

Analyse des images

E. Angelini, I. Bloch, Y. Gousseau, S. Ladjal, H. Maître, D. Matignon,

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

Polycopié du cours SI241 - Analyse des Images
Département TSI - Ecole Nationale Supérieure des Télécommunications

23 septembre 2008

Table des matières

| | | |
|----------|--|-----------|
| 1 | Vision et perception | 11 |
| 1.1 | Anatomie du système visuel humain | 11 |
| 1.1.1 | L'œil | 11 |
| 1.1.2 | Les voies optiques | 12 |
| 1.1.3 | Les aires supérieures de la vision | 12 |
| 1.2 | Comment percevons nous les images ? | 12 |
| 1.2.1 | Comportement passif du système visuel | 13 |
| 1.2.2 | Comportement actif du système visuel | 15 |
| 2 | Propriétés statistiques des images | 23 |
| 2.1 | Introduction | 23 |
| 2.1.1 | Pourquoi étudier des propriétés statistiques | 23 |
| 2.1.2 | Quelles images | 24 |
| 2.2 | L'amplitude | 24 |
| 2.2.1 | Propriétés | 24 |
| 2.2.2 | Sensibilité aux modifications d'histogrammes | 25 |
| 2.3 | Les sauts de l'amplitude | 26 |
| 2.4 | La fonction d'autocorrélation | 27 |
| 2.4.1 | À une dimension | 27 |
| 2.4.2 | En multi-dimensions | 30 |
| 2.4.3 | Le spectre de densité de puissance | 31 |
| 2.5 | Entropie | 32 |
| 2.5.1 | Entropie d'ordre 0 | 32 |
| 2.5.2 | Entropie des sauts | 33 |
| 2.5.3 | Entropie d'ordre N | 33 |
| 2.5.4 | Entropie conditionnelle, redondance | 34 |
| 2.5.5 | Retour sur un paradoxe | 34 |
| 2.6 | Un modèle d'image | 35 |
| 2.6.1 | Processus de Markov Gauss | 35 |
| 2.6.2 | Mais | 35 |

| | | |
|----------|---|-----------|
| 2.7 | Critique de la démarche précédente | 36 |
| 2.7.1 | La stationnarité | 36 |
| 2.7.2 | La causalité | 36 |
| 2.7.3 | La monodimensionnalité | 36 |
| 2.7.4 | Vers un modèle complet d'image ? | 37 |
| 3 | L'échantillonnage des images, la représentation fractale | 39 |
| 3.1 | Les signaux monodimensionnels à bande limitée | 39 |
| 3.1.1 | Échantillonnage du signal | 40 |
| 3.1.2 | Reconstruction du signal échantillonné à la fréquence de Nyquist | 41 |
| 3.1.3 | Reconstruction dans le cas de suréchantillonnage | 42 |
| 3.2 | Signaux réels | 42 |
| 3.2.1 | Les limitations physiques du spectre | 42 |
| 3.2.2 | Filtrage du signal | 43 |
| 3.3 | L'extension en dimension N | 44 |
| 3.4 | Le modèle fractal | 45 |
| 3.4.1 | La dimension fractale | 46 |
| 3.4.2 | L'homothétie interne | 48 |
| 3.4.3 | Que faire de la théorie fractale ? | 48 |
| 4 | Représentations discrètes | 53 |
| 4.1 | Introduction | 53 |
| 4.2 | Pavages et maillages | 54 |
| 4.2.1 | Définitions et contraintes | 54 |
| 4.2.2 | Pavages plans réguliers | 55 |
| 4.2.3 | Pavages plans semi-réguliers | 55 |
| 4.2.4 | Dualité pavage / maillage | 56 |
| 4.3 | Topologie discrète | 56 |
| 4.3.1 | Quelques approches | 56 |
| 4.3.2 | Topologie à partir de la notion de voisinage élémentaire | 58 |
| 4.3.3 | Nombre d'Euler : un exemple de caractéristique topologique d'un objet | 63 |
| 4.4 | Représentations géométriques | 65 |
| 4.4.1 | Discrétisation d'une droite continue | 66 |
| 4.4.2 | Caractérisation d'un segment de droite discret | 67 |
| 4.4.3 | Droites analytiques discrètes | 69 |
| 4.4.4 | Cercles discrets | 70 |
| 4.4.5 | Pavage de Voronoï et triangulation de Delaunay | 72 |
| 4.5 | Exemples de structures aléatoires | 75 |
| 4.6 | Fonction distance | 76 |
| 4.6.1 | Définition de distances discrètes | 76 |
| 4.6.2 | Exemples | 77 |

| | | |
|----------|--|-----------|
| 4.6.3 | Algorithmes de calcul | 77 |
| 5 | Restauration des images | 79 |
| 5.1 | Filtrage inverse | 80 |
| 5.1.1 | Approche analytique | 80 |
| 5.1.2 | Approche algébrique | 80 |
| 5.1.3 | Modifications pour des défauts singuliers | 83 |
| 5.1.4 | Conclusions (provisoires) | 84 |
| 5.2 | Régularisation pour les approches algébriques | 85 |
| 5.2.1 | Décomposition du système linéaire en SVD | 85 |
| 5.2.2 | Mesure du conditionnement du système | 85 |
| 5.2.3 | Régularisation par moindres carrés | 86 |
| 5.3 | Régularisation pour les approches analytiques | 87 |
| 5.4 | Mise en place calculatoire | 88 |
| 5.4.1 | Wiener : estimation des spectres de densité de puissance | 88 |
| 5.4.2 | Prise en compte de la décomposition en SVD | 89 |
| 5.4.3 | Restauration des signaux bidimensionnels | 89 |
| 5.5 | Méthodes de restauration itératives | 90 |
| 5.5.1 | Méthode de Jacobi | 90 |
| 5.5.2 | Méthode de Gauss-Seidel | 91 |
| 5.5.3 | Méthode de plus grande pente | 91 |
| 5.6 | Estimation des défauts | 92 |
| 5.6.1 | Détermination par calibrage | 92 |
| 5.6.2 | Détermination statistique | 93 |
| 5.7 | 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 | Quelques applications de l'érosion et de la dilatation | 106 |
| 6.3.1 | Mesures | 106 |
| 6.3.2 | Erodé ultime | 107 |
| 6.3.3 | Rehaussement de contraste | 107 |
| 6.3.4 | Gradient morphologique | 108 |
| 6.4 | Quelques applications de l'ouverture et de la fermeture | 108 |
| 6.4.1 | Filtres alternés séquentiels | 108 |

| | | |
|----------|--|------------|
| 6.4.2 | Chapeau haut-de-forme | 109 |
| 6.4.3 | Granulométries | 109 |
| 6.5 | Introduction à la ligne de partage des eaux | 110 |
| 7 | Définition et simulation d'un champ de Markov | 113 |
| 7.1 | Définition et simulation d'un champ de Markov | 113 |
| 7.1.1 | Description de l'image | 113 |
| 7.1.2 | Modélisation probabiliste de l'image | 114 |
| 7.1.3 | Champs de Markov - Champs de Gibbs | 115 |
| 7.1.4 | Echantillonnage de MRF | 117 |
| 7.1.5 | Le recuit simulé | 119 |
| 7.1.6 | Quelques MRF fondamentaux | 122 |
| 7.2 | Applications : restauration et segmentation | 123 |
| 7.2.1 | Cas de la restauration | 125 |
| 7.2.2 | Cas de la segmentation | 126 |
| 8 | Equations aux dérivées partielles et traitement d'images | 129 |
| 8.1 | L'équation de la chaleur et ses limitations | 130 |
| 8.1.1 | Quelques notations | 130 |
| 8.1.2 | Pourquoi l'équation de la chaleur | 130 |
| 8.1.3 | Equation de la chaleur et noyau gaussien | 131 |
| 8.1.4 | Application aux images | 131 |
| 8.1.5 | L'équation de la chaleur inverse | 131 |
| 8.1.6 | Limitations, invariance par changement de contraste | 133 |
| 8.2 | Equations de diffusion non-linéaires | 134 |
| 8.2.1 | Notations et structure locale des images | 134 |
| 8.2.2 | L'équation de Malik et Perona | 135 |
| 8.2.3 | Mouvement par courbure moyenne et variantes | 136 |
| 8.2.4 | L'équation de Rudin-Osher-Fatemi | 137 |
| 8.2.5 | Le filtre de choc de Rudin | 138 |
| 8.3 | EDP et analyse multi-échelles | 138 |
| 8.3.1 | Définition et propriétés des analyses multi-échelles | 138 |
| 8.3.2 | Pourquoi les analyse multi-échelles sont régies par des E.D.P. | 139 |
| 8.3.3 | Classification des analyses multi-échelles | 140 |
| 9 | Ondelettes et traitement d'images | 147 |
| 9.1 | Principes de l'analyse linéaire des images | 147 |
| 9.2 | Trames | 151 |
| 9.2.1 | Objectif | 151 |
| 9.2.2 | Définition | 151 |
| 9.2.3 | Propriétés | 152 |

| | | |
|-----------|---|------------|
| 9.2.4 | Trame duale | 152 |
| 9.2.5 | Algorithme des trames | 154 |
| 9.2.6 | Exemples de trames | 155 |
| 9.3 | Poursuite adaptative | 156 |
| 9.3.1 | Motivation | 156 |
| 9.3.2 | Algorithme de poursuite | 157 |
| 9.3.3 | Améliorations | 158 |
| 10 | Les prétraitements | 163 |
| 10.1 | Les traitements photométriques ou colorimétriques | 163 |
| 10.1.1 | Linéarité | 163 |
| 10.1.2 | Homogénéité | 163 |
| 10.1.3 | Binarisation, seuillage | 165 |
| 10.1.4 | Augmentation de contraste | 165 |
| 10.2 | Suppression des bruits | 165 |
| 10.2.1 | Filtrages linéaires | 165 |
| 10.2.2 | Filtrage de rang | 167 |
| 10.2.3 | Filtrages morphologiques | 167 |
| 10.2.4 | Filtrages par équations de diffusion | 167 |
| 10.3 | Les filtres adaptatifs | 170 |
| 10.3.1 | Les filtres à coefficients adaptatifs | 170 |
| 10.3.2 | Les filtres à fenêtres adaptatives | 171 |
| 10.4 | Le rééchantillonnage des images | 172 |
| 10.4.1 | Interpolation avec des polynômes | 172 |
| 10.4.2 | Interpolation par des B-splines | 173 |
| 10.4.3 | Interpolation adaptative | 175 |
| 11 | La détection des contours dans les images | 177 |
| 11.1 | Un modèle continu de contour. | 178 |
| 11.2 | Les approches classiques | 179 |
| 11.2.1 | Les détecteurs de gradient par filtrage | 179 |
| 11.2.2 | Les détecteurs de gradient par masques | 180 |
| 11.2.3 | Pré- et post traitements | 181 |
| 11.2.4 | Les détecteurs de passage par zéro du laplacien | 182 |
| 11.2.5 | Les détecteurs par masquage adapté | 183 |
| 11.3 | Les approches analytiques | 184 |
| 11.3.1 | Les critères de Canny | 184 |
| 11.3.2 | Les filtres de Deriche et Shen et Castan | 185 |
| 11.3.3 | L'extension à 2D | 186 |
| 11.3.4 | Variantes autour des filtres analytiques | 188 |
| 11.4 | Les contours actifs = les <i>snakes</i> | 188 |

| | | |
|-----------|--|------------|
| 11.4.1 | Une approche physique des contours | 188 |
| 11.4.2 | Mise en équation des contours actifs | 189 |
| 11.4.3 | Les ensembles de niveaux (<i>level sets</i>) | 191 |
| 11.4.4 | Les modèles géodésiques actifs | 192 |
| 11.5 | La poursuite et la fermeture des contours | 192 |
| 11.5.1 | Méthodes de recherche dans des graphes | 193 |
| 11.5.2 | Les automates | 194 |
| 12 | La segmentation par régions | 197 |
| 12.1 | Les méthodes sur histogramme | 197 |
| 12.1.1 | Avec apprentissage | 197 |
| 12.1.2 | Seuillage sans apprentissage | 198 |
| 12.1.3 | Méthodes de classification | 199 |
| 12.1.4 | Sélection sur l'histogramme et dans l'image | 199 |
| 12.1.5 | Sélection sur histogramme et régularisation | 199 |
| 12.2 | Les méthodes par transformation de régions | 201 |
| 12.2.1 | La croissance de région | 202 |
| 12.2.2 | Le partage de région | 202 |
| 12.2.3 | La réunion de région | 203 |
| 12.2.4 | Les pyramides | 204 |
| 12.3 | Les graphes d'adjacence | 204 |
| 12.4 | La méthode MDL = Minimum Description Length | 205 |
| 12.5 | L'approche de Mumford et Shah | 207 |
| 12.5.1 | Approche formelle | 207 |
| 12.5.2 | Les variantes autour de la formulation de Mumford-Shah | 208 |
| 13 | Les textures | 211 |
| 13.1 | Qu'est ce qu'une texture ? | 211 |
| 13.1.1 | Distribution aléatoire ou régulière ? | 212 |
| 13.1.2 | Les échelles des textures | 213 |
| 13.1.3 | Analyse ou synthèse | 213 |
| 13.2 | Modèles de texture | 213 |
| 13.2.1 | Un modèle biologiquement plausible | 213 |
| 13.2.2 | Modèles stochastiques | 214 |
| 13.3 | Analyse et reconnaissance de textures | 214 |
| 13.3.1 | Schéma général | 214 |
| 13.3.2 | Approches par mesures statistiques | 215 |
| 13.3.3 | La détection d'énergie dans des canaux fréquentiels | 219 |
| 13.3.4 | Les filtres optimisés | 220 |
| 13.3.5 | Les modélisations autorégressives | 221 |
| 13.4 | Les approches par champs markoviens | 221 |

| | | |
|-----------|---|------------|
| 13.4.1 | La méthode de Manjunath et Chellappa | 222 |
| 13.4.2 | La méthode de Kervrann et Heitz | 223 |
| 13.5 | Les méthodes structurales | 224 |
| 13.6 | Textures hétérogènes | 224 |
| 14 | Description de contours et de formes | 225 |
| 14.1 | Fonction caractéristique | 225 |
| 14.2 | Description de formes | 227 |
| 14.2.1 | Représentation par les moments | 227 |
| 14.2.2 | Représentation par les moments invariants | 227 |
| 14.2.3 | Boîtes englobantes et boîtes minimales | 228 |
| 14.3 | Polygones de Guzman | 228 |
| 14.4 | Chaînes de Freeman | 229 |
| 14.4.1 | Définition | 230 |
| 14.4.2 | Les propriétés des chaînes de Freeman | 231 |
| 14.4.3 | Reconnaissance des formes par des chaînes de Freeman | 233 |
| 14.5 | Descripteurs de Fourier | 233 |
| 14.5.1 | Descripteur par tangente | 233 |
| 14.5.2 | Représentation complexe | 234 |
| 14.6 | Approximations polynomiales | 235 |
| 14.6.1 | Approximation d'un nuage de points par une droite unique | 236 |
| 14.6.2 | Approximations polygonales, simplification de contours polygonaux | 238 |
| 14.6.3 | Approximation par des splines | 240 |
| 14.7 | Transformation de Hough | 241 |
| 14.7.1 | Définitions | 242 |
| 14.7.2 | Mise en œuvre de la TH | 244 |
| 14.7.3 | Détections par TH | 245 |
| 14.8 | Conclusion | 246 |
| 15 | Formation de l'image et couleur | 247 |
| 15.1 | Formation des images | 247 |
| 15.1.1 | Imagerie cohérente et incohérente | 247 |
| 15.1.2 | Interaction matière rayonnement | 249 |
| 15.1.3 | Les sources de lumière | 252 |
| 15.2 | Définition des différents espaces couleur | 252 |
| 15.2.1 | Trivariance et primaires | 252 |
| 15.2.2 | Choix d'un espace de couleur | 253 |
| 15.2.3 | L'espace RVB de la CIE 1931 | 254 |
| 15.2.4 | L'espace XYZ | 256 |
| 15.2.5 | L'espace Lab et le repère TLC | 257 |
| 15.2.6 | L'espace YC1C2 | 259 |

| | | |
|----------|--|------------|
| 15.2.7 | L'espace TLS | 259 |
| 15.2.8 | Le système visuel hypothétique | 260 |
| 15.2.9 | L'espace TLC | 262 |
| 15.2.10 | Les espaces de la télévision | 262 |
| 15.3 | Modélisation de la perception colorée des images | 263 |
| 15.4 | Les traitements des images en couleur | 265 |
| 15.4.1 | Un ordre pour l'espace R^3 ? | 265 |
| 15.4.2 | Qu'est ce qu'une couleur représentative d'un nuage? | 267 |
| 15.4.3 | Comment calculer une palette? | 267 |
| 15.4.4 | Comment filtrer 2 couleurs? | 268 |
| 15.4.5 | La détection des contours des images en couleur | 268 |
| 15.4.6 | Représentation par quaternions | 269 |
| 15.4.7 | Couleur et quaternion | 269 |
| A | Décomposition en Valeurs Singulières : SVD | 271 |
| A.1 | Définition | 271 |
| A.2 | Lien avec le rang et interprétation de la décomposition | 271 |
| A.3 | Normes | 272 |
| B | Applications de la SVD à la résolution des systèmes linéaires | 273 |
| B.1 | Conditionnement d'un système linéaire non singulier | 273 |
| B.2 | Pseudo-inverse | 273 |
| B.3 | Solutions robustes d'un système linéaire, moindres carrés | 274 |
| C | Optimisation sous contraintes | 277 |
| C.1 | Contraintes égalités | 277 |
| C.2 | Contraintes inégalités | 277 |

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.

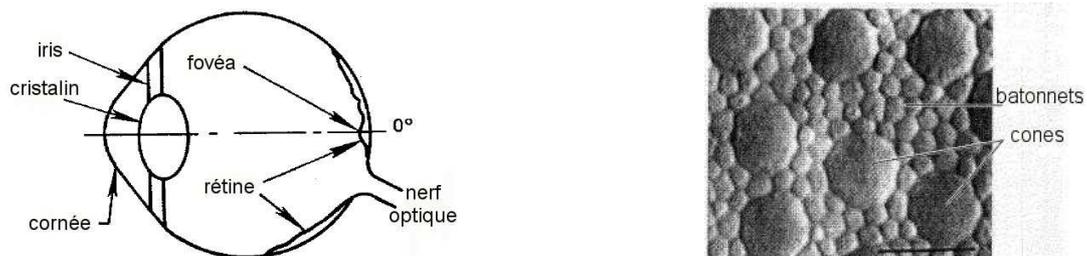


FIG. 1.1 – Schéma de l'œil humain et photographie du fond de la rétine.

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.

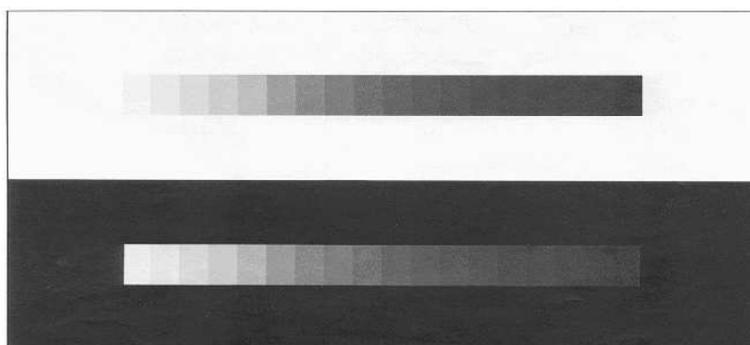


FIG. 1.2 – La perception des différentes gradations d'une échelle de gris dépend du niveau lumineux du fond.

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*

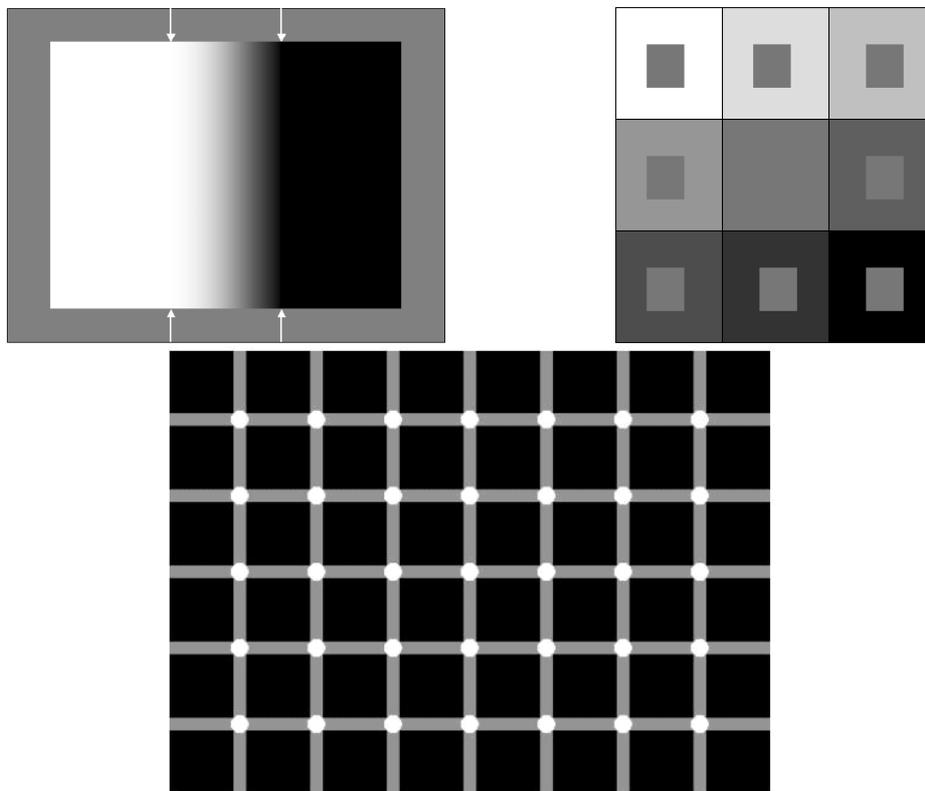


FIG. 1.3 – Phénomènes dus aux interactions spatiales dans l'image.

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

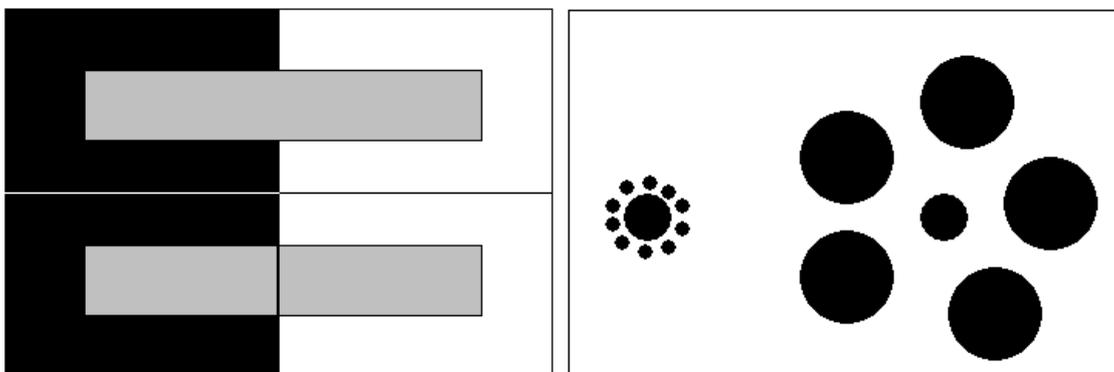


FIG. 1.4 – Phénomènes dus aux interactions spatiales dans l’image. À gauche, la seule adjonction d’un contour au milieu de la barre grise au motif du haut permet d’obtenir un net contraste simultané (motif du bas). À droite, dans l’illusion de Titchener les deux cercles centraux sont de taille égale.

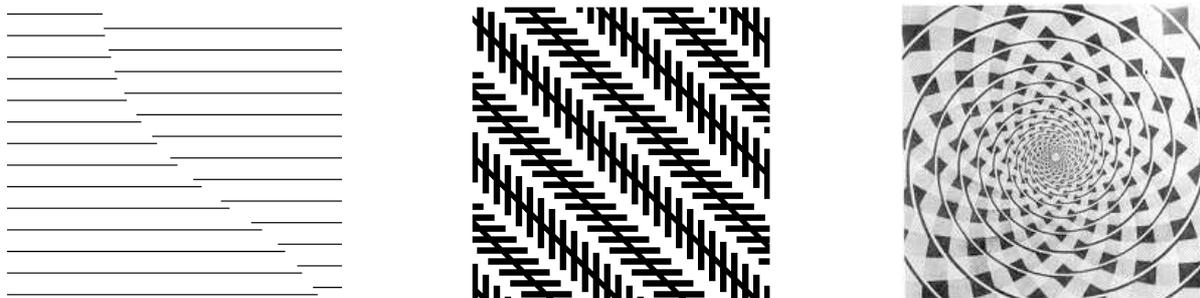


FIG. 1.5 – Phénomènes dus aux interactions spatiales dans l’image. Les contours vrais ou faux (à gauche) jouent par leur orientation un rôle primordial dans certains effets : les diagonales dans l’illusion de Zöllner (au centre) sont parallèles et la spirale de Snearl (à droite) n’est formée... que de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

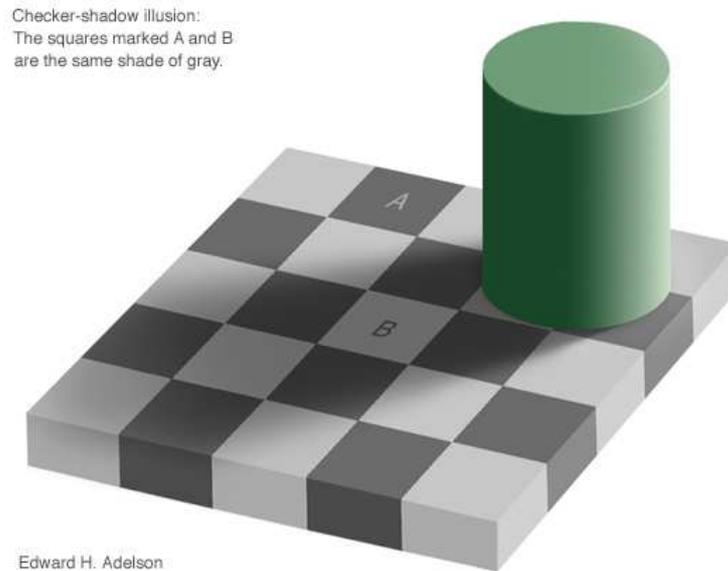


FIG. 1.6 – Dans cet échiquier, les deux carrés marqués A et B ont le même niveau de gris. Plusieurs phénomènes se conjuguent pour renforcer cette illusion. certains relèvent de processus de contraste simultané, d'autres de phénomènes d'interprétation de la scène. © E. H. Adelson.

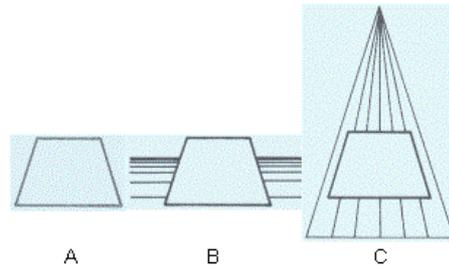


FIG. 1.7 – La perception des formes est modifiée selon le contexte spatial où elle est située.

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 rétiniennes 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

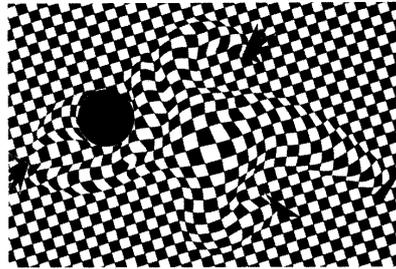


FIG. 1.8 – Rôle des gradients de texture dans l'« Arlequin » de Vasarely (cf. texte).

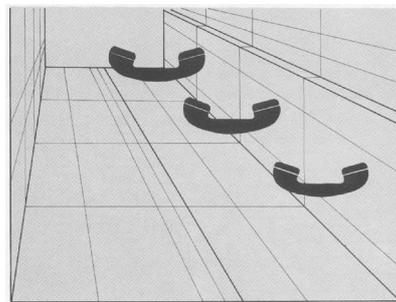


FIG. 1.9 – Illusion géométrique sur la taille provoquée par la perspective.

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.

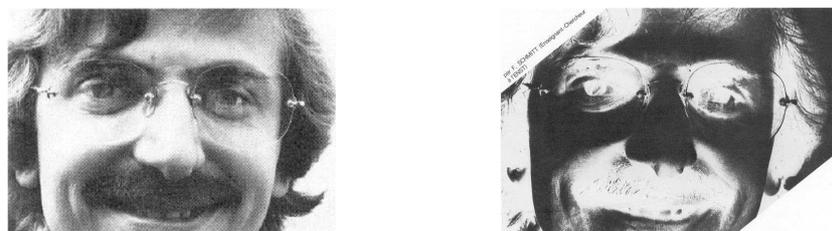


FIG. 1.10 – Positif et négatif de la photo de l'auteur. Vous pouvez vérifier.

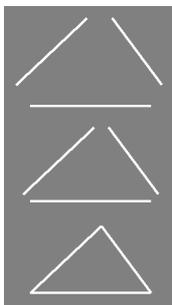


FIG. 1.11 – Le triangle est perçu comme une unité graphique en soi.

Vous constatez que le cerveau ne se contente pas de percevoir que 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 apprenez 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.

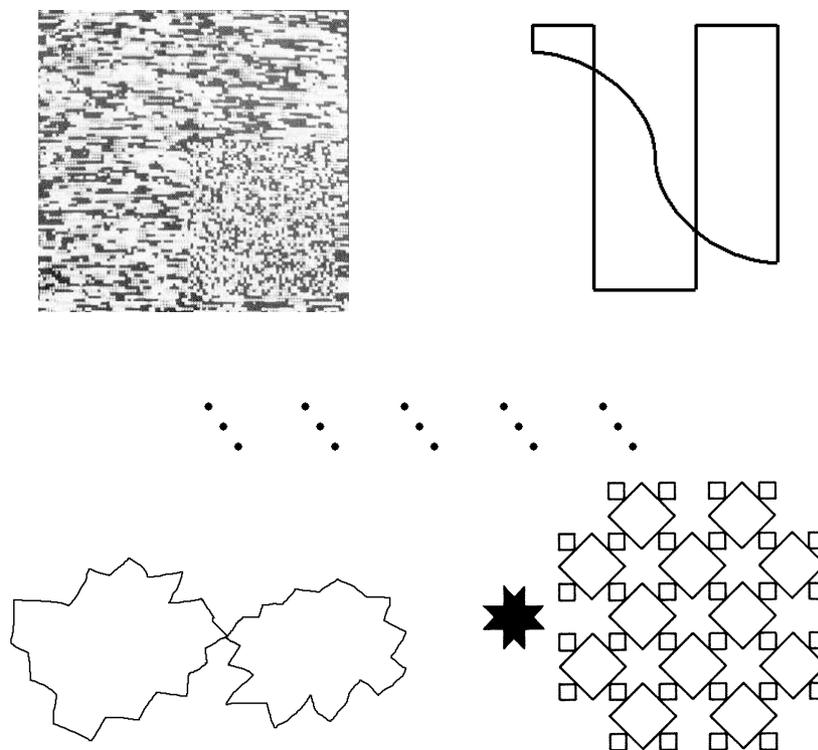


FIG. 1.12 – Lois de la perception (d’après les gestaltistes).

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.

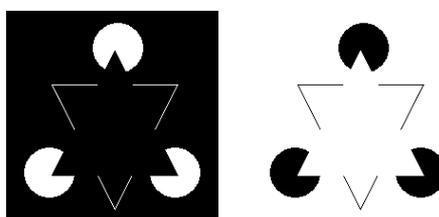


FIG. 1.13 – Deux triangles subjectifs qui n’ont pas de contours (d’après G. Kanizsa, Scientific American, avril 76).



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

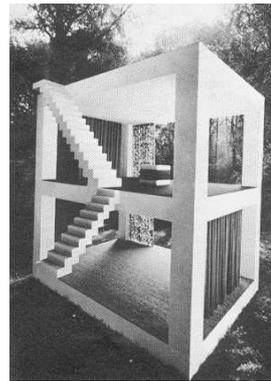
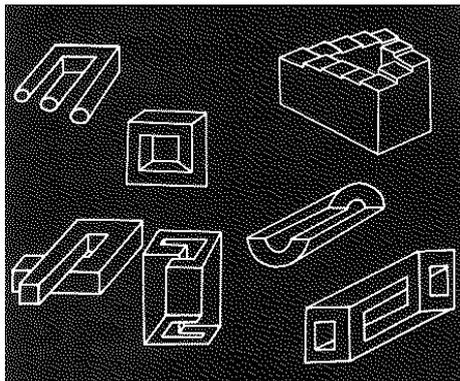


FIG. 1.15 – Quelques constructions impossibles (à gauche), ... qui n'ont pas arrêté cet architecte (photographie A.P. Lamothe), à droite.

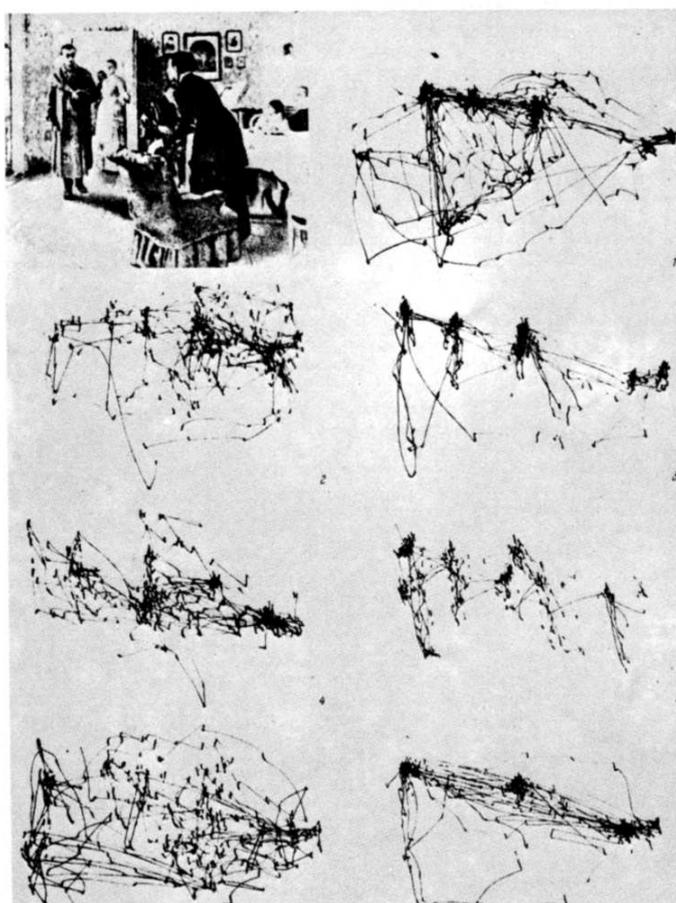
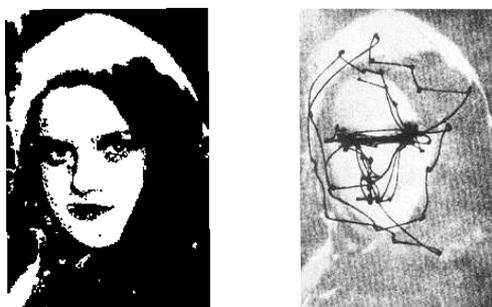


FIG. 1.16 – Chemin que suit l'axe du regard (d'après Yarbus et Kolers [Kolers, 1972]) : en haut, lorsqu'on observe une image ; en bas lorsqu'on demande à l'observateur de se souvenir de diverses choses concernant l'image.

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.

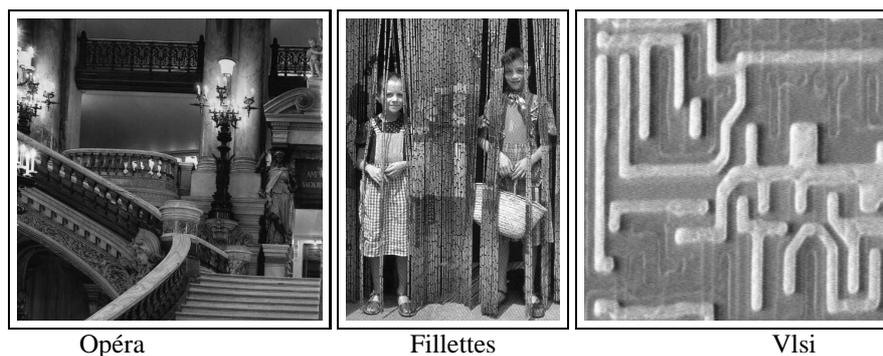


FIG. 2.1 – Trois images « quelconques ».

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 préciserons 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

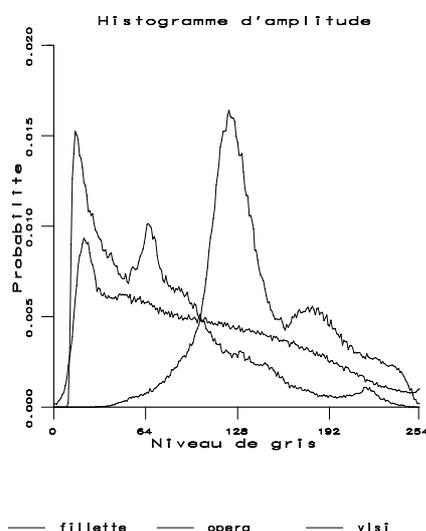


FIG. 2.2 – Les histogrammes des 3 images de la figure 2.1

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-

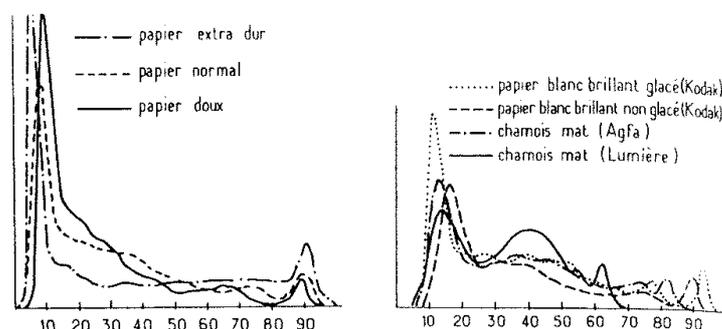


FIG. 2.3 – Histogrammes d’une même image obtenus pour des présentations sur des papiers de natures différentes : à gauche, dureté du papier, à droite, aspect du papier (d’après [Estournet, 1969]).

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 réfléchi 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 :

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

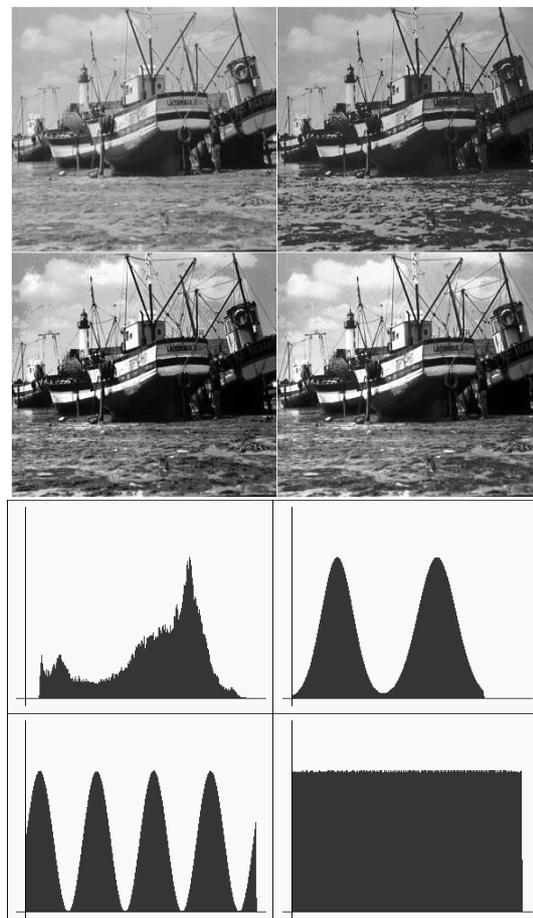


FIG. 2.4 – Modification d'histogramme. En haut à gauche : image originale, en haut à droite somme de 2 gaussiennes, en bas à gauche sinusoïde, en bas à droite répartition uniforme.

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

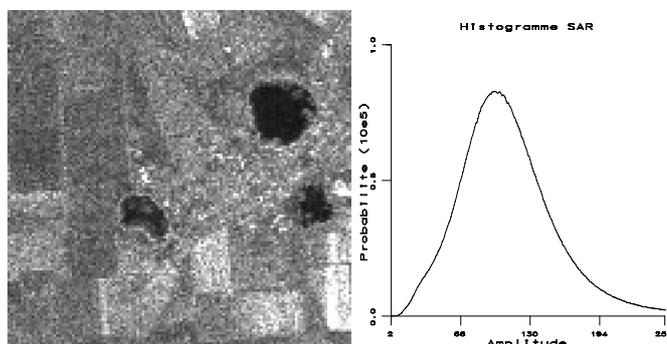


FIG. 2.5 – Agrandissement d’une image de radar à vision latérale (radar ERS-1) sur une région agricole de l’Ukraine (©Cnes) montrant l’importance du bruit de cohérence, et histogramme de la scène complète, très proche d’une loi de Rayleigh.

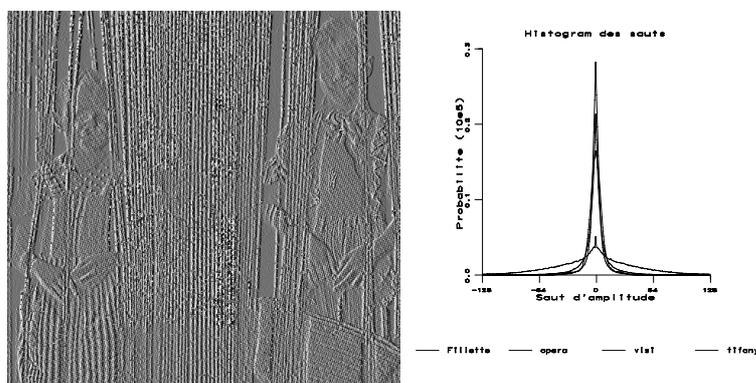


FIG. 2.6 – Une image des gradients horizontaux d’un extrait de l’image « Fillettes » (la valeur 127 a été ajoutée à toute l’image pour l’affichage) et 4 histogrammes de sauts de 4 images différentes.

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.

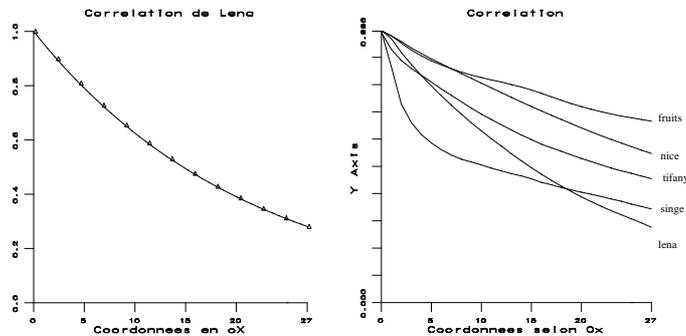


FIG. 2.7 – A gauche sont superposées les valeurs mesurées d'une fonction d'autocorrélation (sur l'image « Lena ») à la loi exponentielle théorique la plus proche, à droite 5 fonctions de corrélation d'images différentes.

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).

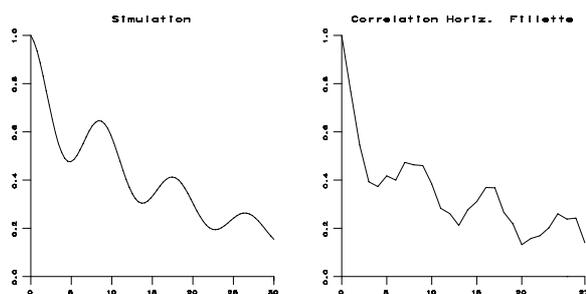


FIG. 2.9 – Allure de l'autocorrélation d'une image avec un fort contenu périodique : théorique à gauche et mesurée (sur *Fillette*), à droite.

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.

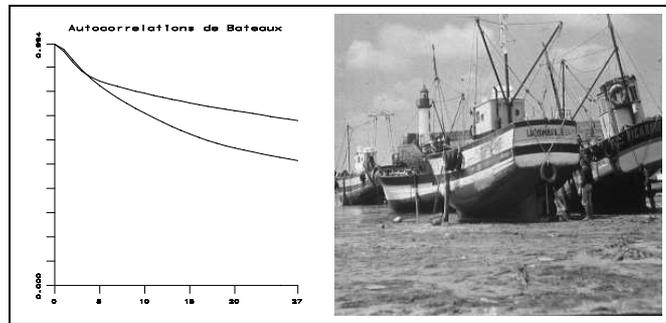


FIG. 2.10 – Corrélation obtenue dans les directions horizontales (en haut) et verticales (en bas) pour l'image « Bateaux ».

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 ,

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

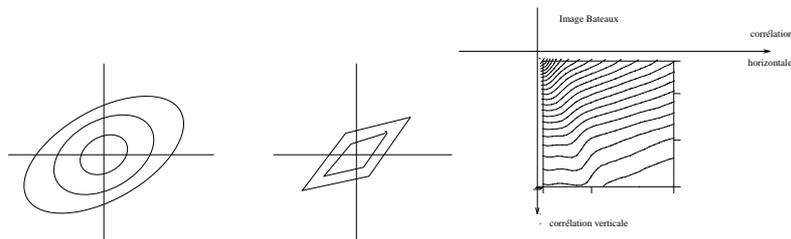


FIG. 2.11 – Courbes d'isocorrélation pour le modèle non-séparable (à gauche) et pour le modèle séparable (au centre) et courbes d'isocorrélation mesurée sur l'image « Bateaux » (le quart inférieur du plan est représenté).

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 :

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

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

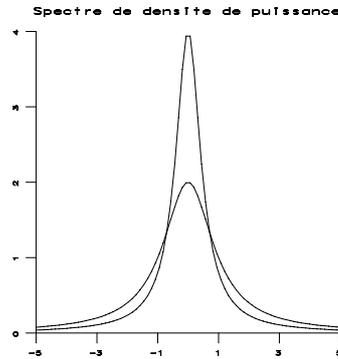


FIG. 2.12 – Spectre de densité de puissance d'images à autocorrélation exponentielle ($\alpha = 0,5$ et $\alpha = 1$)

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'entropie² 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 Bennett 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).

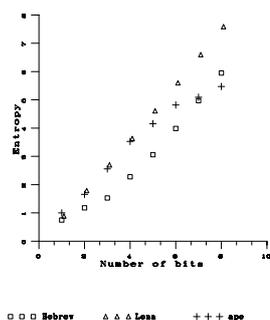


FIG. 2.13 – Décroissance de l'entropie pour 3 images en fonction de la quantification des niveaux de gris.

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'instantanés (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 :

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

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 :

$$P(f_i) = M^i \cdot P(f_0)$$

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 arbitraires 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.

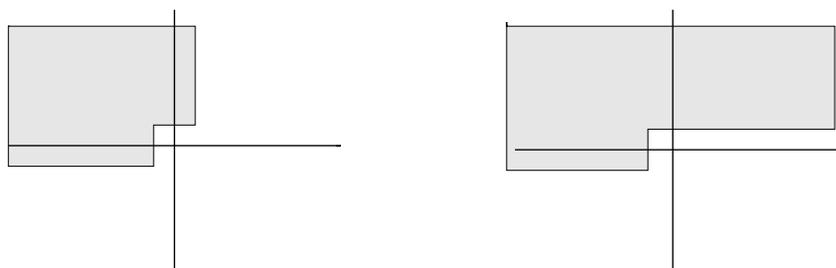


FIG. 2.14 – Prédiction d'un point par « passé quart de plan », à gauche, et « passé demi-plan », à droite. Les zones grisées constituent le passé du point au centre du repère, au sens de la causalité markovienne.

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] :

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

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 $\frac{1}{2U}$.

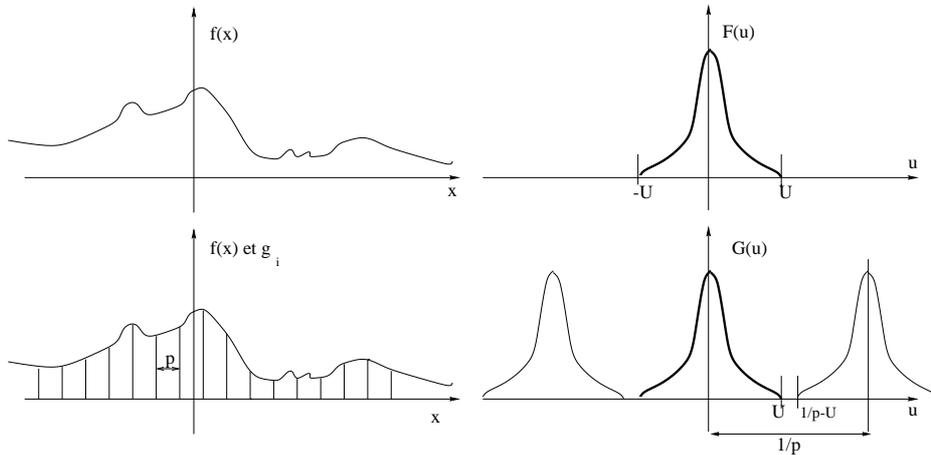


FIG. 3.1 – Echantillonnage d'un signal continu $f(x)$, à spectre $F(u)$ à support borné, avec un pas d'échantillonnage p . Le signal résultant discret a un spectre $|G(u)|^2$.

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 :

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

la fonction $\text{III}(x/p)$ 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 $p < \frac{1}{2U}$.

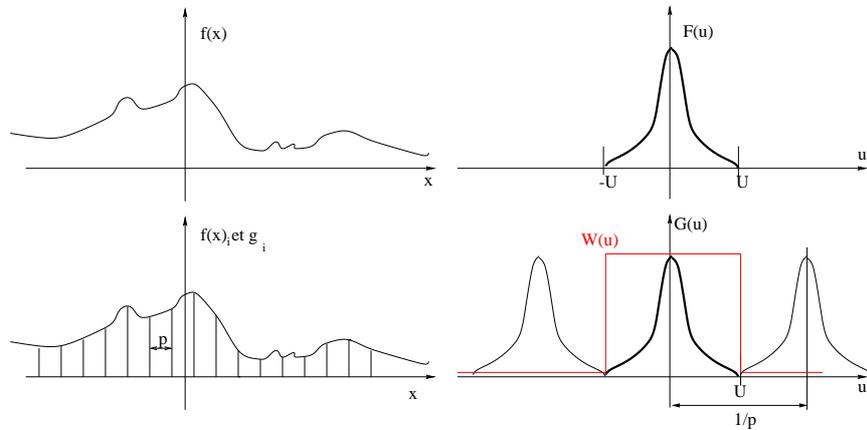


FIG. 3.2 – Échantillonnage d'un signal continu $f(x)$ à la fréquence de Nyquist. Le filtre de reconstruction $W(u)$ est le filtre canonique $W_N(u)$, c'est-à-dire la fonction porte.

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.

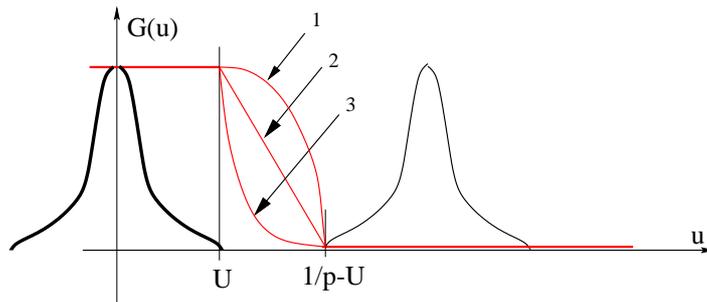


FIG. 3.3 – Trois filtres possibles pour reconstruire exactement un signal suréchantillonné (les filtres sont symétriques pour les u négatifs). Ils diffèrent par leurs valeurs dans l'intervalle compris entre U et $1/p - U$.

3.2 Signaux réels

À quelque échelle que nous l'observons, 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) :

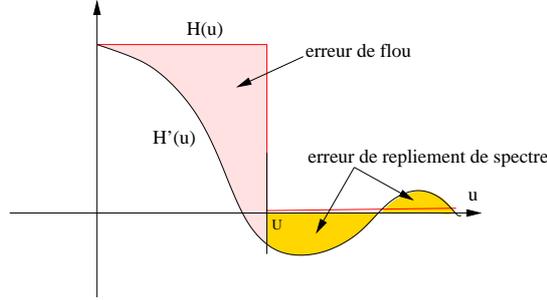


FIG. 3.4 – Les deux types d’erreurs introduites par le filtrage d’un signal par un filtre $h'(x)$ différent du passe-bas idéal $h(x)$ lors de son échantillonnage. Ils sont présentés ici dans l’espace de Fourier. À l’intérieur de la bande, l’erreur est une distorsion des fréquences appelée flou en traitement 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 $F(\mathbf{u})$.

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

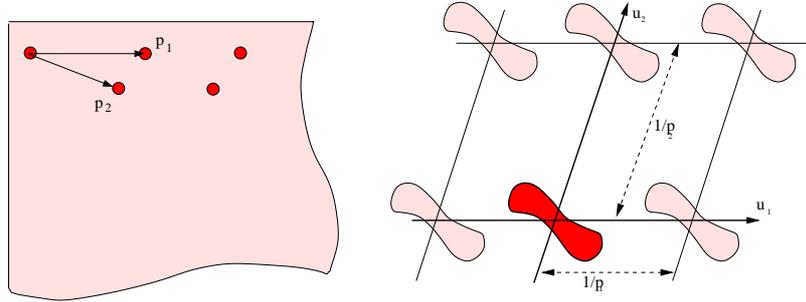


FIG. 3.5 – Echantillonnage d’un signal bidimensionnel selon 2 pas d’échantillonnage p_1 et p_2 (à gauche). Le spectre de l’image est reproduit aux nœuds du réseau réciproque (à droite). Le domaine \mathcal{S} est le domaine où $F(u)$ n’est pas nul ; le domaine \mathcal{V} est la maille de répétition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

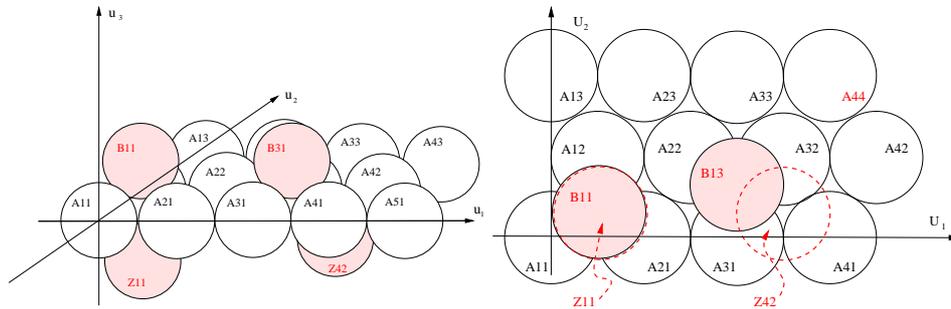


FIG. 3.6 – Echantillonnage de signaux tridimensionnels à spectre isotrope. Les spectres sont représentés par des sphères. la vue de gauche représente une vue perspective du domaine de Fourier, celle de droite une vue de dessus depuis la direction des u_3 positifs. On a représenté un échantillonnage hexagonal dans le plan u_1u_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, étagelement 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 au

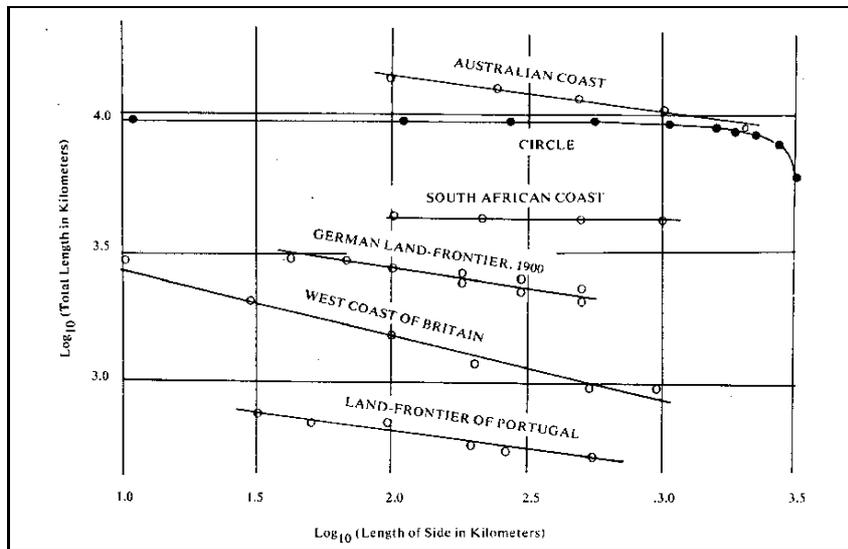


FIG. 3.7 – Évolution de la longueur de quelques frontières et côtes en fonction de la mesure utilisée pour calculer cette longueur (d'après [Richardson, 1961] citée dans [Mandelbrot, 1977]) - échelles logarithmiques.

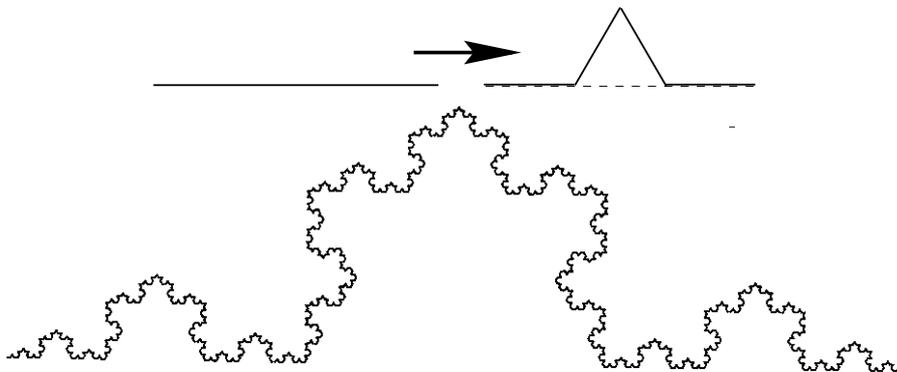


FIG. 3.8 – La courbe de von Koch est un cas d'école de courbe fractale. Elle se construit à partir d'un segment de droite qui est divisé en 3 et remplacé par 4 segments de taille $1/3$. La dimension fractale vaut alors $D = \log 4 / \log 3$.

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 :

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

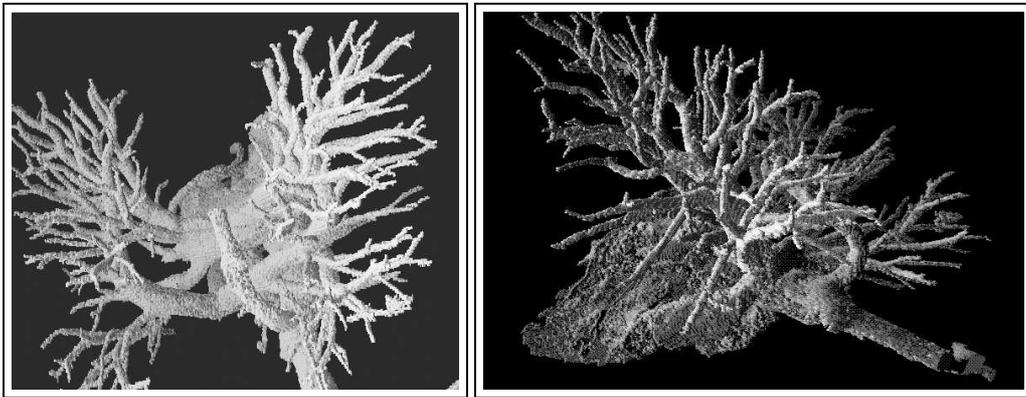


FIG. 3.9 – Quelques exemples de fractales : arbre bronchique de l'homme recalculé à partir du corps du Human Visual Project par Jorge Marquez.

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

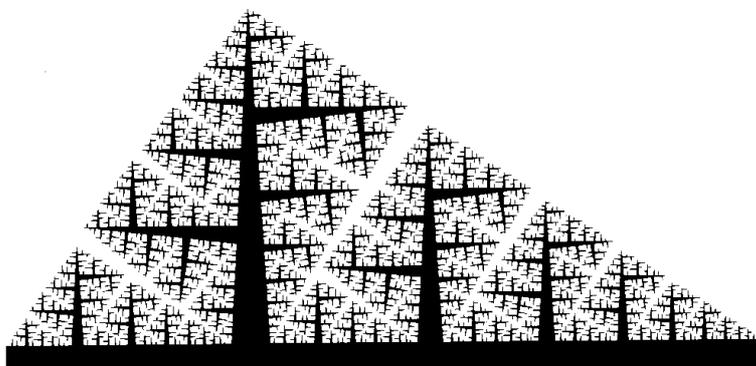


FIG. 3.10 – La forêt étagée de Peano, une forêt fractale artificielle.

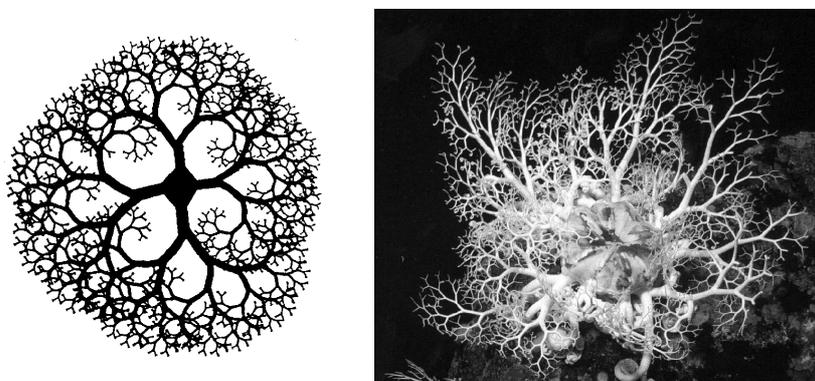


FIG. 3.11 – Corail synthétique et naturel

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).

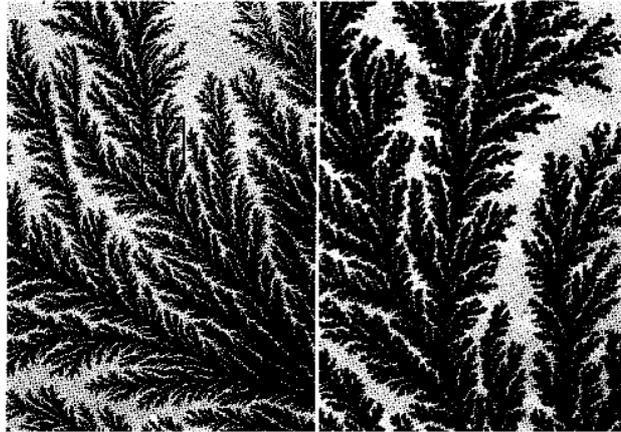


FIG. 3.12 – Electrodeposition de cuivre, c'est un exemple de percolation fractale. La cascade d'homothétie est illustrée par la petite fenêtre de l'image gauche, agrandie sur l'image droite (d'après [Bunde et Havlin, 1994]).

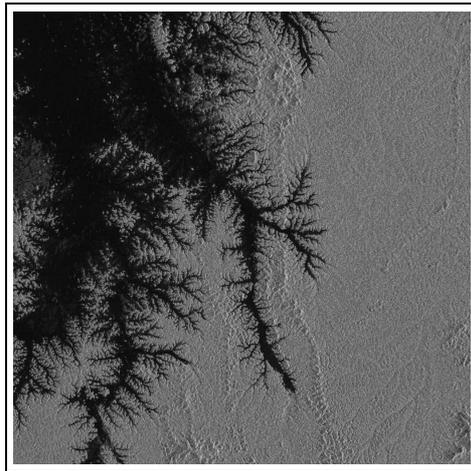


FIG. 3.13 – Un lac en forêt guyanaise vu par le satellite radar ERS-1 ©ESA.

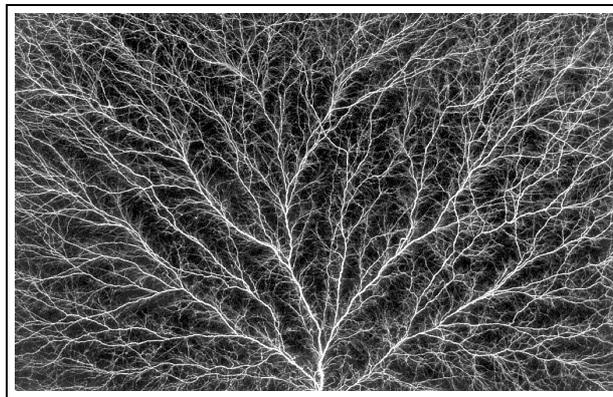


FIG. 3.14 – Décharge électrique claquant un diélectrique (d'après [Bunde et Havlin, 1991]).

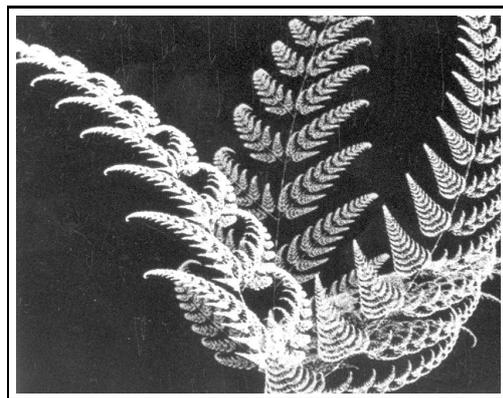


FIG. 3.15 – Fougère simulée par 5 attracteurs superposés (d'après [Peitgen et Saupe, 1988]).

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 appartenait 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, Cœurjolly et al., 2007].

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).

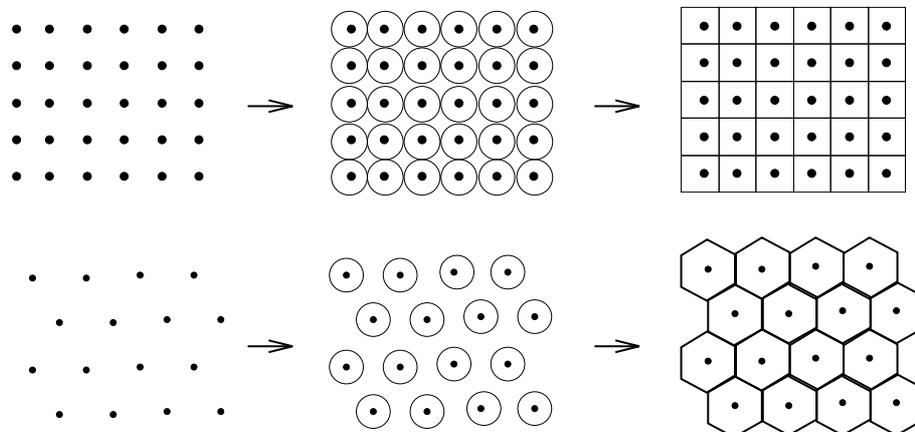


FIG. 4.1 – Deux exemples de construction d'un pavage par propagation à partir d'une distribution de points régulière.

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).

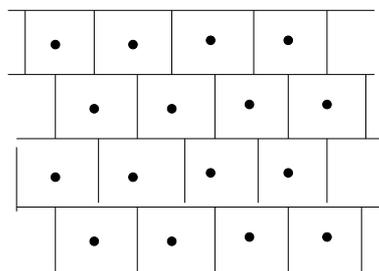


FIG. 4.2 – Exemple d'une configuration où les sommets du pavage ne sont pas en coïncidence avec d'autres sommets. Notons qu'en déformant ce maillage, on peut obtenir un pavage hexagonal qui est le pavage régulier associé à cette distribution de points, comme on le verra par la suite.

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.

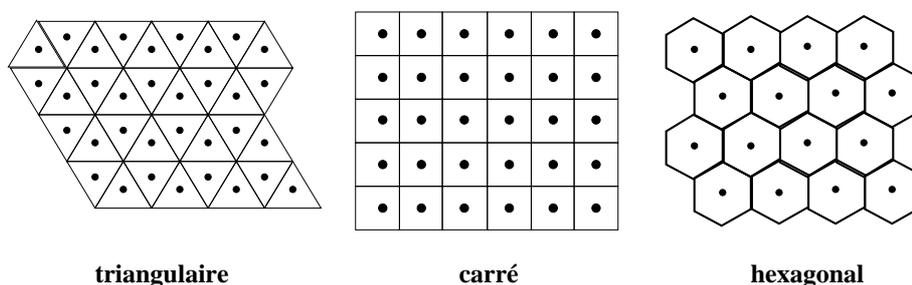


FIG. 4.3 – Les trois pavages réguliers du plan.

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.

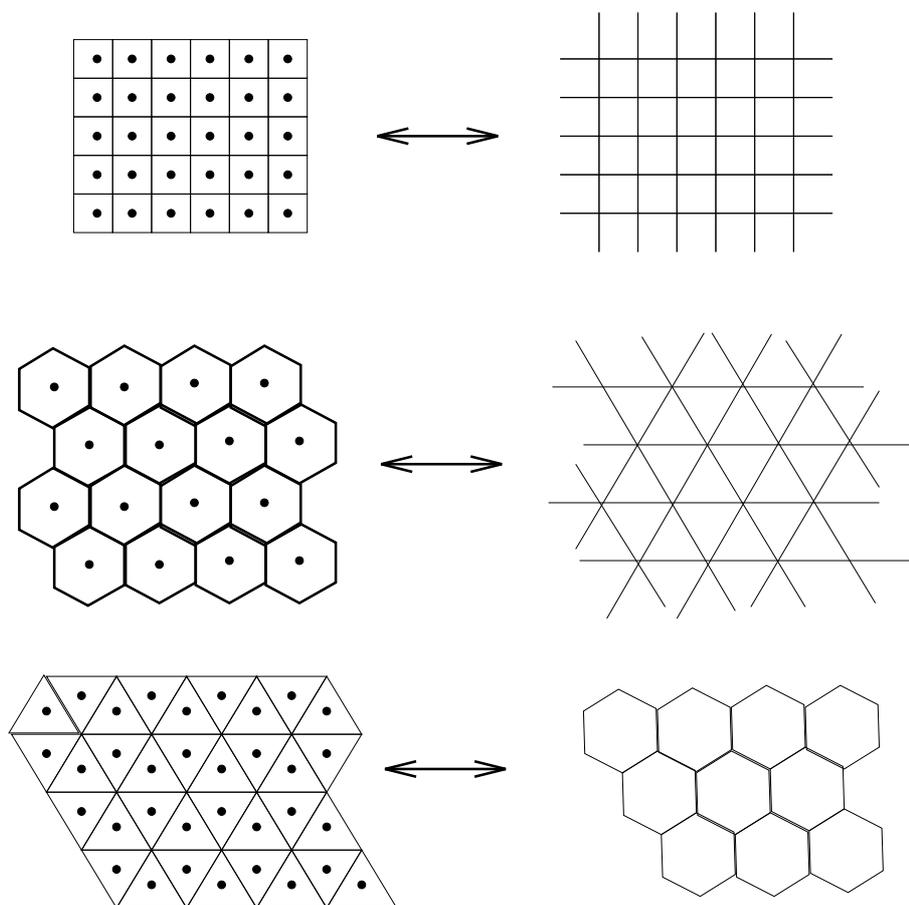


FIG. 4.5 – Dualité entre les maillages et les pavages réguliers du plan.

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.

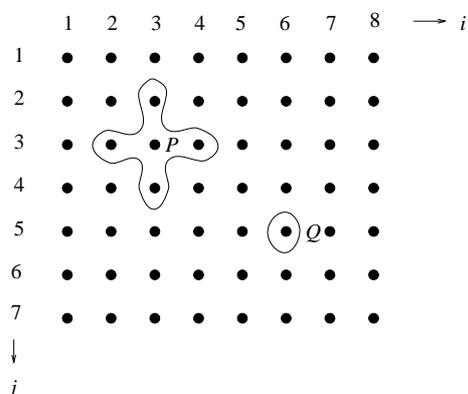


FIG. 4.6 – Définition d'une base topologique dans le plan discret sur une trame carrée.

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.

