l'étude du ferro-magnétisme en physique statistique. L'espace des descripteurs est celui des états des spins, i.e $E = \{-1, 1\}$ (espace binaire), et le voisinage est constitué par les 4 ou 8 plus proches voisins dans un espace bidimensionnel. Les potentiels sont des potentiels en tout ou rien :

$$\begin{aligned} U_{c=(s,t)}(x_s, x_t) &= -\beta \text{ si } x_s = x_t \\ &= +\beta \text{ si } x_s \neq x_t \end{aligned}$$

Ce qui s'écrit également $U_{c=(s,t)}(x_s, x_t) = -\beta x_s x_t$.

Les potentiels des cliques d'ordre 1 (clique constituée par un seul spin) sont de la forme $-B x_s$. L'énergie totale s'écrit :

$$U(x) = - \sum_{c=(s,t) \in \mathcal{C}} \beta x_s x_t - \sum_{s \in \mathcal{S}} B x_s$$

β est la constante de couplage entre sites voisins et B représente un champ magnétique externe. Lorsque β est positif, les configurations les plus probables (i.e d'énergies plus faibles) sont celles pour lesquelles les spins sont de même signe (ferro-magnétisme), alors que dans le cas de β négatif, au contraire, on favorisera l'alternance de spins de signes opposés (anti-ferromagnétisme). La valeur (signe et valeur absolue) de β conditionne donc la régularité du modèle d'Ising. Quant au champ magnétique externe relatif au potentiel d'ordre 1, il favorise a priori par son signe un spin ou un autre.

La figure 7.2 montre des réalisations de modèles d'Ising pour différents paramètres (la régularisation appelée critique correspond à l'apparition des zones homogènes).

◇ Modèle de Potts :

Il s'agit d'une extension du modèle d'Ising [Wu, 1982] pour un espace m -aire, i.e $E = \{0, m - 1\}$. Il peut s'agir de plusieurs niveaux de gris, mais plus souvent pour ce modèle, d'étiquettes (labels) pouvant représenter une classification de l'image (par exemple les classes *eau*, *forêt*, *champ*, *ville*). Le voisinage considéré est 4- ou 8-connexe et les potentiels sont comme précédemment en tout ou rien mais définis seulement pour les cliques d'ordre 2 :

$$\begin{aligned} U_{c=(s,t)}(x_s, x_t) &= -\beta \text{ si } x_s = x_t \\ &= +\beta \text{ si } x_s \neq x_t \end{aligned}$$

Lorsque β est positif, les configurations les plus probables correspondent à des sites voisins de même niveau de gris ou descripteur, ce qui donne des réalisations constituées par des larges zones homogènes. La taille de ces régions est gouvernée par la valeur de β . Des exemples de réalisations pour différentes valeurs de β sont montrés figure 7.3.

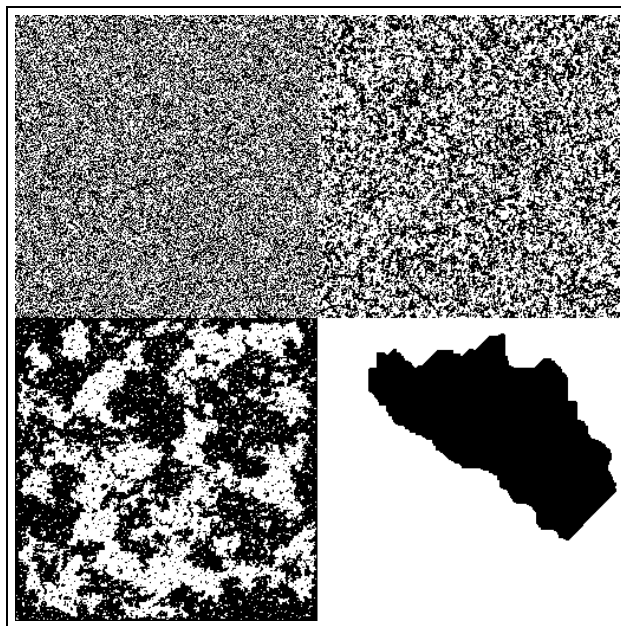


FIG. 7.2 – Le modèle d’Ising plan et 4-connexe pour différentes valeurs de paramètres - Simulations en raccordement torique ; a. Image aléatoire : $\beta = 0$, b.régularisation faible, $\beta = 0, 2$; c. régularisation critique : $\beta \approx 0, 44$; d. régularisation forte : $\beta = 4$.

Il est possible de définir des modèles utilisant des pondérations β différentes en fonction des directions des cliques, et de privilégier ainsi certaines directions.

Ce modèle permet également de prendre en compte différentes relations entre les régions (i.e. entre différentes valeurs des descripteurs). On peut par exemple définir des pondérations $\beta(e_s, e_t)$ pour $e_s, e_t \in E$. Dans notre exemple de classification en 4 étiquettes *eau*, *forêt*, *champ*, *ville*, une configuration de sites avec les étiquettes *champ/forêt* peut être supposée plus probable qu’une configuration *ville/forêt*, d’où des valeurs $\beta(\text{champ}, \text{forêt})$ et $\beta(\text{ville}, \text{forêt})$ différentes [Sigelle, 1993].

◇ Modèle markovien gaussien :

Ce modèle est réservé aux images en niveaux de gris $E = \{0, \dots, 255\}$ et ne convient pas bien aux images d’étiquettes. Le voisinage est 4 ou 8-connexe et l’énergie est de la forme :

$$U(x) = \beta \sum_{c=(s,t)} (x_s - x_t)^2 + \alpha \sum_{s \in S} (x_s - \mu_s)^2$$

Le premier terme correspondant aux cliques d’ordre 2 est un terme de régularisation, qui favorise les faibles différences de niveaux de gris entre sites voisins pour $\beta > 0$. Le second terme peut correspondre à un terme d’attache aux données dans le cas où on possède une image de données extérieures. Le rapport $\frac{\alpha}{\beta}$ pondère les influences respectives de l’attache aux données et de la régularisation, et les valeurs absolues des paramètres caractérisent le caractère plus ou moins piqué ou équiréparti au contraire de la distribution.

7.1.7 Applications : restauration et segmentation

Nous avons étudié dans la section précédente des algorithmes de simulation permettant de générer des réalisations d’un champ de Markov donné. Nous abordons ici leur utilisation en tant qu’outils de traitements d’images pour

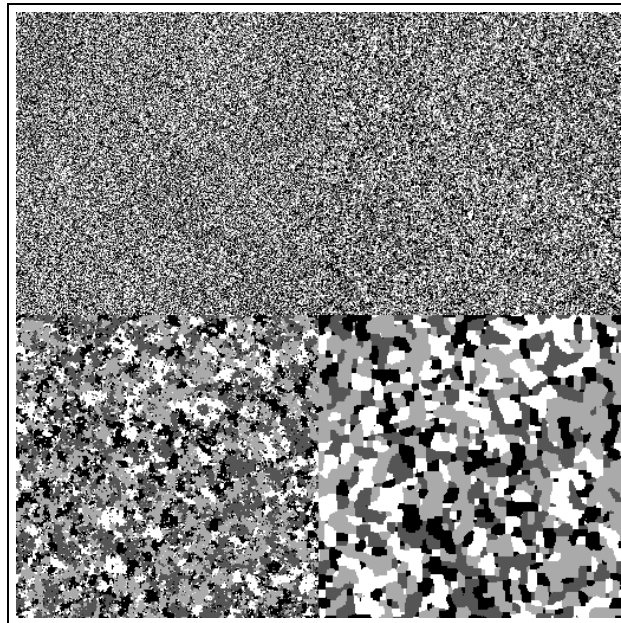


FIG. 7.3 – Le Modèle de Potts 2D et 4-connexe pour différentes valeurs de paramètres ($m = 4$) - Simulations en raccordement torique ; a. Image aléatoire : $\beta = 0$, b.régularisation faible, $\beta = 0.2$; c. régularisation critique : $\beta \approx 1,099$; d. régularisation forte : $\beta = 4$.

deux grandes applications de bas niveau : la restauration d'images et la segmentation.

Cadre bayésien

Pour ces deux applications, on peut modéliser le problème dans un cadre bayésien de la façon suivante. Nous disposons d'une certaine donnée (image) que nous noterons y et que nous pouvons considérer comme une réalisation d'un champ aléatoire Y . Nous cherchons une réalisation x de l'image restaurée ou segmentée, que nous pouvons modéliser comme un champ de Markov X . X est le champ des étiquettes (labels) dans le cas de la segmentation, le champ des intensités dans le cas de la restauration. Les espaces de configurations ne sont donc pas nécessairement les mêmes pour X et Y . Ces deux champs sont liés par le processus d'acquisition de l'image, qui conduit du champ idéal X , le processus image originel que nous cherchons, au champ bruité Y que nous observons. La restauration ou la segmentation ont pour objectif d'inverser le processus et donc de remonter à une réalisation de X à partir de l'observation des données bruitées y . On parle dans ce contexte de champ de Markov caché pour X , ou de données incomplètes puisque y n'est pas une réalisation de X .

On peut par exemple utiliser le critère du maximum a posteriori et rechercher la configuration \hat{x} maximisant la probabilité de X conditionnellement à la donnée y i.e $P(X = x / Y = y)$. Or la règle de Bayes permet d'écrire :

$$P(X = x / Y = y) = \frac{P(Y = y / X = x)P(X = x)}{P(Y = y)}$$

Expression dans laquelle il s'agit alors d'analyser chacun des termes $P(Y = y / X = x)$ et $P(X = x)$, sachant que $P(Y)$ est une constante (indépendante de la réalisation x). Le premier terme $P(Y = y | X = x)$ décrit justement le processus d'observation et d'acquisition des données. L'hypothèse la plus courante (dont la validité

reste à justifier) consiste à supposer l'indépendance conditionnelle des pixels (bruit non corrélé par exemple) :

$$P(Y = y / X = x) = \prod_s P(Y_s = y_s / X_s = x_s)$$

Cette écriture n'est plus valable lorsqu'il y a une convolution par la fonction de transfert du système d'acquisition, mais on peut montrer que le champ a posteriori reste markovien.

Par ailleurs, on fait sur le champ X recherché une hypothèse markovienne selon un voisinage \mathcal{V} et un modèle donné dépendant de l'application. On peut alors écrire :

$$P(X = x) = \frac{\exp(-U(x))}{Z}$$

Si on revient maintenant à la distribution a posteriori, celle-ci s'exprime par :

$$\begin{aligned} P(X = x / Y = y) &\propto P(Y / X)P(X) \propto e^{\ln P(Y / X) - U(x)} \\ &\propto e^{-\mathcal{U}(x / y)} \end{aligned}$$

avec :

$$\mathcal{U}(x / y) = \sum_{s \in \mathcal{S}} -\ln p(y_s / x_s) + \sum_{c \in \mathcal{C}} U_c(x) \quad (7.2)$$

Par conséquent, sous les hypothèses précédentes, on constate que la distribution a posteriori est une distribution de Gibbs et que donc le champ X conditionnellement à y est également un champ de Markov (théorème de Hammersley-Clifford). Ainsi, il est possible de simuler des réalisations de ce champ à l'aide de l'échantillonneur de Gibbs ou de l'algorithme de Metropolis. Mais la configuration x qui nous intéresse est celle maximisant la probabilité a posteriori, donc la réalisation la plus probable du champ de Gibbs, ou encore celle qui minimise l'énergie $\mathcal{U}(x / y)$. L'algorithme du recuit simulé décrit plus haut permet d'atteindre ce (ou ces) état(s) d'énergie minimale.

Cas de la restauration

Reprenons la démarche précédente et exprimons plus en détails l'énergie $\mathcal{U}(x / y)$ dans un cas particulier de restauration.

Dans le cas où le processus d'acquisition entraîne une dégradation de l'image sous forme d'un bruit blanc gaussien de variance σ^2 , on a la probabilité conditionnelle suivante :

$$p(y_s / x_s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_s - y_s)^2}{2\sigma^2}\right)$$

La probabilité a priori $P(X = x)$ permet d'introduire les contraintes que nous souhaitons imposer à la solution (i.e. que nous supposons pour le processus originel). En faisant l'hypothèse que X est markovien nous nous restreignons à des contraintes locales, le plus souvent de régularité entre sites voisins. On choisit fréquemment un modèle avec des potentiels d'ordre 2 :

$$P(X = x) = \frac{1}{Z} \exp\left(-\beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t)\right)$$

On a alors l'énergie suivante correspondant à la distribution de Gibbs du champ a posteriori :

$$\mathcal{U}(x / y) = \sum_{s \in \mathcal{S}} \frac{(x_s - y_s)^2}{2\sigma^2} + \beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t) \quad (7.3)$$

Le champ X conditionnellement à y est donc un champ de Gibbs pour le même système de voisinage que X . La constante β pondère l'influence entre le terme d'attache aux données (cliques d'ordre 1) qui impose des niveaux de gris x_s de l'image restaurée proches de ceux y_s de la donnée bruitée, et le terme de régularisation (cliques d'ordre 2) qui impose une solution constituée de zones homogènes. Le modèle pour X peut être soit markovien gaussien, soit plus adapté à la restauration des contours avec une fonction ϕ appropriée. En effet, le modèle gaussien qui correspond à un fonction ϕ quadratique favorise des niveaux de gris proches pour des pixels voisins dans tous les cas. Or si on considère une image naturelle cet aspect est néfaste à proximité des contours car il favorisera la présence d'un dégradé. Aussi, de nombreuses fonctions ϕ ont été proposées pour modéliser les potentiels des cliques d'ordre 2 : $U_{c=(s,t)} = \phi(x_s - x_t)$. L'idée est de supprimer la pénalisation lorsque la variation de niveaux de gris est supérieure à une certaine valeur considérée comme représentant un contour. La partie 7.4 détaille ces aspects.

Cas de la segmentation

Dans ce contexte, le champ markovien X est défini sur un autre espace de configurations que Y car seulement quelques étiquettes sont considérées : $E = \{1, \dots, m - 1\}$ (correspondant aux différentes classes cherchées). Dans ce cas le processus de passage de X (champ des labels) à Y ne décrit pas tant le processus d'acquisition que l'apparence des classes dans l'image. Le terme $P(Y = y / X = x)$ traduit donc la probabilité de réalisation d'une configuration donnée connaissant son étiquetage (i.e. connaissant la classe de chaque pixel). En supposant l'indépendance des sites les uns par rapport aux autres, et en supposant que le niveau de gris y_s en un site s ne dépend que de l'étiquette x_s en ce site, on a :

$$P(Y = y / X = x) = \prod_s P(y_s / x_s)$$

Les valeurs des probabilités conditionnelles sont données par l'histogramme conditionnel des niveaux de gris pour une classe donnée. Par exemple, si on suppose que chaque classe i a une distribution gaussienne de moyenne μ_i et d'écart-type σ_i , on a :

$$P(y_s / x_s = i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_s - \mu_i)^2}{2\sigma_i^2}\right)$$

Si comme précédemment on fait une hypothèse markovienne sur X et qu'on se limite aux cliques d'ordre 2, on a :

$$P(X = x) = \frac{1}{Z} \exp\left(-\beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t)\right)$$

D'où l'énergie a posteriori :

$$\mathcal{U}(x / y) = \sum_s \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \log \sqrt{2\pi}\sigma_{x_s} + \beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s, x_t) \quad (7.4)$$

Le champ des étiquettes conditionnellement à y est markovien et d'énergie de Gibbs $\mathcal{U}(x / y)$. Là encore, comme pour la restauration, le terme d'ordre 1 exprime le respect des données (le niveau de gris doit correspondre à la classe), et le terme d'ordre 2 la contrainte de régularisation introduite. On choisit souvent un modèle de Potts pour X , ce qui donne une image segmentée avec de larges zones homogènes.

La figure 7.4 montre un exemple de segmentation d'une image satellitaire obtenue par le radar à ouverture synthétique ERS-1. L'utilisation du modèle de Potts pour le terme d'attache aux données donne des régions compactes.

Dans les deux applications précédentes il est nécessaire de pouvoir déterminer le ou les états d'énergie minimale qui correspondent au maximum de la probabilité d'un champ markovien. L'algorithme du recuit simulé présenté permet de trouver ces configurations. Nous reviendrons sur ce point en présentant d'autres estimateurs de la solution dans la section suivante.

7.2 Estimateurs dans un cadre markovien

7.2.1 Introduction

Nous avons vu précédemment comment il était possible d'utiliser le formalisme markovien à des fins de restauration et de segmentation. On se situe alors dans le cadre de données incomplètes (on parle aussi de champs de Markov cachés) car la réalisation dont on dispose est une réalisation bruitée (ou plus généralement vue à travers le système d'acquisition) du champ de Markov originel. En notant Y le champ dont on observe une réalisation, et X le champ initial, l'objectif est alors d'obtenir la meilleure réalisation \hat{x} de X connaissant l'observation y , autrement dit, reconstruire x de manière optimale vis-à-vis d'un certain critère. Dans le précédent paragraphe, nous nous étions intéressés à la réalisation maximisant la probabilité a posteriori $P(X = x / Y = y)$, et nous avons vu un algorithme permettant d'obtenir cette réalisation : le recuit simulé. En réalité d'autres choix sont possibles, auxquels correspondent d'autres méthodes de résolution, et que nous allons aborder dans ce chapitre.

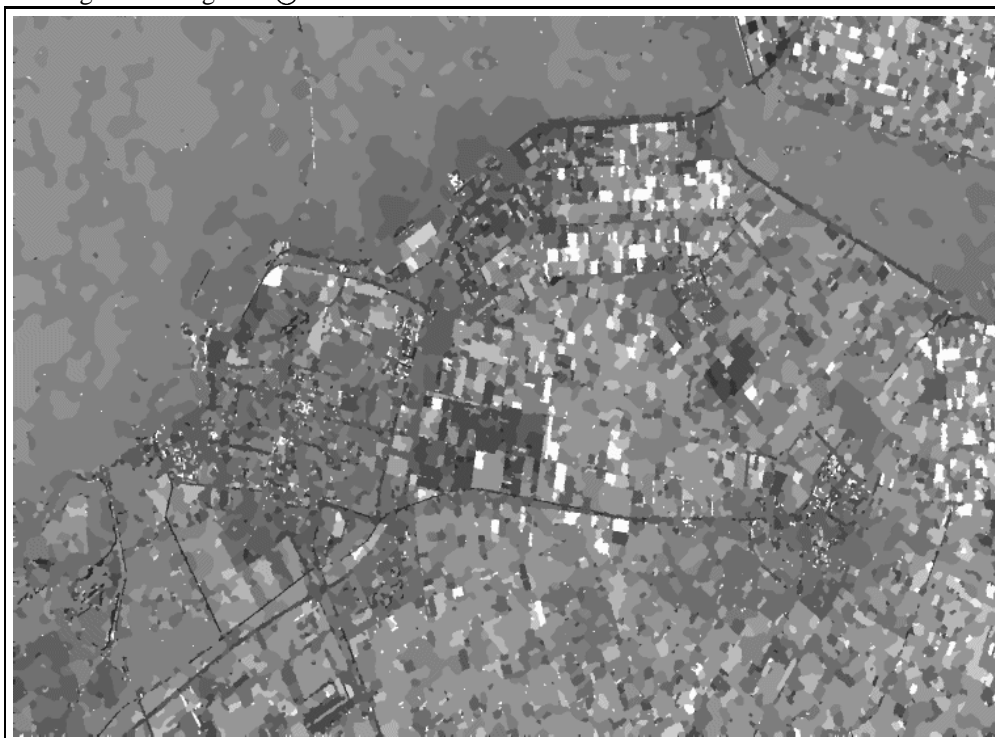
7.2.2 Modélisation bayésienne et fonction de coût

Si nous reprenons rapidement le raisonnement effectué précédemment, on peut écrire, en appliquant la règle de Bayes :

$$P(X = x / Y = y) = \frac{P(Y = y / X = x)P(X = x)}{P(Y = y)}$$



a. Image radar originale ©ERS-1



b. Image segmentée en régions

FIG. 7.4 – Exemple de segmentation markovienne sur une image ERS-1 du Flevoland.

On montre alors que sous certaines hypothèses (indépendance des sites dans la probabilité conditionnelle $P(Y / X)$ et hypothèse markovienne pour le champ X), la distribution a posteriori est une distribution de Gibbs et donc que le champ X conditionnellement à la donnée y est markovien. Cette propriété n'est pas nécessaire pour les notions suivantes, mais elle sera primordiale pour les algorithmes de résolution.

Le problème est alors de déterminer une estimation \hat{x} optimisant un certain critère, où \hat{x} est une fonction déterministe ϕ de la donnée y :

$$\hat{x} = \phi(y) \text{ avec } \phi : \Omega \rightarrow \Omega$$

L'estimation bayésienne procède alors comme suit. On se donne une fonction de coût, L définie de $\Omega \times \Omega$ dans \mathbb{R}^+ , qui représente le coût de remplacer x par $\phi(y)$, et qui possède les propriétés suivantes :

$$\begin{aligned} \forall x, x' \in \Omega \times \Omega : \\ \bullet L(x, x') \geq 0 \\ \bullet L(x, x') = 0 \Leftrightarrow x = x' \end{aligned}$$

L'estimateur optimal, i.e la fonction ϕ optimale est alors la fonction minimisant l'espérance (notée $E[\cdot]$) du coût, c'est à dire :

$$E[L(X, \phi(y)) / Y = y] = \sum_{x \in \Omega} L(x, \phi(y)) P(x / y)$$

La fonction ϕ^{opt} minimise donc l'erreur moyenne conditionnellement à y , et l'estimateur optimal est $\hat{x} = \phi^{\text{opt}}(y)$.

Suivant les fonctions de coût envisagées, on obtient différents estimateurs et différentes méthodes de résolution associées.

7.2.3 Estimateur MAP

Considérons la fonction de coût suivante :

$$\begin{aligned} L(x, x') &= 1 \text{ si } x \neq x' \\ L(x, x') &= 0 \text{ sinon} \end{aligned}$$

Cette fonction consiste donc à pénaliser toute différence entre deux configurations, et ce, quel que soit le nombre de sites en lesquels elles diffèrent. Nous pouvons alors écrire :

$$\begin{aligned} E[L(X, \phi(y)) / y] &= \sum_{x \in \Omega} L(x, \phi(y)) P(x / y) \\ &= 1 - P(X = \phi(y) / y) \end{aligned}$$

Par conséquent, la fonction ϕ^{opt} minimisant l'espérance pour cette fonction de coût est celle qui maximise la probabilité a posteriori :

$$\hat{x} = \phi^{\text{opt}}(y) = \text{Argmax}_{\phi} [P(X = \phi(y) / y)]$$

Il nous faut donc trouver la réalisation \hat{x} , fonction de y , maximisant la probabilité a posteriori $P(X / y)$. On parle de l'estimateur MAP (maximum a posteriori) ou de maximum de vraisemblance a posteriori.

On retrouve donc, avec cette fonction de coût en tout ou rien, la démarche intuitive que nous avons présentée dans la partie précédente 7.1.7 à savoir chercher la configuration maximisant la probabilité conditionnellement à la donnée disponible.

Les solutions algorithmiques associées à cet estimateur sont le recuit simulé et l'ICM que nous avons présentés dans le paragraphe 7.1.5, utilisés avec la distribution a posteriori avec paramètre de température.

7.2.4 Estimateur MPM

Considérons maintenant la fonction de coût définie par :

$$L(x, x') = \sum_{s \in S} L(x_s, x'_s) = \sum_{s \in S} \mathbf{I}_{x_s \neq x'_s}$$

La fonction de coût pénalise cette fois-ci une configuration proportionnellement au nombre de différences entre deux configurations. Elle paraît donc plus naturelle que la fonction de coût en tout ou rien précédente.

Dans le cas d'une fonction de coût définie comme somme de coûts en chaque site, on peut montrer le résultat suivant :

$$\begin{aligned} E[L(X, \phi(y)) / Y = y] &= \sum_{x \in \Omega} L(x, \phi(y)) P(x / y) \\ &= \sum_{s \in S} \sum_{x_s} L(x_s, \phi(y)_s) \sum_{x^s} P(x^s, x_s / y) \end{aligned}$$

Or $\sum_{x^s} P(x^s, x_s / y) = P(X_s = x_s / y)$, donc on peut faire apparaître les probabilités conditionnelles et espérances en chaque site s :

$$\begin{aligned} E[L(X, \phi(y)) / Y = y] &= \sum_{s \in S} \sum_{x_s} L(x_s, \phi(y)_s) P(X_s = x_s / y) \\ &= \sum_{s \in S} E[L(X_s, \phi(y)_s) / y] \end{aligned}$$

On passe donc de la probabilité conditionnelle globale d'une configuration à la probabilité conditionnelle en un site. Il s'agit d'une somme de termes positifs, et par conséquent la fonction ϕ optimale minimise en chaque site l'espérance conditionnelle du coût local $E[L(X_s, \phi(y)_s) / y]$. Ce résultat est valable pour toutes les fonctions de coût définies par une somme de coûts en chaque site.

Dans le cas de la fonction définie ci-dessus, on a alors comme précédemment :

$$E[L(X_s, \phi(y)_s) / y] = 1 - P(X_s = \phi(y)_s / y)$$

Ainsi, la valeur optimale de $\phi(y)$ ou de \hat{x} en chaque site est telle que :

$$\hat{x}_s = \phi^{\text{Opt}}(y)_s = \text{Argmax}_{\phi} [P(X_s = \phi(y)_s / y)]$$

i.e. on maximise en chaque site la marginale a posteriori $P(X_s / y)$.

On obtient donc des estimateurs du maximum a posteriori locaux, à calculer en chaque pixel contrairement à la recherche précédente qui était globale. L'estimateur est appelé maximum de vraisemblance a posteriori local ou maximum posterior marginal (pour maximum a posteriori de la marginale) abrégé en MPM.

D'un point de vue algorithmique, la taille de l'espace des configurations Ω ne permet pas un calcul direct des quantités $P(x_s / y)$. Aussi réalise-t-on en pratique des approximations de type Monte-Carlo. En effet, supposons que l'on soit capable de tirer des réalisations de X selon sa loi conditionnelle à y , et notons les $x(1), \dots, x(N)$. Il est alors possible de calculer une approximation de l'estimateur MPM. Le tirage des réalisations ne pose quant à lui pas de problème particulier car nous avons vu dans le paragraphe précédent 7.1.4 comment tirer des réalisations d'un champ de Gibbs avec l'échantillonneur de Gibbs et l'algorithme de Métropolis. Or sous les hypothèses rappelées en début de la section 7.2.1, la probabilité a posteriori $P(X / y)$ est une distribution de Gibbs.

Supposons donc que nous disposions de N échantillons de X tirés selon la loi a posteriori et que nous cherchions à estimer la distribution conditionnelle en chaque site $P(X_s = \lambda / y) \forall \lambda \in E$. Nous allons estimer cette quantité par la fréquence empirique de λ au site s dans les échantillons $x^{(k)}$ de X , i.e :

$$\hat{P}(X_s = \lambda / y) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{x^{(k)}_s = \lambda}$$

L'estimation au sens du MPM est alors donnée en chaque site en choisissant la valeur de x_s dans Ω maximisant $P(X_s / y)$.

7.2.5 Estimateur TPM

Considérons maintenant la fonction de coût définie par :

$$L(x, x') = \|x - x'\|^2 = \sum_{s \in S} (x_s - x'_s)^2$$

Il s'agit de l'erreur quadratique et elle pénalise cette fois-ci directement la somme des différences entre les deux configurations. Elle peut donc être plus adaptée dans certains cas que les précédentes, puisqu'elle tient compte non seulement du nombre de différences comme le MPM, mais aussi de leurs valeurs.

Dans ce cas, on a en utilisant le résultat établi précédemment :

$$E[L(X, \phi(y)) / Y = y] = \sum_{s \in S} E[(X_s - \phi(y)_s)^2 / y]$$

On cherche donc ϕ , telle que $E[(X_s - \phi(y)_s)^2 / y]$ soit minimum. Nous allons écrire cette espérance sous une nouvelle forme en utilisant la moyenne conditionnelle au site s , $\bar{x}_s = E[X_s / y] = \sum_{x_s \in E} x_s P(x_s / y)$:

$$\begin{aligned} E[(X_s - \phi(y)_s)^2 / y] &= \sum_{x_s \in E} (x_s - \bar{x}_s)^2 P(x_s / y) + \sum_{x_s \in E} (\bar{x}_s - \phi(y)_s)^2 P(x_s / y) \\ &= K + (\bar{x}_s - \phi(y)_s)^2 \end{aligned}$$

où K est une constante ne dépendant pas de ϕ donc n'intervenant pas dans la minimisation. Par conséquent, le minimum de l'erreur est atteint pour la fonction ϕ telle que :

$$\phi^{\text{opt}}(y)_s = E[X_s / y]$$

Cet estimateur consiste à prendre en chaque site la moyenne conditionnelle locale donnée par la loi a posteriori, d'où le nom de TPM (*Thresholded Posteriori Mean*).