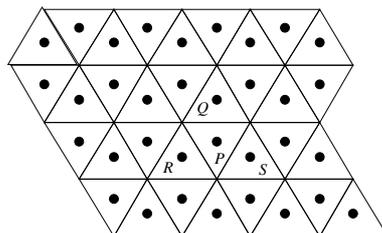


FIG. 4.7 – Définition d'une base topologique dans le plan discret sur une trame triangulaire. Sur cette figure, $\mathcal{U}(P) = \{P, Q, R, S\}$, $\mathcal{U}(Q) = \{Q\}$, $\mathcal{U}(R) = \{R\}$, $\mathcal{U}(S) = \{S\}$.

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.

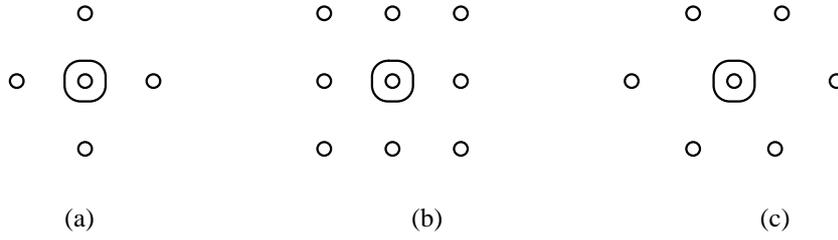


FIG. 4.8 – Connexité sur une trame discrète carrée (a : 4-connectivité, b : 8-connectivité), ou hexagonale (c : 6-connectivité).

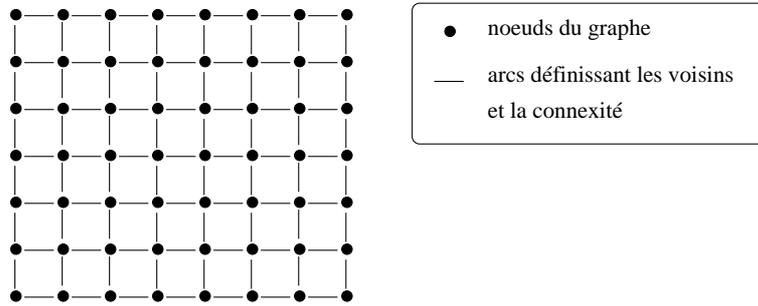


FIG. 4.9 – Graphe définissant la 4-connectivité sur une trame carrée.

Sur une trame hexagonale, la connexité la plus naturelle est la 6-connectivité (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-connectivité. 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-connectivité : les points de coordonnées $(i - 1, j)$, $(i + 1, j)$, $(i, j - 1)$, $(i, j + 1)$,
- en 8-connectivité : 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-connectivité. 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

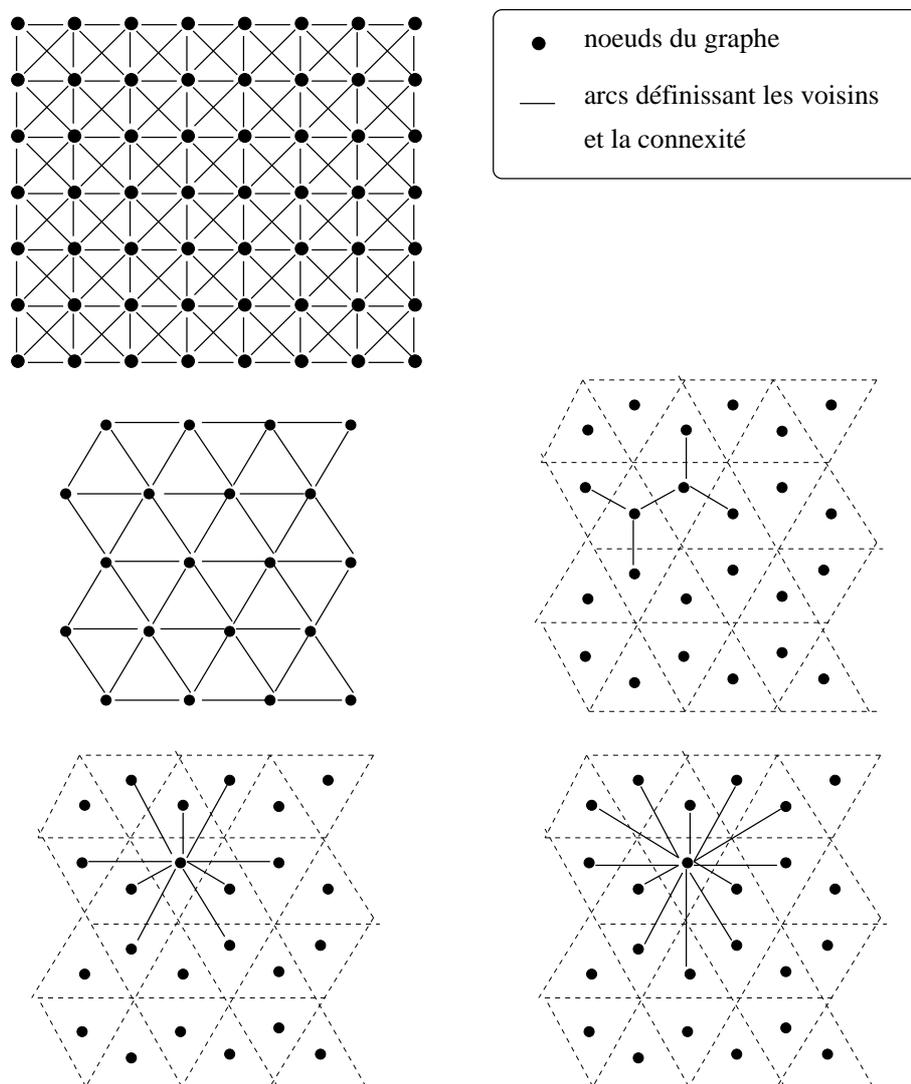


FIG. 4.10 – Graphes définissant la 8-connectivité sur une trame carrée, la 6-connectivité sur une trame hexagonale, et les 3-, 9- et 12-connectivités sur une trame triangulaire (seuls quelques arcs sont représentés pour la trame triangulaire).

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 connectivité 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 connectivité 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 :

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

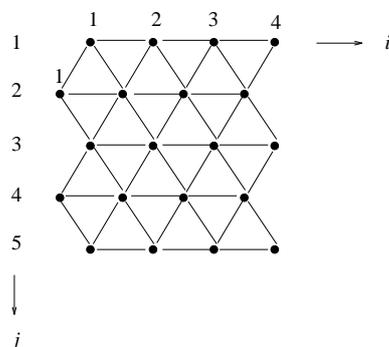


FIG. 4.11 – Les indices des voisins d'un point sur une trame hexagonale dépendent de la parité de la ligne sur laquelle se trouve le point.

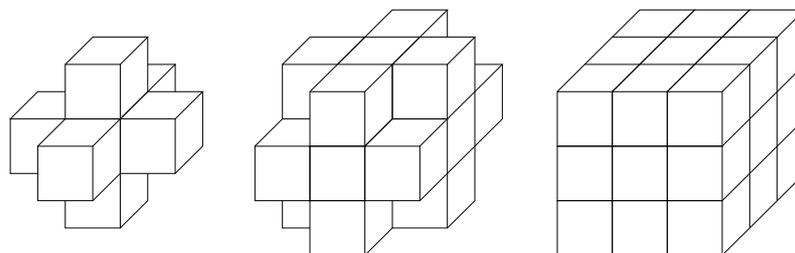


FIG. 4.12 – Les trois types de voisinages élémentaires pour une image 3D définie sur une trame cubique.

– 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,

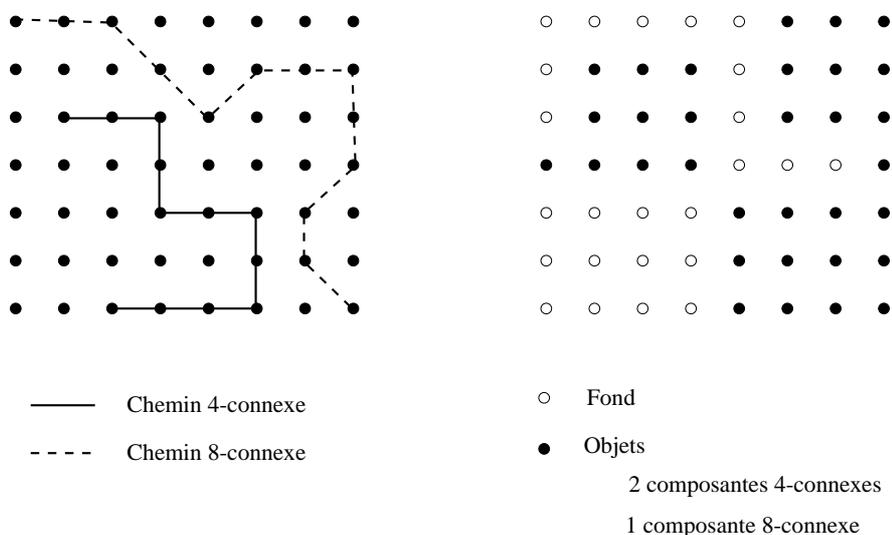


FIG. 4.13 – Illustration des définitions des chemins et composantes connexes sur une trame carrée, en 4- et 8-connexités.

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.

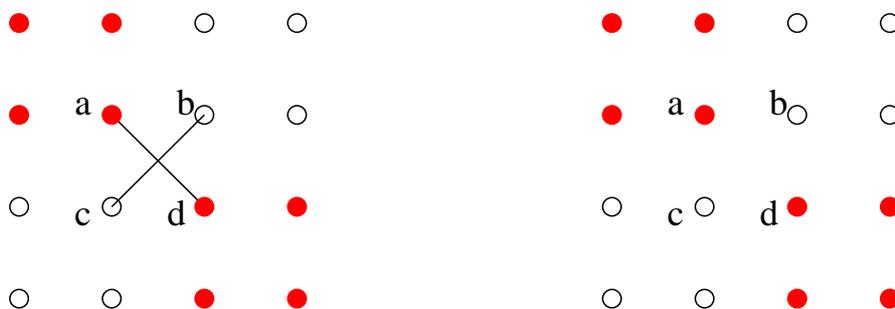


FIG. 4.14 – Paradoxe sur une trame discrète carrée en 8- et 4-connexité.

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 $i = j \pm 1[n + 1]$.

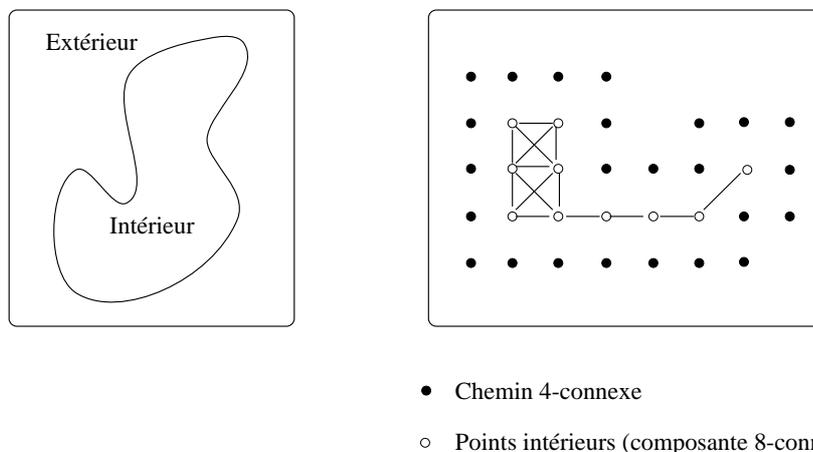


FIG. 4.15 – Illustration du théorème de Jordan dans le cas continu et dans le cas discret sur une trame carrée (les points extérieurs ne sont pas représentés).

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 utilisée.

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

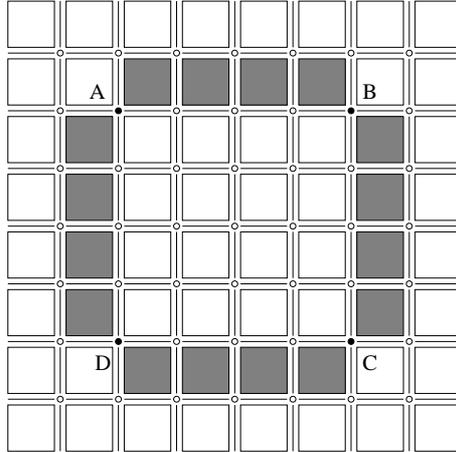


FIG. 4.16 – Le théorème de Jordan est toujours vérifié dans le modèle des complexes cellulaires. Chaque pixel est décomposé en carré, arête et sommet. La couleur des points A, B, C et D détermine la connexité de la courbe noire et celle du fond.

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.

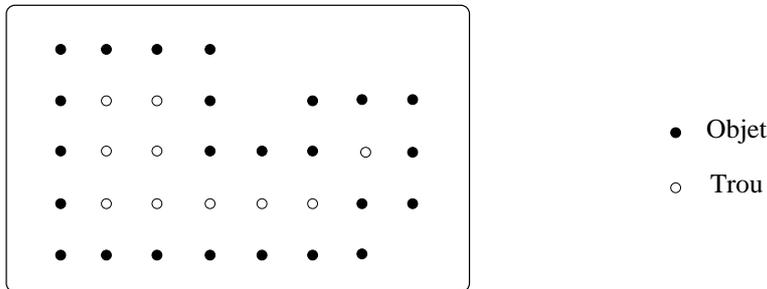


FIG. 4.17 – Nombre d'Euler sur une trame carrée : si on considère les objets en 8-connexité et les trous en 4-connexité, $N_{cc} = 1$ et $N_t = 2$, donc $E = -1$. Avec les conventions inverses, $N_{cc} = 1$ et $N_t = 1$, donc $E = 0$.

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

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.

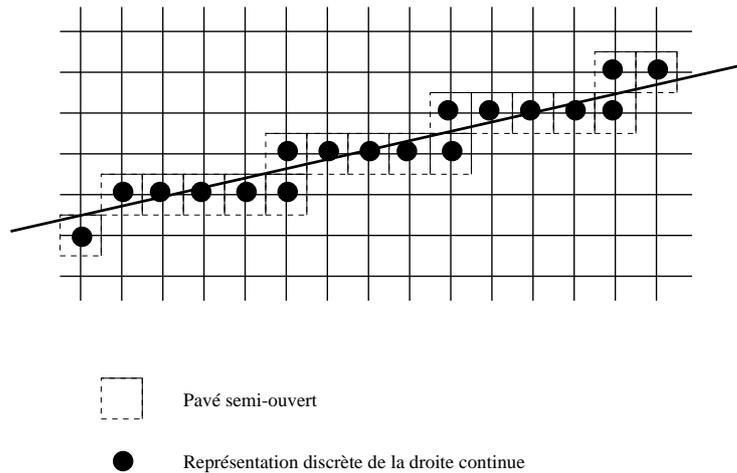


FIG. 4.19 – Discrétisation d'un segment de droite par la méthode du pavé semi-ouvert.

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.

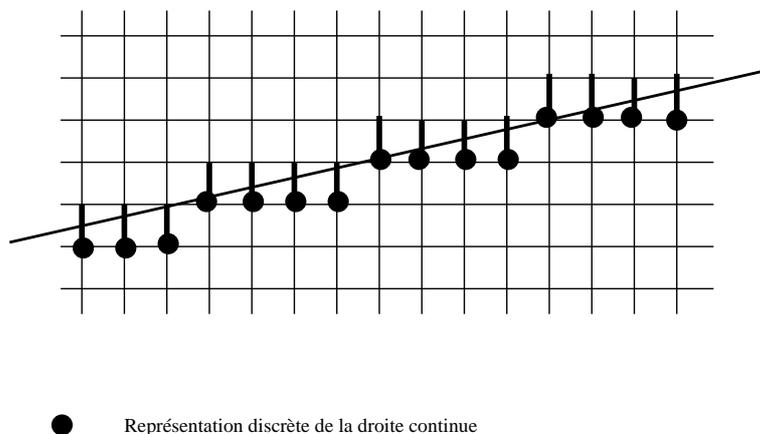


FIG. 4.20 – Discrétisation d'un segment de droite par la contrainte d'unilatéralité.

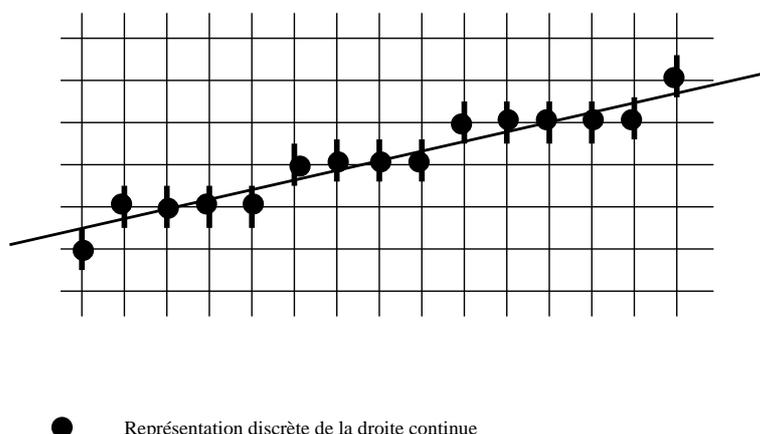


FIG. 4.21 – Discrétisation d'un segment de droite par la méthode de segment vertical semi-ouvert.

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

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

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 [Hégon, 1985].

la propriété de la corde si et seulement si :

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

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.

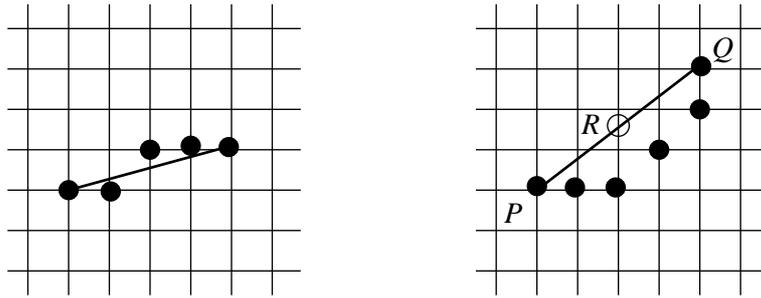


FIG. 4.22 – L'ensemble de points discrets de gauche vérifie la propriété de la corde, celui de droite ne la vérifie pas (le point R du segment $[P, Q]$ est à une distance de plus de 1 de tout point discret de \mathcal{S}).

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.

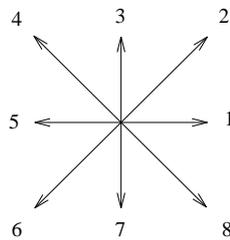


FIG. 4.23 – Les 8 directions possibles sur une trame carrée. Deux directions sont consécutives si leurs numéros diffèrent de 1 (modulo 8).

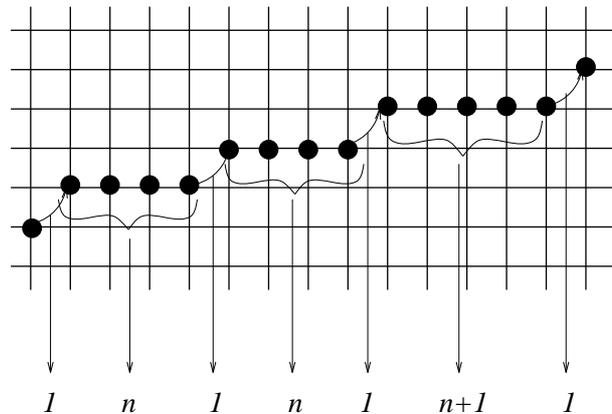


FIG. 4.24 – Caractérisation syntaxique d'un segment de droite discret.

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. 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

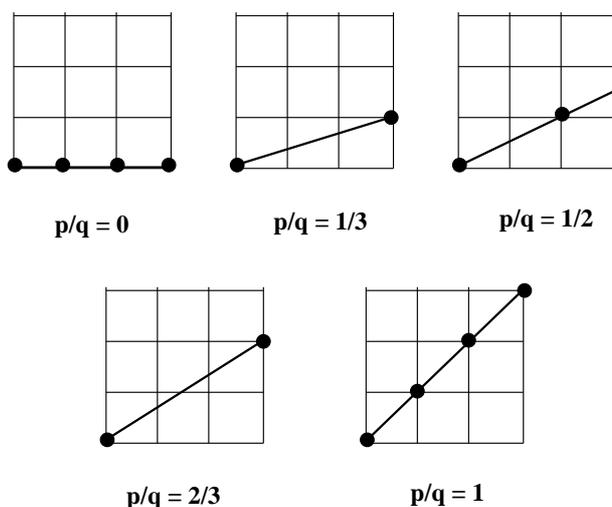


FIG. 4.25 – Les droites possibles (d’intersection non vide avec l’ensemble des points de la trame) sur une image 4×4 .

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 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 :

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

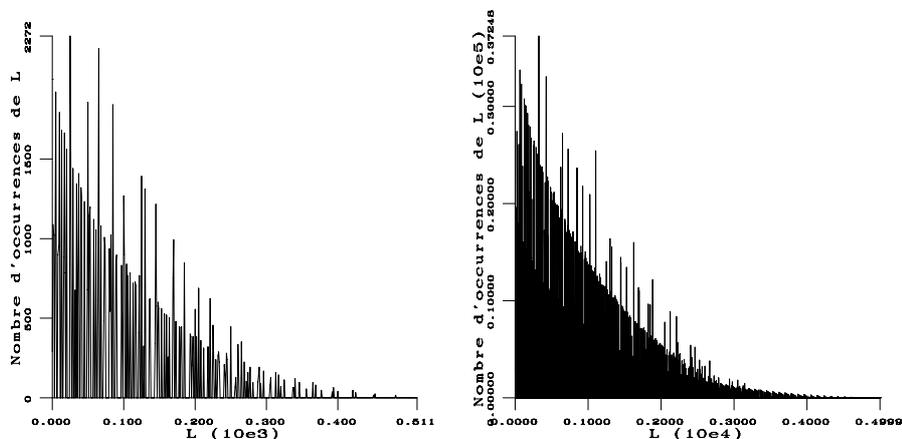


FIG. 4.26 – Distribution des carrés des longueurs des segments discrets sur une image 16 × 16 et sur une image 50 × 50.

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.

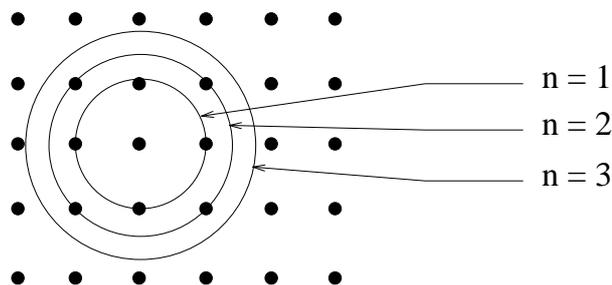


FIG. 4.27 – Intersection de cercles continus avec des points de la trame carrée.

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.

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,

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

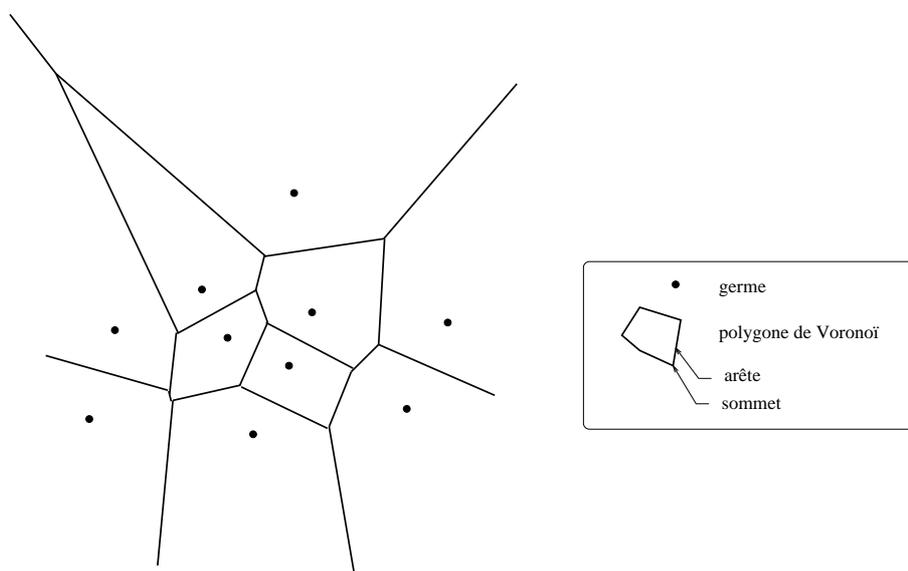


FIG. 4.28 – Pavage de Voronoï à partir d'un ensemble de points.

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.

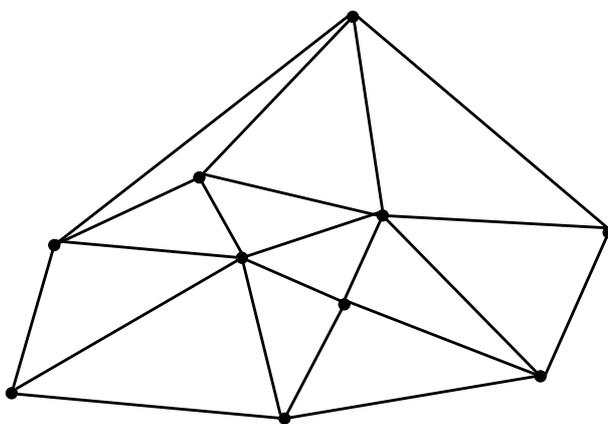


FIG. 4.29 – Triangulation de Delaunay d'un ensemble de points.

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.

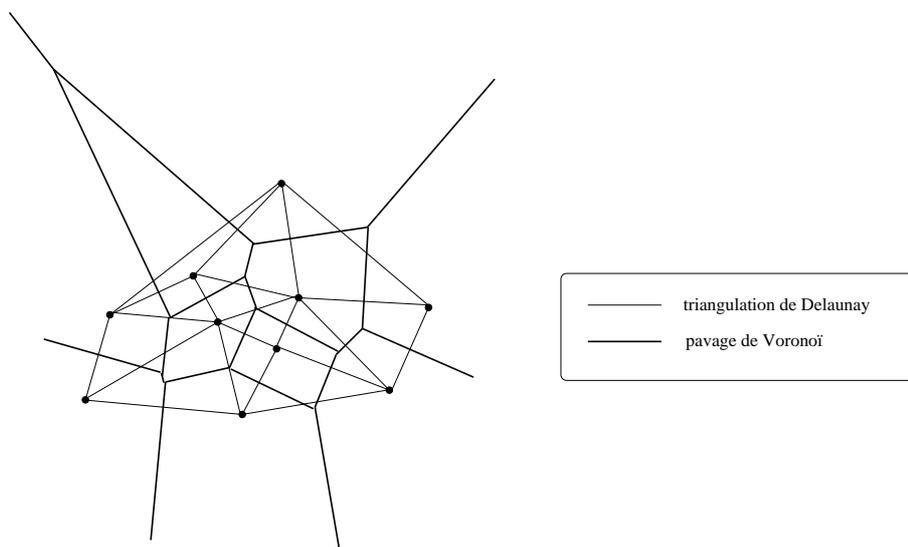


FIG. 4.30 – Dualité entre triangulation de Delaunay et pavage de Voronoï.

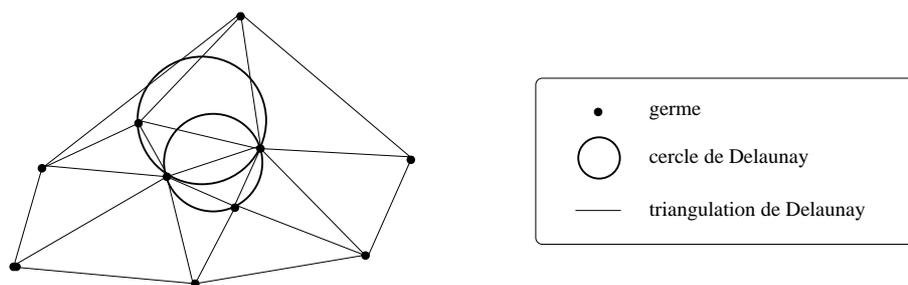


FIG. 4.31 – Propriétés géométriques de la triangulation de Delaunay.

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 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.

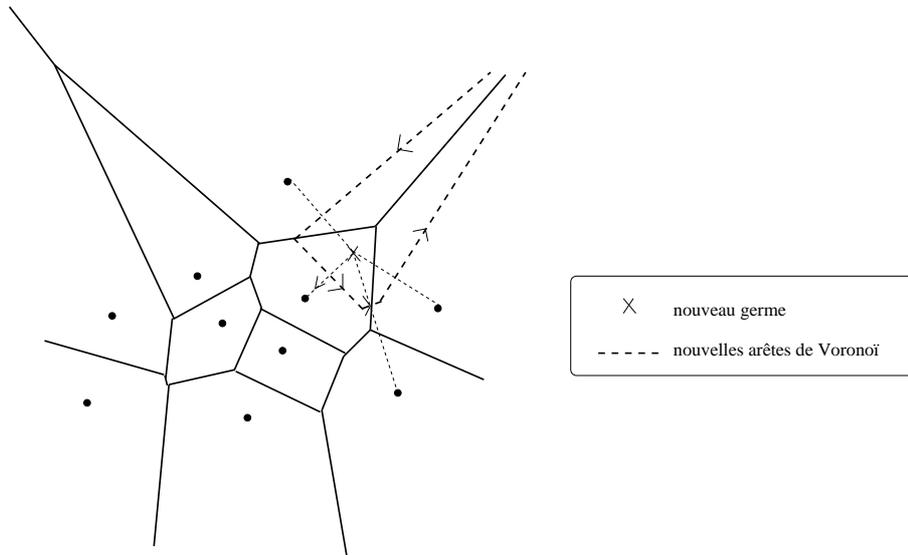


FIG. 4.32 – Construction incrémentale du pavage de Voronoï.

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 donnés 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.

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

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 $(\alpha, \alpha + d\alpha)$, 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).

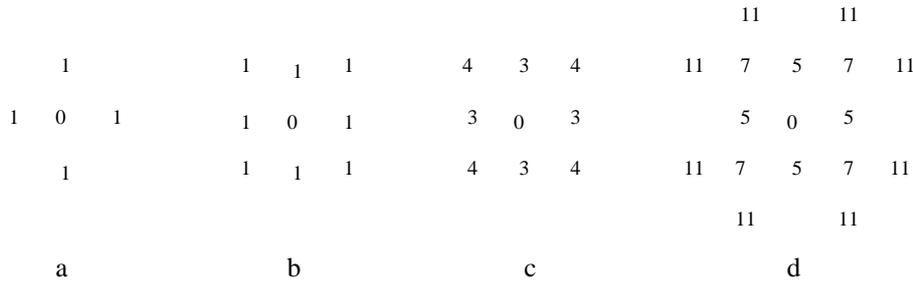


FIG. 4.33 – Exemples de masques pour la distance du chanfrein.

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 ré-écriture, 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, Elsa ANGELINI et Saïd LADJAL

La restauration des images est l'opération qui corrige des défauts engendrés lors du processus d'enregistrement numérique. Des exemples de défauts incluent le flou de mise au point ou de mouvement, ou le bruit de source électroniques. Les méthodes de restauration proposent des algorithmes d'élimination ou de réduction des défauts, permettant de reconstruire une image de bonne qualité à partir d'une acquisition de médiocre qualité. Plus précisément, on peut distinguer l'**amélioration** d'image qui corrige des défauts généralement inconnus par des techniques heuristiques et la **restauration** d'images qui corrige les effets d'un défaut connu par inversion de celui ci. Des méthodes récentes proposent des approches de « restaurations aveugles » dans lesquelles le défaut, inconnu, est estimé pendant le processus de restauration. 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. Ce chapitre présente les méthodes de restaurations qui utilisent un modèle du défaut ou une estimation de celui ci pour éliminer les dégradations engendrées par lui dans l'image. Les approches d'estimations de défauts sont également présentées.

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. Nous ne traiterons dans ce chapitre que de problèmes de restauration d'images dégradées par un système linéaire et spatialement invariant. 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 Filtrage inverse

Le filtrage inverse, ou déconvolution, a commencé à être utilisé dans les années 1960 pour la restauration d'images spatiales, acquises au cours des premières missions d'exploration de planètes, et d'images de télescopes, dégradées par des turbulences atmosphériques.

La théorie des systèmes linéaire fut exploitée en optique ou traitement du signal bien avant l'avènement des méthodes de traitement d'image. Des techniques de déconvolution avaient été mises en place pour la conception de filtres numériques et l'analyse de séries temporelles de mesures par exemple. Ainsi, les premières méthodes de restauration d'image furent inspirées d'approches monodimensionnelles utilisées en traitement du signal analogique et digital. Nous considérerons donc tout d'abord la restauration des signaux monodimensionnels et repousserons l'aspect bi- et tri-dimensionnel à la section 5.4.3. Deux approches du problème sont présentées : analytique, et algébrique, dont certaines correspondances sont mises en valeur.

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 :

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

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 :

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

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 :

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

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

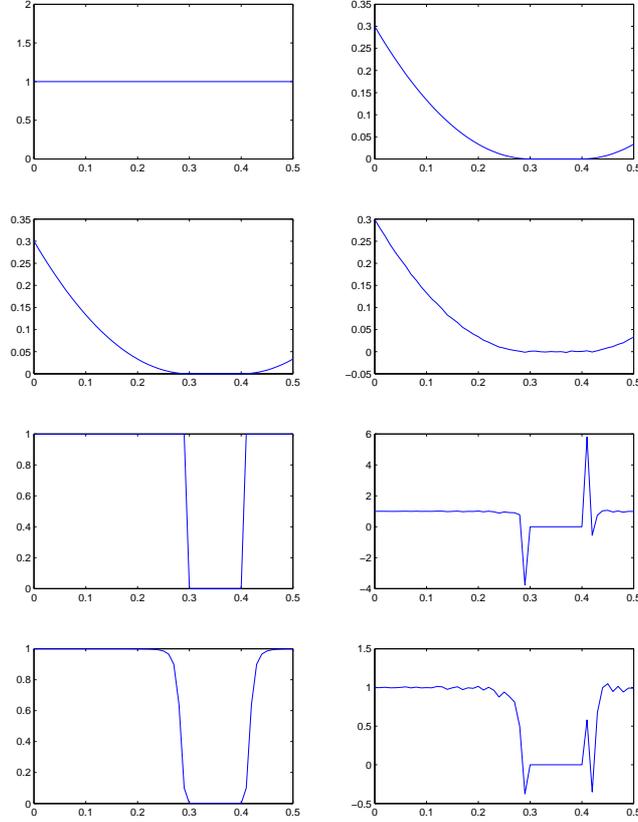


FIG. 5.1 — Illustration du filtrage par pseudo inverse et régularisation. On se place dans le domaine spectral. Sur la première ligne : Le signal d'origine et la fonction de transfert du défaut. Seconde ligne : L'observation sans bruit et l'observation avec bruit (1 millième). Troisième ligne : La restauration par pseudo inverse avec et sans bruit. Notez que le pic dans le cas bruité se situe à 6 pour un signal d'origine de maximum 1. Notez aussi que la pseudo inverse renvoie zéro dans les zones où le filtre est nul. Quatrième ligne : La restauration avec régularisation de Wiener dans les deux cas. Conclusion : même si le bruit est très faible, la pseudo inverse l'amplifie dans des proportions inacceptables quand le défaut est mal conditionné.

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) :

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

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 :

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

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 :

$$\Delta_c = \{d_{(j-k)_{\text{mod}N}}\}$$

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

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

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 :

$$\mathbf{p}_k = \left[\left\{ \left(e^{-\frac{2i\pi}{N}} \right)^k \left(e^{-\frac{2i\pi}{N}} \right)^{2k} \dots \left(e^{-\frac{2i\pi}{N}} \right)^{Nk} \right\} \right] \quad (5.12)$$

Les valeurs propres associées sont :

$$\lambda_k = \sum_j d_{jk} e^{\frac{2ik\pi}{N}j} \quad (5.13)$$

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.

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.

5.1.3 Modifications pour des défauts singuliers

Le problème principal de l'approche par **filtrage inverse** est la prise en compte de défauts non inversibles. 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.

1. 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 :

$$W_B(u) \begin{cases} = 0 & \text{si } D(u) = 0 \\ = \frac{1}{D(u)} & \text{sinon} \end{cases} \quad (5.14)$$

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

2. Dans l'**approche algébrique** avec une représentation vectorielle (équation 5.8), deux types de problèmes se présentent :

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.

- (a) d'une part la matrice Δ peut être rectangulaire,
- (b) d'autre part, elle peut ne pas être de rang plein.

On résoud ces impossibilités mathématiques par deux choix raisonnables

- (a) On veut que Δf soit le plus proche possible de g .
- (b) On veut (s'il y a plus d'un choix possible pour f suivant l'exigence précédente) choisir f de norme la plus petite possible.

On vérifie que ces deux exigences nous mènent à définir une application linéaire qui transforme l'objet observé g en une image restaurée f qui respecte nos deux conditions. Cette application (et la matrice qui la représente) est appelée **pseudo inverse**, que l'on va noter Δ^- .

On peut donner la forme de Δ^- dans trois cas particuliers (exercice !) :

- (a) Si Δ est injective : $\Delta^- = (\Delta^T \Delta)^{-1} \Delta^T$. Ce cas correspond à une situation où l'objet g est de plus grande (ou égale) dimension que l'objet que l'on recherche f . Par exemple, si f est décrit par un petit nombre de paramètres.
- (b) Si Δ est surjective : $\Delta^- = \Delta^T (\Delta \Delta^T)^{-1}$. Ceci correspond au cas où l'objet recherché f vit dans un espace plus grand que l'objet observé g .
- (c) Si Δ est diagonale alors Δ^- l'est aussi et on remplace les valeurs propres non nulles de Δ par leur inverse, les valeurs propres nulles sont conservées. Ce cas correspond, par exemple, au cas où Δ est circulaire et donc diagonalisable dans la base de Fourier, on retrouve alors la même formulation que dans 5.14

Remarquez que quand Δ est carrée, chacune des deux premières hypothèses conduit au fait que Δ est inversible et alors $\Delta^- = \Delta^{-1}$. Dans les cas où Δ n'est pas de rang maximal, la décomposition SVD donne la bonne définition de la pseudo inverse (voir plus loin).

Cette matrice a la propriété suivante : son rang est égal à celui de la plus grande matrice carrée régulière issue de Δ .

5.1.4 Conclusions (provisoires)

Nous avons présenté trois approches différentes pour restaurer le signal si celui-ci est représenté sous forme échantillonnée :

1. Inversion du système avec une matrice circulante Δ_c , qui ne permet de corriger que les **défauts spatialement invariants** en négligeant les effets de bord. Cette approche permet 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].
2. 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].
3. Inversion ordinaire de la matrice Δ , sans tenir compte de ses propriétés spécifiques par méthode du pivot de Gauss, ou méthode de Gauss Seidel par exemple [Durant, 1961, Porte et Vignes, 1974, Golub et van Loan, 1996]. Cette approche 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 $\mathcal{O}(N^3)$.

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 5.7 comment on peut atténuer les effets de bord lors de la restauration d'images.

| | 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.

5.2 Régularisation pour les approches algébriques

5.2.1 Décomposition du système linéaire en SVD

Le système linéaire étudié $g = \Delta f$ peut être étudié pour évaluer son conditionnement, par décomposition en valeurs singulières (SVD) [Golub et van Loan, 1996] de la matrice Δ de taille $[M \ N]$ en supposant simplement que $M > N$. Cette décomposition s’écrit :

$$\begin{aligned}
 \Delta &= V \Sigma U^T & \Sigma &= \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N) \\
 U^T \times V &= Id & V &\in \mathfrak{R}^{M \times N} \\
 V^T \times U &= Id & U &\in \mathfrak{R}^{N \times N} \\
 \sigma_1 &\geq \sigma_2 \geq \dots \geq \sigma_N
 \end{aligned} \tag{5.15}$$

Cette décomposition existe toujours, et les valeurs singulières $(\sigma_1, \sigma_2, \dots, \sigma_N)$ sont uniques (mais pas forcément la décomposition SVD elle-même, et les vecteurs singuliers associés). De nombreuses propriétés du système linéaire défini par Δ peuvent être déduites de sa SVD. Ainsi, le rang r du système est donné par le rang de la plus petite valeur singulière σ_r non nulle. Le noyau de Δ est donné par les lignes de V de $r+1$ à M : $\text{Ker}(A) = \{v_{r+1}, \dots, v_M\}$ et l’image de Δ est définie par les lignes de U de 1 à r : $\text{Im}(A) = \{u_1, \dots, u_r\}$. Pour les systèmes **réguliers** (i.e. Δ inversible), la décomposition SVD de Δ^{-1} est immédiatement obtenue à partir de celle de Δ .

5.2.2 Mesure du conditionnement du système

Le conditionnement d’un système linéaire caractérise son comportement en présence de petite perturbation : il est bien conditionné si la sortie g varie peu pour des petites variations sur Δ ou g , et est mal conditionné dans le cas contraire.

1. Conditionnement des systèmes réguliers Le degré de conditionnement d’un système linéaire régulier peut être évalué par le nombre $C = \|\Delta\| \times \|\Delta^{-1}\|$. les normes appliquées aux matrices sont le maximum de $\|Mx\|$ lorsque x est de norme 1. Le nombre C est toujours plus grand ou égal à 1 ($C = 1$ correspond à un excellent conditionnement). Par la SVD de Δ on obtient $C = \sigma_1/\sigma_N$, qui est le rapport entre la plus grande et la plus petite valeur singulière.

Un système est mal conditionné si C est de l’ordre de l’inverse de la précision de la machine sur laquelle on travaille.

2. Conditionnement des systèmes singuliers Dans le cas d’un système singulier Δ , de rang $r < N$, la SVD permet de définir la pseudo inverse de Δ par :

$$\Delta^- = U_r \times \Sigma_r^{-1} \times V_r^T \tag{5.16}$$

qui ne garde que les r premières composantes de la SVD de Δ . Les raisonnements précédents sur le conditionnement du système peuvent alors être faits, à partir de la norme de la pseudo inverse.

Dans le cadre de la restauration par filtrage inverse, la matrice Δ est souvent mal conditionnée [DUR], dû à la présence de défauts ou perturbations dans le système d'acquisition, et le système numérique à résoudre doit être **régularisé**, en imposant de contraintes de norme minimale sur la solution ou sur l'erreur *a posteriori* commise entre l'estimation de f (via Δf) et l'observation g . De nombreuses techniques ont été proposées pour régulariser l'équation 5.2 [Twomey, 1965, Phillips, 1962, Tikonov et Arsenine, 1974]. Pour formaliser les différentes approches de régularisation, il est pratique d'ajouter dans l'équation 5.8 une composante de bruit $\mathbf{n}(\mathbf{x})$ additif, aléatoire, de moyenne nulle, non corrélé au signal. L'équation 5.8 devient alors :

$$\mathbf{g} = \Delta \mathbf{f} + \mathbf{n} \quad (5.17)$$

Les méthodes de régularisation visent alors à minimiser la norme de cette composante de bruit.

5.2.3 Régularisation par moindres carrés

Un premier résultat fondamental à connaître est que, dans le cas des systèmes singuliers, non-directement inversibles, la solution obtenue avec la pseudo-inverse de Δ :

$$\mathbf{f}^r = \Delta^{-} \mathbf{g} \quad (5.18)$$

est la solution de norme L_2 $\|\mathbf{f}\|$ minimale et qui minimise l'erreur $\|g - \Delta f\|$, égale à $\|\mathbf{n}\|$ (comme nous l'avons déjà dit).

Suivant ce qui a été dit précédemment, la régularisation du système $g = \Delta f$ peut être vue comme l'estimation de f^r qui minimise la fonctionnelle $W(\mathbf{f}^*)$ suivante :

$$W(\mathbf{f}^r) = \|\mathbf{Q}(\mathbf{f}^r)\|^2 + \lambda \|\mathbf{n}\|^2 \quad (5.19)$$

où Q est un opérateur de contrainte sur le signal restauré \mathbf{f}^r , qui permet de contrôler l'apparence finale de l'image, et λ est un multiplicateur de Lagrange.

On minimise $W(\mathbf{f}^r)$ en annulant la dérivée de la fonctionnelle par rapport à \mathbf{f}^r :

$$\frac{\partial W(\mathbf{f}^r)}{\partial \mathbf{f}^r} = 2Q^T Q \mathbf{f}^r - 2\lambda \Delta^T (\mathbf{g} - \Delta \mathbf{f}^r) = \mathbf{0} \quad (5.20)$$

La solution pour \mathbf{f}^r est donnée par :

$$\mathbf{f}^r = \left[\Delta^T \Delta + \left(\frac{1}{\lambda} \text{Id} \right) Q^T Q \right]^{-1} \Delta^T \mathbf{g} \quad (5.21)$$

Cette equation fournit la solution générale de restauration contrainte par moindres carrés. Nous pouvons regarder plusieurs cas particuliers :

1. $Q = 0$. La restauration est effectuée sans contrainte sur la norme de l'image et on obtient l'estimés par filtrage pseudo-inverse.
2. $Q = Id$. On a pour l'image restaurée :

$$\mathbf{f}^r = \left[\Delta^T \Delta + \left(\frac{1}{\lambda} \text{Id} \right) \right]^{-1} \Delta^T \mathbf{g} \quad (5.22)$$

3. $Q = SNR^{-1}$ (Filtre de Wiener Paramétrique) Si on traite f^r et n comme des variables aléatoires et on prend comme contrainte de maximiser le ratio signal sur bruit (SNR) dans l'image restaurée f^r . Ce ratio est calculé avec les matrices de covariances : $R_{f^r} = E[f^r f^{rT}]$ et $R_n = E[nn^T]$, comme :

$$Q = \left(\frac{R_n}{R_{f^r}} \right)^{\frac{1}{2}} \quad (5.23)$$

Dans ce cas on obtient

$$f^r = \left[\Delta^T \Delta + \left(\frac{1}{\lambda} \frac{R_n}{R_{f^r}} \right) \right]^{-1} \Delta^T g \quad (5.24)$$

Dans le cas stationnaire, invariant par translation, ceci est égale au filtre de Wiener, qui sera défini dans le domaine de Fourier un peu plus tard.

4. $Q =$ Contrainte de lissage Dans la plupart des cas, on utilise Q pour lisser l'image restaurer et éviter de reconstruire le bruit présent dans l'image observée. Par exemple, Q peut contraindre le Laplacien de l'image restaurée $\nabla^2(f^r)$ en choisissant Q comme la convolution par la matrice suivante (attention Q n'est pas une matrice 3×3 ! $Q^T Q$ est de la même taille que $\Delta^T \Delta$) :

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (5.25)$$

5.3 Régularisation pour les approches analytiques

Comme présenté précédemment pour les approches algébriques, pour formaliser les contraintes de régularisation, il est pratique d'ajouter dans l'équation 5.2 une composante de bruit $n(x)$ additif, aléatoire, de moyenne nulle, non corrélé au signal. L'équation 5.2 devient alors :

$$g(x) = f(x) * d(x) + n(x) \quad (5.26)$$

Nous présentons ici l'approche analytique proposée par **Wiener** [Hellstrom, 1967]. .

Wiener recherche une solution linéaire f^r (donc obtenue 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 n :

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

Appliquant le théorème de Parseval, on obtient dans le domaine de Fourier :

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

remplaçant f^r par $G \times W_W$, permutant intégrale et espérance, et imposant à chaque composante fréquentielle u de minimiser sa contribution à ϵ , on obtient :

$$\frac{\partial \epsilon}{\partial W_W(u)} = 0 \quad \implies \quad W_W(u) = \frac{D^*}{DD^* + \Phi_n / \Phi_{f^r}} \quad (5.27)$$

où Φ_n et Φ_{f^r} sont respectivement les densités de puissance du bruit et du signal :

$$\Phi_b(u) = \langle N(u)N^*(u) \rangle \quad (5.28)$$

$$\Phi_f(u) = \langle F(u)f^r(u) \rangle \quad (5.29)$$

On peut alors distinguer plusieurs comportements du filtre de Wiener :

1. Défaut singulier et absence de bruit : la solution est équivalente à celle du filtre inverse.
2. Défaut singulier : la solution proposée est celle de Bracewell.
3. Système mal conditionné : quand D 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.

La restauration effectuée sur un défaut mal conditionné par le filtre de Wiener est illustrée Fig. 5.2

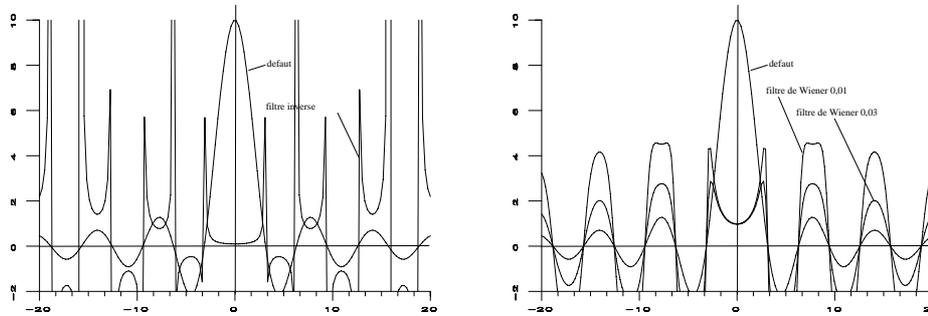


FIG. 5.2 – A gauche : fonction de transfert ($\sin u/u$) d'un défaut de bougé à vitesse constante et le filtre inverse permettant de le corriger. A droite, deux filtres de Wiener de ce même défaut obtenus à partir de deux valeurs différentes du rapport Φ_b/Φ_f supposés constants.

5.4 Mise en place calculatoire

5.4.1 Wiener : estimation des spectres de densité de puissance

Les deux termes de densité spectrale $\Phi_n = R_n$ et $\Phi_{f^r} = R_{f^r}$ revêtent une importance particulière dans la régularisation de la restauration par l'approche de Wiener. Ils sont malheureusement le plus souvent inconnus lors de la résolution d'un problème particulier. Deux approches peuvent être alors choisies :

1. Une façon très répandue de circonvenir cette ignorance consiste à considérer le **rapport signal à bruit** $SNR = \frac{\Phi_n}{\Phi_{f^r}}$ 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 SNR variables.
2. 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_n , 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 :

$$\frac{\Phi_n(u)}{\Phi_{f^r}(u)} = \frac{\sigma_n^2(\alpha^2 + u^2)}{2\alpha}$$

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

5.4.2 Prise en compte de la décomposition en SVD

Afin de calibrer le système à restaurer et d'éviter des problèmes numériques dus à son mauvais conditionnement, on peut utiliser les bases de vecteurs identifiées par la SVD du système pour effectuer les pré-corrrections suivantes :

1. pour toutes les valeurs singulières σ_i de Σ très supérieures au bruit, on peut inverser le système, et multiplier par $\frac{1}{\sigma_i}$ chaque composante du signal ;
2. pour toute valeur singulière σ_i nulle ou très inférieure au bruit, et en l'absence d'une information a priori sur le signal, on annule la composante du signal correspondante ;
3. pour les valeurs singulières très proches du bruit, on met en œuvre une technique de régularisation qui apparaîtra appropriée, en tirant parti au mieux des connaissances existant sur le signal.

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

5.4.3 Restauration des signaux bidimensionnels

1. Dans le cas de l'approche **analytique**, 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.
2. Dans le cas de l'approche **algébrique**, l'extension est plus délicate, puisque l'image ne se présente pas sous forme d'un vecteur monodimensionnel. Il faut la ramener à cette forme, à l'aide de ce que l'on appelle la forme développée des images [Pratt, 1975]. Cette forme s'obtient en rangeant les lignes de l'image dans un vecteur $\bar{\mathbf{f}}$, de dimension très grande, (à N^2 composantes si l'image est de côté N), les unes après les autres.

$$f_{jk} \longrightarrow \bar{f}_m = f_{jk} \quad \text{ssi} \quad m = k + jN$$

La relation 5.6 devient alors :

$$\bar{\mathbf{g}} = \Delta_{TbT} \bar{\mathbf{f}}$$

où Δ_{TbT} est maintenant une matrice Töplitz-blocs-Töplitz, c'est-à-dire une matrice de taille $N^2 \times N^2$, composée de $N \times N$ matrices de Töplitz, chacune de taille $N \times N$, agencées de façon Töplitz, c'est à dire de façon que des matrices identiques se retrouvent sur une même diagonale :

$$\Delta_{TbT} = \begin{bmatrix} \Delta_T^0 & \Delta_T^1 & \Delta_T^2 & \dots & \dots & & \\ \Delta_T^{-1} & \Delta_T^0 & \Delta_T^1 & \Delta_T^2 & \dots & \dots & \\ \vdots & \Delta_T^{-1} & \Delta_T^0 & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & & & \\ & & & & \ddots & & \vdots \\ & & & & & \ddots & \Delta_T^1 \\ & & & & & \dots & \Delta_T^{-1} \\ & & & & & & \Delta_T^0 \end{bmatrix} \quad (5.30)$$

Les conditions de cette convergence sont aisément discutées à partir de la TF de l'équation 5.33 :

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

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 $|1-D| < 1$. 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 5.33 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 ($\alpha < 1$) :

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

5.5.2 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 :

$$f^{k+1} = g + \Delta'_2 f^k + \Delta''_2 f^{k+1}$$

5.5.3 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).

5.6 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 :

$$d(x, y) = \frac{J_1^2(\pi \rho D \sqrt{x^2 + y^2} / \phi)}{(\pi \sqrt{x^2 + y^2} \rho D / \phi)^2} \quad (5.34)$$

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].

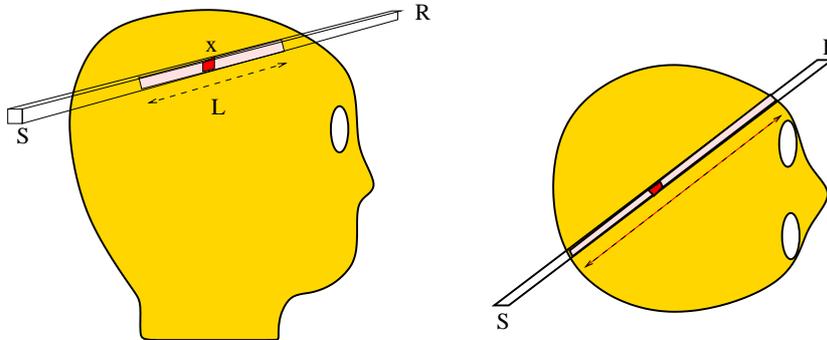


FIG. 5.3 – Construction de l'image en tomographie médicale. S est la source de rayons X, R est le récepteur. Une mesure g est l'intégrale le long de L de l'atténuation des rayons X due aux tissus.

5.6.1 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.

5.6.2 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 :

$$P_g = \langle |G|^2 \rangle = \langle |FD|^2 \rangle = \langle |F|^2 \rangle \langle |D|^2 \rangle = P_f |D|^2 \quad (5.35)$$

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 5.34, mais dans le plan des fréquences.

On voit sur l'équation 5.35 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.

5.7 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 décrit les opérations de base et leurs principales propriétés, et présente quelques applications immédiates. 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 :

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

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

Semi-continuité : ce principe 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 :

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

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

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 :

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

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

Ψ est anti-extensive si :

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

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

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 :

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

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 :

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

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 (non développé dans ce cours, voir UEs de spécialité).

L'équation 6.2 est équivalente aux expressions suivantes :

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

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

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

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

- 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.

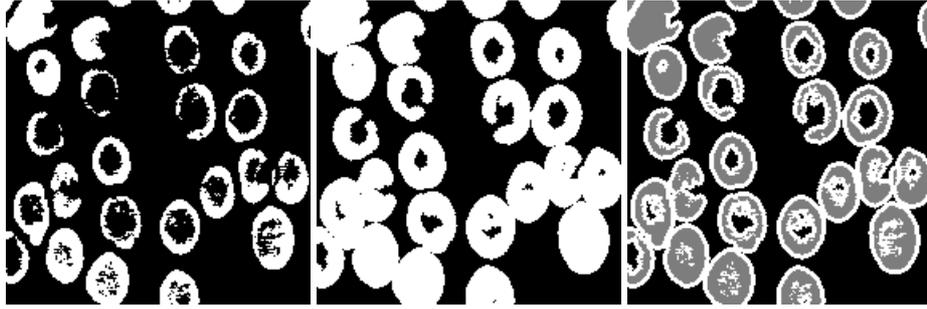


FIG. 6.1 – Exemple de dilatation binaire (de gauche à droite : image initiale, dilatation par un disque de taille 3, différence : en blanc, les parties rajoutées par la dilatation).

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

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

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

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,
- $B \subseteq B' \Rightarrow E(X, B') \subseteq E(X, B)$,
- elle satisfait les relations suivantes par rapport à la réunion et à l'intersection (en particulier, elle commute avec l'intersection) :

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

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

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

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

- 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 :

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

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.

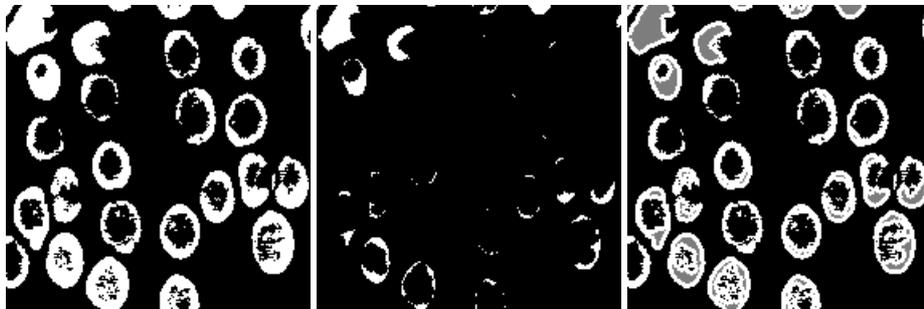


FIG. 6.2 – Exemple d'érosion binaire (de gauche à droite : image initiale, érosion par un disque de taille 3, différence : en blanc, les parties supprimées par l'érosion).

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 dilatations 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 dilatations 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 dilatations 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 :

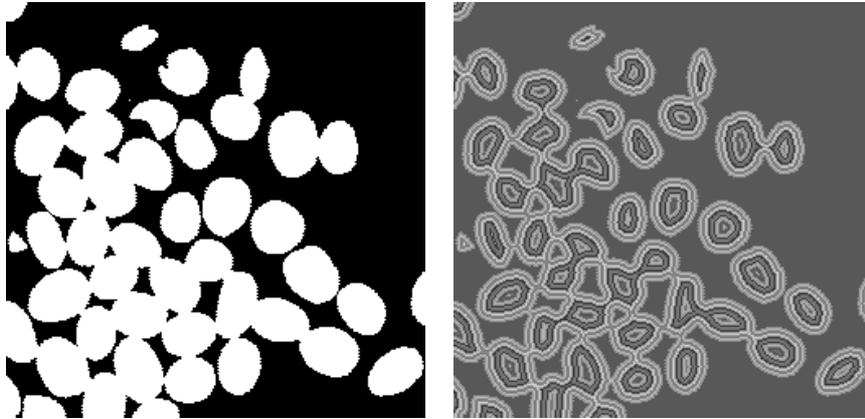


FIG. 6.3 – Les courbes d’iso-distances à l’intérieur des objets donnent les érodés successifs.

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

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

2. soit en considérant le sous-graphe de la fonction 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 :

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

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.

²On a le même résultat pour la dilatation binaire.

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.



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 :

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

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



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

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.

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 UEs de spécialité) : 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 :

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

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.

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).

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 recouvrées 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 :

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

Propriétés. L'ouverture a les propriétés algébriques suivantes :

- elle est anti-extensive ($X \supseteq X_B$)⁴,
- elle est croissante ($X \subseteq Y \Rightarrow X_B \subseteq Y_B$),
- 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 $B \subseteq B' \Rightarrow X_{B'} \subseteq X_B$, 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.

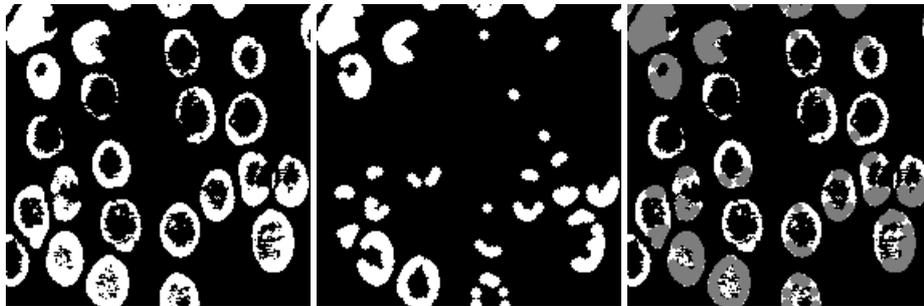


FIG. 6.6 – Exemple d'ouverture binaire (de gauche à droite : image initiale, ouverture par un disque de taille 3, différence : en blanc, les parties supprimées par l'ouverture).

⁴Contrairement au cas de la dilatation et l'érosion, cette propriété est vraie sans restriction sur l'élément structurant.

Définition 12. Fermeture binaire : La fermeture de l'ensemble X par l'élément structurant B est définie par :

$$X^B = E[D(X, B), \check{B}]. \quad (6.14)$$

Propriétés. La fermeture a les propriétés algébriques suivantes :

- elle est extensive ($X \subseteq X^B$),
- elle est croissante,
- elle est idempotente.

De même que pour l'ouverture, ces propriétés font de l'ouverture un filtre morphologique.

On a de plus $B \subseteq B' \Rightarrow X^B \subseteq X^{B'}$, 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.

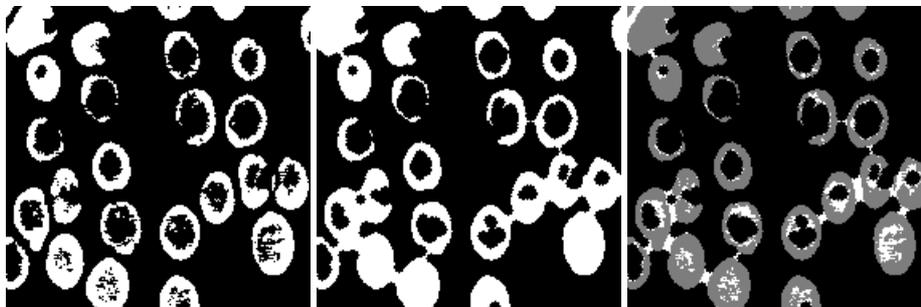


FIG. 6.7 – Exemple de fermeture binaire (de gauche à droite : image initiale, fermeture par un disque de taille 3, différence : en blanc, les parties rajoutées par la fermeture).

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.



FIG. 6.8 – Exemple d'ouverture numérique (de gauche à droite : image initiale et ouverture par un disque de taille 3).

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 :

$$f^B = E[D(f, B), \tilde{B}]. \quad (6.17)$$

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.

6.3 Quelques applications de l'érosion et de la dilatation

6.3.1 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,



FIG. 6.9 – Exemple de fermeture numérique (de gauche à droite : image initiale et fermeture par un disque de taille 3).

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.

6.3.2 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 :

$$EU(X) = \cup_n \{E(X, B_n) \setminus R[E(X, B_{n+1}); E(X, B_n)]\} \quad (6.18)$$

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).

6.3.3 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.

6.3.4 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.

6.4 Quelques applications de l'ouverture et de la fermeture

6.4.1 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 :

$$(\dots(((f_{B_1})^{B_1})_{B_2})^{B_2})\dots_{B_n})^{B_n}.$$



FIG. 6.11 – Exemple de gradient morphologique.

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.

6.4.2 Chapeau haut-de-forme

La transformation du « chapeau haut-de-forme » 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.

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.

6.4.3 Granulométries

Les granulométries sont des opérations de « tamisage » qui servent à sélectionner successivement des particules de tailles données croissantes.

Définition 15. 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. $\forall X \in \mathcal{A}, \forall \lambda > 0, \phi_\lambda(X) \subseteq X$ (ϕ_λ anti-extensive),



FIG. 6.12 – Filtres alternés séquentiels : image initiale, tailles maximales de 1, 2 et 3.

2. $\forall (X, Y) \in \mathcal{A}^2, \forall \lambda > 0, X \subseteq Y \Rightarrow \phi_\lambda(X) \subseteq \phi_\lambda(Y)$ (ϕ_λ croissante),
3. $\forall X \in \mathcal{A}, \forall \lambda > 0, \forall \mu > 0 \lambda \geq \mu \Rightarrow \phi_\lambda(X) \subseteq \phi_\mu(X)$ (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.

6.5 Introduction à la ligne de partage des eaux

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

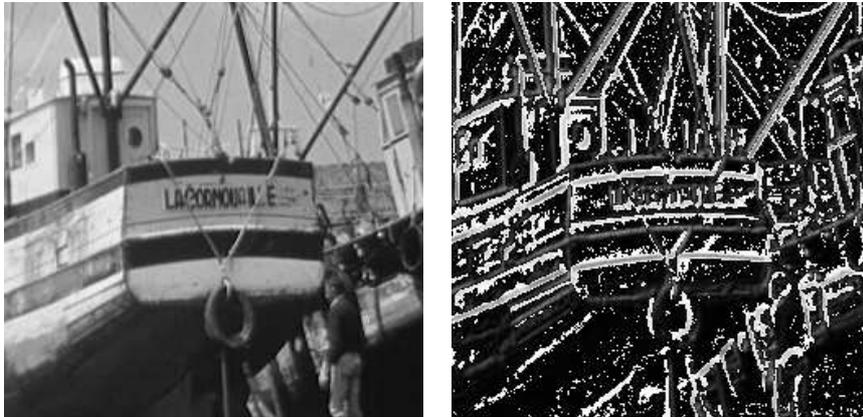


FIG. 6.13 – Exemple de chapeau haut de forme obtenu pour une ouverture de taille 3.

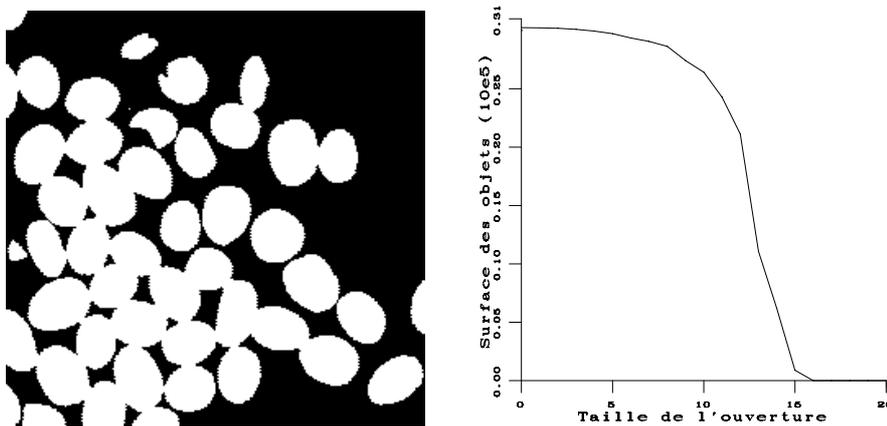


FIG. 6.14 – Image binaire et sa courbe de granulométrie.

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 6.15).

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. L'inconvénient est qu'elle fournit le plus souvent trop de régions (une sur-segmentation) et qu'il faut donc l'associer en pratique à des méthodes de filtrage et de marquage des zones d'intérêt. Tous ces aspects seront développés dans les UEs de spécialité.

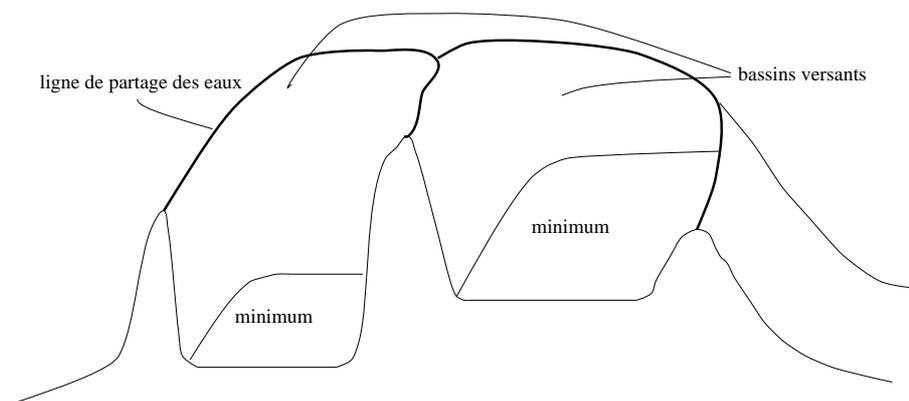


FIG. 6.15 – Ligne de partage des eaux.

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 .

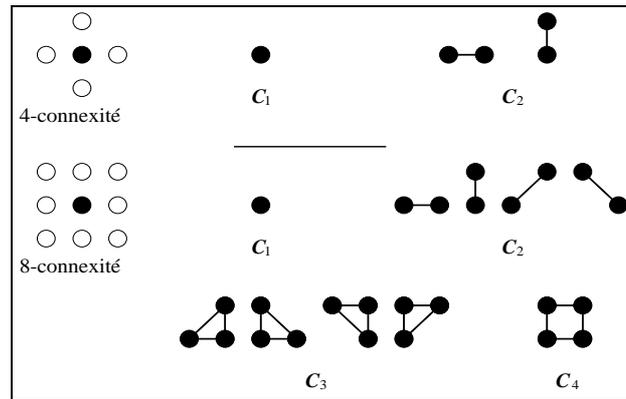


FIG. 7.1 – Les cliques associées à deux systèmes de voisinage en dimension 2.

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 :

$$\begin{aligned} P(X_s = x_s \mid X^s = x^s) &= \frac{\exp(-\sum_{c \in \mathcal{C} / s \notin c} U_c(x) - U_s(x_s / V_s))}{\sum_{x_s \in E} \exp(-\sum_{c \in \mathcal{C} / s \notin c} U_c(x) - U_s(x_s / V_s))} \\ &= \frac{\exp(-U_s(x_s / V_s))}{\sum_{x_s \in E} \exp(-U_s(x_s / V_s))} \end{aligned} \quad (7.1)$$

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 :

$$\forall x^{(0)} \quad \forall x \in \Omega \quad \lim_{n \rightarrow \infty} P(X^{(n)} = x | X^{(0)} = x^{(0)}) = P(x)$$

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 $\exp\left(\frac{U(x) - U^*}{T}\right) \rightarrow 0$ 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.

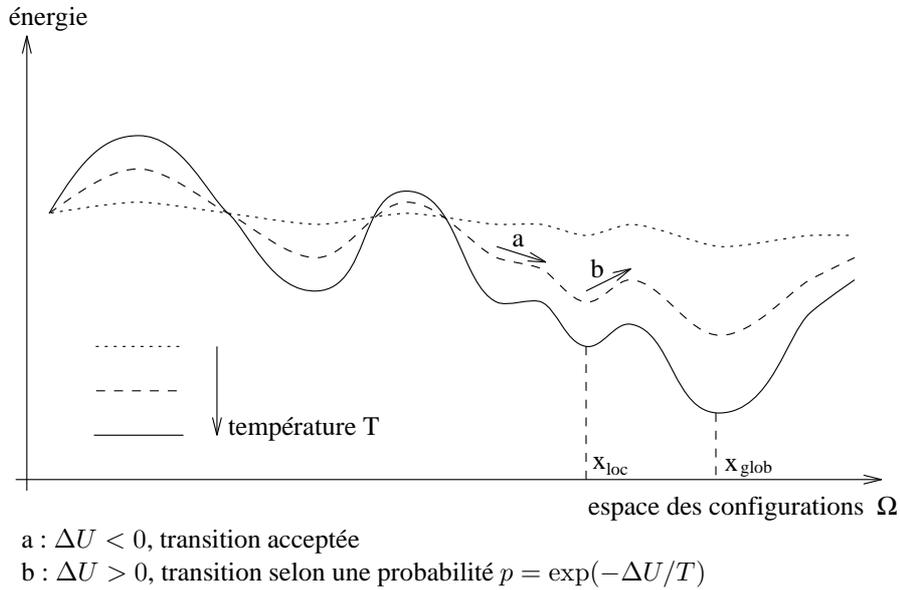
Notons que contrairement aux algorithmes de l'échantillonneur de Gibbs et de Métropolis qui échantillonnent selon la loi de Gibbs et qui sont en mesure de donner toutes les configurations possibles, les images obtenues par recuit simulé sont uniques et doivent en théorie correspondre aux minima globaux 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 $-Bx_s$. L'énergie totale s'écrit :

$$U(x) = - \sum_{c=(s,t) \in \mathcal{C}} \beta x_s x_t - \sum_{s \in S} B x_s$$

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.

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].

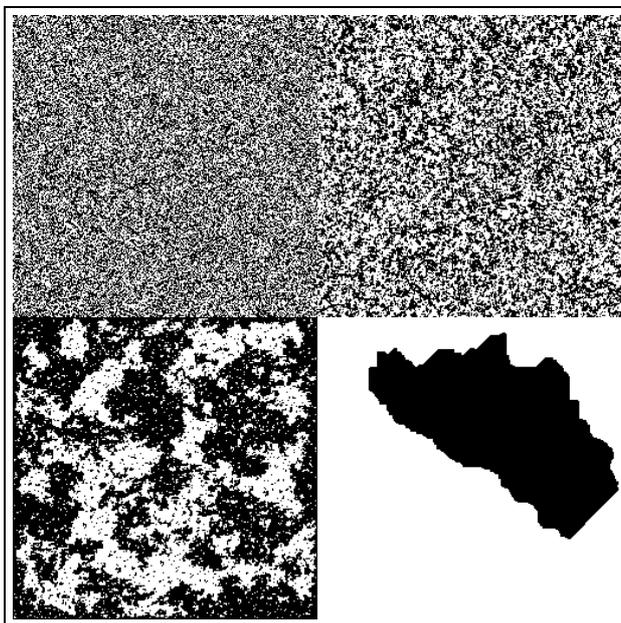


FIG. 7.2 – Le modèle d’Ising plan et 4-connexe pour différentes valeurs de paramètres - Simulations en raccordement torique ; a. Image aléatoire : $\beta = 0$, b.régularisation faible, $\beta = 0, 2$; c. régularisation critique : $\beta \approx 0, 44$; d. régularisation forte : $\beta = 4$.

◇ 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.

7.2 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 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

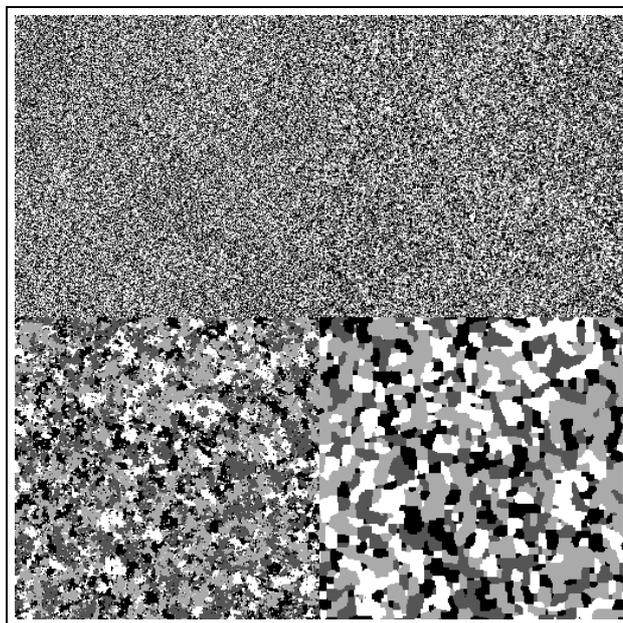


FIG. 7.3 – Le Modèle de Potts 2D et 4-connexe pour différentes valeurs de paramètres ($m = 4$) - Simulations en raccordement torique ; a. Image aléatoire : $\beta = 0$, b.régularisation faible, $\beta = 0.2$; c. régularisation critique : $\beta \approx 1,099$; d. régularisation forte : $\beta = 4$.

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 :

$$\mathcal{U}(x / y) = \sum_{s \in S} -\ln p(y_s / x_s) + \sum_{c \in \mathcal{C}} U_c(x) \quad (7.2)$$

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.

7.2.1 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 :

$$p(y_s / x_s) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x_s - y_s)^2}{2\sigma^2}$$

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 :

$$P(X = x) = \frac{1}{Z} \exp(-\beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t))$$

On a alors l'énergie suivante correspondant à la distribution de Gibbs du champ a posteriori :

$$\mathcal{U}(x / y) = \sum_{s \in S} \frac{(x_s - y_s)^2}{2\sigma^2} + \beta \sum_{(s,t) \in \mathcal{C}_2} \phi(x_s - x_t) \quad (7.3)$$

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.

7.2.2 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.



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.

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, où 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 $n r^2$ tend 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'indiquions 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 bruit 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.



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 $t = 0, 2$.

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 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 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

$$T(\mathbb{1}_{\chi_\lambda(f)}) = \mathbb{1}_{\chi_\lambda(T(f))}. \quad (8.6)$$

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 :

$$Hf = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

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 :

$$\frac{\partial u}{\partial t} = \text{div}(g(|\nabla u|)\nabla u), \quad (8.7)$$

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 :

$$\frac{\partial u}{\partial t} = |\nabla u| \text{curv}(u).$$

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 $\text{curv}(u)(\mathbf{a})$. 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 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 inva-

riantes 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.

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 :

$$\frac{\partial u}{\partial t} = \text{div} \left(\frac{\nabla u}{|\nabla u|} \right) + \alpha(f_0 - u),$$

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 :

$$\frac{\partial u}{\partial t} = \text{curv}(u).$$

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 5.3). 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 déviance standard connue. La restauration par Wiener se fait comme expliqué au paragraphe 5.3, 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 :

$$\frac{\partial u}{\partial t} = -\text{sign}(\Delta u)|\nabla u|, \quad (8.14)$$

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 filtrages itérés, filtrages que nous avons déjà rencontrés au paragraphe 8.1.2.

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 **récursive**, 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 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 tends vers 0, tends 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

$$T_{t,t+h}(0) = 0, \quad \text{et } T_{t,t+h}(u + C) = T_{t,t+h}(u) + C, \quad (8.19)$$

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].

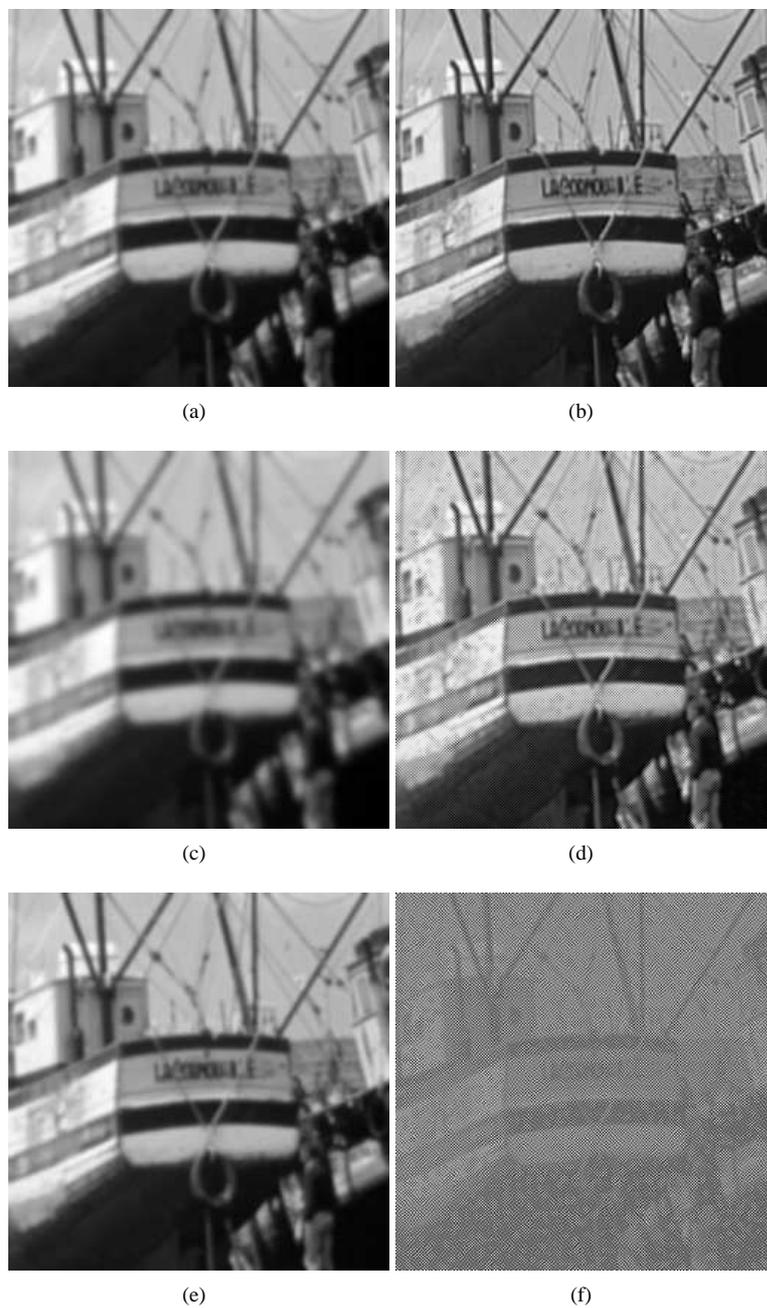


FIG. 8.4 – **(a)** : diffusion selon l'équation de la chaleur (8.1) pour $t = 1$; **(b)** : diffusion " inverse " selon (8.4) à partir de l'image (a), pour $t = 1$; **(c)** et **(d)** : même expérience, avec $t = 3$, on remarque un début d'instabilité ; **(e)** et **(f)** : même expérience avec $t = 1$, 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.

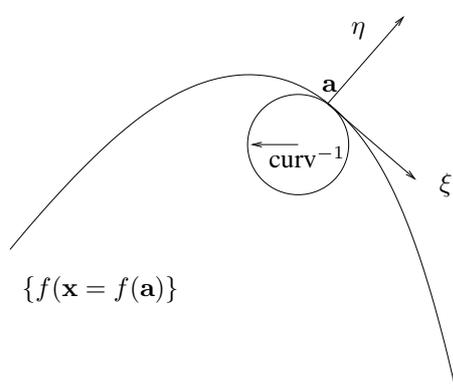


FIG. 8.5 –

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 : $t = 1, 3, 5, 10$.



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).



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).

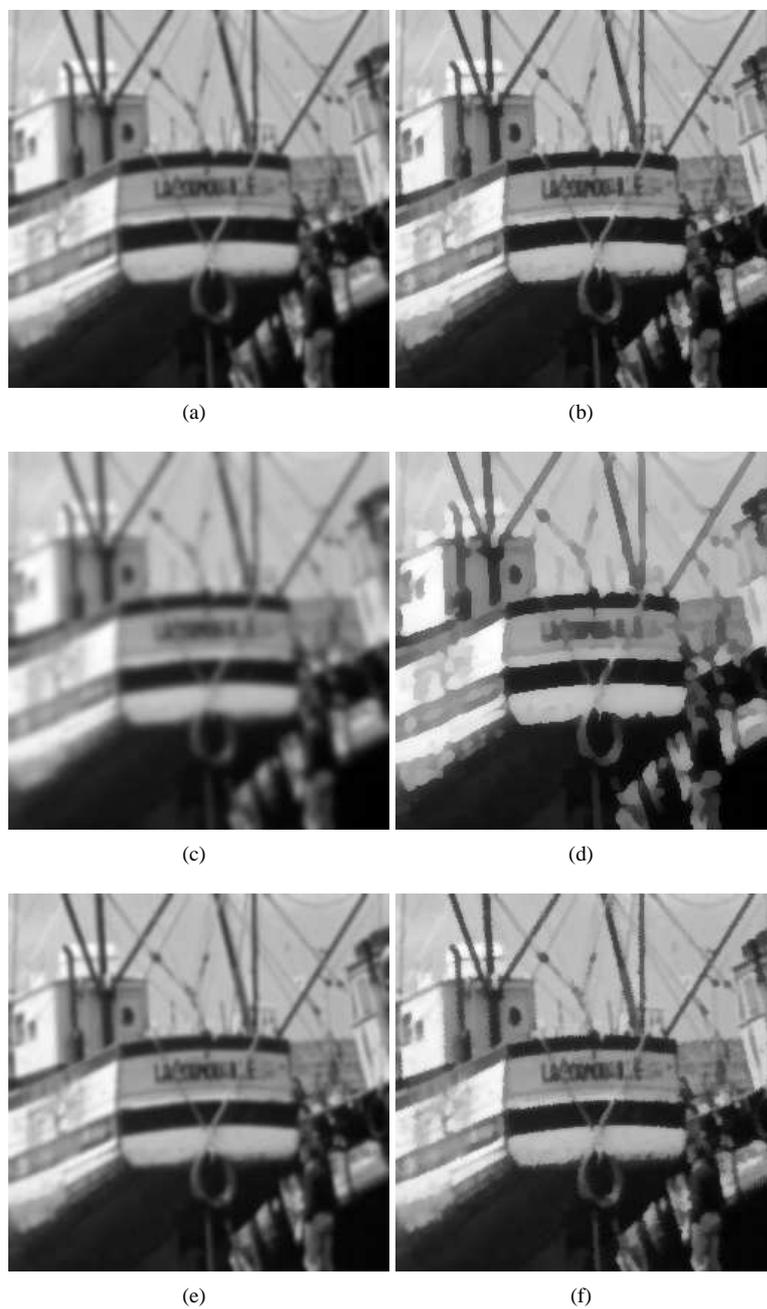


FIG. 8.10 – Filtre de choc de Rudin. Les images (a), (c), (e) sont identiques à celles de la figure 8.3. Les images (b), (d), (f) sont les applications correspondantes de l'équation (8.14) (respectivement $t = 1, t = 3, t = 1$).

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 :

$$\psi_\lambda(x, y) = g(x - x_0, y - y_0) e^{i 2\pi(\mu_x x + \mu_y y)}. \quad (9.2)$$

La TF de cette fonction est :

$$\Psi_\lambda(\nu_x, \nu_y) = G(\nu_x - \mu_x, \nu_y - \mu_y) e^{-i 2\pi[x_0(\nu_x - \mu_x) + y_0(\nu_y - \mu_y)]},$$

où G désigne celle de g (supposée définie). En choisissant g centrée et de type « passe-bas » (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 « ondelette mère » [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 :

$$\Psi_\lambda(\nu_x, \nu_y) = a \Psi^\theta(a\nu_x, a\nu_y) e^{-i 2\pi(x_0\nu_x + y_0\nu_y)},$$

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 « basse-fréquence » 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].

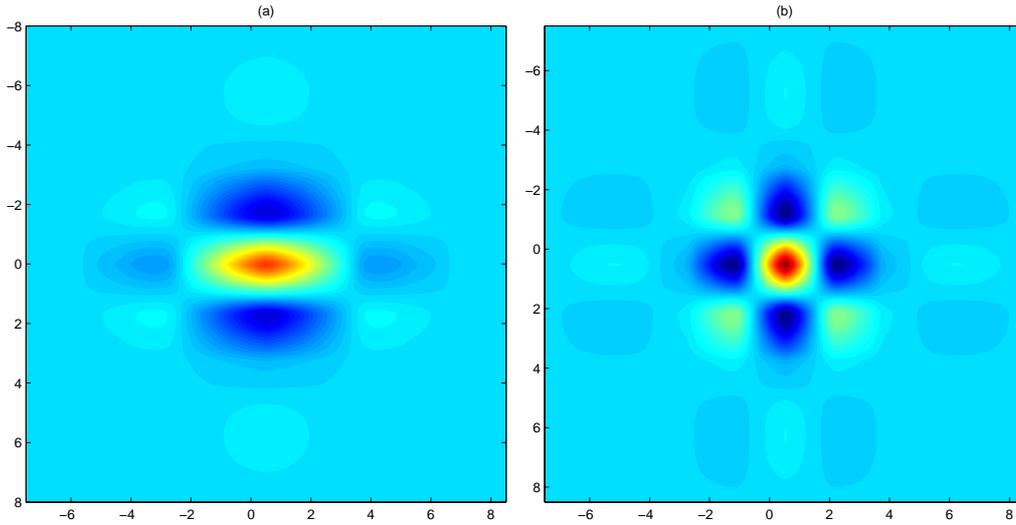


FIG. 9.1 – Exemple d'ondelettes séparables splines : (a) ψ^0 , (b) $\psi^{\pi/4}$. L'ondelette ψ^0 (resp. $\psi^{\pi/2}$) permet d'extraire les informations de type « contours » horizontaux (resp. verticaux) alors que $\psi^{\pi/4}$ sert à l'analyse des détails résiduels.

- *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

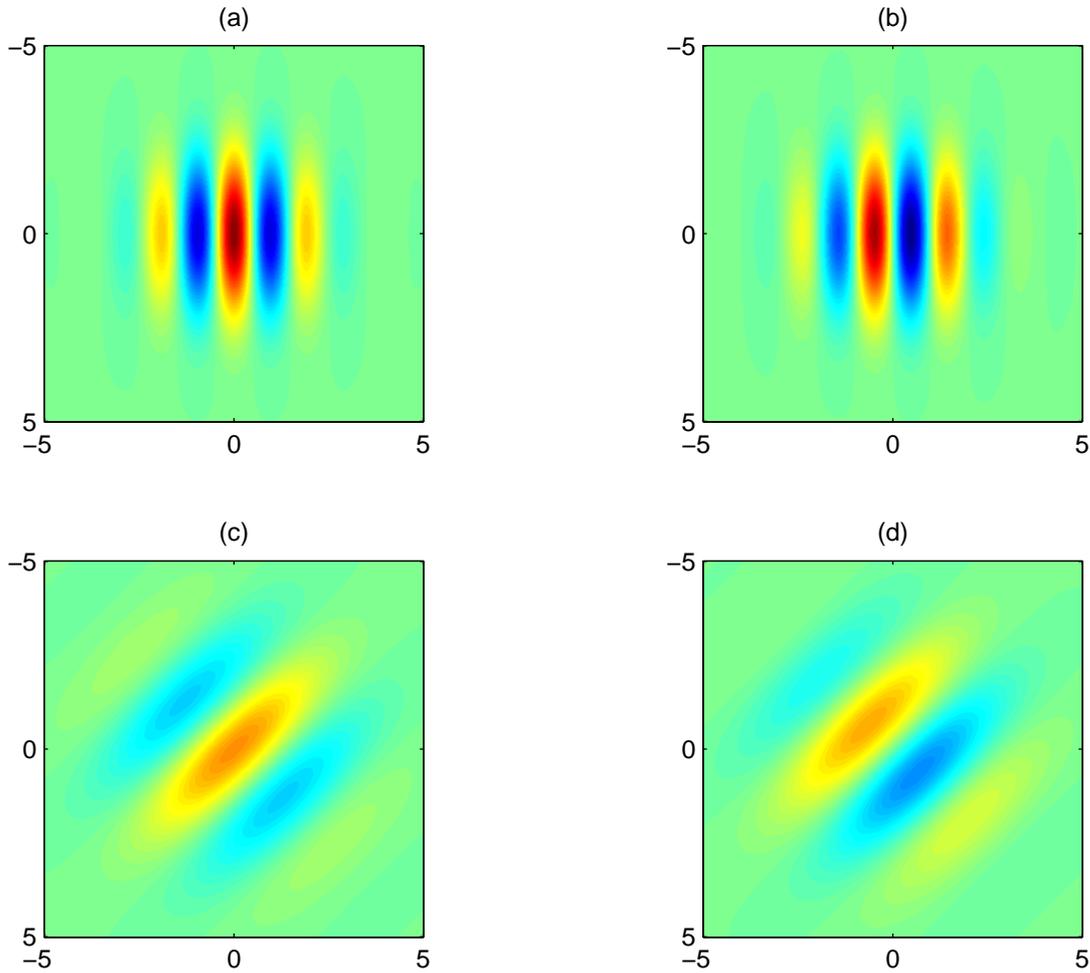


FIG. 9.2 – Exemple de fonctions de Gabor modifiées localisées en $(x_0, y_0) = (0, 0)$: atomes à l'échelle $a = 1$ pour $(\mu_x, \mu_y) = (\frac{1}{2}, 0)$ (orientation suivant l'axe des ordonnées) et (a) $\Theta = 0$ ou (b) $\Theta = \pi/2$; atomes à l'échelle $a = 2$ pour $(\mu_x, \mu_y) = \frac{1}{2\sqrt{2}}(1, 1)$ (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

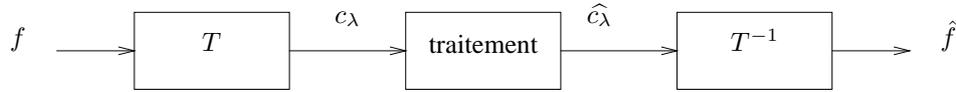


FIG. 9.3 – Traitement dans le domaine transformé (f : image d’origine, c_λ : coefficients transformés, \hat{c}_λ : coefficients traités, \hat{f} : image traitée).

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 « plus de coefficients $\langle s, \psi_i \rangle$ que cela n’est strictement nécessaire » !). 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 :

$$\forall f \in \mathcal{H}, \quad B^{-1} \|f\|^2 \leq \sum_{i \in I} \left| \langle f, \tilde{\psi}_i \rangle \right|^2 \leq A^{-1} \|f\|^2. \quad (9.9)$$

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 :

$$\langle (T^* T)^{-1} f, f \rangle = \sum_{i \in I} \left| \langle f, \tilde{\psi}_i \rangle \right|^2.$$

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 « base de Riesz », qui formalise cette notion de « non-redondance ».

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 :

$$\frac{1}{B} \|\psi_{i_0}\|^2 \leq \sum_{i \in I} \left| \langle \psi_{i_0}, \tilde{\psi}_i \rangle \right|^2 \leq \frac{1}{A} \|\psi_{i_0}\|^2 \iff \frac{1}{B} \leq 1 \leq \frac{1}{A} \iff A \leq 1 \leq B.$$

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

$$\widehat{f}_{j_0} = \sum_{\theta} \sum_{j=j_0}^{\infty} c_j^\theta[n, m] \widetilde{\psi}_{j,n,m,\theta},$$

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, \widehat{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

$$I = \left\{ i = (n\Delta, m\Delta, 2^j, \nu_0 \cos \varphi, \nu_0 \sin \varphi, \Theta), (n, m, j) \in \mathbb{Z}^3, \right. \\ \left. \varphi \in \left\{ -\frac{\pi}{2K}(K-1), \dots, \frac{\pi}{2K}(K-1), \frac{\pi}{2} \right\}, \Theta \in \left\{ 0, \frac{\pi}{2} \right\} \right\}, \quad (9.10)$$

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

$$I = \left\{ (n2^j, 2^j, 2\pi\ell2^j), n \in \mathbb{Z}, j \in \mathbb{Z}, j \leq j_M, \ell \in \{0, \dots, 2^{-j} - 1\} \right\},$$

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 « résidu »),

$$R_f^1 = f - \hat{f} = f - \langle f, \psi_{i_1} \rangle \psi_{i_1},$$

de façon à isoler le « second ingrédient essentiel », ψ_{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.

À 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].

¹On montre que cela se produit nécessairement au bout d'un nombre fini d'itérations.

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.



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.

Chapitre 10

Les prétraitements

Chapitre rédigé par Henri MAÎTRE

On regroupe souvent sous le terme de prétraitement toutes les opérations qui sont appliquées aux images, indépendamment de leur usage futur, pour leur assurer une bonne qualité. Elles concernent donc essentiellement les corrections de contraste et la suppression du bruit. Cette pratique est souvent critiquée car l'expérience montre qu'il est toujours important d'adapter les traitements de plus bas niveau aux traitements plus élaborés et donc aux objectifs à long terme du projet, mais les prétraitements sont souvent appliqués dans des circonstances où l'on destine les images à de nombreuses applications différentes dont on ignore souvent les besoins exacts.

10.1 Les traitements photométriques ou colorimétriques

10.1.1 Linéarité

Le défaut de non-linéarité apparaît si la valeur du signal d'image en un pixel n'est pas proportionnelle à l'énergie qu'il reçoit. La linéarité des capteurs d'image est très souvent médiocre. C'est particulièrement vrai pour les films et pour les tubes vidéo. Les capteurs solides (CCD par exemple) leur sont sur ce point très supérieurs [Marion, 1997]. On sait pourtant que cette linéarité est indispensable pour un grand nombre de problèmes, en particulier pour la restauration puisque c'est l'un des premiers prérequis des approches, qu'elles soient analytiques ou algébriques (cf. chapitre 5). Corriger une non-linéarité est une opération simple lorsqu'il n'y a pas de saturation du capteur. Cela revient à inverser cette non-linéarité et cela se fait par des tables de transcodage après calibrage du capteur (aussi appelé LUT : *Look-up Tables*). Il n'est pas possible de corriger une saturation sans information supplémentaire sur le signal.

10.1.2 Homogénéité

Ce défaut apparaît lorsque le capteur n'a pas une réponse égale en tous les plans de son champ. Le vignettage [Pérez, 1991] est un exemple de défaut d'homogénéité frappant les systèmes optiques et affectant les points du champ éloignés du centre de l'image. Mais il peut également provenir d'un éclairage inégal de la surface à analyser.

Avec des capteurs en ligne (comme on en trouve par exemple en télédétection ou sur les scanners de documents), les défauts d'homogénéité apparaissent sous forme de trainées. Les corrections d'homogénéité sont alors pratiquées par apprentissage des éléments de calibrage pour chaque cellule sensible, puis correction en associant une table de transcodage à chaque cellule. Si le capteur est linéaire (au sens de la section 10.1.1), le calibrage

revient à apprendre pour chaque cellule i le gain α_i et l'offset β_i qui fait passer du signal d'entrée $e(i)$ au pixel image $f(i)$:

$$f(i) = \alpha_i e(i) + \beta_i$$

Pour les capteurs bidimensionnels (films, matrices) les corrections rigoureuses sont très difficiles car il n'est pas possible de généraliser l'approche précédente en raison de la dimension beaucoup trop grande du problème et de la très grande difficulté d'en apprendre les trop nombreux paramètres. On préfère alors des approches souvent approchées : méthodes par morceaux (subdivision du champ) ou corrections globales paramétriques (par exemple approximation par des plans ou par des quadriques).

Homogénéisation par subdivision du champ

Une méthode souvent employée, en particulier pour corriger les défauts d'éclairage, procède par subdivision du champ de façon à rechercher des statistiques constantes dans le champ de l'image. Cette méthode fait donc l'hypothèse que la scène observée possède une homogénéité que l'image a perdu.

Les corrections par moyenne constante remplacent l'image $f(x, y)$ par une image corrigée $f'(x, y)$ qui vérifie :

$$f'(x, y) = f(x, y) - m(x, y) + M$$

avec m une moyenne locale calculée dans un voisinage \mathcal{V}_{xy} autour de (x, y) :

$$m(x, y) = \frac{1}{s} \int_{\mathcal{V}_{xy}} f(x', y') dx' dy'$$

et M la moyenne d'ensemble sur l'image complète \mathcal{I} :

$$M = \frac{1}{S} \int_{\mathcal{I}} f(x, y) dx dy$$

La surface de \mathcal{V}_{xy} est dénotée par s et celle de \mathcal{I} par S .

Les corrections par contraste constant créent une image f'' telle que :

$$f''(x, y) = \frac{V}{v} \cdot f(x, y)$$

où v et V sont les variances de f sur \mathcal{V}_{xy} et \mathcal{I} .

On peut imposer à l'image d'avoir une moyenne et un contraste constants, mais la formule n'est plus alors directe.

Homogénéisation par approximation par une quadrique

Une autre façon de rendre homogène une image est d'en soustraire une composante qui serait inégale dans l'image, par exemple en raison d'un éclairage peu uniforme. Si le signal d'intérêt est faible autour de la valeur moyenne du signal, fortement variable, on peut par exemple approcher celle-ci à l'aide d'un polynôme dépendant des variables d'espace. Voici comment on s'y prend lorsque le polynôme est une quadrique $Q(x, y)$ de paramètres c_i inconnus :

$$Q(x, y) = c_1 + c_2 x + c_3 y + c_4 xy + c_5 x^2 + c_6 y^2$$

On recherche le meilleur polynôme en minimisant, par rapport aux c_i , l'erreur ϵ :

$$\epsilon = \int_{\mathcal{I}} [f(x, y) - Q(x, y)]^2 dx dy$$

On obtient un système de taille 6×6 avec pour inconnues les c_i . Il se met sous la forme :

$$AC = M$$

où C est le vecteur des coefficients inconnus c_i , A est la matrice des moments de l'image (matrice symétrique mesurée pour chaque image) et M le vecteur des monômes en x et y .

10.1.3 Binarisation, seuillage

Ce sont des cas extrêmes de prétraitements qui ramènent l'image à deux ou à quelques niveaux seulement. Nous ne les évoquons ici que par souci d'exhaustivité. Ils sont traités plus complètement au chapitre 12 et dans les références [Sahoo et al., 1988, Cham et al., 1998].

10.1.4 Augmentation de contraste

Une famille de traitements a pour objectif de donner à l'image un plus grand contraste. Ces méthodes procèdent de trois façons différentes [Pratt, 1978, Zamperoni, 1996] :

1. Par étirement d'histogramme : on cherche à exploiter au mieux la dynamique de l'histogramme, tout d'abord en utilisant toute l'échelle de gris ou de couleurs disponible (par une table de transcodage), mais aussi en modifiant l'histogramme de façon à le transformer en un histogramme de référence. Les plus utilisés sont les histogrammes plats et hyperboliques (nous avons vu une telle transformation au chapitre 2) [Bovik, 2000].
2. Par filtrage passe-haut de l'image : l'idée de base est ici de réduire l'importance du terme continu et des basses fréquences. Un point délicat de ces méthodes est de définir le gabarit du filtre utilisé.
3. Par des méthodes locales : ces méthodes modifient localement l'histogramme pour conserver toujours une bonne dynamique, même dans des zones de fort contraste (par exemple Dorst). Un exemple de réhaussement de contraste local par des méthodes morphologiques est présenté au chapitre 6 sur la figure 6.5.

10.2 Suppression des bruits

C'est l'un des sujets les plus délicats du traitement des images. Il a vu couler beaucoup d'encre et de nombreuses méthodes lui ont été consacrées, tout d'abord très intuitives, mais progressivement de plus en plus complexes. Nous verrons tout d'abord les approches linéaires, puis les méthodes non-linéaires.

10.2.1 Filtrages linéaires

Dans de nombreux cas on considère que l'image non dégradée $f(x, y)$ est affectée d'un bruit additif $n(x, y)$:

$$g(x, y) = f(x, y) + n(x, y)$$

Le cas d'un bruit multiplicatif est beaucoup plus difficile. Il est par exemple traité dans le cas de l'imagerie radar [Maître, 2001].

Dans les conditions les plus simples, on suppose le signal stationnaire, et on adopte donc souvent une approche par filtrage linéaire. Intuitivement, et en l'absence de toute autre information, on recherche des filtres passe-bas, éliminant les variations à très haute fréquence du signal.

Expérimentalement ces filtres ne sont pas très bons (cf. figure 10.1) car ils dégradent considérablement les contours et rendent l'image floue, ce qui est subjectivement désagréable. Une seule exception : si le signal est effectivement stationnaire (et donc sans contours) et si le bruit est gaussien (dans ce cas ce sont les meilleurs filtres au sens des moindres carrés et du maximum de vraisemblance). Les plus utilisés sont le filtre de moyenne (sur de petites fenêtres de 3×3 ou 5×5 pixels) et le filtre gaussien.

Complexité des filtres linéaires

La complexité des filtres linéaires est généralement en $\mathcal{O}(\nu^2 N)$, si N est le nombre de pixels dans l'image et ν le côté de la fenêtre carrée sur laquelle est faite le filtrage. Leur coût grandit donc très vite avec l'étendue du filtre et cela prohibe souvent les filtres larges.



FIG. 10.1 – **Filtres de moyenne** - A gauche, image bruitée par un bruit gaussien de variance 120, au centre, filtrage passe-bas sur une fenêtre 3×3 , à droite sur une fenêtre 7×7 .

Il existe cependant des exceptions notables :

- Le filtre de moyenne a une complexité indépendante de la taille de la fenêtre (si l'on garde en mémoire les traitements appliqués aux points et aux lignes précédentes).
- Le filtre gaussien est séparable, il peut être obtenu par 2 filtrages monodimensionnels. Sa complexité est donc $\mathcal{O}(\nu N)$.

On utilise aussi la propriété des filtres gaussiens d'être à noyau auto-reproducteur (on obtient une nouvelle gaussienne d'écart-type fort, par convolution de deux gaussiennes d'écart-type faible) en filtrant successivement par des gaussiennes étroites pour obtenir des gaussiennes larges. On a aussi beaucoup utilisé des approximations splines des fonctions gaussiennes, plus simples à calculer.

Le filtre de Wiener

Dans l'approche de Wiener (cf. chapitre 5, section 5.3), on recherche le filtre $h(x, y)$ qui minimise :

$$\left\langle \int_{\mathcal{I}} (\hat{f} - f)^2 dx dy \right\rangle \quad (10.1)$$

où l'espérance $\langle . \rangle$ est prise sur toutes les réalisations du bruit, le domaine d'intégration \mathcal{I} est l'image complète et $\hat{f} = f * h$.

Si on connaît le spectre de densité de puissance du bruit $P_n(u, v)$ et le spectre de densité de puissance de l'image $P_f(u, v)$, le filtrage optimal au sens des moindres carrés est le filtre adapté. C'est le filtre de Wiener avec un défaut égal à l'identité. Sa fonction de transfert (TF de sa réponse impulsionnelle) a pour expression :

$$H(u, v) = \left[1 + \frac{P_n(u, v)}{P_f(u, v)} \right]^{-1}$$

Lorsque l'on ne connaît pas ces spectres de densité de puissance, on adopte souvent, pour ces grandeurs, les expressions (cf. chapitre 2) :

$$P_f(u, v) = \frac{1}{1 + \alpha u^2} \frac{1}{1 + \alpha v^2} \quad \text{et} \quad P_n(u, v) = N^2$$

Malheureusement, compte tenu des difficultés à estimer les paramètres exacts du signal et du bruit, ce filtre conduit en pratique à des filtrages de bruit souvent médiocres.

10.2.2 Filtrage de rang

Ces approches reposent sur un raisonnement très différent. On considère que le signal vrai doit être estimé à partir d'échantillons bruités. Ces échantillons appartiennent à une fenêtre \mathcal{V}_{xy} de $n \times n$ pixels entourant le pixel à filtrer. Si la fenêtre est entièrement à l'intérieur d'une zone homogène, on est ramené au problème traité ci-dessus. Si la fenêtre est à cheval sur des champs de statistiques différentes, on doit trouver un **estimateur robuste** qui non seulement écartera le bruit, mais supprimera également l'influence du signal parasite (différence entre « bruit » et « out-lyers »). Les filtres d'ordre ou de rang sont de tels filtres [Kotropoulos et al., 1997, Tagare et Figueiredo, 1985, Kim et Yaroslavski, 1986, Pitas et Venetsanopoulos, 1990, Arce et al., 2000].

On trie les valeurs des pixels dans la fenêtre \mathcal{V}_{xy} par amplitude croissante, on remplace le pixel central par un pixel choisi dans la liste ordonnée :

- si on choisit le plus grand : dilatation en morphologie mathématique (cf. chapitre 6),
- si on choisit le plus petit : érosion en morphologie mathématique,
- si on choisit celui du milieu : filtre médian.

Les avantages des filtres de rang sont les suivants :

- ils ne sont pas sensibles aux valeurs extrêmes qui n'influencent donc pas le filtrage (cela ne concerne bien sûr pas les érosions et les dilatations) ;
- ils attribuent au signal filtré une valeur appartenant à l'ensemble des valeurs du signal bruité et donc ne créent pas de valeur nouvelle (à ce titre, ils permettent de filtrer des signaux portant une information plus *symbolique* qu'un simple niveau de gris, par exemple des images de classes obtenues après une étape de segmentation).

On a développé des filtres de rang plus complexes, par exemple le filtre d_α qui remplace la valeur courante par la valeur qui minimise :

$$d_\alpha = \left[\sum_{i \in \mathcal{V}} |\hat{f} - f_i|^\alpha \right]^{1/\alpha}$$

où $\hat{f} \in \{f_i\}$ et \mathcal{V} est la fenêtre d'estimation. Lorsque α est faible, on donne plus de poids aux valeurs f_i proches de \hat{f} tandis que les grandes valeurs de α donnent plus de poids aux écarts importants.

Ces filtres de rang donnent des résultats souvent satisfaisants dans les cas simples d'élimination du bruit. Ainsi, le filtre médian est pratiquement toujours préféré au filtre de moyenne ou au filtre gaussien pour améliorer les images bruitées.

10.2.3 Filtrages morphologiques

On a déjà vu que les érosions et les dilatations sont des filtres de rang (mais pas des filtres au sens morphologique), mais on leur préfère les ouvertures et les fermetures pour leur propriété d'idempotence (cf. chapitre 6). Pour des filtres de taille croissante on prend des filtres alternés séquentiels (FAS) (cf. figure 10.2) [Schmitt et Mattioli, 1994b, Serra, 1982b].

Enfin, on fait également des filtres d'augmentation de contraste en répartissant judicieusement le niveau de sortie entre l'érodé, le dilaté ou la moyenne selon la distance du pixel à la valeur moyenne (on peut remplacer dans la phrase précédente *moyenne* par *médiane*).

10.2.4 Filtrages par équations de diffusion

Ces opérations ont fait l'objet du chapitre 8 où l'on a mis en évidence leur rôle de filtrage aussi bien pour l'analyse que pour la restauration. Ce sont ces propriétés qui nous intéresseront ici encore. Mais rappelons tout d'abord que les équations de diffusion constituent aussi l'approche duale de la minimisation d'une intégrale totale sur l'image. Dans cette approche, à une équation intégrale d'énergie comme l'équation 10.1, on associe une



FIG. 10.2 – **Filtre Alterné Séquentiel** - Une, deux ou trois itérations de la séquence ouverture puis fermeture, d'éléments structurants croissants.

équation d'Euler-Lagrange et des conditions aux limites (généralement les conditions de Neumann). On résout alors l'équation différentielle par des techniques de gradient (résolution itérative), à partir d'une solution initiale généralement égale à l'image bruitée. S'il y a convergence, c'est alors vers un minimum local.

Diffusion isotrope

On a vu au chapitre 8 que le filtrage par des équations de diffusion isotrope est, dans le cas continu, mathématiquement équivalent au filtrage gaussien [Marr et Hildreth, 1980, Babaud et al., 1986, Witkin, 1983a, Koenderink, 1984]. Il est obtenu à partir de l'équation de la chaleur :

$\frac{\partial f}{\partial t} = \Delta f$ où l'on remplace le temps par des itérations successives. Ce filtrage, comme le filtrage gaussien, a le défaut de rendre les contours d'images très flous. Ce n'est donc pas un bon filtre d'amélioration d'images.

Diffusion anisotrope

On a montré au chapitre 8 que l'on préfère les équations de diffusion anisotrope [Perona et Malik, 1990b, Alvarez, 1996]. Leur équation est :

$$\frac{\partial f}{\partial t} = \text{div}[g(f, t) \cdot \nabla f] \quad (10.2)$$

où $g(f, t)$ est une fonction généralement du module de la dérivée de f , maximale pour une dérivée nulle et décroissante lorsque le gradient croît : on choisit généralement les fonctions suivantes (cf. figure 10.3) :

$$\begin{aligned} - g(f, t) &= \exp\left(-\frac{|\nabla f|^2}{\kappa^2}\right), \\ - g(f, t) &= \left[1 + \left[\frac{|\nabla f|}{\kappa}\right]^{1+\alpha}\right]^{-1}. \end{aligned}$$

Le coefficient κ est proportionnel à un gradient. Il caractérise l'amplitude des gradients qui autoriseront une forte diffusion. Pour des valeurs de gradient très inférieures à κ , la diffusion sera très faible car la zone sera considérée comme déjà lisse. Pour des valeurs très supérieures, la diffusion sera interdite (on considérera qu'il y a présence d'un contour que le filtrage devra préserver). Cela se voit sur la figure 10.3 où l'on a porté l'évolution du produit $g(f, t) \cdot |\nabla f|$ en fonction de $|\nabla f|$.

La mise en œuvre des filtres de diffusion anisotrope se fait en calculant une série d'images f^n telles que :

$$f^n = f^{n-1} + \frac{\partial}{\partial x} \left[g(f, n) \cdot \frac{\partial f^{n-1}}{\partial x} \right] + \frac{\partial}{\partial y} \left[g(f, n) \cdot \frac{\partial f^{n-1}}{\partial y} \right]$$

Deux étapes de ce processus sont présentées sur la figure 10.4.

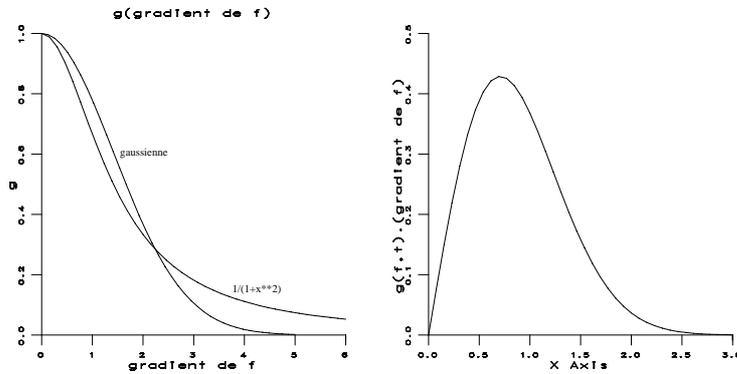


FIG. 10.3 – A gauche, terme contrôlant la diffusion dans les techniques de diffusion anisotropique pour deux exemples de filtres utilisés : la gaussienne $\exp(-\frac{|\nabla f|^2}{\kappa^2})$ ou la lorentzienne : $[1 + \frac{|\nabla f|^2}{\kappa^2}]^{-1}$. À droite, pour le cas de la gaussienne, variation du produit $g(f, t)|\nabla f|$ en fonction de $|\nabla f|$. Le maximum est proche de $|\nabla f| = \kappa$.



FIG. 10.4 – **Filtrage par Diffusion Anisotrope** : image originale bruitée (à gauche) et deux filtrages par 15 et 80 itérations avec $\kappa = 25$.

Filtres de courbure moyenne

On a également vu au chapitre 8 que plusieurs autres formes avaient été proposées pour les filtres par E.D.P. pour garantir la conservation de certaines propriétés des images. D'un intérêt particulier sont les filtres qui garantissent une invariance du traitement par changement de contraste. C'est le cas des filtres de courbure moyenne (ou *mean curvature equation*) [Kimia et al., 1992] :

$$\frac{\partial f}{\partial t} = \text{div}\left[\frac{\nabla f}{|\nabla f|}\right] \cdot \nabla f \tag{10.3}$$

Filtrage à Variation Totale

Dans la plupart des approches de filtrage, on recherche la meilleure solution au sens de l'erreur quadratique moyenne. Ce critère n'est pas très satisfaisant pour un observateur humain qui préfère souvent des optimisations sous d'autres normes. On préfère parfois la norme L^1 , mais celle ci conduit à des solutions souvent difficiles à mettre en œuvre. Cet argument a été poursuivi dans [Rudin et al., 1992b] pour proposer un filtrage dans la classe des signaux à variation totale bornée (ou signaux **BV**), sous-classe des signaux **TV**, c'est-à-dire à contrainte sur la variation totale [Rudin et Osher, 1994].

Ainsi, sous ce formalisme, on minimise :

$$\int_{\Omega} \hat{f}_x^2 + \hat{f}_y^2 dx dy$$

où \hat{f}_x est la dérivée en x de l'estimée de l'image, sous les contraintes :

$$\int_{\Omega} \hat{f} dx dy = \int_{\Omega} f dx dy \quad \text{et} \quad \int_{\Omega} (\hat{f} - f)^2 dx dy = \sigma^2$$

Ces deux contraintes sont des hypothèses sur le bruit (valeur moyenne nulle et écart-type fixé).

Ces relations conduisent aux équations d'Euler-Lagrange :

$$\frac{\partial}{\partial x} \left(\frac{\hat{f}_x}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} \right) + \frac{\partial}{\partial y} \left(\frac{\hat{f}_y}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} \right) - \lambda_1 - \lambda_2(\hat{f} - f) = 0$$

avec, comme condition aux limites $\frac{\partial \hat{f}}{\partial n} = 0$. On la résout itérativement par une descente de gradient :

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x} \left(\frac{\hat{f}_x}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} \right) + \frac{\partial}{\partial y} \left(\frac{\hat{f}_y}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} \right) - \lambda(\hat{f} - f) \quad (10.4)$$

avec une valeur de λ donnée par :

$$\lambda = -\frac{1}{2\sigma^2} \int \left[\sqrt{\hat{f}_x^2 + \hat{f}_y^2} - \left(\frac{f_x \hat{f}_x}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} + \frac{f_y \hat{f}_y}{\sqrt{\hat{f}_x^2 + \hat{f}_y^2}} \right) \right] dx dy$$

Selon cette méthode, on filtre progressivement l'image d'origine, tout en respectant ses contours. Le schéma donne des résultats assez semblables à ceux du filtre de courbure moyenne. On s'en convaincra en notant l'identité des équations 10.3 et 10.4, dans le cas où $\lambda = 0$ et le membre de droite est multiplié par $|\nabla f|$.

10.3 Les filtres adaptatifs

10.3.1 Les filtres à coefficients adaptatifs

Ces filtres s'inspirent des filtres de moyenne, mais chaque terme de la moyenne est pondéré par un coefficient qui décroît avec la similarité entre le pixel considéré et le pixel central de la fenêtre. On peut également interpréter ces deux filtres comme des filtres de diffusion isotrope mais adaptative.

Filtre de gradient inverse

Il est appelé également filtre σ ; il calcule une moyenne des pixels de la fenêtre, pondérés par l'inverse de leur gradient, de façon que les points trop différents du point traités ne contribuent que peu :

$$f'(i, j) = \frac{1}{\alpha} \sum_k \sum_l \frac{f(i+k, j+l)}{|f(i, j) - f(i+k, j+l)| + 1}$$

C'est un filtre d'action très modérée que l'on itère généralement plusieurs fois.

Filtre de Saint Marc

Il repose sur un principe assez semblable, mais la pondération est exponentielle négative :

$$f^{k+1}(i, j) = \frac{1}{\alpha} \sum_l \sum_{l'} f^k(i + l, j + l') \exp(-\beta |\nabla f^k|)$$

Son action est plus nette que celle du précédent. Il conduit à de bons filtrages respectant bien les contours [Saint-Marc et al., 1989].

Toboggan de Fairfield

Il agit de façon très différente des précédents. Il creuse l’histogramme de l’image en attribuant à chaque pixel le niveau de gris du point de minimum de gradient dans son voisinage. On trace tout d’abord une carte du gradient. Puis, à partir d’un point non déjà parcouru, on descend sur cette carte comme sur un toboggan (d’où le nom) jusqu’à ce qu’on tombe dans un minimum local du gradient. On attribue alors la valeur du niveau de gris du point minimum local. Tous les points du chemin parcouru sont mis au même niveau, et tous ces points sont marqués afin de n’être plus traités. On répète ce traitement jusqu’à ce que tous les points aient été parcourus [Fairfield, 1990].

10.3.2 Les filtres à fenêtres adaptatives

Dans le cas de filtres à fenêtres adaptatives, on recherche, autour de chaque pixel, la fenêtre la plus adaptée au filtrage. Cela se fait de deux façons :

1. soit en sélectionnant parmi une famille de fenêtres celle qui convient le mieux (c’est le cas du filtre de Nagao),
2. soit en faisant croître une fenêtre et en contrôlant la croissance (cas du filtre de Wu).

Filtre de Nagao

Dans l’approche de Nagao [Natsuyama, 1979], on travaille sur une fenêtre de taille 5 × 5 entourant le pixel central (qui appartient à toutes ces fenêtres). On définit 9 fenêtres possibles (cf. figure 10.5), toutes de 9 pixels, identifiées par un indice *k*. Sur chaque fenêtre on mesure la moyenne *m_k* et la variance *σ_k²* des niveaux de gris. On choisit alors de remplacer le pixel central par la valeur moyenne de la fenêtre dont la variance est la plus faible.

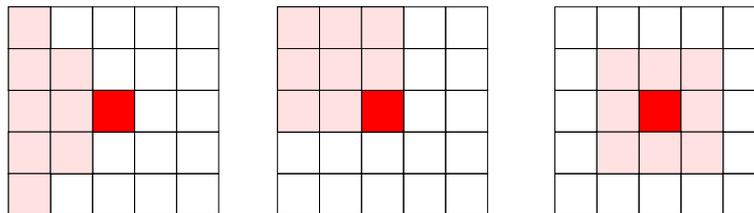


FIG. 10.5 – Le pixel central d’une fenêtre 5 × 5 appartient à 9 fenêtres de 9 pixels chacune : 4 se déduisent de la fenêtre représentée à gauche par rotations de *n* × 90°, 4 de la fenêtre au centre et la fenêtre de droite.

Filtre de Wu : filtrage par fenêtre maximale

Il procède par croissance de région [Wu et Maître, 1992]. Il utilise un prédicat d’homogénéité sur une région (cf. chapitre 12, section 12.2), c’est souvent une mesure de la variance (ce peut être le coefficient de variation

pour une image de radar). On part d'une zone de taille 3×3 , si elle est homogène on l'étend à une fenêtre 5×5 puis 7×7 en lui ajoutant des couronnes successives. Si le prédicat n'est plus vérifié, on recherche la cause de l'hétérogénéité. Si elle n'est présente que sur un côté de la fenêtre, on procède à une croissance sur une fenêtre rectangulaire en interdisant de croître sur le côté hétérogène. Si 2 côtés adjacents sont hétéromogènes, on croît selon un triangle. Lorsque la fenêtre ne peut plus croître, on remplace le point initial par la moyenne sur la plus grande fenêtre obtenue (cf. figure 10.6).

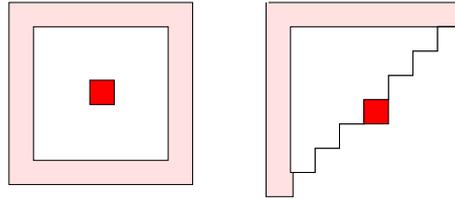


FIG. 10.6 – Méthode de Wu. A gauche, passage d'une fenêtre à la suivante dans le cas d'une croissance isotrope. A droite, s'il y a un contour dans le 1/2 plan inférieur droit, la croissance ne se fait plus que selon un triangle.

10.4 Le rééchantillonnage des images

Cette opération est nécessaire chaque fois que l'on doit transformer une image dans une géométrie prédéfinie (par exemple en télédétection pour se ramener dans une référence cartographique, ou en imagerie médicale pour aligner les images sur les repères anatomiques). Ces opérations de rééchantillonnage s'appuient sur les bases mathématiques de la théorie de l'échantillonnage de Shannon (cf. chapitre 3).

L'idée de principe est assez simple : considérant que l'image d'origine vérifie le théorème de Shannon, on reconstruit tout d'abord le signal continu dont cette image discrète est issue, puis cette image continue est filtrée de façon à donner une image dont le spectre est compatible avec le nouvel échantillonnage que l'on souhaite réaliser, enfin on échantillonne cette image aux points que l'on désire connaître.

Mais cette idée de principe se heurte à des contraintes de mise en œuvre matérielle très difficiles (temps de calcul, place mémoire, etc.) qui nécessitent que l'on trouve des solutions de remplacement.

Cela se fait en combinant reconstruction et filtrage en un seul filtrage. Cela se fait également très souvent en remplaçant les fonctions canoniques de reconstruction (généralement des sinus cardinaux) par des fonctions plus compactes, souvent polynomiales, permettant d'effectuer des sommes finies et non infinies.

10.4.1 Interpolation avec des polynômes

Dans ces approches, on convole la fonction initiale $f(x)$ ¹ par un noyau $h(x)$:

$$g(x) = f(x) * h(x) = \int_{-\infty}^{\infty} f(y)h(x-y)dy$$

ce qui s'exprime dans l'espace de Fourier par :

$$G(u) = F(u)H(u)$$

F et H étant les TF de f et h .

¹Nous la considérons ici monodimensionnelle pour simplifier, pour les fonctions bidimensionnelles il convient généralement de faire le produit d'un polynôme en x par un polynôme en y .

L'interpolation idéale est obtenue pour :

$$h(x) = \frac{\sin(2\pi x)}{x} = \text{sinc } \pi x \quad ; \quad H(u) = \text{rect}(u)$$

Mais d'autres interpolations, plus rapides, sont souvent utilisées :

- Interpolation au plus proche voisin :

$$h(x) = \text{rect}(x) \quad ; \quad H(u) = \text{sinc } \pi u.$$

- Interpolation linéaire :

$$h(x) = (1 - |x|)\text{rect}(2x) \quad ; \quad H(u) = \text{sinc}^2 \pi u.$$

- Interpolation quadratique :

Plusieurs formules ont été proposées. Celle de [Dodgson, 1997] est particulièrement utilisée :

$$\begin{aligned} h(x) &= -2x^2 + 1 & \text{si} & \quad |x| \leq 1/2 \\ h(x) &= x^2 - 5/2|x| + 3/2 & \text{si} & \quad 1/2 \leq |x| \leq 3/2 \\ h(x) &= 0 & \text{sinon} & \end{aligned}$$

dont la fonction de transfert est :

$$H(u) = \frac{6 \sin \pi u - 6 \sin 3\pi u - \cos \pi u + \cos 3\pi u}{(2\pi u)^2}$$

- Interpolation cubique :

Les solutions possibles sont encore plus nombreuses et ont fait l'objet de très nombreux travaux [Maeland, 1988]. On retiendra la famille paramétrée par la variable a :

$$\begin{aligned} h(x) &= (a + 2)|x|^3 - (a + 3)x^2 + 1 & \text{si} & \quad |x| \leq 1 \\ h(x) &= a(|x|^3 - 5x^2 + 8|x| - 4) & \text{si} & \quad 1 \leq |x| \leq 2 \\ h(x) &= 0 & \text{sinon} & \end{aligned}$$

La valeur $a = -1/2$ est souvent retenue. Elle conduit à une interpolation dont la fonction de transfert est :

$$H(u) = \frac{18 - 24 \cos 2\pi u + 6 \cos 4\pi u - \pi u(2 \sin 2\pi u - \sin 4\pi u)}{(2\pi u)^4}$$

10.4.2 Interpolation par des B-splines

Les B-splines forment également une famille de polynômes intéressants [Unser et al., 1993a, Unser, 1999, Chalmond, 2000]. Ils sont de degré k et se déduisent récursivement du polynôme de degré 0 par la formule :

$$\beta^k(x) = *^{k+1} \beta^0(x)$$

où $*^k$ exprime la $k^{\text{ème}}$ convolution. Ainsi :

$$\begin{aligned} \beta^1(x) &= \beta^0(x) * \beta^0(x) \\ \beta^2(x) &= \beta^0(x) * \beta^0(x) * \beta^0(x) \\ &\dots \text{etc.} \end{aligned}$$

La fonction d'ordre 0 est définie par la fonction porte : $\beta^0(x) = \text{rect } x$, et les diverses fonctions sont explicitées dans [Unser et al., 1993b] jusqu'à l'ordre 7. Leurs TF s'expriment simplement par :

$$TF[\beta^k] = (\text{sinc } u/2)^{k+1}$$

Une représentation par les splines d'ordre n prend la forme générale :

$$s(x) = \sum_{k \in \mathbb{Z}} c(k) \beta^n(x - k) \quad (10.5)$$

Lorsqu'on interpole un signal dans l'approche par splines, on choisit généralement des splines **interpolantes**, qui passent exactement par tous les points de l'échantillonnage, et non des splines **approximantes** qui passent au plus près d'un ensemble de points. Calculer une telle approximation consiste donc à déterminer les coefficients inconnus c_k tels que en tout point x appartenant aux échantillons initiaux on ait :

$$\sum_{k \in \mathbb{Z}} c(k) \beta^n(x - k)|_{x=k} = s(k) \quad (10.6)$$

les $s(k)$ étant les valeurs connues de l'échantillonnage.

On peut calculer les approximations par splines de deux façons :

1. par résolution du système d'équations 10.6 liant les coefficients inconnus des splines aux coordonnées et aux amplitudes connues des échantillons ; on est alors amené à inverser une matrice Toeplitz (cf. chapitre 5) (invariante par translation le long de x). Les coefficients de la matrice sont les valeurs des β^n aux nœuds du maillage, ils sont au nombre de $n + 1$. Les inconnues, les $c(k)$, sont plus nombreuses que les équations (on ne dispose pas d'équations complètes pour les $n/2 - 1$ premiers et derniers points de l'échantillonnage). On choisit donc généralement de répéter les premières et les dernières valeurs, ou on les reproduit selon une symétrie miroir.
2. par filtrage (c'est-à-dire par convolution) [Goshtasby et al., 1990] :
Pour cela, utilisons le noyau b_m^n obtenu en échantillonnant la B-spline de degré n dilatée d'un facteur m et associons lui sa transformée en z :

$$b_m^n(k) = \beta(x/m)|_{x=k} \rightarrow B_m^n(z) = \sum_{k \in \mathbb{Z}} b_m^n(z)(k) z^{-k}$$

Nous voulons déterminer les coefficients $c(k)$ des splines qui garantissent l'égalité aux échantillons entiers :

$$\sum_{l \in \mathbb{Z}} c(l) \beta^n(x - l)|_{x=k} = s(k)$$

ce que l'on peut réécrire comme une convolution :

$$s(k) = b_1^n * c$$

Si l'on définit par $\gamma(k)$ la fonction dont la transformée en Z (TZ) est l'inverse de $B_1^n(z)$:

$$\gamma(k) \rightarrow \frac{1}{B_1^n(z)}$$

on obtient une solution du problème par :

$$c(k) = \gamma(k) * s(k)$$

On montre que cette déconvolution est stable et peut être effectuée très rapidement [Unser et al., 1993a, Unser et al., 1993b].

À deux dimensions, l'interpolation spline se déduit du cas monodimensionnel par produit tensoriel de l'interpolation en x et en y (équation 10.5) :

$$s(x, y) = \sum_{k \in Z} \sum_{k' \in Z} c(k, k') \beta^n(x - k) \beta^n(y - k') \quad (10.7)$$

10.4.3 Interpolation adaptative

Des méthodes complètement différentes des précédentes effectuent des interpolations adaptatives, dans le but de préserver les objets présents dans la scène et leurs contours en particulier. Ces techniques contrôlent souvent l'interpolation par une détection de contours ou une segmentation de l'image. Un exemple de telles techniques est donné par [Karabassis et Spetsakis, 1995].

D'autres méthodes [Calle et Montanvert, 1998] s'appuient sur des décompositions fractales qui permettent de retrouver les propriétés fines - filtrées par l'échantillonnage original - mais présentes à travers la pyramide d'homothétie interne (cf. chapitre 3).

Chapitre 11

La détection des contours dans les images

Chapitre rédigé par Henri MAÎTRE

La détection de contours dans les images a débuté de façon extrêmement empirique par des opérateurs locaux qui, soit estimaient un gradient, soit convoluait l'image par des masques caractéristiques des contours [Haralick et Shapiro, 1985]. Dans les années 80, des approches plus systématiques ont été mises en place par Marr [Marr et Hildreth, 1980], puis Canny [Canny, 1986], pour obtenir des contours plus significatifs. Ces travaux ont abouti à une bonne compréhension de ce qu'il faut faire pour détecter les contours, mais la définition même des contours demeure très vague, ce qui rend ces techniques encore peu efficaces sur un problème concret. De plus, on a pu montrer que le problème de détection de contours est généralement mal posé (au sens de la résolution des systèmes) [Torre et Poggio, 1986]. Les seuls modèles de contours utilisables sont ceux de contours idéalisés, comme ceux présentés sur la figure 11.1 ; ils sont bien loin de la réalité. C'est pourquoi, même si de très gros progrès ont été accomplis dans ce domaine, les techniques empiriques d'estimation du gradient proposées dans les années 70-80 restent souvent encore employées en concurrence de techniques plus modernes. Une excellente référence à ce problème est l'ouvrage collectif [Cocquerez et Philipp, 1996]. Plusieurs articles existent sur la comparaison des performances de détecteurs de contours, en particulier [Palmer et al., 1996, Heath et al., 1997].

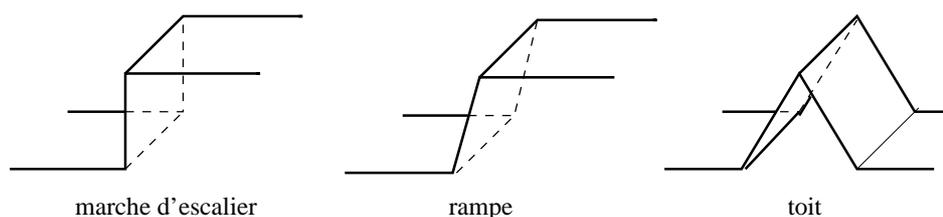


FIG. 11.1 – Quelques modèles de contours. Le plus utilisé est celui en marche d'escalier.

Nous présenterons tout d'abord une approche formelle (mais malheureusement stérile !) de la détection des contours, puis les méthodes empiriques encore utilisées seront vues dans la partie 2. Une approche plus analytique, proposée par Canny, sera ensuite présentée, avec ses dérivées qui sont aujourd'hui les plus employées en raison de leur efficacité. Enfin les techniques fondées sur le principe des **contours actifs** de Kass, Witkins et Terzopoulos seront présentées, ainsi que les approches par **ensembles de niveaux** (*level sets*). Nous terminerons par une brève présentation des méthodes de poursuite et de fermeture des contours.

11.1 Un modèle continu de contour.

Dans une image supposée continue $f(x, y)$, un contour apparaît comme une ligne où sont localisées les très fortes variations de f . Soit \vec{G} le gradient de f :

$$\vec{G} = \vec{\nabla}f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

On associe à f une image du module du gradient de f :

$$G = |\vec{\nabla}f| = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2} \quad (11.1)$$

ainsi qu'une image des directions du gradient à partir de :

$$\vec{g} = \frac{\vec{\nabla}f}{|\vec{\nabla}f|}$$

et par application de $[-\pi, \pi[$ sur $\{0, 255\}$. Un contour est alors défini comme **le lieu des maximums du gradient dans la direction \vec{g} du gradient**. Un point de contour vérifie donc :

$$\frac{\partial G}{\partial g} = 0 \quad \text{et} \quad \frac{\partial^2 G}{\partial g^2} \leq 0 \quad (11.2)$$

avec :

$$\frac{\partial}{\partial g} = \vec{g} \cdot \vec{\nabla} \quad (11.3)$$

L'équation obtenue est complexe et non linéaire :

$$\frac{\partial f}{\partial x} \cdot \frac{\partial}{\partial x} \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} + \frac{\partial f}{\partial y} \cdot \frac{\partial}{\partial y} \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} = 0$$

Elle n'a donc en général pas de solution explicite et ne peut être résolue que par tâtonnement. On procède alors en deux étapes : on calcule tout d'abord le gradient, puis on recherche les extrémums dans la direction du gradient.

Afin de garantir la double dérivation même en présence de discontinuités de type *marche d'escalier*, on prétraite l'image f par convolution avec une fonction au moins deux fois dérivable. On a beaucoup utilisé pour cela la gaussienne, mais on peut également prendre des polynômes de faible degré.

On peut également se placer dans les axes locaux définis par la tangente \vec{t} et la normale \vec{g} à la surface $f(x, y)$ (cf. figure 11.2). Ce repère est tourné d'un angle Φ par rapport au repère $\{x, y\}$:

$$\Phi = \text{Arctg} \left[\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} \right] \quad (11.4)$$

On a alors :

$$\frac{\partial f}{\partial g} = \frac{\partial f}{\partial x} \cos \Phi + \frac{\partial f}{\partial y} \sin \Phi$$

et les équations 11.2 donnent :

$$\frac{\partial^2 f}{\partial g^2} = \frac{\partial^2 f}{\partial x^2} \cos^2 \Phi + 2 \cdot \frac{\partial^2 f}{\partial x \partial y} \cos \Phi \sin \Phi + \frac{\partial^2 f}{\partial y^2} \sin^2 \Phi$$

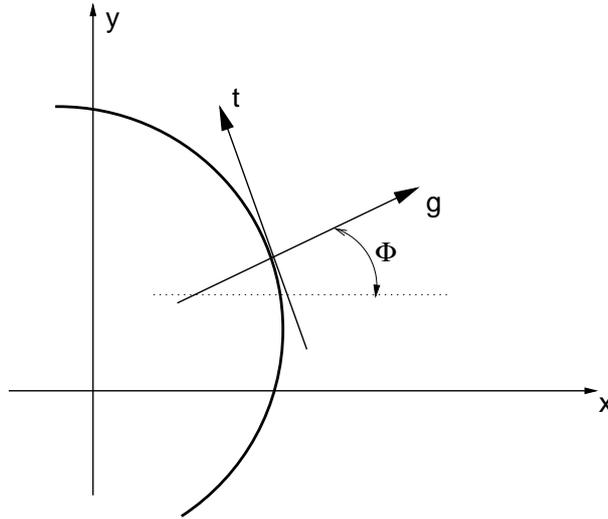


FIG. 11.2 – Repère local défini par le vecteur normal et le vecteur tangent à l'image.

Cette équation n'est pas plus simple à résoudre que la précédente. Cependant elle se relie assez aisément à l'équation du laplacien de f :

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

qui s'exprime également (par invariance du laplacien par changement de repère euclidien) dans le repère local $\{t, g\}$:

$$\Delta f = \frac{\partial^2 f}{\partial g^2} + \frac{\partial^2 f}{\partial t^2}$$

On voit que si le terme $\frac{\partial^2 f}{\partial t^2}$ est négligeable (c'est-à-dire si le contour a une courbure très faible), alors :

$$\Delta f = 0 \quad \approx \quad \frac{\partial^2 f}{\partial g^2} = 0 \quad (11.5)$$

Cette propriété sera utilisée dans les approches dites par passage par zéro du laplacien (cf détecteur de Marr).

11.2 Les approches classiques

Cette section présente un ensemble de méthodes qui ont eu historiquement une grande importance en traitement des images. Bien que certaines soient encore régulièrement employées, un lecteur intéressé par les techniques plus actuelles pourra se rendre directement à la section 11.3.

11.2.1 Les détecteurs de gradient par filtrage

Ces détecteurs reposent tous sur une recherche d'un extremum de la dérivée première (ou d'un passage par zéro d'une dérivée seconde), celle-ci étant calculée de diverses manières, mais généralement par un filtrage passe-haut précédé d'un léger filtrage passe-bas pour s'affranchir des bruits. Ces approches par **filtrage linéaire** sur des critères très simples de sélection en fréquence ont reçu une base théorique dans [Modestino et Fries, 1977, Shanmugam et Green, 1979]. La somme des connaissances sur ce sujet se trouve dans [Torre et Poggio, 1986].

Ainsi, Modestino et Fries ont proposé de détecter les contours par un filtrage de Wiener, le but de ce filtrage étant d'effectuer sur l'image non bruitée une opération de laplacien qui sera suivie par la suite d'une détection de zéros alliée à un seuillage (nous verrons ces étapes ultérieurement).

Le filtre de base de Modestino et Fries est un filtre récursif de la forme :

$$H_0(z_1, z_2) = \frac{\sum_i \sum_j b_{i,j} z_1^{-i} z_2^{-j}}{1 + \sum_i \sum_j a_{i,j} z_1^{-i} z_2^{-j}}$$

où les coefficients $a_{i,j}$ et $b_{i,j}$ sont déterminés à partir de 3 paramètres spécifiques de l'image à traiter : λ , ρ et ϵ . λ est le paramètre poissonnien de densité de contours (il exprime la distance moyenne entre 2 contours lors d'un balayage de l'image), ρ la constante de la loi exponentielle de l'auto-corrélation de l'image (cf. chapitre 2) et ϵ la puissance du bruit. Comme un tel filtre ne filtre que du coin supérieur gauche au coin inférieur droit, on lui associe les 3 autres filtres obtenus par rotation de $\pi/2$, pour obtenir un filtre final. Dans la pratique pour assurer des temps de calcul raisonnables, on choisit des filtres d'ordre 1 ($n = 1$).

11.2.2 Les détecteurs de gradient par masques

A côté de ces approches très inspirées du traitement du signal, des filtres de dérivation plus empiriques ont été proposés à partir d'estimateurs locaux de l'image f ou de ses dérivées $\frac{\partial f}{\partial x}$. Ces estimées sont obtenues à l'aide de masques (cf. figure 11.3) appliqués sur des fenêtres de 2×2 pixels ou 3×3 pixels (exceptionnellement, en cas d'images très bruitées, sur des fenêtres plus grandes). On note sans surprise que la somme des coefficients de ces filtres est nulle (fonction de transfert nulle à la fréquence 0), et que les coefficients sont anti-symétriques. Les filtres les plus utilisés sont, dans l'ordre décroissant :

Sobel > Roberts > Gradient > Prewitt

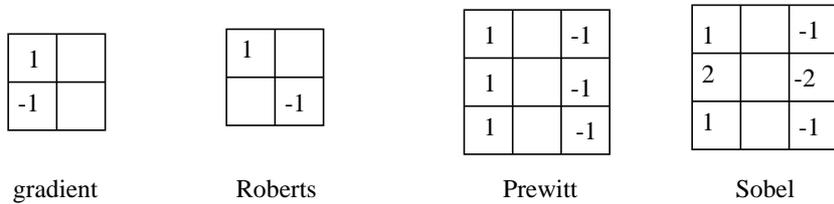


FIG. 11.3 – Quatre filtres de détection de contours par estimation du gradient. Les filtres représentés estiment une seule dérivée. Par rotation de $\pi/2$, on calcule la seconde dérivée.

Les filtres 3×3 sont un peu moins précis (c'est-à-dire que les contours qu'ils détectent sont moins bien localisés et souvent épais), mais les images ainsi obtenues sont généralement plus fiables et permettent des post-traitements plus poussés, ils sont également centrés sur un pixel et non entre des pixels (filtres à phase nulle).

La réponse de l'un quelconque de ces filtres s'obtient de la façon suivante : le filtre est centré en chaque pixel successivement (au coin supérieur gauche pour les filtres 2×2 , au point central pour les filtres 3×3). Le produit du masque par les valeurs des pixels correspondant étant fait, la valeur absolue de la somme est retenue. Puis le masque est tourné de 90° autour de son centre et la même mesure est répétée. Les deux mesures sont alors ajoutées et leur somme constitue la mesure du gradient en ce point selon la formule de l'équation 11.1, ou selon la formule approchée (en norme \mathcal{L}_1), un peu plus rapide à calculer :

$$\hat{G} = \left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right|.$$

On peut également mesurer l'orientation Φ du contour au point donné en faisant le rapport des réponses des deux filtres comme dans l'équation 11.4.

De tels filtres sont particulièrement simples de mise en œuvre, et rapides de calcul, permettent en particulier d'assurer des calculs en temps réel pour des applications industrielles.

11.2.3 Pré- et post traitements

La qualité de la détection est très liée à la qualité des contours dans l'image (cela sera vérifié pour toutes les méthodes, mais particulièrement pour ces méthodes peu élaborées). Ainsi, elles sont souvent bonnes pour les applications de robotique, en environnement artificiel, pour des images aériennes à forte résolution (en milieu urbain par exemple), elles se dégradent très vite pour les scènes d'extérieur, les images médicales, et deviennent totalement inapplicables en imagerie bruitée (thermographie, imagerie ultrasonore, radar SAR, etc.). On a alors intérêt à revenir à des démarches moins intuitives comme celles exposées dans [Modestino et Fries, 1977, Shanmugam et Green, 1979].

Ces détecteurs doivent toujours être suivis d'une étape de post-filtrage, et souvent précédés d'une étape de pré-filtrage.

Parmi les filtres de **pré-traitement**, on recherche ceux qui diminuent le bruit tout en préservant les discontinuités. Ils ont fait l'objet du chapitre 10 : filtre médian par exemple, mais aussi filtre toboggan [Fairfield, 1990] (cf. section 10.3.1), ou filtre de diffusion anisotrope de Perona et Malik [Perona et Malik, 1987] (cf. section 10.2.4). Ces filtres ont cependant le défaut de créer des contours artificiels qui peuvent être par la suite difficiles à éliminer.