D'un point de vue algorithmique, la démarche est similaire à celle effectuée dans le paragraphe précédent pour le MPM. On approxime l'espérance conditionnelle en chaque site par la moyenne empirique en ce site des N échantillons tirés selon la loi a posteriori :

$$\hat{E}(X_s / y) = \frac{1}{N} \sum_{k=1}^N x^{(k)}_s$$

L'estimation au sens du TPM est alors donnée en chaque site par sa moyenne empirique. Remarquons que cet estimateur est mal adapté à une problématique de segmentation car la moyenne des étiquettes n'a alors aucun sens.

7.2.6 Comparaison des estimateurs MAP, MPM, et TPM

Nous comparons dans cette section les trois estimateurs dans le cadre de la restauration. Dans le cas du MAP, les résultats sont obtenus par ICM et par recuit simulé. L'image à restaurer (figure 7.5.a) est une image bruitée par un bruit blanc gaussien. L'énergie a posteriori utilisée s'écrit :

$$\mathcal{U}(x / y) = \sum_s \frac{(y_s - x_s)^2}{\sigma^2} + \sum_{c \in \mathcal{C}} \beta \phi(x_r - x_s) \text{ avec } \beta > 0$$

$$\text{et } \phi(u) = \frac{u^2}{1 + u^2}$$

Comme indiqué dans le paragraphe 7.1.7, cette fonction permet de seuilier les pénalités imposées par le terme de régularisation en présence de contours dans l'image.

Les paramètres utilisés sont fixés aux valeurs suivantes : $\sigma = 28$ (écart-type du bruit), $\delta = 10$ (saut en amplitude à partir duquel on considère qu'il y a un contour), $\beta = 0,5$ pondération de l'influence relative des deux termes). L'initialisation est donnée par l'image à restaurer. Pour tous les estimateurs, on réalise 600 itérations. Dans le cas du recuit simulé, la température initiale est de 6. Les résultats sont montrés sur la figure 7.5.

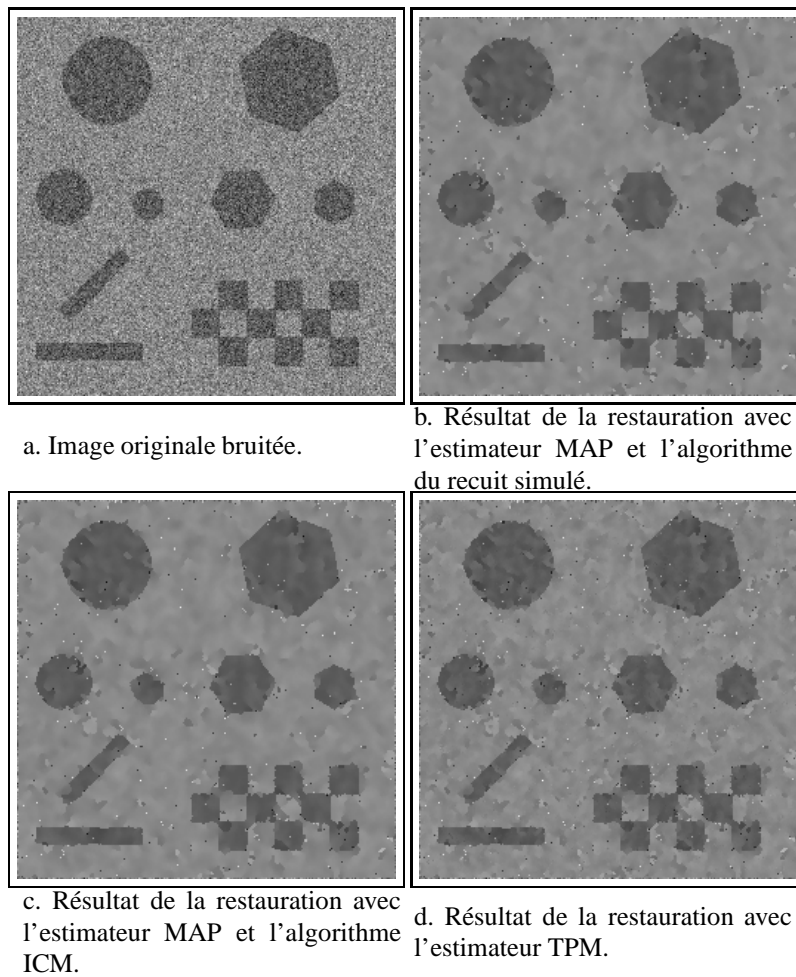


FIG. 7.5 – Comparaison des algorithmes ICM, Recuit Simulé et TPM en restauration.

On constate visuellement que le meilleur résultat est obtenu par le MAP du recuit simulé. Comme l'image originale est une assez bonne initialisation, il n'y a pas de grandes différences entre les algorithmes de recuit simulé et d'ICM pour l'estimateur du MAP. Ce n'est pas vrai dès que l'initialisation s'écarte du résultat à obtenir et les différences avec le recuit simulé peuvent être très importantes. Par ailleurs, on constate que l'estimateur TPM, qui est par définition plus local, donne un résultat plus bruité et moins régularisé. Cette analyse visuelle est confirmée par l'étude statistique qui peut être effectuée sur des zones homogènes de l'image. Le tableau ci-dessous donne les statistiques d'une zone sombre et d'une zone claire de l'image originale et pour les différents résultats de restauration. Les écarts-types les plus faibles sont obtenus pour l'estimateur MAP.

Statistiques sur l'image originale		
zone 1	135.7	27.8
zone 2	90.9	27.1
Estimateur MAP ICM		
zone 1	136.7	11.5
zone 2	91.6	6.9
Estimateur TPM		
zone 1	136.2	11.3
zone 2	91.6	6.9
Estimateur MAP recuit simulé		
zone 1	136.2	10.3
zone 2	91.9	6.1

TAB. 7.1 – Statistiques sur des zones homogènes pour les différents résultats de restauration

On notera également que des points isolés de faible ou fort niveau de gris subsistent dans l'image restaurée. Cela est lié à l'utilisation de la ϕ fonction qui ne régularise plus au-delà d'un certain seuil contrôlé par la valeur de δ .

En ce qui concerne les temps de calcul, les méthodes se répartissent comme suit : l'algorithme le plus rapide pour converger est sans conteste l'ICM, les algorithmes de recuit simulé et de TPM (ou MPM) étant à peu près équivalents. En effet, plus le nombre d'itérations est grand, meilleure est l'estimation de la moyenne a posteriori.

Les conclusions qui sont données ici ne sont pas nécessairement valables pour une application en segmentation. L'estimateur du TPM peut en effet dans certains cas donner de meilleurs résultats que le MAP. Par ailleurs, l'ICM peut s'avérer très utile lorsqu'on connaît une configuration proche de la configuration optimale.

7.3 Estimation des paramètres

7.3.1 Introduction

Le problème de l'estimation de paramètres (encore appelés hyperparamètres dans la littérature), revient très fréquemment en traitement d'image par champs de Markov. Donnons-en plusieurs exemples :

1. On se donne une réalisation d'un champ de Markov associé au modèle d'Ising, mais on ne connaît pas ses paramètres. Quels sont-ils ?
2. On veut généraliser ceci à une image de texture donnée dont on connaît le modèle sous-jacent (par exemple un modèle gaussien en 4-connexité) mais pas les paramètres, qui sont du type : moyenne locale, variance locale, poids de la régularisation locale. Quels sont-ils ? Leur connaissance pourrait bien en effet servir à la classification d'images composées de zones texturées en se basant sur l'estimation locale de tels paramètres. On classierait alors selon les valeurs de ces attributs locaux.

3. On veut segmenter une image, et pour cela apprendre les paramètres de chaque classe, ainsi que le coefficient optimal du modèle de régularisation adapté à cette tâche. On sait en effet que le résultat d'une segmentation par estimateur MAP par exemple dépend fondamentalement du poids respectif de la régularisation par rapport à celui de l'attache aux données. Il faut donc là aussi estimer ce poids d'une façon optimale dans un sens à définir.

L'ensemble forme un problème réputé difficile. Nous ne démentirons pas ici la difficulté de ces problèmes, accentuée par le fait que de très nombreuses variantes ont été mises au point dans la littérature pour le résoudre. Certaines tentatives de comparaison de ces approches ont déjà été effectuées [Pieczynski, 1994]. Nous exposerons ici les variantes les plus fréquemment utilisées.

Elles se décomposent en deux classes fondamentales :

- le cas des données dites complètes, correspondant aux deux premiers problèmes cités plus haut : un échantillon d'une distribution de Gibbs est connu. Il s'agit de remonter aux paramètres de cette distribution.
- le cas des données dites incomplètes. Là non seulement le résultat de traitement est inconnu, mais les paramètres sont également à estimer.

7.3.2 Données complètes

Notons dans ce qui va suivre x une configuration observée relativement à une distribution de Gibbs donnée P_θ dont l'énergie associée puisse s'écrire sous la forme d'une fonction linéaire d'un paramètre θ , par exemple $U(x) = \theta \Phi(x)$, où Φ est un potentiel donné. Un principe naturel en vue de la recherche de θ est d'écrire la vraisemblance de la donnée x :

$$L(\theta) = P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta}$$

et de chercher par exemple la valeur de l'hyperparamètre $\hat{\theta}$ maximisant cette vraisemblance $L(\theta)$. Le problème essentiel est que l'on ne sait en général pas calculer exactement la fonction de partition Z_θ . Même pour des modèles aussi simples et fondamentaux que ceux d'Ising et de Potts, le résultat (analytique) est obtenu après des calculs excessivement compliqués [Onsager, 1944, Landau et Lifschitz, 1961]. Dans les autres cas on sera donc amené :

- soit à effectuer des approximations de la fonction de partition globale au moyen des fonctions de partition conditionnelles locales (codages paragraphe 7.3.2, pseudo-vraisemblance paragraphe 7.3.2)
- soit à employer des algorithmes itératifs (gradient stochastique paragraphe 7.3.2) à partir de la vraisemblance exacte, mais dont il s'agit alors de prouver la convergence ainsi que le type d'optimum trouvé (local, global).

Méthode des codages

Le principe de la méthode des codages [Besag, 1974] est le suivant. Une fois défini un système de voisinage pour un champ de Markov, nous sommes capables de définir un certain nombre de sous réseaux, chacun formé de sites/pixels indépendants les uns des autres : chacun de ces sous réseaux est appelé un codage. Par exemple avec un voisinage en 4 connexité il existe deux codages comme le montre la figure ci dessous (à gauche), et 4 codages différents dans le cas de la 8-connexité (à droite) :

					3	4	3	4	3
2	1	2	1	2	2	1	2	1	2
1	2	1	2	1	3	4	3	4	3
2	1	2	1	2	2	1	2	1	2
					3	4	3	4	3

Nous allons poser le problème d'estimation dans le cadre de chaque codage. Pour un codage donné les différents sites/pixels le constituant sont indépendants les uns des autres puisqu'ils ne sont pas des voisins pour le champ de Markov de départ. La probabilité globale d'un codage se trouvera donc être le produit des probabilités individuelles de chacun des sites du codage. Or du fait de la structure de Markov du champ de départ cette probabilité individuelle se trouve être la probabilité conditionnelle locale du site/pixel dans le champ de Markov. Pour un codage Cod_n donné nous pouvons donc écrire :

$$P_\theta(\{X_s = x_s\}_{s \in \text{Cod}_n} / \{X_r = x_r\}_{r \notin \text{Cod}_n}) = \prod_{s \in \text{Cod}_n} P_\theta(X_s = x_s / V_s) \quad (7.5)$$

Etant donné la structure des probabilités locales telles qu'elles ont été décrites dans les chapitres précédents la fonction de vraisemblance devient calculable, puisque les fonctions de partition conditionnelles locales le sont. Dans le cas où la dépendance des énergies locales est linéaire vis-à-vis des paramètres, la log-vraisemblance associée,

$$\log P_\theta(\{X_s = x_s\}_{s \in \text{Cod}_n} / \{X_r = x_r\}_{r \notin \text{Cod}_n}) = -\theta \sum_{c \in \mathcal{C}} U_c(x) - \sum_{s \in \text{Cod}_n} \log(Z_s)$$

est une fonction concave du (des) paramètre(s), car somme de fonctions concaves [Sigelle, 1997]. Elle se prête donc bien à la recherche d'un optimum par une méthode classique de type gradient. On peut également montrer qu'il s'agit dans ce cas d'un simple problème de moindres carrés [Derin et al., 1985].

Pseudo-vraisemblance

Il apparaît en fait expérimentalement que la méthode des codages n'est pas fiable. La méthode du maximum de vraisemblance vrai paraît quant à elle incalculable. Des algorithmes ont cependant été étudiés pour tenter de résoudre ce problème [Younes, 1988], voir paragraphe 7.3.2. En fait nous allons utiliser une méthode intermédiaire qui aura de bonnes propriétés : la méthode du pseudo-maximum de vraisemblance [Graffigne, 1987]. Du maximum de vraisemblance vrai, nous allons prendre l'idée de travailler sur l'ensemble de l'image et non séparément sur des réseaux indépendants. De la méthode de codage, nous conservons l'idée de manipuler une fonction de vraisemblance produit des probabilités locales de chacun des sites/pixels. Cette fonction sera appelée pseudo-maximum de vraisemblance, et elle s'écrira :

$$PL_\theta(X = x) = \prod_{s \in S} P(X_s = x_s / V_s) \quad (7.6)$$

ou, encore, en considérant le logarithme de cette fonction, et l'expression de la probabilité locale de chaque site/pixel :

$$\log PL_\theta(X = x) = -\theta \sum_{c \in \mathcal{C}} U_c(x) - \sum_{s \in S} \log(Z_s) \quad (7.7)$$

Maintenant l'expression $-\log(Z_\theta)$ devient calculable, puisque reliée à la fonction de normalisation de la probabilité conditionnelle locale telle qu'elle a été décrite dans le paragraphe 7.1.3. Donc un raisonnement identique à celui du paragraphe précédent conduit au fait que la log-pseudo-vraisemblance est une fonction concave des paramètres lorsque l'énergie en dépend de façon linéaire. Les algorithmes usuels de type gradient ou gradient conjugué s'appliquent donc aussi ici naturellement à la recherche de l'optimum (qui est unique comme précédemment).

Il s'agit alors de qualifier la valeur des paramètres obtenus par cette méthode par rapport à la valeur vraie. Des résultats théoriques importants ont été obtenus à ce sujet [Graffigne, 1987, Guyon, 1992] : la méthode de la pseudo-vraisemblance est consistante et convergente.

Algorithme du gradient stochastique

Partons de la vraisemblance exacte du paramètre. Elle s'écrit bien sûr :

$$L(\theta) = P_\theta(x) = \frac{\exp -\theta \Phi(x)}{Z_\theta}$$

La valeur de l'hyperparamètre satisfaisant au principe du maximum de vraisemblance $\hat{\theta} = \arg \max_{\theta} P_\theta(x)$ doit donc vérifier l'équation :

$$\left(\frac{\partial \log P_\theta(x)}{\partial \theta} \right)_{\hat{\theta}} = -\Phi(x) - \left(\frac{\partial \log Z_\theta}{\partial \theta} \right)_{\hat{\theta}} = 0 \quad (7.8)$$

Comme $E_{\hat{\theta}}[\Phi] = - \left(\frac{\partial \log Z_\theta}{\partial \theta} \right)_{\hat{\theta}}$ [Sigelle, 1997], cette valeur optimale est donc unique et doit satisfaire l'équation suivante :

$$E_{\hat{\theta}}[\Phi] = \Phi(x) \quad (7.9)$$

Il s'agit là de ce que l'on appelle une équation stochastique. Comme remarqué en introduction de la Section 7.3.2 cette équation ne peut être résolue exactement, pour la raison que $E_{\theta}[\Phi]$, qui dérive de la fonction de partition Z_θ , ne peut en général être calculé de façon analytique exacte. On est donc amené à employer des algorithmes basés sur un schéma itératif de type Newton-Raphson, mais adapté à ce cadre stochastique. Un schéma rigoureux conduirait à chaque étape (n) à :

$$\theta_{n+1} = \theta_n - \frac{E_{\theta_n}[\Phi] - \Phi(x)}{\left(\frac{\partial (E_{\theta}[\Phi])}{\partial \theta} \right)_{\theta_n}}$$

c'est-à-dire au schéma itératif :

$$\theta_{n+1} = \theta_n + \frac{E_{\theta_n}[\Phi] - \Phi(x)}{\text{var}_{\theta_n}(\Phi)}$$

L'idée est alors de remplacer les grandeurs statistiques mises en jeu par leurs valeurs empiriques approchées. Ainsi pour l'espérance du potentiel de régularisation Φ , on prendra sa moyenne empirique au cours d'une seule itération (c'est à dire la valeur effective obtenue !) d'un échantillonneur de Gibbs ou de Metropolis mené avec la valeur courante du paramètre. Quant à la variance de ce potentiel, on l'estime encore plus crûment par une grandeur positive fixée V ! On peut montrer que le prix à payer pour cette approximation est l'introduction d'un terme correctif supplémentaire en $\frac{1}{n+1}$ dans le schéma itératif à l'itération (n). C'est le principe de l'algorithme de gradient stochastique [Younes, 1988] :

$$\begin{cases} \theta_0 & \text{arbitraire} \\ x^{(0)} & \text{tiré au hasard} \end{cases}, \quad \theta_{n+1} = \theta_n + \frac{\Phi(x^{(n)}) - \Phi(x)}{(n+1)V} \quad \text{pour } n \geq 1 \quad (7.10)$$

Il est très important de noter ici que $x^{(n)}$ ($n \geq 1$), échantillon de la distribution P_{θ_n} obtenu par une dynamique de Gibbs (ou de Metropolis) associée à la valeur courante du paramètre θ_n , est généré à partir de l'échantillon $x^{(n-1)}$

obtenu à l'itération précédente (c'est à dire lors de la valeur précédente du paramètre). On peut alors montrer que cet algorithme stochastique converge presque sûrement, en termes de probabilité, vers la valeur optimale $\hat{\theta}$ lorsque le coefficient V est choisi suffisamment grand.

7.3.3 Données incomplètes

Dans cette section, nous abordons le problème de l'estimation des paramètres dans le cas des données incomplètes dites encore manquantes.

Dans ce cas nous connaissons une observation y , échantillon de la v.a. Y , mais elle est appelée incomplète (ou dégradée), car reliée à une scène originale x , non-dégradée, dont le champ aléatoire correspondant sera noté X . La relation entre y et x s'effectue via une loi de probabilité conditionnelle représentant l'attache aux données (cf. 7.1.7) dont nous explicitons la dépendance p.r. à un paramètre λ positif :

$$P_\lambda(Y = y / X = x) = \frac{\exp - U_\lambda(y / x)}{Z_\lambda}$$

Ainsi, dans le cas d'un bruit blanc gaussien additif en restauration ou en déconvolution et l'approximation discrète finie, on a :

$$P_\lambda(Y = y / X = x) = \frac{\exp - \lambda \|y - Rx\|^2}{Z_\lambda}$$

où R est la matrice associée à la réponse impulsionnelle de la fonction de flou (l'identité en restauration) et $Z_\lambda = \left(\sqrt{\frac{\pi}{2\lambda}}\right)^{|S|}$. On notera que la fonction de partition $Z_\lambda = \sum_{y \in \Omega} \exp - U_\lambda(y / x)$ est ici indépendante de la variable cachée x .

On suppose aussi que l'on dispose d'une connaissance a priori sur la scène à retrouver x , que ce soit en segmentation ou restauration, via la distribution de Gibbs suivante :

$$P_\theta(x) \frac{\exp - \theta \Phi(x)}{Z_\theta}$$

Nous aurons besoin dans la suite de ce chapitre d'utiliser la distribution de Gibbs a posteriori :

$$P_{\theta,\lambda}(X = x / Y = y) = \frac{\exp - U_\lambda(y / x) - \theta \Phi(x)}{Z_{\theta,\lambda}}$$

de fonction énergie $\mathcal{U}(x / y) = U_\lambda(y / x) + \theta \Phi(x)$, et de fonction de partition associée $Z_{\theta,\lambda}$.

Nous commencerons par généraliser, dans un cadre de Maximum de Vraisemblance (MV), la méthode de gradient stochastique vue au paragraphe précédent lorsque la loi d'observation (attache aux données) est complètement connue, c'est-à-dire que nous nous focaliserons sur l'estimation du meilleur paramètre de régularisation θ . Puis nous comparerons cette méthode à des variantes importantes similaires répertoriées dans la littérature.

Nous aborderons ensuite l'estimation des paramètres d'attache aux données, en particulier dans le cadre de la segmentation d'images. Cela nous permettra de décrire ensuite la seconde grande classe de méthodes adaptée à l'estimation des paramètres : l'EM (Expectation-Maximisation).

Gradient stochastique généralisé [Younes, 1989]

On va prouver dans cette partie que l'estimation du paramètre de régularisation connaissant la forme du potentiel a priori ne peut être dissociée dans la plupart des cas de la forme de l'attache aux données, supposée connue ici par simplicité, c'est à dire que λ est connu. On suppose accéder par exemple d'une façon ou d'une autre à la variance d'un bruit gaussien en restauration d'image bruitée (et à la réponse impulsionnelle du flou si l'on est en

déconvolution).

Dans le cas supposé où aucune information *a priori* sur le paramètre de régularisation n'est disponible, c'est-à-dire lorsque θ suit la distribution uniforme sur \mathbb{R} , la vraisemblance de ce paramètre est la grandeur adéquate à étudier connaissant l'observation incomplète y et le paramètre λ . Elle peut se calculer de façon élégante [Peyrard, 2001] en remarquant que la loi jointe de l'observation et des variables cachées s'écrit :

$$\begin{aligned} P_{\theta,\lambda}(X = x, Y = y) &= P_\lambda(X = x / Y = y)P_\theta(X = x) \\ &= \frac{\exp - U_\lambda(y / x)}{Z_\lambda} \cdot \frac{\exp - \theta \Phi(x)}{Z_\theta} = \frac{\exp - U_\lambda(y / x) - \theta \Phi(x)}{Z_\lambda \cdot Z_\theta} \end{aligned}$$

Si au moins une valeur de x n'annule pas $P(Y = y / X = x)$, on en déduit que :

$$L(\theta) = P_{\theta,\lambda}(Y = y) = \frac{P_{\theta,\lambda}(Y = y, X = x)}{P_{\theta,\lambda}(X = x / Y = y)} = \frac{Z_{\theta,\lambda}}{Z_\lambda \cdot Z_\theta} \quad (7.11)$$

Toute valeur optimale du paramètre de régularisation $\hat{\theta}$ satisfait donc l'équation suivante :

$$\left(\frac{\partial \log L(\theta)}{\partial \theta} \right)_{\hat{\theta}} = E_{\hat{\theta}}[\Phi] - E_{\hat{\theta},\lambda}[\Phi] = 0 \Rightarrow E_{\hat{\theta}}[\Phi] = E_{\hat{\theta},\lambda}[\Phi] \quad (7.12)$$

où l'on rappelle que $E_\theta[\cdot]$ est l'espérance d'une v.a. sous la distribution de Gibbs a priori d'énergie $\theta \Phi(x)$, tandis que $E_{\theta,\lambda}[\cdot]$ signifie l'espérance statistique sous la distribution de Gibbs a posteriori. Notons également que les deux grandeurs statistiques $E_{\theta,\lambda}[\Phi]$ et $E_\theta[\Phi]$ sont des fonctions monotones décroissantes de θ , ce qui implique que plusieurs valeurs optimales de l'hyperparamètre peuvent exister (voir fig.7.3.3).

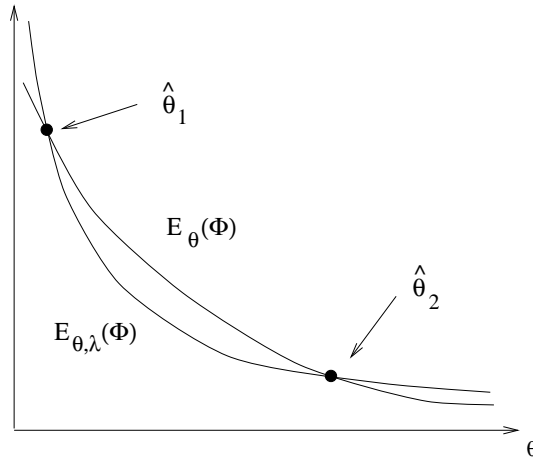


FIG. 7.6 – Valeur(s) optimale(s) $\hat{\theta}$ du paramètre de régularisation en données incomplètes.

Implémentation : Si l'on applique un schéma de Newton-Raphson à l'équation stochastique (7.12) on obtient directement :

$$\theta_{n+1} = \theta_n - \frac{E_{\theta_n}[\Phi] - E_{\theta_n,\lambda}[\Phi]}{\left(\frac{\partial(E_{\theta_n}[\Phi] - E_{\theta_n,\lambda}[\Phi])}{\partial \theta_n} \right)} = \theta_n + \frac{E_{\theta_n}[\Phi] - E_{\theta_n,\lambda}[\Phi]}{\text{var}_{\theta_n}(\Phi) - \text{var}_{\theta_n,\lambda}(\Phi)} \quad (7.13)$$

Il faut d'abord noter que dans ce schéma le dénominateur du dernier terme peut être de signe quelconque contrairement au cas des données complètes (7.10), ce qui est relié à l'existence de plusieurs solutions possibles en données

incomplètes. Maintenant, en s'inspirant du raisonnement effectué pour le gradient stochastique pour les données complètes, on s'aperçoit qu'il est nécessaire d'approximer des quantités statistiques (moyenne, variance) reliées aux distributions a posteriori et a priori. On a donc besoin d'échantillonner deux champs de Markov en général : celui lié à la régularisation pure (champ a priori) et le champ de Markov postérieur comprenant l'attachement aux données. Le schéma empirique qui en résulte naturellement est donc le suivant [Younes, 1989] :

$$\begin{cases} \text{un échantillon } x^{(n)} \text{ généré par } P_{\theta_n} \text{ (distribution a priori)} \\ \text{un échantillon } \tilde{x}^{(n)} \text{ généré par } P_{\theta_n, \lambda} \text{ (distribution a posteriori)} \end{cases}$$

avec la procédure d'évolution de l'hyperparamètre :

$$\theta_{n+1} = \theta_n + \frac{1}{n} \cdot \frac{\Phi(x^{(n)}) - \Phi(\tilde{x}^{(n)})}{(\langle \text{var}_{\theta_n}(\Phi) \rangle - \langle \text{var}_{\theta_n, \lambda}(\Phi) \rangle)}$$

Il faut noter en particulier la présence des estimateurs empiriques des variances $\langle \text{var}_{\theta_n}(\Phi) \rangle$ et $\langle \text{var}_{\theta_n, \lambda}(\Phi) \rangle$ (qui nécessitent donc en fait au moins deux échantillons pour chacune des distributions et à chaque étape du schéma d'évolution de l'hyperparamètre). Ainsi en segmentation, on doit échantillonner aussi bien la distribution a posteriori, ce qui sera par ailleurs utile en vue de la segmentation à convergence des paramètres par l'un des estimateurs MPM ou TPM, que la distribution a priori qui correspond au modèle de régularisation pur : modèles d'Ising, de Potts, généralisation, et ceci pour la valeur courante du (des) hyperparamètre(s).

La convergence de cette méthode est à montrer dans le cadre général des algorithmes stochastiques [Métivier et Priouret, 1987], qui dépasse largement l'objet de ce chapitre.

Comparaison avec d'autres variantes d'estimation

Dans ce paragraphe on présente d'autres méthodes usuelles d'estimation des hyperparamètres en données incomplètes, qui sont en fait reliées à d'autres estimateurs que le maximum de vraisemblance. Toute la subtilité réside ici dans le choix de la fonction de type vraisemblance à maximiser. On va voir en effet que de très légères différences conduisent à des formes fort différentes d'équations stochastiques, et donc des résultats d'estimation a priori fort différents.

On suppose dans les deux premiers cas qu'un résultat "optimal" de restauration (ou segmentation) x^* est connu à l'étape d'estimation où l'on se situe.

Une première approche consiste à utiliser la loi jointe de l'observation et du résultat connaissant la valeur courante des hyperparamètres [Lakshmanan et Derin, 1989a, Descombes et al., 1999]. On peut alors écrire que le paramètre optimal doit satisfaire :

$$\hat{\theta} = \arg \max_{\theta} P_{\theta, \lambda}(X = x^*, Y = y)$$

En utilisant comme précédemment le fait que l'attache aux données (resp. la régularisation) est indépendante de l'hyperparamètre θ (resp. de λ), on obtient :

$$P_{\theta, \lambda}(X = x^*, Y = y) = P_{\lambda}(Y = y / X = x^*) P_{\theta}(X = x^*)$$

Et comme le premier terme (attache aux données) ne dépend pas de θ ,

$$\hat{\theta} = \arg \max_{\theta} P_{\theta}(X = x^*) = \frac{\exp - \theta \Phi(x^*)}{[Z_{\theta} = \sum_{x \in \Omega} \exp - \theta \Phi(x)]}$$

On retombe alors sur l'estimateur au sens du maximum de vraisemblance pour le champ de Markov a priori d'énergie $\theta \Phi(x)$ et pour la donnée ici complète x^* , c'est à dire : $E_{\hat{\theta}}[\Phi] = \Phi(x^*)$. On peut donc utiliser une technique de gradient stochastique classique pour estimer ce paramètre (cf. paragraphe 7.3.2). On peut ensuite itérer

en effectuant le traitement désiré (restauration, segmentation) avec les nouvelles valeurs des hyperparamètres obtenues. On obtient donc une nouvelle configuration optimale à partir de laquelle on peut re-estimer les paramètres, et ainsi de suite [Khoumri, 1997, Zerubia et Blanc-Féraud, 1998]. La méthode est théoriquement convergente vers les valeurs optimales des paramètres et de la configuration.