Les **post-traitements** commencent le plus souvent par l'élimination des points de contour trop faibles par un seuillage (seuillage adaptatif si l'image est très hétérogène). On élimine ensuite les points qui ne sont pas des extrémums locaux dans la direction du gradient. Cela s'obtient à partir de la direction Φ mesurée précédemment et une comparaison simple des points rencontrés (ou interpolés si l'on souhaite une précision meilleure que le pixel). On peut opérer également des seuillages *par hystérésis*, afin de ne conserver que les composantes les plus importantes des contours. Pour cela, on procède à deux seuillages, le second étant très tolérant (c'est-à-dire laissant un très grand nombre de contours candidats) ; on ne garde des contours sélectionnés par ce second seuil que les points connexes d'un point de contour préservé par le premier (cela se fait aisément par un double étiquetage des images seuillées) cf. figure 11.4.

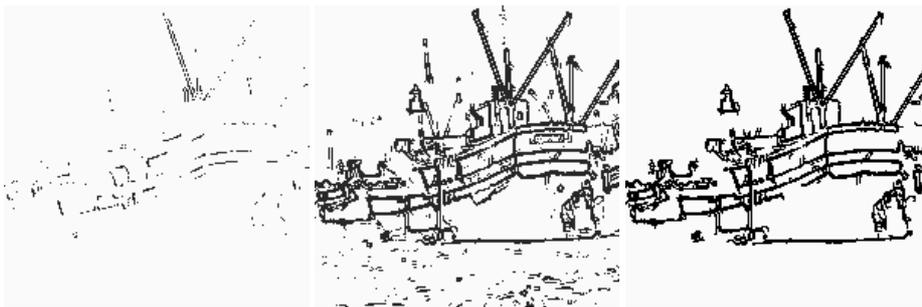


FIG. 11.4 – Filtrage de contours par hystérésis : à gauche, application d'un seuil sévère (70) au gradient de l'image bateau, seuls les contours les plus fiables sont détectés. Au centre, application d'un seuil tolérant (25), beaucoup de contours sont détectés, mais également beaucoup de bruit. A droite, résultat du filtre par hystérésis : seuls les contours du second filtrage connexes à ceux du premier sont conservés. Le détecteur de contours est un détecteur de Sobel. Aucun filtrage n'est appliqué pour réduire l'épaisseur des contours.

Enfin on procède à des étapes de **poursuite** et de **fermeture** des contours. La première opération a pour objectif

de supprimer les petites disparitions de contours qui peuvent se produire par suite du bruit ou d'une perte de contraste, la seconde se propose de détecter des objets topologiquement clos, c'est-à-dire des zones fermées. Ces étapes seront vues plus en détail plus loin.

11.2.4 Les détecteurs de passage par zéro du laplacien

Ces méthodes ont été proposées en 1976 [Marr et Hildreth, 1980]. Elles ont eu une grande importance historique, étant considérées comme le prototype du détecteur de contour inspiré des systèmes biologiques (*primal sketch* de Marr). Elles utilisent le fait que le passage par zéro du laplacien permet de bien mettre en évidence les extrémums de la dérivée. Nous avons eu l'occasion de voir (cf. équation 11.5) qu'en l'absence de forte courbure, le passage par zéro du laplacien correspond en effet bien au maximum du gradient dans la direction du gradient. Ces méthodes tirent en outre profit du fait que les zéros de la dérivée seconde constituent un réseau de lignes fermées (évitant donc, en principe, les étapes de poursuite et de fermeture). Il en est de même du réseau des lignes de crête du gradient, mais le premier est plus aisément détecté à partir d'un simple étiquetage des zones positives et négatives. Mais l'estimation de la dérivée seconde étant très sensible aux bruits, il convient de filtrer très fortement l'image avant d'en mesurer le laplacien. Cela conduit au filtrage suivant :

$$\psi * \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) [\psi * f].$$

où ψ est un filtre passe-bas ; ce qui se réécrit symboliquement :

$$\text{image des contours} = \text{passage par zéro} \left(f * \left[\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \right] \right).$$

cela exprime qu'une image de contours est obtenue par filtrage de l'image par la dérivée seconde d'un filtre passe-bas, puis détection des zéros de la fonction ainsi obtenue.

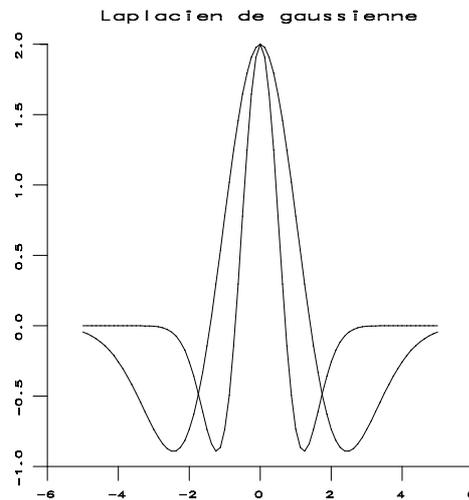


FIG. 11.5 – Filtre LOG pour deux valeurs différentes de σ .

Les filtres les plus utilisés pour ces filtrages passe-bas sont les filtres gaussiens [Marr et Hildreth, 1980]. Marr a montré en effet qu'avec de tels filtres on pouvait approcher de très près les effets donnés par le système visuel

humain. Le filtre obtenu par convolution avec le laplacien d'une gaussienne est connu sous le nom de LOG (cf figure 11.5). Un autre filtre très utilisé et très proche du LOG est le DOG (*Difference of Gaussians*) (figure 11.6, qui procède par différence de deux gaussiennes dont les écarts-type sont dans un rapport 1,6 (pour simuler au mieux le système visuel). Les gaussiennes sont bien sûr affectées des facteurs adéquats pour que la somme des parties positives du filtre soit égale à la somme des parties négatives (valeur nulle à la fréquence 0 de la fonction de transfert). Sa mise en œuvre peut tirer profit des diverses qualités de la gaussienne : séparabilité, noyau auto-reproducteur, limite de suites de polynômes, approximation par des splines, etc., ce qui permet d'en accélérer la réalisation.

Dans des approches très sommaires, le DOG est parfois remplacé par le DOB (*Difference of Box*), filtre très rapide même pour des grandes tailles (on tire profit alors du fait que les filtres de moyenne (*Box*) ont une complexité de calcul indépendante de la taille de la fenêtre d'intégration au prix d'un coup mémoire important). Les résultats de ces filtres est cependant généralement de piètre qualité.

Les filtres LOG et DOG, n'ont plus aujourd'hui qu'un intérêt historique. On leur préfère les filtres analytiques que nous verrons plus loin.

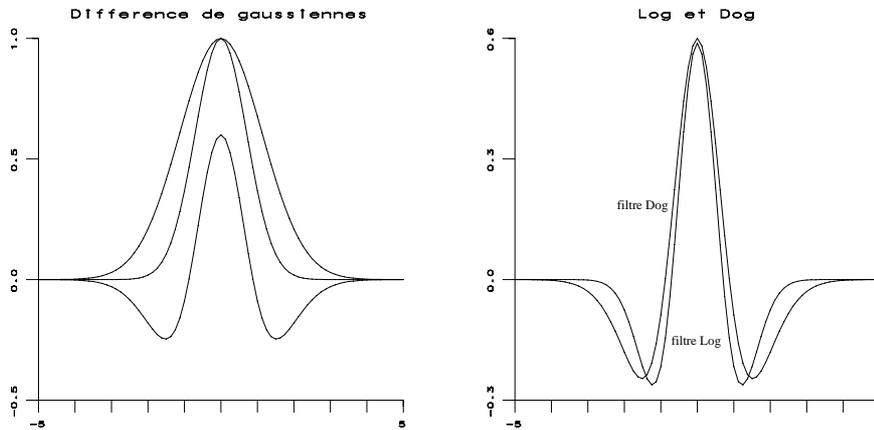


FIG. 11.6 – A gauche : filtre DOG et les deux gaussiennes dont il est la différence (leurs écarts-types sont dans un rapport 1,6). A droite filtre LOG et filtre DOG comparés.

11.2.5 Les détecteurs par masquage adapté

Dans ces techniques, on recherche en chaque point de l'image la présence d'une configuration conforme à un gabarit appartenant à un *dictionnaire de contours*. Pour cela, on définit une distance entre un contour type et une fenêtre de l'image. C'est par exemple l'inverse du produit scalaire des deux fenêtres : si l'image est dénotée $f(x, y)$, et le $k^{\text{ème}}$ gabarit $g_k(x, y)$, on calcule par exemple :

$$r_k(x, y) = \frac{\left(\sum_{x,y \in fen\hat{e}tre} f^2(x, y) \cdot \sum_{x,y \in fen\hat{e}tre} g_k^2(x, y) \right)^{1/2}}{\sum_{x,y \in fen\hat{e}tre} f(x, y)g_k(x, y)}$$

pour toutes les valeurs de k . On ne conserve que la meilleure des valeurs et seulement si elle est suffisamment faible. L'approche par masques adaptés (*template matching*), permet de connaître aisément la direction du contour, elle permet également de soumettre, en parallèle, ces calculs à des architectures adaptées. Mais la qualité des contours, ainsi que les post- traitements qu'ils requièrent sont très semblables à ceux que l'on a vus pour les filtres de maximum du gradient.

Les filtres adaptés les plus utilisés sont le filtre de Kirsch (cf. tableau 11.1), le filtre de Nevatia et Babu (cf. tableau 11.2) et les boussoles directionnelles de Prewitt (cf. tableau 11.3).

| | | |
|----|----|----|
| 5 | 5 | 5 |
| -3 | 0 | -3 |
| -3 | -3 | -3 |

| | | |
|----|----|----|
| -3 | 5 | 5 |
| -3 | 0 | 5 |
| -3 | -3 | -3 |

TAB. 11.1 – Masques de Kirsch : il y a 8 filtres issus de ces 2 fenêtres par rotation de $\pi/4$.

| | | | | |
|------|------|---|-----|-----|
| -100 | -100 | 0 | 100 | 100 |
| -100 | -100 | 0 | 100 | 100 |
| -100 | -100 | 0 | 100 | 100 |
| -100 | -100 | 0 | 100 | 100 |
| -100 | -100 | 0 | 100 | 100 |

| | | | | |
|------|------|------|-----|-----|
| -100 | 32 | 100 | 100 | 100 |
| -100 | -78 | 92 | 100 | 100 |
| -100 | -100 | 0 | 100 | 100 |
| -100 | -100 | -92 | 78 | 100 |
| -100 | -100 | -100 | 32 | 100 |

TAB. 11.2 – Masques de Babu et Nevatia : il y a 12 filtres issus de ces 2 fenêtres par rotation, la première de $\pi/2$, la seconde de $\pi/4$.

| | | |
|----|----|----|
| 1 | 1 | 1 |
| 1 | -2 | 1 |
| -1 | -1 | -1 |

| | | |
|----|----|---|
| 1 | 1 | 1 |
| -1 | -2 | 1 |
| -1 | -1 | 1 |

TAB. 11.3 – Masques de boussole directionnelle : il y a 8 filtres issus de ces 2 fenêtres par rotation.

Au-delà de ces approches souvent heuristiques, Hueckel a proposé une démarche beaucoup plus rationnelle, s'appuyant sur la décomposition du signal d'image sur une base de fonctions orthogonales en coordonnées polaires (polynômes d'Hermite utilisés également pour l'oscillateur harmonique), tronquée à ses 8 premiers termes [Hueckel, 1971]. Malheureusement les résultats n'ont pas été au niveau des investissements et cette voie est maintenant abandonnée.

11.3 Les approches analytiques

Nous allons voir maintenant une approche qui a permis une bien meilleure compréhension des conditions d'une bonne détection de contours et qui a ainsi conduit à des détecteurs de très bonne qualité. On les voit émerger dans les années 85, à partir des travaux : [Torre et Poggio, 1986, Shen et Castan, 1986, Canny, 1986].

11.3.1 Les critères de Canny

Canny, dans une approche originale, [Canny, 1986] a proposé un filtre déterminé analytiquement à partir de 3 critères :

1. garantir une bonne détection, c'est-à-dire une réponse forte même à de faibles contours,
2. garantir une bonne localisation,
3. assurer que pour un contour il n'y aura qu'une seule détection (éviter les effets de *rebonds* dus, par exemple, à la troncature des filtres).

Ces 3 critères s'expriment par l'optimisation conjointe de 3 fonctionnelles qui permettent de définir le **filtre linéaire optimal** pour la détection d'une **marque d'escalier** sous l'hypothèse d'un bruit additif indépendant du signal ¹.

Si l'on considère que le filtre a pour réponse impulsionnelle $\psi(x)$, ces fonctionnelles s'écrivent :

$$\text{bonne détection : } \Sigma = \frac{\int_0^{\infty} \psi(x) dx}{\sqrt{\int_{-\infty}^{\infty} \psi^2(x) dx}}$$

$$\text{bonne localisation : } \Lambda = \frac{|\psi'(0)|}{\sqrt{\int_{-\infty}^{\infty} \psi'^2(x) dx}}$$

$$\text{réponse unique : } \frac{|\psi'(0)|}{\sqrt{\int_{-\infty}^{\infty} \psi'^2(x) dx}} = k \frac{\int_{-\infty}^0 \psi(x) dx}{\sqrt{\int_{-\infty}^{\infty} \psi^2(x) dx}}$$

La résolution du système est assez complexe (on maximise le produit $\Sigma\Lambda$ sous la contrainte du 3^{ème} terme constant). Par ailleurs, Canny souhaite obtenir un filtre monodimensionnel (dans la direction orthogonale au contour), à réponse impulsionnelle finie (RIF) ; celui-ci a alors une expression complexe, composée de 4 termes, chacun combinant des lignes exponentielles et trigonométriques :

$$\psi(x) = a_1 e^{x/\sigma} \sin \omega x + a_2 e^{x/\sigma} \cos \omega x + a_3 e^{-x/\sigma} \sin \omega x + a_4 e^{-x/\sigma} \cos \omega x$$

où les coefficients a_i et ω sont déterminés à partir de la taille du filtre. Le paramètre σ est un paramètre de grande importance que nous retrouverons dans tous les autres filtres dérivés de l'approche de Canny. C'est un paramètre d'échelle qui indique en-deça de quelle distance deux contours parallèles seront confondus en un seul. Canny montre que la dérivée d'une gaussienne est une bonne approximation de son filtre.

$$\psi(x) \approx -x \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

C'est donc la réduction à 1 dimension du filtre proposé par Marr à 2 dimensions (le maximum de la dérivée de la gaussienne est obtenu pour le passage par zéro du laplacien de gaussienne). Ce filtre donne des résultats de bonne qualité. Le critère de Canny (critère $\Sigma\Lambda$) vaut 1,12 pour le filtre RIF optimal, 0,97 pour le filtre gaussien. Remarquons cependant que le filtre gaussien n'est pas RIF.

11.3.2 Les filtres de Deriche et Shen et Castan

Au filtre de Canny, on préfère souvent le détecteur de Deriche, qui répond exactement aux mêmes critères de qualité que celui de Canny, mais qui possède une réponse impulsionnelle infinie (filtre RII). Il a pu donc être synthétisé de façon récursive particulièrement efficace [Deriche, 1987]. Le filtre de Deriche a une expression générale de la forme :

$$\psi(x) = -cx \exp(-\alpha|x|),$$

avec :

$$c = \frac{[1 - \exp(-\alpha)]^2}{\exp(-\alpha)}.$$

avec les critères même de Canny il est supérieur au filtre de Canny ($\Sigma\Lambda = 2$). Le paramètre α de Deriche représente alors l'inverse de l'écart type σ de la gaussienne du filtrage de Canny ($\alpha = \frac{\sqrt{\pi}}{\sigma}$). Le filtre de Deriche s'écrit de

¹Par rapport aux approches proposées précédemment, on note que le filtre reste linéaire, mais que le critère d'optimisation est plus complet.

façon récursive en fonction des valeurs $f(x)$ de l'image, par l'intermédiaire de 2 filtres l'un décrivant l'image de gauche à droite et l'autre de droite à gauche :

$$z^+(x) = c\epsilon + 2\epsilon z^+(x-1) - \epsilon^2 z^+(x-2)$$

$$z^-(x) = -c\epsilon + 2\epsilon z^-(x+1) - \epsilon^2 z^-(x+2)$$

$$z(x) = z^+(x) + z^-(x)$$

avec :

$$\epsilon = \exp(-\alpha)$$

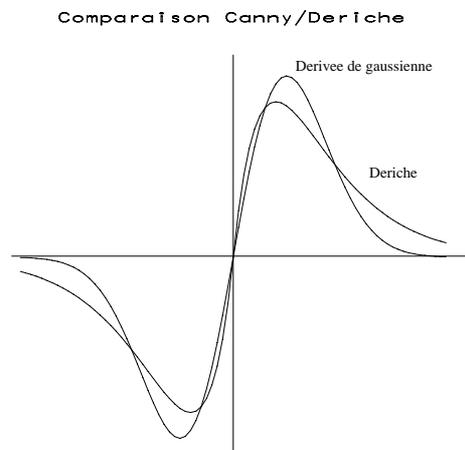


FIG. 11.7 – Le filtre de Deriche : $x \cdot \exp(-|x|)$ et la dérivée de la gaussienne : $\frac{x}{\sqrt{\pi}} \cdot \exp(-x^2/\pi)$.

Le filtre de Shen et Castan [Shen et Castan, 1986] possède également de très bonnes qualités de détection, il s'apparente à la même famille des filtres exponentiels (mais comme il n'est pas dérivable à l'origine, il ne se prête pas à une comparaison avec le critère de Canny) :

$$f(x) = c \cdot \text{sgn}(x) \exp(-\alpha|x|)$$

Le filtre de Shen et Castan, en raison de sa forme à l'origine, donne une localisation très précise des contours, mais il est sensible aux bruits.

Dans [Demigny et Kamlé, 1997], Demigny a étendu les critères d'optimalité des filtres de Canny à des images discrètes. Il les a ensuite étendus à des profils en rampe et non plus en marche d'escalier.

11.3.3 L'extension à 2D

Tous ces filtres sont mono-dimensionnels, le filtrage 2D est donc obtenu par l'action de deux filtres croisés, l'un en x , l'autre en y .

Pour définir l'un de ces deux filtres, lorsque l'on a choisi le gabarit du filtre dans la direction perpendiculaire au contour, afin de mieux intégrer le signal, on utilise, dans la direction du contour, une **fonction de prolongement**

qui permet de filtrer les bruits. C'est souvent une gaussienne de même écart-type que celle qui est dérivée dans le détecteur. Par exemple la dérivée en x pourra être estimée par convolution de l'image par :

$$\psi(x, y) = -x \exp(-(\alpha x^2)) \exp(-(\alpha y^2))$$

Ces filtres ne sont donc généralement pas isotropes. C'est-à-dire qu'ils ne répondent pas de façon identique à des contours identiques mais d'orientations différentes. Les seules fonctions séparables et isotropes sont les fonctions gaussiennes. Mais même si l'on utilise des filtrages gaussiens, la mise en œuvre discrète du filtre sur une maille carrée rend les filtres anisotropes. Des travaux importants ont été consacrés à la réduction de cet effet [Kunduri et al., 1999] utilisant une combinaison de deux détecteurs à $\pi/4$ l'un de l'autre.

En conclusion, dans une approche moderne de la détection des contours, un filtre de détection se compose de deux estimateurs de dérivées, l'un selon Ox , l'autre selon Oy . L'un de ces détecteurs (choisissons celui selon Ox) se compose du produit de 2 fonctions :

- selon Oy c'est une fonction passe-bas, symétrique, (la fonction de prolongement) dont l'étendue est fonction de l'importance du bruit et de la distance que l'on souhaite respecter entre deux contours ;
- selon Ox c'est une fonction passe-haut (anti-symétrique), souvent égale à la dérivée de la précédente.

Les couples « fonctions de prolongement - dérivées » sont typiquement issus de la gaussienne (Canny), de l'exponentielle décroissante (Shen et Castan) ou du produit de l'exponentielle par x (Deriche) (cf. figure 11.8), mais on peut également les concevoir très librement pour une application spécifique.

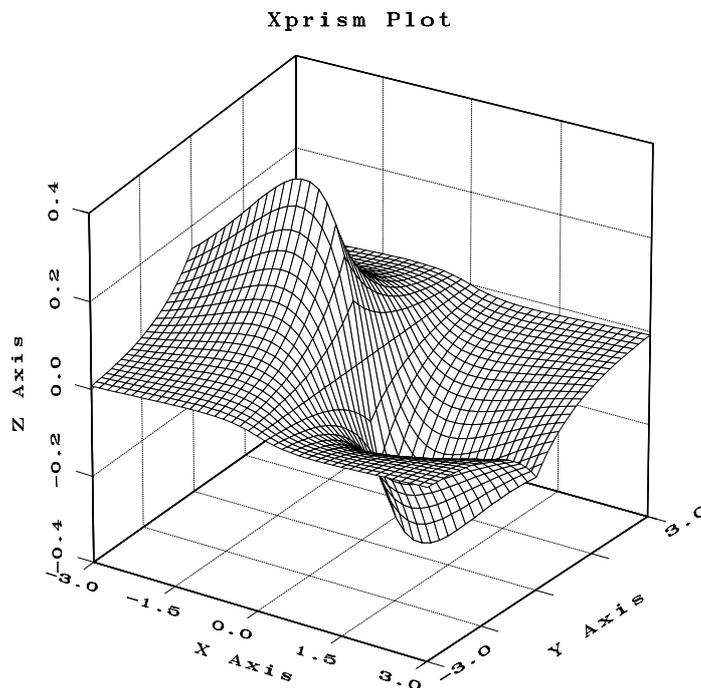


FIG. 11.8 – Profil général d'un filtre d'estimation de la dérivée dans la direction Ox . Selon Oy le filtre est un intégrateur (ici $\exp(-|y|)$), selon Ox , c'est un dérivateur (ici $x.\exp(-|x|)$).

11.3.4 Variantes autour des filtres analytiques

D'autres filtres aux performances comparables, mais résultats d'optimisations différentes, ont été proposés dans :

- [Spacek, 1986] : On modifie le critère de Canny, et on cherche pour filtre une spline cubique de forme $x^3 - 2x^2 + x$.
- [Petrou et Kittler, 1991] : On considère alors que le contour est également le résultat d'un flou, ce n'est donc plus une marche, mais une rampe. Les critères de sélection sont ceux de Spacek et le filtre est une combinaison non-linéaire de lignes trigonométriques et d'exponentielles décroissantes.
- [Sarkar et Boyer, 1991] : On modifie dans ce cas les critères de Canny, on adopte une approche variationnelle et on choisit la représentation récursive d'un filtre RII comme Deriche.

Une excellente comparaison et analyse des filtres de détection des contours est donnée dans [Cocquerez et Philipp, 1996].

Utilisant les résultats précédents ainsi que des méthodes éprouvées d'optimisation semi-globale développées dans les contours actifs (cf. section 11.4), des solutions variationnelles ont également été proposées qui sont fort efficaces et permettent d'ajouter des informations connues a priori sur l'allure des contours lors de leur détection. Voir par exemple [Fua et Leclerc, 1990].

11.4 Les contours actifs = les *snakes*

11.4.1 Une approche physique des contours

Une approche très différente des méthodes antérieures de détection de contours a été proposée en 1987 par Kass, Witkins et Terzopoulos [Kass et al., 1988] (voir aussi [Blake et Isard, 1998]), appelée contours actifs ou *snakes*. Il s'agit d'une méthode semi-interactive dans laquelle l'opérateur place dans l'image, au voisinage de la forme à détecter, une ligne initiale de contour. Cette ligne sera amenée à se déformer sous l'action de plusieurs forces :

- une énergie propre, assimilée à l'énergie mécanique de tension et de torsion d'une ligne matérielle,
- une énergie potentielle imposée par l'image qui vise à plaquer la courbe sur les contours,
- une énergie externe, introduite par l'utilisateur pour traduire les contraintes spécifiques du problème qu'il se pose.

Sous ces énergies, le contour *actif* va évoluer pour rechercher la position d'énergie minimale, qui sera ainsi un compromis entre les diverses contraintes du problème.

L'écriture formelle du problème passe par la définition paramétrique du contour, en fonction d'une variable s généralement l'abscisse curviligne :

$$v(s) = [x(s), y(s)]^t \quad s \in [0, 1]$$

$$E_{totale} = \int_0^1 [E_{interne}(v(s)) + E_{image}(v(s)) + E_{externe}(v(s))] ds \quad (11.6)$$

avec :

$$E_{interne} = \alpha(s) \left(\frac{dv}{ds} \right)^2 + \beta(s) \left(\frac{d^2v}{ds^2} \right)^2$$

où :

- la première dérivée prend en compte les variations de longueur de la courbe, (c'est donc un terme de tension (résistance à la rupture), qui est contrôlé par l'élasticité que l'on attribue au contour),
- tandis que la seconde exprime les variations de la courbure (c'est un terme de flexion contrôlé par la raideur du contour).

Ces deux termes agiront donc pour produire une courbe régulière (cf. figure 11.9).

Le second terme d'énergie : E_{image} caractérise les lignes que l'on souhaite suivre. Dans le cas de détection de contours, ce sont des lignes de fort gradient, il vaut donc généralement :

$$E_{image} = -\nabla f$$

mais il peut être adapté pour suivre les maximums des niveaux de gris (dans ce cas $E_{image} = f$), ou toute autre fonction définie à partir de ceux-ci.

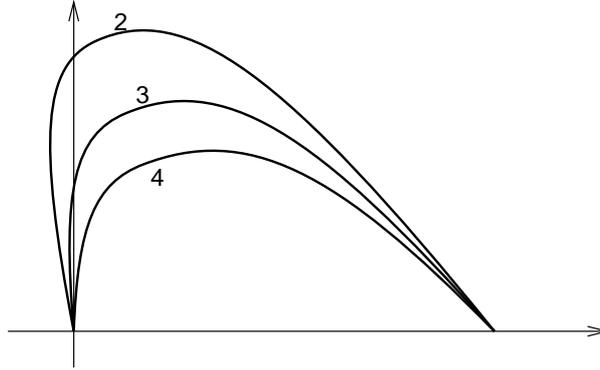


FIG. 11.9 – Déformation d'un contour actif à extrémités fixes pour des valeurs du rapport α/β (élasticité/raideur) égales à 2, 3 et 4, en l'absence d'une énergie liée à l'image.

Enfin le dernier terme d'énergie : $E_{externe}$ est choisi par l'utilisateur. Il peut avoir des formes très variées afin, par exemple, de contraindre le contour à ressembler à un gabarit donné, à s'approcher d'un contour déjà détecté sur une autre image (suivi de séquences, ou images à 3D)².

En l'absence de ce dernier terme, et dans les nombreux cas où l'on recherche les lignes de fort gradient, l'équation 11.6 s'écrit alors :

$$E_{totale} = \int_0^1 \left[-\nabla f + \alpha(s) \left(\frac{dv}{ds} \right)^2 + \beta(s) \left(\frac{d^2v}{ds^2} \right)^2 \right] ds \quad (11.7)$$

11.4.2 Mise en équation des contours actifs

L'équation intégrale 11.7 est résolue généralement de façon variationnelle. On suppose que le contour évolue vers un minimum d'énergie, soit vers un zéro de $\frac{\partial E_{totale}}{\partial t}$. En désignant par v' et v'' les dérivées de v le long de la courbe, on obtient une équation différentielle vectorielle :

$$\gamma \frac{\partial v}{\partial t} + (\alpha v')' - (\beta v'')'' = \frac{\partial |\nabla f|^2}{\partial v}$$

où α , β et γ sont potentiellement variables le long de s . Il y a plusieurs façons de concevoir la discrétisation de la courbe :

1. selon les différences finies : les éléments de la courbe sont réduits en des points auxquels sont attachés les éléments mécaniques (masse, raideur, etc.) de la courbe considérée concentrée en ces points ;

²On notera cependant que, dans le cas de volumes numériques tridimensionnels $f(x, y, z)$, il est possible d'étendre les contours actifs à des surfaces actives pour lesquelles l'énergie s'exprime comme pour des plaques minces.

- selon les éléments finis on remplace chaque portion de courbe par le segment élémentaire et les éléments mécaniques sont calculés sur ces segments.

Nous ne décrivons ici que la première approche la plus fréquemment employée. Après discrétisation de la courbe en un nombre n de points et en posant :

$$V^t = [v_0^t, v_1^t, v_2^t, \dots, v_{n-1}^t]^t$$

et :

$$h_v = -|\nabla f(v)|^2$$

on obtient une équation matricielle de la forme :

$$(A + \gamma I)V^t = \gamma V^{t-1} - h_v(V^{t-1})$$

où A est une matrice pentadiagonale³ (figure 11.11) de taille $n \times n$ fonction de α et β , I est la matrice unité de taille $n \times n$ et γ exprime une *inertie* de la courbe aux déplacements. Sa résolution donne :

$$V^t = (A + \gamma I)^{-1}(\gamma V^{t-1} - h_v(V^{t-1}))$$

Si les paramètres α , β et γ sont invariants le long de la courbe, il suffit alors de calculer une seule fois $(A + \gamma I)^{-1}$ pour résoudre le système pour tous les temps t . Sinon, on inverse la matrice à chaque instant. Les choix des paramètres α , β et γ qui ne sont pas dictés par le problème demeurent souvent délicats pour garantir une convergence convenable, ainsi que le choix de l'estimateur de ∇f . L'utilisation d'un estimateur par filtre de Deriche est parfois préconisée.

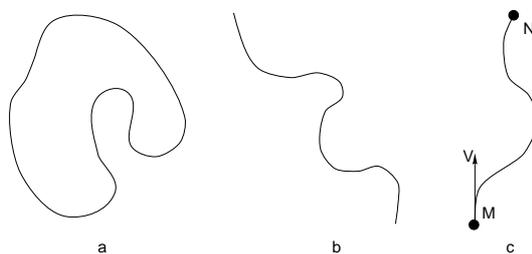


FIG. 11.10 – Contour actif fermé en a, ouvert à extrémités libres en b et ouvert à extrémités fixes en c (on impose en M une extrémité fixe et une orientation fixe donnée par le vecteur V).

Malgré les difficultés de réglage de la convergence, les contours actifs apportent une solution heureuse à la détection de contour, intermédiaire entre les solutions purement locales (opérateurs de type Sobel) et les segmentations globales. Trois types de contours actifs différents sont utilisés (cf. figure 11.10) :

- les contours actifs fermés (où $v_0^t = v_{n-1}^t$),
- les contours actifs à extrémités libres,
- les contours actifs à extrémités fixes (où les positions v_0 et v_{n-1} sont fixes dans le temps, ainsi que, potentiellement les premières dérivées en ces points).

Selon le type de contours actifs, les matrices A présentent la propriété d'être circulante (contours actifs fermés), Töplitz (contours actifs à extrémités libres) ou quelconques (figure 11.11).

Laissé libre d'évoluer seul, en l'absence de forces d'attraction dues au gradient dans l'image, le contour actif a tendance à se réduire à un point s'il est fermé, à une droite s'il est à extrémités libres. Pour compenser cette tendance, on est parfois amené à introduire des forces internes de *gonflage*. On a alors affaire à des *ballons*

³Dans le cas d'une discrétisation par éléments finis, la matrice est heptadiagonale

$$\begin{bmatrix}
 2\alpha+6\beta & -\alpha-4\beta & \beta & 0 & 0 & \dots & \dots \\
 -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & 0 & \dots & \dots \\
 \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & & \\
 0 & \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & & \\
 0 & 0 & \beta & -\alpha-4\beta & 2\alpha+6\beta & & \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 & & & & & & \dots
 \end{bmatrix}$$

FIG. 11.11 – Un contour actif fermé est représenté par une matrice circulante pentadiagonale.

[Cohen, 1991]. Comme les équations des forces de gonflage ne dérivent plus d'un potentiel, les écritures s'en trouvent un peu complexifiées.

Des solutions élégantes ont également été proposées pour rechercher la meilleure position du snake en en déplaçant ses nœuds sur toute la grille discrète de l'image par une technique de programmation dynamique [Amini et al., 1990] (cf. figure 11.12).

Par ailleurs, le couplage de techniques de contours actifs avec des méthodes s'appuyant sur des propriétés statistiques distinguant la forme de son fond, et optimisées par un maximum de vraisemblance au sens bayésien, a montré sa très grande efficacité, permettant de faire le lien entre techniques de contours et de régions [Chesnaud et al., 1999].

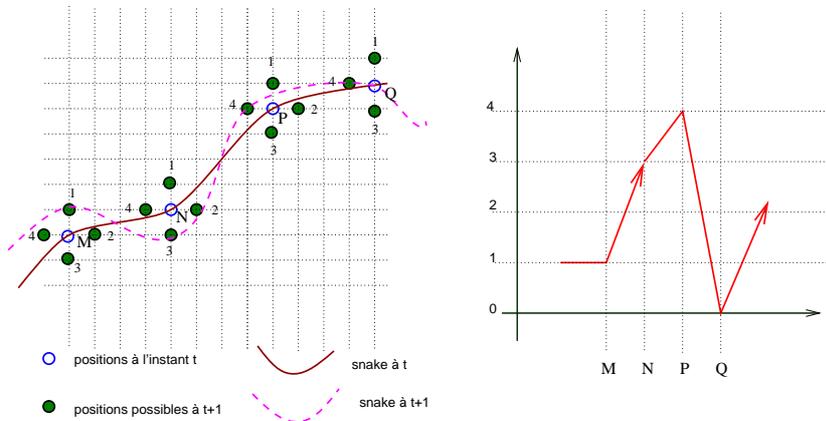


FIG. 11.12 – Dans la technique d'optimisation des contours actifs par programmation dynamique, chaque nœud M, N, P, Q peut se déplacer dans l'un des nœuds 1, 2, 3, ou 4, ou rester en place (position 0). L'optimisation se fait dans le graphe de droite par propagation de gauche à droite.

11.4.3 Les ensembles de niveaux (level sets)

Ce sont également des représentation variationnelles des contours qui conduisent donc à des solutions qui évoluent au cours du temps dans l'image, régies par un critère global. Ces modèles ont cependant le mérite remarquable de pouvoir changer de topologie si les contours l'imposent : par exemple un contour simple peut évoluer en deux contours séparés, ou, à l'inverse, deux contours séparés se réunir en un seul contour. Cela est rendu pos-

sible par l'utilisation de fonctions d'une dimension supérieure à celle des contours recherchés (si l'on cherche des lignes de contours, on introduira pour inconnue une surface de \mathbb{R}^3 , si on cherche des surfaces de \mathbb{R}^3 , on introduit des volumes de \mathbb{R}^4). Le contour est alors défini comme l'ensemble de niveau zéro (*level sets*) de cette fonction : $v(s)$ est une ligne de niveau $z = 0$ de la surface $z = f(x, y)$, souvent prise comme la distance au contour [Osher et Sethian, 1988, Sethian, 1996].

Sur le contour :

$$\frac{dz}{ds} = 0 = \frac{\partial z}{\partial x} \frac{dx(s)}{ds} + \frac{\partial z}{\partial y} \frac{dy(s)}{ds}$$

et la courbe de niveau évolue en fonction des itérations t selon l'équation fondamentale :

$$\frac{\partial z}{\partial t} = \|\vec{grad}(z)\| \vec{N} \wedge \frac{d\vec{v}}{dt}$$

où \vec{N} représente la normale au contour v .

Des équations semblables existent à 3D [Zhao et al., 2000].

11.4.4 Les modèles géodésiques actifs

Les approches par contours actifs peuvent être reliées à la famille des problèmes d'optimisation géodésique dans des espaces dont la géométrie est fixée par la dynamique de l'image (*cf.* chapitre 12, section 12.5.1). Ces techniques, appelées modèles géodésiques actifs, ont été proposées dans [Caselles et al., 1997] et le lien avec les contours actifs y est discuté.

Le problème du snake, s'il n'y a pas d'élasticité, peut s'écrire comme la minimisation de l'énergie :

$$J_1(v) = \int_a^b |v'(s)| ds + \lambda \int_a^b \psi(|\nabla f(v(s))|)^2 ds$$

où ψ représente le détecteur de contour. Caselles a introduit une fonctionnelle du type :

$$J_2(v) = 2\sqrt{\lambda\alpha} \int_a^b |v'(s)| \psi(|\nabla f(v(s))|) ds$$

Minimiser J_2 revient à trouver le chemin de longueur minimale dans l'espace riemannien dont la métrique est induite par l'image. On a pu montrer un certain nombre d'équivalences entre les propriétés des minimums de J_1 et de J_2 dans [Aubert et Blanc-Féraud, 1999].

Les contours géodésiques utilisent des solutions proches de celles des ensembles de niveaux pour le codage et la représentation de la ligne de contours.

11.5 La poursuite et la fermeture des contours

Si l'on excepte les contours actifs et les passages par zéro des laplaciens, la plupart des détecteurs de contours fournissent des contours ouverts, c'est-à-dire qu'ils ne séparent pas les composantes de l'image dans des objets topologiquement distincts. Cela rend souvent plus difficile l'étape de reconnaissance des formes qui doit suivre. Pour cela, on a proposé plusieurs méthodes permettant d'obtenir des contours clos. C'est ce que l'on appelle la **fermeture de contours**.

De nombreuses solutions ont été proposées pour cette étape fondamentale, reposant sur des principes très différents (les contours actifs à extrémités fixes que nous venons de voir en sont un exemple). La qualité des

résultats obtenus est généralement directement liée au coût informatique consenti, de très bonnes solutions existant, mais à des coûts prohibitifs pour la plupart des applications.

11.5.1 Méthodes de recherche dans des graphes

Ce sont les méthodes qui reposent sur les bases les plus solides et susceptibles de donner les meilleurs résultats, mais leurs coûts sont souvent très élevés, car elles se proposent d'explorer l'espace de toutes les solutions et de retenir la meilleure en fonction d'un critère que l'opérateur s'est donné. L'écueil de la recherche combinatoire exhaustive peut être évité par des recherches polynomiales assez complexes, surtout fondées sur la programmation dynamique.

La technique de base, présentée par Martelli [Martelli, 1972] et Montanari [Montanari, 1971], est celle d'une optimisation combinatoire minimisant la fonction de coût (ou maximisant une fonction de mérite). Les algorithmes de programmation dynamique (Viterbi, A^*), sont particulièrement bien adaptés à ce problème, mais, malgré leur complexité polynomiale sont généralement limités par des profondeurs de recherche de quelques pixels si le graphe est la totalité de l'image.

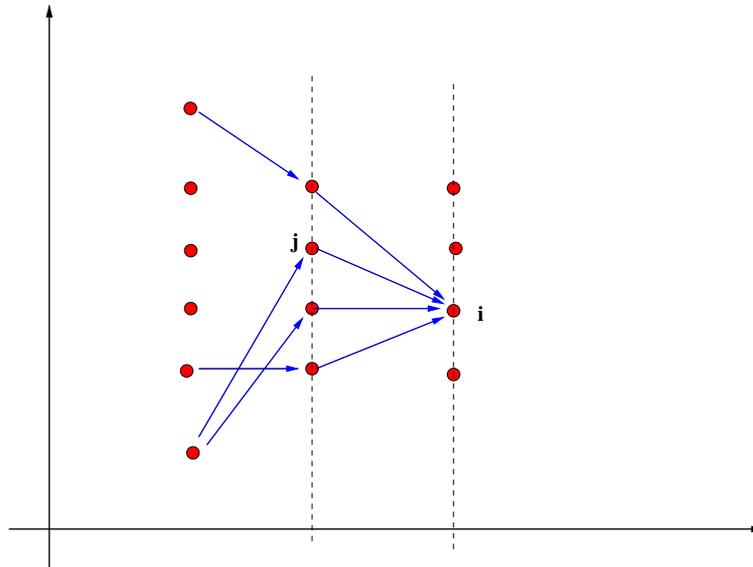


FIG. 11.13 – Résolution d'un problème de recherche de chemin optimal dans un graphe par programmation dynamique : la propagation se fait de la gauche vers la droite. On trouve le meilleur chemin arrivant à l'état i en optimisant la fonction de coût sur tous les états j qui ont été calculés à l'étape antérieure.

Dans une approche de type Viterbi, on considère l'évolution pas à pas d'un contour. À chaque pas le coût du contour est évalué pour toutes les provenances possibles du contour et n'est retenu pour tout point du parcours que le seul chemin qui minimise le coût pour arriver à ce point (cf. figure 11.13). On garde alors en chaque point ce coût minimum ainsi que l'adresse du *père* qui a permis d'y aboutir. Un nouveau pas est alors fait et toutes les destinations sont explorées que l'on peut atteindre en un pas à partir des points déjà atteints. La fonction de coût se calcule de la façon suivante :

$$\Gamma(i) = \gamma(i) + \min_{j \in \text{antécédents}} [\Gamma(j) + \delta(i, j)] = \gamma(i) + \epsilon(i)$$

où $\gamma(i)$ est inversement proportionnel à la qualité du point i comme point de contour et $\delta(i, j)$ exprime l'incompatibilité de i et j le long d'un contour. Le terme γ est un terme d'innovation le long du chemin tandis que ϵ est un

terme de mémoire. La pondération de ces deux termes est un problème délicat en programmation dynamique.

Un contour est fermé lorsque le point i est confondu avec un point d'un autre contour. Le coût de la fermeture est alors l'intégrale le long de la chaîne des coûts de liaison δ et des coûts individuels γ . Pour éviter de trop pénaliser les chaînes longues, il est possible de calculer le *coût moyen* le long du parcours en pondérant le terme de mémoire ϵ par le chemin parcouru.

En limitant les domaines de recherche à des champs convexes, ou plus généralement en introduisant des limitations sur les types de chemin que l'on s'autorise à trouver, on peut étendre notablement la portée de ces fermetures (cf. figure 11.14). Des approches multi-résolutions permettent de travailler de façon hiérarchique, d'une fermeture grossière à une fermeture précise, en limitant progressivement le domaine de recherche à une bande étroite autour du contour détecté à l'étape précédente. Dans d'autres cas, on se limite à des horizons faibles pour entreprendre une recherche par programmation dynamique qui est validée, puis relancée à partir du résultat précédent.

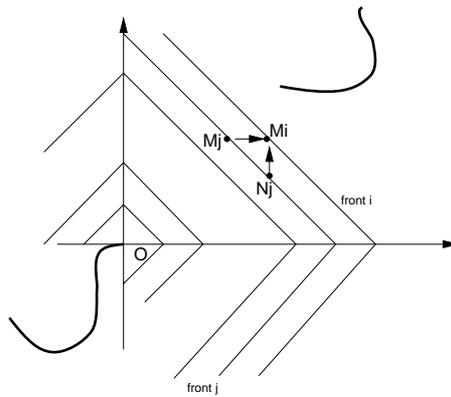


FIG. 11.14 – Exemple d'une fermeture de contours par recherche d'un chemin optimal dans un graphe. Le contour à fermer s'arrête en O . Les horizons de recherche à chaque étape de la programmation dynamique sont des carrés (fronts d'ordre i), centrés en O . Chaque point M_i peut avoir 2 antécédents (M_j et N_j). Une fermeture s'obtient lorsqu'un des points appartient à un autre contour. Pour éviter que le contour ne se ferme trop rapidement sur lui-même (près de O), il faut invalider les points du contour lui-même. Une telle propagation contraint les contours à une courbure toujours de même signe.

Dans une recherche de type A^* , on dispose d'une estimée de la fonction de coût ($\hat{\Gamma} > \Gamma_{opt}$) et la fonction de coût doit être croissante le long d'une fermeture. On fait une recherche en largeur d'abord en comparant à chaque nœud la valeur locale du coût à cette estimée et on invalide le nœud si le coût dépasse l'estimée. On a tout intérêt à trouver donc une estimée très proche du coût optimal réel.

Les fonctions de coût sont les points déterminants de la qualité de la méthode. Cette fonction de coût doit être locale pour préserver le caractère markovien nécessaire à la programmation dynamique (le coût en un point ne dépend que de γ et du coût au point précédent), mais elle doit également incorporer toutes les informations que l'on connaît sur un *bon contour* : contraste entre les deux plages séparées par le contour bien sûr, mais aussi, longueur et courbure, voire écart à un contour modèle dans certains cas. C'est dans cette fonction de coût que l'on rajoute l'information qui a manqué au détecteur local placé en amont pour détecter le contour.

11.5.2 Les automates

Ces solutions, au contraire des précédentes, ne proposent que des solutions très sous-optimales, fondées sur de seuls critères locaux (il n'y a pas d'optimum global comme par programmation dynamique). Elles sont particulièrement rapides de mise en œuvre [Giraudon, 1987], mais peuvent parfois faire diverger la combinatoire si l'on

gère mal la prolifération des chemins par ramification. Les résultats d'une détection par masquage adapté sont les candidats idéaux à ces techniques de fermeture puisqu'ils fournissent des directions de recherche.

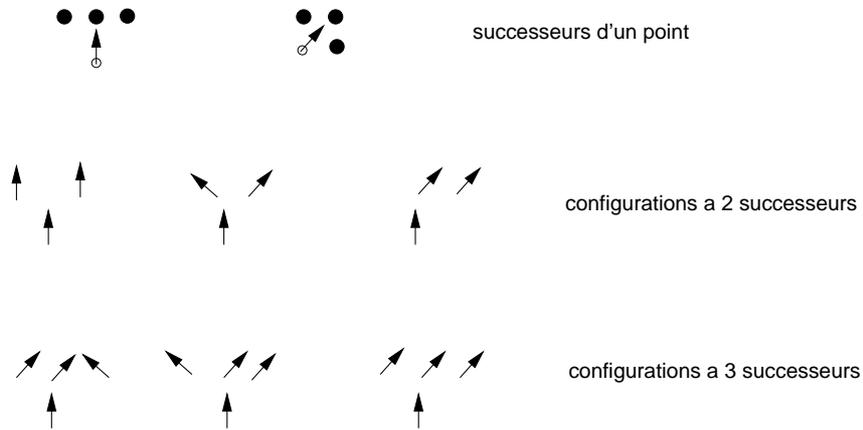


FIG. 11.15 – Configurations retenues par l'automate de Nevatia et Babu pour orienter un contour vers 3 successeurs au plus.

Les diverses méthodes se distinguent par le choix des successeurs d'un point (parmi des configurations pré-définies dans [Robinson, 1977, Cocquerez et Philipp, 1996, Nevatia et Babu, 1980], par prédiction de type Kalman) et le mode de balayage de l'image (par exemple 1 balayage dans [Nevatia et Babu, 1980], 2 balayages, l'un horizontal l'autre vertical dans [Cocquerez et Philipp, 1996], traitement de listes chaînées dans [Giraudon, 1987]). Il est également important de définir des fonctions de mérite pour valider les chaînes ainsi trouvées, car les automates, comme la programmation dynamique, trouvent toujours un contour, mais ne garantissent pas toujours qu'il correspond véritablement au contenu de l'image.

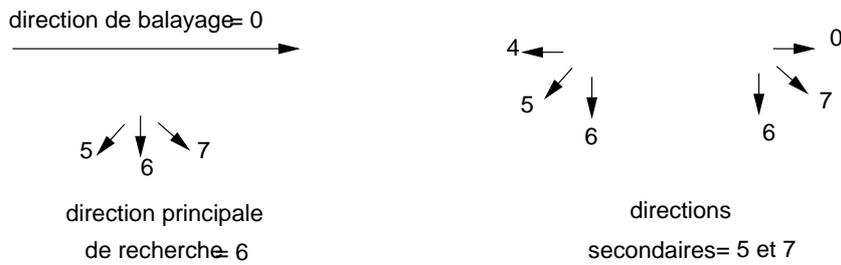


FIG. 11.16 – L'automate de Cocquerez lors d'un balayage horizontal utilise les directions 5, 6 et 7 de Freeman (cf. chapitre 14, section 14.4) comme directions principales. Lorsque l'une d'elle a été retenue, les directions adjacentes (par exemple le 4 et le 6 pour la direction principale 5) sont utilisées pour élargir la recherche. Cocquerez propose de faire 2 balayages successifs, l'un horizontal, l'autre vertical.

Chapitre 12

La segmentation par régions

Chapitre rédigé par Henri MAÎTRE

L'approche duale de la détection des contours pour la décomposition d'une image en ses formes élémentaires est l'approche par régions. Elle repose sur la recherche de zones possédant des attributs communs, soit de luminosité, soit, plus rarement, de textures (nous n'aborderons pratiquement pas ce second point dans ces lignes, mais au chapitre 13).

Nous allons voir ici, tout d'abord, les méthodes utilisant l'histogramme, qui sont souvent très proches des méthodes de classification conventionnelles, mais dans lesquelles serait renforcé l'aspect de cohérence de zone. Nous verrons ensuite les méthodes pyramidales, prototypes des méthodes de subdivision hiérarchique descendante, puis les méthodes de croissance de régions, qui s'inspirent des méthodes de conquête de l'optimisation combinatoire, et les méthodes de partage et réunion qui tirent profit simultanément des avantages des deux méthodes précédentes.

Enfin, nous donnerons très brièvement un aperçu de méthodes à base de règles qui se proposent de combiner approches par contours et approches par régions dans un seul vaste programme de segmentation.

12.1 Les méthodes sur histogramme

Ces méthodes sont de mise en œuvre assez simple et de performances souvent réduites car elles ne tirent pas profit de l'aspect spatial de l'information d'image. Elles sont recommandées dans les cas suivants :

1. lorsque les images présentent des classes évidentes : documents écrits ou schémas en noir et blanc ou en couleur, objets très contrastés (par exemple cellules d'une biopsie ou avion sur un ciel), etc.
2. lorsque les images sont définies sur de nombreux canaux (images multi- ou hyper-spectrales), ce qui enrichit l'information portée par l'histogramme.

L'idée générale de ces méthodes consiste à isoler des pics de l'histogramme. À une dimension on procède donc à des seuillages ou des multi-seuillages. À n-dimensions on procède à des classifications [Dubuisson, 1990].

12.1.1 Avec apprentissage

Seuillage bayésien

Si l'on dispose d'une connaissance sur les classes que l'on recherche (en particulier la probabilité d'occurrence d'un niveau de gris pour chaque classe, et la probabilité a priori des classes), alors on peut aisément se replacer

dans les conditions de la théorie bayésienne de la décision [Dubuisson, 1990, Duda et Hart, 1973] et choisir les seuils qui minimisent le coût des fausses erreurs.

Pour des histogrammes à 1 dimension et pour 2 populations X et Y , en dénotant par :

- $P(X)$ et $P(Y)$ les probabilités à priori des classes X et Y (sous la contrainte $P(X) + P(Y) = 1$),
- $P(n/X)$ et $P(n/Y)$ les probabilités conditionnelles qu'un pixel de la classe X ait un niveau de gris n (et idem pour Y),
- C_X et C_Y les coûts des mauvaises classifications de X et Y ,

en supposant ces grandeurs toutes connues, et sous l'hypothèse de stationarité de l'image (c'est-à-dire que les propriétés statistiques sont invariantes en tout point de l'image), le seuil optimal est défini comme celui minimisant le coût global :

$$\Gamma(s) = C_x \int_s^N P(X) \cdot P(n/X) dn + C_y \int_0^s P(Y) \cdot P(n/Y) dn \quad (12.1)$$

Ce seuil est obtenu en testant pour tout n le rapport de vraisemblance $\Lambda(n)$ face au seuil s :

$$\Lambda(n) = \frac{P(n/Y)}{P(n/X)} \quad \text{et} \quad s = \frac{C_y P(Y)}{C_x P(X)}$$

Seuillage de Neyman-Pearson

Une autre décision peut se faire selon le critère de Neyman-Pearson. Dans ce cas on s'intéresse en particulier à une classe (ici on choisit X). On définit la probabilité de fausse alarme pour cette classe :

$$P_f = \int_0^s P(Y) P(n/Y) dn$$

On maximise alors la probabilité de détection pour une probabilité de fausse alarme donnée. Dans ce cas on considère que la fausse alarme est l'erreur la plus grave. On fixe sa probabilité à une valeur acceptable : $P_f = \alpha$. La probabilité de détection de X est donnée par :

$$P_d = \int_0^s P(X) P(n/X) dn$$

Le seuil de Neyman-Pearson est donné par la méthode du lagrangien :

$$L = P_d - \lambda(P_f - \alpha) \quad (12.2)$$

$$= \lambda\alpha - \int_0^s [P(n/Y)P(Y) - P(n/X)P(X)] dn \quad (12.3)$$

et la décision se fait en comparant le rapport de vraisemblance $\Lambda(n)$ à α .

Cette décision est moins sensible aux probabilités à priori et conduit en particulier à des décisions plus proches de nos choix intuitifs que la décision bayésienne dans le cas où un événement est très rare.

D'autres critères existent bien sûr qui s'appuient sur d'autres choix statistiques (par exemple minimum d'entropie).

12.1.2 Seuillage sans apprentissage

Lorsque l'on dispose du seul histogramme pour extraire des classes, on recherche généralement des modes de cet histogramme qu'on isole par des seuillages au creux des vallées. Souvent les histogrammes sont trop irréguliers et il convient alors de les filtrer, soit par des fenêtres glissantes, soit par des équations de diffusion (éventuellement isotropes). De nombreuses règles empiriques ont été proposées pour choisir les seuils automatiquement qui ne sont guère généralisables [Kittler et al., 1985, Weszka, 1978].

12.1.3 Méthodes de classification

Disposant d'un histogramme, éventuellement multidimensionnel, la plupart des techniques de classification s'appliquent à sa segmentation. Les plus utilisées sont :

- les techniques de nuées dynamiques (*k-means*) qui procèdent alternativement en classifiant au plus proche voisin le nuage des points, selon une distance à des noyaux donnés, puis en estimant la position des meilleurs noyaux de ces classes ainsi obtenues. Il est important pour cette méthode de disposer du nombre de classes recherchées. Si l'on ne le connaît pas, on choisit souvent des critères entropiques (comme le critère d'Akaike ou le critère de Hannan et Quinn [Olivier et al., 1997]) qui permettent d'évaluer des classifications obtenues avec des nombres de classes différents¹.
- les réseaux neuromimétiques et en particulier les cartes de Hopfield.
- et, si l'on dispose d'un certain nombre d'échantillons déjà classés, les plans, ou les courbes, séparateurs, ainsi que les k-plus-proches voisins.

Dans les espaces de grande dimension (imagerie hyperspectrale par exemple), on peut avoir intérêt à réduire tout d'abord la dimension des espaces pour éviter d'avoir à estimer des probabilités très faibles. On peut le faire par ACP (analyse en composantes principales) ou par analyse de Karhunen-Loeve. Ces deux transformations sont identiques à une normalisation près des axes. Elles procèdent en projetant le nuage de points sur le sous-espace construit à partir d'un nombre limité des vecteurs propres de la matrice de covariance des données.

Dans les espaces de dimension p , les distances utilisées entre deux nuages de points caractérisés par leur moyenne (vectorielle) m_i et leur matrice de covariance Γ_i sont, par ordre de complexité croissante [Fukunaga, 1990, Zhang et Modestino, 1990] :

- la distance euclidienne qui pondère également toutes les variables :

$$(m_1 - m_2)^t (m_1 - m_2)$$

- la distance de Mahalanobis (par exemple du nuage 2 par rapport au nuage 1), qui rend compte de l'orientation des inerties des nuages dans l'espace :

$$(m_1 - m_2)^t \Gamma_2^{-1} (m_1 - m_2)$$

- la distance de Bhattacharyya, qui permet de distinguer des lois de mêmes moyennes, mais de variances différentes. :

$$1/4(m_1 - m_2)^t (\Gamma_1 + \Gamma_2)^{-1} (m_1 - m_2) + \frac{1}{2} \ln \frac{\frac{1}{2} |\Gamma_1 + \Gamma_2|}{\sqrt{|\Gamma_1| |\Gamma_2|}}$$

12.1.4 Sélection sur l'histogramme et dans l'image

Le défaut des approches par classification est de complètement négliger les relations spatiales entre pixels, pour ne s'attacher qu'à leurs propriétés de radiométrie. Pour pallier ce défaut, dans une approche proposée dans [Ohlander et al., 1978] et souvent reprise, on procède de façon itérative dans l'espace image et sur l'histogramme :

1. sur l'histogramme on sélectionne un mode isolé,
2. parmi les zones de l'images contribuant à ce mode on sélectionne la zone connexe la plus grande.

On itère ce processus, soit en subdivisant à nouveau l'histogramme de la zone retenue, soit en s'occupant d'un autre mode de l'histogramme original.

12.1.5 Sélection sur histogramme et régularisation

Mais si l'on souhaite améliorer les propriétés spatiales des zones obtenues par sélection de modes sur l'histogramme ou par classification, l'une des méthodes les plus appropriées est de modéliser le problème par un champ

¹On note cependant avec [Olivier et al., 1997] que ces modèles ont tendance à sur-évaluer le nombre de classes trouvées.

markovien (voir chapitre 7). On considère que les régions forment une partition sur l'image (cf. section 12.2). Chaque région est représentée par une fonction caractéristique et identifiée par une étiquette $\lambda_i \in E$. Le champ d'étiquettes est un champ caché à l'utilisateur qui n'observe que la réalisation bruitée de l'image $f(x, y)$ en chaque pixel, et qui va essayer d'estimer la meilleure distribution des étiquettes λ_i connaissant les n (généralement selon le critère du Maximum A Posteriori de Λ (MAP)).

Dans la formalisation markovienne on passe d'une représentation probabiliste à une représentation en énergie (généralement au moyen d'une formalisation bayésienne de la probabilité à postériori). Très souvent l'énergie (représentant la probabilité à postériori de la classe sachant l'observée) est constituée de deux termes : l'un traduit la similarité de la classification aux données que l'on observe, c'est le terme d'attache aux données (probabilité des observées conditionnellement aux classes), le second exprime les a priori que nous avons sur les classes (par exemple classes compactes, aux contours réguliers ou de forme prédéfinie, dans des relations de voisinage particulières, probabilité à priori de classe, etc.). Des hypothèses peuvent également être faites sur les distributions des niveaux de gris (par exemple lois gaussiennes sur les régions), des étiquettes (dépendance ou indépendance spatiale), ainsi que des distributions conditionnelles des niveaux de gris sachant les étiquettes.

Dans la démarche la plus simple on utilise alors souvent un terme d'attache aux données de la forme :

$$U_0 = (f(x, y) - \mu_i)^2$$

où $f(x, y)$ est le niveau de gris du pixel, μ_i le niveau de gris de la classe λ_i dans laquelle on a placé le pixel (x, y) . Un tel terme est équivalent à une probabilité gaussienne d'appartenance à la classe λ_i :

$$p(f(x, y)|\lambda_i) \propto \exp(-U_0) = \exp\left[-\frac{(f(x, y) - \mu_i)^2}{2\sigma_i^2}\right]$$

Le terme de régularisation est calculé sur les cliques $\{c\}$ constituant le voisinage que l'on souhaite donner à (x, y) (cf. figure 7.1). Il prend en compte la compatibilité $\Phi(\lambda_i, \lambda_k(x', y'))$ entre l'étiquette λ_i attribuée à (x, y) et celle $\lambda_k(x', y')$ attribuée à (x', y') :

$$E_1 = \sum_{(x', y') \in \{c\}} \Phi(\lambda_i, \lambda_k(x', y'))$$

Par exemple, dans le cas le plus simple, on peut attribuer une énergie négative si $\lambda_i = \lambda_k$ et positive dans le cas contraire (modèles de Potts et d'Ising, voir chapitre 7). Des modèles plus complexes peuvent prendre en compte des compatibilités plus subtiles (classes plus ou moins proches). On recherche alors un minimum de $U = U_0 + \theta U_1$ par toutes les techniques usuelles d'optimisation itératives (ICM, recuit simulé, etc.). Le coefficient θ permet de pondérer les effets de la régularisation par rapport à ceux de la classification. Pour un θ fort, on obtiendra des classes très compactes et grandes au détriment d'une bonne séparation des classes sur l'histogramme [Derin et Elliott, 1987b, Lakshmanan et Derin, 1989, Heitz et al., 1994, Won et Derin, 1992].

Dans d'autres approches, on peut se dispenser d'introduire des étiquettes de classes en introduisant explicitement un processus de bord [Geman et Geman, 1984b], processus b discret binaire (0 ou 1) ou continu (entre 0 et 1) qui prend, généralement, ses valeurs entre les pixels. Dans le cas binaire, lorsqu'il vaut 1 le voisin concerné n'a plus d'effet sur le site, selon la formule :

$$U_1 = \sum_{(x', y') \in \{c\}} (1 - b) \cdot \Phi(g_i, g_k(x', y'))$$

C'est donc l'équivalent d'un contour introduit dans l'image. Mais on pénalise usuellement l'introduction d'un contour par un terme de la forme :

$$U'_1 = \sum_{(x', y') \in \{c\}} b$$

ou par un coût de type MDL (Minimum Description Length) (cf. section 12.4).

Enfin, on peut également supprimer le processus de bord en le « cachant » à l'intérieur du potentiel U_1 . Pour cela on choisit une fonction Φ qui sature pour des valeurs fortes des différences entre niveaux de gris de sites voisins. On perd alors, avec la convexité de la fonctionnelle d'énergie, des bonnes propriétés de convergence du champ markovien et l'optimisation devient plus complexe [Blake et Zisserman, 1987, Black et Rangarajan, 1996].

12.2 Les méthodes par transformation de régions

Les méthodes que l'on va examiner maintenant s'appuient toutes sur la notion de **prédicat** et sur celle de **partition**.

Un prédicat \mathcal{P} est une proposition logique dont la valeur dépend de son argument. Nous nous intéresserons aux prédicats sur les régions \mathcal{R}_i de l'image.

Le prédicat de base de la segmentation est : *la région \mathcal{R}_i est homogène*. Pour vérifier le prédicat, parmi les arguments les plus utilisés en segmentation d'images, nous retiendrons les suivants :

- le contraste sur la région : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \max_{\mathcal{R}_i}[f(x, y)] - \min_{\mathcal{R}_i}[f(x, y)] < \sigma$
- l'écart-type de la région : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \sqrt{\frac{1}{N} \sum_{\mathcal{R}_i} (f(x, y) - m)^2} < \sigma$ (avec $N = \text{Card}(\mathcal{R}_i)$ et $m = \frac{1}{N} \sum_{\mathcal{R}_i} n(x, y)$),
- la distance interquartile sur la région, c'est-à-dire la distance séparant les 25 % inférieurs des 25 % supérieurs de l'histogramme,
- les différences limitées : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff \forall (k, j) \text{ voisins}, |f(k) - f(j)| < \sigma$,
- l'entropie : $\mathcal{P}(\mathcal{R}_i) = \text{vrai} \iff - \sum_{\mathcal{R}_i} p(f) \log(p(f)) < \sigma$.
- etc.

Un autre prédicat également utilisé est : *la région \mathcal{R}_i est distincte de ses voisines*. On utilise des arguments qui mettent en jeu les différences de valeurs moyennes, les distances inter-médianes, les contrastes moyens ou minimums aux frontières, etc.

Une partition Π est un ensemble de sous-ensembles \mathcal{R}_i (appelées régions) de l'image, vérifiant :

$$\Pi \iff \begin{cases} \forall i, j \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \\ \bigcup_i \mathcal{R}_i = \text{support}(\text{image}) \\ \forall i \quad \mathcal{R}_i \neq \emptyset \end{cases}$$

Lorsque toutes les régions d'une partition vérifient le prédicat, on dit que la partition le vérifie. La partition triviale où tous les pixels sont dans des régions différentes de cardinal 1 vérifie tout prédicat qui peut se vérifier sur une image. Tout prédicat qui ne peut être vérifié par la partition triviale est impossible. On connaît donc au moins une partition qui vérifie tout prédicat non impossible.

Il existe bien sûr un très grand nombre de partitions d'une image et il existe également généralement un très grand nombre de partitions qui vérifient le prédicat (il suffit, partant d'une telle partition, de subdiviser n'importe quelle région pour obtenir une nouvelle partition vérifiant le prédicat). On ne sait pas trouver toutes les partitions vérifiant le prédicat (la combinatoire est généralement inaccessible) mais on ne sait non plus généralement pas choisir entre des partitions vérifiant un prédicat. Des critères empiriques peuvent être utiles :

- le cardinal de la partition (à minimiser),
- la taille de la plus petite région (à maximiser),
- une « distance entre régions » (par exemple somme des distances entre zones adjacentes) (à maximiser).

Dans de nombreux cas par exemple, on ne recherche que des partitions maximales, c'est-à-dire telles que deux régions adjacentes ne vérifient pas le prédicat :

$$\forall (i, j : i \text{ et } j \text{ adjacents}) \Rightarrow \mathcal{P}(\mathcal{R}_i \cup \mathcal{R}_j) = \text{faux}$$

En l'absence de stratégie pour trouver la meilleure partition vérifiant un prédicat donné, nous allons décrire des stratégies de base qui sont très souvent utilisées.

12.2.1 La croissance de région

Les méthodes de croissance de région (*region growing*) n'aboutissent pas à des partitions car la propriété $\bigcup_i \mathcal{R}_i = image$ n'est généralement pas satisfaite. Partant de **germes**² on applique successivement à l'image des prédicats plus sévères que le prédicat \mathcal{P} . Ainsi on commence à associer aux germes les seuls pixels qui sont en très bon accord avec le prédicat. On réduit cette sévérité progressivement, et on se rapproche ainsi du prédicat objectif. La décision d'associer un pixel à une région se fait alors le plus souvent sans ambiguïté à moins que ses « distances » à deux régions soient égales (et en ce cas un choix quelconque est peu décisif). Plus importante est la décision de regrouper 2 régions qui sont adjacentes et vérifient le prédicat. Il a été montré qu'il convient alors d'être assez sévère pour fusionner des régions et qu'il est souvent préférable de traiter ce point lors d'une étape ultérieure en acceptant donc une **sur-segmentation** de l'image, plutôt qu'une fusion abusive qui ne serait plus récupérable.

Les méthodes de croissance de région cessent souvent de créer de nouveaux germes bien avant que le prédicat ne vaille \mathcal{P} et rejettent les régions de cardinal 1. Ainsi tous les pixels ne se retrouvent pas dans des régions (cf. figure 12.1) [Adams et Bishof, 1994, Beaulieu et Goldberg, 1989, Chang et Li, 1994].

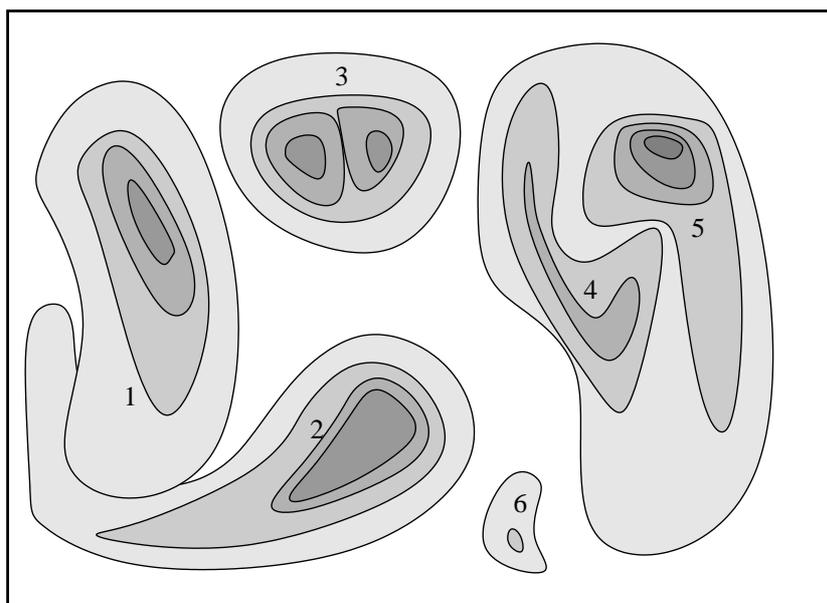


FIG. 12.1 – Croissance de région : 6 régions ont été obtenues par 4 prédicats successifs. La zone 6 n'a été créée qu'au second prédicat. La zone 3 regroupe 2 sous-régions qui ont été fusionnées car leur union vérifie le postulat et des critères annexes sur la forme résultante sont vérifiés. Ce n'est pas le cas de 1 et 2.

12.2.2 Le partage de région

Dans une démarche par partage de région (*region splitting*), on part de l'image entière (à laquelle généralement le prédicat \mathcal{P} ne s'applique pas). On appelle \mathcal{R} cette région. On lui applique plusieurs divisions δ produisant de nouvelles régions \mathcal{R}_i^δ . On teste \mathcal{P} sur chaque \mathcal{R}_i^δ , et on retient la meilleure subdivision δ , c'est-à-dire :

²les germes sont souvent des régions où le prédicat est trivialement vérifié, par exemple des zones où l'image est constante

- celle qui conduit à des sous-régions vérifiant toutes \mathcal{P} ,
- ou celle qui donne le plus de sous-régions vérifiant toutes \mathcal{P} ,
- ou, si aucune \mathcal{R}_i^δ ne vérifie \mathcal{P} , celle qui fournit la meilleure valeur à un critère \mathcal{E} dit critère d'échec.

Pour δ choisi, chaque sous-région \mathcal{R}_i^δ ne vérifiant pas \mathcal{P} devient alors une région \mathcal{R} passible du traitement ci-dessus.

Les stratégies de subdivision sont peu nombreuses, on utilise généralement les 2 bipartitions régulières horizontale ou verticale (cf. figure 12.2), parfois des tripartitions régulières.

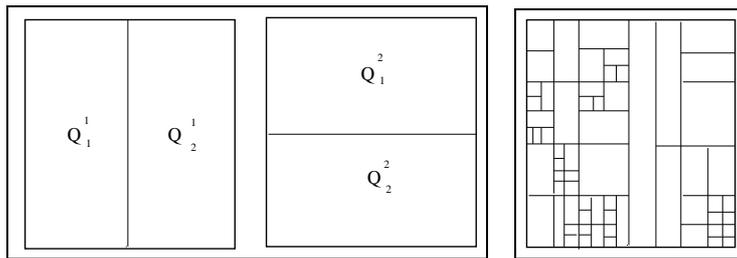


FIG. 12.2 – A gauche : partition d'une zone : on choisit entre la partition verticale ou la partition horizontale. A droite, à la fin du partage, l'image partitionnée.

Les critères d'échec \mathcal{E} sont souvent les mesures de variance ou de dynamique qui sont testées dans les prédicats \mathcal{P} . Lorsque toutes les régions ont été récursivement testées, il peut se trouver que des zones adjacentes le long d'une frontière d'ordre supérieur vérifient le prédicat. On peut alors les réunir avec les mêmes précautions que celles signalées au paragraphe 12.2.1.

Les algorithmes de partage sont mal adaptés à une mise en œuvre informatique sur machine séquentielle car les calculs effectués sur tous les pixels de la zone \mathcal{Q} ne sont généralement pas réutilisables au niveau inférieur \mathcal{Q}_i .

12.2.3 La réunion de région

Les techniques de réunion (*region merging*) prennent l'exact contre-pied des précédentes. Ce sont des méthodes ascendantes (*bottom-up*) lorsque les autres étaient descendantes (*top-down*). Les pixels sont visités systématiquement. Pour chaque carré de 2×2 pixels le postulat \mathcal{P} est testé, et s'il est accepté les pixels sont regroupés en une région. Après le parcours de toute l'image, les groupes de 2×2 régions se voient appliquer le même test et, éventuellement, les mêmes conséquences (réunion en une région de niveau 2) [Zhu et Yuille, 1996].

Les tests de réunion de région sont fréquemment faits sur des tests statistiques [Saporta, 1978]. On se place souvent dans l'hypothèse de bruits gaussiens sur des fonctions à valeur moyenne constante. Les tests les plus utilisés sont :

- le Chi2 (χ^2),
- le test de Wilcoxon, il travaille sur les pixels triés par ordre croissant de niveaux de gris des deux régions. Pour chaque pixel de la liste issue de la région 1 on compte combien de pixels de 2 sont plus sombres. Ces nombres sont additionnés pour donner une variable U de distribution asymptotiquement gaussienne de moyenne $n_1 n_2 / 2$ et de variance $n_1 n_2 (n_1 + n_2 + 1) / 12$. On teste donc cette valeur U face à sa distribution asymptotique.
- le test de Student d'égalité des espérances, on teste la variable de Student :

$$T(n_1 + n_2 - 2) = \frac{(m_1 - m_2)\sqrt{n_1 + n_2 - 2}}{\sqrt{(n_1\sigma_1^2 + n_2\sigma_2^2)(1/n_1 + 1/n_2)}}$$

– le test de Fisher-Snedecor d'égalité des moyennes et des variances, on teste la variable de Fisher :

$$F(n_1 - 1; n_2 - 1) = \frac{n_1 \sigma_1^2 (n_2 - 1)}{n_2 \sigma_2^2 (n_1 - 1)}$$

12.2.4 Les pyramides

Afin de bénéficier des avantages des deux méthodes précédentes, Horowitz a proposé une approche par pyramide [Horowitz et Pavlidis, 1976]. L'image est représentée sur une pyramide appelée *quad-tree*, constituée de N niveaux, l'image originale étant au niveau 0. Chaque pixel au niveau ν a 4 fils au niveau $\nu - 1$. Le pixel au niveau ν est la moyenne de ses 4 fils. Un schéma plus complet a été proposé par Burt [Burt, 1981] où les passages entre niveaux se font par filtrage gaussien. Une abondante littérature a été proposée depuis pour améliorer ces filtrages par une approche en ondelettes.