Une seconde approche consiste à utiliser la probabilité du résultat conditionnellement à l'observation et aux hyperparamètres. Dans ce cas, le paramètre optimal doit vérifier :

$$\hat{\theta} = \arg \max_{\theta} P_{\theta, \lambda}(X = x^* / Y = y)$$

En appliquant la seconde formule de Bayes et en utilisant le même argument d'indépendance des différentes lois de probabilités vis-à-vis des différents hyperparamètres mis en jeu, on obtient :

$$\hat{\theta} = \arg \max_{\theta} \frac{\exp - \lambda U(y / x^*) - \theta \Phi(x^*)}{[Z_{\theta, \lambda} = \sum_{x \in \Omega} \exp - \lambda U(y / x) - \theta \Phi(x)]}$$

On tombe cette fois sur l'estimateur au sens du maximum de vraisemblance pour la distribution de Gibbs a posteriori d'énergie $\theta \Phi(x) + \lambda U(y / x)$ et pour la donnée complète x^* , c'est-à-dire : $E_{\hat{\theta}, \lambda}[\Phi] = \Phi(x^*)$. Les deux méthodes aboutissent donc à l'estimation d'un paramètre unique, et peuvent donc être implémentées avec une technique de gradient stochastique classique [Younes, 1988, Khoumri, 1997]. Nous citerons également l'alternative des méthodes MCMCML (Monte Carlo Markov Chains Maximum Likelihood) [Descombes et al., 1999].

Il est très intéressant de comparer les deux résultats précédents avec le véritable gradient stochastique généralisé vu précédemment. On rappelle que celui-ci permet de maximiser la probabilité du paramètre de régularisation conditionnellement à l'observation uniquement [Younes, 1989] :

$$\hat{\theta} = \arg \max_{\theta} P_{\theta, \lambda}(Y = y)$$

En toute rigueur, il est l'estimateur exact au sens du MV du paramètre de régularisation. Les deux autres cités précédemment ont essentiellement l'avantage d'être plus rapides, car ne nécessitant qu'un échantillonnage à chaque étape, et permettant de faire alterner le traitement d'image désiré avec l'estimation des paramètres.

En pratique, les trois estimateurs cités aboutissent à des valeurs du paramètre de régularisation semblables, ce qui provient du fait que les modèles adoptés pour l'attache aux données et pour la régularisation sont souvent choisis en cohérence avec l'observation fournie.

Estimation des hyperparamètres d'attache aux données : le cas de la segmentation

Supposons que l'on veuille estimer également les paramètres intervenant dans l'attache aux données, que nous regrouperons sous une variable λ . Ainsi, dans le cas important de la segmentation où le niveau de gris de chacune des m régions d'une image est supposée suivre une loi gaussienne de moyenne μ_i et de variance σ_i^2 , λ est l'ensemble $\{\sigma_i, \mu_i\}_{i=1, m}$. Nous étudierons ici l'estimation de ces paramètres dans le cas gaussien par raison de commodité. La probabilité des observations conditionnellement aux labels s'écrit donc :

$$P_{\lambda}(Y = y / X = x) = \prod_{s \in S} \frac{1}{\sqrt{2\pi}\sigma_{x_s}} \exp - \frac{1}{2\sigma_{x_s}^2} (y_s - \mu_{x_s})^2$$

On peut écrire la distribution a posteriori sous la forme :

$$P_{\theta, \lambda}(X = x / Y = y) = \frac{\exp - \mathcal{W}_{\theta, \lambda}(x / y)}{Z_{\theta, \lambda}}$$

avec comme définition de la fonction énergie a posteriori effective :

$$\mathcal{W}_{\theta,\lambda}(x/y) = \sum_{s \in S} \left(\frac{1}{2\sigma_{x_s}^2} (y_s - \mu_{x_s})^2 + \log \sigma_{x_s} \right) + \theta \Phi(x) \quad (7.14)$$

et comme fonction de partition a posteriori associée $Z_{\theta,\lambda}$. La vraisemblance des paramètres d'attache aux données s'écrit donc selon le même principe que précédemment

$$L(\lambda) = P_{\theta,\lambda}(Y = y) = \frac{Z_{\theta,\lambda}}{(\sqrt{2\pi})^{|S|} Z_{\theta}}$$

Intéressons nous à maximiser cette quantité vis-à-vis d'un des paramètres de λ_i . Il faut remarquer ici que les fonctions énergie mises en jeu dépendent de façon non-linéaire de l'ensemble de paramètres $\lambda = \{\lambda_i\}$. D'où l'on déduit pour chacun des paramètres λ_i :

$$\frac{\partial \log L(\lambda)}{\partial \lambda_i} = \frac{\partial \log Z_{\theta,\lambda}}{\partial \lambda_i} = -\mathbb{E}_{\theta,\lambda} \left[\frac{\partial \mathcal{W}_{\theta,\lambda}}{\partial \lambda_i} \right] = 0 \quad (7.15)$$

Précisons qu'il s'agit bien de l'espérance a posteriori. Pour analyser la dépendance précise de l'énergie a posteriori vis-a-vis des paramètres $\{\sigma_i, \mu_i\}_{i=1,m}$, il est commode de re-écrire la formule (7.14) en employant les fonctions caractéristiques d'appartenance à chaque classe :

$$\mathcal{W}_{\theta,\lambda}(x/y) = \sum_{i=1}^m \left[\sum_{s \in S} \left(\frac{1}{2\sigma_i^2} (y_s - \mu_i)^2 + \log \sigma_i \right) \cdot \mathbf{I}_{x_s=i} \right] + \theta \Phi(x)$$

Il en résulte des formules importantes pour la suite :

$$\begin{cases} \frac{\partial \mathcal{W}_{\theta,\lambda}(x/y)}{\partial \mu_i} = \frac{1}{\sigma_i^2} \sum_{s \in S} (\mu_i - y_s) \mathbf{I}_{x_s=i} \\ \frac{\partial \mathcal{W}_{\theta,\lambda}(x/y)}{\partial \sigma_i} = \frac{1}{\sigma_i} \sum_{s \in S} \left(\frac{-(y_s - \mu_i)^2}{\sigma_i^2} + 1 \right) \mathbf{I}_{x_s=i} \end{cases} \quad (7.16)$$

- Examinons d'abord le cas d'un paramètre de moyenne μ_i donné. On peut montrer que (si $\sigma_i < +\infty$) :

$$\forall i \in [1..m], \quad \mu_i = \frac{\sum_{s \in S} y_s P_{\theta,\lambda}(X_s = i)}{\sum_{s \in S} P_{\theta,\lambda}(X_s = i)} \quad (7.17)$$

L'interprétation physique en est claire : on obtient le barycentre des observations en chaque site de l'image pondérées par la probabilité a posteriori d'avoir une classe déterminée.

- De la même façon, on obtient pour les variances de chaque classe :

$$\forall i \in [1..m], \quad \sigma_i^2 = \frac{\sum_{s \in S} (y_s - \mu_i)^2 P_{\theta,\lambda}(X_s = i)}{\sum_{s \in S} P_{\theta,\lambda}(X_s = i)} \quad (7.18)$$

les μ_i pouvant être calculés par la formule (7.17). Le résultat est similaire à une variance empirique, mais avec pondération en chaque site par la probabilité a posteriori que ce site ait le label étudié.

En définitive on aboutit donc à des équations (7.17) et (7.18) auto-cohérentes, de forme bien plus complexe que celles des gradients stochastiques simple et généralisé. On aimerait pouvoir les remplacer par des formes itératives simples : c'est ce qui va justifier l'emploi des méthodes de type EM, à adapter dans le cadre gibbsien défini ici.

Expectation-maximization (EM)

Récapitulons les résultats précédents obtenus concernant l'estimation au Maximum de Vraisemblance des hyperparamètres dans le cas des données incomplètes :

- pour l'hyperparamètre de régularisation nous sommes arrivés à l'équation stochastique :

$$\mathbf{E}_{\hat{\theta}}[\Phi] = \mathbf{E}_{\hat{\theta}, \lambda}[\Phi]$$

- pour les paramètres d'attache aux données en segmentation avec m classes nous sommes arrivés au système d'équations :

$$\forall i \in [1..m] , \left\{ \begin{array}{l} \mu_i = \frac{\sum_{s \in S} y_s P_{\theta, \lambda}(X_s = i)}{\sum_{s \in S} P_{\theta, \lambda}(X_s = i)} \\ \sigma_i^2 = \frac{\sum_{s \in S} (y_s - \mu_i)^2 P_{\theta, \lambda}(X_s = i)}{\sum_{s \in S} P_{\theta, \lambda}(X_s = i)} \end{array} \right.$$

Nous en avons conclu qu'il serait désirable de résoudre ces équations de manière itérative. Par exemple si les probabilités a posteriori $P_{\theta, \lambda}(\cdot)$ étaient connues (ou apprises) à une étape donnée on pourrait les injecter dans la première équation à condition de pouvoir estimer pour toute valeur de θ l'espérance a priori du potentiel de régularisation ainsi que dans le deuxième système. Ces deux arguments vont nous fournir les principes de l'EM, avec un certain nombre de résultats théoriques très puissants à la clé.

La méthode EM est par essence itérative. Des résultats théoriques très généraux montrent que la vraisemblance des paramètres estimés croît à chaque itération de l'algorithme associé [Baum et al., 1970, Redner et Walker, 1984]. On pourra également consulter en particulier [Chalmond, 1999] pour une présentation claire de cette méthode. De plus cette méthode permet d'une certaine façon l'optimisation séparée des hyperparamètres d'attache aux données et de celui de régularisation. En effet elle considère essentiellement la probabilité jointe de l'observation et des données cachées (un étiquetage de l'image par exemple), c'est-à-dire le produit de la loi d'observation et de la probabilité a priori.

Supposons que les paramètres θ_n et λ_n soient connus à l'étape courante n . On suppose aussi que l'on a accès d'une façon ou d'une autre aux statistiques liées à la loi a posteriori des données cachées courantes (partie *Expectation* de l'EM, voir plus loin). On définit alors la notation :

$$Q(\theta, \lambda, \theta_n, \lambda_n) = \mathbf{E}_{\theta_n, \lambda_n}[\log P_{\theta, \lambda}(X = x, Y = y)]$$

La partie optimisation (*Maximisation*) de l'EM consiste à rechercher :

$$(\theta_{n+1}, \lambda_{n+1}) = \arg \max_{\theta, \lambda} Q(\theta, \lambda, \theta_n, \lambda_n)$$

Du fait de la séparabilité de la loi jointe, la fonction objectif $Q(\theta, \lambda, \theta_n, \lambda_n)$ s'écrit :

$$Q(\theta, \lambda, \theta_n, \lambda_n) = \mathbf{E}_{\theta_n, \lambda_n}[\log P_{\lambda}(Y = y / X = x) + \log P_{\theta}(X = x)]$$

Cela correspond donc à l'optimisation séparée :

$$\left\{ \begin{array}{l} \theta_{n+1} = \arg \max_{\theta} \mathbf{E}_{\theta_n, \lambda_n}[\log P_{\theta}(X = x)] \\ \lambda_{n+1} = \arg \max_{\lambda} \mathbf{E}_{\theta_n, \lambda_n}[\log P_{\lambda}(Y = y / X = x)] \end{array} \right.$$

On rappelle que la vraisemblance des paramètres $L(\theta_n, \lambda_n)$ croît en fonction de l'itération n par cette méthode ce qui aboutit à terme à un optimum local pour les valeurs des paramètres.

Examinons plus précisément ce qui se passe pour chacune des catégories de paramètres :

- en ce qui concerne le paramètre de régularisation, cela correspondra à :

$$E_{\theta_n, \lambda_n} \left[\frac{\partial \log P_\theta(X = x)}{\partial \theta} \right] = E_{\theta_n, \lambda_n} \left[-\Phi - \frac{\partial \log Z_\theta}{\partial \theta} \right] = 0$$

c'est-à-dire en vertu de résultats supposés maintenant acquis :

$$E_{\theta_{n+1}}[\Phi] = E_{\theta_n, \lambda_n}[\Phi] \quad (7.19)$$

Supposons que l'on sache d'une manière ou d'une autre calculer ou estimer l'espérance a posteriori du potentiel de régularisation pour la valeur courante des paramètres. Ainsi, en pratique, on approxime cette quantité par sa moyenne empirique au cours d'un échantillonneur de Gibbs (ou de Metropolis) pris pour la valeur courante des paramètres³. On est alors ramené à un problème d'estimation du paramètre de régularisation pour la donnée complète $E_{\theta_n, \lambda_n}[\Phi]$! On peut donc appliquer la technique du pseudo-maximum de vraisemblance [Chalmond, 1989] ou bien encore celle du gradient stochastique [Younes, 1991] pour re-estimer la nouvelle valeur du paramètre de régularisation.

- pour les paramètres d'attache aux données, on a vu plus haut que

$$\frac{\partial \log P_\lambda(Y = y / X = x)}{\partial \lambda_i} = - \frac{\partial \mathcal{W}_{\theta, \lambda}(x)}{\partial \lambda_i}$$

Le principe EM s'écrit donc

- pour les paramètres de moyenne :

$$\begin{aligned} E_{\theta_n, \lambda_n} \left[\frac{\partial \mathcal{W}_{\theta, \lambda}}{\partial \mu_i} \right] &= \frac{1}{\sigma_i^2} \sum_{s \in S} (\mu_i - y_s) P_{\theta_n, \lambda_n}(X_s = i) = 0 \\ \Rightarrow \forall i \in [1..m], \mu_i(n+1) &= \frac{\sum_{s \in S} y_s P_{\theta_n, \lambda_n}(X_s = i)}{\sum_{s \in S} P_{\theta_n, \lambda_n}(X_s = i)} \end{aligned} \quad (7.20)$$

- pour les paramètres de variances :

$$\begin{aligned} E_{\theta_n, \lambda_n} \left[\frac{\partial \mathcal{W}_{\theta, \lambda}}{\partial \sigma_i} \right] &= \sum_{s \in S} \left(\frac{-(y_s - \mu_i)^2}{\sigma_i^2} + 1 \right) P_{\theta_n, \lambda_n}(X_s = i) = 0 \\ \Rightarrow \forall i \in [1..m], \sigma_i(n+1)^2 &= \frac{\sum_{s \in S} (y_s - \mu_i(n+1))^2 P_{\theta_n, \lambda_n}(X_s = i)}{\sum_{s \in S} P_{\theta_n, \lambda_n}(X_s = i)} \end{aligned} \quad (7.21)$$

Deux remarques insistantes :

1. Ce sont bien les distributions a posteriori qui sont mises en jeu dans l'estimation itérative (eqs. 7.20 et 7.21). Elles sont donc à re-estimer à chaque itération n . Une manière bien naturelle est de remplacer ces probabilités à l'étape courante du processus EM par la fréquence empirique d'apparition des labels lors d'une série d'échantillonnages de la distribution de Gibbs a posteriori courante menés à l'aide d'un échantillonneur de type Gibbs ou Metropolis

³Ceci correspond à l'étape Estimation de l'algorithme EM.

[Chalmond, 1989]. Notons pour cela N le nombre d'itérations effectué avec l'échantillonneur ainsi sélectionné :

$$\begin{array}{ccc} \bullet & \text{-----} & \bullet \\ (n) & N & (n+1) \end{array}$$

Notons également $N_s(i)$ le nombre de fois que le label i a été tiré en un site s au cours de l'ensemble de ces N échantillonnages. On peut donc écrire : $P_{\theta_n, \lambda_n}(X_s = i) \approx \frac{N_s(i)}{N}$, et il en résulte les estimations empiriques des paramètres de moyennes et de variance de chaque classe :

$$\forall i \in [1..m], \left\{ \begin{array}{l} \mu_i(n+1) = \frac{\sum_{s \in S} y_s N_s(i)}{\sum_{s \in S} N_s(i)} \\ \sigma_i(n+1)^2 = \frac{\sum_{s \in S} (y_s - \mu_i(n+1))^2 N_s(i)}{\sum_{s \in S} N_s(i)} \end{array} \right.$$

De la même façon, notons $x^{(n)}(k)$ la série d'images échantillons ainsi obtenue pour $k = 1..N$. On obtient pour l'espérance a posteriori du potentiel de régularisation

$$E_{\theta_n, \lambda_n}[\Phi] \approx \frac{1}{N} \left[\sum_{k=1}^N \Phi(x^{(n)}(k)) \right]$$

C'est donc ensuite que l'on procède à l'étape d'estimation EM comme indiqué plus haut, à partir de ces valeurs empiriques a posteriori.

2. On retrouve le caractère itératif que nous voulions anticiper à propos des équations plus haut. Là aussi, à convergence, on doit satisfaire à la forme exacte des équations (7.17) et (7.18) obtenues par le Maximum de Vraisemblance.

Variantes Une variante importante, appelée ICE ou Iterative Conditional Estimation est la suivante [Pieczyński, 1994]. Considérons les N échantillons $x^{(n)}(k)$ obtenus précédemment à une étape courante n de l'EM. On pourrait songer à estimer moyenne et variance des classes pour chacun d'eux considéré comme segmentation courante, puis prendre la moyenne empirique (arithmétique) des valeurs ainsi obtenues

$$\forall i \in [1..m], \left\{ \begin{array}{l} \mu_i(n+1) = \frac{1}{N} \sum_{k=1}^N \left(\frac{\sum_{s \in S} y_s \mathbf{I}_{x_s^{(n)}(k)=i}}{\sum_{s \in S} \mathbf{I}_{x_s^{(n)}(k)=i}} \right) \\ \sigma_i(n+1)^2 = \frac{1}{N} \sum_{k=1}^N \left(\frac{\sum_{s \in S} (y_s - \mu_i(n+1))^2 \mathbf{I}_{x_s^{(n)}(k)=i}}{\sum_{s \in S} \mathbf{I}_{x_s^{(n)}(k)=i}} \right) \end{array} \right.$$

Cette estimation est différente de l'EM. On peut montrer qu'elle correspond en fait à l'estimateur de la moyenne a posteriori des paramètres.

Conclusion pour l'estimation en données incomplètes

On voit donc au terme de cette partie que l'estimation en données incomplètes se prête à un nombre très riche de variantes. On pourrait ainsi parfaitement juger préférable d'estimer certains types de paramètres comme

moyenne et variance par l'EM, tandis que l'on adopterait la méthode de Lakhsmann-Derin pour le paramètre de régularisation, ou réciproquement ! Ces variantes méritent encore d'être examinées et comparées entre elles de façon exhaustive, dans la lignée de l'approche suivie dans [Pieczynski, 1994].

Il faut aussi préciser que les méthodes utilisant un échantillonnage de la distribution a posteriori (comme l'EM) permettent en même temps de fournir une série de configurations échantillons se prêtant favorablement à la segmentation ou à la restauration dans le cadre des estimateurs de type TPM ou MPM, lorsque la convergence des paramètres vers leur valeur optimale est supposée atteinte.

7.4 Processus de bords

L'un des grands domaines d'application des modèles markoviens est la restauration comme nous l'avons mentionné au paragraphe 7.1.7. Connaissant une réalisation y , on cherche une solution x minimisant le critère :

$$\mathcal{U}(x / y) = U_1(x, y) + \beta U_2(x)$$

Le terme de régularisation $U_2(x)$ permet alors d'agir directement sur les dérivées de la solution cherchée. Dans le cas d'un potentiel quadratique, il s'agit d'une régularisation de Tikhonov qui entraîne un fort lissage de la solution au détriment des discontinuités souvent naturellement présentes dans l'image.

De nombreux travaux ont porté sur l'introduction de discontinuités et sur les avantages potentiels que cela impliquait dans la recherche de la solution. Nous nous limiterons ici à des cliques d'ordre deux, c'est-à-dire à l'introduction de contraintes sur la dérivée première de la solution cherchée. Nous mettons en évidence l'équivalence entre processus de bords et certaines fonctions avant de présenter quelques algorithmes déterministes d'optimisation.

7.4.1 Processus de bords explicites et implicites

Dès l'article fondateur de Geman et Geman [Geman et Geman, 1984a], l'introduction de processus de bords permettant de désactiver l'effet de lissage entre deux sites est envisagé. Celui-ci, que nous noterons B dans la suite est défini sur une grille duale de celle de l'image et est souvent décomposé en deux sous champs, l'un représentant les interactions verticales B^v et l'autre les interactions horizontales B^h . Il s'agit souvent (mais pas nécessairement) de processus de bords booléens, prenant la valeur 1 en présence de discontinuités et 0 en son absence. Le terme de régularisation s'écrit alors [Blake et Zissermann, 1987] [Geiger et Girosi, 1991] :

$$U_2(x, b) = \sum_{(s,t) \in \mathcal{C}^v} [(x_s - x_t)^2 (1 - b_{st}^v) + \gamma b_{st}^v] + \sum_{(s,t) \in \mathcal{C}^h} [(x_s - x_t)^2 (1 - b_{st}^h) + \gamma b_{st}^h] \quad (7.22)$$

D'une façon intuitive, en l'absence de discontinuités, on a $b_{st} = 0$ et on retrouve un potentiel quadratique visant à lisser la solution. En revanche, en présence de discontinuité ($b_{st} = 1$), la pénalité est arbitraire et vaut γ^4 de façon à limiter le nombre de discontinuités introduites par le modèle. Ce modèle (modèle de la membrane mince -*weak membrane*-) est très simple et n'introduit pas d'interactions entre les éléments du processus bords.

En réalité, il n'est pas nécessaire d'introduire explicitement un processus de bords et un choix de fonction judicieux peut procurer le même résultat. En effet, notons $\phi^v(x_s, x_t, b_{st}^v)$ le terme $(x_s - x_t)^2 (1 - b_{st}^v) + \gamma b_{st}^v$ (et

⁴Ce paramètre peut varier pour des modèles anisotropes ou non stationnaires (on a alors γ_{st}^v et γ_{st}^h).

de même pour ϕ^h), alors on cherche la solution (x, b) telle que $U_2(x, b)$ soit minimale. Or :

$$\begin{aligned} \min_{(x,b)} U_2(x, b) &= \min_{(x,b)} \sum_{(s,t) \in \mathcal{C}^v} \phi^v(x_s, x_t, b_{st}^v) + \sum_{(s,t) \in \mathcal{C}^h} \phi^h(x_s, x_t, b_{st}^h) \\ &= \min_x \sum_{(s,t) \in \mathcal{C}^v} \min_{b_{st}^v} \phi^v(x_s, x_t, b_{st}^v) + \sum_{(s,t) \in \mathcal{C}^h} \min_{b_{st}^h} \phi^h(x_s, x_t, b_{st}^h) \\ &= \min_x \sum_{(s,t) \in \mathcal{C}^v} \psi(x_s - x_t) + \sum_{(s,t) \in \mathcal{C}^h} \psi(x_s - x_t) \end{aligned}$$

en notant :

$$\psi(u) = \min(u^2, \gamma)$$

Par conséquent, la quadratique est remplacée dans ce modèle par une quadratique tronquée (figure 7.7). L'utilisation d'un processus de bords explicite est dans ce cas équivalente à l'utilisation de la fonction ψ (on parle de processus de bords implicite). La valeur de γ détermine à partir de quelle valeur du gradient on introduira une discontinuité.

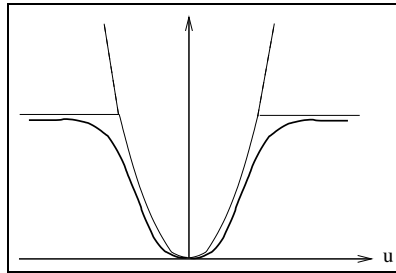


FIG. 7.7 – Fonctions quadratique, quadratique tronquée et la fonction de régularisation proposée par Geman et McClure préservant les discontinuités.

Un très grand nombre de fonctions de régularisation préservant les discontinuités ont été proposées et étudiées (on pourra se référer à [Charbonnier et al., 1996] pour une étude comparative). La fonction suivante, qui présente l'intérêt d'être dérivable partout, est très utilisée en restauration [Geman et McClure, 1985] :

$$\phi(u) = \frac{u^2}{1 + u^2}$$

Notons qu'il est intéressant d'utiliser des fonctionnelles convexes pour la recherche de la solution, mais la prise en compte des discontinuités repose sur le comportement de $\frac{\phi'(u)}{2u}$.

Le choix d'un processus de bords implicite ou explicite va donner lieu à différents algorithmes de minimisation. En effet, le recuit simulé permettant d'accéder au minimum global de l'énergie est un algorithme long et coûteux. Il peut dans de nombreux cas être remplacé par un algorithme déterministe pour accélérer la recherche de la solution.

7.4.2 Algorithmes de minimisation

Nous donnons ici le principe des quelques algorithmes déterministes :

- GNC [Blake et Zissermann, 1987] : Blake et Zisserman ont proposé un algorithme déterministe appelé le GNC (Graduated Non Convexity) pour l'optimisation du critère (sous sa forme implicite). Le principe de l'algorithme consiste à approximer le critère par une fonction convexe qui permet de définir une bonne solution initiale (calculée par une descente de gradient par exemple, puisque la solution est dans ce cas unique).

Puis le critère est modifié graduellement perdant sa propriété de convexité pour se rapprocher du critère initial. À chaque étape une solution est trouvée de façon déterministe en utilisant pour initialisation la solution donnée par l'étape précédente. Si des preuves de convergence peuvent être obtenues dans certains cas particuliers, cette démarche n'assure cependant pas de trouver le minimum global pour toutes les fonctions d'énergie.

- MFA [Geiger et Girosi, 1991] : une autre approche est le recuit par champ moyen MFA (Mean Field Annealing) qui utilise cette fois un processus bords explicite. L'algorithme met en œuvre une descente de température au cours de laquelle à chaque étape une solution (image restaurée et processus de bords) est estimée au sens du champ moyen. On a des expressions explicites pour le processus de bords, tandis que x est estimée de façon itérative.
- Artur et Legend [Charbonnier, 1994] : Charbonnier a également proposé deux algorithmes déterministes appelés Artur et Legend pour l'optimisation. Il utilise un processus de bords explicite et exploite le fait qu'à b fixé, le critère est quadratique en x (donc convexe et minimisable rapidement par une descente de type gradient), tandis qu'à x fixée, il existe une expression analytique du processus de bords.

Malgré de nombreuses différences (notamment sur les propriétés de convergence et sur leur généralité), il s'agit toujours d'algorithmes itératifs minimisant des suites d'énergies (GNC ou MFA) ou faisant varier les variables auxiliaires (le processus de bords pour Artur et Legend) au cours des itérations. Une comparaison est effectuée dans [Charbonnier, 1994] et [Zerubia et Chellapa, 1993].

7.5 Quelques applications des champs markoviens

Cette partie présente quelques applications des champs markoviens en traitement d'images pour illustrer les potentialités de ce domaine. On s'intéressera dans un premier temps à des applications de bas niveau (analyse de textures et segmentation) avant de présenter des applications manipulant des graphes construits à partir de primitives plus complexes (régions, segments).

7.5.1 Applications sur le graphe des pixels

Cette première partie présente trois applications de bas-niveau, i.e qui travaillent sur le graphe des pixels :

- La première est une analyse de textures s'appuyant sur un modèle markovien ; elle permet de discriminer différentes textures de l'image à l'aide des paramètres du champ extraits ; un exemple d'applications pour la détection des zones urbaines en imagerie satellitaire est donné ;
- La seconde est un simple exemple de segmentation appliquée en imagerie radar mettant en évidence la souplesse de ce modèle pour prendre en compte des statistiques très variées dans les images et détaillant quelques aspects pratiques de segmentation ; on se placera dans un cadre supervisé pour le paramètre de régularisation, c'est à dire que celui-ci sera fixé de façon empirique ;
- La troisième application présentée est un schéma de fusion markovien, relativement général qui permet de combiner plusieurs sources d'informations ; un exemple est donné dans le cas de l'analyse d'images satellitaires NOAA dans plusieurs bandes spectrales.

Analyse de textures

Nous nous intéressons dans cette partie à l'utilisation des modèles markoviens pour discriminer différents types de textures dans les images. L'idée est de se placer dans le cas de données complètes, i.e de faire l'hypothèse que l'image dont nous disposons est la réalisation d'un champ markovien, et d'extraire les paramètres de ce champ. La variabilité des paramètres en fonction du type de textures devant alors permettre de réaliser une classification de l'image.