La segmentation procède directement à un niveau intermédiaire (par exemple $\nu = 2$), et tous les pixels fils sont testés avec le prédicat \mathcal{P} . Si le prédicat n'est pas vérifié, le pixel considéré est étiqueté comme candidat à un partage. Sinon il est candidat à une réunion avec ses voisins. C'est donc une technique de partage et réunion, où l'on profite du passage au niveau supérieur (ici $\nu = 3$) pour accélérer la procédure (cf. figure 12.3).

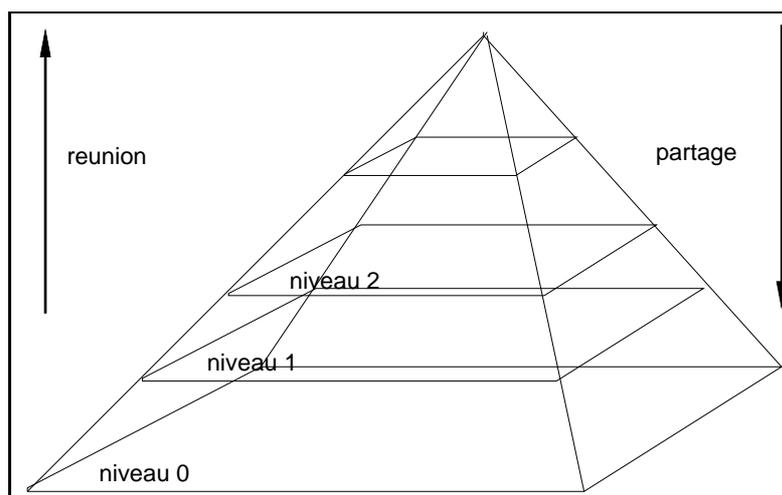


FIG. 12.3 – Méthode de pyramidage par quad-tree.

Mais la stratégie du quad-tree est trop rigide et conduit à des partitions trop régulières que les techniques de croissance ne permettent plus de rattraper. Des techniques inspirées de la pyramide ont été proposées qui donnent de bien meilleurs résultats. Ainsi, la méthode proposée dans [Suk et Chung, 1993]), autorise beaucoup plus de fusion que les simples réunions de pixels 2×2 . Douze fenêtres différentes sont autorisées, avec une priorité aux plus grandes (cf figure 12.4) (voir aussi [Chen et al., 1991]).

12.3 Les graphes d'adjacence

Les techniques par graphes d'adjacence sont beaucoup utilisées à partir de sur-segmentations (c'est-à-dire de segmentations où les zones sont subdivisées trop finement). Ces sur-segmentations sont par exemple le résultat d'un prédicat trop sévère dans la phase de segmentation ou d'un algorithme très sensible aux variations locales comme la technique des lignes de partage des eaux (cf. chapitre 6 et [Schmitt et Mattioli, 1994b]).

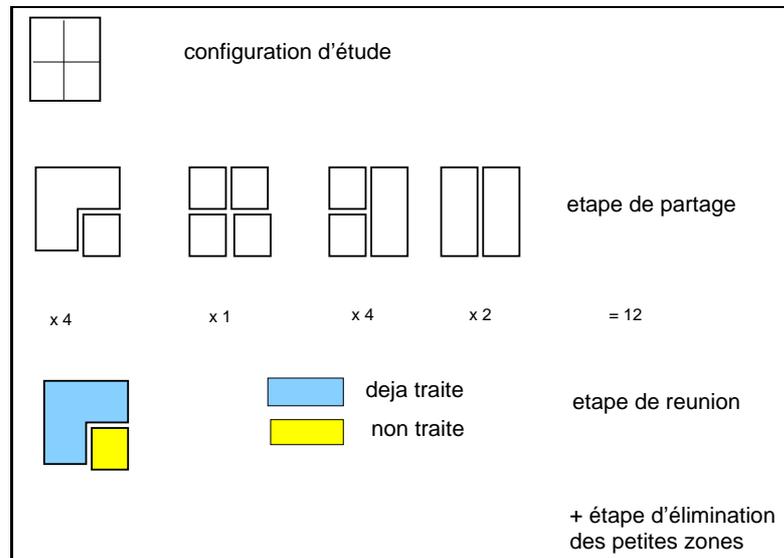


FIG. 12.4 – Les 3 étapes de la méthode de partage et réunion de Suk. Pour chaque étape il y a un prédicat différent. Celui du partage et de la réunion est généralement un prédicat de contraste, celui sur l'élimination des petites zones est un prédicat de taille. Les petites zones sont associées à la zone la plus proche en niveaux de gris.

L'idée de base consiste à plonger les régions obtenues dans une structure de graphe où une région est un nœud et un arc une relation d'adjacence. Puis on définit une fonction de similarité entre 2 nœuds. On trie tous les couples de nœuds adjacents dans une liste ordonnée. On regroupe les 2 meilleurs candidats. On remet à jour la liste et on itère.

La méthode proposée par Beaulieu et Goldberg [Beaulieu et Goldberg, 1989] est un bon exemple d'une telle technique. Le critère d'homogénéité $H(\mathcal{R}_i)$ d'une zone \mathcal{R}_i est le résidu de son approximation par un polynôme de degré d fixé :

$$H(\mathcal{R}_i) = \sum_{(x,y) \in \mathcal{R}_i} \left[f(x,y) - \sum_{p+q \leq d} a_{p,q}^i x^p y^q \right]^2$$

et plus particulièrement d'une approximation par une constante. En partant d'une partition très fine en régions très homogènes, et en acceptant progressivement des réunions de régions de moins en moins semblables, on s'approche de l'optimisation du critère global de minimum de variance sur les partitions.

D'autres schémas plus complexes sont mis en œuvre, par exemple dans [Wang, 1998] pour obtenir des segmentations meilleures.

12.4 La méthode MDL = Minimum Description Length

C'est une technique issue de la théorie stochastique de l'information [Rissanen, 1984, Rissanen, 1987, Rissanen, 1989] pour optimiser la représentation de données. Elle a été reprise de façon simplifiée dans divers domaines du traitement des images (en particulier en codage). L'idée du MDL consiste à exploiter l'analogie, au sens de la théorie de l'information, entre longueur minimale de description et quantité d'information. On cherche donc à optimiser le choix d'un modèle pour décrire les données conduisant à la plus courte description :

- d'une part par le choix d'un modèle qui s'adapte bien aux données,
- d'autre part par le choix d'un modèle simple qui nécessite peu de paramètres.

L'optimisation au sens du MDL procure un compromis entre ces deux termes. Dans le cas de la description d'une image, le modèle doit décrire les contours des régions et approcher au mieux l'image entre les contours.

On se place donc dans le domaine de la représentation paramétrique des données. On dispose d'un modèle dépendant de paramètres et on recherche le meilleur jeu de paramètres pour approcher les données. Soit $\phi = (\phi_1, \dots, \phi_k)$ le vecteur à k composantes des paramètres. Soit x l'observée (l'image). Dans le cadre de l'estimation au maximum de vraisemblance, on cherche le meilleur $\phi \in \mathbb{R}^k$ qui maximise $p(x/\phi)$. Cela revient à maximiser la log-vraisemblance :

$$\mathcal{L}(x/\phi) = -\log p(x/\phi) \quad (12.4)$$

Si l'on dispose de plusieurs modèles, chacun ayant des jeux différents de ϕ_i , la transmission de la représentation de x nécessite la transmission d'une part des paramètres décrivant la forme dans ce modèle, d'autre part les éléments permettant de reconstruire le modèle puisqu'il existe de nombreux modèles potentiellement disponibles. On sait que selon la théorie de l'information, il existe des codes optimaux pour transmettre le premier terme en utilisant un nombre de bits égal au logarithme de la probabilité $p(x/\phi)$, c'est à dire le terme de l'équation 12.4. On devra ensuite transmettre le second terme. Pour rechercher la meilleure description au sens de la théorie de l'information, on minimisera donc une expression plus complète que celle utilisée pour la représentation au maximum de vraisemblance (équation 12.4) :

$$\mathcal{L}'(x/\phi) = -\log p(x/\phi) + \mathcal{L}(\phi) = \mathcal{L}(x/\phi) + \mathcal{L}(\phi) \quad (12.5)$$

où $\mathcal{L}(\phi)$ représente la quantité d'information nécessaire pour transmettre le modèle. Dans le cadre général de la classification ou du traitement du signal, il est parfois possible de mener une optimisation complète et explicite du MDL.

Utilisé dans le cadre de la segmentation d'image, le MDL introduit dans la segmentation un terme $\mathcal{L}'(x/\phi)$ de pénalité lors d'une représentation en régions. Il vient limiter d'une part le nombre de contours (le premier terme $\mathcal{L}(x/\phi)$), d'autre part la complexité des contours par le choix si possible d'un petit nombre de paramètres très simples à coder (le second terme : $\mathcal{L}(\phi)$).

Dans la pratique, il n'est généralement pas possible en traitement des images d'obtenir une solution explicite de l'équation 12.5 et l'on recherche très souvent des solutions itératives (par modification des lignes de contour) sous le contrôle du terme 12.5. Le MDL est utilisé en segmentation d'images de différentes façons. On exprime généralement :

1. le coût de la représentation d'une région par une fonction (par exemple constante). Ce coût s'exprime par le nombre de bits nécessaires à coder la constante plus le nombre de bits nécessaires à coder l'erreur résiduelle,
2. le coût de la représentation des contours (coût de codage de la chaîne des contours). Une chaîne de contours peut être codée de façons très diverses : Freeman (cf. chapitre 14.4), approximations polygonales, splines, processus de bords, etc. et conduira alors à des codages très différents.

Le MDL tend à procurer un compromis entre ces deux termes, c'est-à-dire à donner peu de contours réguliers mais au bon endroit. Les techniques mises en œuvre en MDL sont à base de champs de Markov (optimisation par recuit simulé) [Lee, 1998, Leclerc, 1989, Zhu et Yuille, 1996], ou d'optimisation de courbes [Darell et al., 1990, Keren et al., 1990].

12.5 L'approche de Mumford et Shah

12.5.1 Approche formelle

C'est une façon formelle de poser le problème de segmentation [Mumford et Shah, 1989] qui n'a malheureusement pas de solution exacte. Soit $f(x, y)$ l'image³, de support \mathcal{I} , on la remplace sur des domaines \mathcal{R}_i par des fonctions régulières $g_i(x, y)$ qui se rejoignent le long de contours Γ_j (cf. figure 12.5). La fonctionnelle à minimiser se compose de :

$$U(\Gamma, g, f) = \mu^2 \iint_{\mathcal{I}} (f(x, y) - g_i(x, y))^2 dx dy + \iint_{\mathcal{R}_i} \|\nabla g_i(x, y)\|^2 dx dy + \nu \int_{\Gamma_i} dl \quad (12.6)$$

On montre aisément que les 3 termes sont simultanément nécessaires sous peine de convertir l'expression en une formule triviale. Mumford et Shah ont déduit des propriétés particulières des solutions Γ_j et g , que nous examinons ci-dessous.

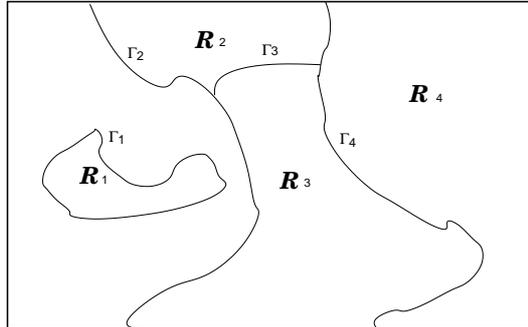


FIG. 12.5 – Approche de Mumford et Shah : on recouvre chaque domaine \mathcal{D}_i d'une fonction continue, on cherche à minimiser la distance entre la fonction de représentation et l'image, ainsi que la longueur des côtés.

On sait résoudre ce problème dans quelques cas simples (par exemple si g_i est une constante qui ne prend que deux valeurs (généralement choisies comme +1 et -1), on retrouve le problème d'Ising qui a une solution - en théorie exacte - par recuit simulé (cf. chapitre 7)⁴.

Mumford et Shah ont montré par ailleurs que l'équation générale 12.6 peut évoluer vers deux formes limites.
– Tout d'abord la minimisation de l'énergie U_o :

$$U_o(\Gamma, f) = \frac{1}{\mu^2} U(\Gamma, f, g) = \sum_i \iint_{\mathcal{R}_i} (f - g_i)^2 dx dy + \nu \int_{\Gamma_j} dl$$

³Attention dans le texte original de Mumford et Shah, [Mumford et Shah, 1989] les rôles de f et g sont interchangés, afin de garder la cohérence de nos notations nous avons préféré désigner par f l'image et par g son approximation.

⁴On se souvient cependant que les conditions de convergence vers le maximum absolu par recuit simulé sont liées à une descente en température infiniment lente.

que l'on obtient en imposant à g d'être constante par morceaux sur les \mathcal{R}_i : $g(x, y) = g_i$ ssi $(x, y) \in \mathcal{R}_i$. Il est alors aisé de montrer que $g_i = \frac{1}{s_i} \iint_{\mathcal{R}_i} f(x, y) dx dy$ où s_i représente la surface de \mathcal{R}_i (c'est donc la valeur moyenne sur la surface).

U est la limite de U_o quand μ tend vers 0. La minimisation de U_o est un problème bien posé. Le problème d'Ising évoqué plus haut en est un sous-problème réduit au cas à 2 classes, le problème de Potts l'étend à un plus grand nombre de classes.

– Puis la minimisation de l'énergie :

$$U_\infty(\Gamma) = \int_\Gamma \left[\nu_\infty - \left(\frac{\partial f}{\partial n} \right)^2 \right] dl$$

où ν_∞ est une constante et $\frac{\partial f}{\partial n}$ mesure la composante normale à Γ du gradient de f .

Ce problème peut se ramener à un problème de géodésique : on cherche simultanément à déterminer les trajectoires Γ_i de longueur minimale (terme en ν_∞) et à maximiser le gradient transverse de f le long de Γ_i ⁵. Ce problème est généralement mal posé si les Γ sont quelconques ; sur des contours de forme restreinte, le problème peut être bien posé.

On peut montrer que U_∞ est la limite de U quand μ tend vers l'infini.

On montre que si les Γ sont fixés (c'est-à-dire si l'on a fait par ailleurs une détection de contours), alors la minimisation de U est la recherche du minimum d'une fonction quadratique définie positive possédant un minimum unique. Elle est solution de :

$$\Delta g = \mu^2(g - f) \quad (12.7)$$

avec pour conditions aux limites :

$$\frac{\partial g}{\partial n} |_{\Gamma_j} = 0$$

Si les contours Γ_j sont des courbes régulières (généralement on recherche des solutions parmi les courbes de classe \mathcal{C}^2)⁶, on déduit alors les propriétés suivantes :

- les contours Γ_j ne peuvent posséder que des points singuliers de deux types :
 - soit des points triples où 3 courbes se joignent à 120° ,
 - soit des points extrémités d'où démarre une courbe unique (démarrage d'un pli) ;
- les contours se rattachent aux bords de l'image selon des angles droits ;
- par ailleurs, les fonctions g ont une dérivée normale aux Γ_j horizontale le long des Γ_j .

12.5.2 Les variantes autour de la formulation de Mumford-Shah

Des variantes du problème de Mumford et Shah ont été proposées qui abordent le problème de nombreuses façons. Par exemple des méthodes déterministes issues de la mécanique et de la résistance des matériaux ont été proposées pour apporter des solutions en termes de *membranes* ou de *plaques minces* [Blake et Zisserman, 1987]. On cherche alors à résoudre l'équation 12.7. Ces solutions conduisent à des équations données par :

– dans le cas d'une membrane :

$$U(g, f) = \alpha_1 \cdot \sum_i \iint_{\mathcal{D}_i} (f(x, y) - g_i(x, y))^2 dx dy + \alpha_2 \cdot \sum_i \iint_{\mathcal{D}_i} (|\nabla g_i(x, y)|)^2 dx dy \quad (12.8)$$

⁵On a également abordé ce problème lorsqu'on a évoqué les modèles géodésiques actifs (cf. paragraphe 11.4.4)

⁶Une courbe de classe \mathcal{C}^n possède n dérivées continues.

– dans le cas d'une plaque mince :

$$U(g, f) = \alpha_1 \cdot \sum_i \iint_{\mathcal{D}_i} (f(x, y) - g_i(x, y))^2 dx dy + \alpha_2 \cdot \sum_i \iint_{\mathcal{D}_i} (\Delta g_i(x, y))^2 dx dy \quad (12.9)$$

Il est également possible de prendre en compte dans ce second cas un terme de torsion (dérivées croisées) dans le terme de régularisation.

Dans les 2 cas les discontinuités Γ sont fixées (par exemple par un détecteur de contour préalable).

La résolution de ces problèmes (connaissant la position des contours on recherche la valeur en tout point de g_i) se fait traditionnellement par des techniques convolutionnelles. À chaque fonctionnelle est attachée une réponse impulsionnelle - ou fonction de Green - G . C'est-à-dire que, sur chaque domaine, \mathcal{D}_i , $g(x, y)$ est de la forme :

$$g(x, y) = \int \int_{\mathcal{D}_i} G(x, y, u, v) \cdot f(u, v) du dv$$

Par exemple en l'absence de contours, la membrane (équation 12.8) a pour fonction de Green K_o , fonction de Bessel modifiée de seconde espèce :

$$G(\rho = \sqrt{(x - x')^2 + (y - y')^2}) = \frac{1}{2\pi\lambda^2} K_o\left(\frac{\rho}{\lambda}\right)$$

équivalente à $\log(1/\rho)$ près de Γ et équivalente à $\frac{1}{\rho} \exp(-\rho)$ pour ρ grand. C'est la solution de l'équation sans second membre issue de l'équation 12.7 :

$$g - \frac{\alpha_2}{\alpha_1} \cdot \nabla^2 g = \delta(x, y)$$

La plaque mince (équation 12.9) a pour fonction de Green :

$$G(\rho) = \frac{i}{4\pi\mu^2} \left[K_o\left(\frac{\rho\sqrt{i}}{\mu}\right) - K_o\left(\frac{\rho}{\mu\sqrt{i}}\right) \right]$$

et elle est solution de l'équation :

$$g + \mu^4 \nabla^4 g = \delta(x, y)$$

avec $\mu^4 = \frac{\alpha_2}{\alpha_1}$.

Dans le cas où il n'y a pas de discontinuités de g le long des Γ , le problème est assez simple mais peu satisfaisant en termes de segmentation d'image [Grimson, 1981, Terzopoulos, 1983]. La solution est obtenue par des techniques de relaxation [Grimson, 1981] ou de relaxation multi-grilles [Terzopoulos, 1983]⁷.

Dans le cas discret on étend les solutions continues par des approches par éléments finis permettant de calculer de façon discrète les dérivées (on est alors dans un schéma très semblable à celui adopté pour les contours actifs vus au paragraphe 11.4).

On définit un processus de bord b booléen et on calcule les divers termes de l'énergie en fonction des valeurs $b = 0$ (pas de bord) ou $b = 1$ (présence d'un bord) de ce processus.

Si l'on ne connaît pas la position exacte des contours, on met en œuvre des techniques itératives et (souvent) sous-optimales qui partent de contours estimés et les améliorent. La résolution se fait par des techniques

⁷Les techniques multi-grilles sont des techniques d'optimisation qui utilisent la régularité spatiale de la solution pour trouver tout d'abord une solution exacte sur une grille d'échantillonnage très grossière, puis par un raffinement itératif de la solution, de la calculer sur des grilles de plus en plus fines pour aboutir à la résolution ultime de l'image.

de Markov avec processus de bord explicites [Geman et Geman, 1984b] - cf paragraphe 12.1.5 - ou implicites et des schémas d'optimisation adaptés (par exemple la méthode GNC) ou des techniques à base de réseaux neuro-mimétiques de Hopfield. La méthode GNC (*Graduated Non Convexity* de Zisserman [Blake et Zisserman, 1987]), est particulièrement adaptée à ce problème. Elle remplace les fonctions d'énergie non convexes à optimiser par une famille de fonctions convexes qui, successivement optimisées par une technique de gradient, permettent d'atteindre l'optimum global.

Dans ces schémas, on peut aussi introduire un terme de pénalité sur la longueur des contours comme dans l'équation 12.6. On trouve dans la littérature des formes variées de cette énergie :

1. la plus simple est celle de Mumford et Shah et s'exprime par :

$$U = \int_{\Gamma} dl$$

2. une minimisation de la courbure des contours :

$$U = \int_{\Gamma} \left(\frac{d\theta}{dl} \right) dl$$

3. une pénalisation pour les seuls forts changements de direction des contours.

Des techniques où l'on mélange des termes de dérivées premières et secondes (comme dans les contours actifs - cf. paragraphe 11.4) sont également possibles, mais réputées moins bonnes.

Chapitre 13

Les textures

Chapitre rédigé par Henri MAÎTRE

13.1 Qu'est ce qu'une texture ?

Dans le domaine du traitement de l'image et de la vision, il n'existe pas de définition satisfaisante de la texture. Les définitions mathématiques construites à partir de propriétés statistiques sont soit trop générales et imprécises soit trop restrictives pour s'adapter à la diversité des cas rencontrés. La définition que nous proposons n'est pas opérationnelle et prête sur ce point à critique. Elle s'appuie sur une constatation expérimentale : une texture est un champ de l'image qui apparaît comme un domaine cohérent et homogène, c'est-à-dire formant un tout pour un observateur. C'est cette propriété de cohérence de la texture placée dans son contexte d'être perçue comme un tout homogène par l'œil humain qui sera recherchée le plus souvent par le traiteur des images, dans le but d'isoler les textures, soit pour segmenter l'image, soit pour reconnaître des régions. La figure 13.1 illustre la diversité des textures et l'ouvrage de référence de Brodatz [Brodatz, 1966] offre une collection de textures naturelles qui constituent d'excellents exemples et sont souvent utilisées pour tester les algorithmes et les méthodes.

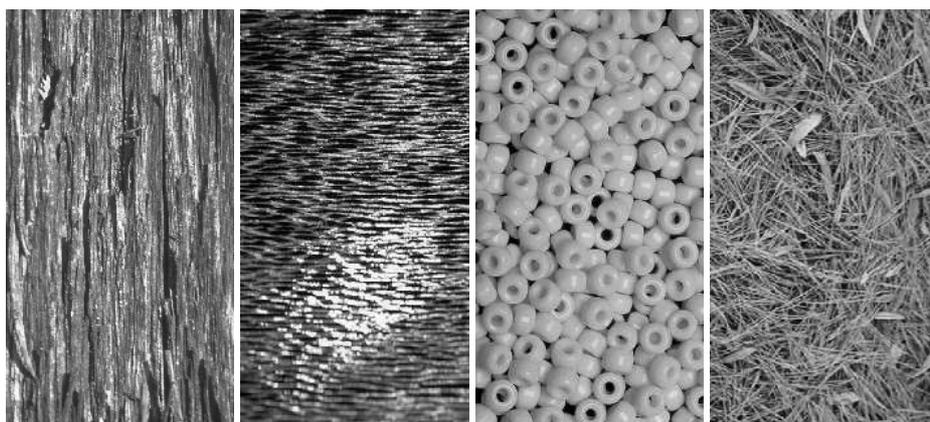


FIG. 13.1 – Quelques exemples de textures naturelles : écorce d'arbre, poil court, perles, tapis d'aiguilles de pin et de feuilles.

Les capacités du système visuel humain à discerner les textures différentes sont remarquables et très mal ex-

pliquées. Nos capacités à mémoriser et discriminer des textures sont très variables. Elles deviennent excellentes dès lors que l'on a reconnu l'objet de la vie courante dont la texture est issue. Notre mémorisation est alors remarquable et invariante à de très nombreuses transformations (changement d'échelle, d'orientation, d'éclairage, de couleur, etc.). Au contraire, des textures non reconnues (champs aléatoires créés mathématiquement par exemple, ou textures placées hors de leur contexte) sont mémorisées de façon très fugitive et sont peu robustes aux transformations. Nos capacités à distinguer des textures différentes peuvent parfois s'expliquer par les capacités de filtrage du système visuel (bande passante des voies optiques, discrimination angulaire, sensibilité aux variations locales de luminosité [Hubel et Wiesel, 1969, Julesz, 1971]). Parfois elles relèvent de mécanismes de psycho-vision et peuvent trouver des explications par exemple dans la Gestalt Theorie qui propose des mécanismes d'association (groupement perceptuel) ou de discrimination, à partir des symétries, proximités, similarités des stimulus élémentaires composant la texture [Wertheimer, 1944, Grossberg et Mingolla, 1985].

L'une des remarquables capacités du système visuel est de reconnaître des textures déjà mémorisées même sous de très fortes distorsions géométriques (vues perspectives, recouvrement de surfaces gauches, etc.). Dans ces conditions, et même en l'absence d'un contexte favorable (par exemple présence de silhouettes d'objet ou d'alignements perspectifs), le système visuel interprète la déformation de la texture, non comme une perturbation des propriétés géométriques du champ texturé, mais comme une variation de la géométrie qui supporte la texture, celle-ci restant globalement invariante dans un repère imprécisé (cf. figure 13.2). Ainsi un tronc d'arbre est vu comme une texture homogène enroulée sur un cylindre et non comme une texture plane dont les fréquences spatiales s'accroissent en approchant des bords.

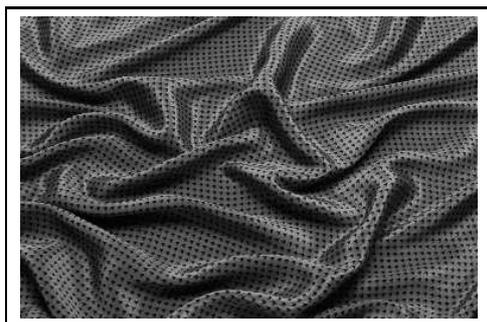


FIG. 13.2 – Lorsque la texture n'est pas homogène, on interprète naturellement les variations locales de ses propriétés statistiques comme des déformations de la surface sur laquelle elle est projetée.

Pour toutes ces raisons, (mais aussi pour simplifier considérablement la tâche du traiteur d'image), nous considérerons par la suite les textures comme spatialement invariantes, c'est-à-dire que nous négligerons généralement les effets de perspectives et les variations d'homogénéité du champ de distribution des intensités. Nous lèverons cette hypothèse au paragraphe 13.6.

13.1.1 Distribution aléatoire ou régulière ?

Une première constatation que l'on fait en examinant des textures naturelles est le rôle particulier que joue l'aléatoire dans la texture. On distingue assez naturellement deux modèles extrêmes de textures, entre lesquels se positionnent un peu toutes les textures :

1. les textures régulières, dans lesquelles la périodicité du motif est évidente : grilles, murs, tissus, etc.
2. les textures aléatoires pour lesquelles la distribution des intensités n'est l'objet d'aucune régularité apparente : sable, nuages, herbe, foule.

La première famille sera bien décrite par des approches fréquentielles ou des approches structurelles dans lesquelles on associera un motif et des règles de placement sur un pavage régulier.

La seconde approche sera plutôt décrite par des lois statistiques, des moments, une description spectrale en termes de densité de puissance, des propriétés de corrélation ou d'isotropie.

Mais une texture n'est jamais strictement périodique ni totalement aléatoire et les deux modèles se complètent naturellement. L'une des difficultés est de les mêler dans des modèles capables de s'adapter à la variabilité des textures étudiées.

13.1.2 Les échelles des textures

Une autre particularité très importante des textures est qu'elles présentent généralement plusieurs niveaux d'échelle auxquels on peut les étudier.

À petite échelle (donc pour des détails fins), on observe souvent un objet élémentaire qui constitue la base de la texture (poil, perle, aiguille de pin, sur la figure 13.1). Cette primitive peut être très régulière géométriquement (perle), ou photométriquement (aiguille de pin). Elle peut être au contraire relativement variable (personne dans une foule, caillou sur une plage, nuage). Elle possède une taille et des propriétés statistiques caractéristiques (qui joueront sur sa fonction de corrélation, son isotropie, ...). Cette composante élémentaire de la texture a tendance naturellement à disparaître lorsqu'on observe le champ global de la texture. Elle est intégrée à la texture.

À plus grande échelle (donc pour une analyse plus grossière), la texture apparaît comme une juxtaposition plus ou moins régulière des motifs ci-dessus. Cette juxtaposition possède ses propres lois d'isotropie, de périodicité, de régularité et induit donc d'autres lois statistiques (corrélation, densité de puissance) qui se combinent à celles du motif de base.

Une bonne analyse de texture donnera des informations sur ces deux composantes également.

13.1.3 Analyse ou synthèse

Un effort très important a été fait dans le domaine de la synthèse des images pour créer artificiellement ou copier des textures naturelles : forêts ou cultures pour des simulateurs de vols, bois, tissus, métaux pour des représentations d'intérieur, murs, toitures, eau, pour des jeux vidéo, etc. Ces travaux se distinguent notablement de ceux conduits en traitement d'image qui visent plutôt à extraire des paramètres discriminants et robustes permettant de séparer des textures différentes. Mais ces travaux se rejoignent en ce qu'ils contribuent simultanément à une meilleure connaissance et compréhension des textures à travers la boucle « analyse/synthèse » qui a fait ses preuves en reconnaissance des formes. Dans la suite, nous empruntons les méthodes aussi bien à l'analyse qu'à la synthèse des images en précisant lorsque c'est nécessaire les limites des méthodes dans l'une ou l'autre application.

13.2 Modèles de texture

13.2.1 Un modèle biologiquement plausible

Compte tenu du rôle important de la perception humaine dans la définition même de texture, des modèles de textures ont été proposés, s'inspirant de ce que l'on connaît aujourd'hui de cette perception. Ces travaux s'appuient sur les mécanismes d'adaptation du système visuel aux fréquences spatiales et aux orientations [Hubel et Wiesel, 1969, de Valois et al., 1982] ainsi que sur des expérimentations sur les effets de masquage psychovisuel [Phillips et Wilson, 1984].

On trouve par exemple dans [Bergen et Landy, 1991] un tel modèle. On décompose tout d'abord le flot optique en une pyramide (cf. paragraphe 12.2.4) par une succession de filtrages gaussiens et de sous-échantillonnages adaptés (semblables aux pyramides de Burt et Adelson [Burt, 1981]). Chaque niveau de la pyramide est alors filtré par 4 filtres qui produisent une dérivée seconde directionnelle selon l'horizontale, la verticale et les deux diagonales. On procède ensuite à une mesure d'énergie dans chaque image par une intégration locale des sorties

précédentes et l'on soustrait ces énergies, dans un même niveau de résolution, entre orientations voisines (pour exprimer le contraste relatif entre directions). On procède enfin à une étape de normalisation pour tenir compte d'une sorte de gain variable de contrôle de contraste vérifié expérimentalement.

Un tel schéma de fonctionnement du système visuel humain est très schématisé et controversé. Il existe d'autres schémas tout aussi intéressants [Mayhew et Frisby, 1978], mais il a le mérite de comporter la plupart des étapes que nous reverrons plus loin pour analyser les textures.

13.2.2 Modèles stochastiques

Ces modèles, au contraire, mettent l'accent sur la distribution statistique des pixels et sur leur dépendance spatiale. La texture est alors considérée comme la réalisation d'un processus aléatoire gouverné par ses lois. Nous avons vu dans les chapitres antérieurs deux modèles qui ont été abondamment utilisés pour modéliser les textures :

- le schéma booléen, et plus généralement les ensembles fermés aléatoires, qui considèrent la texture comme une distribution poissonnienne de motifs et qui est particulièrement bien analysé par les outils développés en morphologie mathématique qui permettent de définir sa capacité de Choquet [Schmitt et Mattioli, 1994b] ;
- les modèles autorégressifs et leurs dérivés, chaînes de Markov et champs de Markov, ces derniers étant aujourd'hui les plus universellement adoptés. Nous nous attarderons particulièrement sur ce point à la section 13.4 [Cross et Jain, 1983, Chellappa et Kashyap, 1985].

13.3 Analyse et reconnaissance de textures

Il existe de nombreuses revues des diverses approches de l'analyse de textures, et nous renvoyons le lecteur à trois textes qui couvrent assez bien l'évolution du domaine : [VanGool et al., 1985, Reed et du Buf, 1993, Randen et Husøy, 1999]

13.3.1 Schéma général

Le principe le plus général de l'analyse statistique des textures est le suivant.

1. On définit un voisinage \mathcal{V}_{ij} de tout pixel (i, j) , de taille et de forme appropriées.
2. Sur le voisinage \mathcal{V}_{ij} on mesure des propriétés particulières de l'image : soit π_k la mesure attachée à la configuration k parmi les N mesurées. Ces mesures seront attachées au pixel (i, j) .
3. On classe les pixels à partir du vecteur formé par les π_k par l'une des nombreuses méthodes de la reconnaissance des formes (rappelées en section 12.1.3).
4. Éventuellement, avant cette classification, et si le vecteur de mesures est trop grand, on réduit la dimension de l'espace par une sélection judicieuse des composantes les plus significatives ou par analyse en composantes principales.

La fenêtre d'analyse

La dimension du voisinage \mathcal{V}_{ij} est importante. C'est l'un des choix délicats de l'analyse des textures (cf. figure 13.3). Elle doit comprendre au moins un motif de base de la texture mesurée pour fournir des statistiques homogènes sur la texture, mais, si \mathcal{V}_{ij} est trop grand, la précision de localisation des frontières des textures détectées sera médiocre. La forme est généralement carrée. Taille et forme peuvent éventuellement s'adapter au signal à mesurer si l'on a quelque méthode pour le faire (pré-segmentation, masque de contours, etc.), mais il faut s'assurer alors que la technique de classification n'est pas perturbée par des mesures faites sur des échantillons variables.

La classification des données

Ce sont celles que nous avons vues au chapitre 12.1.3 : nuées dynamiques, plans séparateurs, réseaux neuro-mimétiques, C moyennes floues, etc. Dans l'idéal on aimerait pouvoir obtenir à partir de l'image $f(i, j)$, une nouvelle image $g(i, j)$ dont les plages seront très distinctes et pourront, idéalement, être segmentées par simple seuillage (cas de la mesure d'un seul paramètre discriminatoire) ou par des seuils multiples sur les diverses composantes. En pratique il est rare que l'on puisse séparer les nuages des mesures et il faut trouver des compromis lors des décisions. Les erreurs de classification peuvent bénéficier de techniques contextuelles de reclassification comme la relaxation ou les champs de Markov comme nous le verrons plus bas (section 13.4).

Lorsque l'on cherche non seulement à discriminer des plages de textures différentes, mais aussi à reconnaître des textures particulières, on utilise des techniques de classification supervisées : k-plus proches voisins, classificateurs bayésiens, réseaux neuro-mimétiques, algorithmes génétiques, préalablement entraînés sur des ensembles d'apprentissage [Vickers et Modestino, 1982].

La réduction des dimensions

Les diverses variables mesurées sont rarement indépendantes et le vecteur des π_k occupe rarement toutes les dimensions de N . On calcule alors la matrice Π de taille $N \times N$ et telle que $\Pi(k, l) = \langle \pi_k \pi_l \rangle$. On diagonalise Π et on ne retient que ses p plus grands vecteurs propres sur lesquels on projette toutes les mesures. La classification se fait alors dans l'espace de dimension p , ce qui réduit les calculs et permet de prendre de meilleures décisions car les composantes sont alors non corrélées.

13.3.2 Approches par mesures statistiques

Dans ces approches on mesure les propriétés statistiques autour du point (i, j) dans le voisinage \mathcal{V}_{ij} .

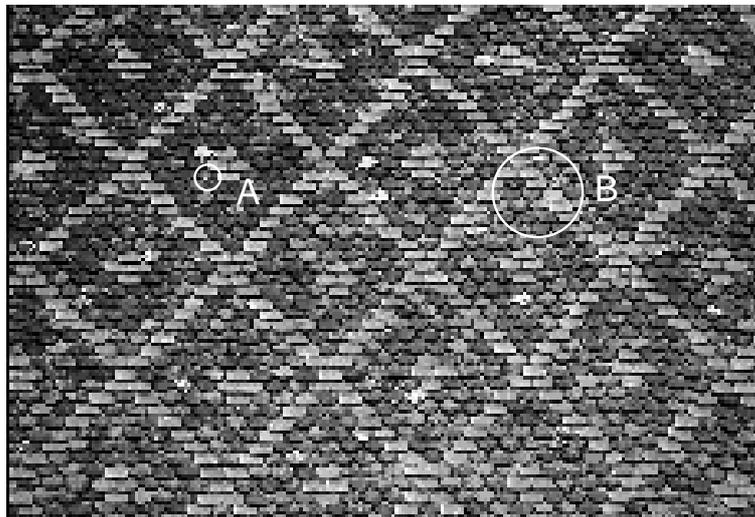


FIG. 13.3 – Le choix de la taille de la fenêtre d'analyse est important pour sélectionner les détails que l'on souhaite préserver. La fenêtre A permettra de déterminer des propriétés caractéristiques d'une tuile de ce toit et distinguera tuiles claires de tuiles sombres, la fenêtre B mesurera des propriétés moyennes sur tout le toit et ignorera les différences entre tuiles claires et sombres.

Statistiques du premier ordre

Ce sont celles qui ne prennent en compte qu'un pixel à la fois¹. Elles se déduisent de la probabilité empirique $p(n)$ du niveau de gris n dans la fenêtre ou de l'histogramme $h(n) \simeq \nu p(n)$ où ν est le nombre total de pixels dans l'image. Ce sont :

- les moments d'ordre k (non centrés) : $\mu_k = \sum_n n^k p(n)$,
- les moments centrés d'ordre k : $\tilde{\mu}_k = \sum_n (n - \mu_1)^k p(n)$, et en particulier :
 1. la moyenne : μ_1 ,
 2. la variance : $\sigma^2 = \tilde{\mu}_2$,
 3. le biais : $\gamma_1 = \frac{\tilde{\mu}_3}{\sigma^3}$,
 4. l'aplatissement (ou kurtosis) : $\gamma_2 = \frac{\tilde{\mu}_4}{\sigma^4} - 3$.
- l'énergie : $W = \sum_n |p(n)|^2$,
- l'entropie : $E = -\sum_n p(n) \log p(n)$,
- le contraste : $C = \frac{\max(n) - \min(n)}{\max(n) + \min(n)}$,
- la dynamique : $D = \max(n) - \min(n)$,
- le coefficient de variation (surtout pour les images cohérentes dont le bruit est multiplicatif : radar, images ultra-sonores) : $v = \frac{\mu}{\sigma}$,
- l'exposant de Holder, caractéristique de la dimension fractale : [Pentland, 1984].

Si les fenêtres de mesure sont petites, les statistiques mesurées sont souvent peu significatives (pour une fenêtre de 30×30 pixels et un histogramme de 256 niveaux de gris, l'occurrence moyenne des pixels est inférieure à 4 par niveau de gris). On choisit alors souvent de réduire la dynamique des images. On a montré que les propriétés de discrimination des textures se conservaient remarquablement même pour des quantifications très fortes des textures (jusqu'à 8 ou 4 niveaux de gris seulement), à la condition d'adapter la quantification. Une quantification brutale se fait en ne conservant que les bits de poids fort. Une quantification adaptée choisit de partager les niveaux de gris en fonction des statistiques de la texture :

- par moyenne et écart type par exemple (cela donne 4 classes non équiréparties),
- par médiane et distance interquartile², ce qui donne des classes équiréparties.

Les statistiques de points particuliers

On recherche dans ce cas, dans la fenêtre \mathcal{V}_{ij} , la densité moyenne de certains points d'intérêt. Les plus souvent utilisés sont :

- les maximums locaux de l'intensité,
- les points de contour (après application d'un détecteur et seuillage),

Les statistiques d'ordre élevé

Ce sont surtout les statistiques d'ordre 2 qui sont exploitées, c'est-à-dire celles qui mettent en jeu deux pixels simultanément. En effet, B. Julesz a émis une conjecture, fondée sur une très vaste étude expérimentale, que le système perceptif humain ne distingue pas les textures qui ont des statistiques similaires aux ordres 1 et 2, même si elles diffèrent aux ordres supérieurs [Julesz, 1971]. Cette conjecture a été démontrée fautive par A. Gagalowicz qui a produit des textures artificielles ne différant qu'à l'ordre 3 [Gagalowicz et Tournier-Lasserre, 1986], mais perceptivement différentes (cf. figure 13.4). La conjecture de Julesz demeure cependant un guide important pour les études sur les textures car elle est assez bien vérifiée pour les textures naturelles.

¹Attention, on désigne également parfois par premier ordre les statistiques comme la moyenne, qui ne mettent en jeu que des moments du premier ordre.

²La distance interquartile est la distance qui, dans l'histogramme, sépare les niveaux de gris des 25% de pixels les plus sombres de ceux des 25% de pixels les plus clairs. C'est une mesure robuste de l'étalement d'une loi.

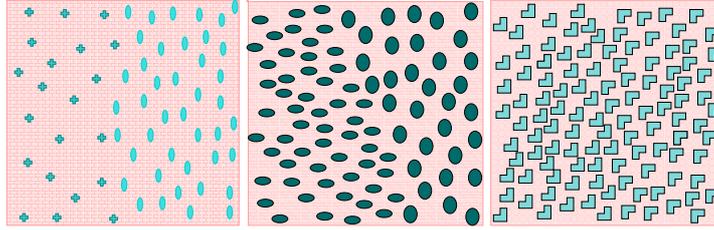


FIG. 13.4 – A gauche, les 2 textures diffèrent dès le premier ordre. Au centre, les 2 textures sont identiques au premier ordre mais diffèrent au second. A droite les 2 textures sont identiques à l'ordre 1 et à l'ordre 2, mais différentes à l'ordre 3. Ces deux textures sont distinguables visuellement, en contradiction avec la conjecture de Julesz.

Les statistiques d'ordre 2 sont bien appréhendées par la **fonction d'autocorrélation**³ de la texture, et plus particulièrement par la fonction d'autocorrélation normée centrée :

$$C_f(k, l; i, j) = \frac{\sum_{\mathcal{V}_{ij}} (f(i, j) - \bar{f}) \cdot (f(i + k, j + l) - \bar{f})}{\sum_{\mathcal{V}_{ij}} (f(i, j) - \bar{f})^2}$$

On déduit de nombreuses propriétés de cette fonction :

1. dans une direction donnée, et au voisinage de l'origine, elle peut être fréquemment approchée par une fonction régulière, par exemple :

$$C_f(k, l; i, j) \sim e^{-\alpha(i, j)|k|} e^{-\beta(i, j)|l|}$$

et les paramètres α et β sont très significatifs de la dépendance spatiale (en i et j) des niveaux de gris à l'intérieur de la texture ;

2. les deux directions d'inertie maximale $\lambda_0(i, j)$ et minimale $\lambda_1(i, j)$ extraites de $C_f(k, l; i, j)$ expriment bien les directions privilégiées de la texture et, si le rapport $\frac{\lambda_0}{\lambda_1}$ est fort, la texture est fortement anisotrope.
3. si la fonction $C_f(k, l; i, j)$ possède des maximums locaux différents de $k = 0, l = 0$, alors la texture est fortement périodique et les périodes de la fonction de corrélation permettent d'estimer de façon robuste la période de la texture.
4. enfin par transformation de Fourier, la fonction d'autocorrélation donne accès au spectre de densité de puissance (théorème de Wiener-Kinchine, cf. section 2.4), et à tous les traitements spectraux que nous verrons ci-dessous.

Plus encore que la fonction d'autocorrélation, les **matrices de cooccurrence** sont des outils adaptés à l'analyse des textures. Ces matrices sont définies pour un vecteur de translation V_i donné comme la probabilité jointe de l'amplitude d'un point M et d'un point $N = M + V_i$. Elles sont obtenues pratiquement par une estimation sur la texture, c'est-à-dire un décompte des occurrences des niveaux m en M et n en N . C'est donc un tableau généralement de taille 256×256 si la texture à 256 niveaux de gris :

$$\Gamma(m, n; V_i) = \text{proba}(f(M) = m, f(N = M + V_i) = n)$$

En raison de sa taille même, la matrice de cooccurrence est sous cette forme un outil ni très pratique, ni très fiable (les valeurs $\Gamma(m, n; V_i)$ sont des estimations très médiocres des probabilités réelles). On a donc intérêt à compresser ces matrices. Cela se fait de plusieurs façons :

- en diminuant le nombre de niveaux de gris par une quantification régulière ou adaptative, par exemple 8 niveaux de gris déterminés de façon adaptée (par moyenne et écart-type ou par médiane et distance inter-quartile) donnent souvent une bonne qualité de représentation ;

³certains auteurs préfèrent étudier le **variogramme** qui s'exprime comme $1 - C_f(k, l)$.



FIG. 13.5 – Construction d’une matrice de cooccurrence : on définit un vecteur V_i , puis on mesure l’occurrence des couples de niveaux de gris m en M et n en N . La matrice de cooccurrence ainsi calculée est $\Gamma(m, n, V_i)$.

- en tirant profit des symétries existant dans la texture pour regrouper des vecteurs V_i contribuant de façon identique à l’apparence de la texture (par exemple des vecteurs symétriques le long des axes, ou des vecteurs de même module si la texture est isotrope) ;
- en ne conservant que les seuls vecteurs significatifs ;
- enfin, on a pu montrer que quelques descripteurs des matrices de cooccurrences pouvaient très bien les représenter : position du centre de gravité, rapport des inerties, énergie, entropie, facteurs de symétrie, etc. [Chen et Pavlidis, 1979].

Les configurations particulières de pixels

On s’intéresse dans cette méthode à mesurer la probabilité d’apparition de certaines configurations particulières de pixels (cf. figure 13.6). Ce sont donc des techniques qui s’appliquent surtout à des textures binaires ou à très peu de niveaux de gris (ou rendues telles). Les mesures ainsi faites permettent de construire un vecteur d’attributs qui permettra de discriminer des textures différentes. Des configurations bien choisies pour un problème particulier peuvent être très discriminantes. Mais il est souvent difficile d’éviter que des mesures ne soient statistiquement dépendantes car les probabilités de certaines fenêtres sont évidemment dépendantes. Il convient donc souvent de réduire la dimension de l’espace des paramètres par analyse en composantes principales.

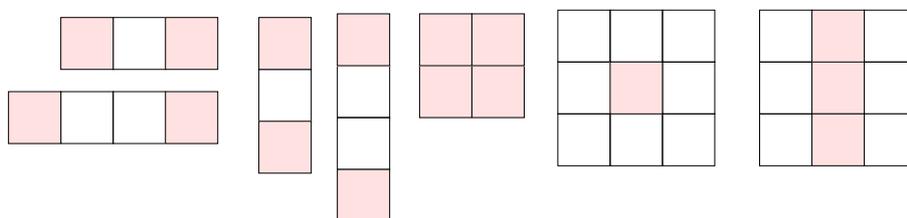


FIG. 13.6 – Quelques configurations particulières de pixels qui peuvent être mesurées pour fournir un descripteur statistique : par exemple on mesure le nombre d’occurrences de trois pixels alignés horizontalement, les 2 des bords étant noirs et celui du centre blanc (en haut à gauche).

13.3.3 La détection d'énergie dans des canaux fréquentiels

L'idée sous-jacente à toutes ces techniques est d'extraire l'énergie portée par le signal dans des bandes de fréquences diverses. Il n'est alors pas nécessaire d'utiliser de fenêtre d'estimation, car les analyses fréquentielles utilisées disposent généralement de leurs propres fonctions de fenêtrage, souvent gaussiennes. Il existe donc deux familles de techniques :

- celles qui analysent systématiquement tout l'espace fréquentiel (par exemple par une décomposition en ondelettes),
- celles qui sélectionnent un petit nombre de domaines fréquentiels significatifs (c'est par exemple l'analyse par filtres de Gabor).

Les filtres de Law

Ce sont des filtres très simples qui sont appliqués dans l'espace de l'image par des masques, dans l'esprit des filtres de détection de contour par masquage adapté. Les filtres sont au nombre de 25, obtenus par produit en x et y de 5 filtres de base (cf. figure 13.7). Leurs performances sont modestes, mais ils sont très rapides.

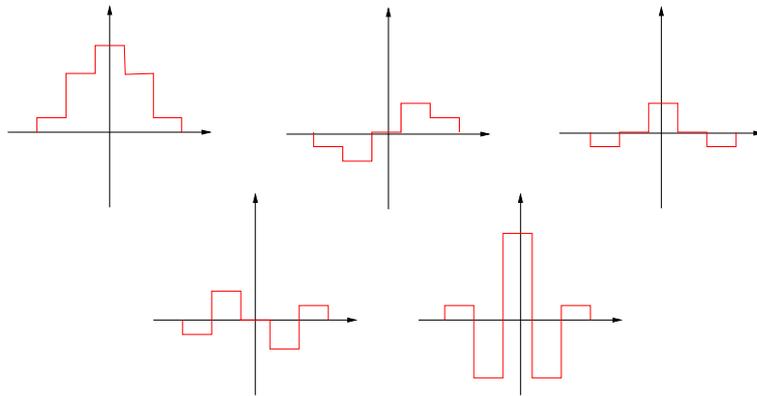


FIG. 13.7 – Les 5 filtres de base de Law. Un filtre de détection est le produit séparable de deux de ces filtres. Il est appliqué dans l'espace de l'image.

Les filtres en anneau et en coin

Ce sont des filtres définis dans l'espace de Fourier à 2 dimensions [Coggins et Jains, 1985]. Ils sont constitués du produit de deux familles élémentaires de filtres :

- 7 filtres en anneau, échelonnés de façon dyadique (c'est-à-dire centrés autour de fréquences suivant une progression géométrique de fréquences, le premier de fréquence $\nu_1 = \nu_0$, le second de fréquence $\nu_2 = 2\nu_0$, le troisième de fréquence $\nu_3 = 4\nu_0$, etc.), et de profil gaussien (cf. figure 13.8) :

$$\Phi_i(\nu = \sqrt{u^2 + v^2}) = \exp \left[-\frac{(\nu - \nu_i)^2}{2\alpha\nu_i^2} \right]$$

le recouvrement de ces filtres est contrôlé par le paramètre α .

- 4 filtres directionnels selon les axes et les diagonales, également de profils gaussiens.

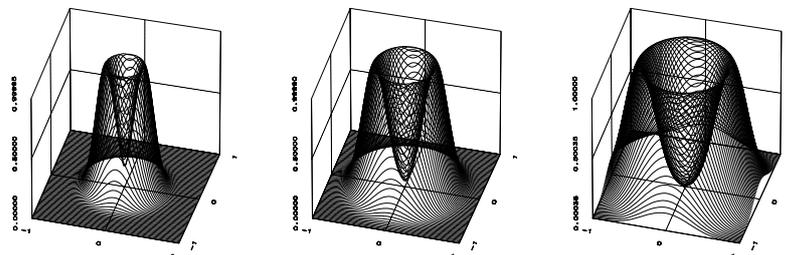


FIG. 13.8 – Les filtres en anneau dans le domaine spectral centrés sur les fréquences : 0,25, 0,40 et 0,60. Ce ne sont donc pas des filtres dyadiques.

Les filtres dyadiques de Gabor

Ce sont des filtres qui essaient de copier les fonctions de la vision des mammifères. Ils sont assez proches des précédents. Ce sont aussi des filtres à couverture dyadique qui se déduisent par rotation et homothétie du filtre de base qui a pour réponse impulsionnelle [Jain et Farrokhnia, 1991] :

$$\phi(x, y) = \exp \left[-1/2 \left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right] \right] \cos(2\pi\nu_0 x)$$

Dans [Jain et Farrokhnia, 1991], on conseille l'emploi de 5 fréquences radiales différentes et de 4 orientations. Dans [Randen et Husøy, 1999], on utilise aussi 4 orientations et les fréquences radiales suivantes :

$$\frac{\sqrt{2}}{2^6} \quad \frac{\sqrt{2}}{2^5} \quad \frac{\sqrt{2}}{2^4} \quad \frac{\sqrt{2}}{2^3} \quad \frac{\sqrt{2}}{2^2}$$

Les ondelettes et les QMF

La transformation en ondelettes discrètes permet de choisir une famille complète de filtres de décomposition en sous-bandes. Ces méthodes ont été présentées au chapitre 9. Les ondelettes de Mallat [Mallat, 1989] utilisent elles aussi une décomposition dyadique (mais deux directions d'analyse seulement). Des travaux suggèrent qu'une décomposition plus serrée que l'octave (c'est-à-dire des décompositions non-dyadiques) serait plus efficace pour l'analyse des textures [Chang et Kuo, 1993]. On propose ainsi l'usage de paquets d'ondelettes [Laine et Fan, 1993, Saito et Coifman, 1995]. Les ondelettes de Daubechies demeurent l'une des bases d'ondelettes les plus utilisées pour l'analyse de textures en raison de leur bonne efficacité de calcul, de leurs bonnes performances pour séparer les fréquences ainsi que de la grande variété des fonctions qu'elles autorisent [Daubechies, 1992b, Unser, 1995].

Des modèles ont également été créés alliant les propriétés de sélection des ondelettes et les modèles statistiques [Portilla et Simoncelli, 2000].

13.3.4 Les filtres optimisés

Lorsque l'on cherche à distinguer 2 textures, ou un petit nombre de textures connues par avance, la recherche systématique de l'énergie dans de nombreuses décompositions en sous-bandes peut être très lourde et hasardeuse. Il est possible de mettre en place des techniques adaptées pour les seules textures que l'on recherche qui permettent de se concentrer sur quelques mesures seulement de l'espace fréquentiel.

Les filtres de Gabor, qui disposent d'un petit nombre de paramètres seulement se prêtent assez bien à ce type

d'optimisation. On cherche alors à optimiser un critère comme :

$$J_1 = \frac{\mu_{t_1}}{\mu_{t_2}} \quad J_2 = \frac{(\mu_{t_1} - \mu_{t_2})^2}{\mu_{t_1}\mu_{t_2}} \quad \text{ou} \quad J_3 = \frac{(\mu_{t_1} - \mu_{t_2})^2}{\sigma_{t_1}^2 + \sigma_{t_2}^2}$$

où μ_{t_i} est la valeur moyenne de la mesure donnée par la texture t_i et σ_{t_i} sa variance [Unser, 1986]. Il est parfois possible de déterminer des valeurs optimales analytiques des filtres par ces critères, dans les autres cas on optimise par essai-erreur.

13.3.5 Les modélisations autorégressives

De la même façon que les signaux monodimensionnels sont analysés par des modélisations autorégressives uni-dimensionnelles (AR ou ARMA) afin de mettre en évidence les périodicités qu'ils contiennent, les signaux bidimensionnels peuvent être analysés par des modèles autorégressifs bi-dimensionnels. Une étude très complète de ces méthodes de description est présentée dans [Garello, 2001].

Dans une approche autorégressive à moyenne ajustée (ARMA), l'image est décrite par la formule :

$$f(i, j) = \sum_{(k, l) \in \mathcal{D}_s} a_{k, l} f(i - k, j - l) + b_{0, 0} B(i, j) + \sum_{(k, l) \in \mathcal{D}_e} b_{k, l} B(i - k, j - l) \quad (13.1)$$

où \mathcal{D}_s représente le domaine de prédiction lié à la sortie du filtre, \mathcal{D}_e , celui lié à l'entrée du filtre et B décrit un processus de bruit. Si les coefficients $b_{k, l}$ sont nuls pour tout (k, l) appartenant à \mathcal{D}_e , alors le processus est AR. Si les coefficients $a_{k, l}$ sont nuls pour tout (k, l) appartenant à \mathcal{D}_s , alors le processus est MA.

Le choix des domaines \mathcal{D}_s et \mathcal{D}_e incombe à l'utilisateur. Dans de nombreux cas on choisit $\mathcal{D}_s = \mathcal{D}_e = \mathcal{D}$. Afin de conserver aux modèles ARMA une bonne localité, il est souhaitable que le nombre de termes dans \mathcal{D} soit faible, mais pour bien représenter des signaux à longue périodicité, il vaut mieux qu'il soit grand. Ces modèles expriment naturellement une dépendance causale et suscitent donc des discussions comme celles que nous avons abordées au chapitre 2.

Les modèles ARMA permettent de représenter très bien n'importe quelle densité spectrale de puissance. Ceci se fait en utilisant les équations normales associées de Yule Walker [Alata et Cariou, 2001]. Les représentations par modélisation autorégressive des textures est particulièrement efficace pour des textures très périodiques et de large extension.

13.4 Les approches par champs markoviens

Les champs de Markov (cf. chapitre 7) peuvent être vus comme une sous-classe des processus ARMA, mais leur adaptation à traiter des images les distinguent. Ils se prêtent doublement à l'analyse des textures :

1. tout d'abord parce qu'ils comportent naturellement des descriptions des dépendances spatiales entre pixels par le choix des cliques et des potentiels d'interaction au sein des cliques,
2. mais aussi parce qu'ils ont, dans le terme d'énergie lié aux connaissances à priori, les éléments qui permettent de décrire les régions et leurs interactions.

Les champs de Markov sont particulièrement adaptés pour synthétiser et modéliser des textures, puisqu'il suffit de se donner les potentiels correspondant aux propriétés de dépendance, d'isotropie, de distance aux centres de classe, etc. que l'on souhaite voir représenter. Il y a donc une abondante littérature qui exploite le formalisme des champs markoviens pour segmenter des textures : [Hu et Fahmy, 1992, Derin et Elliott, 1987b] [Won et Derin, 1992, Andrey et Tarroux, 1998] très faciles d'emploi en synthèse des textures, ils sont cependant d'un usage plus difficile en analyse car il n'existe pas de formulation explicite permettant de déterminer les potentiels correspondant à une texture dont on dispose.

En règle générale, pour retrouver un champ de Markov à partir de la réalisation d'une texture, il faut définir les cliques significatives, puis les potentiels qui lient les sites dans ces cliques [Azencott et al., 1997].

1. La détermination des cliques utiles : on peut l'obtenir soit par l'analyse des fonctions de corrélation, soit par des techniques de croissance de voisinage, accompagnées de tests d'indépendance au fur et à mesure que l'on agrandit le voisinage.
2. La détermination des potentiels : elle passe souvent par la réduction des potentiels possibles à des classes étroites (par exemple les potentiels quadratiques ou les potentiels polynomiaux), elle peut alors se faire par des techniques de moindres carrés ou par des méthodes de filtrage (techniques de renormalisation).

13.4.1 La méthode de Manjunath et Chellappa

C'est une technique assez typique des approches markoviennes. Elle repose sur l'hypothèse de probabilités gaussiennes dans chaque classe (donc de potentiels quadratiques) [Manjunath et Chellappa, 1991]. Les énergies d'attache aux données pour une classe $L_s = l$ sont de la forme :

$$U_1(y_s|y_r, l) = \frac{1}{2\sigma_l^2} \left[y_s^2 - 2 \sum_r \theta_l^r y_s y_r \right] \quad (13.2)$$

dont les inconnues sont, pour chaque texture l , les θ_i , μ et σ^2 , où les θ_i correspondent aux cliques des directions S, SE, E, etc. du point s (cf. figure 13.9).

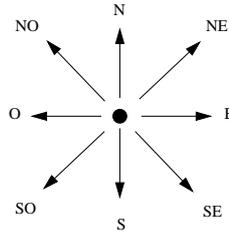


FIG. 13.9 – Le 8-voisinage utilisé par Manjunath et Chelappa. Seules les cliques d'ordre 2 sont utilisées.

Dans une première étape, il faut apprendre les paramètres des textures. L'image est arbitrairement subdivisée en petites fenêtres Ω . Prenons une seule fenêtre de taille $n \times n$ dont on va supposer la texture homogène. Sur cette fenêtre on estime $\hat{\Theta}$ aux moindres carrés par la formule⁴ :

$$\hat{\Theta} = \left[\sum_{\Omega} Q_s Q_s^t \right]^{-1} \left[\sum_{\Omega} Q_s y_s \right]$$

avec $Q_s = [y_{s+1}, y_{s-1}, y_{s+2}, \dots]^t$

et $\hat{\sigma}^2 = \frac{1}{n^2} \sum_{\Omega} [y_s - \hat{\Theta}^t Q_s]^2$.

Une tâche délicate est l'estimation du nombre de classes qui se fait soit de façon supervisée, soit par analyse du nuage des points obtenus sur toutes les fenêtres. Supposons que l'on recherche N classes. On détermine les N meilleures classes du nuage des points obtenus sur toutes les fenêtres par exemple par un algorithme de nuées dynamiques. Cela nous fournit N vecteurs de paramètres $f_k = \theta_1^k, \theta_2^k, \theta_3^k, \theta_4^k, \mu^k, \sigma_2^k$.

On peut alors procéder à la phase de segmentation. On va désormais travailler au niveau des sites du champ de Markov. On introduit un champ de liaisons entre les pixels, décrit par la seule variable binaire (processus de

⁴C'est ici que l'hypothèse gaussienne simplifie le calcul : l'estimation peut se faire par moindres carrés de façon explicite.

bord) b qui vaut 0 si deux pixels voisins sont dans une même classe et 1 s'ils sont dans des classes différentes. On recherche la distribution des b et des $V_{s\lambda}$, variables caractéristiques de l'appartenance du pixel s à la classe λ ($V_{s\lambda} = 0$ ssi le pixel au site s appartient à la classe λ) qui minimise l'énergie :

$$U = \sum_s \sum_\lambda u(s, \lambda) V_{s\lambda} - (1 - b) \sum_s \sum_\lambda \sum_{t \in N_s} V_{t\lambda} V_{s\lambda}$$

où le premier terme exprime l'attache aux données et le second l'a priori sur les régions. Les u incluent toute l'information connue sur le site s appartenant à la classe λ : par exemple $u(s, \lambda) = w(\lambda) + u_1(Ys, \lambda)$, u_1 étant comme dans la formule 13.2 et w étant un biais propre à la classe λ . L'optimisation d'une telle fonction d'énergie se fait soit par relaxation déterministe (par exemple l'ICM *Iterated Conditional Mode* qui remplace à chaque étape la classe λ par celle qui minimise l'énergie U au site considéré), soit par recuit simulé (relaxation stochastique) (cf. chapitre 7).

13.4.2 La méthode de Kervrann et Heitz

Dans cette méthode [Kervrann et Heitz, 1995], on calcule sur des fenêtres recouvrant l'image, des matrices de cooccurrence (cf. section 13.5). En chaque fenêtre, on représente la texture par un vecteur d'attributs issus de ces matrices $O_s = \{o_s^1, o_s^2, \dots, o_s^k\}$. Le terme d'attache aux données s'écrit sous la forme :

$$U_1 = \sum_{s \in \Omega} V(O_s, O_\lambda)$$

où O_s est le vecteur d'attributs du site s et O_λ est le vecteur d'attributs de la région d'étiquette λ affectée au site s . Le potentiel V s'écrit à partir de la distance Δ de Smirnov-Kolmogorov des composantes o_s^i et o_λ^i des vecteurs :

$$V(O_s, O_\lambda) = \sum_{i=1}^k [2\Gamma(\Delta(o_s^i, o_\lambda^i) > c^i) - 1]$$

où la fonction Γ vaut 0 ou 1 selon que les valeurs de Δ sont supérieures ou inférieures à une valeur de seuil c^i issue de tables du test de Smirnov-Kolmogorov [Saporta, 1990].

Le champ évolue vers le critère du MAP (Maximum A Posteriori) par la technique de relaxation sous-optimale de l'ICM : pour chaque site on choisit la meilleure classe parmi les seules classes des voisins du site considéré (cf. chapitre 7).

La construction des classes o_λ se fait de la façon suivante. Initialement il n'y a qu'une classe, identifiée au vecteur de la première fenêtre rencontrée. Les autres vecteurs de texture sont associés à cette classe si leur distance est inférieure au seuil donné. Afin de créer de nouvelles classes, on crée une classe « fourre-tout » dans laquelle sont rangées toutes les textures trop éloignées de celle déjà trouvée. Après un passage sur l'image, on détermine les paramètres de la classe trouvée (en calculant un vecteur O_λ sur toutes les réalisations trouvées) et on accroît de 1 le nombre de classes. On reprend le balayage de l'image. On itère ce procédé jusqu'à ce qu'un nombre pré-établi de classes soit trouvé ou qu'il n'y ait plus de candidats dans la classe fourre-tout.

Comme dans le cas précédent, la taille des fenêtres doit être :

1. assez grande pour permettre une caractérisation statistique réaliste des textures (et de leurs motifs) ;
2. assez petite pour conduire à des segmentations fines des détails.

Notons que cette méthode, au contraire de la précédente, ne permet pas de déterminer les potentiels et les cliques des textures retenues, autrement que par référence aux primitives de la matrice de cooccurrence.

13.5 Les méthodes structurales

Ces méthodes ne s'appliquent qu'aux textures construites sur une primitive bien identifiable qui se répète sur une maille régulière. L'analyse d'une texture homogène se fait en plusieurs étapes :

1. l'image est découpée en sous-fenêtres, chacune contenant plusieurs primitives (typiquement de l'ordre de 5×5 primitives),
2. sur chaque sous fenêtre on détermine les 2 vecteurs moyens de périodicité, par exemple par l'examen de la fonction d'auto-corrélation,
3. connaissant ces vecteurs, chaque sous-image est normalisée par un rééchantillonnage de façon à ce que les nouveaux vecteurs de périodicité soient tous dans une position de référence,
4. on calcule alors par TF le motif de la primitive en sommant dans l'espace de Fourier toutes les sous-fenêtres normalisées et en isolant le sous-domaine du fondamental,
5. on est alors en mesure de reconstruire une texture parfaite en répétant le motif moyen sur la maille moyenne, ou des textures plus ou moins parfaites en mélangeant des motifs réels ou moyens sur des mailles réelles ou moyennes.

Des méthodes plus structurales encore s'appuient sur des techniques de graphes ou de grammaires. Dans les approches par graphes, on représente les dépendances entre les diverses composantes de la texture par des graphes (arbres, cycles) mettant en évidence les dépendances entre pixels. Dans les approches par grammaires, on parcourt la texture ou des sous-ensembles de la texture par un balayage systématique de tous les sites, construisant ainsi une chaîne de descripteurs. Les relations entre les descripteurs successifs sont exprimées par des règles qui indiquent l'enchaînement des niveaux de gris le long du parcours. Les chaînes font alors l'objet d'un traitement syntaxique qui extrait ces dépendances [Lu et Fu, 1978, Lu et Fu, 1979].

13.6 Textures hétérogènes

Si l'on fait l'hypothèse que la texture sous-jacente est stationnaire, on peut déduire des variations de ses propriétés statistiques dans l'image, des éléments suffisants pour remonter à une connaissance de certains éléments de relief et d'orientation 3D des objets de la scène observée. Cela est naturellement fait par système visuel humain (cf. figure 13.2). Pour faire cela par traitement numérique, il convient de mesurer en chaque point de l'image (c'est-à-dire en des zones suffisamment petites autour de chaque point) des propriétés statistiques d'ordre 2 (prenant en compte les pixels 2 par 2, comme par exemple la décroissance de la corrélation à $1/2$), puis d'étudier les variations spatiales de ces propriétés. Ces variations sont alors reliées à l'orientation de la surface observée (généralement sous l'hypothèse que celle-ci fait un angle constant avec la direction d'observation).

Chapitre 14

Description de contours et de formes

Chapitre rédigé par Henri MAÎTRE

Dans de nombreuses applications de traitement des images, une fois l'image segmentée (on peut le faire, de la façon la plus simple, en seuillant l'image [Sahoo et al., 1988, Cham et al., 1998] ou à partir des méthodes décrites aux chapitres 11 et 12), on essaie de reconnaître les divers objets qui la composent à partir de leur seule silhouette. C'est particulièrement vrai dans les applications de tri dans lesquelles on est intéressé à saisir ou trier des objets arrivant sur un convoyeur¹, dans les applications de reconnaissance de caractères, ainsi que pour des applications de surveillance ou de guidage, par exemple en imagerie militaire.

Il est utile alors que l'utilisateur dispose d'une représentation de la forme vérifiant plusieurs propriétés :

1. une bonne fidélité à la forme initiale,
2. une bonne discrimination de formes différentes,
3. une bonne adaptation aux opérations de reconnaissance des formes, et en particulier une insensibilité aux déformations qui sont susceptibles d'entacher l'objet,
4. une certaine compacité pour permettre l'archivage de nombreuses formes et potentiellement de ces mêmes formes sous divers aspects.

De nombreuses représentations des formes ont été développées concurremment, chacune pour répondre à un problème parfois assez spécifique et donc mettant l'accent sur l'une ou l'autre des propriétés ci-dessus.

14.1 Fonction caractéristique

La représentation naturelle d'une forme dans une image est une représentation par fonction caractéristique, c'est à dire sous forme d'une image binaire, les seuls pixels non-nulles étant ceux couverts par la forme :

$$\begin{aligned} f(i, j) &= 0 && \text{si } (i, j) \notin \text{objet} \\ &= 1 && \text{si } (i, j) \in \text{objet} \end{aligned}$$

Dans une image où de nombreux objets existent, on utilise une image d'étiquettes (ou de *labels*) qui généralise la notion de fonction caractéristique :

$$f(i, j) = k \quad \text{si } (i, j) \in \text{objet } k$$

¹Le tri se partage en tri planaire (où les objets sont préalablement mis à plat, généralement sur un tapis roulant), et en tri en vrac, où ils sont présentés dans un désordre total.

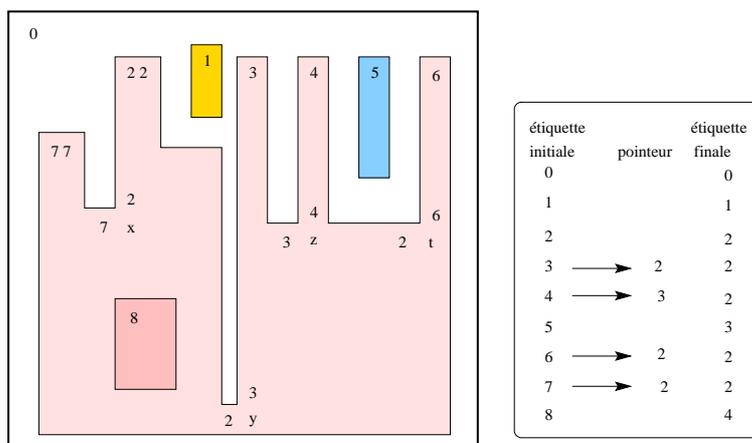


FIG. 14.1 – Etiquetage d'une image en 4 connexité. L'image de gauche est le résultat du premier passage jusqu'au point noté x. Les étiquettes ont été attribuées à chaque point soit en donnant une nouvelle étiquette si les 2 antécédents du point sont à 0, soit en donnant l'étiquette des antécédents si ces antécédents ont une étiquette semblable différente de 0. En x les étiquettes sont de valeur différente. Le pixel x se voit attribué l'étiquette 2 car $2 = \min(2,7)$, le pixel y l'étiquette 2, le pixel z l'étiquette 3 et le pixel t l'étiquette 2. Le point x crée le pointeur $7 \sim 2$, y le pointeur $3 \sim 2$, $z \ 4 \sim 3$ et $t \ 6 \sim 2$. Après un second passage sur l'image, on détermine la liste finale des étiquettes de 0 à 4. Pour recréer une image d'étiquettes en conformité avec cette liste, il faut donc faire un deuxième passage sur l'image.

Une image d'étiquettes résulte d'une procédure d'étiquetage. Elle se fait par le choix d'une connexité qui permet de définir les propriétés de topologie des divers objets (4- ou 8- connexité sur les trames carrées, 6-connexité en trame hexagonale). L'étiquetage vient attribuer une étiquette semblable à tous les points connexes au sens de la segmentation. Par exemple en 4-connexité un algorithme d'étiquetage fonctionne de la façon suivante (cf. figure 14.1).