Nous avons déjà vu au paragraphe 7.3.2 un certain nombre de techniques (méthode des codages, maximum de pseudo-vraisemblance, gradient stochastique) pour calculer les paramètres d'un champ markovien. Toutes ces méthodes, qui ne présupposent pas de modèle pour le champ, sont assez lourdes à mettre en œuvre. Nous allons supposer ici que nous nous situons dans le cadre d'un champ markovien gaussien pour lequel nous pouvons obtenir une expression exacte des paramètres. Dans le cas d'un champ markovien gaussien, nous pouvons écrire l'énergie $U(x)$ sous la forme :

$$U = \frac{1}{T} \left(\lambda \sum_{s \in S} (x_s - \mu)^2 + \sum_{c=(s,t) \in \mathcal{C}} (x_s - x_t)^2 \right)$$

Avec cette écriture, les paramètres caractérisant le champ et que nous cherchons à estimer sont donc la température T (correspondant à une sorte de "variance généralisée"), la moyenne μ et la pondération du terme d'attache aux données de l'énergie, λ .

La méthode des queues de comètes La méthode que nous allons voir a été proposée dans [Descombes, 1993]. Considérons la probabilité conditionnelle en un site :

$$P(X_s = x_s / x_t, t \in \mathcal{V}_s) = \frac{1}{Z(V_s)} \exp \left\{ -\frac{1}{T} \left(\lambda (x_s - \mu)^2 + \sum_{t \in \mathcal{V}_s} (x_s - x_t)^2 \right) \right\}$$

En notant n le nombre de sites voisins de s , et m_s la moyenne locale des niveaux de gris des voisins du site s ,

$m_s = \frac{\sum_{t \in \mathcal{V}_s} x_t}{n}$, on peut montrer qu'on a :

$$\begin{aligned} P(X_s = x_s / x_t, t \in \mathcal{V}_s) &= \frac{1}{Z(m_s)} \exp \left\{ -\frac{n + \lambda}{T} \left(x_s - \frac{1}{n + \lambda} (nm_s + \lambda\mu) \right)^2 \right\} \\ &= P(X_s = x_s / m_s) \end{aligned}$$

On a donc remplacé le conditionnement local par l'ensemble des voisins de s , par une seule variable conditionnante m_s , ce qui améliorera la robustesse des estimateurs. En effet, le nombre de sites concernés par un conditionnement par m_s sera bien plus grand que celui des sites concernés par une configuration $V_s = (x_t, t \in \mathcal{V}_s)$ du voisinage.

Nous constatons que $P(X_s / m_s)$ est une loi gaussienne, de moments (moyenne et variance) suivants :

$$\begin{aligned} E(X_s / m_s) &= \frac{nm_s + \lambda\mu}{n + \lambda} \\ \text{var}(X_s / m_s) &= \frac{T}{2(n + \lambda)} \end{aligned}$$

L'espérance et la variance de X_s conditionnellement à m_s étant fonction des paramètres λ , T et μ que nous recherchons, il nous suffit d'estimer ces moments de façon empirique. La variance ne dépendant pas de la valeur conditionnante m_s , on notera le moment théorique par σ^2 . Pour faire ces estimations, on construit l'image dite des "queues de comètes" en raison de son aspect, qui est simplement l'image des fréquences normalisées des niveaux de gris x_s conditionnellement à la moyenne m_s du voisinage. Si on met les valeurs de m_s en colonne et celles de x_s en ligne, chaque ligne de l'image des queues de comètes représente $P(X_s / m_s)$, i.e une gaussienne de moyenne $\frac{nm_s + \lambda\mu}{n + \lambda}$ et de variance $\frac{T}{2(n + \lambda)}$.

On considère alors pour la variance, l'estimateur suivant :

$$\widehat{\sigma}^2 = \sum_{m_s} P(m_s) \widehat{\text{var}}(X_s / m_s)$$

en notant $\widehat{\text{var}}(X_s / m_s)$ l'estimation empirique de la variance faite selon une ligne des queues de comètes. Cet estimateur permet d'accorder à chaque probabilité conditionnelle locale une importance proportionnelle au nombre d'échantillons qui la constituent et entraîne une plus grande robustesse de l'estimation.

En ce qui concerne l'espérance, nous avons la relation suivante :

$$\begin{aligned} \mathbf{E}(X_s / m_s) &= \frac{n}{n + \lambda} m_s + \frac{\mu}{n + \lambda} \\ &= \alpha m_s + \beta \end{aligned}$$

Par conséquent, les espérances des probabilités conditionnelles se situent sur une droite. L'estimation empirique des moyennes $\widehat{\mathbf{E}}(X_s / m_s)$ permet de faire une estimation aux moindres carrés des paramètres α et β de la droite. Une fois estimés empiriquement $\widehat{\sigma}^2$, $\widehat{\alpha}$, $\widehat{\beta}$, on peut déduire λ , μ , et T par les relations :

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{\widehat{T}}{2(n_s + \widehat{\lambda})} \\ \widehat{\alpha} &= \frac{n_s}{n_s + \lambda} \\ \widehat{\beta} &= \frac{\widehat{\lambda}\widehat{\mu}}{n_s + \widehat{\lambda}} \end{aligned}$$

En explicitant l'estimation aux moindres carrés des paramètres de la droite, on peut exprimer directement λ , μ , et T en fonction des moments d'ordre 1 et 2, conditionnés ou non. On retrouve dans ce cas pour μ , la moyenne empirique des x_s .

Application à la détection des zones urbaines

Cet exemple d'application est tiré de la thèse de X. Descombes [Descombes, 1993].

On peut utiliser les résultats précédents pour analyser les textures présentes sur une image satellitaire SPOT. L'image étant par essence non stationnaire (sinon l'analyse aurait peu d'intérêt!), le calcul des paramètres se fait localement, sur une fenêtre glissante centrée en chaque pixel. La fiabilité des estimateurs s'accroît avec la taille de la fenêtre, en même temps, et de façon antagoniste, que le risque de considérer des mélanges de textures différentes à l'intérieur de la fenêtre d'étude. Une solution pour remédier à ce problème peut être de ne considérer que les échantillons les plus représentés. En cas de mélange, les échantillons appartiendront à la texture la plus présente dans la fenêtre.

Le paramètre de température, est un bon indicateur du milieu urbain qui se présente sur une image SPOT sous une forme assez texturée, type "poivre et sel" avec alternance de niveaux de gris faibles et forts. En effet, ce paramètre qui mesure en quelque sorte le cahut de la zone, a des valeurs plus élevées dans les régions urbaines. Il permet d'obtenir une bonne discrimination du milieu urbain.

7.5.2 Segmentation

Nous allons aborder dans cette partie une application très classique des champs markoviens qui est la segmentation. Nous commençons par rappeler le principe de la segmentation markovienne en prenant l'exemple des images radar [Tupin, 1997], puis présentons une méthode de fusion dans un cadre markovien [Descombes, 1993].

Segmentation d'une image radar *Cet exemple d'application est tiré de la thèse de F. Tupin [Tupin, 1997].*

La modélisation est similaire à celle qui a été présentée dans le paragraphe 7.1.7. Ecrivons à nouveau les deux termes intervenant dans la probabilité a posteriori. Pour la probabilité du champ des observations conditionnellement au champ des étiquettes, en supposant l'indépendance des pixels, on a :

$$P(Y / X = x) = \prod_s P(Y_s = y_s / X_s = x_s)$$

Les images radar sont des images très bruitées par le phénomène de speckle. En revanche, le processus d'acquisition est bien modélisé statistiquement et on a l'expression suivante pour une image radar en amplitude :

$$p(Y_s = y_s / X_s = i) = \frac{2L^L}{\mu_i^L \Gamma(L)} y_s^{(2L-1)} \exp\left(-\frac{L y_s^2}{\mu_i}\right)$$

avec L un paramètre du système connu⁵ appelé nombre de vues, Γ la fonction Gamma, et μ_i les moyennes en intensité (carré de l'amplitude) des différentes classes i considérées.

Le champ des étiquettes est supposé markovien avec un modèle de Potts qui vise à obtenir des zones homogènes compactes sur l'image segmentée :

$$P(X = x) = \frac{1}{Z} \exp\left(-\beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t)\right)$$

avec $\beta > 0$, $\phi(0) = 1$ et $\phi(x) = 1 \quad \forall x \neq 0$.

Le champ a posteriori résultant est donc markovien et son énergie s'écrit :

$$\mathcal{U}(x / y) = L \sum_s \left(\frac{y_s^2}{\mu_{x_s}} + \ln \mu_{x_s} \right) + \sum_{c=(s,t)} \phi(x_s - x_t)$$

En choisissant l'estimateur MAP, la solution est obtenue par recuit simulé avec une décroissance géométrique en température, et une température initiale fixée arbitrairement. Le choix des classes se fait de la façon suivante. On se fixe le nombre de classes (15 dans les illustrations ci-dessous) et on applique un algorithme de k-moyennes dont le résultat⁶ sert à calculer les valeurs des moyennes en intensité μ_i des différentes classes. Notons que ces classes n'ont pas de contenu sémantique et que la segmentation correspond ici à un "découpage" de l'image.

Le choix de β qui pondère l'influence entre attache aux données et régularisation se fait de façon ad hoc après quelques essais. L'augmentation de β entraîne une augmentation de la taille des zones obtenues par la segmentation. Il serait bien sûr possible d'estimer ce paramètre à l'aide d'une des méthodes décrites au chapitre 3. Notons que ce modèle n'est pas adapté à la préservation des cibles ponctuelles et des lignes qui sont des configurations de forte énergie pour le champ des étiquettes (il faut donc une forte attache aux données pour que ces configurations subsistent dans le résultat final). Il est bien sûr possible de prendre en compte un champ externe pour mieux respecter les lignes par exemple [Tupin et al., 1996] ou d'utiliser un modèle plus approprié [Descombes et al., 1995].

Le tableau ci-dessous 7.2 résume les paramètres utilisés et la figure 7.4 montre le résultat de la segmentation.

Schéma de fusion dans un cadre markovien • Principe du schéma de fusion

Nous nous plaçons maintenant dans le cas où plusieurs sources d'information sont disponibles pour réaliser la classification de l'image. Ces différentes sources peuvent provenir de l'extraction de différents paramètres à partir d'une même image ou directement de plusieurs capteurs.

⁵Pour les produits PRI du satellite ERS1 $L = 3$.

⁶Le résultat des k-moyennes est très bruité à cause du bruit multiplicatif présent sur les images radars et l'absence de modes dans l'histogramme. Il ne peut donc pas dans ce cas être utilisé directement comme résultat de segmentation.

TAB. 7.2 – Valeurs des paramètres de la segmentation

Température initiale	5
Facteur de décroissance géométrique	0,95
Paramètre de régularisation β	0,4
Nombre de classes	15
Nombre d'itérations pour les k-moyennes	20

Notons y le vecteur d'attributs correspondant aux différentes sources d'informations $y = (y^1, \dots, y^K)$ avec K le nombre de données (ou canaux) et M le nombre de classes $E = \{\lambda_1, \dots, \lambda_M\}$. La probabilité a posteriori s'écrit :

$$p(X / Y = y) \propto p(Y / X)p(X)$$

Si nous faisons l'hypothèse que les sources sont **indépendantes** entre elles, et les pixels de chaque source entre eux, on a :

$$\begin{aligned} p(Y / X = x) &= \prod_{s \in S} P(Y_s / X_s = x_s) \\ &= \prod_{s \in S} P(\{Y_s^1, Y_s^2, \dots, Y_s^K\} / X_s = x_s) \\ &= \prod_{s \in S} P(Y_s^1 / X_s = x_s) \dots P(Y_s^K / X_s = x_s) \\ &= \prod_{s \in S} \prod_{k=1}^K P(Y_s^k / X_s = x_s) \end{aligned}$$

Le potentiel d'attache aux données résultant s'exprime alors sous forme d'une somme des potentiels individuels de chaque source :

$$U_s(y_s = (y_s^k)_k / x_s = \lambda) = \sum_k U_s(y_s^k / \lambda)$$

Les différentes sources ne renseignant pas de la même façon sur toutes les classes, il est possible d'introduire des coefficients de pondération exprimant la confiance (la fiabilité) qu'on veut accorder à chacune des sources par rapport à une classe. On notera $\gamma_{(k,\lambda)}$ la "confiance" accordée à la source k pour la classe λ . Il faut par ailleurs ne favoriser aucune classe, donc les coefficients doivent vérifier $\sum_k \gamma_{(k,\lambda)} = 1 \quad \forall \lambda$. L'expression de l'attache aux données s'écrit alors :

$$U_s(y_s = (y_s^k)_k / x_s = \lambda) = \sum_k \gamma_{(k,\lambda)} U_s(y_s^k / \lambda)$$

En pratique, il n'est pas toujours nécessaire de faire des modélisations statistiques compliquées pour calculer les potentiels d'attache aux données. Souvent des potentiels très simples, linéaires par morceaux, permettent d'obtenir de bons résultats. On attache un potentiel faible (typiquement 0) à la plage de niveaux de gris qui correspond à la classe considérée et un potentiel élevé (typiquement 1) ailleurs. Cette définition peut se faire de façon supervisée en analysant l'histogramme ou de façon automatique par une recherche des modes de l'histogramme de l'image (analyse multi-échelle [Aurdal, 1997], recuit simulé sur l'histogramme [Bloch et Maitre, 1997], etc.).

La définition des coefficients de fiabilité des sources est souvent plus problématique. On se limite en général à des remarques de bon sens en affectant $\gamma_{k,\lambda} = 0$ lorsque la source k n'est pas significative pour la classe λ , 0.5 si l'information délivrée est approximative, et 1 si la source est pertinente (avant normalisation).

- **Application à l'analyse des images SPOT**

Cet exemple d'application est tiré de la thèse de X. Descombes [Descombes, 1993]. On trouvera d'autres exemples dans [Tupin et al., 1996] [Aurdal, 1997].

Nous décrivons ici l'application de ce schéma à l'analyse de plusieurs canaux délivrés par le satellite NOAA. Il s'agit de 5 canaux de basse résolution (1.1km) correspondant aux domaines visible, proche infra-rouge, moyen infra-rouge et 2 canaux thermiques. Les classes qu'on cherche à discriminer sont les suivantes : mer, nuages, continent sans relief, continent avec relief et icebergs.

Une première étape consiste à définir les "sources" que nous allons utiliser. La méthode des queues de comète que nous avons présentée plus haut permet en effet de déduire de chaque image, 3 images de paramètres (moyenne locale, température et paramètre d'attache aux données λ). Au total 20 images sont donc disponibles, dont les plus significatives pour notre objectif, sont sélectionnées : l'image 1 (visible), l'image de température et l'image de moyenne associées, l'image 3 (moyen infra-rouge) et l'image de température associée (soit 5 en tout).

Pour chacune de ces images, une analyse supervisée de l'histogramme est effectuée, permettant de définir les potentiels (linéaires par morceaux) et la pertinence du canal pour chaque classe. Le terme d'attache aux données de l'énergie a posteriori est alors défini comme mentionné précédemment. Quant au terme de régularisation, les classes recherchées étant relativement compactes, on utilise un modèle de Potts.

7.6 Applications sur des graphes de primitives

Comme nous l'avons mentionné dans la section 7.1, le formalisme markovien est défini sur tout graphe et son champ d'applications est donc bien plus vaste que la simple grille des pixels. De nombreux problèmes se prêtent à la manipulation de primitives plus complexes, soit pour des raisons de rapidité comme dans le premier exemple que nous décrivons ci-dessous, soit parce qu'il s'agit d'un problème plus proche de l'interprétation d'image, difficile à traiter au niveau du pixel et nécessitant l'introduction d'informations de haut niveau comme dans le second exemple décrit.

7.6.1 Graphes de régions : application à la segmentation d'une image d'IRM cérébrale

Cet exemple d'application est tiré de la thèse de T. Géraud [Géraud et al., 1995].

Pour accélérer la segmentation des images, plusieurs applications partent d'une sur-segmentation de l'image qui est ensuite améliorée en fusionnant les régions. Cette fusion, qui procure la segmentation finale, peut se faire dans un cadre markovien. Le graphe est construit à partir des régions de la sur-segmentation, chaque région correspondant à un sommet du graphe et la relation de voisinage étant définie par la relation d'adjacence entre les régions. Le terme d'attache aux données dépend alors des attributs de la région (niveau de gris moyen des pixels la constituant, moments d'ordre supérieur, etc.) et le terme de régularisation dépend de l'application, un potentiel de Potts pouvant être utilisé lorsqu'on essaye de trouver des zones relativement compactes.

L'objectif de l'exemple décrit ici est de réaliser une segmentation d'images IRM cérébrales. Les classes considérées sont la matière grise, blanche, le liquide céphalo-rachidien, les ventricules et une classe ASI représentant les autres structures internes (noyaux caudés, thalamus, putamen) qui sont difficiles à segmenter et auxquelles on s'intéresse particulièrement dans cette application. Le nombre de pixels de ces images volumiques ($256 \times 256 \times 128$) limite l'utilisation de méthodes markoviennes en raison du temps de calcul. Par contre, l'utilisation d'un graphe de régions construit à partir d'une sur-segmentation, en réduisant drastiquement le nombre de sites permet de réaliser un recuit simulé à un coût raisonnable.

- **Sur-segmentation** : L'étape de sur-segmentation 3D est réalisée par un algorithme calculant la ligne de partage des eaux sur l'image du gradient après fermeture morphologique pour réduire le nombre de bassins. L'image résultat est constituée de zones de niveaux de gris homogènes, auxquelles on associe les attributs suivants : volume

(noté vol), niveau de gris moyen (qui sera l'observation y du champ des données), coordonnées du centre du bassin. Le résultat de cette méthode appliquée à l'image 7.8.a est montré sur la figure 7.8.b.

• **Relaxation markovienne** : Un graphe est construit comme indiqué précédemment à partir de la sursegmentation (fig. 7.8.c). On associe aux arcs du graphe la surface d'adjascence entre les deux régions (notée surf). Les potentiels du champ markovien sont alors définis comme suit :

- Terme d'attache aux données :

$$P(Y_s / x_s) = \exp \left(- \sum_s \frac{\text{vol}_s}{2\sigma_{x_s}^2} (y_s - \mu_{x_s})^2 \right)$$

(dédit d'une étude statistique des classes [Géraud et al., 1995])

- Terme de régularisation :

$$U_{c=(s,t)}(x) = \text{surf}_{s,t} Q_{(x_s, x_t)}$$

Seuls les potentiels des cliques d'ordre 2 sont choisis non nuls. La matrice Q est une matrice d'adjascence permettant de pondérer les voisinages entre classes suivant qu'ils sont favorisés ou non [Sigelle, 1993].

Le critère utilisé est un critère MAP et la solution est obtenue par un recuit simulé. Le temps de calcul est de moins d'une minute pour un graphe de 32 000 sites. Les résultats pour les ASI sont montrés sur la figure 7.8.c.

7.6.2 Graphes de segments : application à la détection du réseau routier

Cet exemple d'application est tiré de la thèse de F. Tupin [Tupin, 1997].

Nous nous intéressons dans cette application à la détection du réseau routier et hydrographique dans le cas des images radar. Le bruit de speckle présent sur ces images ne permet pas d'obtenir de très bons résultats de bas niveau et les détecteurs de lignes même adaptés à ce type d'imagerie ont des taux de fausses alarmes élevés si on veut obtenir des taux de détection suffisants.

L'idée est donc de faire suivre l'étape de bas niveau de détection des lignes par une étape de plus haut niveau, dans laquelle on injectera des informations a priori sur la forme des routes. Le cadre markovien, par l'intermédiaire du terme a priori se prête bien à l'introduction de connaissances sur les objets recherchés, à condition que celles-ci puissent s'exprimer de façon locale à l'échelle du graphe. Dans le cas du réseau cette hypothèse (qui assure que le champ soit markovien) est vérifiée, puisque que la pratique montre qu'il nous suffit d'informations locales au niveau des segments pour prendre une décision (présence ou absence de réseau routier).

La démarche adoptée pour la détection du réseau est donc la suivante. L'étape de bas niveau permet de détecter des segments candidats. Parmi ceux-ci, certains appartiennent aux objets à détecter, quand d'autres sont de fausses détections. On fait alors l'hypothèse que les segments détectés et toutes les connexions possibles entre ces segments contiennent le réseau routier. Les connexions "possibles" ne sont pas toutes les connexions, mais les connexions raisonnables : entre des segments suffisamment proches, et à peu près "alignés" par exemple. L'ensemble de ces segments (ceux détectés et les connexions) constituent les sommets du graphe (voir figure 7.9). La relation de voisinage entre deux segments est définie par le partage d'une extrémité par ces 2 segments. Les figures ci-dessous illustrent les étapes de la construction du graphe sur un extrait d'image radar (fig. 7.12).

Une fois ce graphe construit, on définit un champ d'observation associé Y et un champ d'étiquettes X (les étiquettes étant simplement 0 pour "non-route" et 1 pour "route") dont on cherche la configuration optimale au sens du critère MAP.

Les termes de l'énergie sont définis de la façon suivante :

- Terme d'attache aux données : l'observation en un site est définie comme la réponse moyenne du détecteur de lignes le long de ce segment ; les potentiels $U_s(y_s / x_s)$ pour les labels 0 et 1 sont alors obtenus par une étape d'apprentissage le long de quelques segments appartenant à une vraie route et de quelques segments de "fausse alarme" ; la figure 7.10 montre les fréquences des observations y (approximant les $P(y_s / x_s)$) et les potentiels linéaires par morceaux qui en sont déduits (fig. 7.11) ;

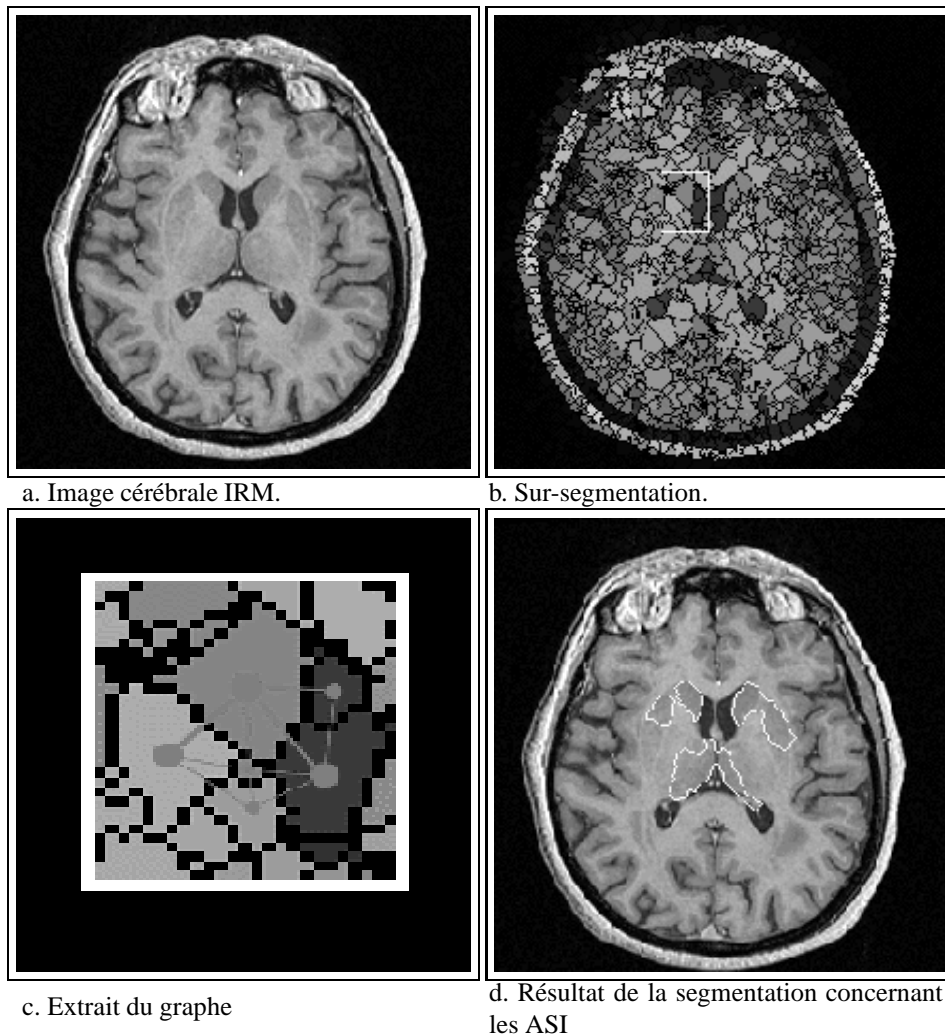


FIG. 7.8 – Etapes de la segmentation sur le graphe des régions

- Énergie a priori : elle permet justement d’intégrer toutes les connaissances a priori qu’on veut prendre en compte pour la détection du réseau, par exemple du type :
 - (i) les routes sont longues et, dans l’absolu, elles ne s’arrêtent pas ;
 - (ii) elles ont une courbure relativement faible ;
 - (iii) un segment de route est plus souvent connecté en une extrémité à un unique segment de route qu’à plusieurs.

Ces a priori peuvent s’exprimer en définissant les potentiels des cliques d’ordre maximal (rappelons qu’une clique est un ensemble de sites tous voisins les uns des autres, donc dans notre cas, un ensemble de segments qui se rejoignent en un même point) ; quatre paramètres suffisent pour les contraintes mentionnées précédemment : un paramètre contrôlant les extrémités (qui vise à défavoriser la configuration d’une clique où un seul segment a le label “route”) ; deux paramètres contrôlant la longueur des routes (qui visent à favoriser des configurations où il y a deux segments routes qui se joignent en une extrémité et qui sont “alignés”) ; un paramètre contrôlant les carrefours (qui vise à défavoriser les configurations où une multitude de segments sont “route”).

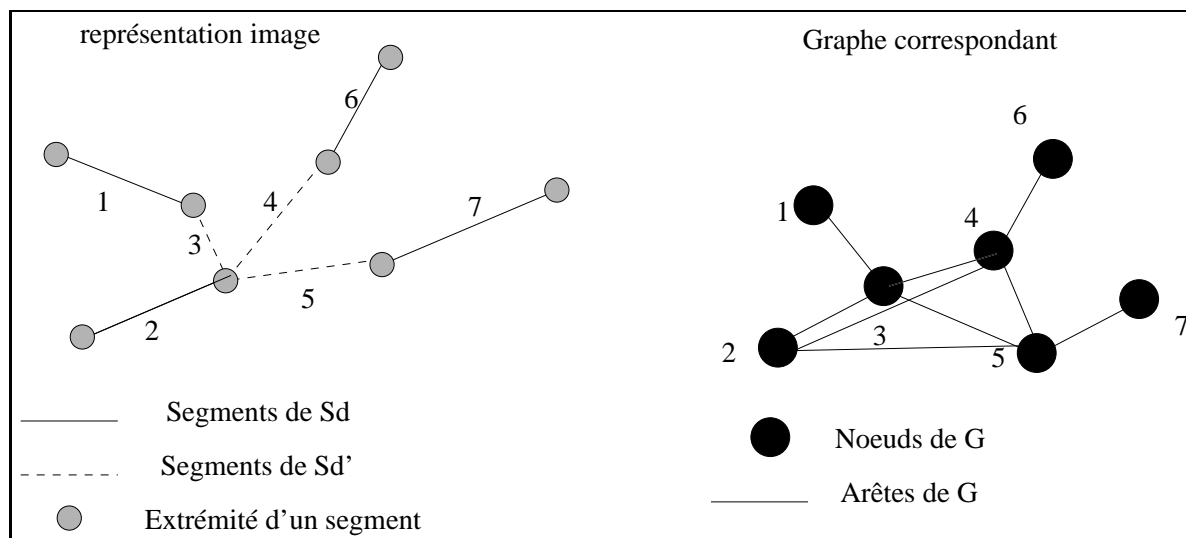


FIG. 7.9 – Construction du graphe de segments

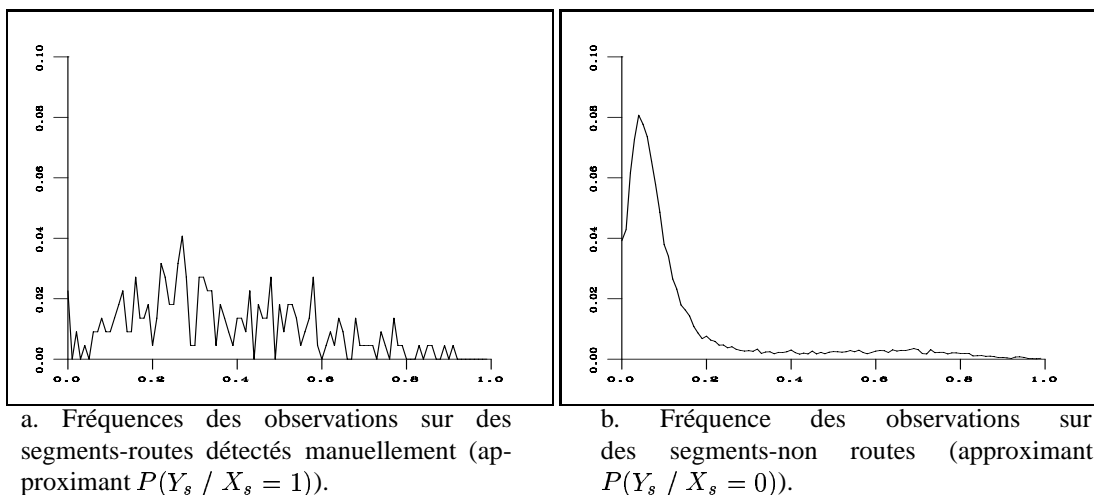


FIG. 7.10 – Fréquences conditionnelles des observations sur une zone test.

L'apprentissage de ces paramètres se fait difficilement par les méthodes du chapitre 3. En revanche l'étude de configurations extrêmes (chaîne de segments tous à "route", etc.), plus connue sous le nom de "Boîtes qualitatives d'Azencott" [Azencott, 1992], permet de fixer des intervalles de valeurs pour ces paramètres. Un exemple de résultat obtenu par recuit simulé est montré sur la figure 7.13.

7.6.3 Conclusion

Ce chapitre ne se veut pas une présentation exhaustive des applications possibles des champs markoviens, mais présente quelques unes de leurs potentialités dans des domaines et à des niveaux de traitement d'image variés. Leur grande souplesse permet en effet d'introduire toutes sortes d'informations, que ce soit pour le terme d'attache aux données, que ce soit pour les a priori possibles, ou même encore sur la forme du graphe à utiliser.

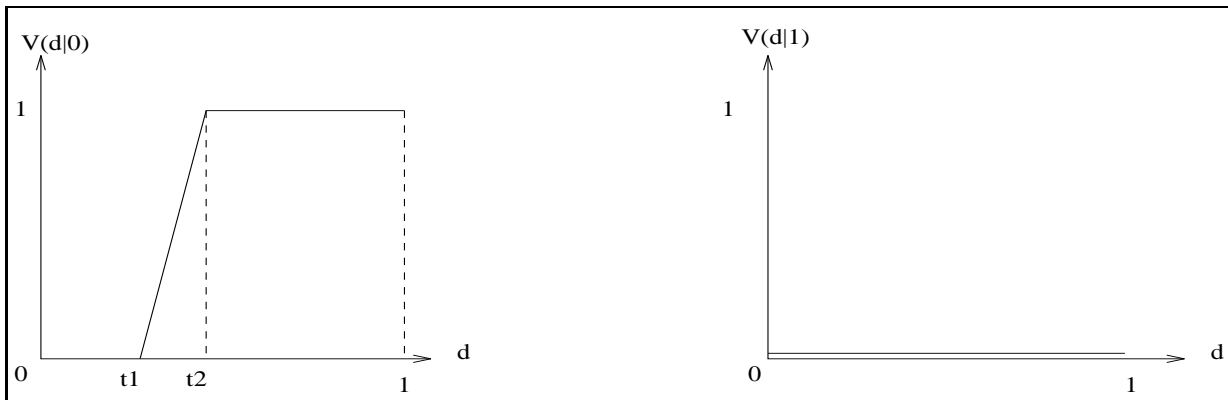
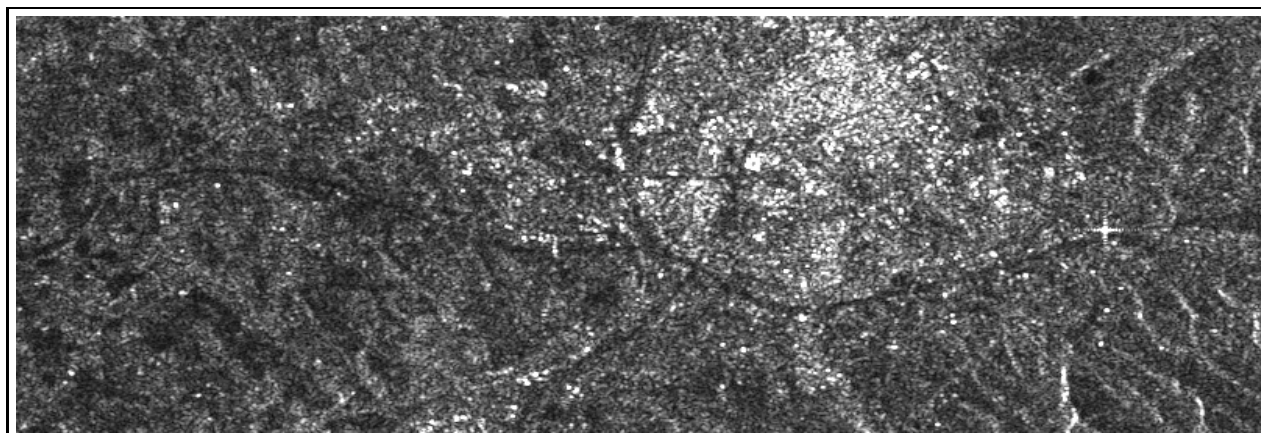
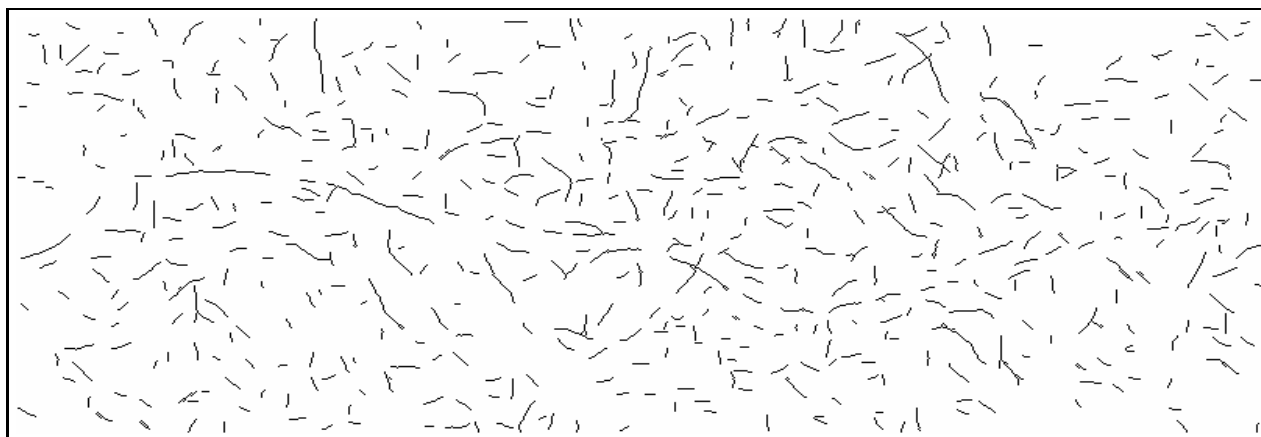


FIG. 7.11 – Potentiels linéaires par morceaux utilisés

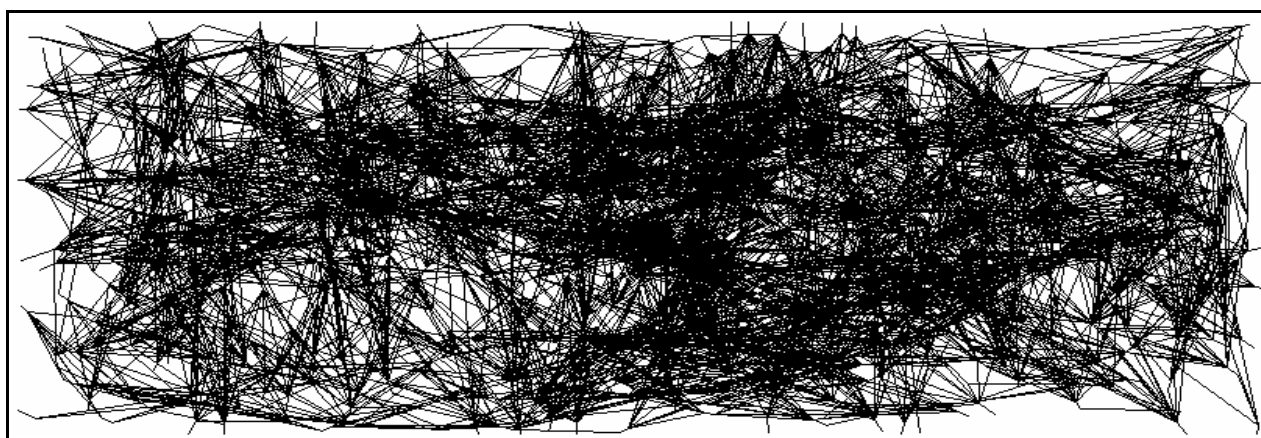
A côté de la multitude d'applications qui peuvent trouver une solution dans un cadre markovien, s'ouvrent des champs de recherche plus théoriques sur l'estimation des paramètres et l'accélération des techniques de recherche de solution.



a. Image originale centrée sur Aix en Provence ©ESA .

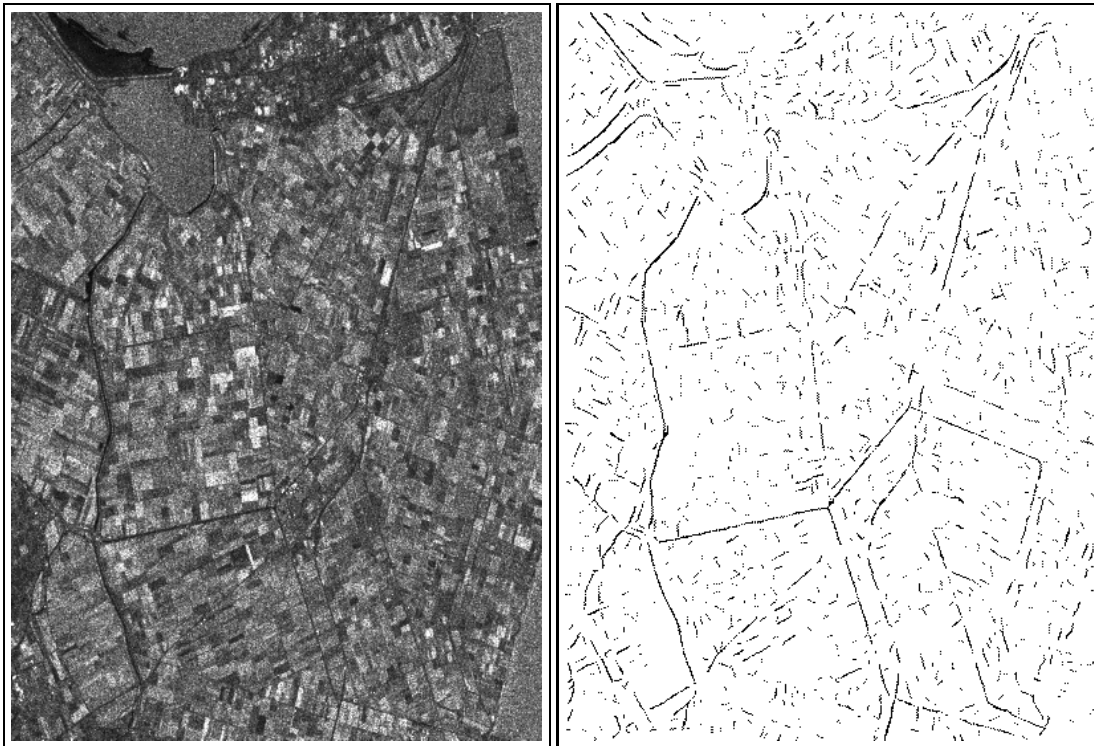


b. Ensemble des 839 segments détectés.



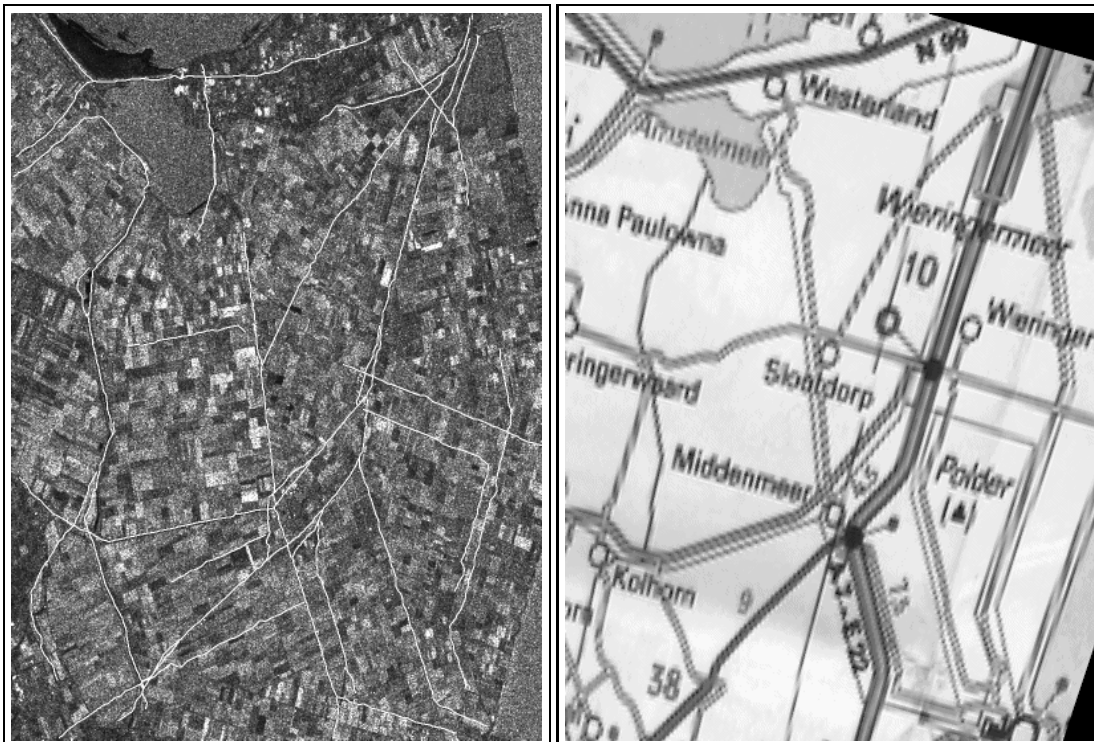
c. Ensemble des 8891 segments constituant les sites du graphe.

FIG. 7.12 – Les étapes de la construction du graphe



a. Image originale (Wieringermeer) ©ESA.

b. Image des segments détectés.



c. Résultat de la détection markovienne.

d. Carte recalée de la zone de l'image.

FIG. 7.13 – Détection des routes sur le Wieringermeer

Chapitre 8

Equations aux dérivées partielles et traitement d'images

Chapitre rédigé par Yann GOUSSEAU

Un exemple classique d'équation aux dérivées partielles est l'équation de la chaleur, qui s'écrit, pour une fonction u_0 de $\Omega \subset \mathbb{R}^2$ dans \mathbb{R} :

$$\begin{cases} \frac{\partial u}{\partial t} = \Delta u \stackrel{\text{def}}{=} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \\ u(x, 0) = u_0(x) \end{cases} \quad (8.1)$$

avec des conditions aux bords de Ω . La solution de cette équation est une fonction u de $\Omega \times \mathbb{R}_+$ dans \mathbb{R} , qui décrit la propagation au cours du temps t de la chaleur dans un milieu isotrope ($u(x, t)$ est la quantité de chaleur au point x et à l'instant t). Les équations aux dérivées partielles (E.D.P.) apparaissent naturellement dans de nombreux domaines de la physique, mais leur utilisation en traitement des images peut sembler étrange au premier abord : à quoi correspond l'évolution dans le temps ou l'état stationnaire ? Lorsque l'équation (8.1) est appliquée à une image u , t ne désigne plus le temps mais est un paramètre d'échelle. Partant d'une image u_0 et résolvant cette équation, nous obtenons une suite d'images $u(\cdot, t)$, images "simplifiées" à l'échelle t . Les E.D.P. apparaissent donc dans le cadre d'analyses multi-échelles, et sont à ce titre utilisées en débruitage, en restauration, ou en traitement préalable à des applications en reconnaissance des formes. Ces équations sont également essentielles dans le cadre des problèmes variationnels, ou l'on cherche une image (ou une représentation associée à l'image) qui minimise une certaine énergie. La descente de gradient associée à cette énergie fait alors apparaître une E.D.P. dont la résolution fournit une solution au problème d'optimisation considéré. Cette approche est utilisée dans des domaines aussi variés que la détection de contours, la restauration, le recallage d'images, etc.

Dans ce chapitre, nous envisagerons les E.D.P. essentiellement sous l'angle de l'analyse multi-échelles. En particulier nous n'aborderons qu'en passant le lien entre E.D.P. et problèmes variationnels. Les résultats mathématiques seront admis, et l'accent sera mis sur quelques exemples d'équations illustrant les points nous paraissant les plus importants. Après avoir précisé l'effet de l'équation de la chaleur sur les images, nous expliquerons pourquoi certaines équations non-linéaires sont mieux adaptées à leur structure, et donnerons un aperçu de l'importance des E.D.P. dans le cadre de l'analyse multi-échelles.

8.1 L'équation de la chaleur et ses limitations

8.1.1 Quelques notations

Avant de rentrer dans le vif du sujet, il est nécessaire de préciser quelques notations. Un point de \mathbb{R}^2 sera noté $\mathbf{a} = (a_1, a_2)$ et sa norme $|\mathbf{a}| = (a_1^2 + a_2^2)^{\frac{1}{2}}$. Le produit scalaire entre \mathbf{a} et \mathbf{b} est $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2$. La fonction indicatrice d'un ensemble A de \mathbb{R}^2 sera notée $\mathbb{1}_A$.

Pour une fonction réelle f , que nous supposons C^2 (deux fois continuellement différentiable) nous noterons ses dérivées partielles :

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_{xy} = \frac{\partial^2 f}{\partial x \partial y}, \quad \dots$$

Le gradient de f est défini par :

$$\nabla f = (f_x, f_y).$$

Le laplacien de f est défini par :

$$\Delta f = f_{xx} + f_{yy}.$$

8.1.2 Pourquoi l'équation de la chaleur

Dans les années 60, Gabor a remarqué que la différence entre une image nette u_0 et cette même image floue était en première approximation proportionnelle au laplacien de u_0 , Δu_0 . Le petit calcul qui suit explique ce phénomène dans le cas où l'image est rendue floue par moyennage locale. Pour r positif, nous définissons l'opérateur de moyennage local par :

$$M_r(u_0)(\mathbf{x}) = \frac{1}{\pi r^2} \int_{D(\mathbf{x}, r)} u_0(\mathbf{x}') d\mathbf{x}',$$

où $D(\mathbf{x}, r)$ est le disque de centre \mathbf{x} et de rayon r . En effectuant un développement de Taylor au voisinage de \mathbf{x} il est facile de voir que

$$\frac{M_r(u_0)(\mathbf{x}) - u_0(\mathbf{x})}{r^2} = \frac{1}{8} \Delta u_0(\mathbf{x}) + \epsilon(r),$$

où ϵ est une fonction qui tend vers 0 si r tend vers 0. Si nous notons alors M_r^n l'opérateur correspondant à n applications successives de M_r , il est possible de montrer que, si nr^2 tends vers un nombre positif t et n vers l'infini, alors $M_r^n(u_0)(\mathbf{x})$ converge vers $u(\mathbf{x}, t)$, solution de l'équation

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{8} \Delta u, \\ u(\cdot, t) = u_0. \end{cases} \quad (8.2)$$

Ce résultat signifie que si l'on construit une suite d'images de plus en plus lissées en partant de u_0 , on obtient asymptotiquement une suite $u(\cdot, t)$ solution de l'équation de la chaleur (le facteur 1/8 peut être supprimé en normalisant le filtre). Plus généralement, ce type de relation reste vérifié si M_r est remplacé par un opérateur de convolution avec un noyau isotrope g de masse unité (c'est à dire que $\int g = 1$) suffisamment régulier (voir [Guichard et Morel, 2002]). En ce sens, l'équation de la chaleur est caractéristique d'un grand nombre de filtrages linéaires, et nous allons donc maintenant nous intéresser à cette équation, en commençant par donner une construction explicite de ses solutions.

8.1.3 Equation de la chaleur et noyau gaussien

Soit $C = [0, 1]^2$ le cube unité. Soit u_0 une fonction intégrable de C dans \mathbb{R} (une image). Nous prolongeons u_0 par symétrie par rapport aux axes Ox et Oy , de façon à obtenir une fonction définie sur $[-1, 1]^2$. Puis nous prolongeons cette fonction à \mathbb{R}^2 entier par périodisation. Ce prolongement a pour but de pouvoir appliquer des convolutions à u_0 . Nous obtenons ainsi une fonction symétrique par rapport aux axes, 2-périodique en chacune de ses variables. Nous noterons dans la suite L_C l'ensemble des fonctions intégrables ayant ces propriétés. C'est dans le cadre de cet espace (dont un élément peut être vu comme une image représentée par une fonction de C dans \mathbb{R}) que nous donnons la forme des solutions de l'équation de la chaleur.

Soit

$$G_t(\mathbf{x}) = \frac{1}{4\pi t} \exp\left(-\frac{|\mathbf{x}|^2}{4t}\right),$$

et $u_0 \in L_C$. Alors $u(\cdot, t) = G_t * u_0$ est solution de l'équation de la chaleur dans L_C , c'est à dire que

$$\begin{cases} \frac{\partial u}{\partial t} = \Delta u, \\ \int_C |u(t, \mathbf{x}) - u_0(\mathbf{x})| \xrightarrow[t \rightarrow 0]{} 0. \end{cases} \quad (8.3)$$

Ce résultat est ici admis, le lecteur intéressé peut se reporter à [Guichard et Morel, 2002]. Les bases mathématiques nécessaires à la compréhension de ces résultats pourront être trouvées dans [Rudin, 1998]. Pour une étude plus complète de l'équation de la chaleur et de nombreuses autres E.D.P., on se reportera à [Evans, 1998]. Numériquement, nous obtenons donc les fonctions $u(\cdot, t)$ en convoluant l'image originale avec des noyaux gaussiens et en utilisant des conditions de symétrie miroir aux bords de l'image. Dans ce chapitre, nous n'aborderons pas les aspects numériques liés aux E.D.P., et nous renvoyons le lecteur intéressé aux références [Weickert et al., 1998] et [Guichard et Morel, 2002].

8.1.4 Application aux images

Comme nous l'indiquons en introduction, les E.D.P. servent entre autre à simplifier graduellement les images, selon le paramètre t , de manière à en simplifier la structure et à supprimer le bruit. Sur la figure 8.1, nous présentons une série d'images auxquelles a été appliquée l'équation de la chaleur (8.1) pour différents temps t . Nous remarquons que les images sont de plus en plus simplifiées lorsque t augmente. En particulier, une telle série peut être utilisée lors de la recherche de structures géométriques, tels des contours. A chaque valeur de t ("échelle") correspond un ensemble de structures, qui doivent ensuite être traitées ensemble. Une telle approche est présentée dans l'article [Witkin, 1983b]. Cependant, nous remarquons également que la structure des images est fortement altérée par l'équation de diffusion (8.1). Les images deviennent floues. En particulier, les discontinuités ("contours") sont lissés. Ces défauts de l'équation de la chaleur seront précisés au paragraphe 8.1.6.

8.1.5 L'équation de la chaleur inverse

Nous avons vu précédemment que l'équation de la chaleur permet de lisser une image à l'échelle t , et que partant d'une image u_0 , nous obtenons une suite d'images $u_t = G_t * u_0$, images de plus en plus floues. Nous avons également mentionné que cette suite d'images était asymptotiquement représentative d'un grand nombre de filtrages linéaires. Dans le cadre de la restauration d'images, il semble donc raisonnable, partant d'une image originale floue, de lui appliquer l'équation de la chaleur inverse, c'est à dire

$$\frac{\partial u}{\partial t} = -\Delta u. \quad (8.4)$$

Cette méthode donne des résultats satisfaisant pour des valeurs de t petites (figure 8.3), mais est numériquement très instable ce qui limite grandement son utilisation. En effet, l'équation (8.4) est mal posée, et est très sensible



FIG. 8.1 – Application de l'équation de la chaleur. De haut en bas et de gauche à droite : image originale u_0 ; images auxquelles a été appliquée l'équation (8.1), respectivement pour des temps $t = 1, 3, 5, 10$. On remarque en particulier que les détails disparaissent et que les images deviennent floues lorsque t augmente.

aux conditions initiales. Ainsi, le résultat explose pour des temps t très rapidement atteints en pratique, et cette explosion est encore accélérée si l'on bruite l'image, ou si on la quantifie. Cette instabilité est illustrée sur la figure 8.4, où l'on applique successivement l'équation de la chaleur et l'équation inverse. Pour $t = 3$ (images (c) et (d)), nous remarquons l'apparition d'oscillations sur l'image restaurée. En présence de bruit, (images (e) et (f)), nous ne voyons presque plus rien, car certaines valeurs sont devenues très grandes en valeur absolue devant les valeurs maximales de l'image originale. Il est ainsi impossible d'utiliser cette équation pour résoudre un problème du type déconvolution en présence de bruit, et son utilité se limite à l'amélioration d'image, avec des temps t très petits (voir figure 8.3, (a), (b)). Au chapitre 8.2.5, nous présenterons une équation non-linéaire qui permet une restauration stable numériquement.

8.1.6 Limitations, invariance par changement de contraste

Comme nous l'avons vu sur les exemples qui précèdent, l'équation de la chaleur permet de simplifier graduellement une image et de se débarrasser d'oscillations indésirables, au prix d'une dégradation de leur structure. Cette

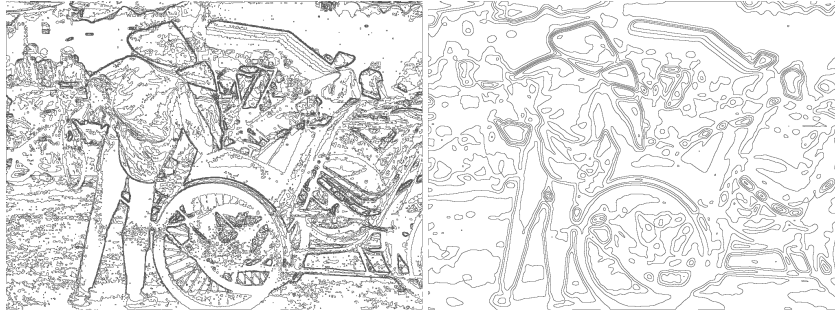


FIG. 8.2 – Gauche : quelques lignes de niveau de l'image originale (pour les niveaux multiples de 30); droite, lignes de niveau de l'évolution au temps $t = 3$. On remarque que le long des contours, endroit où les lignes de niveau s'accumulent, l'équation de la chaleur 8.1 les éloigne les unes des autres.



FIG. 8.3 – Equation de la chaleur inverse. A gauche : image originale u_0 ; à droite : application de l'équation (8.4) pour $t = 0, 2$.

dégradation est flagrante sur les ensembles de niveau de l'image. Nous définissons l'ensemble de niveau λ , associé à l'image f définie sur le domaine Ω , par

$$\chi_\lambda(f) = \{x \in \Omega \mid f(x) \geq \lambda\}.$$

Un contour dans une image correspond à un saut brusque de niveau de gris, et donc à une accumulation des frontières des ensembles de niveau, appelées lignes de niveau. Ceci est visible sur la figure 8.2, où nous montrons une image et quelques-unes de ses lignes de niveau (pour tous les niveaux λ multiples de 30). Sur la même figure, à droite, nous montrons ce que sont devenues ces lignes après lissage par l'équation de la chaleur. Nous voyons nettement que les lignes se sont écartées les unes des autres, ce qui se traduit sur l'image par un effet de flou. L'image a donc été simplifiée, mais ses contours ont été détruits.