1. une liste d'étiquettes est initialisée à 0 ;
2. le balayage (par exemple vidéo) de l'image est initialisé ; au premier point est attribuée l'étiquette 0 ;
3. un point nouveau est pris dans l'ordre du balayage ; on examine sa connexité à son voisin supérieur et à son voisin de gauche (ses antécédents) :
 - si le point courant est connexe à ses deux antécédents (au sens de la segmentation), et si ses deux antécédents ont la même étiquette, on lui attribue l'étiquette de ces antécédents ;
 - si le point courant est est connexe à ses deux antécédents, mais que ses deux antécédents n'ont pas la même étiquette, on donne au point l'étiquette la plus petite parmi les deux, puis dans la liste des étiquettes, on crée un pointeur qui renvoie de l'étiquette la plus grande vers l'étiquette la plus petite ;
 - si le point courant est connexe à l'un seulement de ses antécédents, on lui attribue l'étiquette de cet antécédent ;
 - si le point courant n'est pas connexe à ses deux antécédents, on incrémente la liste des étiquettes, puis on attribue au point courant cette nouvelle étiquette ;
4. on retourne en 3 jusqu'à atteindre le dernier point de l'image.

Lorsque l'ensemble de l'image a été ainsi traité (après donc un passage complet sur l'image), on réorganise l'image des étiquettes. Pour cela on parcourt cette image et chaque valeur d'étiquette est remplacée par l'étiquette terminale rencontrée en remontant la liste des pointeurs. La liste des étiquettes est alors filtrée de toutes les étiquettes qui possèdent au moins un pointeur. On constitue ainsi une liste d'étiquettes qui occupe tous les entiers entre 0 et k_{max} s'il y a k_{max} objets se découpant sur un fond.

14.2 Description de formes

Les descripteurs (ou paramètres) de forme sont des nombres qui représentent chaque forme et permettent de les classer (par exemple par nuées dynamiques ou plans séparateurs). Ces paramètres ont été abondamment proposés dans la littérature [Coster et Chermant, 1985], beaucoup étant adaptés à des formes particulières. Citons les plus utilisés, par exemple :

- le rapport iso-périmétrique, proportionnel au rapport du carré du périmètre de l'objet à sa surface (dans des images continues, il est maximum pour le cercle) ;
- l'indice d'allongement [Schmitt et Mattioli, 1994b], proportionnel au rapport du carré du diamètre géodésique à la surface de l'objet, (minimal pour les disques au sens du voisinage choisi).

Les paramètres de forme sont bien adaptés pour des objets de taille assez grande car ils ont souvent été définis pour des formes continues. Pour des objets de petite taille (quelques pixels), la discrétisation du maillage induit des comportements souvent peu satisfaisants.

14.2.1 Représentation par les moments

Connaissant l'objet par sa fonction caractéristique $f(x, y)$, une représentation classique de sa forme consiste à en mesurer les divers moments :

$$M_{mn} = \iint x^m y^n f(x, y) dx dy$$

En particulier les moments centrés (rapportés au centre de gravité (X_g, Y_g) de la forme) sont invariants par translation :

$$\bar{M}_{mn} = \iint (x - X_g)^m (y - Y_g)^n f(x, y) dx dy$$

Sur une image discrète, ces moments s'écrivent :

$$\bar{M}_{mn} = \frac{1}{J^m K^n} \sum_{j=1}^J \sum_{k=1}^K (j - X_g)^m (k - Y_g)^n f(j, k)$$

Les moments d'inertie (valeurs propres de la matrice d'inertie, matrice 2×2 de terme courant \bar{M}_{mn} , $m+n = 2$) sont invariants par rotation. Les moments d'inertie normés par la plus grande valeur propre sont invariants par similitude (rotation et facteur d'échelle). Les moments d'inertie décrivent bien l'allongement de formes régulières comme des ellipses ou des distributions gaussiennes. Ils sont plus ambigus sur des formes complexes (cf. figure 14.2).

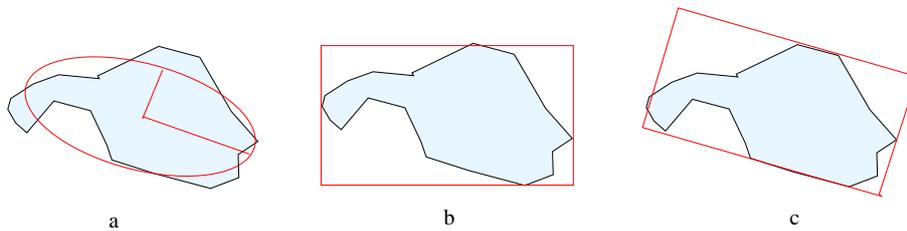


FIG. 14.2 – Description de formes par : a - ellipse et axes d'inertie, b - boîte englobante, c - boîte minimale.

14.2.2 Représentation par les moments invariants

Des moments invariants ont été proposés (nommés moments de Hilbert), invariants par translation, rotation et changement d'échelle (dans la limite de l'approximation par une maille discrète) [Hu, 1962]. Ils se construisent à

partir du moment normé :

$$N_{nm} = \frac{J^m K^n}{\bar{M}_{00}^\alpha} \bar{M}_{mn}$$

où $\alpha = \frac{1}{2}(m + n) + 1$, par les formules :

$$\begin{aligned} h_1 &= N_{20} + N_{02} \\ h_2 &= (N_{20} - N_{02})^2 + 4N_{11}^2 \\ h_3 &= (N_{30} - 3N_{12})^2 + (N_{03} - 3N_{21})^2 \\ h_4 &= (N_{30} + N_{12})^2 + (N_{03} - N_{21})^2 \\ h_5 &= (N_{30} - 3N_{12})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{03} + N_{21})^2] \\ &\quad + (3N_{21} - N_{03})(N_{03} + N_{21})[3(N_{30} + N_{12})^2 - (N_{03} + N_{21})^2] \\ h_6 &= (N_{20} - N_{02})[(N_{30} + N_{12})^2 - (N_{03} + N_{21})^2] \\ &\quad + 4N_{11}(N_{30} + N_{12})(N_{03} + N_{21}) \\ h_7 &= (3N_{12} - N_{30})(N_{03} + N_{21})[3(N_{30} + N_{12})^2 - (N_{03} + N_{21})^2] \\ &\quad + (3N_{21} - N_{03})(N_{30} + N_{12})[(N_{30} + N_{12})^2 - 3(N_{03} + N_{21})^2] \end{aligned}$$

14.2.3 Boîtes englobantes et boîtes minimales

Lorsque les formes sont plus irrégulières, on préfère souvent des descriptions de l'allongement par la forme de la boîte englobante :

- soit la boîte alignée sur les axes et donc simplement définie par ses dimensions ($x_{max} - x_{min}$) et ($y_{max} - y_{min}$),
- soit la boîte orientée selon le diamètre minimal de l'objet (de calcul un peu plus complexe) (cf. figure 14.2).

14.3 Polygones de Guzman

C'est l'une des méthodes les plus anciennes [Guzman, 1968]. L'approche de Guzman consiste à envelopper l'objet à reconstruire dans des « boîtes » de formes de plus en plus précisément adaptées. Elle peut donc s'appliquer soit à partir d'une image d'étiquettes de la forme à analyser, soit à partir d'une liste des pixels de contour de la forme. Les formes prototypes auxquelles seront comparées les objets sont construites sur un maillage carré de façon systématique et se classent en niveaux en fonction de la longueur de leur périmètre. Cette longueur (exprimée en pixels) étant obligatoirement paire, on décrit chaque niveau par le demi-cardinal du nombre de pixels constituant la frontière. Dans chaque niveau, les divers prototypes sont repérés par un indice qui les identifie uniquement. Ainsi, la forme la plus simple (le carré unitaire) a une longueur 4 (et donc appartient au niveau 2). C'est la forme $\Phi(2, 1)$. Afin de tenir compte des symétries et des rotations de $\pi/2$, toutes les formes identiques par rotation modulo $\pi/2$ et toutes les formes identiques par symétrie droite (par rapport à un axe horizontal ou vertical) sont rapportées à une même prototype du dictionnaire des formes (cf. figure 14.3). Un objet quelconque est donc décrit par une succession d'indices décrivant l'objet à divers niveaux de résolution. Deux objets sont généralement identiques jusqu'à un niveau n , puis différents à partir du niveau $n + 1$ (cf. figure 14.4).

Cette approche rencontre plusieurs limites :

- pour les ordres grands, les formes deviennent très nombreuses et le dictionnaire trop grand. La recherche du représentant se faisant par comparaison au prototype devient très longue ;
- les distances d'une forme à un prototype ne s'imposent pas de façon unique (cf. figure 14.5), laissant place à des classements différents selon le critère adopté ;

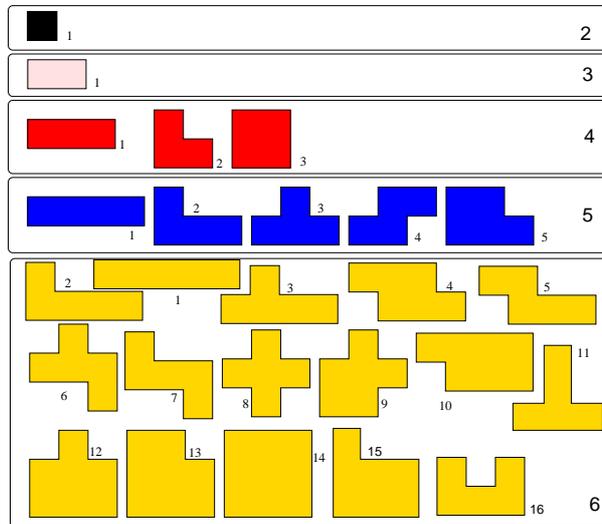


FIG. 14.3 – Les premières formes du dictionnaire de Guzman.

- les relations « verticales » entre niveaux sont complexes et ne permettent pas d’accélérer le parcours de l’arbre de façon commode ;
- certaines formes se retrouvent identiques dans des niveaux différents, à un facteur d’échelle près (par exemple le carré élémentaire se retrouve dans tous les niveaux impairs).

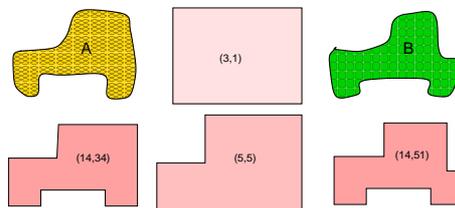


FIG. 14.4 – Représentation de Guzman. Les 2 formes A et B sont semblables aux niveaux de représentation 3 et 5, mais différent au niveau 14, où l’une est représentée par le prototype 34 et l’autre par le prototype 51.

L’approche de Guzman a été abandonnée aujourd’hui, mais elle apporte des concepts intéressants à la description des formes :

- la notion de raffinement de la description,
- la notion de similarité entre tous les objets « vus de très loin »,
- la notion d’invariance par rotation, symétrie et changement d’échelle que l’on pourrait étendre éventuellement à d’autres transformations.

14.4 Chaînes de Freeman

C’est la méthode la plus ancienne de description des contours dans les images et aussi la plus utilisée encore aujourd’hui [Freeman, 1961, Freeman, 1977]. Si elle est moins utilisée en reconnaissance des formes qu’elle le fut, elle est très fréquemment utilisée dans des applications récentes comme la transmission des images par zones (par exemple dans MPEG-).

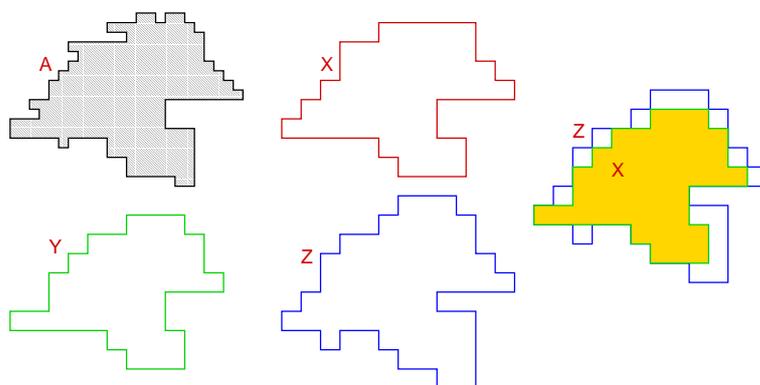


FIG. 14.5 – L'arbitraire de l'approximation des formes : la forme A est l'objet. Les 3 formes X, Y et Z en sont des approximations. La distance d'une représentation à la forme initiale peut se mesurer par la surface entre les 2 contours. Il y a de nombreuses formes à distance minimale. X est une forme la plus proche en terme de surface et « la plus simple » au sens du nombre d'angles droits lorsqu'il y a plusieurs solutions de même distance. Y est la forme la plus proche par défaut (entièrement contenue dans A), Z est celle par excès. Z et Y sont superposées à droite.

14.4.1 Définition

C'est une technique de représentation des directions du contour (on code la direction le long du contour dans un repère absolu lors du parcours du contour à partir d'une origine donnée). Les directions peuvent se représenter en 4-connextité (codage sur 2 bits) ou en 8-connextité (codage sur 3 bits)².

Le codage d'un contour se fait donc de la façon suivante :

1. transmission des coordonnées absolues du point de départ,
2. transmission de la liste des codes de déplacement d'un point du contour au suivant sur le maillage.

Les codes des contours sont donnés par la figure 14.6.

Dans d'autres techniques, on code de façon différentielle le changement de direction d'un point au suivant. Cela peut se justifier en codage sur 3 bits si une direction est très dominante par rapport aux autres.

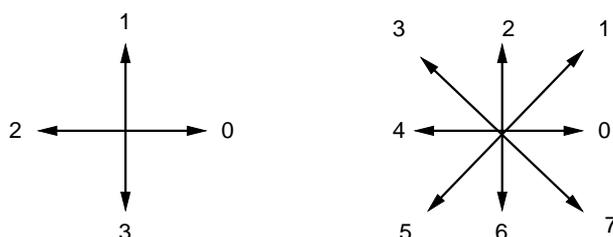


FIG. 14.6 – Les codes de Freeman en 4-connextité (à gauche) et en 8-connextité (à droite).

Deux exemples de codages par chaînes de Freeman sont présentés sur la figure 14.7 (le codage du positionnement absolu a été omis). Dans le cas de la 4-connextité, le codage de la suite de contours occupe 124 bits (62×2), tandis qu'en 8 connextité elle occupe 126 bits (42×3). Il n'y a pas de règle générale sur l'efficacité des représentations en termes de compression. Si la courbe est très complexe (beaucoup de changements de direction),

²des essais ont été faits pour coder également les chaînes de Freeman sur 4 bits en ajoutant les directions reliant le point central à la seconde couronne des voisins qui ne sont pas dans la direction des voisins de la première couronne. Mais cette approche n'a pas eu un très grand succès.

le codage en 8-connextité est généralement plus compact. La qualité du codage est cependant toujours au moins aussi bonne en 8-connextité.

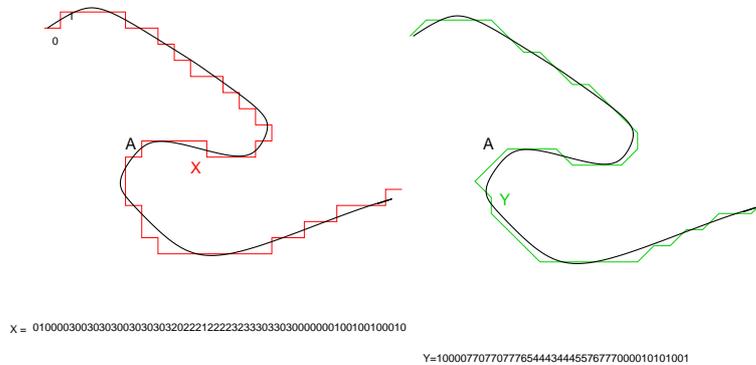


FIG. 14.7 – Codage de Freeman. Approximation de la courbe A par la chaîne X (en 4-connextité) et par la chaîne Y (en 8-connextité).

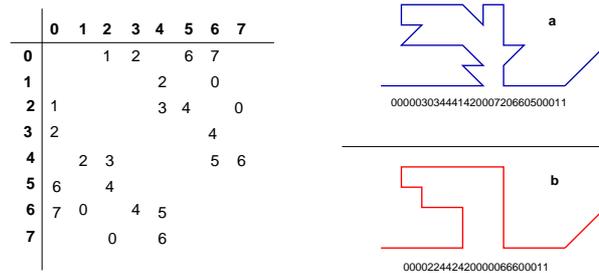


FIG. 14.8 – En 8-connextité, il est possible de simplifier les courbes en remplaçant certains couples de descripteurs par le descripteur de la table ci-dessus. Dans la figure de droite, la figure b est une simplifiée de la figure a par remplacement des seuls angles aigus. Il existe plusieurs simplifiées selon l'ordre de substitution que l'on adopte.

14.4.2 Les propriétés des chaînes de Freeman

Les chaînes de Freeman se prêtent à un certain nombre de manipulations commodes.

1. On obtient une dilatation de la courbe d'un facteur k en répétant k fois chaque descripteur.
2. On ne peut généralement pas réduire une courbe sans distorsion.
3. On fait tourner une courbe de $k \times \frac{2\pi}{n}$ (dans le cas d'une chaîne de Freeman en n -connextité) en ajoutant (ou retranchant) k modulo n à la chaîne initiale.
4. On mesure la longueur d'une courbe par les formules suivantes :
 - en 4-connextité : L = nombre de descripteurs,
 - en 8-connextité : L = nombre de descripteurs pairs + $\sqrt{2}$ nombre de descripteurs impairs.
5. Inversion d'un chemin : on inverse tous les descripteurs et on inverse la séquence. L'inverse d'un descripteur j est $\bar{j} = n/2 + j \text{ mod}(n)$.
 exemple en 4-connextité : $X = 001321 \rightarrow \bar{X} = 301322$.

6. Simplification d'un chemin : c'est un chemin dont on a supprimé des détails sans changer globalement la forme. Cela s'obtient en remplaçant des séquences de p descripteurs consécutifs par des descripteurs équivalents reliant les mêmes points (cf. figure 14.8) :
 exemple en 4-connexité : $\{012\} \rightarrow \{1\}$,
 en 8-connexité : $\{03\} \rightarrow \{2\}$.
7. Réduction d'un chemin : c'est l'un des chemins de longueur minimale reliant les 2 extrémités de la courbe initiale. On associe 2 par 2 des descripteurs inverses de la chaîne et on les supprime.
 exemple en 4-connexité : $X = \{00132122\} \rightarrow \tilde{X} = \{21\}$. On obtient tous les chemins réduits en changeant l'ordre des associations.
 en 8-connexité la réduction est un peu plus complexe car il faut aussi regrouper des ensembles de 3 ou 4 descripteurs qui s'annulent : exemple $\{025\}$ ou $\{7225\}$.
8. Fermeture d'un contour : on teste la fermeture d'un contour en vérifiant que la chaîne réduite est nulle.
 On ferme un contour en lui ajoutant le chemin inverse d'un de ses chemins réduits. Il y a beaucoup d'autres fermetures possibles que par addition de l'inverse d'un réduit, mais les fermetures obtenues ainsi sont de longueur minimale. Il y a d'autres fermetures minimales que celles obtenues par addition de l'inverse d'un réduit.
9. Courbe qui s'intersecte : pour savoir si une courbe décrite par sa chaîne se recoupe, on procède à une réduction de chemin systématique en partant de son origine et en testant si chaque nouveau descripteur possède un inverse dans la chaîne déjà parcourue. Si à un instant la chaîne déjà parcourue se réduit à une chaîne nulle, on a trouvé un point double.
10. Changement d'origine : le changement de l'origine d'une chaîne de longueur L revient à une permutations circulaires des descripteurs modulo L .
11. Sens de parcours d'un contour : le contour fermé d'une forme simplement connexe peut être décrit dans le sens direct ou dans le sens inverse. Pour connaître le sens d'un contour fermé, on réduit ce contour jusqu'à n'avoir que 4 descripteurs. En 4-connexité, il n'existe que 2 chaînes possibles (et celles qui s'en déduisent par changement d'origine) :
 - le carré direct : $X_D = 0123$,
 - le carré inverse : $X_I = 0321$.
 En 8-connexité les configurations sont un peu plus nombreuses.
12. Surface d'une région simplement connexe : la technique est assez complexe mais ne nécessite pas la reconstruction de la région [Freeman, 1977].

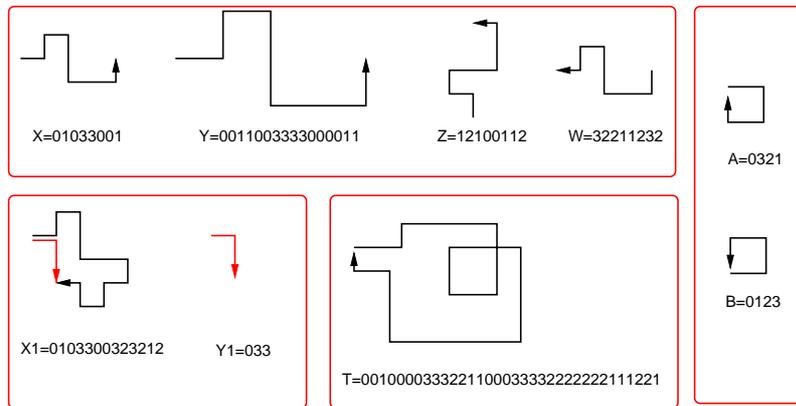


FIG. 14.9 – Quelques propriétés des chaînes de Freeman : X = chaîne originale, Y = contour double, Z = rotation de $\pi/2$, W = contour inverse. X1 = chaîne originale et Y1 = chaîne réduite. T = exemple de boucle dans une chaîne. A et B sont les deux contours élémentaires (inverse et direct) en 4-connexité.

14.4.3 Reconnaissance des formes par des chaînes de Freeman

Pour que 2 formes X et Y soient identiques, il faut tout d’abord qu’elles aient le même nombre de descripteurs. Puis, comme elles peuvent différer par le sens de description, il faut comparer X à Y d’une part et à \bar{Y} d’autre part. Enfin, comme elles peuvent avoir des origines différentes, il convient de tester les deux chaînes descripteur à descripteur sous toutes les hypothèses de permutation circulaires. Des techniques rapides ont été développées pour faire ces recherches (d’une complexité sous-linéaire en L), [Boyer et Moore, 1977, Miclet, 1984]³.

Dans le cas où l’on ne recherche pas une similarité exacte mais seulement approchée (par exemple pour prendre en compte la présence du bruit lors de la détection des contours), on est amené à utiliser des techniques de **distance d’édition** pour lesquels on chiffre les coûts des erreurs possibles : omission de contours, adjonction de contours, permutation de descripteurs, etc. On utilise alors généralement des techniques de programmation dynamique (algorithme de Wagner et Fisher [Wagner et Fisher, 1974, Miclet, 1984], élargi pour des substitutions de plusieurs descripteurs consécutifs dans [Lowrance et Wagner, 1975]) ou de relaxation (algorithme de Davis) pour retrouver la distance minimale entre X et Y .

14.5 Descripteurs de Fourier

14.5.1 Descripteur par tangente

Dans cette approche on considère le contour comme une courbe continue qui peut être décrite par son abscisse curviligne s à partir d’une origine A choisie (cf. Fig 14.10). On paramètre la courbe par l’angle fait par le vecteur tangent en chaque point et celui au point origine : $\phi(s)$ et on crée la variable réduite t qui prend ses valeurs entre 0

³Ce sont des automates finis semblables à ceux qui sont utilisés en particulier pour les compilateurs et les éditeurs de texte.

et $2\pi : t = 2\pi s/L$, où L est la longueur complète du contour. On construit alors la fonction $\Phi(t)$:

$$\Phi(t) = \phi \left[\frac{2\pi \cdot s}{L} \right] - \frac{2\pi s}{L}$$

le terme correctif prenant en compte l'enroulement de 2π de la tangente pour un tour de contour. La fonction $\Phi(t)$ est une fonction périodique sur $[0, 2\pi[$ qui admet donc une série de Fourier : $\Phi(t) = \sum_{k=0}^{\infty} a_k \exp(-ikt)$. On appelle descripteurs de Fourier l'ensemble des modules des a_k : $\{|a_k|\}$.

Ils bénéficient des propriétés suivantes :

1. ils sont invariants par translation de la forme,
2. ils sont invariants par changement d'échelle (puisque t est normalisé),
3. ils sont invariants par rotation, puisque l'on a choisi la différence d'angle entre 2 tangentes,
4. ils sont invariants par changement d'origine, car passer d'une origine A à une origine A' revient à :
 - retrancher $\Phi(t_{A'})$ à toutes les valeurs $\Phi(t)$,
 - changer t en $t - t_{A'}$.

Donc : $\Phi_{A'}(t) = \Phi_A(t) * \delta(t - t_{A'}) - \Phi(t_{A'})$ et : $a_k \rightarrow a_k \exp(-kt_{A'})$ si $k \neq 0$

Pour comparer des formes on compare leurs descripteurs par ordre croissant. Si de plus on veut simplifier le contour, il suffit de supprimer les ordres k élevés dans le développement. malheureusement, dans cette représentation, si un contour est fermé, le contour obtenu en filtrant les hautes fréquences ne l'est généralement plus. C'est pourquoi on préfère souvent les descripteurs de Fourier par représentation complexe qui n'ont pas ce défaut.

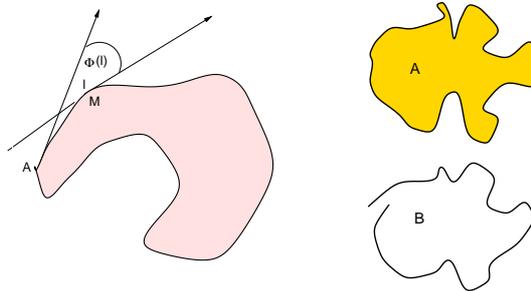


FIG. 14.10 – A gauche : descripteurs de Fourier par tangente. Le point de départ A est choisi arbitrairement. Son vecteur tangent sert de référence à la paramétrisation de la courbe. A droite : après troncature du développement de Fourier, la forme A , initialement fermée devient B , plus régulière, mais non fermée.

14.5.2 Représentation complexe

Dans cette représentation, on décrit la forme par un ensemble $\{M_j\}$ de points de contours, et on représente la forme dans le plan complexe. On attache donc à chaque M_j un nombre complexe $z_j = x_j + iy_j$. On appelle alors descripteurs de Fourier, les coefficients de la TF Z de z :

$$Z_k = \sum_{j=1}^N z_j \exp(-2\pi i j k)$$

Les coefficients Z_k , pour $k \in [-N/2 + 1, N/2]$, jouissent d'intéressantes propriétés [Bertrand et al., 1982].

1. Pour $k = 0$, Z_0 est le centre de gravité de la forme. Si l'on l'omet, la description est invariante par translation.

2. Si tous les Z_k sont nuls sauf pour $k = 1$, la forme est un cercle de rayon Z_1 (ou un polygone régulier à N côtés), donc Z_1 joue le rôle de facteur d'échelle. La normalisation par Z_1 rend la forme invariante par homothétie.
3. Les coefficients $Z_{|k|}$ et $Z_{|1-k|}$ (pour $k \neq 1$ et $k \neq 0$) jouent des rôles symétriques (mais opposés) de la façon suivante :
 - l'ordre k indique le nombre d'actions sur le cercle unité (entre 0 et 2π) : 1 action pour $k = 2$ et $k = -1$, 2 actions pour $k = 3$ et $k = -2$, 3 actions pour $k = 4$ et $k = -3$, etc. Ces actions sont réparties régulièrement autour du cercle unité,
 - les valeurs de $k > 0$ indiquent des actions de traction sur la courbe, pour la déformer vers l'extérieur du cercle unité, les valeurs de $k < 0$ indiquent des actions de pression sur la courbe, pour creuser la courbe vers son centre,
 - la phase du nombre complexe Z_k : ϕ_k exprime le lieu, sur le cercle unité, où s'exerce l'action.
4. Plus les coefficients sont nombreux, plus la forme est complexe. Et plus les coefficients sont élevés plus les détails sont fins sur la courbe.

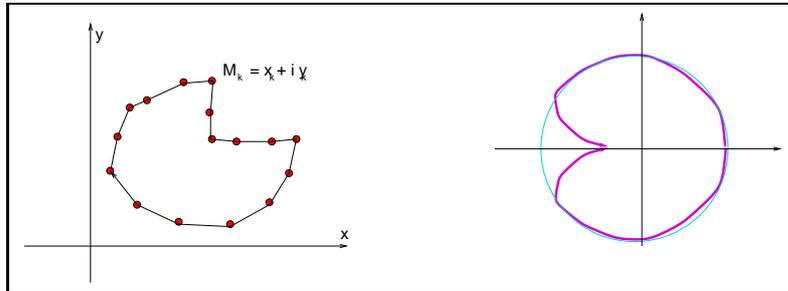


FIG. 14.11 – A gauche : descripteurs de Fourier par représentation complexe. Le point courant M_k est décrit par ses coordonnées complexes dans le plan image. A droite : le cercle résulte de la prise en compte du seul coefficient Z_1 , la courbe présentée résulte de l'adjonction d'un coefficient de pression (donc à $k < 0$), en 1 seul point (donc $k = 2$ ou $k = -1$, ici $k = -1$), de déphasage nul (puisque situé en π).

On voit que les descripteurs de Fourier par représentation complexe ont le même type de comportement que les descripteurs par tangente. Ils sont mieux adaptés aux formes discrètes puisqu'ils garantissent toujours que la forme demeure fermée après troncature du développement de Fourier. Ils peuvent être invariants par rotation si l'on s'intéresse aux seuls modules des coefficients Z_k .

Parmi leurs inconvénients, il faut noter que l'on ne peut pas aisément garantir qu'un contour de forme simplement connexe ne donnera pas, après troncature, un contour qui s'auto-intersectera.

14.6 Approximations polynomiales

Il y a essentiellement 3 types de telles approximations :

1. les approximations analytiques par ajustements de nuages de points par des polynômes du premier ordre (conduisant donc à des polygones) ;
2. les approximations polygonales par des critères géométriques obtenues en parcourant la courbe des points ordonnés selon une abscisse curviligne ;
3. les approximations par des polynômes de degré > 1 et en particulier les approximations par des fonctions splines.

Les approximations polygonales des formes décrites par des points de contour sont les représentations les plus employées. Dans les cas les plus simples, on se contente de relier 2 à 2 des points de contour dans un ordre déterminé au préalable (donc selon une abscisse curviligne croissante). Deux problèmes se posent alors :

- peut-on éliminer d'éventuels points parasites qui n'appartiennent pas au contour ?
- peut-on faire l'économie de certains sommets de la ligne polygonale pour obtenir des formes plus simples ?

Ces problèmes seront abordés dans le paragraphe 14.6.2.

Mais avant d'en arriver à cette situation simple, se posent les problèmes difficiles de déterminer quels points appartiennent à un même contour et dans quel ordre les relier lorsque ces points sont issus d'un détecteur de contour par nature local. C'est ce que nous abordons tout d'abord.

14.6.1 Approximation d'un nuage de points par une droite unique

C'est un problème bien classique mais qui mérite pourtant un peu d'attention. Soit $M_i = (x_i, y_i)$ les points, en nombre N , que l'on cherche à approcher par une droite. Il y a 2 façons de prendre le problème.

Approximation par régression linéaire

C'est une approche aux moindres carrés : on recherche la droite $\Delta : y = a_0 + a_1x$ qui minimise la distance (cf. Fig 14.12) :

$$d_1^2 = \sum_i^N [y_i - (a_0 + a_1x_i)]^2 \quad (14.1)$$

La solution est donnée par : $A = X^\#Y = (X^tX)^{-1}X^tY$, où $X^\#$ dénote la matrice pseudo-inverse de la matrice X , où X , Y et A sont donnés par :

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_N \end{bmatrix}$$

$$Y = [y_1, y_2, \dots, y_N]^t$$

$$A = [a_0, a_1]^t$$

Ces formules s'étendent très aisément aux espaces de dimensions supérieures, ainsi qu'aux approximations par des polynômes d'ordre plus élevé. la distance minimisée est celle mesurée selon le seul axe y . C'est donc une mesure généralement mal adaptée en traitement d'image, puisque x et y jouent habituellement un rôle équivalent. On lui préfère donc les méthodes par axe d'inertie.

Approximation par axe principal d'inertie

C'est aussi une approche aux moindres carrés, mais on minimise dans ce cas la somme des distances de tous les points à la droite Δ :

$$d_2^2 = \sum_{i=1}^N \frac{[(a_0 + a_1x_i - y_i)]^2}{a_1^2 + 1} \quad (14.2)$$

C'est l'équation de l'axe d'inertie des points, qui passe par leur centre de gravité X_g, Y_g et qui est donné comme vecteur propre de plus grande valeur propre de la matrice de forme quadratique :

$$S = \sum_N V_i V_i^t = \begin{bmatrix} \sum x_i x_i & \sum x_i y_i \\ \sum x_i y_i & \sum y_i y_i \end{bmatrix}$$

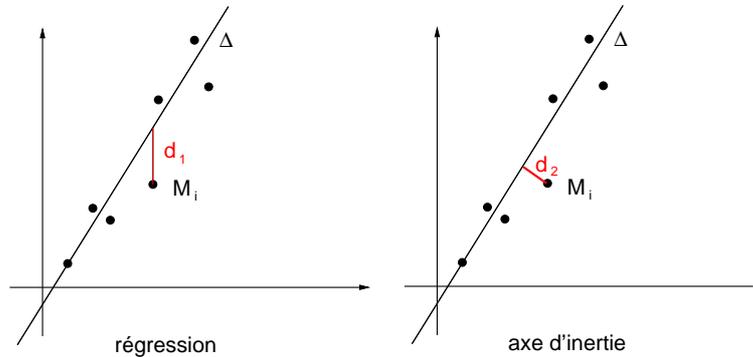


FIG. 14.12 – Deux approximations linéaires aux moindres carrés minimisant des distances différentes : à gauche par régression linéaire, à droite par axe d’inertie.

Ces équations s’écrivent sans problème en dimensions supérieures (pour estimer des variétés d’ordre variable). Elles se transcrivent beaucoup plus difficilement à des polynômes d’ordre plus grand car on ne sait pas, en règle générale, exprimer la distance d’un point courant à une telle fonction.

Estimations robustes

Si l’on repose le problème précédent dans le cadre plus général de l’estimation robuste, on est amené à considérer ce problème de façon un peu plus complète [Meer et al., 1991, Rousseeuw et Leroy, 1987]. On cherche ici la meilleure droite représentant au mieux l’ensemble des points M_i sous l’hypothèse d’un bruit entachant la position des points.

Lorsque le bruit est gaussien, les estimations aux moindres carrés nous assurent d’une bonne qualité de l’estimée. Mais il faut souvent tenir compte de la présence, parmi les points, de détections erronées dont les écarts à la droite ne sont pas gaussiens (ce sont des *outliers*⁴).

Une première façon de prendre en compte ces points consiste à faire une première estimation aux moindres carrés, puis à éliminer les points trop éloignés de la droite (par exemple à l’aide d’un test statistique tel celui de Cramer-von Mises⁵). Ces méthodes ont reçu une base théorique dans l’approche dite par **moindres carrés tronqués** (LTS : *Least Trimmed Squares*) ainsi que toutes ses formes dérivées : moindres quantiles carrés (LQS), moindre k^e carré (LKS) [Lee et al., 1998].

Mais cette approche est réputée dangereuse (on peut éliminer successivement les points appartenant à la droite si un *outlier* est trop éloigné de celle-ci). On lui préfère des techniques d’estimation par médiane ou par médianes itérées, plus coûteuses en temps de calcul, mais capables d’éliminer beaucoup plus de points erronés.

Dans une estimation par médiane, on choisit 2 points quelconques M_i et M_j de l’ensemble. On leur associe une droite qui fournit des coefficients a_{ij} et b_{ij} . Par combinatoire sur l’ensemble des points, on obtient deux ensembles de coefficients : $a = \{a_{ij}\}$ et $b = \{b_{ij}\}$. Les ensembles a et b sont triés séparément et l’on choisit les valeurs médianes \bar{a} et \bar{b} comme estimateurs des paramètres de la droite. Le filtrage médian peut tolérer jusqu’à 50% – ϵ d’*outliers*. Dans le cas de l’estimation simultanée de deux variables, l’estimation par médiane permet de tolérer jusqu’à 25% d’*outliers* ($0,5^2$).

De nombreux estimateurs robustes ont été calculés à partir de la notion de médiane. Ainsi, le minimum des médianes des distances quadratiques (LMS *Least Median of Squares*) détermine le centre d’un nuage de points en

⁴on appelle *outlier* un point qui n’est pas régi par la même distribution statistique que les points dont on recherche une approximation.

⁵Le test de Cramer-von Mises calcule une distance entre la loi gaussienne théorique et la répartition expérimentale des erreurs. Un point est rejeté si sa distance à la distribution des autres points est trop improbable sous l’hypothèse gaussienne [Saporta, 1990].

calculant pour chaque point la distance à tous les autres, puis en caractérisant la distance de ce point au nuage par la valeur moyenne des distances et enfin en choisissant comme centre du nuage celui dont la médiane est minimale [Rousseeuw et Leroy, 1987].

D'autres approches robustes reviennent à une estimation par filtrage (comme le font la moyenne ou la régression), mais en pondérant chaque point i par un coefficient $\rho(x_i)$ qui tend vers 0 pour des points loin de la solution. Une autre approche consiste à intégrer cette pondération des points lointains directement dans la distance (c'est le cas des M-estimateurs [Huber, 1981]). Dans ce cas, les « distances »⁶ croissent moins vite que la loi quadratique ou décroissent même pour tendre vers 0 pour les grandes distances. On les appelle des M-estimateurs.

Une autre famille d'estimateurs robustes s'appuie sur des recherches systématiques. c'est par exemple le cas de la méthode RanSac (*Random Sample Consensus*) [Fischler et Bolles, 1981]. Cette méthode consiste à choisir 2 points et à adopter la droite qui les joint comme approximation. On mesure alors le nombre de points qui sont en accord avec cette hypothèse (par exemple à l'aide d'un critère de distance du point à la droite). On choisit finalement le couple de points qui conduit au plus fort consensus. On montre qu'il n'est pas nécessaire de tester tous les couples mais qu'un petit nombre choisis aléatoirement (de l'ordre de quelques pourcents) est usuellement suffisant.

Estimation d'un mélange de droites

Si l'on sait que n droites existent dans l'ensemble des points, il est possible de faire une classification au sens des nuées dynamiques par exemple. Pour cela on initialise le processus en choisissant n droites (issues par exemple de l'ensemble des points par tirage aléatoire) représentées par leurs paramètres $\{a_0^i, a_1^i, i = 1, \dots, n\}$. On attribue chaque point à la droite qui minimise la distance 14.1 ou mieux 14.2. Après classification de tous les points on estime pour chaque droite ses paramètres $\{a_0^i, a_1^i\}$ par l'une des méthodes vues ci-dessus. On itère le processus enchaînant classification et estimation. On converge vers une solution qui dépend assez fortement de l'initialisation.

Une autre famille de méthodes utilise la version floue des nuées dynamiques (les C-moyennes-floues ou *Fuzzy-C-means*), dans laquelle chaque point k « appartient » à l'une des droites i avec une appartenance $\mu_i(k)$ fonction de sa distance à la droite. Pour tout point k , on vérifie : $\sum_{i=1}^n \mu_i(k) = 1$ [Bezdek, 1981].

Extension de l'estimation à des coniques

Dans le cas où l'on recherche des formes représentées par des cercles ou des ellipses, on peut choisir une paramétrisation matricielle de la forme quadratique les représentant. Chaque conique indiquée par i est alors décrite par son centre q_i , son rayon r_i et sa matrice d'ellipticité A_i :

$$[(\mathbf{x} - q_i)^t A_i (\mathbf{x} - q_i)] = r_i^2$$

où \mathbf{x} est le point courant du plan. La distance d'un point \mathbf{x}_k à cette conique s'exprime par :

$$d_i^2(k) = \left[[(\mathbf{x}_k - q_i)^t A_i (\mathbf{x}_k - q_i)]^{1/2} - r_i \right]^2$$

L'estimation des paramètres se fait de la même façon que précédemment à partir d'une conique initiale souvent circulaire.

14.6.2 Approximations polygonales, simplification de contours polygonaux

Lorsque l'on dispose d'une courbe continue ou finement échantillonnée et que l'on souhaite la réduire à une ligne polygonale, on dispose de très nombreux algorithmes qui proposent soit des critères d'approximation soit des

⁶Les distances ainsi définies ne vérifient en fait plus les propriétés d'une distance.

prises en œuvre différents. Le plus connu est l'algorithme de la corde.

Algorithme de la corde (ou de Ramer)

C'est un processus de subdivision qui peut être entrepris soit de façon récursive (chaque segment créé fait l'objet d'une nouvelle subdivision), soit de façon itérative (la courbe Γ est considérée globalement à chaque étape). C'est cette dernière version que nous examinons sur l'exemple de la figure 14.13. Les sommets du polygone sont choisis successivement comme les points de Γ les plus éloignés des cordes précédemment tirées. Le processus s'arrête lorsque la nouvelle distance candidate est inférieure à un seuil ϵ fixé. Très employé dans de nombreuses applications par la simplicité de sa mise en œuvre, l'algorithme de la corde n'est pas très rapide. Il ne garantit pas non plus une convergence uniforme vers la courbe finale car, dans son implémentation itérative, il se peut que la distance à l'étape n soit supérieure à celle à l'étape $n - 1$.

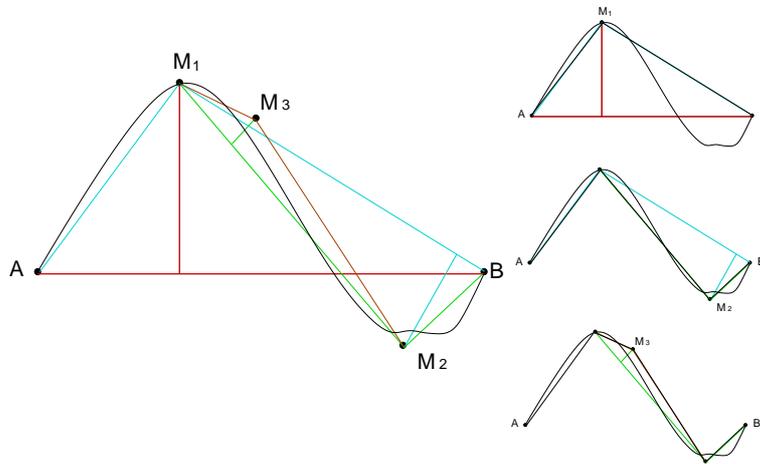


FIG. 14.13 – Algorithme de la corde. Les points M_1 , M_2 et M_3 sont successivement sélectionnés pour créer la ligne polygonale représentant la courbe AB . Le critère de sélection est la distance maximale à la corde polygonale précédemment obtenue.

Algorithme de Dunham

Dans cette méthode on cherche à supprimer les cordes successives qui sont presque alignées. On se fixe donc une tolérance angulaire ϵ . Partant de A , on couvrira par une même corde tous les points de la courbe qui se trouvent dans la tolérance et l'on choisira le premier point du polygone comme le dernier point de cet ensemble. Le processus est répété en ce point.

Algorithme de Wahl et Danielsson

Son objectif est de minimiser l'aire laissée entre la courbe et la ligne polygonale.

Algorithme progressif

Enfin, un algorithme progressif utilise le critère de distance de tout point de la courbe au segment d'approximation, mais dans un schéma progressif. Dans ce schéma, partant d'une extrémité, la courbe est parcourue, et le

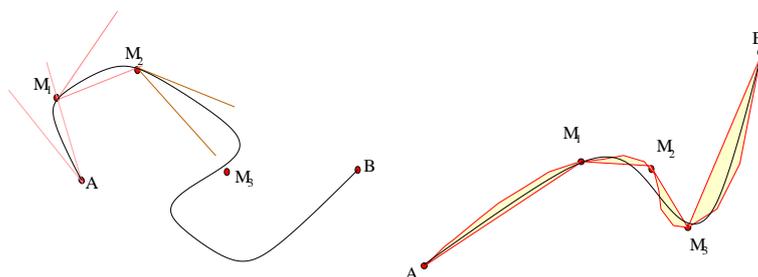


FIG. 14.14 – Algorithme de Dunham, à gauche, les points M_1 , M_2 et M_3 sont choisis lorsque la tolérance angulaire ϵ est insuffisante à approximer la courbe. A droite, algorithme de Wahl et Danielsson : il cherche un minimum des surfaces laissées entre la ligne polygonale et la courbe.

point courant est retenu comme extrémité d'un segment d'approximation s'il est le dernier point visité permettant une approximation de tous les points déjà visités, à une tolérance ϵ donnée.

14.6.3 Approximation par des splines

L'approximation d'un ensemble de points de contours par des polynômes d'ordre plus élevé que les droites est une bonne façon de représenter ce contour [Chalmond, 2000]. Le choix du degré du polynôme se fait généralement de façon à assurer des propriétés de continuité au contour obtenu :

- les segments de droite (degré 1) assurent la continuité du contour,
- les approximations par polynômes d'ordre 2 permettent d'avoir des dérivées continues,
- les polynômes d'ordre 3 permettent d'avoir des courbures continues.

Les fonctions splines sont de bons candidats pour ce type d'approximation. Nous en avons déjà vu l'usage à diverses occasions : au chapitre 10.4.2 lorsque nous avons discuté du rééchantillonnage des images, au chapitre 12.5.2 lorsque nous avons estimé des fonctions régulières lors de segmentation par régions. Deux types de fonctions peuvent être utilisées [Unser et al., 1993a, Unser, 1999] :

1. les **fonctions interpolantes** qui passent exactement par tous les points de contours qui sont utilisés pour calculer la spline ;
2. les **fonctions approximantes** qui ne passent pas nécessairement par les points mais qui s'en approchent de façon contrôlée.

Approximation

Dans le cas des fonctions approximantes, on définit la spline d'ordre k par morceaux en fonction d'une variable continue u , et des m points de contrôle P_i qui la déterminent. Q_i et P_i sont des fonctions de même dimension : des vecteurs de \mathbb{R}^2 si les points P_i sont définis par leurs coordonnées du plan ou des vecteurs de \mathbb{R}^3 dans l'espace.

Le morceau i est défini par :

$$Q_i(u) = \sum_{r=0}^{k-1} P_{i+r} b_r^k(u) \quad i = 0, \dots, m - k + 1 \quad (14.3)$$

Les fonctions $b_r^k(u)$ sont définies à partir des fonctions $\beta^k(u)$ qui ont été introduites au chapitre 10.4.2 équations 10.5 et suivantes :

$$b_r^k(u) = \beta^k \left[u - \frac{r-1}{2} \right]$$

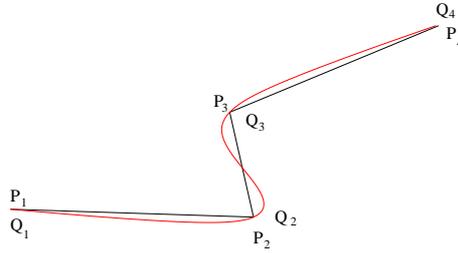


FIG. 14.15 – Dans le cas d’une interpolation spline, les jonctions entre les morceaux de splines sont les points de contrôle eux-mêmes.

Les positions des points de raccord entre splines sont données par $u = 0$ dans l’équation 14.3 :

$$Q_i(0) = \sum_{r=0}^{k-1} P_{i+r} b_r^k(0) \tag{14.4}$$

Il y a de nombreuses façons de définir les variables u . La plus naturelle consiste à considérer que les échantillons sont régulièrement distribués en fonction de u (attention cela conduit à une paramétrisation où u n’est pas l’abscisse curviligne le long du contour), alors $b_0^k(0) = 0$ et $b_{k-1}^k(1) = 0$, on peut réécrire cette équation [Goshtasby et al., 1990] sous la forme :

$$Q_i = \sum_{r=0}^{k-2} P_{i+r} b_r^k(0)$$

qui prend la forme d’une convolution : $Q = P * H_k$. Ceci montre que les Q sont des points filtrés issus des P .

Interpolation

Dans ce cas, les coefficients des splines de l’équation 14.3 sont inconnus et l’équation devient :

$$Q_i(u) = \sum_{r=0}^{k-1} c_{i+r} b_r^k(u) \quad i = 0, \dots, m - k + 1$$

et les coefficients inconnus c_i sont déterminés en contraignant les splines à passer par les points de contrôle P_i :

$$P_i(0) = \sum_{r=0}^{k-1} c_{i+r} b_r^k(0)$$

dont on a vu au chapitre 10.4.2 les divers modes de résolution soit par inversion de matrice soit par filtrage (cf. figure 14.15).

14.7 Transformation de Hough

La transformation de Hough est un outil de traitement des images dont l’usage déborde largement l’applica-tion pour laquelle nous la présentons maintenant [Ballard, 1978, Sklansky, 1978, Maître, 1985b]. Ces applications seront vues plus tard. Nous nous intéressons actuellement au sous-problème suivant :

Connaissant un ensemble de pixels P (potentiellement bruité) appartenant à une frontière que l’on sait poly-gonale, comment déterminer le nombre de segments impliqués dans cette ligne polygonale et leur position ?

Exprimé dans les termes que nous avons vus précédemment, on se retrouve devant un difficile problème de **mélange**. Les solutions combinatoires (où l'on recherche systématiquement les classes possibles) sont généralement inacceptables. La transformation de Hough nous propose au contraire une solution élégante.

Elle consiste à transformer un problème inconnu (trouver des droites) en un problème mieux connu : retrouver des nuages de points. Ceci s'obtient en associant à l'espace de l'image (dénnoté \mathcal{I} et défini par ses variables d'espaces $\{x, y\}$), un espace de paramètres, dénoté par \mathcal{H} (dans le cas où l'on recherche des droites représentées par l'équation $y = ax + b$, on a $\mathcal{H} = \{a, b\}$). Pour faire cette transformation il y a 2 façons : la transformation de Hough de I à m et la transformation de Hough de m à I .

14.7.1 Définitions

Transformation de I à m

Dans cette définition, on associe à tout point $M_i = (x_i, y_i)$ de P toutes les droites du plan. Elles sont définies dans \mathcal{H} par : $b = -x_i a + y_i$, ce qui est l'équation d'une droite. A un point de \mathcal{I} est associée une droite de \mathcal{H} (d'où le nom de I à m : à 1 point sont associés m points, *many* en anglais) et à un point de \mathcal{H} une droite de \mathcal{I} .

Lorsque l'on transforme tous les points de P par la transformation, on associe à \mathcal{I} un ensemble de droites qui, idéalement, se coupent en des points H_k . Le nombre de transformations à faire est $\nu = N = \text{Card}(P)$. Les transformés des H_k sont les droites cherchées dans \mathcal{I} . Ces points H_k s'obtiennent aisément en ne conservant que les intersections des droites de \mathcal{H} et en recherchant les nuages d'intersections (figure 14.16).

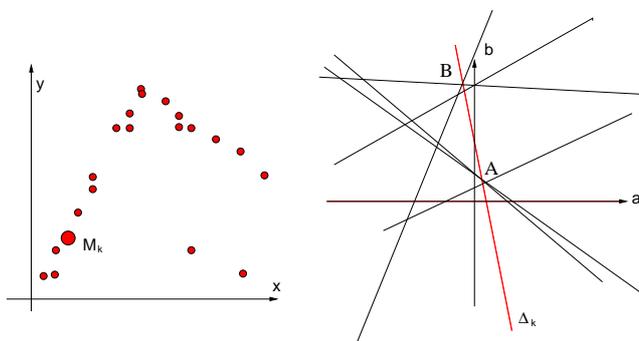


FIG. 14.16 – Transformation de Hough de I à m . Un point M_k est transformé en une droite Δ_k de l'espace des paramètres $\{a, b\}$. On a représenté quelques-unes seulement de toutes les droites existantes. Elles définiront finalement 2 points d'intersection, l'un en A, l'autre en B.

Transformation de m à I

Dans cette définition, on choisit d'associer à tout bi-point $\{M_i, M_j\}$, une droite Δ_{ij} de \mathcal{I} qui se transforme en un unique point⁷ Q_{ij} de \mathcal{H} . Les coordonnées a_{ij} et b_{ij} du point Q_{ij} se déduisent immédiatement de l'équation de Δ_{ij} . En combinant tous les couples de points de P , on obtient tous les points Q de \mathcal{H} . Le nombre de transformations à accomplir est $\nu' = \frac{1}{2}N(N - 1)$. L'espace de Hough est alors constitué de ν' points que l'on identifie immédiatement aux intersections des ν droites obtenues par la transformation de I à m .

On se ramène donc, comme précédemment, à la recherche des nuages de points les plus denses du plan \mathcal{H} .

⁷la dénomination « m à I » se traduirait alors précisément par « 2 à 1 », attention, m n'a pas la même signification dans les 2 définitions.

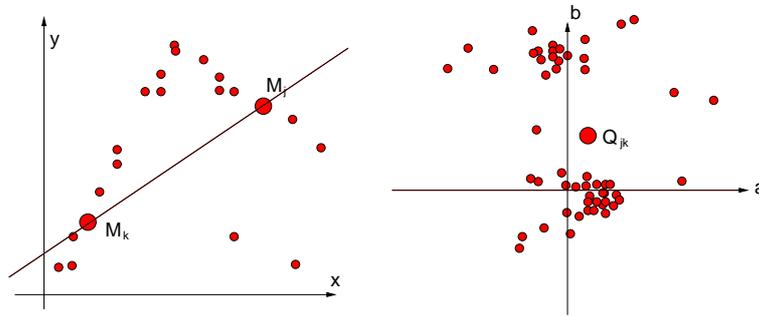


FIG. 14.17 – Transformation de Hough de m à 1 . Deux points M_j, M_k définissent une droite qui se transforme en 1 point Q_{jk} de l'espace des paramètres (a, b) .

Passage d'une TH à l'autre

Il est assez simple de montrer que l'accumulateur de la TH de m à 1 se déduit de celui de la TH de 1 à m par un seuillage au niveau 1. Ces deux transformations sont donc tout à fait équivalentes dans leurs résultats.

Généralisation

On appellera transformation de Hough associée à une forme paramétrée par les variables $\{a_i\}$ la transformation qui fait passer de l'espace image \mathcal{I} à l'espace des paramètres \mathcal{H} .

La transformation est de 1 à m si on associe à un point de \mathcal{I} la variété de \mathcal{H} décrivant toutes les formes passant par ce point. La transformation est de m à 1 si l'on associe à toutes les combinaisons m points de \mathcal{I} une seule forme (donc un point unique de \mathcal{H}).

Il existe généralement de nombreuses transformations de n à m ($n < m$), où l'on se place dans une situation intermédiaire : en associant n points de \mathcal{I} on limite le sous-espace d'arrivée à une partie de la variété de la transformation de 1 à m . Par exemple, on peut utiliser simultanément la position d'un point de contour et la direction de la tangente en ce point. Ces problèmes n'étant pas directement liés à la recherche de contours, des exemples seront vus plus loin.

TH et filtrage adapté

Sklansky [Sklansky, 1978] a démontré qu'il y a équivalence entre la TH et la détection par filtrage adapté (par corrélation) dans une situation particulière mais fréquente. Si les seules inconnues d'un problème de détection sont les deux translations du plan (δx et δy), c'est à dire si la forme ne subit ni changement d'échelle ni rotation, ni transformation perspective. Soit $f(x, y)$ la forme caractéristique de l'objet disponible ($f(x, y) = 1$ si (x, y) appartient au contour de l'objet, et 0 sinon). Soit $g(x, y)$ la forme recherchée. Le filtrage adapté conduit à une détection optimale au sens de l'erreur quadratique moyenne, en recherchant les maximums de la fonction de corrélation :

$$C_{fg}(\delta x, \delta y) = \iint f(x, y)g(x - \delta x, y - \delta y)dx dy = f(x, y) * g(-x, -y)$$

Cette démonstration est assez simple si l'on voit que chaque point M de coordonnées (x, y) de l'image f est candidat à être associé à un point quelconque de g de coordonnées (u, v) si l'on décale g du vecteur $x - u, y - v$ (cf. figure 14.18). Cette contribution à l'accumulateur de position $\delta x = x - u, \delta y = y - v$ vaut exactement :

$f(x, y)g(u, v)$. Sommé sur toute la courbe f on obtient :

$$\begin{aligned}\gamma(u, v) = \gamma(-\delta x, -\delta y) &= \iint f(x, y)g(x - \delta x, y - \delta v)dx dy \\ &= f(x, y) * g(-x, -y)\end{aligned}$$

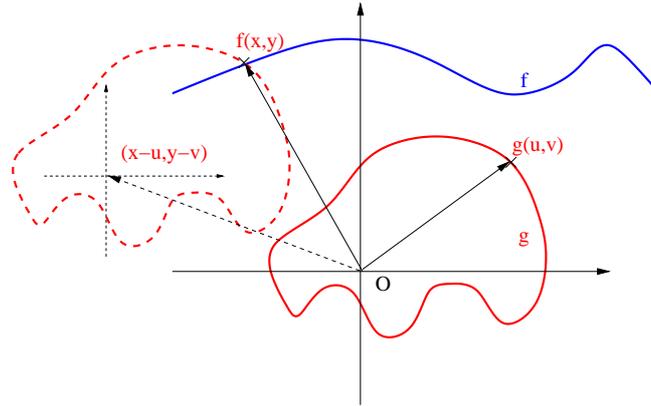


FIG. 14.18 – Transformation de Hough : le point de coordonnées x, y de f sera associé au point $g(u, v)$ de g si l'on fait subir à cette courbe g une translation de $(x - u, y - v)$.

Ce qui est bien le résultat cherché. En fait on montre que la TH agit en pratique en inversant l'ordre des intégrales par rapport au filtrage adapté. Celui-ci procède usuellement en calculant l'intégrale ci-dessus pour un décalage u, v donné, en sommant sur tous les points du plan tandis que la TH choisit un point du plan et incrémente toutes les intégrales (les compteurs) qui lui sont associées par un déplacement u, v quelconque. Cette inversion des opérations peut permettre de tirer profit d'a priori importants pour accélérer les calculs.

Grâce à cette propriété la TH peut hériter des nombreux résultats obtenus en filtrage adapté : pour prendre en compte le rôle du bruit si celui-ci à une densité de probabilité connue, pour déterminer la taille des accumulateurs lorsque la PSF est connue, etc. [Maître, 1985b, Princen et al., 1992].

14.7.2 Mise en œuvre de la TH

Discretisation de \mathcal{H}

Dans la pratique le plan \mathcal{H} est discrétisé. Chaque cellule de position a_i, b_i de largeur $\delta a, \delta b$ est appelée accumulateur. Elle contiendra un nombre d'autant plus important que la droite $y = a_i x + b_i$ sera plus probable dans \mathcal{I} .

Au début de la transformation, tous les accumulateurs sont à zéro. Les points sont alors visités successivement (individuellement dans la TH de l à m , combinatoirement pour les autres). Chaque hypothèse donne naissance à des votes pour des cellules particulières. Pour chaque vote, l'accumulateur de la cellule est incrémenté de 1.

Lorsque tous les candidats ont voté, on recherche les accumulateurs de \mathcal{H} de comptes localement maximaux.

Le choix de la taille des accumulateurs est donc un problème délicat.

1. Des accumulateurs trop petits recevront très peu de votes et donneront des statistiques peu significatives. On ne saura décider de la position d'un maximum entre plusieurs cellules voisines de comptes faibles et proches.

2. Des accumulateurs trop grands donneront une mauvaise précision dans la détection de la droite.

Par ailleurs, la mise en œuvre de la TH nécessite un espace mémoire de taille $\frac{\Delta a \Delta b}{\delta a \delta b}$ où Δa et Δb représentent les domaines de variation de a et b .

Des solutions ont donc été proposées pour adapter localement la taille des accumulateurs (très précis près des lieux où les votes sont nombreux). Cela peut se faire dynamiquement par des approches de subdivision récursive des cellules (selon le principe des **quad-tree** cf. 12.2.4), ou par des passages successifs sur l'ensemble des points de P , le premier passage permettant de dimensionner les cellules.

Ce problème est particulièrement important dans le cas où l'on recherche des formes possédant de nombreux paramètres, puisque l'espace \mathcal{H} est l'espace produit de tous ces paramètres. Il est alors de grande dimension.

Paramétrisation des formes

Le choix d'une bonne paramétrisation des formes recherchées apparaît immédiatement à la lecture du paragraphe ci-dessus si l'on souhaite appliquer la démarche proposée à la recherche de droites. Dans le cas d'une paramétrisation cartésienne ($y = ax + b$), les domaines de variation de a et b sont potentiellement de moins l'infini à plus l'infini, conduisant à un insoluble problème de résolution de cellule. Par ailleurs, si l'on examine la variable a par exemple, on voit que, statistiquement, elle décrit 25% des droites du plan pour $a \in [0, 1[$ et 25% pour $a \in]1, \infty[$, conduisant donc à des cellules qui seront très inégalement remplies si leur taille est uniforme.

On peut cependant choisir une paramétrisation plus heureuse pour la droite, par exemple la paramétrisation normale (cf. figure 14.19). Dans ce cas la droite $y = ax + b$ est repérée par la distance ρ et l'angle θ .

La transformation de I à m s'écrit : $\rho = x_i \cos \theta + y_i \sin \theta$, et la forme associée à un point est donc une sinusoïde dans l'espace $\rho = f(\theta)$, d'équation :

$$\rho = \sqrt{x_i^2 + y_i^2} \cos(\theta + \phi)$$

avec : $\cos(\phi) = x / \sqrt{x_i^2 + y_i^2}$.

Cette paramétrisation est tout d'abord homogène puisque les angles θ sont a priori équiprobables, ainsi que les distances ρ . Elle fournit également des domaines de variation finis : $\Delta\theta = [-\pi, +\pi[$ et $\Delta\rho = [0, \sqrt{2}L]$ si L est le côté de l'image, puisqu'une droite dont ρ serait supérieur à la longueur de la diagonale de l'image ne serait pas visible dans l'image.

La transformation de m à I associe au doublet M_i, M_j le point de \mathcal{H} de coordonnées :

$$\begin{aligned} \rho_{ij} &= \frac{|x_i y_j - x_j y_i|}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \\ \theta_{ij} &= -\text{Arctg} \left[\frac{x_j - x_i}{y_j - y_i} \right] \end{aligned}$$

14.7.3 Détections par TH

La TH est très souvent utilisée, comme nous l'avons vu précédemment, pour détecter des droites. Les autres courbes paramétrées détectées par la TH sont :

1. les paraboles (3 paramètres si l'on connaît la direction de leur axe, 5 dans le cas général). Par exemple pour détecter des côtes dans les radiographies du corps humain) ;
2. les cercles (3 paramètres) par exemple pour détecter des cuves ou des citernes en imagerie aérienne. On paramètre alors le cercle par son centre et son rayon et la forme associée dans cet espace est un cône, centré au centre du cercle.
3. les ellipses (5 paramètres : les 3 du cercle + la direction de l'axe et l'aplatissement ou l'ellipticité), également utilisée pour détecter des formes circulaires vues en perspectives.

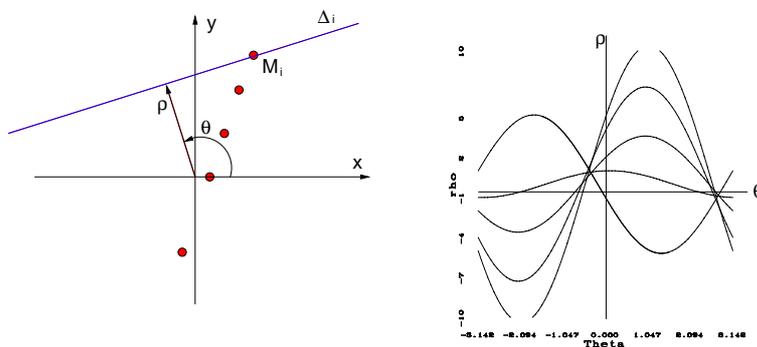


FIG. 14.19 – Transformation de Hough de I à m en coordonnées normales. Une droite Δ_i est repérée par l'angle θ et la distance au pied de sa normale ρ . Dans l'espace \mathcal{H} , chaque point donne une sinusoïde.

4. des traces sinusoïdales (5 paramètres), forme fréquemment recherchée en expérimentation en physique ou en géologie.

La TH est également utilisée pour suivre des formes quelconques sous des hypothèses de transformations particulières. Les paramètres sont alors ceux qui caractérisent la déformation : changement d'échelle, rotation à 2 ou 3 D, projection perspective, etc.

14.8 Conclusion

Il n'est pas possible d'épuiser toutes les variétés des représentations et des modélisations qui ont été adoptées pour décrire les formes des objets en traitement des images. en effet, beaucoup de ces représentations tirent profit de spécificités attachées à la famille particulière des objets à traiter pour une application donnée dans un contexte précis. Nous avons cependant vu que toutes les représentations cherchent à marier la simplicité de la description et la fidélité à la forme. En cela, elles relèvent déjà toutes de la reconnaissance des formes, puisque ces deux objectifs sont aussi ceux que la reconnaissance des formes se donne.

Chapitre 15

Formation de l'image et couleur

Chapitre rédigé par Henri MAÎTRE et Anne-Catherine CARRILERO

15.1 Formation des images

15.1.1 Imagerie cohérente et incohérente

Deux grandes classes d'images doivent être distinguées qui ont des propriétés très différentes (voir chapitre 1) : les images obtenues en imagerie cohérente et celles obtenues en imagerie incohérente. Pour compliquer les choses, des images sont faites en imagerie "partiellement cohérente", comme celles obtenues en microscopie par exemple ; elles présentent des propriétés intermédiaires entre celles des deux types principaux.

Imagerie cohérente

Une source est cohérente si elle présente deux types de propriétés [Pérez, 1991] :

1. elle est **temporellement cohérente**, c'est-à-dire monochromatique (une seule longueur d'onde), et les photons qu'elle émet présentent une relation de phase qui se maintient dans le temps au delà de la durée de vie des photons ;
2. elle possède une **cohérence spatiale**, c'est-à-dire que les divers points de la source émettent des rayonnements en phase.

Les sources naturelles dans le domaine visible ne sont pas cohérentes. On peut les rendre cohérentes en les filtrant par des filtres monochromatiques très étroits et en les diaphragmant.

Les principales sources cohérentes sont :

- dans le domaine visible : les lasers, les lampes spectrales (Hg, Cd, etc.). Dans la vie quotidienne certaines lampes utilisées en éclairage urbain ou des tubes en éclairage domestique ou professionnel peuvent se comporter, dans des conditions très particulières, comme des sources cohérentes ;
- dans les autres domaines des ondes électromagnétiques : le radar est le type-même de source cohérente ; il émet un rayonnement dans les gammes centimétriques (voir [Maître, 2001]), certains rayonnements utilisés en médecine ou en contrôle non destructif : PET (Imagerie par émission de Positons), rayons γ , etc.
- dans le domaine des ondes acoustiques : en imagerie ultrasonore, en prospection sismique, etc. Les sources sont alors des transducteurs piézo-électriques ou des générateurs électro-mécaniques.

Lors de la formation d'une image en éclairage cohérent, une source primaire est réfléchiée ou diffractée par la scène que l'on veut imager, chaque point de cette scène est une source secondaire de rayonnement caractérisée

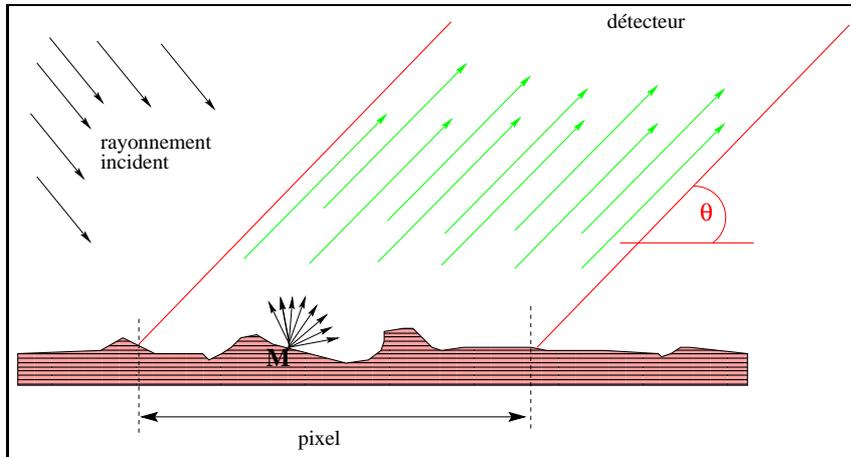


FIG. 15.1 – Formation d’une image : le rayonnement renvoyé par chaque point de la scène est collecté (par exemple par une sélection angulaire en fonction de θ) et sommé sur le capteur. En éclairage cohérent on somme des amplitudes complexes, en éclairage incohérent, on somme des densités spectrales d’énergie.

par sa phase et son amplitude dans chaque direction de l’espace (principe d’Huygens). Le capteur sélectionne les seules contributions provenant du pixel dont il fera l’image (sur des critères angulaires ou sur des critères de temps de propagation) et construira un signal $R(k, l)$ égal à (nous considérons ici une sélection angulaire) :

$$R(k, l) = \left[\left| \int_{x-\delta x}^{x+\delta x} \int_{y-\delta y}^{y+\delta y} \rho(x, y) a(x, y) \exp(-2\pi\phi(x, y)) dx dy \right| \right]^2 \quad (15.1)$$

Le terme $a(x, y)$ exprime les effets des absorptions et des atténuations dues à la réflexion sur le pixel ; il s’exprime en fonction de l’amplitude de l’onde incidente et des propriétés du matériau. Le terme de phase comporte d’une part les retards dus à la propagation (de la source au point de la scène, puis de la scène au capteur), ainsi que les déphasages introduits par l’interaction avec la scène (ce que l’on appelle fréquemment la phase propre du pixel). Le terme $\rho(x, y)$ traduit les atténuations dues à la propagation entre le point du sol et le capteur. La sommation se fait sur les amplitudes complexes. Le capteur mesure une énergie, c’est-à-dire le carré du module de l’intégrale ainsi obtenue.

En imagerie cohérente, tous les effets de phase sont très importants car ils régissent pour une grande part le bilan de l’équation 15.1. Ainsi la contribution d’un point du pixel peut-elle être positive ou négative selon que sa phase est en accord avec celle de ses voisins ou non. La variation très rapide de l’amplitude crée le phénomène de **granularité**¹. L’imagerie cohérente est considérée comme très bruitée par rapport à l’imagerie incohérente (voir [Nicolas, 1999, Maître, 2001]), mais son traitement bénéficie de la connaissance de très fortes propriétés statistiques qui n’existent pas en optique incohérente.

Imagerie incohérente

Le processus de formation des images en optique incohérente est finalement beaucoup plus simple, puisque le capteur intègre les énergies issues de chaque point source du pixel, et cela pour chacune des longueurs d’onde

¹Des termes différents sont employés dans des domaines différents : *granularité* est le terme employé en optique, *tavelure* en astronomie, *chatoiement* en radar. Le terme anglo-saxon consacré est *speckle*.

auquel il est sensible (avec une sensibilité $s(\lambda)$) :

$$R(k, l) = \int_{\lambda_0}^{\lambda_{max}} s(\lambda) \int_{x-\delta x}^{x+\delta x} \int_{y-\delta y}^{y+\delta y} A(x, y, \lambda) dx dy d\lambda \quad (15.2)$$

où $A(x, y, \lambda)$ représente la densité spectrale d'énergie réémise par le point M dans la direction du capteur, et $s(\lambda)$ la sensibilité spectrale du capteur.

En comparant les équations 15.1 et 15.2 on comprend que les effets géométriques interviennent beaucoup moins dans ce second cas, puisqu'ils n'interviennent que par la modulation du terme A (par exemple en traduisant la nature plus ou moins directionnelle des réflexions), cette modulation étant, toutes choses égales par ailleurs, aussi présente dans le terme a de l'équation 15.1.

15.1.2 Interaction matière rayonnement

Cette partie a été rédigée en partie à partir du livre de Patrick Callet [Callet, 1998].