Plus formellement, l'équation de la chaleur n'est pas invariante par changement de contraste. Nous appelons changement de contraste une transformation

$$f \rightarrow g \circ f,$$

où g est une fonction croissante de \mathbb{R} dans \mathbb{R} . Il est bien connu que ce type de changement n'altère pas notre perception de la structure d'une image, et il suffit pour s'en convaincre de se munir de lunettes de soleil ou de changer les réglages d'une télévision. Insistons sur le fait que nous parlons ici de perception de la structure, et que ceci n'est pas contradictoire avec le fait qu'un changement de contraste peut améliorer la qualité d'une image, voir

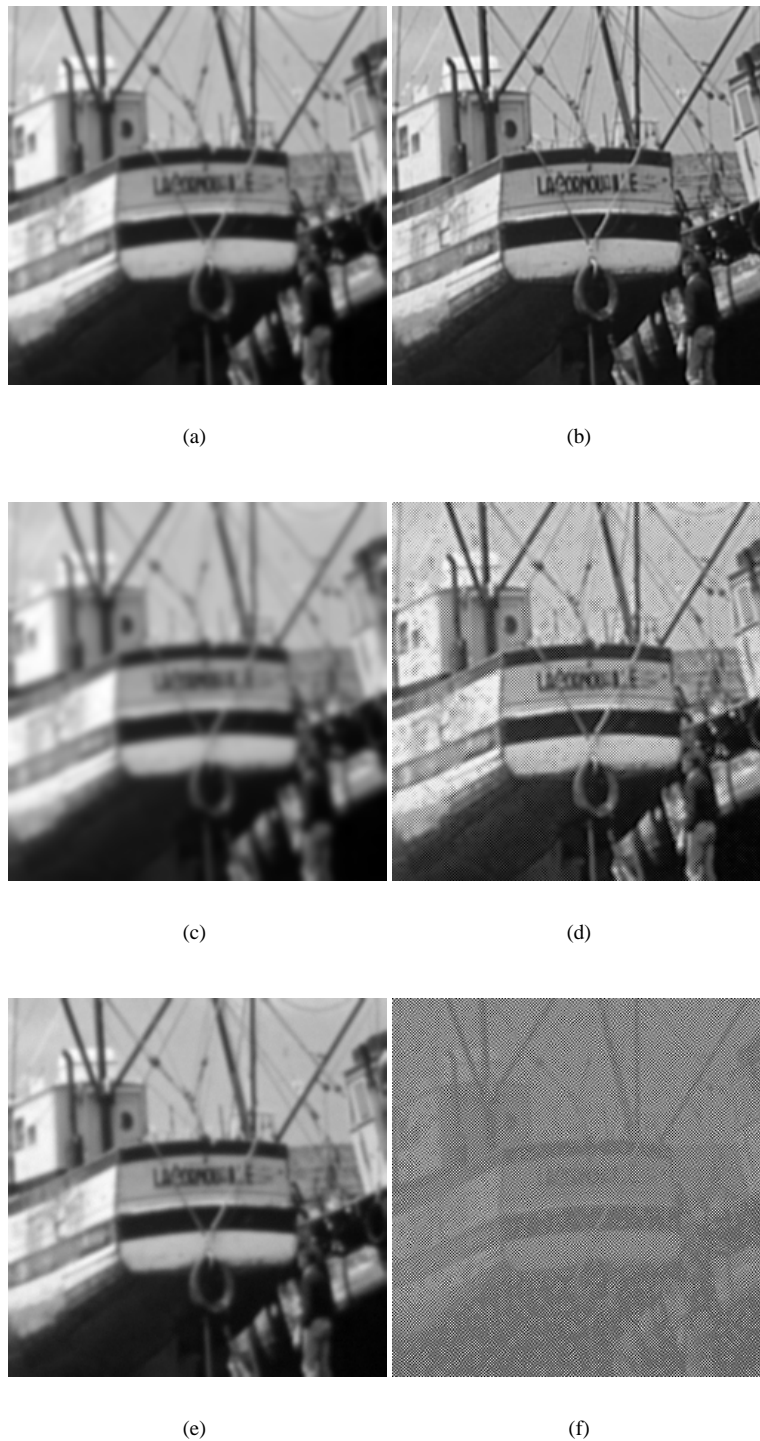


FIG. 8.4 – **(a)** : diffusion selon l'équation de la chaleur (8.1) pour $t = 1$; **(b)** : diffusion " inverse " selon (8.4) à partir de l'image (a), pour $t = 1$; **(c)** et **(d)** : même expérience, avec $t = 3$, on remarque un début d'instabilité ; **(e)** et **(f)** : même expérience avec $t = 1$, mais un bruit gaussien d'écart type 1 a été ajouté à la première image. Les valeurs de niveau de gris explosent en plusieurs points et le résultat est inexploitable.

le chapitre 10. Par ailleurs, des changements de contraste interviennent dans la chaîne d'acquisition des images. Pour ces raisons, nous nous intéressons aux modifications d'image invariantes par changement de contraste : nous dirons qu'un opérateur T (agissant sur les images) est invariant par changement de contraste si :

$$T(g \circ f) = g \circ T(f), \quad (8.5)$$

pour toute fonction réelle g , croissante et continue. Or si T_t désigne l'opérateur envoyant une image u_0 sur $u(\cdot, t)$, solution de l'équation de la chaleur au temps t avec u_0 comme condition initiale, il est facile de voir que $g \circ T_t(u_0)$ n'est plus solution de l'équation (8.1) (le vérifier !), et donc que T_t n'est pas invariant par changement de contraste. Précisons que l'invariance par changement de contraste est liée à l'évolution des ensembles de niveau de la façon suivante : si T est un opérateur vérifiant (8.5) et monotone (c'est à dire que si $u > v$, alors $Tu > Tv$), alors

$$T(\mathbb{I}_{\chi_\lambda(f)}) = \mathbb{I}_{\chi_\lambda(T(f))}. \quad (8.6)$$

L'invariance par changement de contraste est donc équivalente à une action sur les lignes de niveau. Plus généralement, l'équation (8.6) permet de ramener l'étude des opérateurs sur les images invariants par changement de contraste à celle d'opérateurs géométriques agissant sur des ensembles du plan. Bien que nous ne détaillons pas ici cet aspect, nous voyons poindre le lien entre ces équations aux dérivées partielles et la morphologie mathématique (voir le chapitre 6) : toute l'information d'une image est contenue dans ses ensembles de niveau. Le lecteur intéressé se reportera à l'ouvrage [Guichard et Morel, 2002].

Nous verrons au paragraphe 8.3.3 que, sous des hypothèses assez faibles, la préservation des discontinuités implique l'utilisation d'équations non-linéaires. Nous allons maintenant voir comment l'utilisation de ces équations permet de simplifier les images en respectant mieux leur structure.

8.2 Equations de diffusion non-linéaires

La recherche d'E.D.P. permettant une analyse multi-échelles des images tout en préservant leurs discontinuités a été à l'origine d'un grand nombre d'équations dans les années 90. Dans ce chapitre, nous présentons quelques-unes de ces équations, qui ont en commun de modifier les images en fonction de leur structure locale. Pour mieux comprendre leurs effets, il est nécessaire de donner quelques définitions et résultats sur la structure locale des fonctions de deux variables, ce qui est l'objet du paragraphe suivant.

8.2.1 Notations et structure locale des images

Nous considérons une fonction f de Ω dans \mathbb{R} , supposons $f \in C^2$, et nous plaçons en un point \mathbf{a} de Ω . En ce point, passe la ligne de niveau associée au paramètre $\lambda = f(\mathbf{a})$, c'est à dire la frontière de $\chi_{f(\mathbf{a})}(f)$, courbe deux fois différentiable. Le long de cette ligne, la fonction est constante égale à $f(\mathbf{a})$. Un premier résultat important est que le gradient de f en \mathbf{a} , c'est à dire le vecteur $\nabla f(\mathbf{a}) = (f_x(\mathbf{a}), f_y(\mathbf{a}))$, est perpendiculaire à cette ligne de niveau (lorsqu'il est non nul). Si nous supposons que $|\nabla f(\mathbf{a})| \neq 0$, nous pouvons définir un système de coordonnées locales en ce point (voir la figure 8.5) :

$$\eta = \frac{\nabla f}{|\nabla f|} \quad \text{et} \quad \xi = \frac{\nabla f^\perp}{|\nabla f^\perp|},$$

avec $\nabla f^\perp = (-u_y, u_x)$.

Nous définissons la matrice hessienne de f :

$$Hf = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

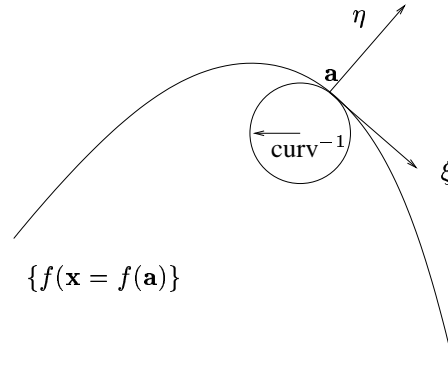


FIG. 8.5 –

et notons également Hf pour la forme quadratique associée, c'est à dire que pour deux vecteurs X et Y , nous avons $Hf(X, Y) = X^t Hf Y$, où X^t est le transposé de X .

Nous pouvons montrer les formules suivantes, qui permettent de calculer les dérivées secondes selon η et ξ avec les coordonnées euclidiennes :

$$Hf \left(\frac{\nabla f}{|\nabla f|}, \frac{\nabla f}{|\nabla f|} \right) = f_{\eta\eta} \quad \text{et} \quad Hf \left(\frac{\nabla f^\perp}{|\nabla f|}, \frac{\nabla f^\perp}{|\nabla f|} \right) = f_{\xi\xi}.$$

Pour finir, nous définissons la courbure de l'image au point \mathbf{a} (où l'on a supposé que $|\nabla f| \neq 0$) par :

$$\text{curv}(f)(\mathbf{a}) = \frac{1}{|\nabla f|^3} Hf(\nabla f^\perp, \nabla f^\perp)(\mathbf{a}) = \frac{f_{xx}f_y^2 - 2f_{xy}f_xf_y + f_{yy}f_x^2}{(f_x^2 + f_y^2)^{\frac{3}{2}}}(\mathbf{a}).$$

Ceci est équivalent à :

$$\text{curv}(f)(\mathbf{a}) = \text{div} \left(\frac{\nabla f}{|\nabla f|} \right)(\mathbf{a}) = \frac{f_{\xi\xi}}{|\nabla f|}.$$

formule où la divergence est définie, pour une fonction vectorielle $A(x, y) = (A_1(x, y), A_2(x, y))$ par :

$$\text{div}(A) = \frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y}.$$

On peut montrer que $\text{curv}(f)(\mathbf{a})$ est la courbure de la ligne de niveau passant par \mathbf{a} , c'est à dire l'inverse du rayon du cercle osculateur à la courbe $\{f(\mathbf{x}) = f(\mathbf{a})\}$ en \mathbf{a} (voir la figure 8.5).

8.2.2 L'équation de Malik et Perona

L'introduction des équations de diffusion non-linéaires dans le domaine du traitement des images remonte à un article de Malik et Perona de 1987, ([Perona et Malik, 1990a]). Remarquons que l'équation de la chaleur (8.1) peut s'écrire

$$\frac{\partial u}{\partial t} = \Delta u = \text{div}(\nabla u).$$

La faiblesse de cette équation est que la diffusion est identique en tout point de l'image (cette équation modélise initialement la diffusion de la chaleur dans un milieu isotrope). En particulier, comme nous l'avons vu au paragraphe précédent, l'image est lissée aussi bien dans les zones homogènes que le long des contours. L'idée de Malik

et Perona est de lisser l'image dans les zones homogènes, et de ne pas faire évoluer l'image le long des contours, voire de rehausser ces derniers, comme nous allons le voir plus précisément. L'équation correspondante s'écrit :

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(|\nabla u|)\nabla u), \quad (8.7)$$

avec g une fonction décroissante, valant 1 en zéro, et tendant vers 0 en l'infini. L'équation se rapproche donc de (8.1) aux points où $|\nabla u|$ est proche de 0. A titre d'exemple, nous considérerons $g(s) = \frac{1}{1+(\lambda s)^2}$, un des choix de fonction g proposé par l'article original de Malik et Perona. Pour comprendre ce qui se passe aux points où $|\nabla u|$ est grand, nous réécrivons l'équation sous la forme

$$\frac{\partial u}{\partial t} = \frac{1}{1 + \lambda^2 |\nabla u|^2} u_{\xi\xi} + \frac{(1 - \lambda^2 |\nabla u|^2)}{(1 + \lambda^2 |\nabla u|^2)^2} u_{\eta\eta}, \quad (8.8)$$

avec les notations du paragraphe 8.2.1. Contrairement au cas de l'équation de la chaleur (8.1), qui peut se réécrire $\frac{\partial u}{\partial t} = \Delta u = u_{\xi\xi} + u_{\eta\eta}$, l'évolution varie en fonction des valeurs de $|\nabla u|$ (norme et direction). Le premier terme de l'équation (8.8) représente une diffusion uni-dimensionnelle dans la direction orthogonale au gradient, alors que le deuxième terme prend des valeurs positives ou négatives selon la valeur de $|\nabla u|$, ce qui correspond respectivement à une équation de la chaleur directe ou inverse (comme étudié au paragraphe 8.1.5). Remarquons l'importance du paramètre λ , qui impose le type de diffusion en fonction des valeurs de $|\nabla u|$. Ainsi, cette équation se propose de résoudre à la fois un problème d'analyse (l'image est simplifiée à l'échelle t), et un problème de restauration (rehaussement des contours), qui dépend d'un seuil sur le contraste. Signalons également que rien ne garantit l'existence et l'unicité des solutions éventuelles de cette équation. Enfin remarquons que lors de l'utilisation de cette équation comme prétraitement avant un détecteur de contours du type Canny (voir le chapitre 11), le choix du paramètre λ correspondra au seuil que l'on se fixe pour $|\nabla u|$. Sur la figure 8.6, nous montrons les évolutions de l'image u_0 de la figure 8.1 sous l'effet de l'équation (8.7). A nouveau, nous obtenons une suite d'images de plus en plus simplifiées, ce qui sera utile par exemple pour une recherche de structures. Remarquez que les contours sont mieux préservés que sur la figure 8.1, mais qu'ils ne sont pas traités de la même façon selon les valeurs du contraste (de $|\nabla u|$). En particulier, les bords du chapeau très contrastés sont renforcés (comportement du type équation de la chaleur inverse), alors que les bords de la roue moins contrastés deviennent flous lorsque t augmente.

8.2.3 Mouvement par courbure moyenne et variantes

Le type le plus simple d'équation de diffusion du type anisotrope s'écrit :

$$\frac{\partial u}{\partial t} = u_{\xi\xi}, \quad (8.9)$$

et a originellement été introduite par Sethian, [Sethian, 1985]. Il y a diffusion dans la direction perpendiculaire au gradient, alors qu'il ne se passe rien dans la direction du gradient. Avec les résultats de la section 8.2.1, cette équation peut se réécrire :

$$\frac{\partial u}{\partial t} = |\nabla u| \operatorname{curv}(u).$$

Signalons que l'équation (8.9) est invariante par changement de contraste. Nous verrons plus loin que ce type de mouvement selon la courbure est inévitable si l'on s'intéresse aux équations invariantes par changement de contraste. Rappelons que les modifications d'images invariantes par changement de contraste peuvent être vues comme une évolution des ensembles de niveau de l'image, ce qui est exprimé par l'équation (8.6). Il est possible de montrer que l'équation (8.9) fait évoluer les lignes de niveau de l'image selon leur courbure, c'est à dire qu'en chaque point \mathbf{a} , la ligne de niveau passant par \mathbf{a} (c'est à dire la courbe $\{u(\mathbf{x}) = u(\mathbf{a})\}$) évolue dans la direction de sa normale en \mathbf{a} (la direction de $\nabla u(\mathbf{a})$), à la vitesse $\operatorname{curv}(u)(\mathbf{a})$. En particulier, les lignes droites n'évoluent pas sous l'effet de (8.9), et les oscillations brusques tendent à disparaître. Du fait de ces propriétés, cette équation a été



FIG. 8.6 – Equation de Malik et Perona. Même image de départ que sur la figure 8.1. De haut en bas et de gauche à droite : $t = 1, 3, 5, 10$.

proposée dans le cadre de l'analyse de formes, voir [Kimia et al., 1992]. Nous illustrons l'effet de cette équation figure 8.7, pour plusieurs temps t . Remarquez en particulier l'évolution du chapeau : les bords rectilignes évoluent peu tandis que la pointe voit sa courbure diminuer avec le temps t . Observez également la différence avec la figure 8.6 : les contours restent nets tant qu'ils sont présents, et ce indépendamment du contraste (des valeurs de $|\nabla u|$) le long du contour. Sur la figure 8.8, nous montrons les lignes de niveau de l'image u au temps $t = 3$; notez en particulier les différences avec la figure 8.2.

Remarque : il est possible de montrer (voir [Guichard et Morel, 2002]) que, de même que l'équation de la chaleur est l'équation associée asymptotiquement à l'itération de moyennages locaux (paragraphe (8.1.2)), le mouvement par courbure moyenne (8.9) est asymptotiquement équivalent à l'itération de filtres médians (filtres non-linéaires définis au paragraphe 10.2.2).

Nous verrons au paragraphe 8.3 que les équations "raisonnables" pour l'analyse des images qui sont invariantes par changement de contraste et isométries (rotations-translations) s'écrivent sous la forme :

$$\frac{\partial u}{\partial t} = |\nabla u| G(\text{curv}(u), t), \quad (8.10)$$

avec $G(x, y)$ une fonction réelle continue et décroissante en x . Le mouvement par courbure moyenne correspond au cas $G(x, y) = x$. Si l'on impose de plus l'invariance par transformations affines, il ne reste plus qu'une seule équation, introduite en 1993 par Alvarez, Guichard, Lions et Morel (voir [Alvarez et al., 1993]) :

$$\frac{\partial u}{\partial t} = |\nabla u| \text{sign}(\text{curv}(u)) (\text{curv}(u))^{1/3}. \quad (8.11)$$

Ces différents résultats sont précisés au paragraphe 8.3.3.

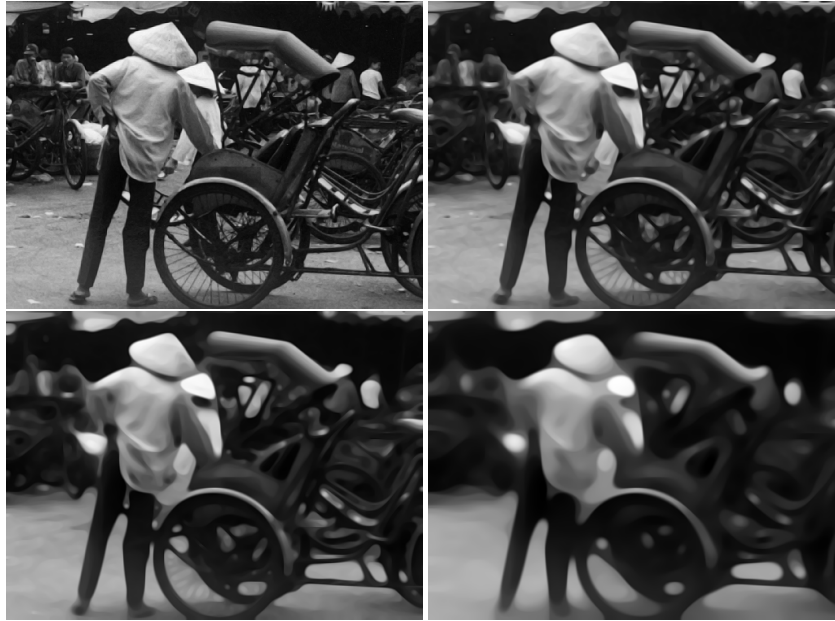


FIG. 8.7 – Evolution par courbure moyenne. L'image originale est l'image u_0 de la figure 8.1. De haut en bas et de gauche à droite, les solutions de l'équation (8.9) respectivement pour $t = 1, t = 3, t = 5, t = 10$.



FIG. 8.8 – Lignes de niveau de la solution de (8.9), mouvement par courbure moyenne, pour $t = 3$ (niveaux multiples de 30).

8.2.4 L'équation de Rudin-Osher-Fatemi

Les équations aux dérivées partielles apparaissent dans le cadre d'approches variationnelles de problèmes de traitement des images, c'est à dire lorsque l'on cherche à résoudre un problème en minimisant une énergie. Rudin, Osher et Fatemi, dans un article de 1992 ([Rudin et al., 1992a]) s'intéressent à l'énergie :

$$E(f) = \alpha \int_{\Omega} (f - f_0)^2 + \int_{\Omega} |\nabla f|, \quad (8.12)$$

où Ω désigne toujours le domaine de définition des images, et α un paramètre réel positif. Minimiser cette énergie sur un ensemble de fonctions (par exemple les fonctions C^1 égales à f_0 sur la frontière de Ω) revient à chercher

une fonction proche de f_0 tout en pénalisant les fonctions trop oscillantes grâce au deuxième terme. L'équation d'Euler-Lagrange associée à cette énergie est $\alpha(f - f_0) - \nabla f = 0$, et la descente de gradient correspondante est :

$$\frac{\partial u}{\partial t} = \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) + \alpha(f_0 - u),$$

avec comme précédemment u une fonction de \mathbf{x} et de t . Si l'on fait abstraction du terme de fidélité aux données ($\alpha = 0$), et en utilisant les résultats du paragraphe 8.2.1, nous pouvons réécrire cette équation sous la forme :

$$\frac{\partial u}{\partial t} = \operatorname{curv}(u).$$

Ces équations sont très utilisées en restauration d'images, car la minimisation d'énergies du type (8.12), comprenant un terme $\int |\nabla f|$ (la "variation totale" de l'image), permet d'obtenir des solutions aux bords francs, sans les oscillations résiduelles qui peuvent apparaître, par exemple, avec l'utilisation du filtre de Wiener (voir paragraphe 5.2.2). Dans le cas de la restauration d'image convoluée par un noyau g et bruitée, l'énergie (8.12) prend la forme

$$E(f) = \alpha \int_{\Omega} (g * f - f_0)^2 + \int_{\Omega} |\nabla f|. \quad (8.13)$$

Ces méthodes sont en particulier utilisées par le Centre National d'Etudes Spatiales (CNES), dans le cadre de l'imagerie satellitaire. Figure 8.9, nous comparons les résultats obtenus par variation totale et filtre de Wiener dans le cas d'une image satellitaire. L'image est convoluée avec un filtre connu, puis lui est ajouté un bruit gaussien de déviation standard connue. La restauration par Wiener se fait comme expliqué au paragraphe 5.2.2, la densité de puissance du signal étant estimée sur l'image dégradée. La restauration par la méthode de Rudin-Osher-Fatemi s'effectue en minimisant l'énergie (8.13) par évolution suffisamment longue de l'équation associée.

Remarque : de la même manière, on montre que l'équation de la chaleur (8.1) est associée à l'énergie $\int |\nabla u|^2$.

8.2.5 Le filtre de choc de Rudin

Nous avons vu au paragraphe 8.1.5 que l'équation de la chaleur pouvait être utilisée à des fins d'amélioration d'images, mais que ceci posait de gros problèmes de stabilité. Une amélioration de cette méthode, appelée "filtre de choc" et due à Rudin ([Rudin, 1987]), consiste à résoudre l'équation :

$$\frac{\partial u}{\partial t} = -\operatorname{sign}(\Delta u) |\nabla u|, \quad (8.14)$$

avec l'image à restaurer comme condition initiale. La diffusion des niveaux de gris s'effectue dans la même direction qu'avec l'équation de la chaleur inverse (8.4), mais est bien plus stable. Il est possible de montrer que cette équation fait évoluer les lignes de niveau à une vitesse constante. Sur la figure 8.10, nous reprenons les expériences présentées pour l'équation de la chaleur inverse, pour illustrer l'effet du filtre de choc ainsi que sa stabilité. Remarquons en particulier que la restauration se fait au prix d'une perte des détails et de la texture : pour des valeurs croissantes de t , l'image filtrée prend une allure de type bande dessinée. Sur la dernière expérience (8.10 (f)) on remarque que la présence de bruit sur l'image initiale (e) entraîne l'apparition d'effilochement le long des contours.

8.3 EDP et analyse multi-échelles

Dans les chapitres précédents, nous avons vu des exemples d'analyse multi-échelles des images, sous la forme d'équations aux dérivées partielles. Dans ce chapitre, nous formalisons cette notion, et donnons, sans preuves, les éléments permettant une classification de ces analyses en fonction de leurs invariances, notamment géométriques. Nous allons voir que les E.D.P. apparaissent naturellement dans ce cadre, et nous allons retrouver des équations vues ci-dessus. Plus précisément, nous allons voir comment ramener le concept d'analyse multi-échelles à des filtres itérés, filtres que nous avons déjà rencontrés au paragraphe 8.1.2.

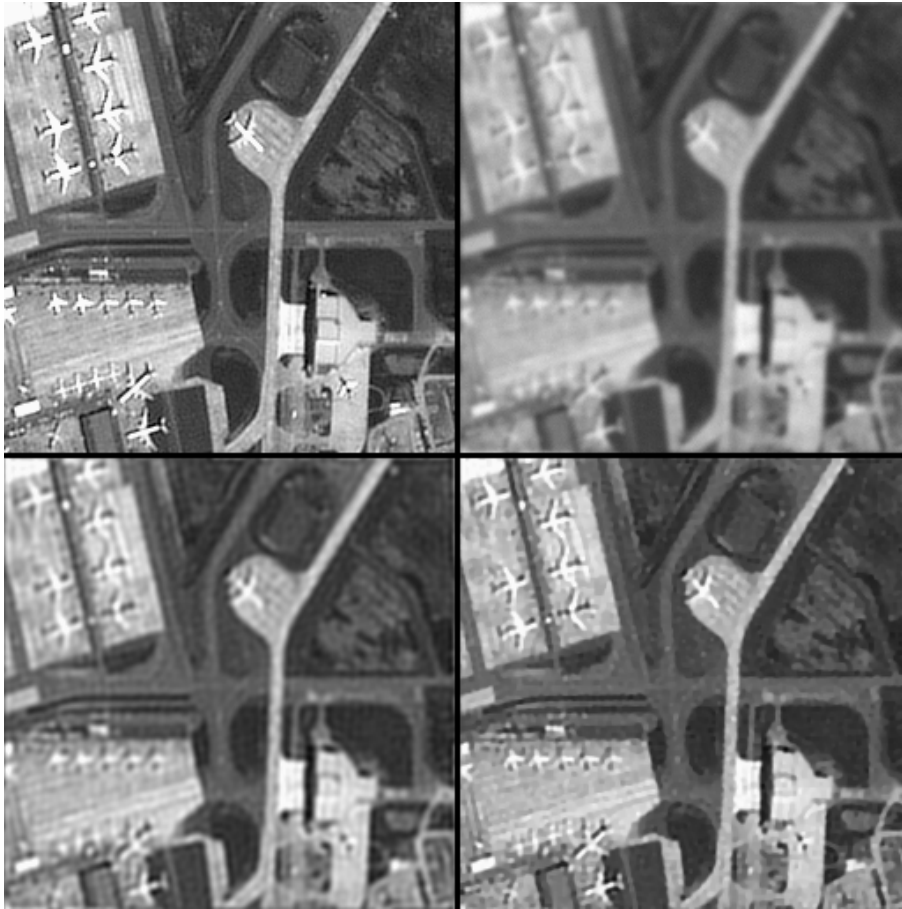


FIG. 8.9 – Exemple de restauration. De haut en bas et de gauche à droite : image originale (image Spot 5 ©CNES 2002); image dégradée par convolution et ajout de bruit ; image restaurée par Wiener ; image restaurée par minimisation de l'énergie (8.13).

8.3.1 Définition et propriétés des analyses multi-échelles

Nous définissons une "analyse multi-échelles" comme une famille d'opérateurs T_t sur les images, le paramètre t étant appelé paramètre d'échelle. Etant donné une image $u_0(\mathbf{a})$, $(T_t u_0)(\mathbf{a}) = u(t, \mathbf{a})$ est l'image u_0 analysée à l'échelle t . Nous donnons maintenant quelques propriétés désirables pour ces familles d'opérateurs :

Structure pyramidale. La famille T_t a cette structure s'il existe une famille d'opérateurs $T_{t+h,t}$, appelés "filtres de transition", tels que :

$$T_{t+h} = T_{t+h,t} T_t, \quad T_0 = Id. \quad (8.15)$$

Cette propriété signifie que l'image à une échelle plus grossière peut se déduire d'une version à une échelle plus fine, sans nécessiter l'image originale. Un cas particulier de cette propriété est le cas où l'analyse multi-échelles est **récurive**, c'est à dire que $T_{t+h,t} = T_h$ (propriété de semi-groupe).