C'est l'un des problèmes les plus complexes de la formation de l'image. Les processus mis en jeu sont très nombreux et opèrent fréquemment de façon simultanée rendant leurs contributions individuelles très difficiles à évaluer. Par ailleurs ces phénomènes peuvent s'expliquer par des modélisations à divers niveaux de la physique, ces niveaux n'étant pas compatibles. Ainsi des explications peuvent provenir d'une simple interprétation en optique géométrique, en optique ondulatoire, en électromagnétique classique, en électromagnétique quantique, en thermodynamique ou en chimie [Callet, 1998]. Dans de nombreux cas il est très difficile de mettre en évidence les phénomènes dominants et des interprétations à des niveaux différents peuvent conduire à des résultats très différents. Pour compliquer le tableau des connaissances actuelles, et devant l'ampleur et la complexité des bilans et des modèles qui cherchent à s'appuyer sur la physique, des modélisations complètement artificielles et purement mathématiques ont été proposées pour simuler les **apparences** dans des domaines comme la synthèse des images. Ces modèles ont été hybridés avec des modèles physiques conduisant à un tableau extrêmement confus.

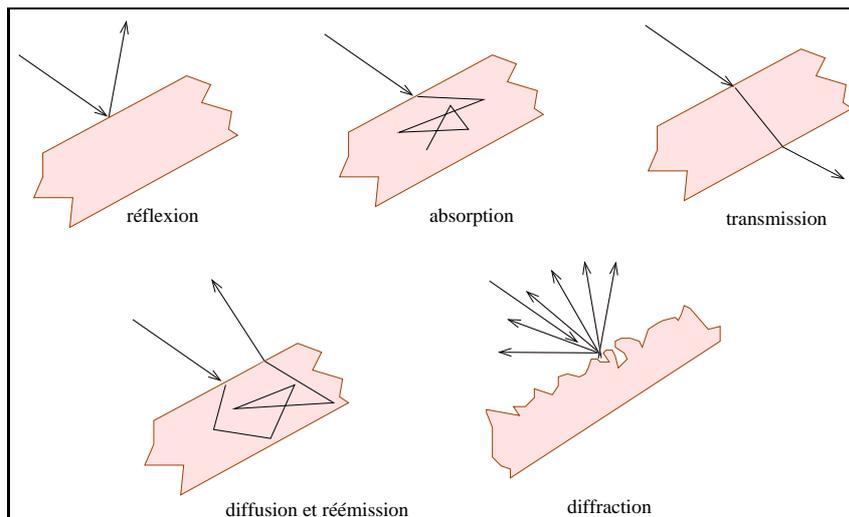


FIG. 15.2 – Les principaux mécanismes de propagation des rayons lumineux

Réflexion et réfraction

Elles résultent d'une modélisation en optique (géométrique : sans interférences, ou ondulatoire : avec interférences et diffraction). Les relations de Descartes ($n_1 \sin i_1 = n_2 \sin i_2$) régissent les réflexions spéculaires (miroirs ou dioptrés parfaits). Des relations empiriques comme la formule de Phong [Phong, 1975] s'en inspirent pour proposer des modèles de miroirs moins parfaits pour lesquels la décroissance de l'intensité autour de l'angle spéculaire est de la forme $I(\theta) \propto \cos^\nu(i_2 - \theta)$, ν étant une constante adaptée à chaque matériau.

Des approches photométriques suivant ces directions s'appuient sur une fonction dite de **réflexion bidirectionnelle** (FRB ou BDRF) qui relie l'énergie reçue dans une direction repérée par rapport à la normale à la surface en fonction de l'énergie émise par la source (elle même repérée par rapport à la normale). Cette FRB est normalement donc fonction de 4 angles [Nicodemus et al., 1977]. Des propriétés d'isotropie peuvent parfois être mises en avant pour ramener ces variables de 2 à 4.

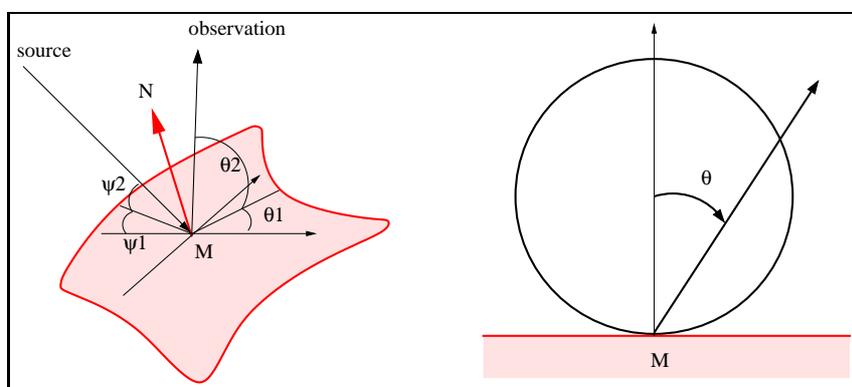


FIG. 15.3 – La fonction de réflectance bidirectionnelle exprime l'intensité réémise par la surface en fonction de l'intensité de la source et des 4 angles repérant source et récepteur par rapport à la normale. A droite, dans le cas de la diffusion lambertienne, l'intensité réémise est proportionnelle au cosinus de l'angle θ

Outre la réflexion spéculaire, la forme la plus usitée de FRB est la loi de Lambert qui correspond à une surface isotrope et parfaitement diffusante. Pour une surface vérifiant la loi de Lambert, la luminance est indépendante de la direction d'observation et ne dépend que de la direction ψ de la source. Pour une surface lambertienne l'énergie réémise par unité d'angle solide est proportionnelle au cosinus de l'angle fait par cette direction avec la normale.

La représentation vectorielle des champs électro-magnétiques conduit des équations de Descartes aux équations de Fresnel. Un modèle de ce type est par exemple utilisé dans les modèles de milieu rugueux de Torrance-Sparrow [Torrance et Sparrow, 1967] qui considère une distribution gaussienne de micro-facettes aléatoires. C'est un modèle de ce type qui a été exploité en synthèse d'images par Blinn [Foley et al., 1995], souvent en concurrence au modèle de Phong dont il ne diffère que peu.

Les lois de Descartes-Fresnel permettent en particulier de relier les champs absorbés et dispersés par le milieu (relations entre parties réelles et imaginaires de ϵ (relations de Kramers-Kronig)). En généralisant les relations de Fresnel aux milieux métalliques (d'indice ϵ complexe), on peut justifier un modèle très classiquement utilisé en synthèse d'images, celui de Cook et Torrance [Cook et Torrance, 1982] qui donne au matériau un indice complexe effectif du type :

$$\hat{n} = n(1 + i)$$

On construit également des modèles de milieux rugueux (rugosité de Bekmann ou de Beckmann et Spizzichino, formules de Goodman pour l'imagerie cohérente et la granularité totalement développée) [Beckmann et Spizzichino, 1963].

On peut également traiter les milieux périodiques ainsi que les réseaux (équations de Kirchoff, diffraction de Bragg).

Dans le cas de propagation dans des matériaux volumiques épais on retrouve là aussi la théorie des réseaux épais de Bragg à partir d'une modélisation en optique ondulatoire.

Dispersion dans les milieux homogènes

Cette modélisation considère la propagation des ondes électromagnétiques dans la matière. On cherche à exprimer les modifications du champ électro-magnétique incident sous l'effet des perturbations induites dans le milieu par ce champ. Les dipôles étant les structures sensibles au champ EM, les plus simples à modéliser, elle s'appuie donc sur la résolution des équations de Maxwell dans un milieu caractérisé par la présence de dipôles [Born et Wolf, 1975]. On établit ainsi des relations entre la fonction diélectrique complexe ϵ et la polarisabilité des dipôles. Une modélisation en est faite par la méthode de Lorentz qui conduit aux équations de Clausius-Mossotti (liant ϵ à la polarisabilité α des dipôles élémentaires). Pour la formation des images, ceci a conduit à la définition d'un **milieu effectif**, c'est-à-dire d'un milieu virtuel permettant de rendre compte des mêmes effets coloriques, défini uniquement par son indice de réfraction effectif. On explique ainsi l'apparence des vitraux et le rôle des ions métalliques comme Co^+ et Fe^+ (équations de Maxwell Garnett) [Berthier, 1993]. On peut également appliquer la théorie du champ moyen (méthode de Bruggeman) qui permet d'éviter de connaître les fonctions diélectriques des particules participant au milieu.

Diffusion et diffraction par des particules

Une autre famille de problèmes provient de la diffusion et de la diffraction de la lumière par des particules de dimension comparable ou inférieure à la longueur d'onde : poussières, molécules ou atomes.

Dans l'approche de Rayleigh, on traite des populations de particules incohérentes dans un milieu très dilué (cela entraîne qu'il n'y aura qu'une interaction au plus pour un rayon donné) de taille très petite devant λ . On montre alors que la puissance diffusée par unité de volume est proportionnelle à λ^{-4} , entraînant d'importants effets colorés en fonction de l'angle d'observation. L'intensité de l'onde traversant un espace de longueur l prend alors la forme :

$$I = I_0 \exp(-ld)$$

où d est la densité optique, proportionnelle à la concentration des particules, et à λ^{-4} .

La théorie de Rayleigh a été complétée par Mie qui a considéré non seulement les dipôles, mais aussi les multipôles. Sa théorie conduit à des résultats asymptotiquement équivalents à ceux de Rayleigh quand la taille des particules tend vers 0.

La diffusion par les particules explique de très nombreux phénomènes physiques : couleur du ciel ou des mers, fumées, etc.

Ces théories ne prenant en compte qu'une seule interaction sur un faisceau, ont été complétées par des approches plus complexes de diffusion multiple, mais celles-ci n'aboutissent pas à des expressions explicites et on leur préfère des démarches plus empiriques comme par exemple les équations du transfert radiatif [Kerker, 1969].

Parallèlement à ces travaux, des approches phénoménologiques ont été faites du phénomène de diffusion pour rendre compte des résultats sans en expliquer les causes exactes. C'est souvent des approches qui sont retenues pour résoudre des problèmes concrets. Une telle approche est par exemple celle de Melamed qui essaie de rendre compte de l'effet de poudres pulvérulentes (en ce cas on ramène le milieu à une seule couche faiblement absorbante de particules de taille grande devant λ) [Melamed, 1963]. Il rend compte ainsi de la neige, de la poussière, du sable, du café, etc. Le modèle de Kubelka-Munk² est particulièrement adapté à présenter les effets des matériaux

²On se reportera à l'article [Duntley, 1942] plutôt qu'à l'original de Kubelka-Munk en allemand.

synthétiques : pigments dans les peintures, plastiques, papier coloré, etc. Il découpe le milieu en couches successives et met en évidence le rôle important du rapport du coefficient d'absorption sur le coefficient de diffusion.

Approche chimique

Dans une telle démarche on s'intéresse essentiellement à la structure chimique des composés. Dans un composé chimique, il y a une étroite relation entre l'énergie de liaison et la position des bandes d'absorption des matériaux qui contrôlent l'aspect extérieur du matériau. Ceci est décrit par les équations de la mécanique quantique appliquées à la structure en question. Ainsi l'augmentation de l'intensité de la liaison de covalence favorise l'apparition de niveaux peu serrés et donc l'absorption se fera aux plus grandes longueurs d'ondes, plutôt vers le rouge que le bleu. Une bande d'absorption dans le violet donnera au matériau une apparence jaune ou rouge très vive.

Ces propriétés expliquent les couleurs vives de certains oxydes. D'autres apparences découlent de la position relative des anions et des cations au sein de la structure cristalline, de l'existence de transitions électroniques particulières, de la dimension de certaines structures, ou de la présence et de la concentration d'impuretés.

15.1.3 Les sources de lumière

1. Illuminant A : température $T = 2856\text{K}$,
2. Illuminant B : $T = 4800\text{K}$, il est proche d'une lumière solaire directe,
3. Illuminant C : $T = 6500\text{K}$, il est proche d'une lumière du jour moyenne,
4. Illuminants D50, D55, D65 et D75 : les températures respectives sont 5000K , 5500K , 6500K et 7500K , ils correspondent à quatre rayonnements diurnes.

15.2 Définition des différents espaces couleur

15.2.1 Trivariance et primaires

Maxwell a montré qu'une lumière blanche peut être reproduite "en apparence" par une combinaison linéaire de trois primaires. Grassman a étendu cette propriété à toute couleur et a donc établi la "trivariance" de l'espace coloré : **tout rayonnement peut être remplacé par une combinaison linéaire de 3 primaires qui aura la même apparence colorée pour un observateur.**

Dans la pratique, la plupart du temps, le choix des primaires se porte sur le rouge, le vert, et le bleu dans les synthèses additives (les seules que nous aborderons ici), et le jaune, le magenta et le cyan (leurs couleurs complémentaires) dans les synthèses soustractives. Dans la synthèse additive, on additionne les primaires pour obtenir la teinte cherchée ; dans la synthèse soustractive, partant d'une lumière considérée blanche, on soustrait à l'aide de filtres les primaires.

Deux rayonnements qui donnent une même impression colorée mais n'ont pas le même contenu spectral sont dits **métamères**. Le métamérisme forme une classe d'équivalence. Le métamérisme naturel exact est extrêmement rare.

Si l'on mélange les primaires rouge, vert et bleu, on obtient les couleurs secondaires qui sont également les primaires soustractives :

- rouge + vert = jaune,
- rouge + bleu = magenta (couleur mauve),
- vert + bleu = cyan (couleur bleu-turquoise),
- rouge + vert + bleu = blanc.

Les couleurs complémentaires (qui additionnées en bonne proportion donnent du blanc) sont donc :

- jaune et bleu,
- magenta et vert,
- cyan et rouge.

Les lois de la colorimétrie [Marion, 1997] sont complétées par :

- l'égalisation visuelle des couleurs demeure si l'on change également les quantités de chaque lumière du mélange. Cette égalisation ne disparaît qu'aux très faibles éclaircissements (perte de la sensation colorée) et aux forts éclaircissements (éblouissement).
- La luminance visuelle d'un mélange de couleurs est la somme des luminances visuelles des luminances de ses composantes.

Un espace colorimétrique est défini par un blanc de référence (c'est-à-dire une répartition $W(\lambda)$ donant une sensation visuelle de blanc) et 3 primaires P_1, P_2 et P_3 . Il existe une infinité d'espaces colorimétriques. On appelle composante trichromatique de la couleur C dans le système colorimétrique (P_1, P_2, P_3, W) , les 3 rapports :

$$T_i(C) = \frac{L_i(C)}{L_i(W)} \quad i = 1, 2, 3$$

où les $L_i(C)$ sont les luminances de la primaire nécessaires pour égaler visuellement la sensation de la lumière C (et idem pour W). La luminance, exprimée en candéla par m^2 , se calcule comme :

$$L = K \int_{400nm}^{800nm} S(\lambda)\sigma(\lambda)d\lambda \quad (15.3)$$

où $S(\lambda)$ est la densité spectrale de puissance de la lumière réfléchiée sur le capteur, K un facteur ne dépendant que de la géométrie de prise de vue et $\sigma(\lambda)$ la courbe de sensibilité spectrale de l'œil.

15.2.2 Choix d'un espace de couleur

Nous désirons créer l'image d'un objet éclairé par une source primaire de lumière (par exemple le Soleil). L'objet est alors considéré, du point de vue du capteur, comme une source secondaire de lumière émettant un rayonnement qui dépend du rayonnement incident auquel il est soumis. Le capteur, par exemple une caméra, est conçu pour recevoir l'information photométrique émise par l'objet et traduire ces données énergétiques en une image. On interpose des filtres entre l'objet et la caméra pour sélectionner le contenu spectral de la lumière mesurée.

Suivant le schéma de Maxwell, on décompose donc le rayonnement arrivant sur le capteur, suivant les trois primaires choisies, à l'aide de filtres appliqués successivement sur le capteur³. On obtient ainsi trois images, par exemple rouge, verte et bleue qui nous permettent de recréer l'image en couleur. Tout point de l'image est donc repéré dans un espace tridimensionnel par ses trois composantes chromatiques R, V et B (dénommées **composantes trichromatiques**), propres au capteur choisi, dont les expressions sont les suivantes :

$$R = \int_{-\infty}^{\infty} S(\lambda) T_r(\lambda) s_R(\lambda) d\lambda$$

$$V = \int_{-\infty}^{\infty} S(\lambda) T_v(\lambda) s_V(\lambda) d\lambda$$

$$B = \int_{-\infty}^{\infty} S(\lambda) T_b(\lambda) s_B(\lambda) d\lambda$$

où $S(\lambda)$ est la distribution spectrale de la source lumineuse secondaire,

³ou de trois capteurs masqués par un filtre chacun

$T_r(\lambda)$, $T_v(\lambda)$, $T_b(\lambda)$ sont les transmittances des filtres rouge, vert et bleu,

$s_R(\lambda)$ est la sensibilité du capteur rouge, et de même pour le vert et le bleu.

L'espace RVB peut être représenté sous forme d'un cube dont trois des arêtes portent les trois primaires. La diagonale reliant les points caractérisés par les trois composantes $R = V = B = 0$ (noir), et $R = V = B = 1$ (blanc) est appelée axe achromatique. Cette représentation conventionnelle laisse croire que les composantes portées par les 3 axes orthogonaux sont indépendantes, mais on sait expérimentalement qu'il n'en est rien.

15.2.3 L'espace RVB de la CIE 1931

La CIE⁴ a défini un espace RVB de référence (CIERV B 1931) à partir des éléments suivants :

1. les 3 primaires ont pour puissance 1 Watt et sont monochromatiques, de longueurs d'ondes : 700 nm pour le rouge, 546 nm pour le vert et 436 nm pour le bleu ;
2. le blanc de référence est choisi comme W_0 de densité de puissance constante et égale à $5,3 \cdot 10^{-2}$ W/nm ; c'est le blanc d'égale énergie.

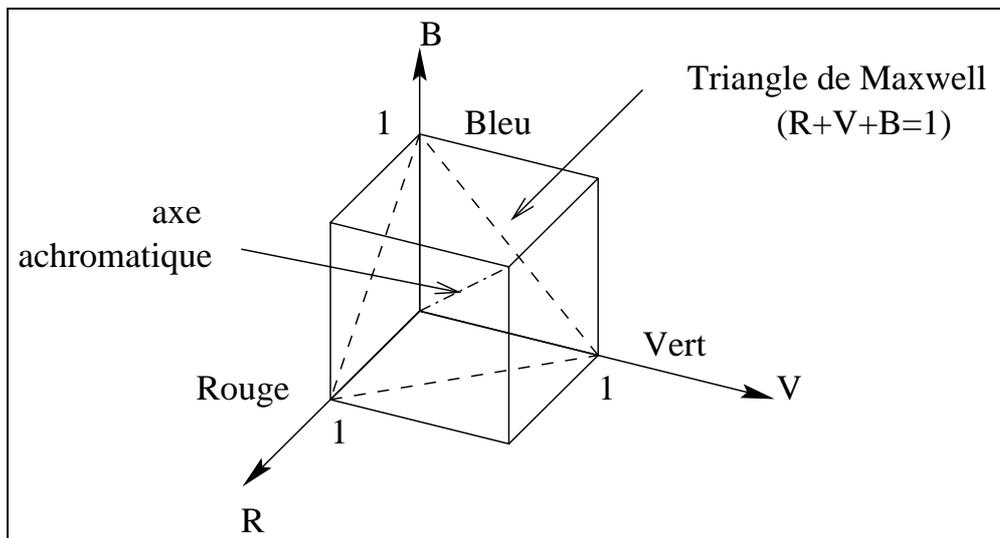


FIG. 15.4 – Le cube des couleurs et le triangle de Maxwell

Le triangle de Maxwell est le triangle défini, dans ce cube, par l'équation $R + V + B = 1$

Les fonctions colorimétriques de RVB CIE 1931

Les fonctions colorimétriques, définies pour le triplet rouge, vert, bleu, ont des valeurs nulles en dehors du domaine visible des longueurs d'onde $[400 \text{ nm}, 800 \text{ nm}]$. Elles sont égales aux **composantes trichromatiques** lorsque le stimulus à caractériser est de luminance constante et d'intégrale unité. Elles sont alors notées : $\overline{r(\lambda)}$, $\overline{v(\lambda)}$, $\overline{b(\lambda)}$. On en déduit l'expression des composantes chromatiques lorsque le stimulus n'est pas de luminance unité, mais de luminance $L(\lambda)$:

$$R = \int L(\lambda) \overline{r(\lambda)} d\lambda$$

⁴CIE=Commission Internationale de l'Eclairage

$$V = \int L(\lambda) \overline{v(\lambda)} d\lambda$$

$$B = \int L(\lambda) \overline{b(\lambda)} d\lambda$$

Les fonctions colorimétriques sont également nommées courbes d'excitation ou fonctions de mélange. Elles sont représentées figure 15.5.

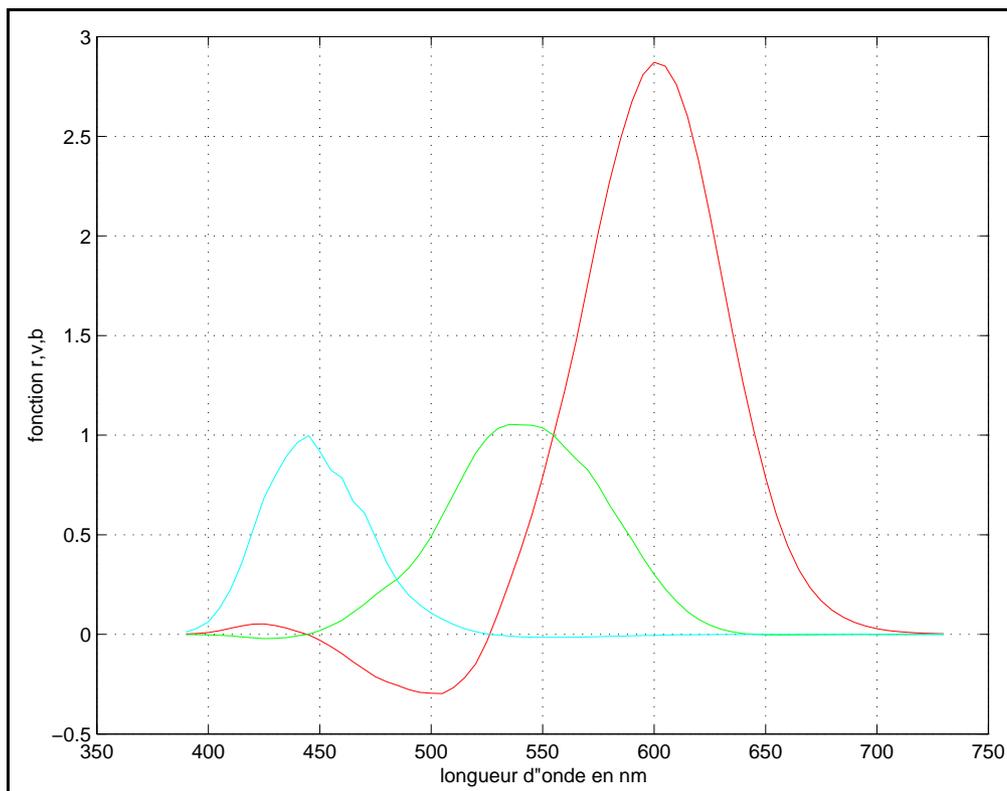


FIG. 15.5 – Fonctions colorimétriques. De gauche à droite : Bleue, Verte et Rouge. Notez les valeurs négatives (en particulier de Rouge), qui caractérisent une inhibition.

Limites de cet espace

1. L'espace RVB n'est pas le meilleur espace pour reproduire la perception visuelle humaine : il est par exemple difficile d'attacher simplement une couleur à un point de cet espace.
2. Les trois composantes RVB sont fortement corrélées (diminuer la composante V, fait apparaître la teinte plus rouge), et il est difficile de séparer la notion d'intensité de la notion de chromaticité.
3. une couleur pure a toujours au moins une composante négative (voir fig 15.5).

Pour éviter ces inconvénients, il a été proposé de passer à des espaces différents : tout d'abord l'espace XYZ, puis des espaces dont la représentation se rapproche de la perception humaine des couleurs, et dans lequel on peut décorréler l'intensité de la couleur (Lab, TLS, visuel hypothétique, et TLC).

Dans ces différents espaces, on définit mathématiquement les caractéristiques de la couleur : la Teinte, la Saturation et le Chroma. On peut également les définir littéralement indépendamment des espaces [Sève, 1996].

Définitions des grandeurs chromatiques

La Teinte est définie de la façon suivante : “Attribut de la sensation visuelle qui peut être décrit par des qualificatifs tels rouge, jaune...”. La teinte détermine donc la couleur d’une surface.

Le Chroma est “le niveau de coloration d’une surface, évalué relativement à **la lumière qu’elle reçoit**. Le chroma d’une surface donnée est un attribut perceptif indépendant du niveau d’éclairement. Pour une surface de chromaticité constante, le chroma augmente avec la clarté de la surface, contrairement à la saturation.”

La Saturation quant à elle peut être définie de la façon suivante : “Niveau de coloration d’une surface, évalué relativement à **sa luminosité**. Il permet d’estimer la proportion de couleur chromatique pure dans la sensation totale. La saturation d’une surface de chromaticité constante est un attribut perceptif indépendant de sa clarté.”

La Clarté est “un attribut d’une sensation visuelle, selon laquelle une surface paraît diffuser plus ou moins de lumière relativement à celle reçue. La clarté s’évalue relativement au diffuseur parfait éclairé dans les mêmes conditions.”

La Luminance est “une grandeur physique qui caractérise une surface émettant un rayonnement dans une direction donnée. C’est le quotient du flux émis par l’angle solide et l’aire apparente de la surface depuis la direction d’émission.”

Les notions de chroma et de saturation sont souvent confondues, leurs expressions formelles seront définies par la suite et leurs notations seront respectivement C et S.

Dans les paragraphes suivants les différents espaces utilisés, faisant intervenir ces grandeurs, seront présentés. Ils nécessitent que l’on définisse tout d’abord l’espace intermédiaire XYZ.

15.2.4 L’espace XYZ

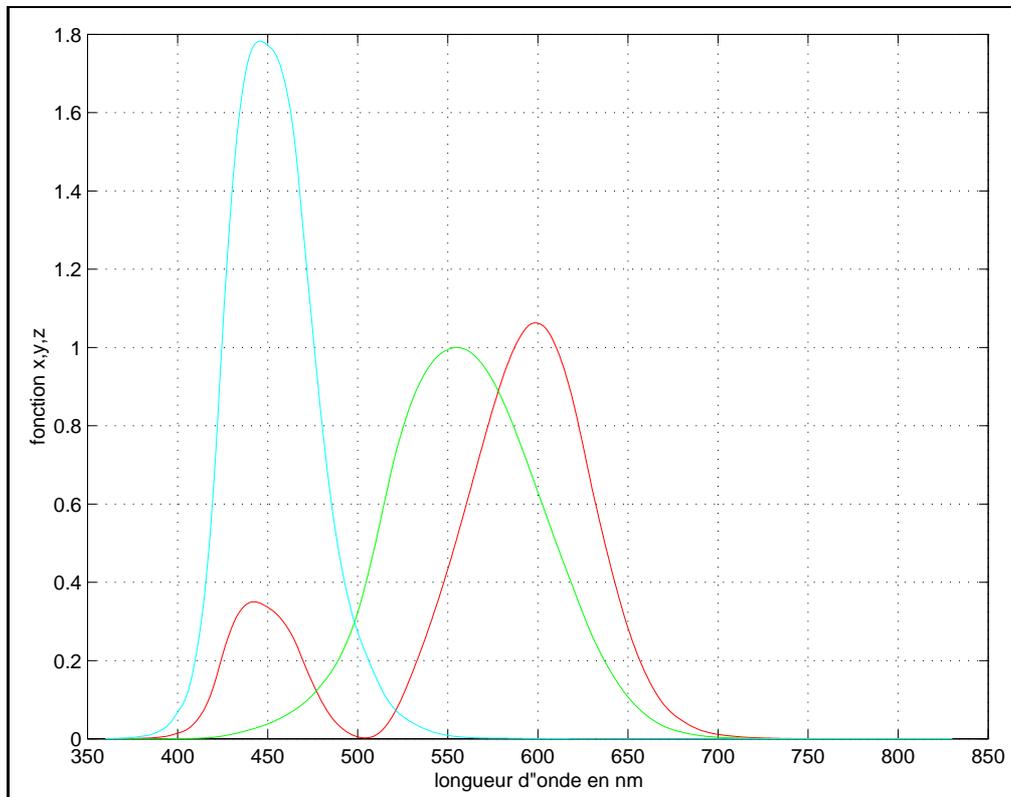
Comme nous pouvons le voir sur la figure 15.5, les fonctions colorimétriques prennent des valeurs négatives, qui correspondent donc à des inhibitions. La CIE proposa un changement de primaires afin d’avoir des fonctions colorimétriques à valeurs positives. Ces primaires furent nommées : $[X]$, $[Y]$, $[Z]$ et les fonctions colorimétriques associées \bar{x} , \bar{y} , \bar{z} . Les nouvelles fonctions colorimétriques sont représentées sur le schéma 15.6. Ces primaires ne sont pas physiques mais irréelles. On choisit en particulier \bar{y} égale à la sensibilité relative spectrale $\sigma(\lambda)$ de l’œil, ce qui permet de relier directement la luminance de la couleur à la valeur Y (voir équation 15.3) : $L(C) = KY$.

Le passage de l’espace RVB à l’espace XYZ se fait par une transformation linéaire. On définit donc une matrice de passage inversible nommée : RVB_XYZ (dont les valeurs dépendent du système physique d’acquisition et d’affichage de l’image si l’on ne se place pas dans RVB 1931). La transformation réciproque se fait grâce à la matrice inverse de RVB_XYZ nommée XYZ_RVB. Ces deux matrices sont, dans le cas du passage des espaces CIE 1931, définies de la façon suivante :

$$RVB_XYZ = \begin{bmatrix} 0.42 & -0.16 & -0.08 \\ -0.09 & 0.25 & 0.16 \\ 0.001 & -0.002 & 0.179 \end{bmatrix}$$

$$XYZ_RVB = RVB_XYZ^{-1} = \begin{bmatrix} 2.78 & 1.75 & 1.13 \\ 1 & 4.59 & 0.06 \\ 0 & 0.057 & 5.59 \end{bmatrix}$$

Les espaces RVB et XYZ n’ayant pas les mêmes blancs (ils sont en fait proportionnels), on définit un espace

FIG. 15.6 – Primaires \bar{x} , \bar{y} , \bar{z} de la CIE.

$X'Y'Z'$ qui se déduit de XYZ par changement de blanc et qui est relié à RVB par :

$$RVB_X'Y'Z' = \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.18 & 0.81 & 0.01 \\ 0.0 & 0.01 & 0.99 \end{bmatrix}$$

$$X'Y'Z'_RVB = RVB_X'Y'Z'^{-1} = \begin{bmatrix} 2.37 & -0.9 & -0.47 \\ -0.52 & 1.43 & 0.09 \\ 0.004 & -0.014 & 1.01 \end{bmatrix}$$

Des matrices semblables existent, permettant de passer par exemple des espaces NTSC, PAL ou SECAM aux espaces de la CIE et d'un type de télévision à l'autre (les deux systèmes européens ont les mêmes blancs et les mêmes luminophores).

A partir de cet espace nous pouvons définir l'espace Lab.

15.2.5 L'espace Lab et le repère TLC

L'espace Lab est un espace uniforme, ce qui signifie que les écarts de couleur dans cet espace sont égaux aux écarts de couleur perçus par un observateur (en première approximation). C'est un espace normalisé par le CIE. L'origine de cet espace vient de la propriété suivante : la luminance d'un rayonnement est indépendante de la chromaticité de ce dernier.

Il est important également de constater que dans le cas de l'espace RVB les composantes trichromatiques sont corrélées. Ainsi il est impossible d'obtenir l'information de luminance indépendamment de la chromaticité, en revanche, cela est possible dans le cas de l'espace Lab. On dispose de trois composantes pour un seul point de l'espace : L la clarté, a et b les coordonnées de chrominance. Celles-ci sont obtenues à partir des composantes X , Y , Z . On définit de plus un blanc de référence dans l'espace XYZ dont les coordonnées X_n, Y_n, Z_n sont déterminées en fonction de la nature de l'illuminant choisi [Sève, 1996].

- Illuminant A : $X_n = 109.85, Y_n = 100, Z_n = 35.58$.
- Illuminant C : $X_n = 98.07, Y_n = 100, Z_n = 118.23$.
- Illuminant D65 : $X_n = 95.04, Y_n = 100, Z_n = 108.88$.

L'espace Lab, tel qu'il a été adopté par le CIE en 1976, porte le nom de CIELab. Il est défini par les équations suivantes :

$$\begin{aligned} L &= 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \\ a &= 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \\ b &= 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right] \end{aligned}$$

Pour des valeurs de l'argument inférieures à 0.008856, la fonction puissance $x^{\frac{1}{3}}$ est remplacée par $7.787x + \frac{16}{116}$.

On peut ainsi voir que le plan orthogonal à l'axe des clartés est le plan chromatique contenant les deux axes a et b . On remarque que l'axe a peut être décomposé en deux couleurs : vert et rouge et l'axe b en bleu et jaune. On a donc :

bleu pour $a = 0$ et $b < 0$ et jaune pour $a = 0$ et $b > 0$

vert pour $a < 0$ et $b = 0$ et rouge pour $a > 0$ et $b = 0$

L'axe achromatique L est donc défini par $a = 0$ et $b = 0$.

Plutôt que de caractériser la couleur par les valeurs a et b , on préfère souvent utiliser les coordonnées cylindriques. Elles permettent d'introduire deux caractéristiques de la couleur : la teinte T et le chroma C dont les expressions sont les suivantes :

$$C = \sqrt{a^2 + b^2}$$

$$\begin{aligned} T &= a \cos \frac{a}{\sqrt{a^2 + b^2}} \quad \text{si } b > 0 \\ &= 2\pi - a \cos \frac{a}{\sqrt{a^2 + b^2}} \quad \text{sinon} \end{aligned}$$

La signification du chroma est la suivante : si dans l'espace Lab on effectue une coupe orthogonale à l'axe des luminances, à une valeur de L fixée, on définit un plan particulier de chromaticité. Le chroma est alors la distance d'un point de ce plan à l'axe achromatique. Les points d'isochroma se situent sur un cercle dont le centre est le point d'intersection entre l'axe achromatique et le plan de chrominance choisi.

La teinte est définie comme un angle, ses valeurs varient donc entre 0 et 2π ou entre $-\pi$ et π .

Remarque : la teinte est une notion délicate car une valeur de teinte à 0 semble très éloignée d'une valeur de teinte égale à 2π , alors que ce sont deux teintes identiques. Ces deux grandeurs sont représentées sur le schéma 15.7.

Caractéristiques de cet espace

1. L'espace Lab est un espace uniforme, donc par définition tout écart entre deux couleurs est égal à l'écart perçu par l'homme c'est donc un espace dont la métrique est proche de celle de l'espace visuel humain.

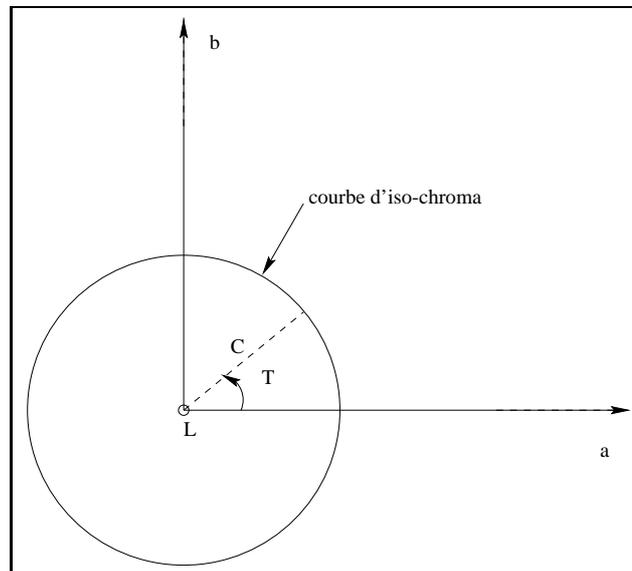


FIG. 15.7 – Coupe orthogonale à l'axe des luminances dans l'espace Lab : le repère TLC.

2. La composante L est décorrélée des composantes chromatiques, Malheureusement le chroma dépend de la clarté.
3. Le passage de RVB à Lab entraîne des pertes de précision numérique qui sont dues d'une part aux changements successifs d'espaces, et d'autre part à l'approximation causée par l'exposant $\frac{1}{3}$.
4. L'accès aux informations de teinte et de chroma est faussé par les différentes étapes nécessaires au calcul. C'est pourquoi il est préférable de passer par l'intermédiaire d'autres espaces pour définir ces grandeurs, notamment l'espace TLC.

15.2.6 L'espace YC1C2

C'est un espace générique. Il est obtenu à partir de l'espace RVB par transformation linéaire. La composante Y est définie comme étant la luminance et $C1$, $C2$ les valeurs de chrominance. On peut écrire cette transformation sous forme matricielle, on obtient alors :

$$\begin{bmatrix} Y \\ C2 \\ C1 \end{bmatrix} = [M] \begin{bmatrix} R \\ V \\ B \end{bmatrix}$$

La matrice M est définie suivant l'espace utilisé : TLS ou visuel hypothétique.

15.2.7 L'espace TLS

C'est un espace créé par les traiteurs d'images pour sa commodité d'emploi. Les composantes d'un point de cet espace sont : Teinte (T) Luminance (L) et Saturation (S). On l'appelle en anglais l'espace HIS (*Hue, Intensity, Saturation*). Les composantes sont définies, en passant à l'espace YC1C2, par la matrice M définie par :

$$M = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$$

On obtient alors les valeurs de T , L , S de la façon suivante :

$$L = Y$$

$$\begin{array}{lll} \text{si } R = V = B & \text{alors } S = 0 & \text{et } T = 0 \\ \text{sinon} & S = 1 - \frac{3\text{Min}(R,V,B)}{R+V+B} & \text{et si } C^2 > 0 \text{ alors } T = \text{acos}\left(\frac{C_1}{\sqrt{C_1^2+C_2^2}}\right) \\ & & \text{sinon } T = 2\pi - \text{acos}\left(\frac{C_1}{\sqrt{C_1^2+C_2^2}}\right) \end{array}$$

La *saturation* est définie comme une distance séparant un point quelconque du point achromatique, on remarquera que ce n'est pas une distance euclidienne. Soit D' la droite passant par un point quelconque et par le point origine de l'axe achromatique. La teinte est définie comme l'angle compris entre la droite D' et l'axe de référence C_1 .

Cet espace présente un inconvénient : la fonction Min , qui définit la saturation, compromet la transformation réciproque passant de l'espace TLS à YC1C2. Pour éviter cela on peut utiliser le système visuel hypothétique [Carron, 1995].

15.2.8 Le système visuel hypothétique

Pour correspondre à la perception visuelle Garbay a défini ce système dont la sensibilité est indépendante de la longueur d'onde, du facteur de pureté et de l'intensité de la couleur considérée [Carron, 1995].

Dans ce modèle on définit une matrice M de passage à l'espace YC1C2 par :

$$M = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ \sqrt{3}/2 & -\sqrt{3}/2 & 0 \\ -1/2 & -1/2 & 1 \end{bmatrix}$$

Par la suite on définit les deux grandeurs suivantes : la *teinte* et le *chroma*.

Soit P un point de l'espace dont on désire connaître les propriétés chromatiques. Soit P' la projection orthogonale de P sur l'axe achromatique. Soit D la droite passant par P et P' .

La teinte est l'angle que font la droite D et l'axe de référence C_1 . Le chroma est la distance séparant le point considéré, de coordonnées (C_1, C_2) , de l'axe achromatique. D'où l'expression suivante pour le chroma : $C = \sqrt{(C_1^2 + C_2^2)}$

Remarques : Différences entre l'espace TLS et le système visuel hypothétique.

L'espace TLS est également appelé système de coordonnées triangulaires. Cela vient du fait que le plan de chrominance est défini par le triangle de Maxwell, l'axe achromatique est l'axe lui étant orthogonal, donc la diagonale principale du cube dans l'espace RVB. On remarquera qu'on différencie les termes de *Chroma* et de *Saturation*. En effet dans l'espace TLS, étant donné la définition de S ($S=f(\text{Min}(R,V,B))$) on peut dire que les courbes d'iso-saturation sont des triangles homothétiques au triangle de Maxwell et de taille inférieure à ce dernier, figure 15.8. Or dans le cas du système visuel hypothétique, le chroma n'est pas défini par une fonction Min , il est défini comme une distance euclidienne, donc les courbes d'iso-chroma sont des cercles, figure 15.8.

On peut noter également une différence dans les matrices de passage de RVB à YC1C2, la matrice M de l'espace visuel hypothétique introduit une rotation de 30° des axes C_1 et C_2 défini dans l'espace TLS, ce qui

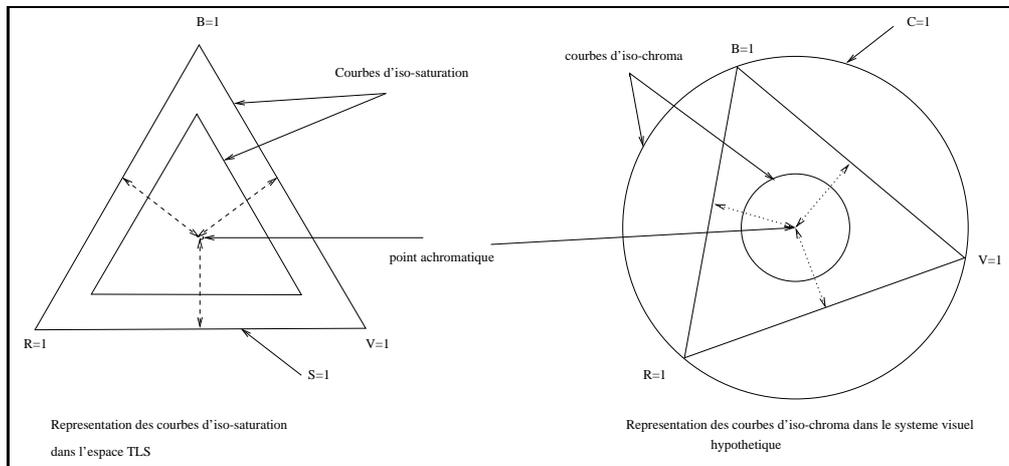


FIG. 15.8 – Les courbes d'iso-saturation dans l'espace TLS et d'iso-chroma dans l'espace visuel hypothétique.

implique que l'axe de référence C1 ne passe plus par le sommet rouge du triangle de Maxwell, mais correspond à l'axe Rouge-Vert. Si on reprend la schématisation de l'espace RVB (voir figure 15.4), on peut représenter les axes C1 et C2 des deux espaces, en considérant l'axe achromatique perpendiculaire au plan de la figure, voir schéma 15.9.

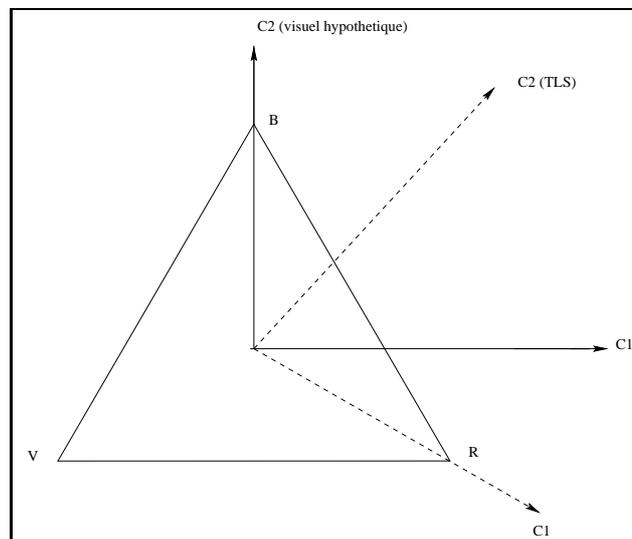


FIG. 15.9 – Les plans C1C2 de l'espace TLS et C1C2 du système visuel hypothétique

On voit qu'il existe des différences entre l'espace TLS et l'espace visuel hypothétique, cependant on constate que la formule de la saturation dans TLS ne permet pas la transformation inverse, pour cela il est préférable de choisir la formule du chroma définie dans l'espace visuel hypothétique. En revanche, dans ce système les axes C1 et C2 ne portent aucun sommet du triangle de Maxwell, ce qui privilégie le choix de la représentation par TLS. Mais plutôt que d'avoir à choisir entre ces deux espaces, on peut adopter la voie décrite par Carron [Carron, 1995] : faire un compromis entre les deux espaces, d'où l'apparition de l'espace TLC.

15.2.9 L'espace TLC

Cet espace est obtenu en effectuant un passage de l'espace RVB à l'espace YC1C2 par l'intermédiaire de la transformation linéaire exprimée à l'aide de la matrice M définie de la façon suivante :

$$M = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & -1/2 & -1/2 \\ 0 & -\sqrt{3}/2 & \sqrt{3}/2 \end{bmatrix}$$

La matrice M ainsi définie est très proche de celle définie pour l'espace TLS, on constate uniquement un changement d'orientation pour l'axe C2.

La matrice de passage de l'espace YC1C2 à l'espace RVB est définie par :

$$M^{-1} = \begin{bmatrix} 1 & 2/3 & 0 \\ 1 & -1/3 & \sqrt{3}/3 \\ 1 & -1/3 & -\sqrt{3}/3 \end{bmatrix}$$

Grâce au composantes Y, C1 et C2 nous pouvons définir les grandeurs suivantes : la luminance, la teinte et le chroma .

On obtient ainsi :

$$\begin{aligned} L &= Y \\ C &= \sqrt{(C1^2 + C2^2)} \\ \text{si } C1 > 0 \text{ alors } T &= \arccos\left(\frac{C1}{C}\right) \\ \text{sinon } T &= 2\pi - \arccos\left(\frac{C1}{C}\right) \end{aligned}$$

Ces formules sont inversibles ce qui permet un retour après traitement à l'espace YC1C2 puis à l'espace RVB.

De surcroît, on constate que le passage aux valeurs de teinte, luminance et chroma se fait en peu d'étapes et donc nécessite moins d'approximations que les expressions obtenues à partir de l'espace Lab. C'est pourquoi on préfère parfois utiliser l'espace TLC, afin d'éviter la perte de précision au cours des changements d'espaces.

15.2.10 Les espaces de la télévision

Afin de garantir une reproductivité universelle des couleurs, les systèmes de télévision se sont dotés de normes, aussi bien pour l'affichage que pour la transmission.

Système NTSC

Il propose un espace RVB relié à l'espace XYZ par les matrices :

$$RVB_XYZ_{NTSC} = \begin{bmatrix} 0.607 & 0.174 & 0.020 \\ 0.299 & 0.287 & 0.114 \\ 0 & 0.066 & 0.117 \end{bmatrix}$$

$$XYZ_RVB_{NTSC} = RVB_XYZ_{NTSC}^{-1} = \begin{bmatrix} 1.910 & -0.533 & -0.288 \\ -0.987 & 2.00 & -0.028 \\ 0.058 & -0.118 & 0.896 \end{bmatrix}$$

Systèmes PAL et SECAM

Ils proposent le même espace RVB relié à l'espace XYZ par les matrices :

$$RVB_XYZ_{PAL/SECAM} = \begin{bmatrix} 0.429 & 0.343 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.019 & 0.132 & 0.939 \end{bmatrix}$$

$$XYZ_RVB_{PAL/SECAM} = RVB_XYZ_{PAL/SECAM}^{-1} = \begin{bmatrix} 3.073 & -1.403 & -0.475 \\ -0.970 & 1.877 & 0.042 \\ 0.074 & -0.235 & 1.068 \end{bmatrix}$$

NTSC : les signaux en composantes

La transmission se fait par Y et I et Q : liées à RVB de NTSC par la matrice :

$$\begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.512 \end{bmatrix}$$

PAL : les signaux en composantes

On transmet Y et $U = 0.493(B - Y)$ ainsi que $V = 0.877(R - Y)$.

SECAM : les signaux en composantes

On transmet Y et $R - Y$ $B - Y$.

15.3 Modélisation de la perception colorée des images

Devant la complexité des expérimentations sur la couleur, un grand intérêt est apparu dans la modélisation de la perception de façon à évaluer automatiquement les performances d'un opérateur humain ou d'apprécier la qualité d'un document sans passer par une longue étape expérimentale.

Nous présentons ici un modèle de la vision - assez ancien - mais très représentatif de ceux que l'on a pu élaborer sur ce sujet. Ce modèle est cependant moins complet d'un point de vue colorimétrique que des modèles plus récents comme celui de Nayatani [Nayatani et al., 1990] ou celui de Guth [Guth, 1991], qui néanmoins ne permettent pas aisément de faire varier la distance d'observation.

Le modèle de Frei-Faugeras fait passer d'une image représentée par ses coordonnées RVB à une image dans un espace perceptuel $A C_1 C_2$, où A code la luminance et C_1 et C_2 les antagonismes rouge/vert et jaune/bleu. Un tel modèle comporte un étage de récepteurs non-linéaires, une séparation de l'information colorée en un 1 canal achromatique et 2 canaux chromatiques, un filtre spatial sur chaque canal (il est représenté sur la figure 15.10).

Le signal RVB , transformé en signal CIE 1931 XYZ , est tout d'abord converti en valeurs LMS . Celles-ci expriment assez bien l'absorption par ce qui semble être les trois familles de récepteurs des cônes des primates, L , M et S dénotant les longueurs d'ondes longues, moyennes et courtes. Cette absorption est suivie d'une étape de compression (traduite par un filtrage non-linéaire logarithmique) conduisant aux valeurs \tilde{L} , \tilde{M} , \tilde{S} . Un second étage donne naissance au canal achromatique A et aux valeurs d'antagonismes chromatiques C_1 et C_2 ($C_1 = \tilde{L} - \tilde{M}$, et $C_2 = \tilde{L} - \tilde{S}$). Enfin 3 filtres différents $H(f)$, $H_1(f)$ et $H_2(f)$ viennent rendre compte des interactions spatiales

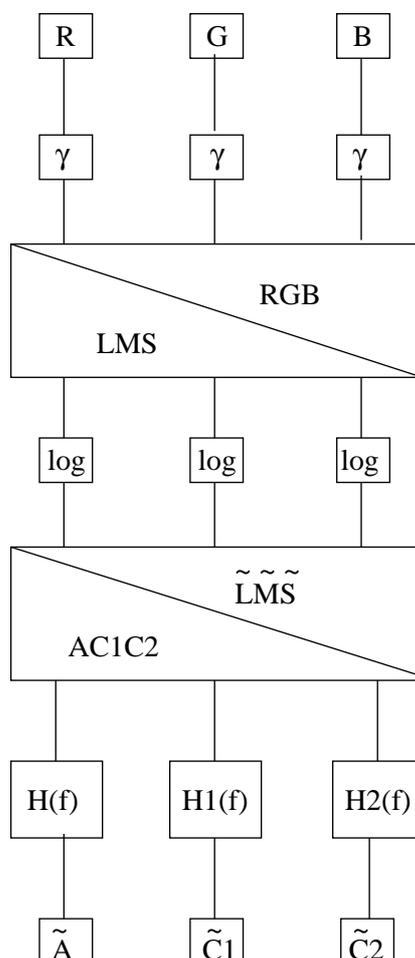


FIG. 15.10 – Le modèle de vision proposé par Frei et développé par Faugeras [Faugeras, 1976].

de chaque canal (respectivement A , C_1 et C_2) indépendamment. Les valeurs de ces divers filtres, ainsi que des matrices de composition et des non-linéarités du modèle sont proposées dans [Faugeras, 1976].

Les filtres spatiaux H_1 et H_2 agissant sur les canaux antagonistes, sont assez étroits, exprimant une nette sélection fréquentielle dont le rôle sera déterminant dans notre expérience (maximum de réponse respectivement à 4 cycles par degré et 2,6 cycles par degré).

Ces filtres ont été réalisés par différences de gaussiennes, conformément à l'approche de Marr [Marr et Hildreth, 1980] (voir chapitre sur la segmentation, paragraphe 11.5). Ils sont présentés sur la figure 15.11.

Le modèle développé par von Kreis reprend le modèle de Frei-Faugeras, mais tient compte du phénomène d'adaptation chromatique qui permet de conserver une perception quasi-permanente de la couleur même lorsque l'éclairage incident varie fortement en teinte. Pour cela von Kreis procède tout d'abord à une normalisation de la couleur sur l'ensemble de la scène observée et introduit des gains chromatiques pour rattraper un possible déséquilibre de l'éclairage.

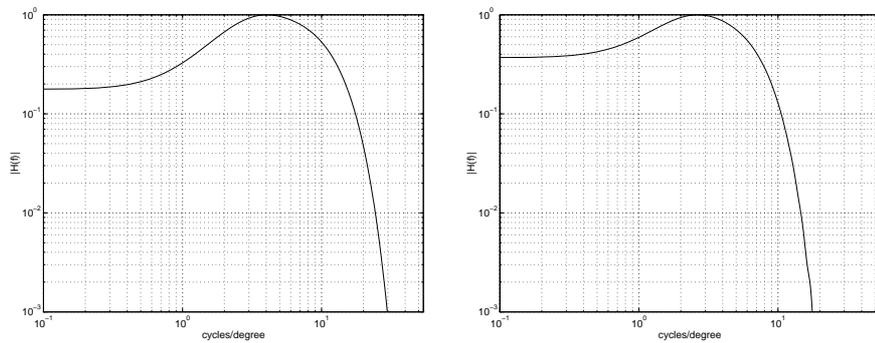


FIG. 15.11 – A gauche la fonction de transfert $H_1(f)$ du canal C_1 (à une dimension), calculée à partir d'une différence de deux gaussiennes. A droite : la fonction de transfert $H_2(f)$ du canal C_2 .

15.4 Les traitements des images en couleur

15.4.1 Un ordre pour l'espace R^3 ?

Un espace est **ordonné** s'il est muni d'une relation entre ses éléments (notée \leq) qui vérifie les 3 propriétés suivantes :

1. réflexivité : $x \leq x$,
2. transitivité : $x \leq y$ et $y \leq z$ implique $x \leq z$,
3. anti-symétrie : si $x \leq y$ et $y \leq x$ alors $x = y$.

Si seules les relations 1 et 2 sont vérifiées, alors la relation \leq est un **pré-ordre**.

Les entiers naturels sont ordonnés dans N . Les couples de N^2 ou les triplets de N^3 ne bénéficient pas d'ordre naturel.

On peut induire un ordre sur un espace non ordonné T' à partir d'un espace ordonné T si l'on fait correspondre à tout élément $x' \in T'$ un élément $x \in T$ par une relation : $x' = h(x)$ injective. Si la relation n'est pas injective alors elle définit un pré-ordre sur T' . On peut alors créer un ordre sur T' en remplaçant les x' par leur classe d'équivalence définie par : $X' = x' : h(x') = a$

On dit qu'il existe un **ordre total** sur un espace T si quelque soit $x, y \in T \times T$ soit $x \leq y$ soit $y \leq x$ (soit $x \leq y$ et $y \leq x$).

Si un ordre n'est pas total, il est dit **partiel**.

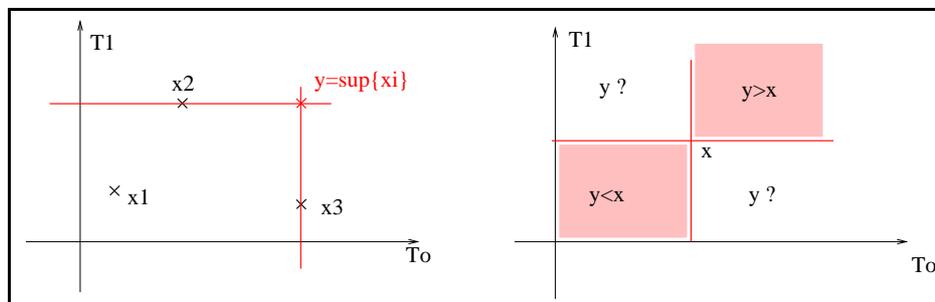


FIG. 15.12 – A gauche : y , le sup des x_i n'appartient pas à l'ensemble qui n'a donc pas de plus grand élément. A droite : l'ordre canonique définit les éléments supérieurs et inférieurs à x , mais ne classe pas les autres.

Si un espace $T' = \{X = \{x_0, x_1\}\}$ est espace produit de 2 espaces ordonnés $T_0 = x_0$ et $T_1 = x_1$, on obtient un ordre partiel dit **ordre canonique** par la relation conjonctive des deux relations d'ordre (voir figure 15.12). On ne peut donc définir au moyen de l'ordre canonique le dilaté (ou l'érodé) d'une image en couleur par cet ordre.

Si l'on veut bénéficier d'un certain nombre d'outils développés en traitement des images, il faut pouvoir induire un ordre total dans l'espace des couleurs. Avec un tel ordre, il est alors possible d'appliquer les algorithmes de la morphologie mathématique ou des filtres non-linéaires de rang , par exemple, (comme les filtres médians en particulier). L'impossibilité de le faire au moyen de l'ordre canonique est illustré sur la figure 15.12 pour la dilatation. On ne peut définir le dilaté (ou l'érodé) d'une image en couleur par cet ordre, puisque ni le min ni le max d'un ensemble de valeurs n'appartient forcément à cet ensemble.

Il a été proposé donc d'induire un ordre dans cet espace par un balayage judicieux des 3 dimensions (le choix d'une application h injective ci-dessus). Plusieurs ordres sont possibles, qui sont illustrés à 2 dimensions dans la figure 15.13. Aucun de ces balayages ne résoud complètement le problème car l'ordre induit dans l'espace transformé respecte mal la notion de voisinage que l'on avait dans l'espace \mathbb{R}^3 initial. Ceci est confirmé par le théorème de Netto :

Toute application injective de $[0, 1]^2$ dans $[0, 1]$ est discontinue.

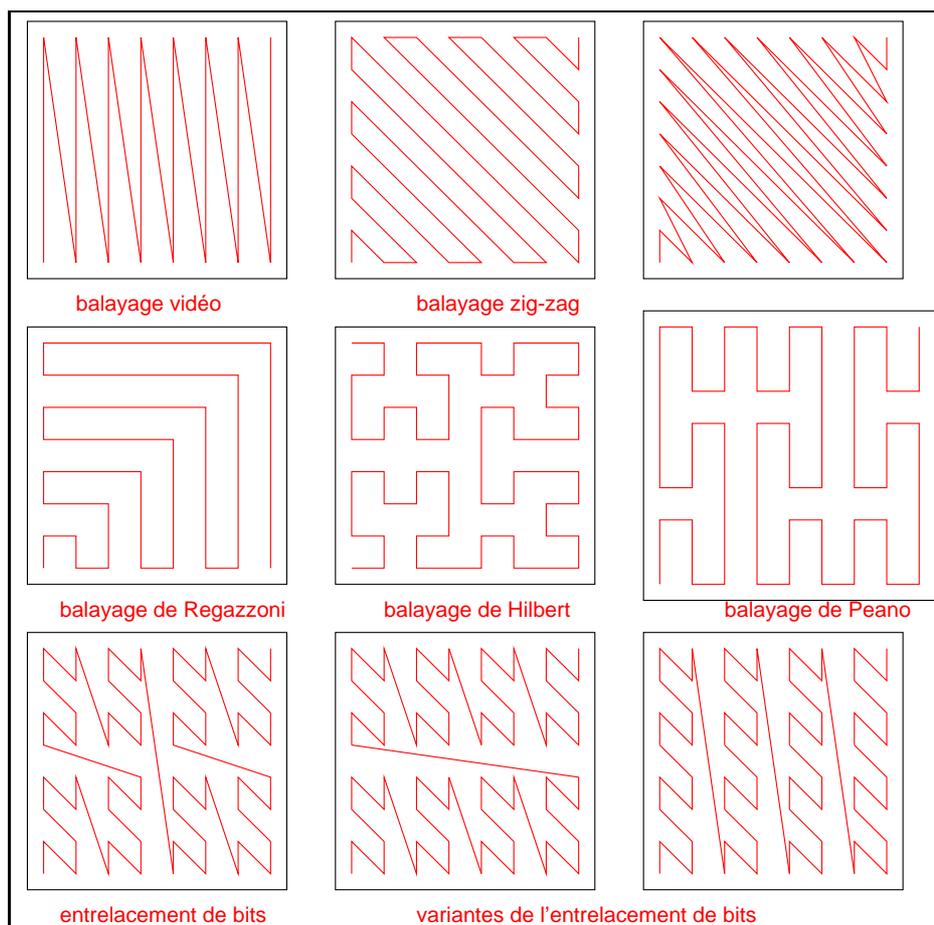


FIG. 15.13 – Les divers modes de balayage de l'espace R^2

Considérons la 8 connexité de l'espace \mathbb{R}^2 . Chaque point x a alors 8 voisins appartenant au voisinage V_x . Définissons une 8-connexité le long du balayage de l'image, qui associe les 4 points précédant x et les 4 points succédant à x le long de la courbe. Soit W_x ce voisinage. Un balayage préservera la topologie de \mathbb{R}^2 si les points de V_x se retrouvent dans W_x . Statistiquement, sur toute la surface $[0, 1]^2$, on trouve les résultats suivants :

| balayage | Hilbert | Peano | Entrelacement | ZigZag | Regazzoni | TV |
|-----------------------------|---------|-------|---------------|--------|-----------|------|
| pourcentage de points | | | | | | |
| du voisinage 1D appartenant | 58 % | 57 % | 45-50 % | 27 % | 26 % | 25 % |
| au voisinage 2D | | | | | | |

A ce titre, les balayages de Peano et de Hilbert (deux balayages à structure fractale puisqu'auto-similaires) et l'entrelacement de bits sont parmi les plus satisfaisants.

15.4.2 Qu'est ce qu'une couleur représentative d'un nuage ?

On suppose que l'on dispose de N points de couleur, on veut en extraire une valeur significative. Une fois choisi l'espace des couleurs, le choix d'une couleur moyenne est trivial puisqu'il revient à filtrer canal par canal et à constituer la valeur moyenne à partir des 3 moyennes des composantes.

On calcule parfois des moyennes "frustrées", c'est-à-dire calculée à partir de l'ensemble privé de ses éléments extrêmes. Pour cela on travaille en deux étapes :

1. à partir de l'ensemble des points on calcule une moyenne,
2. on écarte de l'ensemble tous les points à une distance trop grande de cette valeur moyenne,
3. une autre possibilité consiste à remplacer les points à distance trop grande par leur projection sur la sphère centrée en la moyenne et de rayon égal au seuil de distance,
4. pour les points retenus, on recalcule une moyenne. Cette valeur moyenne est la moyenne frustrée de l'ensemble.

Le choix d'une valeur médiane n'est pas trivial à cause des problèmes d'ordre exposés ci-dessus. Parmi les techniques choisies, on utilise souvent la technique du minimum de distance :

1. pour chaque point de l'ensemble on calcule la somme des distances à tous les autres,
2. on trie ces distances par ordre croissant,
3. on choisit pour point médian celui qui minimise la somme des distances.

On peut également choisir dans l'algorithme ci-dessus, non le point qui minimise la somme des distances, mais celui qui minimise la médiane des distances d'un point à tous les autres.

15.4.3 Comment calculer une palette ?

Les diverses techniques permettant de quantifier la couleur partent généralement d'un signal sur 2^{24} niveaux. Elles réduisent ce nombre par :

1. la quantification uniforme ; ex : palette 5,5,5 pour coder sur 15 bits (on réduit le nombre de couleurs à 36 000 environ, ce qui permet de construire des histogrammes) ou palette 4,2,2 pour coder sur 1 octet une image couleur.
2. la méthode LBG (Linde, Buzo, Gray) ou méthode des k-moyennes ; cette méthode de classification permet de minimiser la distorsion entre la représentation originale et la représentation quantifiée. Elle étend à N dimensions la solution classique de Lloyd et Max en choisissant [Linde et al., 1980] :

- pour classe celle qui est la plus proche de l'échantillon considéré au sens de la norme adoptée pour mesurer la distorsion,
 - pour représentant d'une classe le centre de gravité de cette classe.
3. la méthode des k-moyennes floues,
 4. l'algorithme de population,
 5. le *median-cut*,
 6. l'*octtree*
 7. les réseaux de Kohonen

15.4.4 Comment filtrer 2 couleurs ?

Les expressions classiques du filtrage linéaire s'appliquent bien sûr aux images en couleur et l'extension à des signaux vectoriels ($f \in \mathbb{R}^3$ au lieu de $f \in \mathbb{R}$ comme dans les images à niveaux de gris), ne change pas fondamentalement le principe du filtrage. Il est cependant très important de bien choisir l'espace de représentation des signaux afin que la linéarité soit bien prise en compte dans l'espace souhaité. En effet, les équations fortement non linéaires de changement d'espaces (par exemple de XYZ à Lab) interdisent d'optimiser (par exemple aux moindres carrés) les filtres simultanément dans les divers espaces (figure 15.14).

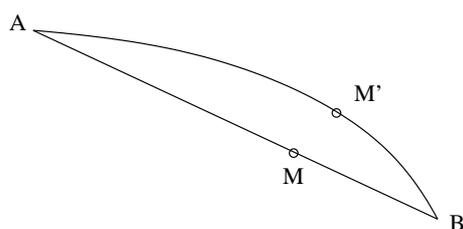


FIG. 15.14 – Lors de l'interpolation de deux couleurs, A et B, le passage par une transformation non-linéaire (par exemple l'exposant $\frac{1}{3}$ du passage de XYZ à Lab) conduit à des couleurs (ici M et M') très différentes.

15.4.5 La détection des contours des images en couleur

La détection des contours des images en couleur peut se faire tout d'abord en appliquant séparément une détection de contours scalaires (voir chapitre 11) sur chacune des composantes [diZenzo., 1986]. On dispose de 3 images de contours qu'il convient alors de composer pour obtenir une seule image de contours (on dit souvent « fusionner »). On choisit fréquemment une combinaison des 3 contours qui accepte tous les contours : par exemple au moyen d'un *max* des 3 images de contours.

$$\text{contour résultant} = \max \{ \text{contours des canaux} \}$$

Si l'on choisit une combinaison en modules :

$$\text{contour résultant} = \sqrt{\text{somme des contours}^2}$$

on se ramène alors souvent à une détection de contours sur le module de l'image (c'est-à-dire à peu près sur l'image de luminance) car beaucoup de détecteurs de contours sont linéaires (détecteur de Deriche, de Canny, de Shen). Attention, la recherche du maximum du gradient dans la direction du gradient n'est - elle - pas linéaire et il n'est pas équivalent de l'appliquer avant ou après la combinaison des canaux.

On peut tirer profit d'informations particulières sur le problème à traiter pour privilégier un type de représentation à d'autres. Par exemple on sait que les ombrages (aussi bien propres que portés) sont peu colorés. On peut alors ne segmenter que les composantes chromatiques des objets si l'on veut séparer des objets sans les fragmenter en fonction des ombres.

15.4.6 Représentation par quaternions

Cette technique particulière et originale peut s'appliquer à n'importe quel espace chromatique. Elle a été proposée dans [Sangwine, 1996]. Elle permet de manipuler la couleur dans une représentation analytique compacte autorisant la plupart des traitements accomplis sur les scalaires par une extension formelle.

Elle utilise la représentation par quaternions, de façon à représenter par un nombre unique un vecteur de \mathbb{R}^4 (et donc a fortiori de \mathbb{R}^3) de la même façon que la représentation par nombre complexe permet de manipuler des objets de \mathbb{R}^2 . Les quaternions forment un anneau qui peut être construit à partir de tout anneau commutatif. Dans le domaine du traitement de l'image on utilise les quaternions construits sur \mathbb{R} dans les applications comme la calibration des caméras où le changement d'espace de représentation ou le matching de points. Mais les quaternions sont également utilisés à partir de leur construction sur \mathbb{Z} et sur \mathbb{C} en théorie des nombres ou en mécanique quantique.

Ils se construisent à partir d'une base constituée de $(1, i, j, k)$, vérifiant des propriétés de multiplicativité semblables à celles des complexes :

$$\begin{aligned} 1.i &= i & 1.j &= j & 1.k &= k \\ ij &= -ji = k & jk &= -kj = i & ki &= -ik = j \\ i^2 &= j^2 = k^2 & &= -1 \end{aligned}$$

Un quaternion p se compose de 2 parties, l'une scalaire a , l'autre vectorielle q :

$$p = a + bi + cj + dk = a + q$$

Les quaternions vectoriels (c'est-à-dire tels que $a = 0$), également appelés quaternions purs, forment un espace vectoriel isomorphe à \mathbb{R}^3 .

Les opérations sur les quaternions permettent de définir le conjugué \bar{p} de $p = a + q$, ($\bar{p} = a - q$), la somme, le produit, le produit scalaire de 2 quaternions ($\langle p', p \rangle = 1/2(p'\bar{p} + p\bar{p}')$) et la norme d'un quaternion ($\|p\| = \sqrt{p\bar{p}}$). Le produit n'est pas commutatif, sauf si les 2 quaternions ont leurs parties vectorielles proportionnelles. En effet :

$$pp' = (aa' - qq', aq' + a'q + q \wedge q')$$

15.4.7 Couleur et quaternion

Représenter un point de couleur R, V, B par un quaternion pur, c'est tout d'abord choisir un espace de couleur (par exemple RVB), puis une origine de l'espace des couleurs (ce peut être le noir $(0,0,0)$, mais dans [Evans et al., 2000] on préfère le point de coordonnées $(127.5, 127.5, 127.5)$ qui se situe au centre du cube des couleurs. On représente ensuite un point de couleur par un quaternion pur. Par exemple comme :

$$q = i(R - O) + j(V - O) + k(B - O)$$

Les opérations conduites sur les quaternions (produits, convolutions, etc.) héritent des propriétés spécifiques du produit de quaternions. En particulier, pour des quaternions purs q et q' , on vérifie aisément que la partie scalaire $Scal[q.q']$ du produit de quaternions q et q' vaut :

$$Scal[qq'] = -q.q'$$

où \cdot dénote le produit scalaire de 2 vecteurs, et la partie vectorielle $Vect$:

$$Vect[qq'] = q \times q'$$

où \times représente le produit vectoriel. Ces propriétés permettent d'obtenir des résultats particuliers par des mises en équation très simples.

Par exemple dans [Evans et al., 2000], on montre que l'on peut filtrer une image en couleur par un filtre passe-bas qui n'agit que sur la composante chromatique parallèle à une couleur donnée quelconque C_1 (et pas sur les composantes orthogonales) en construisant le quaternion unitaire $\mu_{C_1} : \mu_{C_1} = \frac{C_1}{\|C_1\|}$, puis en convolant (à droite ou à gauche) l'image q par un noyau N (par exemple de 3×3 pixels) d'amplitude μ_{C_1} :

$$N = \frac{1}{9} \begin{bmatrix} \mu_{C_1} & \mu_{C_1} & \mu_{C_1} \\ \mu_{C_1} & \mu_{C_1} & \mu_{C_1} \\ \mu_{C_1} & \mu_{C_1} & \mu_{C_1} \end{bmatrix}$$

En effet, ces convolutions : $N * q = u$ et $q * N = v$ fournissent 2 quaternions u et v de même partie réelle et de parties vectorielles opposées. Les parties scalaires obtenues sont égales à la moyenne des composantes chromatiques de l'image parallèles à μ_{C_1} .

$$Scal[N * f](x, y) = \sum_{s,t} -\frac{1}{9} \mu_{C_1} f(x-s, y-t)$$

Les parties vectorielles sont \pm égales au produit vectoriel :

$$Vect[N * f](x, y) = -Vect[f * N](x, y) = \sum_{s,t} \frac{1}{9} \mu_{C_1} \times f(x-s, y-t)$$

Leur somme $u + v$ est un scalaire qui peut être orienté selon C_1 en le multipliant par μ_{C_1} . On aboutit à un filtrage hypercomplexe M qui vaut :

$$M = q - \frac{\mu_{C_1}}{2} [N' * q + q * N']$$

avec

$$N' = \frac{1}{9} \begin{bmatrix} \mu_{C_1} & \mu_{C_1} & \mu_{C_1} \\ \mu_{C_1} & -8\mu_{C_1} & \mu_{C_1} \\ \mu_{C_1} & \mu_{C_1} & \mu_{C_1} \end{bmatrix}$$

Cette représentation par quaternions permet de définir d'une façon semblable des détecteurs de contours [Sangwine, 1998], des filtres passe-bas [Evans et al., 2000], des TF de l'espace des couleurs ([Ell et Sangwine, 2000]).