Nous allons considérer des opérateurs causaux, c'est à dire qui ne créent pas de nouvelle structure lorsque l'échelle augmente. Une autre propriété désirable pour une analyse multi-échelles est la localité. Dans le cas où

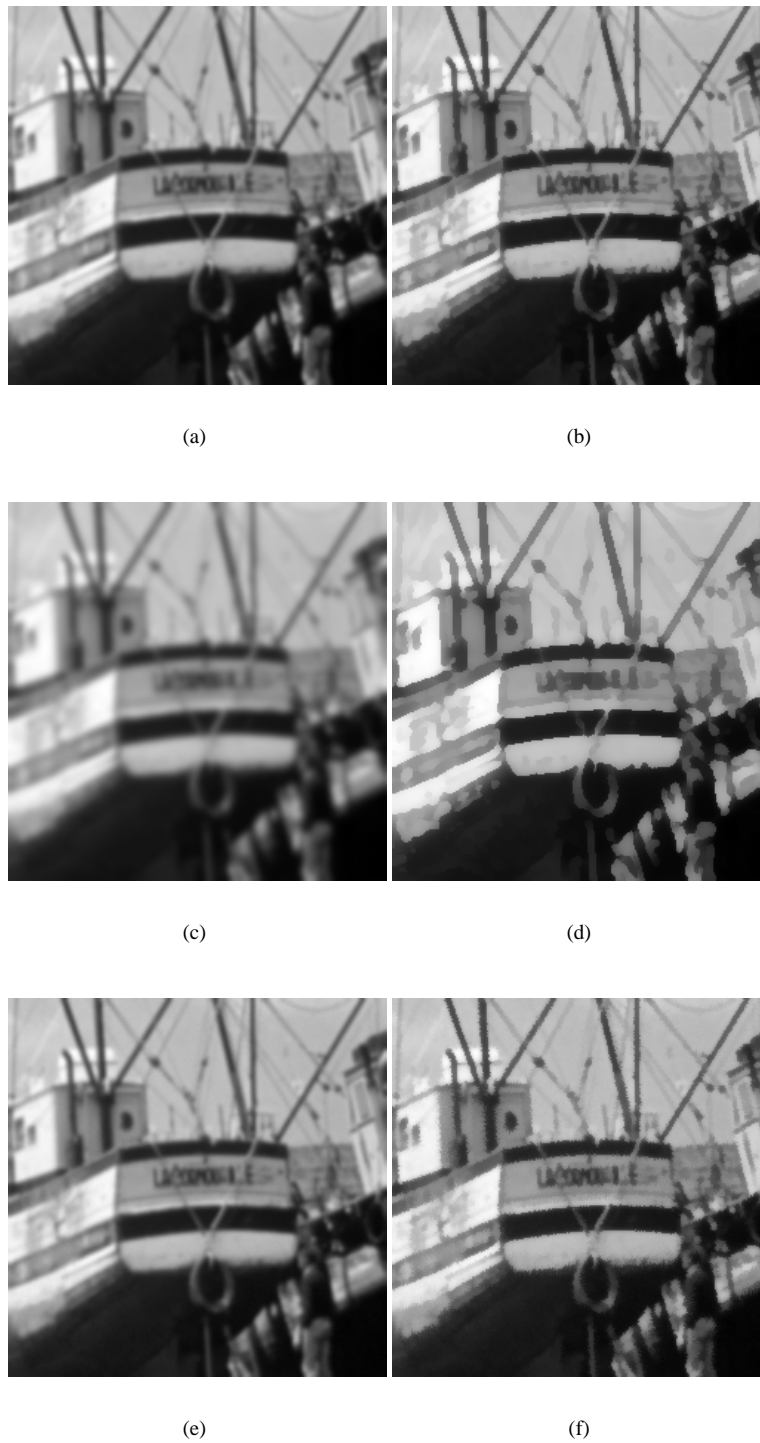


FIG. 8.10 – Filtre de choc de Rudin. Les images (a), (c), (e) sont identiques à celles de la figure 8.3. Les images (b), (d), (f) sont les applications correspondantes de l'équation (8.14) (respectivement $t = 1, t = 3, t = 1$).

les opérateurs T_t sont pyramidaux, le résultat de l'application du filtre de transition en un point, $T_{t+h,t}u_0(\mathbf{a})$, ne doit dépendre que des valeurs de u_0 dans un voisinage de \mathbf{a} . Ces deux principes, de localité et de causalité, sont condensés dans le principe suivant, qui affirme que si une image est localement plus claire qu'une autre, il en sera de même pour les deux familles multi-échelles correspondantes pendant un certain temps :

Principe de comparaison locale. Une analyse multi-échelles (T_t) est dite satisfaire le principe de comparaison locale, si pour toutes fonctions f_1 et f_2 :

- si $f_1(\mathbf{b}) > f_2(\mathbf{b})$ dans un voisinage de \mathbf{a} , $\mathbf{b} \neq \mathbf{a}$, alors pour h suffisamment petit,

$$(T_{t+h,t}f_1)(\mathbf{a}) \geq (T_{t+h,t}f_2)(\mathbf{a}),$$

et

- si $\forall \mathbf{a} \in \mathbb{R}^2$, $f_1(\mathbf{a}) \geq f_2(\mathbf{a})$, alors :

$$\forall \mathbf{b}, \forall h > 0, \quad (T_{t+h,t}f_1)(\mathbf{b}) \geq (T_{t+h,t}f_2)(\mathbf{b}).$$

Une dernière propriété que nous imposerons à une analyse multi-échelles est une propriété de régularité : une image régulière (par exemple différentiable) évolue régulièrement. Il s'avère suffisant de faire cette hypothèse pour des fonctions quadratiques.

Régularité. Soit \mathbf{a} un point de \mathbb{R}^2 , soit la forme quadratique :

$$f(\mathbf{b}) = \frac{1}{2}A(\mathbf{b} - \mathbf{a}) \cdot (\mathbf{b} - \mathbf{a}) + p \cdot (\mathbf{b} - \mathbf{a}) + c. \quad (8.16)$$

L'analyse multi-échelles est dite régulière s'il existe une fonction $F(A, p, c, \mathbf{a}, t)$, continue par rapport à A , telle que :

$$\frac{(T_{t+h,t}f - f)(\mathbf{a})}{h} \rightarrow F(A, p, c, \mathbf{a}, t) \quad \text{quand } h \rightarrow 0. \quad (8.17)$$

Dans la suite, nous dirons qu'une analyse multi-échelles est **causale** si elle est pyramidale, et satisfait aux principes de comparaison locale et de régularité énoncés ci-dessus. Ce sont ce type d'analyses multi-échelles dont nous allons maintenant clarifier le lien avec les E.D.P., et pour lesquelles nous donnons un début de classification.

8.3.2 Pourquoi les analyse multi-échelles sont régies par des E.D.P.

Nous donnons un résultat qui précise le lien entre analyse multi-échelles et équations aux dérivées partielles. Le lecteur intéressé en trouvera la preuve dans [Guichard et Morel, 2002].

Si une analyse multi-échelles T_t est causale, (c'est à dire pyramidale régulière, et satisfaisant au principe de comparaison locale), alors il existe une fonction F telle que

$$((T_{t+h,t}f - f)/h)(\mathbf{a}) \rightarrow F(Hf(\mathbf{a}), Df(\mathbf{a}), f(\mathbf{a}), \mathbf{a}, t) \quad (8.18)$$

lorsque h tend vers 0^+ , pour toute fonction $f \in C^2$, et tout $\mathbf{a} \in \mathbb{R}^2$.

Rappelons alors que l'opérateur $T_{t+h,t}$ permet de passer du temps t au temps $t + h$, et donc que si dans l'équation ci-dessus nous voyons f comme $u(\cdot, t)$, le terme de gauche, lorsque h tend vers 0, tend vers $\frac{\partial u}{\partial t}(t)$. Le résultat est en réalité non trivial, mais l'on peut montrer que $u = T_t(u_0)$ est solution de :

$$\frac{\partial u}{\partial t} = F(D^2u, Du, u, \mathbf{a}, t),$$

avec $u(\cdot, 0) = u_0$. Ces résultats sont prouvés dans [Guichard et Morel, 2002], grâce à une notion de solution particulière, les solutions de viscosité.

Grâce au résultat ci-dessus, la classification des analyses multi-échelles se ramène à la classification des fonctions F . Nous allons maintenant voir qu'en imposant à l'analyse des contraintes liées à la structure des images, les formes possibles de la fonction F sont grandement réduites.

Commençons par l'invariance par addition d'une constante. Supposons que

$$T_{t,t+h}(0) = 0, \text{ et } T_{t,t+h}(u + C) = T_{t,t+h}(u) + C, \quad (8.19)$$

pour toute fonction u et toute constante C (un décalage constant de niveau de gris n'affecte pas l'action de l'opérateur T). Il est facile de voir que si T_t est causal et invariant par addition d'une constante, alors $F(A, p, c, \mathbf{a}, t)$ ne dépend pas de c . Ceci se prouve en appliquant (8.17) et (8.19) à la fonction $u + C$, où u est une forme quadratique du type (8.16).

Similairement, nous avons les résultats suivants, pour un opérateur $T_{t,t+h}$ invariant par l'addition de constantes :

- Si $T_{t,t+h}$ est invariant par translation (c'est à dire commute avec l'opérateur $u(\mathbf{a}) \rightarrow u(\mathbf{a} + \mathbf{c})$, pour tout $\mathbf{c} \in \mathbb{R}^2$), alors F ne dépend pas de son quatrième argument, \mathbf{a} .
- Si $T_{t,t+h}$ commute avec les rotations et translations (invariance euclidienne) alors pour toute matrice d'isométrie O ,

$$F(OHuO^t, O\nabla u, t) = F(Hu, \nabla u, t).$$

(Rappelons que F ne dépend pas de \mathbf{a} dans ce cas en vertu de la propriété précédente).

- Si $T_{t+h,t}$ est invariant par changement de contraste, alors pour toutes constantes λ, μ :

$$F(\mu Hu + \lambda \nabla u \otimes \nabla u, \mu \nabla u, t) = \mu F(Hu, \nabla u, t),$$

où \otimes désigne le produit tensoriel, c'est à dire que :

$$p \otimes p = \begin{bmatrix} p_1^2 & p_1 p_2 \\ p_1 p_2 & p_2^2 \end{bmatrix}.$$

Cette dernière propriété est quelque peu obscure, mais est indiquée comme résultat intermédiaire pour établir la forme des analyses multi-échelles invariantes par changement de contraste, donnée ci-dessous au paragraphe 8.3.3.

8.3.3 Classification des analyses multi-échelles

Grâce à la caractérisation des analyse multi-échelles causales, exprimée par l'équation (8.18), nous avons vu que l'étude de ces analyses se ramène à l'étude d'équations différentielles du type :

$$\begin{cases} \frac{\partial u}{\partial t} = F(D^2 u, Du, u, \mathbf{a}, t), \\ u(0) = u_0. \end{cases} \quad (8.20)$$

De plus, des hypothèses simples d'invariance permettent de grandement simplifier les formes possibles de la fonction F . Ces résultats, détaillés dans [Guichard et Morel, 2002], permettent une classification des analyses multi-échelles. Nous donnons maintenant quelques-uns de ces résultats de classification.

Analyse linéaire isotrope

Le premier résultat indique que si l'analyse est linéaire et isotrope, elle est donnée par l'équation de la chaleur, résultat que nous avons entrevu au début du paragraphe 8.1.2.

Si T_t est un opérateur causal, invariant par rotation et translation, et linéaire, alors (à une renormalisation $t' = h(t)$ près) $u(\mathbf{a}, t) = T_t(u_0)(\mathbf{a})$ est solution de :

$$\begin{cases} \partial u / \partial t = \Delta u \text{ in } \mathbb{R}^2 \times [0, \infty[, \\ u(0, \cdot) = u_0(\cdot) \text{ in } \mathbb{R}^2. \end{cases} \quad (8.21)$$

Remarquons que ce résultat implique en particulier que linéarité et invariance par changement de contraste sont incompatibles, puisque l'équation de la chaleur n'est pas contraste invariante.

Invariance par changement de contraste

Nous avons vu au paragraphe 8.1.6 l'importance de l'invariance par changement de contraste en traitement des images, et il est donc naturel de se demander quelles sont les équations satisfaisant à cette invariance. Le résultat est le suivant : si T_t est une analyse multi-échelle causale, invariante par rotation et translation, et invariante par changement de contraste, alors $u(\mathbf{a}, t) = T_t(u_0)(\mathbf{a})$ est solution de l'équation

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

où G est une fonction croissante et continue par rapport à son premier argument. Remarquons que les équations présentées au paragraphe 8.2.3 sont parmi les équations les plus simples de ce type.

Invariance affine

Pour conclure nous nous intéressons aux analyses invariantes par changement de contraste et par transformations affines. Les transformations affines sont d'une grande importance en traitement des images, en particulier parce qu'elles fournissent une approximation simple des transformations projectives qu'induisent sur les images les mouvements des appareils optiques de capture lors du passage d'une scène tridimensionnelle à une photo bidimensionnelle.

Si l'on suppose que T_t est une analyse multi-échelles causale et invariante par transformation affine, alors (à une renormalisation du paramètre t près)

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

où,

$$\begin{cases} \gamma(s) = C_1 s^{\frac{1}{3}} \text{ si } s > 0, \\ \gamma(s) = -C_2 |s|^{\frac{1}{3}} \text{ si } s < 0, \end{cases} \quad (8.22)$$

C_1 et C_2 étant des constantes ([Alvarez et al., 1993]). En imposant des contraintes assez simples et directement liées au processus de formation des images, nous nous sommes donc ramenés à une seule équation possible pour une analyse multi-échelles, appelée AMSS pour l'anglais "Affine Morphological Scale Space". En vertu de ses invariances géométriques, et parce qu'elle préserve les discontinuités, cette équation est en particulier utilisée en reconnaissance de formes.

Signalons enfin que le lecteur désireux d'approfondir le sujet pourra se référer à [Guichard et Morel, 2002], [Aubert et Kornprobst, 2001], [Weickert, 1998].

Chapitre 9

Ondelettes et traitement d'images

Chapitre rédigé par Béatrice PESQUET-POPESCU et
Jean-Christophe PESQUET (Université de Marne la Vallée)

9.1 Principes de l'analyse linéaire des images

Pour analyser linéairement une image, on considère généralement un dictionnaire $\mathcal{D} = \{\psi_\lambda, \lambda \in \Lambda\}$ de fonctions d'analyse, Λ étant un certain ensemble de paramètres inclus dans \mathbb{R}^p , $p \in \mathbb{N}^*$, propre à la famille choisie. Une transformation linéaire T de l'image f , définie sur \mathbb{R}^2 , est alors obtenue en calculant une corrélation entre f et la fonction ψ_λ :

$$T : f \mapsto (c_\lambda)_{\lambda \in \Lambda},$$

où

$$\forall \lambda \in \Lambda, \quad c_\lambda = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \overline{\psi_\lambda(x, y)} dx dy, \quad (9.1)$$

$\overline{(\cdot)}$ désignant le complexe conjugué de la quantité en argument.

La première question qui se pose naturellement, dans cette approche, est de savoir quelles fonctions ψ_λ utiliser. Ce choix se fait souvent en suivant deux impératifs, qu'on rencontre plus généralement dans tous les problèmes de modélisation :

- définir une famille suffisamment riche pour pouvoir modéliser efficacement des images au contenu relativement complexe (telles que celles qu'on rencontre en vision assistée par ordinateur ou encore en vidéo)
- garder des fonctions d'expression relativement simple et caractérisées par un nombre limité de paramètres.

L'exemple le plus classique d'analyse linéaire est la transformée de Fourier (TF). On a alors $\Lambda = \mathbb{R}^2$, $\lambda = (\mu_x, \mu_y)$ et :

$$\psi_\lambda(x, y) = e^{i2\pi(\mu_x x + \mu_y y)}.$$

Cette transformation est bien adaptée à l'analyse des comportements harmoniques, surtout apparents dans les zones texturées des images, mais elle n'est pas appropriée à l'analyse de formes bien localisées spatialement. Pour avoir la possibilité de conserver une analyse spectrale tout en acquérant un pouvoir de localisation spatiale, il faut adopter une approche espace-fréquence (temps-fréquence en 1D [Flandrin, 1998]), en choisissant des fonctions ψ_λ à la fois oscillantes et de support limité. On dit alors que les fonctions ψ_λ constituent des *atomes espace-fréquence*. Plusieurs choix de tels atomes sont envisageables :

- *famille de Weyl-Heisenberg*. On modifie légèrement la TF, en introduisant une fenêtre g . On prend ainsi $\Lambda = \mathbb{R}^4$, $\lambda = (x_0, y_0, \mu_x, \mu_y)$ et :

$$\psi_\lambda(x, y) = g(x - x_0, y - y_0) e^{i2\pi(\mu_x x + \mu_y y)}. \quad (9.2)$$

La TF de cette fonction est :

$$\Psi_\lambda(\nu_x, \nu_y) = G(\nu_x - \mu_x, \nu_y - \mu_y) e^{-i2\pi[x_0(\nu_x - \mu_x) + y_0(\nu_y - \mu_y)]},$$

où G désigne celle de g (supposée définie). En choisissant g centrée et de type jj passe-bas jj (localisée en espace et en fréquence autour de $(0,0)$), on voit que ψ_λ est localisée spatialement autour de (x_0, y_0) et fréquentiellement autour de (μ_x, μ_y) . Le principe de Gabor-Heisenberg [Flandrin, 1998] nous assure que la localisation espace-fréquence est optimale si :

$$g(x, y) = \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}} e^{-\frac{1}{4}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}$$

où $(\sigma_x, \sigma_y) \in (\mathbb{R}_+^*)^2$. Lorsque g prend une forme gaussienne de ce type, on dit qu'on effectue une analyse de Gabor.

- *famille d'ondelettes*. On prend $\Lambda = \mathbb{R}^2 \times \mathbb{R}_+^* \times [0, 2\pi[$, $\lambda = (x_0, y_0, a, \theta)$ et

$$\psi_\lambda(x, y) = \frac{1}{a} \psi^\theta \left(\frac{x - x_0}{a}, \frac{y - y_0}{a} \right).$$

La fonction ψ^θ est appelée une jj ondelette mère jj [Meyer, 1990]. Il s'agit généralement d'une fonction de carré sommable assez régulière, bien localisée en espace et en fréquence (la décroissance asymptotique de la fonction, ainsi que celle de sa TF sont rapides) et telle que :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi^\theta(x, y) dx dy = 0.$$

La TF de ψ_λ vaut :

$$\Psi_\lambda(\nu_x, \nu_y) = a \Psi^\theta(a\nu_x, a\nu_y) e^{-i2\pi(x_0\nu_x + y_0\nu_y)},$$

où Ψ^θ désigne la TF de ψ^θ . Si ψ^θ est localisée spatialement en (x_0, y_0) et fréquentiellement en (μ_x, μ_y) (i.e. $|\Psi^\theta(\nu_x, \nu_y)|$ a un maximum global en (μ_x, μ_y)) alors ψ_λ est localisée spatialement en (x_0, y_0) et fréquentiellement en $(a\mu_x, a\mu_y)$. En faisant varier les paramètres de localisation spatiale (x_0, y_0) et le facteur d'échelle $a > 0$ (analyse espace-échelle), il y a bien la possibilité de réaliser du même coup une analyse espace-fréquence.

Nous n'avons pas encore parlé du rôle du paramètre θ , qui correspond à un angle indiquant la direction d'analyse privilégiée. On peut, par exemple, définir :

$$\psi^\theta(x, y) = \psi \left[R_\theta \begin{pmatrix} x \\ y \end{pmatrix} \right]$$

où ψ est une fonction à deux variables anisotrope et R_θ est la matrice de rotation plane

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Un autre choix usuel est de limiter θ à prendre ses valeurs dans $\{0, \pi/4, \pi/2\}$ et de poser :

$$\begin{aligned} \psi^0(x, y) &= \phi(x)\psi(y) \\ \psi^{\frac{\pi}{4}}(x, y) &= \psi(x)\psi(y) \\ \psi^{\frac{\pi}{2}}(x, y) &= \psi(x)\phi(y) = \psi^0(y, x) \end{aligned}$$

où ψ désigne maintenant une ondelette mère 1D (telle que $\int_{-\infty}^{\infty} \psi(x) dx = 0$) et ϕ est une fonction ψ basse-fréquence ψ localisée en 0 telle que $\int_{-\infty}^{\infty} \phi(x) dx = 1$. On génère ainsi une famille d'ondelettes 2D séparables. Un exemple de telles fonctions est donné par la figure 9.1.

Comparées à la famille de Weyl-Heisenberg, les familles d'ondelettes s'avèrent souvent mieux adaptées à l'analyse des images naturelles. De plus, l'analyse qu'elles permettent de réaliser a l'avantage d'être plus proche de celle opérée par le système psycho-visuel [Mallat, 2000].

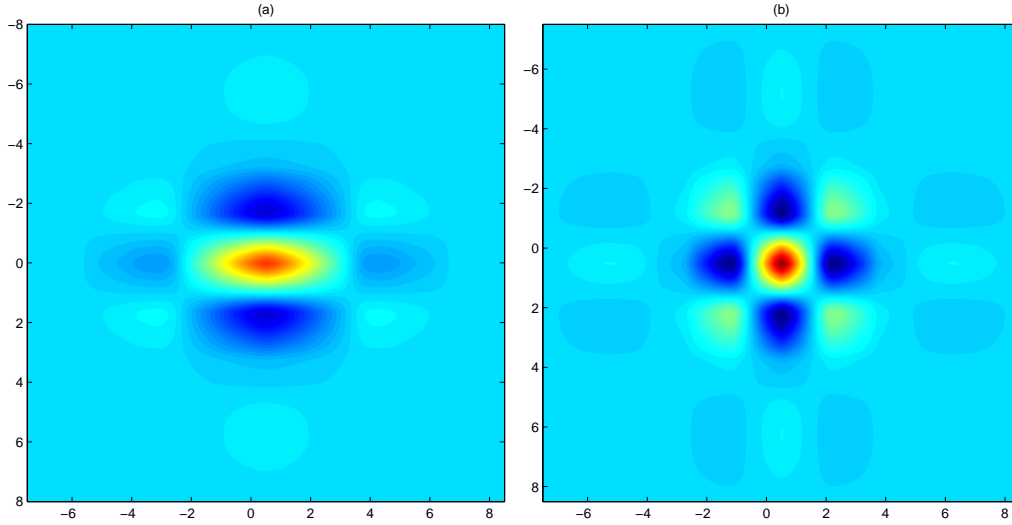


FIG. 9.1 – Exemple d'ondelettes séparables splines : (a) ψ^0 , (b) $\psi^{\pi/4}$. L'ondelette ψ^0 (resp. $\psi^{\pi/2}$) permet d'extraire les informations de type ψ contours ψ horizontaux (resp. verticaux) alors que $\psi^{\pi/4}$ sert à l'analyse des détails résiduels.

- *famille de Gabor modifiée*. En partant de l'idée de combiner les caractéristiques des deux familles précédentes, on peut prendre $\Lambda = \mathbb{R}^4 \times \mathbb{R}_+^* \times [0, 2\pi[$, $\lambda = (x_0, y_0, \mu_x, \mu_y, a, \Theta)$ et

$$\psi_\lambda(x, y) = \frac{1}{a} g\left(\frac{x - x_0}{a}, \frac{y - y_0}{a}\right) \cos\left[\frac{2\pi}{a}(\mu_x(x - x_0) + \mu_y(y - y_0)) + \Theta\right]. \quad (9.3)$$

Notons que, si g est réelle, la transformée d'une image réelle est aussi réelle ($c_\lambda \in \mathbb{R}$). L'analyse obtenue (cf. figure 9.2) est plus flexible que celles de Gabor ou en ondelettes, mais elle comporte un nombre accru de paramètres à gérer.

- *ridgelets*. Les fonctions de cette famille sont fortement apparentées aux ondelettes. Elles sont définies [Candès, 1999] par $\Lambda = \mathbb{R} \times \mathbb{R}_+^* \times [0, 2\pi[$, $\lambda = (b, a, \theta)$ et :

$$\psi_\lambda(x, y) = \frac{1}{\sqrt{a}} \psi\left[\frac{x \cos \theta + y \sin \theta - b}{a}\right]$$

où ψ est une ondelette mère 1D. Les courbes de niveau de la fonction ψ_λ sont les droites d'équations $x \cos \theta + y \sin \theta = \alpha$, $\alpha \in \mathbb{R}$. Ceci explique que cette famille soit particulièrement bien adaptée à l'analyse des contours, qui constituent des éléments sémantiques essentiels des images. Notons que les fonctions ψ_λ ne peuvent être ni sommables, ni de carrés sommables sur \mathbb{R}^2 .

Maintenant que nous avons quelques idées sur la façon de construire une transformation linéaire, un certain nombre d'interrogations doivent nous venir à l'esprit :

1. Quand une transformation linéaire fournit-elle une *représentation complète* ? En d'autres termes, à quelle condition la seule connaissance des coefficients c_λ suffit-elle à caractériser entièrement l'image f ? Le cas

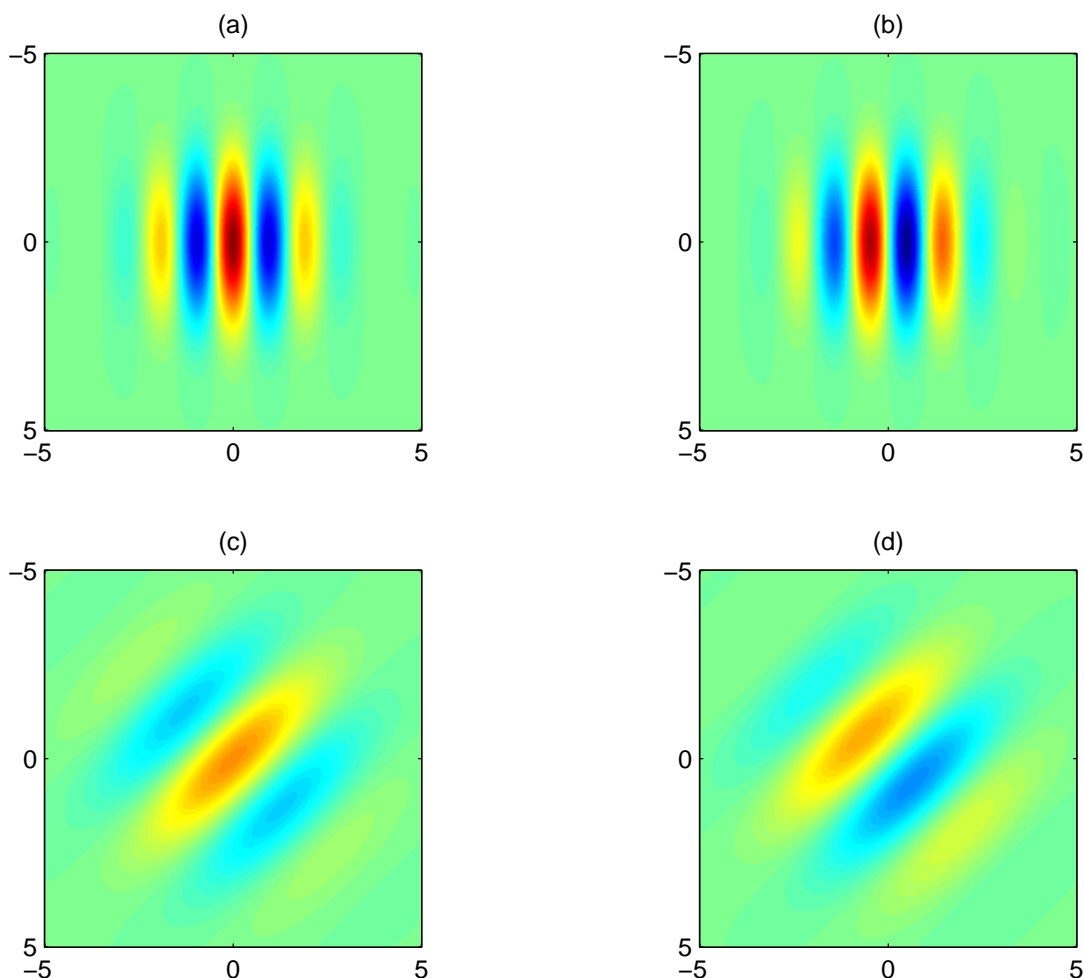


FIG. 9.2 – Exemple de fonctions de Gabor modifiées localisées en $(x_0, y_0) = (0, 0)$: atomes à l'échelle $a = 1$ pour $(\mu_x, \mu_y) = (\frac{1}{2}, 0)$ (orientation suivant l'axe des ordonnées) et (a) $\Theta = 0$ ou (b) $\Theta = \pi/2$; atomes à l'échelle $a = 2$ pour $(\mu_x, \mu_y) = \frac{1}{2\sqrt{2}}(1, 1)$ (orientation suivant la seconde bissectrice) et (c) $\Theta = 0$ ou (d) $\Theta = \pi/2$. La fonction g est ici une gaussienne.

échérant, on doit être à même de reconstruire l'image à partir de ses coefficients. La question revient donc à se demander si l'on peut inverser la transformation T . Un autre intérêt, beaucoup plus pratique, de la définition d'une inverse T^{-1} de T est la possibilité de réaliser des traitements élaborés à partir d'une transformation linéaire, suivant le schéma générique de la figure 9.3. Le traitement peut être, par exemple, un codage ou une estimation afin de compresser ou bien de débruiter l'image. Notons que ces opérations sont généralement plus simples à mettre en œuvre après transformation que directement, sur l'image d'origine.

2. Existe-t-il des méthodes rapides pour le calcul de certaines transformations linéaires ? Peut-on reconstruire ou bien synthétiser de manière efficace une image à partir de ses coefficients c_λ ? Un premier pas dans la voie de l'implémentation est la discrétisation de l'ensemble Λ . Il est clair qu'avec un ordinateur, on ne peut prendre en compte qu'un nombre fini d'atomes espace-fréquence. Il est donc intéressant de se demander si l'on peut se limiter à des familles discrètes de fonctions ψ_λ (*i.e.* des ensembles Λ dénombrables).

Avant d'apporter des éléments de réponses à ces questions, précisons quelques points mathématiques, ne serait-ce

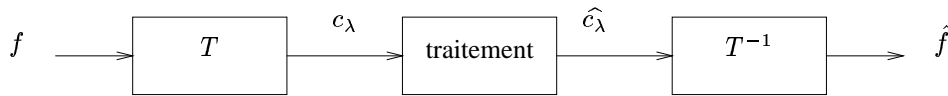


FIG. 9.3 – Traitement dans le domaine transformé (f : image d’origine, c_λ : coefficients transformés, \hat{c}_λ : coefficients traités, \hat{f} : image traitée).

que pour donner un sens rigoureux à l’intégrale (9.1). L’ensemble des images que nous considérerons est $L^2(\mathbb{R}^2)$, l’espace de Hilbert des fonctions f à valeurs complexes, définies sur \mathbb{R}^2 , d’énergie finie, celle-ci étant donnée par :

$$\|f\|^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, y)|^2 dx dy .$$

Cet espace est muni de la norme $\|\cdot\|$ et du produit scalaire associé :

$$\forall (f, f') \in L^2(\mathbb{R}^2), \quad \langle f, f' \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \overline{f'(x, y)} dx dy .$$

9.2 Trames

9.2.1 Objectif

Les trames permettent de définir une représentation complète par décomposition sur une famille discrète $(\psi_i)_{i \in I}$ de fonctions d’un espace de Hilbert \mathcal{H} ($\subset L^2(\mathbb{R}^2)$), I étant un ensemble d’indices dénombrable.

9.2.2 Définition

Une famille $(\psi_i)_{i \in I}$ d’un espace de Hilbert \mathcal{H} est une trame (*frame* en anglais) s’il existe $(A, B) \in (\mathbb{R}_+^*)^2$ tel que :

$$\forall f \in \mathcal{H}, \quad A \|f\|^2 \leq \sum_{i \in I} |\langle f, \psi_i \rangle|^2 \leq B \|f\|^2. \quad (9.4)$$

Les constantes strictement positives A et B sont appelées les *bornes de la trame*. Si $A = B$, on dit que la trame est ajustée (*tight frame*).

Remarques :

- Une base orthonormale $(\psi_i)_{i \in I}$ de \mathcal{H} constitue une trame ajustée puisqu’on a, par conservation de la norme :

$$\sum_{i \in I} |\langle f, \psi_i \rangle|^2 = \|f\|^2 .$$

Inversement, si une trame $(\psi_i)_{i \in I}$ est ajustée avec $A = B = 1$ et si $\forall i \in I, \|\psi_i\| = 1$, alors on montre que c’est une base orthonormale.

- Des éléments d’une trame peuvent être linéairement dépendants. Ceci signifie qu’une trame conduit généralement à une représentation *redondante* de l’information (on calcule \gg plus de coefficients $\langle s, \psi_i \rangle$ que cela n’est strictement nécessaire \llcorner !). Cette redondance peut être pénalisante dans certaines applications telles que la compression d’images, où l’on cherche à compacter au maximum l’information à coder. Au contraire, la redondance est souvent recherchée en analyse d’images, car elle permet d’employer des dictionnaires plus riches de fonctions qui offrent, par exemple, la possibilité de mieux localiser les objets présents dans la scène considérée.

L'un des intérêts de la définition ci-dessus est de conduire à un opérateur de transformation :

$$T : \mathcal{H} \longrightarrow \ell^2(I) = \{(c_i)_{i \in I} \mid \|c\|_{\ell^2(I)} = \sum_{i \in I} |c_i|^2 < \infty\}$$

$$f \longmapsto Tf = (\langle f, \psi_i \rangle)_{i \in I}$$

qui est linéaire et continu. En effet, rappelons que, pour un opérateur linéaire, la continuité est équivalente à la bornitude [Schwartz, 1995] et on vient de voir que :

$$\|Tf\|_{\ell^2(I)}^2 = \sum_{i \in I} |\langle f, \psi_i \rangle|^2 \leq B \|f\|^2.$$

9.2.3 Propriétés

Commençons par ouvrir une petite parenthèse mathématique, en calculant l'adjoint de T . Cet adjoint, noté T^* , est défini par :

$$\forall c = (c_i)_{i \in I} \in \ell^2(I), \forall f \in \mathcal{H},$$

$$\langle T^*c, f \rangle = \langle c, Tf \rangle_{\ell^2(I)} = \sum_{i \in I} c_i \overline{\langle f, \psi_i \rangle} = \left\langle \sum_{i \in I} c_i \psi_i, f \right\rangle$$

et donc :

$$T^*c = \sum_{i \in I} c_i \psi_i. \quad (9.5)$$

Remarquons que, $\forall f \in \mathcal{H}$, on a :

$$\sum_{i \in I} |\langle f, \psi_i \rangle|^2 = \|Tf\|_{\ell^2(I)}^2 = \langle Tf, Tf \rangle_{\ell^2(I)} = \langle T^*Tf, f \rangle.$$

L'équation (9.4) se réécrit donc, sous une forme plus symbolique :

$$A \langle f, f \rangle \leq \langle T^*Tf, f \rangle \leq B \langle f, f \rangle$$

ce qu'on exprime aussi, de manière encore plus synthétique, par :

$$A \text{Id} \leq T^*T \leq B \text{Id}. \quad (9.6)$$

L'opérateur T^*T étant linéaire et continu (puisque T et donc T^* le sont) et minoré par $A \text{Id}$, on peut vérifier qu'il est bijectif et d'inverse continu [Daubechies, 1992a]. De plus, en utilisant des propriétés des opérateurs auto-adjoints, on montre que :

$$B^{-1} \text{Id} \leq (T^*T)^{-1} \leq A^{-1} \text{Id}. \quad (9.7)$$

Utilisons ces résultats pour inverser T . Remarquons tout d'abord que T est bien inversible à gauche. En effet, posons $T^\sharp = (T^*T)^{-1}T^*$. Il est clair que $T^\sharp T = (T^*T)^{-1}T^*T = \text{Id}$. En fait, il s'avère que T n'a pas nécessairement un inverse à gauche *unique* et T^\sharp est le pseudo-inverse de T , *i.e.* l'opérateur de $\ell^2(I)$ vers \mathcal{H} , inverse à gauche de T , qui est de norme minimale [Daubechies, 1992a] (voir aussi le Chapitre 5).

9.2.4 Trame duale

Nous allons maintenant voir que le calcul de la pseudo-inverse de T conduit à la reconstruction de $f \in \mathcal{H}$ à l'aide de ses coefficients $(\langle f, \psi_i \rangle)_{i \in I}$ et des fonctions :

$$\tilde{\psi}_i = (T^*T)^{-1} \psi_i, \quad i \in I.$$

Plus précisément, en utilisant la relation (9.5), on trouve la formule de reconstruction suivante :

$$f = (T^* T)^{-1} T^* T f = (T^* T)^{-1} \sum_{i \in I} \langle f, \psi_i \rangle \psi_i = \sum_{i \in I} \langle f, \psi_i \rangle \tilde{\psi}_i. \quad (9.8)$$

De façon symétrique, on peut vérifier que $f = \sum_{i \in I} \langle f, \tilde{\psi}_i \rangle \psi_i$.

Notons que $(\tilde{\psi}_i)_{i \in I}$ est aussi une trame puisque :

$$\forall f \in \mathcal{H}, \quad B^{-1} \|f\|^2 \leq \sum_{i \in I} |\langle f, \tilde{\psi}_i \rangle|^2 \leq A^{-1} \|f\|^2. \quad (9.9)$$

On dit que $(\tilde{\psi}_i)_{i \in I}$ est la trame *duale* de $(\psi_i)_{i \in I}$. Pour justifier la relation (9.9), on remarque que, pour tout $f \in \mathcal{H}$:

$$(T^* T)^{-1} f = \sum_{i \in I} \langle f, \tilde{\psi}_i \rangle (T^* T)^{-1} \psi_i = \sum_{i \in I} \langle f, \tilde{\psi}_i \rangle \tilde{\psi}_i.$$

Par conséquent :

$$\langle (T^* T)^{-1} f, f \rangle = \sum_{i \in I} |\langle f, \tilde{\psi}_i \rangle|^2.$$

En utilisant l'inégalité (9.7), on en déduit l'inégalité désirée qui est caractéristique d'une trame.

En résumé, de même que l'on dispose de deux formules de reconstruction, on a deux transformations duales :

$$\begin{aligned} T : f &\longmapsto (\langle f, \psi_i \rangle)_{i \in I}, \\ \tilde{T} : f &\longmapsto (\langle f, \tilde{\psi}_i \rangle)_{i \in I}. \end{aligned}$$

A ce point, on peut naturellement se demander dans quelle situation spécifique une trame fournit une analyse non redondante. En premier lieu, il convient d'introduire le concept de \mathbb{J} base de Riesz \mathbb{L} , qui formalise cette notion de \mathbb{J} non-redondance \mathbb{L} .

Définition 1. Une base de Riesz est une trame dont les éléments sont indépendants.

Notons que, si $(\psi_i)_{i \in I}$ est une base de Riesz, alors $(\tilde{\psi}_i)_{i \in I}$ est aussi une base de Riesz, puisque $\forall i \in I, \tilde{\psi}_i = (T^* T)^{-1} \psi_i$. Par ailleurs, compte tenu de la relation duale de (9.8) :

$$\forall i_0 \in I, \quad \psi_{i_0} = \sum_{i \in I} \langle \psi_{i_0}, \tilde{\psi}_i \rangle \psi_i$$

et, la famille $(\psi_i)_{i \in I}$ étant indépendante, on a :

$$\forall i \in I, \quad \langle \psi_{i_0}, \tilde{\psi}_i \rangle = \delta_{i-i_0}.$$

En raison de cette relation, on dit que $(\psi_i)_{i \in I}$ et $(\tilde{\psi}_i)_{i \in I}$ sont des bases bi-orthogonales de \mathcal{H} . Par ailleurs, dans ce cas, si les fonctions $\psi_i, i \in I$, sont normées, on a :

$$\frac{1}{B} \|\psi_{i_0}\|^2 \leq \sum_{i \in I} |\langle \psi_{i_0}, \tilde{\psi}_i \rangle|^2 \leq \frac{1}{A} \|\psi_{i_0}\|^2 \iff \frac{1}{B} \leq 1 \leq \frac{1}{A} \iff A \leq 1 \leq B.$$

Mentionnons que les bases bi-orthogonales sont très utilisées en compression d'images [Antonini et al., 1994] et notamment dans la nouvelle norme JPEG2000 (voir la figure 9.4).



FIG. 9.4 – Compression d'image à l'aide de JPEG2000 : en haut, l'image originale (8×3 bits/pixel) et en bas l'image décodée à 0.25 bit/pixel (c'est-à-dire, une compression d'un facteur 96 pour une image couleur). Le rapport signal sur bruit de codage vaut 28.17 dB en valeur crête (PSNR). JPEG2000 repose sur l'emploi de bases bi-orthogonales d'ondelettes. Les coefficients d'ondelettes sont quantifiés d'autant plus finement que leur amplitude est grande. Une stratégie de codage assez sophistiquée est également mise en œuvre afin d'atteindre de forts taux de compression.

9.2.5 Algorithme des trames

Pour reconstruire une image f à partir des coefficients $(\langle f, \psi_i \rangle)_{i \in I}$, il faut connaître la famille $(\tilde{\psi}_i)_{i \in I}$. Quand les fonctions $\tilde{\psi}_i = (T^* T)^{-1} \psi_i$ ne sont pas calculables analytiquement, il faut les déterminer numériquement. De manière équivalente, on peut directement calculer :

$$f = (T^* T)^{-1} T^* T f = (T^* T)^{-1} r,$$

où $r = T^* T f = \sum_{i \in I} \langle f, \psi_i \rangle \psi_i$.

Dans les deux cas, on est amené à approcher numériquement l'opérateur $L = (T^* T)^{-1}$. Pour cela, on peut poser :

$$T^* T = \frac{A+B}{2} (\text{Id} - R),$$

où $\frac{A+B}{2} \text{Id}$ est une approximation grossière de $T^* T$ à partir des bornes de la trame (cf. relation (9.6)) et $R = \text{Id} - \frac{2}{A+B} T^* T$ correspond à l'erreur d'approximation. D'après l'équation (9.6), on a :

$$-\frac{B-A}{A+B} \text{Id} \leq R \leq \frac{B-A}{A+B} \text{Id}.$$

On peut montrer que cela implique que $\|R\| \leq \frac{B-A}{A+B}$. Sous la condition $\frac{B-A}{A+B} < 1$, qui est, en fait, toujours vérifiée, on a le développement suivant :

$$(\text{Id} - R)^{-1} = \sum_{k=0}^{\infty} R^k,$$

où la série converge en norme. On en déduit que :

$$L = \frac{2}{A+B} \sum_{k=0}^{\infty} R^k.$$

Une approximation d'ordre $N \in \mathbb{N}$ de L est ainsi :

$$L_N = \frac{2}{A+B} \sum_{k=0}^N R^k.$$

L'erreur d'approximation peut être évaluée :

$$L - L_N = \frac{2}{A+B} \sum_{k=N+1}^{\infty} R^k = \frac{2}{A+B} R^{N+1} \sum_{k=0}^{\infty} R^k = R^{N+1} L,$$

soit

$$\|L - L_N\| = \|R\|^{N+1} \|L\| \leq \frac{1}{A} \left(\frac{B-A}{A+B} \right)^{N+1}.$$

On observe que L_N converge d'autant plus rapidement vers L (quand $N \rightarrow \infty$) que $B - A$ est petit, c'est-à-dire que les bornes de la trame sont proches. Par ailleurs, une formule récursive permet de calculer rapidement L_N :

$$L_{N+1} = R L_N + \frac{2}{A+B} \text{Id.}$$

9.2.6 Exemples de trames

Les dictionnaires d'atomes espace-fréquence que nous avons introduits au paragraphe 9.1 permettent de construire des représentations discrètes des images. En effet, on montre que, sous certaines conditions techniques [Daubechies, 1992a, Bergeaud, 1996] qu'il serait fastidieux d'énoncer, on obtient des trames de $L^2(\mathbb{R}^2)$ en choisissant judicieusement $I \subset \Lambda$.

– *trames de Weyl-Heisenberg*

On réalise un échantillonnage régulier des paramètres spatiaux et fréquentiels de la famille de Weyl-Heisenberg (donnée par l'équation (9.2)). On a donc

$$I = \{i = (n\Delta, m\Delta, k\nu_0, l\nu_0), (n, m, k, l) \in \mathbb{Z}^4\}$$

où $(\Delta, \nu_0) \in (\mathbb{R}_+^*)^2$. Une condition nécessaire pour que $(\psi_i)_{i \in I}$ soit une trame de $L^2(\mathbb{R}^2)$ est alors : $\nu_0 \Delta \leq 1$. On montre, par ailleurs, que la trame duale est la trame de Weyl-Heisenberg correspondant à une certaine fenêtre \tilde{g} .

– *trames d'ondelettes*

On discrétise l'échelle suivant une loi exponentielle, alors que les facteurs de translation sont échantillonnés avec un pas proportionnel à cette même échelle [Meyer, 1990]. Ainsi, dans le cas séparable, on a :

$$I = \left\{ i = (na_0^j \Delta, ma_0^j \Delta, a_0^j, \theta), (n, m, j) \in \mathbb{Z}^3, \theta \in \left\{ 0, \frac{\pi}{2}, \frac{\pi}{4} \right\} \right\}$$

où $\Delta \in \mathbb{R}_+^*$ et l'on doit prendre $a_0 > 1$. L'entier j est couramment appelé *niveau de résolution*. Lorsque $j \gg 1$ (resp., $j \ll -1$) la résolution est *grossière* (resp., *fine*), les ondelettes étant dilatées (resp., contractées).

On montre que la trame duale est aussi une trame d'ondelettes.

Très souvent, on se limite à des décompositions dyadiques pour lesquelles $a_0 = 2$. Fait important, il est possible de construire des bases bi-orthogonales [Cohen, 1992] ou même orthonormales d'ondelettes, dans ce cas [Pesquet-Popescu et Pesquet, 2001]. Il peut alors être intéressant de limiter l'analyse en ondelettes à un certain niveau de résolution j_0 , en déterminant les *coefficients d'ondelettes*

$$c_j^\theta[n, m] = \langle f, \psi_{j,n,m,\theta} \rangle, \quad j \geq j_0,$$

ces coefficients, ainsi que les ondelettes associées, ayant été indexés de manière évidente par rapport aux notations précédentes. Ensuite, en calculant

$$\hat{f}_{j_0} = \sum_{\theta} \sum_{j=j_0}^{\infty} c_j^{\theta} [n, m] \tilde{\psi}_{j,n,m,\theta},$$

on peut extraire les structures présentes dans l'image, d'échelle au moins égale à 2^{j_0} . En faisant varier j_0 , on réalise ce qu'on appelle une *analyse multirésolution* (AMR). Quand j_0 décroît, \hat{f}_{j_0} fournit une approximation plus fine de l'image ce qui, en quelque sorte, permet de zoomer sur les détails de la scène (cf. figure 9.5). Grâce à ce concept d'AMR, on peut concevoir des algorithmes efficaces [Mallat, 2000] pour effectuer le calcul d'une décomposition sur une base d'ondelettes. En termes de complexité, ces algorithmes sont compétitifs par rapport à ceux de transformée de Fourier rapide.

A titre illustratif, un exemple de débruitage à l'aide de trames d'ondelettes est donné en figure 9.6.

– *trames de Gabor modifiées*

Dans [Bergeaud, 1996], il est démontré qu'on peut obtenir une trame de $L^2(\mathbb{R}^2)$ à partir d'une famille de Gabor modifiée en prenant

$$I = \{i = (n\Delta, m\Delta, 2^j, \nu_0 \cos \varphi, \nu_0 \sin \varphi, \Theta), (n, m, j) \in \mathbb{Z}^3, \\ \varphi \in \{-\frac{\pi}{2K}(K-1), \dots, \frac{\pi}{2K}(K-1), \frac{\pi}{2}\}, \Theta \in \{0, \frac{\pi}{2}\}\}, \quad (9.10)$$

où $\Delta \in \mathbb{R}_+^*$, $\nu_0 \in \mathbb{R}_+^*$ et $K \in \mathbb{N}^*$.

– *trames de ridgelets*

On peut de manière similaire [Candès, 1999] discrétiser les paramètres d'une famille de ridgelets en prenant

$$I = \{(n2^j, 2^j, 2\pi\ell2^j), n \in \mathbb{Z}, j \in \mathbb{Z}, j \leq j_M, \ell \in \{0, \dots, 2^{-j} - 1\}\},$$

où $j_M \in \mathbb{Z}_-$. On obtient alors une trame de $L^2([0, 1]^2)$, l'espace des images d'énergie finie définies sur le carré unité. Par dilatations et translations des fonctions de cette trame, on peut se ramener à une trame de $L^2(Q)$, où Q est un carré arbitraire de \mathbb{R}^2 . En partitionnant \mathbb{R}^2 en carrés, on peut ainsi générer une trame de $L^2(\mathbb{R}^2)$. Une analyse de ce type (précédée d'une décomposition à l'aide d'un banc de filtres [Vetterli et Kovačević, 1995] isotropes) permet de construire des bases orthonormales de *curvelets*, qui sont mieux adaptées que les ridgelets à l'analyse des contours non rectilignes [Candès et Donoho, 1999].

9.3 Poursuite adaptative

9.3.1 Motivation

Soit $\mathcal{D} = (\psi_i)_{i \in I}$ une trame de $L^2(\mathbb{R}^2)$. On sait que, pour tout $f \in L^2(\mathbb{R}^2)$, il existe $(c_i)_{i \in I}$ tel que

$$f = \sum_{i \in I} c_i \psi_i. \quad (9.11)$$

Cependant, si $(\psi_i)_{i \in I}$ n'est pas une base de Riesz de $L^2(\mathbb{R}^2)$, alors les coefficients $(c_i)_{i \in I}$ ne sont pas définis de manière unique. Parmi toutes les stratégies de choix possibles pour les coefficients $(c_i)_{i \in I}$, la meilleure est sans doute celle qui consiste à concentrer au maximum l'énergie de l'image sur un nombre minimal de termes de la décomposition (9.11). De cette manière, on peut espérer faire apparaître les termes principaux, *i.e.* les structures essentielles composant l'image f .

L'algorithme des trames permet de déterminer des coefficients $c_i = \langle f, \tilde{\psi}_i \rangle$, mais ces derniers ne vérifient généralement pas la propriété de concentration d'énergie dont nous venons de souligner l'importance. Pour pallier ce problème, nous allons maintenant proposer une autre approche.

9.3.2 Algorithme de poursuite

L'algorithme (appelé *matching pursuit*, en anglais) que nous allons présenter vise à extraire itérativement, par ordre décroissant d'importance, les éléments du dictionnaire qui composent le signal.

Supposons que $\mathcal{D} = (\psi_i)_{i \in I}$ soit une trame d'un sous-espace vectoriel fermé \mathcal{H} de $L^2(\mathbb{R}^2)$ (qui peut être $L^2(\mathbb{R}^2)$ lui-même) et que chaque élément de ce dictionnaire soit normé :

$$\forall i \in I, \quad \|\psi_i\| = 1.$$

Si l'on projette $f \in L^2(\mathbb{R}^2)$ (orthogonalement) sur la droite vectorielle engendrée par ψ_i , on obtient une approximation

$$\hat{f} = \langle f, \psi_i \rangle \psi_i$$

de l'image f . D'après le théorème de Pythagore, l'erreur d'approximation est donnée par :

$$\|f - \hat{f}\|^2 = \|f\|^2 - \|\hat{f}\|^2 = \|f\|^2 - |\langle f, \psi_i \rangle|^2.$$

On voit donc que l'erreur est d'autant plus faible que $|\langle f, \psi_i \rangle|$ est grand. Pour déterminer l'élément de \mathcal{D} dont l'importance est prépondérante, il suffit ainsi de rechercher celui qui maximise $|\langle f, \psi_i \rangle|$.

Supposons que :

$$|\langle f, \psi_{i_1} \rangle| = \sup_{i \in I} |\langle f, \psi_i \rangle|.$$

On peut alors appliquer la même méthode à l'erreur d'approximation (appelée aussi $\hat{\psi}$ résidu $\hat{\psi}$),

$$R_f^1 = f - \hat{f} = f - \langle f, \psi_{i_1} \rangle \psi_{i_1},$$

de façon à isoler le $\hat{\psi}$ second ingrédient essentiel $\hat{\psi}_{i_2}, \psi_{i_2}$, entrant dans la composition du signal. Remarquons au passage que R_f^1 est orthogonal à ψ_{i_1} . En procédant de cette manière, itérativement, on obtient à l'étape n :

$$R_f^n = f - \sum_{k=1}^n \langle R_f^{k-1}, \psi_{i_k} \rangle \psi_{i_k} = R_f^{n-1} - \langle R_f^{n-1}, \psi_{i_n} \rangle \psi_{i_n}, \quad (9.12)$$

où

$$\left| \langle R_f^{n-1}, \psi_{i_n} \rangle \right| = \sup_i \left| \langle R_f^{n-1}, \psi_i \rangle \right| \quad (9.13)$$

et l'on a posé $R_f^0 = f$. Toujours en invoquant le théorème de Pythagore, on a :

$$\|R_f^{n-1}\|^2 = \|R_f^n\|^2 + \left| \langle R_f^{n-1}, \psi_{i_n} \rangle \right|^2$$

ce qui conduit, par récurrence, à la relation :

$$\|f\|^2 = \sum_{k=1}^n \left| \langle R_f^{k-1}, \psi_{i_k} \rangle \right|^2 + \|R_f^n\|^2.$$

Remarques :

- Si \mathcal{H} est de dimension infinie, rien ne garantit que $\sup_i |\langle f, \psi_i \rangle|$ est atteint. Il est donc préférable de substituer à la règle de sélection (9.13) la suivante :

$$i_n \in \left\{ \ell \in I \mid |\langle f, \psi_\ell \rangle| \geq (1 - \varepsilon) \sup_{i \in I} |\langle f, \psi_i \rangle| \right\}$$

où $\varepsilon \in]0, 1[$. Notons que, même dans le cas où \mathcal{H} est de dimension finie, cette nouvelle règle de sélection peut être mise à profit afin de faciliter la recherche de i_n .

- On peut montrer [Mallat et Zhong, 1993] que l'algorithme de poursuite converge vers la projection orthogonale de f sur \mathcal{H} . Dans le cas où $\mathcal{H} = L^2(\mathbb{R}^2)$, la limite obtenue est f et on a la décomposition :

$$f = \sum_{k=1}^{\infty} \langle R_f^{k-1}, \psi_{i_k} \rangle \psi_{i_k}.$$

Si \mathcal{H} est de dimension finie, on peut de plus établir que la convergence est exponentielle, *i.e.*

$$\|R_f^n\| \leq \|f\| \rho^n, \quad \rho \in]0, 1[.$$

9.3.3 Améliorations

Réduction de la complexité

La méthode que nous avons présentée au paragraphe précédent pose des problèmes de complexité qui peuvent devenir rédhibitoires en traitement d'image si l'on ne veille pas à optimiser sa mise en œuvre.

A l'itération $n + 1$, on doit déterminer les corrélations du résidu R_f^n avec les fonctions $(\psi_i)_{i \in I}$. Ce calcul peut être réalisé efficacement, de manière récursive, en remarquant que, d'après (9.12),

$$\forall n \in \mathbb{N}^*, \quad \langle R_f^n, \psi_i \rangle = \langle R_f^{n-1}, \psi_i \rangle - \langle R_f^{n-1}, \psi_{i_n} \rangle \langle \psi_{i_n}, \psi_i \rangle. \quad (9.14)$$

On voit donc qu'il est préférable de pré-calculer les produits scalaires $\langle \psi_{i'}, \psi_i \rangle$, $(i', i) \in I^2$. Par ailleurs, pour minimiser le nombre de mises à jour, on a intérêt à limiter le nombre de fonctions $(\psi_i)_{i \in I}$ telles que $\langle \psi_{i'}, \psi_i \rangle \neq 0$, i' étant fixé arbitrairement dans I . Pour le dictionnaire défini par les équations (9.3) à (9.10), cette propriété est vérifiée si g est de support limité. On peut, par exemple, prendre pour g une fonction spline.

Pour initialiser la récurrence (9.14), il faut calculer $\langle f, \psi_i \rangle$. Si des dictionnaires de type (9.3)-(9.10) sont employés, cette corrélation peut être effectuée à l'aide de bancs de filtres [Vetterli et Kovačević, 1995]. La complexité de cette étape de filtrage peut être réduite en imposant à la fonction g d'être séparable.

Notons enfin que l'équation (9.14) doit être mise en œuvre même si $|\langle R_f^{n-1}, \psi_{i_n} \rangle|$ prend de très faibles valeurs, ce qui induit une charge de calculs inutile. Pour accélérer le traitement, on peut se limiter à un sous-dictionnaire restreint formé des éléments $(\psi_i)_{i \in I}$ ayant une influence non négligeable, *i.e.* tels que :

$$|\langle f, \psi_i \rangle| \geq \varepsilon,$$

où $\varepsilon \in]0, 1[$. Dans la suite des itérations de l'algorithme de poursuite, on élimine alors successivement les fonctions ne contribuant pas significativement à l'analyse du résidu, *i.e.*, telles que :

$$|\langle R_f^{n-1}, \psi_i \rangle| < \varepsilon.$$

Quand toutes les fonctions ont ainsi été épuisées¹ et que le sous-dictionnaire est réduit à l'ensemble vide, disons à l'itération m , on peut constituer un nouveau sous-dictionnaire en recherchant les fonctions de \mathcal{D} telles que $|\langle R_f^{m-1}, \psi_i \rangle| \geq \varepsilon$ et itérer la procédure jusqu'à sa convergence [Bergeaud, 1996].

Orthogonalisation

A l'itération $n > 1$ de l'algorithme de poursuite, on sait que le résidu R_f^n est orthogonal à ψ_{i_n} , mais il n'est pas nécessairement orthogonal au sous-espace vectoriel engendré par la famille $\{\psi_{i_1}, \dots, \psi_{i_{n-1}}\}$, si cette dernière n'est pas orthonormale. Ceci montre que rien n'empêche le même élément du dictionnaire d'être retenu plusieurs fois au cours des itérations de l'algorithme de poursuite. Par conséquent, on n'est pas assuré de la convergence de l'algorithme en un nombre fini d'itérations, même si le dictionnaire \mathcal{D} est fini. Pour pallier ce problème, on peut recourir à une orthogonalisation récursive de la famille $(\psi_{i_k})_k$, en employant une procédure de Gram-Schmidt.

¹On montre que cela se produit nécessairement au bout d'un nombre fini d'itérations.



FIG. 9.5 – Exemple d’analyse multirésolution pour $j_0 = 0, j_0 = 1, j_0 = 2, j_0 = 3$ (de haut en bas). L’AMR peut être utilisée pour la transmission progressive d’information ou la recherche hiérarchique dans des bases de données. Dans chacune de ces applications, on part d’une approximation à faible résolution de l’image, qui est raffinée au cours des étapes du traitement.



FIG. 9.6 – Débruitage d'image à l'aide de trames d'ondelettes. A gauche, on peut voir l'image dégradée, après addition d'un bruit non gaussien (mélange de distributions gaussiennes). L'erreur quadratique moyenne normalisée (EQMN) par rapport à l'original vaut $15.48 \cdot 10^{-3}$. L'image restaurée est présentée à droite (EQMN = $6.724 \cdot 10^{-3}$). Un estimateur non linéaire est employé qui consiste simplement en un seuillage des coefficients d'ondelettes. Cette opération permet d'éliminer les coefficients de petite amplitude, très sensibles au bruit, alors qu'on conserve ceux de forte amplitude, sur lesquels se concentre essentiellement l'énergie de l'image.



FIG. 9.7 – Approximation par poursuite adaptative de l'image de la figure (a), en utilisant un dictionnaire de Gabor modifié pour un nombre d'itérations égal à : (b) 650 , (c) 1300 et (d) 1950.