Annexe A

Décomposition en Valeurs Singulières : SVD

Chapitre rédigé par Denis MATIGNON

A.1 Définition

On étudie la décomposition d'une matrice $A \in \mathbb{C}^{p \times m}$, éventuellement dans $\mathbb{R}^{p \times m}$; les résultats donnés pour les matrices réelles se particularisent en changeant les transposées-conjuguées (V^H) en transposées (V^T), et les matrices unitaires sont remplacées par des matrices orthogonales.

Théorème 1. Soit $A \in \mathbb{C}^{p \times m}$, il existe des matrices unitaires $U \in \mathbb{C}^{p \times p}$ et $V \in \mathbb{C}^{m \times m}$ telles que :

$$U^H A V = \Sigma \in \mathbb{R}^{p \times m} \quad \text{avec } \Sigma = \text{diag}(\sigma_1, \dots, \sigma_n) \quad \text{où } n = \min(p, m) \quad \text{et } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

La matrice rectangulaire Σ est constituée de 0 sauf sur la diagonale qui comporte n réels positifs, les σ_i ou *valeurs singulières* de A ; les u_i et les v_i (colonnes de U et V respectivement) constituent les *vecteurs singuliers* de A à gauche et à droite respectivement :

$$A v_i = \sigma_i u_i \quad \text{et} \quad A^H u_i = \sigma_i v_i$$

L'existence de la SVD provient de la diagonalisabilité en base orthonomée des matrices hermitiennes $A^H A$ et $A A^H$, et de leur positivité. On a l'unicité des valeurs singulières (comme on a celle du spectre de $A^H A$); en revanche, l'unicité de la SVD n'est assurée que si la matrice est carrée et possède n valeurs singulières deux à deux distinctes. Mais ceci n'est jamais vraiment gênant pour les propriétés structurelles que l'on peut étudier grâce à la SVD.

A.2 Lien avec le rang et interprétation de la décomposition

On appelle $\sigma_{\max}(A) = \sigma_1$ la plus grande valeur singulière de A et $\sigma_{\min}(A) = \sigma_n$ la plus petite valeur singulière de A (éventuellement nulle). Si le classement se précise comme suit :

$$\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0$$

alors r est le *rang* de A , et on peut caractériser le noyau et l'image de A par :

$$\ker A = \{v_{r+1}, \dots, v_m\} \quad \text{et} \quad \text{im } A = \{u_1, \dots, u_r\}$$

On peut alors donner une expression plus compacte de la décomposition de A , en notant $U_r = U(:, 1 : r)$, $V_r = V(:, 1 : r)$ les sous-matrices rectangulaires extraites, et $\Sigma_r = \Sigma(1 : r, 1 : r) = \text{diag}(\sigma_1, \dots, \sigma_r)$ la matrice carrée diagonale extraite ou partie *utile* de Σ :

$$A = U\Sigma V^H = U_r \Sigma_r V_r^H = \sum_{i=1}^r \sigma_i u_i v_i^H$$

Sur cette dernière décomposition, on voit très bien comment agit A sur chaque vecteur v_i de la base orthornormée de l'espace de départ \mathbb{C}^m : elle le transforme en σ_i fois le vecteur u_i de la base orthornormée de l'espace d'arrivée \mathbb{C}^p .

Pour A^H , on montre de même que $\ker A^H = \{u_{r+1}, \dots, u_p\}$ et $\text{im } A^H = \{v_1, \dots, v_r\}$; et l'on a :

$$A^H = V\Sigma^T U^H = V_r \Sigma_r U_r^H = \sum_{i=1}^r \sigma_i v_i u_i^H$$

A.3 Normes

La norme 2 et la norme Frobenius de A se calculent aisément à partir de la SVD :

$$\begin{aligned} \|A\|_2 &\triangleq \sup_{\|x\|=1} \|Ax\| = \sqrt{\max(\text{spec } A^H A)} = \sigma_1 \\ \|A\|_F &\triangleq \sqrt{\text{tr}(A^H A)} = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\sigma_1^2 + \dots + \sigma_r^2} \end{aligned}$$

La propriété suivante joue un très grand rôle dans les problèmes d'approximation :

Théorème 2. Soit $\{U, \Sigma, V\}$ une SVD de A , considérons pour $1 \leq k \leq r = \text{rang}(A)$ la suite de matrices $A_k = \sum_{i=1}^k \sigma_i u_i v_i^H$ de rang exactement k , alors :

$$\min_{B, \text{rang}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

En fait, les valeurs singulières de A mesurent la distance (en norme 2) de la matrice A par rapport aux matrices de rang inférieur.

Plus précisément, si l'on se donne $\varepsilon > 0$ comme tolérance numérique, et que l'on mesure le rang numérique de A correspondant par $r_\varepsilon = \text{rang}(A, \varepsilon) = \min_{B, \|A-B\|_2 \leq \varepsilon} \text{rang}(B)$, le théorème 2 permet d'écrire :

$$\sigma_1 \geq \dots \geq \sigma_{r_\varepsilon} > \varepsilon \geq \sigma_{r_\varepsilon+1} \geq \dots \geq \sigma_r$$

Annexe B

Applications de la SVD à la résolution des systèmes linéaires

Chapitre rédigé par Denis MATIGNON

B.1 Conditionnement d'un système linéaire non singulier

Considérons le système linéaire $Ax = b$ supposé régulier, c'est-à-dire A inversible. On cherche à évaluer les variations de la solution x relativement aux variations de A et de b (imprécisions sur le système et les données).

Le calcul différentiel permet une analyse directe dans le cas régulier : comme $x = A^{-1}b$, alors $dx = d(A^{-1}b) = d(A^{-1})b + A^{-1}db$; enfin $d(A^{-1}) = -A^{-1}dA A^{-1}$, d'où :

$$dx = -A^{-1}dA A^{-1}b + A^{-1}db = A^{-1}[-dA x + db], \text{ et } \|dx\| \leq \|A^{-1}\| [\|dA\| \|x\| + \|db\|]$$

En divisant le tout par $\|x\|$, et en utilisant que $\|b\| \leq \|A\| \|x\|$, on obtient la majoration suivante :

$$\frac{\|dx\|}{\|x\|} \leq \|A\| \|A^{-1}\| \left[\frac{\|dA\|}{\|A\|} + \frac{\|db\|}{\|b\|} \right]$$

Le nombre $C = \|A\| \|A^{-1}\|$ apparaît clairement comme une mesure de la sensibilité de l'opération d'inversion. Caractérisons la par les valeurs singulières de A .

En introduisant une SVD de $A : A = U\Sigma V^H = \sum_{i=1}^r \sigma_i u_i v_i^H$ (dans le cas régulier, $r = n = m = p$), on trouve immédiatement la décomposition correspondante de $A^{-1} : A^{-1} = V\Sigma^{-1}U^H = \sum_{i=1}^r \sigma_i^{-1} v_i u_i^H$. Il est alors évident que $C = \sigma_1/\sigma_n$; de plus, un raisonnement précis utilisant les vecteurs singuliers montre que la borne proposée est optimale pour certaines directions de A, dA, b, db .

Définition 1. $C = \|A\| \|A^{-1}\| = \sigma_{max}(A)/\sigma_{min}(A)$ est le nombre de condition ou conditionnement de la matrice A .

Remarquons que C est au minimum égal à 1 pour les matrices $A = \lambda U$ où U est unitaire. C peut être très grand, et ce même pour des matrices dont les valeurs propres sont raisonnables.

B.2 Pseudo-inverse

Plusieurs causes de singularité ($\sigma_{min}(A) = 0$) existent dans les systèmes linéaires $y = Ax$:

- A carrée n'est plus inversible ; si $y \in \text{im } A$ alors il y a un sous-espace affine de solutions et il peut être intéressant de chercher celle de norme minimale ; si $y \notin \text{im } A$, il n'y a donc pas de solution exacte, en revanche il y a forcément une solution qui minimise la distance $\|y - Ax\|$.
- A est rectangulaire verticale avec $m < p$, autrement dit il y a plus d'équations que d'inconnues : on parle alors de système surdéterminé ; là encore, il peut n'y avoir aucune solution exacte, mais une solution qui minimise l'erreur $\|y - Ax\|$.
- A est rectangulaire horizontale avec $p < m$, autrement dit il y a plus d'inconnues que d'équations : on parle alors de système sous-déterminé ; bien souvent il y a un sous-espace affine de solutions, et il peut être intéressant de calculer la solution de norme minimale.

Nous donnons à présent la définition de la pseudo-inverse d'une matrice en ayant recours à la SVD puis nous la spécifions dans deux cas particuliers parmi ceux évoqués ci-dessus.

Définition 2. Soit $\{U, \Sigma, V\}$ une SVD de A de rang r , on appelle pseudo-inverse de A la matrice A^\dagger définie par :

$$A^\dagger = V_r \Sigma_r^{-1} U_r^H = \sum_{i=1}^r \sigma_i^{-1} v_i u_i^H$$

Elle vérifie :

$$AA^\dagger A = A \quad \text{et} \quad A^\dagger AA^\dagger = A^\dagger$$

Contrairement à ce que laisserait supposer la définition, A^\dagger est définie de manière unique et ne dépend pas du choix de SVD que l'on fait pour la construire. On a en particulier :

$$A^\dagger A = V_r V_r^H \text{ est la projection orthogonale sur } \text{im } A^H = (\ker A)^\perp = \{v_1, \dots, v_r\} \text{ dans } \mathbb{C}^m,$$

$$AA^\dagger = U_r U_r^H \text{ est la projection orthogonale sur } \text{im } A = (\ker A^H)^\perp = \{u_1, \dots, u_r\} \text{ dans } \mathbb{C}^p.$$

On donne maintenant une expression directe de A^\dagger dans les deux cas particuliers de rang plein, où $r = n$.

Cas de système surdéterminé avec $r = m < p$ dans ce cas $V_r = V$, et l'on a alors $A^H A = V_r \Sigma_r^2 V_r^H$ carrée $m \times m$ inversible, d'inverse $(A^H A)^{-1} = V_r \Sigma_r^{-2} V_r^H$, d'où :

$$(A^H A)^{-1} A^H = V_r \Sigma_r^{-2} V_r^H V_r \Sigma_r U^H = V_r \Sigma_r^{-1} U^H = A^\dagger$$

Cas de système sous-déterminé avec $r = p < m$ dans ce cas $U_r = U$, et l'on a alors $AA^H = U_r \Sigma_r^2 U_r^H$ carrée $p \times p$ inversible, d'inverse $(AA^H)^{-1} = U_r \Sigma_r^{-2} U_r^H$, d'où :

$$A^H (AA^H)^{-1} = V_r \Sigma_r U_r^H U_r \Sigma_r^{-2} U_r^H = V_r \Sigma_r^{-1} U_r^H = A^\dagger$$

Nous voyons maintenant l'utilité de la notion de pseudo-inverse pour calculer des solutions robustes d'un système linéaire général, c'est-à-dire potentiellement singulier ou proche de la singularité.

B.3 Solutions robustes d'un système linéaire, moindres carrés

On cherche à résoudre $y = Ax$, où $x \in \mathbb{C}^m, y \in \mathbb{C}^p$; au cas où ce système n'a pas de solution exacte on cherchera la solution qui minimise l'erreur $\|y - Ax\|$ (ou solution en *moindres carrés* puisque la norme est euclidienne) ; au cas où il y a plusieurs solutions, on prendra celle qui est de norme minimale, autrement dit qui minimise $\|x\|$.

Théorème 3. La solution de norme $\|x\|$ minimale dans \mathbb{C}^m qui minimise l'erreur $\|y - Ax\|$ dans \mathbb{C}^p est exactement donnée par :

$$x^* = A^\dagger y$$

Pour prouver ce théorème fondamental, qui donne tout son sens à la notion de pseudo-inverse, nous allons utiliser $\{U, \Sigma, V\}$ une SVD de A . Comme U, V sont des matrices unitaires, elles conservent la norme euclidienne dans $\mathbb{C}^p, \mathbb{C}^m$ respectivement ; en conséquence :

$\delta = y - Ax$ et $U^H \delta$ ont même norme,

x et $V^H x$ ont même norme,

En notant désormais $a = V^H x$ et $b = U^H y$, le problème se met sous la forme beaucoup plus simple : trouver la solution de norme $\|a\|$ minimale dans \mathbb{C}^m qui minimise l'erreur $\|b - \Sigma a\|$ dans \mathbb{C}^p . Or en décomposant Σ par blocs, il vient :

$$\|b - \Sigma a\|^2 = \sum_{i=1}^r |b_i - \sigma_i a_i|^2 + \sum_{i=r+1}^p |b_i|^2$$

On en déduit que le minimum vaut $\sum_{i=r+1}^p |b_i|^2$, et qu'il est atteint pour les a tels que $a_i = \sigma_i^{-1} b_i$ pour $1 \leq i \leq r$ et a_i quelconques pour $r+1 \leq i \leq m$ (ce qui correspond exactement à la paramétrisation de $\ker A$) ; dans ce sous-espace affine de solutions, celle de norme minimale est celle pour laquelle $a_i = 0$ pour $r+1 \leq i \leq m$: en réinterprétant cela matriciellement, il vient $a^* = \Sigma^\dagger b$. En revenant dans les bases initiales $A^\dagger = V \Sigma^\dagger U^H$, et l'on a alors prouvé le théorème.

Nous regardons plus précisément ce qui se passe dans les deux cas particuliers envisagés précédemment, où $r = n$.

Cas de système surdéterminé avec $r = m < p$ dans ce cas $\ker A = \{0\}$ et le minimum de $\|y - Ax\|$ est atteint en un point unique $x^* = A^\dagger y = (A^H A)^{-1} A^H y$; ce minimum est nul si et seulement si $y \in \text{im } A$.

Cas de système sous-déterminé avec $r = p < m$ dans ce cas $\text{im } A = \mathbb{C}^p$ tout entier et $y = Ax$ possède un sous-espace affine de solutions exactes $x^* + \ker A$, où $x^* = A^\dagger y = A^H (A A^H)^{-1} y$ est celle de norme minimale.

Annexe C

Optimisation sous contraintes

Chapitre rédigé par Denis MATIGNON

C.1 Contraintes égalités

Nous donnons le théorème relatif à la minimisation d'une forme quadratique positive sous contraintes égalités linéaires.

Théorème 4. Soit à trouver le minimum de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ avec $f(x) = \frac{1}{2}x^T Qx + L_0^T x$ sous m contraintes égalités $L_i^T x = y_i$. $Q \geq 0$ et les L_0, L_1, \dots, L_m sont des vecteurs.

On définit le lagrangien du problème :

$$\mathcal{L}(x, \lambda_1, \dots, \lambda_m) \triangleq f(x) - \sum_{i=1}^m \lambda_i (L_i^T x - y_i)$$

f admet un minimum en x^* si et seulement s'il existe des multiplicateurs de Lagrange associés λ_i^* pour lesquels le point $(x^*, \lambda_1^*, \dots, \lambda_m^*)$ est un point critique du lagrangien, soit :

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x}(x, \lambda_1, \dots, \lambda_m) = x^T Q + L_0^T - \sum_{i=1}^m \lambda_i L_i^T \\ 0 &= \frac{\partial \mathcal{L}}{\partial \lambda_i}(x, \lambda_1, \dots, \lambda_m) = L_i^T x - y_i \quad \text{pour } 1 \leq i \leq m \end{aligned}$$

Ce théorème est une version très particulière et fort utile dans la pratique de l'optimisation sous contraintes égalités de fonctions à valeurs réelles ; « condition de Lagrange » optimisation.

C.2 Contraintes inégalités

Nous donnons le théorème relatif à la minimisation d'une forme quadratique positive sous contraintes égalités et inégalités linéaires.

Théorème 5. Soit à trouver le minimum de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ avec $f(x) = \frac{1}{2}x^T Qx + L_0^T x$ sous m contraintes égalités $L_i^T x = y_i$ et p contraintes inégalités $M_j^T x \geq z_j$. $Q \geq 0$ et les $L_0, L_1, \dots, L_m, M_1, \dots, M_p$ sont des vecteurs.

On définit le lagrangien du problème :

$$\mathcal{L}(x, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p) \triangleq f(x) - \sum_{i=1}^m \lambda_i (L_i^T x - y_i) - \sum_{j=1}^p \mu_j (M_j^T x - z_j)$$

f admet un minimum en x^* si et seulement s'il existe des multiplicateurs de Lagrange associés λ_i^* et $\mu_j^* \geq 0$ pour lesquels le point $(x^*, \lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_p^*)$ est un point critique partiel du lagrangien, soit :

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x}(x, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p) = x^T Q + L_0^T - \sum_{i=1}^m \lambda_i L_i^T \\ 0 &= \frac{\partial \mathcal{L}}{\partial \lambda_i}(x, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p) = L_i^T x - y_i \quad \text{pour } 1 \leq i \leq m \\ 0 &\leq \frac{\partial \mathcal{L}}{\partial \mu_j}(x, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p) = M_j^T x - z_j \quad \text{pour } 1 \leq j \leq p \\ \text{si } 0 &\neq M_j^T x^* - z_j \quad \text{alors } \mu_j = 0 \quad \text{pour } 1 \leq j \leq p \end{aligned}$$

Voir des résultats plus généraux « condition de Kuhn et Tucker » en optimisation.

Bibliographie

- [Adams et Bishof, 1994] Adams, R. et Bishof, L. (1994). Seeded region growing. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 16 :641–647.
- [Alata et Cariou, 2001] Alata, O. et Cariou, C. (2001). Modélisation stochastique linéaire 2-D. In Garelo, R., éditeur, *Analyse de signaux bidimensionnels*, chapitre 2, pages 67–116. Hermès - Collection IC2, Paris.
- [Alvarez, 1996] Alvarez, L. (1996). Images and pde's. In Berger, M., Deriche, R., Herlin, I., Jaffré, J., et Morel, J.-M., éditeurs, *Lecture Notes in Control and Information Sciences*, volume 219, pages 3–14.
- [Alvarez et al., 1993] Alvarez, L., Guichard, F., Lions, P. L., et Morel, J. M. (1993). Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis.*, 16(9) :200–257.
- [Amini et al., 1990] Amini, A., Weymouth, T., et Jain, R. (1990). Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Pattern Analysis and Machine Intelligence-12(9) :855–867.
- [Andres, 1994] Andres, E. (1994). Discrete Circles, Rings and Spheres. *Computers & Graphics*, 18(5) :695–706.
- [Andrews et Hunt, 1977] Andrews, H. et Hunt, B. (1977). *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs (USA).
- [Andrey et Tarroux, 1998] Andrey, P. et Tarroux, P. (1998). Unsupervised segmentation of MRF modeled textured images using selectionist relaxation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 20(3) :252–262.
- [Antonini et al., 1994] Antonini, M., Gaidon, T., Barlaud, M., et Mathieu, P. (1994). *Wavelet transform and image communication*. Elsevier, Amsterdam.
- [Arce et al., 2000] Arce, G., Paredes, J. L., et Mullan, J. (2000). Non-linear filtering for image analysis and enhancement. In Bovik, A. C., éditeur, *Handbook of Image and Video Processing*, pages 81–100. Academic Press, San Diego (USA).
- [Attasi, 1975] Attasi, S. (1975). Stochastic state representation of images. *Springer Verlag Lecture Notes in Eco. and Math. Systems.*, 107 :218–230.
- [Aubert et Blanc-Féraud, 1999] Aubert, G. et Blanc-Féraud, L. (1999). Some remarks on the equivalence between 2D and 3D classical snakes and geodesic active contours. *Intern. J. Computer Vision*, 34(1) :19–28.
- [Aubert et Kornprobst, 2001] Aubert, G. et Kornprobst, P. (2001). *Mathematical Problems in Image Processing, Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer.
- [Azencott et al., 1997] Azencott, R., Wang, J., et Younès, L. (1997). Texture classification using windowed Fourier filters. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 19(2) :148–153.
- [Babaud et al., 1986] Babaud, J., Witkin, A., Baudin, M., et Duda, R. (1986). Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8 :26–33.
- [Ballard, 1978] Ballard, D. (1978). Generalizing the Hough transform to detect a arbitrary shapes. *Pattern Recognition*, 13(2) :111–122.

- [Barlaud et Labit, 2002] Barlaud, M. et Labit, C. (2002). *Compression et codage des images et des vidéos*. Collection IC2. Hermès, Paris.
- [Beaulieu et Goldberg, 1989] Beaulieu, J. et Goldberg, M. (1989). Hierarchy in picture segmentation : A stepwise optimization. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11(2) :150–163.
- [Beckmann et Spizzichino, 1963] Beckmann, P. et Spizzichino, A. (1963). *The scattering of electromagnetic waves from rough surfaces*. Artech House, N. Y.
- [Berge, 1958] Berge, C. (1958). *Théorie des graphes et ses applications*. Dunod, Paris.
- [Bergeaud, 1996] Bergeaud, F. (1996). Représentations adaptatives d'images numériques. Thèse de Doctorat de l'Ecole Centrale, Paris.
- [Bergen et Landy, 1991] Bergen, J. et Landy, M. (1991). Computational modeling of visual texture segregation. In Landy, M. S. et Movshon, J. A., éditeurs, *Computational models of visual processing*, pages 253–271. The MIT Press.
- [Berthier, 1993] Berthier, S. (1993). *Optique des milieux continus*. Polytechnica, Paris.
- [Bertrand et al., 1982] Bertrand, G., Queval, R., et Maître, H. (1982). Shape Interpolation by Fourier Descriptors with Application to Animation Graphics. *Signal Processing*, 4 :53–58.
- [Besag, 1974] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, B-36 :192–236.
- [Besag, 1986] Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48 :259–302.
- [Bezdek, 1981] Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithm*. Plenum Press, New York.
- [Black et Rangarajan, 1996] Black, M. et Rangarajan, A. (1996). On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. Comp. Vision*, 19(1) :58–93.
- [Blake et Isard, 1998] Blake, A. et Isard (1998). *Active Contours*. Springer Verlag.
- [Blake et Zisserman, 1987] Blake, A. et Zisserman, A. (1987). *Visual reconstruction*. MIT Press.
- [Blanc-Lapierre et Picinbono, 1981] Blanc-Lapierre, A. et Picinbono, B. (1981). *Fonctions Aléatoires*. Masson, Col. Sc. et Tech. des Télécoms, Paris.
- [Borgefors, 1996] Borgefors, G. (1996). Distance Transforms in the Square Grid. In Maître, H., éditeur, *Progress in Picture Processing, Les Houches, Session LVIII, 1992*, chapter 1.4, pages 46–80. North-Holland, Amsterdam.
- [Born et Wolf, 1975] Born, M. et Wolf, E. (1975). *Principles of Optics*. Pergamon Press, Oxford (UK).
- [Bovik, 2000] Bovik, A. C. (2000). Basic gray-level image processing. In Bovik, A. C., éditeur, *Handbook of Image and Video Processing*, pages 21–36. Academic Press, San Diego (USA).
- [Boyer et Moore, 1977] Boyer, R. et Moore, J. (1977). A fast string searching algorithm. *Comm. of Ass. Comp. Mach.*, 20(10).
- [Bracewell, 1995] Bracewell, R. (1995). *Two-Dimensional Imaging*. Prentice-Hall, Englewood Cliffs (USA).
- [Bracewell, 1978] Bracewell, R. N. (1978). *The Fourier Transform and its Applications*. Mc Graw Hill.
- [Brodatz, 1966] Brodatz, P. (1966). *Textures*. Dover Pub. (New York).
- [Bunde et Havlin, 1991] Bunde, A. et Havlin, S. (1991). *Fractals and Disordered Systems*. Springer Verlag, Berlin.
- [Bunde et Havlin, 1994] Bunde, A. et Havlin, S. (1994). *Fractals in Science*. Springer Verlag, Berlin.
- [Burt, 1981] Burt, P. (1981). Fast filter transforms for image processing. *Comp. Graphics and Image Proc.*, 16 :20–51.

- [Calle et Montanvert, 1998] Calle, D. et Montanvert, A. (1998). Super-résolution inducing of an image. In *IEEE ICIP'98 - Int. Conf. on Image Processing*, volume III, pages 232–236.
- [Callet, 1998] Callet, P. (1998). *Couleur-lumière couleur matière*. Diderot Ed. Arts et Sciences.
- [Candès, 1999] Candès, E. (1999). Harmonic analysis of neural networks. *Applied and Computational Harmonic Analysis*, 6 :197–218.
- [Candès et Donoho, 1999] Candès, E. et Donoho, D. (1999). Ridgelets : a key to higher-dimensional intermitency? *Phil. Trans. R. Soc. Lond., A.*, 357 :2495–2509.
- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Pattern Analysis and Machine Intelligence*, 8(6) :679–698.
- [Carron, 1995] Carron, T. (1995). *Segmentation d'images couleur dans la base Teinte Luminance Saturation : approche numérique et symbolique*. Thèse de doctorat, Université de Savoie.
- [Carterette et Friedman, 1975] Carterette, E. C. et Friedman, M. P. (1975). *Handbook of perception, Vol V, Seeing*. Academic Press, New-York.
- [Caselles et al., 1997] Caselles, V., Kimmel, R., et Sapiro, G. (1997). On geodesic active contours. *Int. J. Computer Vision*, 21(1) :61–79.
- [Chalmond, 2000] Chalmond, B. (2000). *Éléments de modélisation pour l'Analyse d'Images*, volume 33. Springer Verlag, Mathématiques et Applications.
- [Cham et al., 1998] Cham, F., Lam, F., et Zhu, H. (1998). Adaptive thresholding by variational methods. *IEEE Trans on Image Proc.*, 7(3) :468–473.
- [Chang et Kuo, 1993] Chang, T. et Kuo, C. (1993). Texture analysis and classification with tree structured wavelet transform. *IEEE Trans. on Image Processing*, 2(4) :429–441.
- [Chang et Li, 1994] Chang, Y. et Li, X. (1994). Adaptive image region growing. *IEEE Trans on Image Proc.*, 3 :868–872.
- [Chassery et Chenin, 1980] Chassery, J. M. et Chenin, M. I. (1980). Topologies on Discrete Spaces. In Simon et Haralick, éditeurs, *Digital Image Processing*, pages 59–66. Reidel.
- [Chassery et Montanvert, 1991] Chassery, J. M. et Montanvert, A. (1991). *Géométrie discrète*. Hermes, Paris.
- [Chellappa et Kashyap, 1985] Chellappa, R. et Kashyap, R. (1985). Texture synthesis using 2d non-causal AR models. *IEEE Trans on Acoustic Speech and Signal Proc.*, 33 :194–293.
- [Chen et Pavlidis, 1979] Chen, P. et Pavlidis, T. (1979). Segmentation by textures using a cooccurrence matrix and a split and merge algorithm. *Comp. graph. and Image Proc.*, 10 :172–182.
- [Chen et al., 1991] Chen, S., Lin, W., et Chen, C. (1991). Split and merge image segmentation based on localized image features and statistical tests. *Comp. Vision Graphics and Image Proc : Graph. Models and Image Proc.*, 53 :457–475.
- [Chesnaud et al., 1999] Chesnaud, C., Réfrégier, P., et Boulet, V. (1999). Statistical region snake-based segmentation adapted to different physical noise models. *IEEE PAMI*, 21(11) :1145–1157.
- [Cocquerez et Philipp, 1996] Cocquerez, J. et Philipp, S. (1996). *Analyse d'Images : Filtrage et Segmentation*. Masson.
- [Cœurjolly et al., 2007] Cœurjolly, D., Montanvert, A., et Chassery, J. M. (2007). *Géométrie discrète et images numériques*. Hermes, Paris.
- [Coggins et Jains, 1985] Coggins, J. et Jains, A. (1985). A spatial filtering approach to texture analysis. *Pattern Recognition Letters*, 3(3) :195–203.
- [Cohen, 1992] Cohen, A. (1992). *Ondelettes et traitement numérique du signal*. Masson, Paris.
- [Cohen, 1991] Cohen, L. D. (1991). On active contour models and balloons. *Computer Vision, Graphics, and Image Processing : Image Understanding*, 53(2) :211–218.

- [Cointepas et al., 2001] Cointepas, Y., Bloch, I., et Garnero, L. (2001). A Cellular Model for Multi-Objects Multi-Dimensional Homotopic Deformations. *Pattern Recognition*, 34(9) :1785–1798.
- [Cook et Torrance, 1982] Cook, R. et Torrance, K. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1) :7–24.
- [Cooley et Tukey, 1965] Cooley, J. et Tukey, J. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. of Computers*, 19 :297–301.
- [Cornsweet, 1973] Cornsweet, T. N. (1973). *Visual perception*. Academic Press, New-York.
- [Coster et Chermant, 1985] Coster, M. et Chermant, J. L. (1985). *Précis d'analyse d'images*. Editions du CNRS.
- [Cozannet et al., 1981] Cozannet, A., Fleuret, J., Maître, H., et Rousseau, M. (1981). *Optique et Télécommunications*. Eyrolles, Paris.
- [Cross et Jain, 1983] Cross, G. et Jain, A. (1983). Markov random fields texture models in image analysis. *J. Applied Statistics*, 5(1) :25–39.
- [Cullmann, 1975] Cullmann, G. (1975). *Initiation aux chaînes de Markov, méthodes et applications*. Masson.
- [Darell et al., 1990] Darell, T., Sclaroff, S., et Pentland, A. (1990). Segmentation by minimum description length. In *Int. Conf. on Computer Vision (Osaka)*, pages 112–116.
- [Daubechies, 1992a] Daubechies, I. (1992a). *Ten Lectures on Wavelets*. CBMS-NSF, SIAM, Philadelphia, USA.
- [Daubechies, 1992b] Daubechies, I. (1992b). *Wavelets*. S.I.A.M., Philadelphia.
- [de Valois et al., 1982] de Valois, R. L., Yund, E., et Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22 :531–544.
- [Debled-Rennesson et Reveillès, 1995] Debled-Rennesson, I. et Reveillès, J.-P. (1995). A Linear Algorithm for Segmentation of Digital Curves. *International Journal on Pattern Recognition and Artificial Intelligence*, 9 :635–662.
- [Delahaye, 1994] Delahaye, J. (1994). *Information, Complexité et Hasard*. Hermès, Paris.
- [Demigny et Kamlé, 1997] Demigny, D. et Kamlé, T. (1997). A discrete expression of Canny's criteria for step edge detector. Performances evaluation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 19(11) :1199–1211.
- [Deriche, 1987] Deriche, R. (1987). Using canny's criteria to derive recursively implemented optimal edge detector. *Int. J. Computer Vision*, 1(2) :167–187.
- [Derin et Elliott, 1987a] Derin, H. et Elliott, H. (1987a). Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1).
- [Derin et Elliott, 1987b] Derin, H. et Elliott, H. (1987b). Modeling and segmentation of noisy and textured images using gibbs random fields. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 9 :39–55.
- [Descombes, 1993] Descombes, X. (1993). *Champs Markoviens en Analyse d'Images*. Thèse de doctorat, TELECOM Paris.
- [diZenzo., 1986] diZenzo., S. (1986). A note on the gradient of multi-image. *Computer Vision, Graphics and Image Proc.*, 33 :116–125.
- [Dodgson, 1997] Dodgson, N. (1997). Quadratic interpolation for image resampling. *IEEE Trans. on Image Processing*, 6(9) :1322–1326.
- [Dorst et Smeulders, 1984] Dorst, L. et Smeulders, W. M. (1984). Discrete Representation of Straight Lines. *IEEE Trans. on PAMI*, 6(4) :450–463.
- [Dougherty (Ed.), 1992] Dougherty (Ed.), E. (1992.). *Image Processing and Mathematical Morphology*. Marcel Dekker.
- [Dubuisson, 1990] Dubuisson, B. (1990). *Diagnostic et reconnaissance des formes*. Hermès, Paris.

- [Duda et Hart, 1973] Duda, R. et Hart, P. (1973). *Pattern classification and scene analysis*. Interscience, New-York.
- [Duntley, 1942] Duntley, S. (1942). The optical properties of diffusing materials. *J. Optical Society of America*, 32(2).
- [Durant, 1961] Durant, E. (1961). *Solutions numériques des équations algébriques*, volume Tome 1 et 2. Masson, Paris.
- [Ell et Sangwine, 2000] Ell, T. et Sangwine, S. (2000). Hypercomplex Wiener-Khinchine theorem with application to color image correlation. In IEEE, éditeur, *ICIP 2000*, volume II, pages 792–796.
- [Estournet, 1969] Estournet, D. (1969). Etude statistique d'un signal d'images. *L'Onde Electrique*, page 510.
- [Evans et al., 2000] Evans, C., Sangwine, S. J., et Ell, T. A. (2000). Hypercomplex color-sensitive smoothing filters. In IEEE, éditeur, *ICIP 2000*, volume 1, pages 541–544, Vancouver.
- [Evans, 1998] Evans, L. C. (1998). *Partial differential equations*. Graduate Studies in Mathematics. American Mathematical Society.
- [Fairfield, 1990] Fairfield, J. (1990). Toboggan contrast enhancement for contrast segmentation. In *IAPR 10th Int. Conf. on Pattern Recognition*, pages 712–716, Atlantic City.
- [Faugeras, 1976] Faugeras, O. (1976). *Digital Color Image Processing and Psychophysics within the framework of a human visual model*. Thèse de doctorat, University of Utah.
- [Fischler et Bolles, 1981] Fischler, M. et Bolles, R. (1981). Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the ACM*, 24 :381–385.
- [Fisher, 1995] Fisher, Y. (1995). *Fractal Image Compression*. Springer Verlag, Berlin.
- [Flandrin, 1998] Flandrin, P. (1998). *Temps-fréquence*. Hermès, Paris.
- [Fleuret, 1977] Fleuret, J. (1977). *Holographie numérique et application à divers traitements des images*. Thèse d'état, Université Paris Sud.
- [Foley et al., 1995] Foley, J., van Dam, A., Fisher, S. K., Hughes, J. F., et Phillips, R. L. (1995). *Introduction à l'infographie*. Addison-Wesley, Paris.
- [Forsythe et Moler, 1967] Forsythe, G. et Moler, C. (1967). *Computer solutions of linear algebraic systems*. Prentice hall, Englewood Cliffs (USA).
- [Franel, 1924] Franel, J. (1924). Les suites de Farey et les problèmes des nombres premiers. *Göttinger Nachrichten*, pages 198–201.
- [Françon, 1970] Françon, M. (1970). *Vibrations lumineuses - Optique cohérente*. Dunod, Paris.
- [Freeman, 1961] Freeman, H. (1961). On the encoding of arbitrary geometric configurations. *IEEE Trans on Electr. Comput.*, 10 :260–268.
- [Freeman, 1977] Freeman, H. (1977). Computer processing of line drawing images. *ACM Computer Survey*, 6 :57–97.
- [Fua et Leclerc, 1990] Fua, P. et Leclerc, Y. G. (1990). Model driven edge detection. *Machine Vision and Applications*, 3(1) :45–56.
- [Fukunaga, 1990] Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic Press.
- [Gagalowics et Tournier-Lasserre, 1986] Gagalowics, A. et Tournier-Lasserre, C. (1986). Third order model for non homogeneous natural textures. In *Int. Conf. Pattern Recognition*, volume 86, pages 409–411, Vancouver.
- [Gallager, 1966] Gallager, R. (1966). *Information Theory and reliable communication*. John Wiley, New York.
- [Garello, 2001] Garello, R. (2001). *Analyse de signaux bidimensionnels*. Collection IC2. Hermès, Paris.
- [Geman et Geman, 1984a] Geman, S. et Geman, D. (1984a). Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6) :721–741.

- [Geman et Geman, 1984b] Geman, S. et Geman, D. (1984b). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 12(7) :721–741.
- [Giraudon, 1987] Giraudon, G. (1987). Chaînage efficace de contours. Rapport technique 605, INRIA.
- [Golub et van Loan, 1996] Golub, G. et van Loan, C. F. (1996). *The Singular Value Decomposition*. John Hopkins University Press, Baltimore.
- [Goodman, 1976] Goodman, J. (1976). Some fundamental properties of speckle. *Journal Optical Society of America*, 66(11) :1145–1150.
- [Gordon, 1998] Gordon, I. E. (1998). *Theories of Visual Perception*. John Wiley, New York (USA).
- [Goshtasby et al., 1990] Goshtasby, A., Cheng, F., et Barsky, B. A. (1990). B-spline curves and surfaces viewed as digital filters. *Computer Vision Graphics and Image Processing*, 52 :264–275.
- [Grangeat, 2001] Grangeat, P. (2001). *Tomographie médicale*. Collection IC2. Hermès, Paris.
- [Green et Sibson, 1978] Green, P. J. et Sibson, R. (1978). Computing Dirichlet Tessellation in the Plane. *The Computer Journal*, 21 :168–173.
- [Gregory, 1966] Gregory, R. L. (1966). *L'œil et le cerveau, la psychologie de la vision*. Hachette.
- [Grimson, 1981] Grimson, D. (1981). *From images to surface*. MIT Press, Cambridge USA.
- [Grossberg et Mingolla, 1985] Grossberg, S. et Mingolla, E. (1985). Neural dynamics of perceptual grouping : textures, boundaries and emergent segmentations. *perception Psychophys.*, 38 :141–171.
- [Grunbaum et Shepard, 1989] Grunbaum, B. et Shepard, G. C. (1989). *Tilings and Patterns : an Introduction*. Freeman.
- [Guichard et Morel, 2002] Guichard, F. et Morel, J.-M. (2002). *Image Iterating Smoothing and P.D.E.'s*. Cours de DEA, ENS-Cachan. A paraître.
- [Guillois, 1996a] Guillois, J. (1996a). *Techniques de compression des images*. Hermes, Paris, France.
- [Guillois, 1996b] Guillois, J. P. (1996b). *ITechnique de compression des images*. Hermes, Paris.
- [Guth, 1991] Guth, S. (1991). Model for color vision and light adaption. *J. Opt. Soc. Am. A*, 8(6).
- [Guzman, 1968] Guzman, A. (1968). Decomposition of visual scene into three-dimensional bodies. In *Proc J. computer Conf. AFIPS*, volume 33, pages 291–304.
- [Hadwiger, 1957] Hadwiger, H. (1957). *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*. Springer-Verlag, Berlin.
- [Haralick et Shapiro, 1985] Haralick, R. M. et Shapiro, S. G. (1985). Survey : Image segmentation techniques. *Comp. Vision Graphics and Image Processing*, 29 :100–132.
- [Heath et al., 1997] Heath, M., Sarkar, S., Sanocki, T., et Bowyer, K. (1997). A robust visual method for assessing the relative performance of edge detection algorithms. *IEEE Pattern Analysis and Machine Intelligence*, 19(12) :1338–1359.
- [Hégron, 1985] Hégron, G. (1985). *Synthèse d'image : algorithmes élémentaires*. Bordas, Dunod Informatique, Paris.
- [Heitz et al., 1994] Heitz, F., Pérez, P., et Bouthemy, P. (1994). Multiscale minimization of global energy function in some visual recovery problems. *Comp. Vision Graphics and Image Proc : Image Understanding*, 59(1).
- [Hellstrom, 1967] Hellstrom, C. (1967). Image restoration by the method of least squares. *J. Opt. Soc. Amer.*, 57(3) :297.
- [Horowitz et Pavlidis, 1976] Horowitz, S. et Pavlidis, T. (1976). Picture segmentation by a tree traversal algorithm. *J. Ass. Comp. Mach.*, 23(2) :368–388.
- [Hu, 1962] Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Trans on Information Theory*, IT-8(2) :179–187.

- [Hu et Fahmy, 1992] Hu, R. et Fahmy, M. (1992). Texture segmentation based on a hierarchical Markov random field. *Signal Processing*, 26 :285–305.
- [Huang, 1976] Huang, T. S. (1976). Picture Processing and Digital Filtering. *Topics in Applied Physics*, 6.
- [Huang, 1981] Huang, T. S. (1981). Two Dimensional Signal Processing I. *Topics in Applied Physics*, 42.
- [Hubel et Wiesel, 1969] Hubel, D. et Wiesel, T. (1969). Anatomical demonstration of columns in the monkey striate cortex. *Nature*, 221 :747–750.
- [Huber, 1981] Huber, P. (1981). *Robust Statistics*. John Wiley and Sons, New York.
- [Hueckel, 1971] Hueckel, M. (1971). An operator which locates edges in digitized pictures. *J. Ass. Comp. Mach.*, 18(1) :113–125.
- [Ising, 1925] Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift fur Physik*, 31 :253–258.
- [Jain et Farrokhnia, 1991] Jain, A. et Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12) :1167–1186.
- [Julesz, 1971] Julesz, B. (1971). *Foundations of Cyclopean Perception*. The University of Chicago Press, Chicago.
- [Karabassis et Spetsakis, 1995] Karabassis, E. et Spetsakis, M. (1995). An analysis of image interpolation, differentiation and reduction using local polynomial fits. *Graphical Models and Image Processing*, 57(3) :183–196.
- [Kass et al., 1988] Kass, M., Witkin, A., et Terzopoulos, D. (1988). Active contour models. *Int. J. Comp. Vision*, 1 :321–331.
- [Keren et al., 1990] Keren, D., Marcus, R., Werman, M., et Peleg, S. (1990). Segmentation by minimum length encoding. In *10th Int. Conf. on Pattern recogn. (Atlantic City)*, pages 681–683.
- [Kerker, 1969] Kerker, M. (1969). *The scattering of light and other electromagnetic radiation*. Academic Press, New York.
- [Kervrann et Heitz, 1995] Kervrann, C. et Heitz, F. (1995). A Markov random field model-based approach to unsupervised texture segmentation using local and global statistics. *IEEE Trans Image Processing*, 4(6) :856–862.
- [Khalimsky et al., 1990] Khalimsky, E., Koppermann, R., et Meyer, P. R. (1990). Computer Graphics and Connected Topologies as Finite Ordered Sets. *Topology and its Applications*, 36 :1–17.
- [Kim et Yaroslavski, 1986] Kim, V. et Yaroslavski, L. (1986). Rank algorithms for picture processing. *Comp. Vision Graphics and Image Proc.*, 35 :234–258.
- [Kimia et al., 1992] Kimia, B. B., Tannenbaum, A., et Zucker, S. W. (1992). On the evolution of curves via a function of curvature. I. the classical case. *Journal of Mathematical Analysis and Applications*, 163(2) :438–458.
- [Kirkpatrick et al., 1982] Kirkpatrick, S., Gellatt, C. D., et Vecchi, M. P. (1982). Optimization by simulated annealing. *IBM Thomas J. Watson research Center, Yorktown Heights, NY*.
- [Kittler et al., 1985] Kittler, J., Illingworth, J., et Foglein, J. (1985). Threshold selection based on a simple image statistic. *Comp. Vision Graphics and Image Proc*, 30 :125–147.
- [Klein, 1989] Klein, R. (1989). *Concrete and Abstract Voronoi Diagrams*. Springer Verlag.
- [Klette et Rosenfeld, 2004] Klette, R. et Rosenfeld, A. (2004). *Digital Geometry*. Morgan Kaufmann, San Francisco.
- [Koenderink, 1984] Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50 :363–370.
- [Kolers, 1972] Kolers, P. A. (1972). Reading pictures : Some cognitive aspects of visual perception. In Gordon et Breach, éditeurs, *Picture Band with Compression*, pages 97–126. T. S. Huang and O. J. Tretiak.
- [Kolmogorov, 1965] Kolmogorov, A. (1965). Three approaches of the qualitative definition of information. *Information Transmission*, 1(1) :3–11.

- [Kong et Rosenfeld, 1989] Kong, T. Y. et Rosenfeld, A. (1989). Digital Topology : Introduction and Survey. *Computer Vision, Graphics, and Image Processing*, 48 :357–393.
- [Kotropoulos et al., 1997] Kotropoulos, C., Gabrani, M., et Pitas, I. (1997). Adaptive multichannel L-filters with structural constraints. In *IEEE - Int. Work on Non Linear Signal and Image Processing, NSIP'97*, pages 181–183, Michigan.
- [Kovalesky, 1989] Kovalesky, V. A. (1989). Finite Topology as applied to Image Analysis. *Computer Vision, Graphics, and Image Processing*, 46 :141–161.
- [Kunduri et al., 1999] Kunduri, S., Maître, H., Roux, M., Demigny, D., et Karabernou, M. (1999). An improved method for isotropic edge orientation estimation . In *SPIE, éditeur, Europto '99*, Florence.
- [Laine et Fan, 1993] Laine, A. et Fan, J. (1993). Texture classification by wavelet packet signature. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 15 :1186–1190.
- [Lakshmanan et Derin, 1989] Lakshmanan, S. et Derin, H. (1989). Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11(8) :799–813.
- [Leclerc, 1989] Leclerc, Y. (1989). Constructing simple stable descriptions for image partitionning. *Int. J. of Computer Vision*, 3(1989) :73–102.
- [Lee et al., 1998] Lee, K.-M., Meer, P., et Park, R.-H. (1998). Robust adaptive segmentation of range image. *IEEE T. Pattern Analysis and Machine Intelligence*, 20(2) :200–205.
- [Lee, 1998] Lee, T. (1998). Segmenting images corrupted by correlated noise. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 20(5) :481–492.
- [LeGrand, 1956] LeGrand, Y. (1956). *Optique Physiologique*. Editions de la Revue d'Optique, Paris.
- [Lelong-Ferrand et Arnaudès, 1971] Lelong-Ferrand, J. et Arnaudès, J. (1971). *Cours de Mathématiques, Algèbre*, volume 511. Dunod Universités, Paris.
- [Linde et al., 1980] Linde, Y., Buzo, A., et Gray, R. (1980). An algorithm for vector quantizer designs. *IEEE Trans on Communications*, C-28 :84–95.
- [Locher et al., 1972] Locher, J. L., Broos, C. H. A., Escher, M. C., Locher, G. W., et Coxeter, H. S. M. (1972). *L'œuvre de M. C. Escher*. Editions du Chêne, Paris.
- [Lowrance et Wagner, 1975] Lowrance, R. et Wagner, R. (1975). An extension of the string to string correction problem. *J. Ass. Comp. Mach.*, 22(2) :177–183.
- [Lu et Fu, 1978] Lu, S. et Fu, K. (1978). A syntactic approach to texture analysis. *Computer Graphics and Image Proc.*, 7(3) :303–330.
- [Lu et Fu, 1979] Lu, S. et Fu, K. (1979). Stochastic tree grammar inference for texture synthesis and discrimination. *Computer Graphics and Image Proc.*, 9 :234–245.
- [Maeland, 1988] Maeland, E. (1988). On the comparison of interpolation methods. *IEEE Trans. on Medical Imaging*, 7(3) :213–217.
- [Maître, 1977] Maître, H. (1977). Propriétés fondamentales des signaux d'image. *GRETSI*, 1 :34–40.
- [Maître, 1981] Maître, H. (1981). Iterative picture restoration using video optical feedback. *Comp. Graph. and Image Proc.*, 16 :95–115.
- [Maître, 1985a] Maître, H. (1985a). Un panorama de la transformation de Hough. *Traitement du Signal*, 2(4) :305–317.
- [Maître, 1985b] Maître, H. (1985b). Un panorama de la transformation de Hough. *Traitement du Signal*, 2(4) :305–317.
- [Maître, 1986] Maître, H. (1986). Contribution to the Prediction of Performances of the Hough Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(5) :669–674.

- [Maître, 1996] Maître, H. (1996). Entropy, information and image. In Maître, H. et Zinn-Justin, J., éditeurs, *Progress in Picture Processing*, volume LVIII, pages 81–117. North Holland, Amsterdam.
- [Maître, 2001] Maître, H. (2001). *Le traitement des images de Radar à Synthèse d'Ouverture (Ouvrage collectif)*. Hermes, collection IC2, Paris.
- [Mallat, 1989] Mallat, S. (1989). A theory of multiresolution signal decomposition : the wavelet representation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11 :674–693.
- [Mallat, 2000] Mallat, S. (2000). *Une exploration des signaux en ondelettes*. Les éditions de l'Ecole Polytechnique, Palaiseau, France.
- [Mallat et Zhong, 1993] Mallat, S. et Zhong, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12) :3397–3415.
- [Mandelbrot, 1977] Mandelbrot, B. (1977). *Fractals, Form, Chance and Dimension*. W.H. Freeman, San Francisco.
- [Manjunath et Chellapa, 1991] Manjunath, B. et Chellapa, R. (1991). Unsupervised texture segmentation using Markov random field models. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 13(5) :478–482.
- [Marion, 1997] Marion, A. (1997). *Acquisition et Visualisation des Images*. Eyrolles, Paris.
- [Marr et Hildreth, 1980] Marr, D. et Hildreth, E. (1980). Theory of edge detection. In *Proc. Royal Soc. London*, volume 207 of *B*, pages 187–217.
- [Martelli, 1972] Martelli, A. (1972). Edge detection using heuristic search methods. *Comp. Graph. and Image Proc.*, 1 :169–182.
- [Mayhew et Frisby, 1978] Mayhew, J. et Frisby, J. P. (1978). Texture discrimination and Fourier analysis in the human visual system. *Nature*, 275 :438–439.
- [Meer et al., 1991] Meer, P., Mintz, D., Rosenfeld, A., et Kim, D. (1991). Robust regression methods for computer vision : A review. *Int. J. Computer Vision*, 6 :59–70.
- [Melamed, 1963] Melamed, N. (1963). Optical properties of powder. *J. of Applied Physics*, 34 :560–570.
- [Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, N. M., Teller, A. H., et Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chemical Physics*, 21 :1087–1091.
- [Meyer, 1990] Meyer, Y. (1990). *Ondelettes. Ondelettes et opérateurs*. Hermann, Paris. Tome 1.
- [Miclet, 1984] Miclet, L. (1984). *Méthodes structurelles pour la reconnaissance des formes*. Eyrolles, Paris.
- [Modestino et Fries, 1977] Modestino, J. et Fries, R. (1977). Edge detection in noisy images using recursive digital filters. *Comp. Graph. and Image Proc.*, pages 409–433.
- [Montanari, 1971] Montanari, U. (1971). On the optimal detection of curves in noisy images. *Com. ACM*, 14 :335–345.
- [Mumford et Shah, 1989] Mumford, D. et Shah, J. (1989). Optimal approximations by piecewise smoothed functions and associated variational problems. *Com. Pure and Applied Math.*, XLII :577–685.
- [Murch, 1973] Murch, G. M. (1973). *Visual and auditory perception*. Bobbs-Merrill, New-York.
- [Natsuyama, 1979] Natsuyama, M. N. T. (1979). Edge preserving smoothing. *Computer Graphics and Image Processing*, 9 :394–407.
- [Nayatani et al., 1990] Nayatani, Y., Takahama, K., Sobagaki, H., et Hashimoto, K. (1990). Color-appearance model and chromatic-adaption transform. *Color Research and Application*, 15(14) :210–221.
- [Nevatia et Babu, 1980] Nevatia, R. et Babu, K. (1980). Linear feature extraction and description. *Comp. Graphics and Image Proc.*, 13(3) :257–269.
- [Nicodemus et al., 1977] Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., et Limperis, T. (1977). *Geometrical considerations and nomenclature for reflectance, N. 160*. National Bureau of Standards, Washington.

- [Nicolas, 1995] Nicolas, J. M. (1995). Techniques numériques de traitement du signal : les aspects 2d. ENST, cours de dominante SIG/IMA, Paris.
- [Nicolas, 1999] Nicolas, J. M. (1999). Systèmes d'imagerie satellitale, cours de 3a,. Rapport technique, ENST, Paris.
- [Ohlander et al., 1978] Ohlander, R., Price, K., et Reddy, D. (1978). Picture segmentation using a recursive region splitting method. *Comp. Graphics and Image Proc.*, 8 :313–333.
- [Olivier et al., 1997] Olivier, C., Jouzel, F., ElMatouat, A., et Courtellemont, P. (1997). Un nouveau critère pour la sélection de l'ordre d'un modèle. In *GRETSI'97*, pages 451–454, Grenoble.
- [Osher et Sethian, 1988] Osher, S. et Sethian, J. (1988). Front propagating with curvature dependant speed : Algorithm based on the Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79 :12–49.
- [Palmer et al., 1996] Palmer, P., Dabis, H., et Kittler, J. (1996). A performance measure for boundary detection algorithms. *Comp. Vis. and Image Understanding*, 63(3) :476–494.
- [Peitgen et Saupe, 1988] Peitgen, H. et Saupe, D. (1988). *The science of fractal images*. Springer Verlag, Berlin.
- [Pentland, 1984] Pentland, A. (1984). Fractal based description of natural scenes. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 6 :661–674.
- [Perona et Malik, 1987] Perona, P. et Malik, J. (1987). Scale space and edge detection using anisotropic diffusion. In *IEEE Workshop on Computer Vision*, pages 16–22, Miami. IEEE.
- [Perona et Malik, 1990a] Perona, P. et Malik, J. (1990a). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7) :629–639.
- [Perona et Malik, 1990b] Perona, P. et Malik, J. (1990b). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7) :629–639.
- [Pesquet-Popescu et Pesquet, 2001] Pesquet-Popescu, B. et Pesquet, J.-C. (2001). Ondelettes et applications. *Techniques de l'Ingénieur, traité Télécoms*, TE 5 215.
- [Petrou et Kittler, 1991] Petrou, M. et Kittler, J. (1991). Optimal edge detectors for ramp edges. *IEEE-Pattern Analysis and Machine Intelligence*, 13(5) :1483–1491.
- [Pham, 1986] Pham, S. (1986). Digital Straight Segments. *Computer Vision, Graphics, and Image Processing*, 36 :10–30.
- [Phillips, 1962] Phillips, D. (1962). A technique for the numerical solution of certain integral equations of the first kind. *J. Ass. Comp. Mach.*, 9 :84.
- [Phillips et Wilson, 1984] Phillips, G. C. et Wilson, H. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *JOSA - A*, 1 :226–232.
- [Phong, 1975] Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 6(18) :311–317.
- [Pirenne et Crouzy, 1972] Pirenne, M. H. et Crouzy, R. (1972). *Gauthiers-Villars. L'œil et la vision*, Paris.
- [Pitas et Venetsanopoulos, 1990] Pitas, I. et Venetsanopoulos, A. (1990). *Nonlinear digital filters*. Kluwer Academic.
- [Porte et Vignes, 1974] Porte, M. L. et Vignes, J. (1974). *Algorithmes Numériques (Tome 1)*. Technip, Paris.
- [Portilla et Simoncelli, 2000] Portilla, J. et Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelets coefficients. *International Journal of Computer Vision*, 40(1) :49–71.
- [Pratt, 1975] Pratt, W. (1975). Vector formulation of 2D signal processing operations. *Comp. Graph. and Image Proc.*, 4(1) :1–24.
- [Pratt, 1978] Pratt, W. (1978). *Digital image processing*. Wiley Interscience, New-York.
- [Preparata et Shamos, 1988] Preparata, F. P. et Shamos, M. I. (1988). *Computational Geometry, an Introduction*. Springer Verlag.

- [Pérez, 1991] Pérez, J. P. (1991). *Optique géométrique, ondulatoire et polarisation*. Masson, Paris.
- [Princen et al., 1992] Princen, J., Illingworth, J., et Kittler, J. (1992). A formal definition of the Hough transform : properties and relationships. *J. Math. Image and Vision*, 1 :153–168.
- [Randen et Husøy, 1999] Randen, T. et Husøy, J. (1999). Filtering of texture classification : a comparative study. *IEEE Pattern Analysis and Machine Intelligence*, 21(4) :291–310.
- [Reed et du Buf, 1993] Reed, T. et du Buf, H. (1993). A review of recent texture segmentation and feature extraction techniques. *Comp. Vision Graphics and Image Proc : Image Undersrtand.*, 57(3) :359–372.
- [Richardson, 1961] Richardson, L. (1961). The problem of contiguity : an appendix of statistics of deadly quarrels. *General System Year*, 6 :139–187.
- [Rissanen, 1984] Rissanen, J. (1984). Universal coding, information, prediction, and estimation. *IEEE Tr. Inform. Theory*, 30 :629–636.
- [Rissanen, 1987] Rissanen, J. (1987). Minimum description length principle. *Encyclopedia of Statistical Sciences*, 3 :523–527.
- [Rissanen, 1989] Rissanen, J. (1989). *Stochastic complexity in Statistical Inquiry*. World Scientific, Singapore.
- [Robinson, 1977] Robinson, G. (1977). Edge detection by compass gradient masks. *Computer Graphics and Image Processing*, 6 :492–501.
- [Roddiier, 1999] Roddiier, F. (1999). *Adaptive Optics in Astronomy*. Cambridge University Press, Cambridge.
- [Ronse, 1985] Ronse, C. (1985). A Simple Proof of Rosenfeld’s Characterization of Digital Straight Line Segments. *Pattern Recognition Letters*, 3 :323–326.
- [Rosenfeld, 1970] Rosenfeld, A. (1970). Connectivity in Digital Pictures. *Journal of ACM*, 17(1) :146–160.
- [Rosenfeld, 1974] Rosenfeld, A. (1974). Digital Straight Line Segments. *IEEE Trans. on Computers*, 23(12) :1264–1269.
- [Rosenfeld, 1979] Rosenfeld, A. (1979). Digital Topology. *Amer. Math. Monthly*, pages 621–630.
- [Rousseeuw et Leroy, 1987] Rousseeuw, P. et Leroy, A. (1987). *Robust regression and outlier detection*. John Wiley, New-York.
- [Rudin, 1987] Rudin, L. (1987). *Images, Numerical Analysis of Singularities and Shock Filters*. Thèse de doctorat, California Institute of Technology.
- [Rudin et Osher, 1994] Rudin, L. et Osher, S. (1994). Total variation based image restoration with free local constraints. In *IEEE ICIP’94*, volume I, pages 31–35, Dallas (TX).
- [Rudin et al., 1992a] Rudin, L., Osher, S., et Fatemi, E. (1992a). Nonlinear total variation based noise removal algorithms. *Physica D* 60, No.1-4, 259-268. [ISSN 0167-2789].
- [Rudin et al., 1992b] Rudin, L., Osher, S., et Fatemi, E. (1992b). Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4) :259–268.
- [Rudin, 1998] Rudin, W. (1998). *Analyse Réelle et Complexe*. Dunod, 3eme edition.
- [Sahoo et al., 1988] Sahoo, P., Soltani, S., Wong, A., et Chen, Y. (1988). A survey of thresholding techniques. *Comp. Vision Graphics and Image Proc.*, 41(1) :233–260.
- [Saint-Marc et al., 1989] Saint-Marc, P., Chen, J., et Medioni, G. (1989). Adaptive smoothing : A general tool for early vision,. In *CVPR’89*, pages 618–624.
- [Saito et Coifman, 1995] Saito, N. et Coifman, R. (1995). Local discriminant bases and their applications. *J. Mathematical Imaging and Vision*, 5(4) :337–358.
- [Sangwine, 1998] Sangwine, S. (1998). Colour image edge-detector based on quaternion convolution. *Electronics Letters*, 34(10) :969–971.

- [Sangwine, 1996] Sangwine, S. J. (1996). Fourier transforms of colour images using quaternions or hypercomplex numbers. *Electronics Letters*, 32(21) :1979–1980.
- [Santalo, 1940] Santalo, L. A. (1940). Complemento a la Nota : un Teorema Sobre Conjuntos de Paralelepipedos de Aristas Paralelas. *Publ. Int. Mat. Univ. Nac. Litoral*, 2 :49–60.
- [Santalo, 1976] Santalo, L. A. (1976). *Integral Geometry and Geometric Probability*. Addison Wesley.
- [Saporta, 1978] Saporta, G. (1978). *Théorie et Méthodes Statistiques*. Princeton University Press.
- [Saporta, 1990] Saporta, G. (1990). *Probabilités, analyse de données et statistiques*. Editions Technip, Paris.
- [Sarkar et Boyer, 1991] Sarkar, S. et Boyer, K. (1991). Optimal infinite impulse response zero crossing based edge detectors. *Comp. Vision Graphics and Image Proc Image Understanding*, 54(2) :224–243.
- [Schmitt et Mattioli, 1994a] Schmitt, M. et Mattioli, J. (1994.a). *Morphologie mathématique*. Masson, Paris.
- [Schmitt et Mattioli, 1994b] Schmitt, M. et Mattioli, J. (1994b). *Morphologie Mathématique*. Masson.
- [Schwartz, 1995] Schwartz, L. (1995). *Analyse, tome I : Théorie des ensembles et topologie*. Eds. Hermann, Paris.
- [Serra, 1982a] Serra, J. (1982a). *Image Analysis and Mathematical Morphology*. Academic Press, New-York.
- [Serra, 1982b] Serra, J. (1982b). *Image analysis and mathematical morphology*. Academic Press, New York.
- [Serra (Ed.), 1988] Serra (Ed.), J. (1988.). *Image Analysis and Mathematical Morphology, Part II : Theoretical Advances*. Academic Press, London.
- [Sethian, 1985] Sethian, J. (1985). Curvature and the evolution of fronts. *Comm. Math. Phys.*, 101.
- [Sethian, 1996] Sethian, J. (1996). *Level-set methods*. Cambridge University Press.
- [Shanmugam et Green, 1979] Shanmugam, K. et Green, J. (1979). An optimal frequency domain filter for edge detection in digital pictures. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1 :37–49.
- [Shannon, 1948] Shannon, C. (1948). The mathematical theory of communications. *Bell System Technical J.*, 27 :379–423.
- [Shen et Castan, 1986] Shen, J. et Castan, S. (1986). An optimal linear operator for edge detection. In *Proceeding CVPR Miami*.
- [Sigelle, 1993] Sigelle, M. (1993). *Champs de Markov en traitement d'images et modèles de la physiques statistique : applications en relaxation d'images de classification*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications.
- [Sklansky, 1978] Sklansky, J. (1978). On the Hough technique for curve detection. *IEEE Trans. Comp.*, 27(100) :923–926.
- [Soille, 1999] Soille, P. (1999). *Morphological Image Analysis*. Springer-Verlag, Berlin.
- [Spacek, 1986] Spacek, L. (1986). Edge detection and motion detection. *Image and Vision Computing*, 4 :43–56.
- [Suk et Chung, 1993] Suk, M. et Chung, S. (1993). A new image segmentation technique based on partition mode test. *Pattern Recognition*, 16(5) :469–480.
- [Sève, 1996] Sève, R. (1996). *Physique de la couleur : de l'apparence colorée à la technique colorimétrique*. Masson - Physique fondamentale et appliquée, Paris.
- [Tagare et Figueiredo, 1985] Tagare, H. et Figueiredo, R. (1985). Order filters. *Proc of the IEEE*, 73(1) :163–165.
- [Terzopoulos, 1983] Terzopoulos, D. (1983). Multilevel computational process for visual surface reconstruction. *Comp. Vision Graph. and Image Proc.*, 24 :52–96.
- [Tikonov et Arsenine, 1974] Tikonov, A. et Arsenine, V. (1974). *Méthode de résolution de problèmes mal-posés*. Editions MIR, Moscou.
- [Torrance et Sparrow, 1967] Torrance, K. et Sparrow, E. (1967). Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57(9) :1105–114.

- [Torre et Poggio, 1986] Torre, V. et Poggio, T. (1986). On edge detection. *IEEE Pattern Analysis and Machine Intelligence*, 8 :147–153.
- [Toussaint, 1985] Toussaint, G. T. (1985). *Computational Geometry*. North Holland, Amsterdam.
- [Twomey, 1965] Twomey, S. (1965). The application of numerical filtering to the solution of the integral equations encountered in indirect sensing measurements. *J. Franklin Institute*, 279 :95–109.
- [Unser, 1986] Unser, M. (1986). Local linear transforms for texture measurement. *Signal Processing*, 11(1) :61–79.
- [Unser, 1995] Unser, M. (1995). Texture classification and segmentation using wavelet frame. *IEEE Trans on Image Processing*, 4(11) :1549–1560.
- [Unser, 1999] Unser, M. (1999). Splines. a perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6) :22–38.
- [Unser et al., 1993a] Unser, M., Aldroubi, A., et Eden, M. (1993a). B-spline signal processing - Part i. *IEEE Trans. on Signal Processing*, 41(2) :821–833.
- [Unser et al., 1993b] Unser, M., Aldroubi, A., et Eden, M. (1993b). B-spline signal processing - Part II. *IEEE Trans. on Signal Processing*, 41(2) :834–848.
- [VanGool et al., 1985] VanGool, L., Dewaele, P., et Oosterlinck, A. (1985). Textural analysis anno 1983. *Comp. Graph. and Image Processing*, 29 :336–357.
- [Vetterli et Kovačević, 1995] Vetterli, M. et Kovačević, J. (1995). *Wavelets and subband coding*. Prentice Hall, Englewood Cliffs, New Jersey.
- [Vickers et Modestino, 1982] Vickers, A. et Modestino, J. (1982). A maximum likelihood approach to texture classification. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 4 :61–68.
- [Vincent, 1992] Vincent, L. (1992). Morphological Algorithms. In Dougherty, E., éditeur, *Mathematical Morphology in Image Processing*, pages 255–288. Marcel Dekker.
- [Vincent et Soille, 1991] Vincent, L. et Soille, P. (1991). Watersheds in Digital Spaces : an Efficient Algorithm based on Immersion Simulations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6) :583–598.
- [Wagner et Fisher, 1974] Wagner, R. et Fisher, M. (1974). The string to string correction problem. *J. Ass. Comp. Mach.*, 21(1) :168–173.
- [Wang, 1998] Wang, J.-P. (1998). Stochastic relaxation on partitions with connected components and its application to image segmentation. *IEEE Pattern Analysis and Machine Intelligence*, 20(6) :619–636.
- [Weickert, 1998] Weickert, J. (1998). *Anisotropic Diffusion In Image Processing*. Teubner, Stuttgart.
- [Weickert et al., 1998] Weickert, J., ter Haar Romeny, B., et Viergever, M. (1998). Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7 :398–410.
- [Wertheimer, 1944] Wertheimer, M. (1944). Gestalt Theorie. *Social Research*, 11 :78–99.
- [Weszka, 1978] Weszka, J. (1978). A survey of threshold selection techniques. *Comp. Graphics and Image Proc.*, 7 :259–265.
- [Winkler, 1995] Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction*. Applications of mathematics. Springer-Verlag.
- [Witkin, 1983a] Witkin, A. (1983a). Scale-space filtering. In *International Joint Conference on Artificial Intelligence*, pages 1019–1023.
- [Witkin, 1983b] Witkin, A. P. (1983b). Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Conference*, pages 1019–1021.
- [Won et Derin, 1992] Won, C. et Derin, H. (1992). Unsupervised segmentation of noisy and textured images using Markov random fields. *Comp. Vision Graphics and Image Proc. : Graphical Models and Image Processing*, 54(4) :308–328.

- [Wood, 1981] Wood, J. W. (1981). Two-Dimensional Kalman filtering. In Huang, T., éditeur, *Two-Dimensional Digital Signal Processing I*, volume 42 of *Topics in Applied Physics*. Springer Verlag, Berlin.
- [Wu, 1982] Wu, F. Y. (1982). The potts model. *Review of Modern Physics*, 54(1).
- [Wu et Maître, 1992] Wu, Y. et Maître, H. (1992). Smoothing speckled synthetic aperture radar images by using maximum homogeneous region filters. *Optical Engineering*, 31(8) :1785–1792.
- [Zamperoni, 1996] Zamperoni, P. (1996). Feature extraction. In Maître, H. et Zinn-Justin, J., éditeurs, *Progress in Picture Processing, Les Houches 92*, volume LVIII, pages 119–182. North-Holland.
- [Zhang et Modestino, 1990] Zhang, J. et Modestino, J. (1990). A model fitting approach to cluster validation with applications to stochastic model-based image segmentation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 4 :1009–1017.
- [Zhao et al., 2000] Zhao, H.-K., Osher, S., Merriman, B., et Kang, M. (2000). Implicit non-parametric shape reconstruction from unorganized data using variational level set method. *Comp. Vision Graphics and Image Proc*, 80(3) :295–314.
- [Zhu et Yuille, 1996] Zhu, S. et Yuille, A. (1996). Region competition : Unifying snakes, region growing and Bayes/MDL for multiband image segmentation. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18 :884–900.
- [Zohar, 1974] Zohar, S. (1974). The solution of toplitz set of linear equations. *J. of the Ass. Comp. Mach.*, 21(2).

Index

- échantillonnage, 40
- échelle, 154
- élément structurant, 95
- élément structurant fonctionnel, 101
- équation
 - de diffusion, 166
 - de Perona et Malik, 166
- équation de Perona et Malik, 134
- équation de Rudin, Osher, et Fatemi, 137
- équation de la chaleur, 127
- équation de la chaleur inverse, 129
- équations de diffusion, 127
- équations de diffusion non linéaires, 133
- érodé ultime, 105
- érosion à niveaux de gris, 99
- érosion binaire, 97
- 4-connexité, 58
- 6-connexité, 59
- 8-connexité, 58

- adaptation chromatique, 262
- addition de Minkowski, 96
- adjoint, 150
- algorithme
 - de la corde, 237
- algorithme de poursuite, 155
- aliasing, 44
- AMSS, 143
- analyse
 - en composantes principales, 197
- analyse de Gabor, 146
- analyse multi-échelles, 138
- aplatissement, 214
- approximation, 154
- approximations polygonales, 234
- attache aux données, 198
- autorégressif, 219

- B-splines, 171
- bâtonnets, 12
- balayage, 265

- banc de filtres, 156
- bande de Mach, 14
- base
 - bi-orthogonale, 151–153
 - d’ondelettes, 154
 - de Riesz, 151
 - de Riesz, 154
 - orthonormale, 149, 154
- bassins versants, 109
- BDRF, 248
- Bhattacharyya, 197
- blanc de référence, 255
- Blinn, 248
- boîte englobante, 226
- boussoles directionnelles, 181
- Bracewell, 84
- Bresenham, 66
- BV, 167

- C-moyennes floues, 236
- cônes, 12
- calibrage, 91
- Canny, 182
- causalité, 35
- cercle discret, 70
- chaînes de Freeman, 228
- champ
 - de Markov, 198, 220
- chapeau haut-de-forme, 108
- chatoiement, 24, 246
- chemin, 60
- chi², 202
- chiasma optique, 12
- chroma, 254, 256
- chrominance, 255
- CIE, 252
- CIELab, 256
- clarté, 255
- codage fractal, 50
- coefficient de variation, 214
- coefficients d’ondelettes, 153

- cohérence, 24
- complexe cellulaire, 63
- complexité, 83, 86
- composantes trichromatiques, 251, 252
- conditionnement, 84, 271
- coniques, 236
- connexité, 60
- contour, 147, 154, 176
- contours actifs, 186
- contraste morphologique, 105
- cooccurrence, 215, 221
- cortex visuel, 12
- courbes d'excitation, 253
- courbure, 134
- critère
 - de Neyman Pearson, 196
- critère d'Akaike, 197
- croissance de région, 200
- cubique, 171
- curvelets, 154

- décision bayésienne, 196
- décomposition
 - dyadique, 153
- décomposition en valeurs singulières, 86, 269
- détection de contours, 175
- Deriche, 183
- descripteurs de Fourier, 232
- dictionnaire, 145
- diffusion, 127, 166
- diffusion anisotrope, 133, 166
- dilatation à niveaux de gris, 99
- dilatation binaire, 96
- dimension, 48
- distance
 - interquartile, 199, 214
- distance de chanfrein, 76
- distance discrète, 76
- distance frustrée, 265
- distribution de Poisson, 75
- DOG, 181

- EDP, 127
- ensemble
 - de niveau, 189
- ensemble de niveau, 131
- entrelacement, 265
- entropie, 32, 199
- espace Lab, 255
- espace XYZ, 254

- estimation
 - robuste, 235

- famille
 - d'ondelettes, 146
 - de Gabor modifiée, 147
 - de Weyl-Heisenberg, 146
 - discrète, 149
- fausse alarme, 196
- fermeture, 180
 - de contours, 190
- fermeture à niveaux de gris, 103
- fermeture binaire, 102
- FFT, 83
- filtrage de Perona et Malik, 134
- filtrage de Rudin, Osher et Fatemi, 137
- filtrage par fenêtre maximale, 169
- filtre
 - adapté, 164
 - de diffusion anisotrope, 179
 - de Nagao, 169
 - de Perona et Malik, 166
 - de rang, 165
 - de Rudin, Osher et Fatemi, 168
 - de Saint Marc, 168
 - de Wiener, 164
 - de Wu, 169
 - LMS, 236
 - médian, 235
 - morphologique, 165
 - sigma, 168
 - Toboggan, 169, 179
- filtre alterné séquentiel, 107
- filtre de choc, 138
- filtre inverse, 80
- filtre morphologique, 102, 107
- filtres
 - à phase nulle, 178
 - de Gabor, 217
 - miroirs, 218
- Fisher-Snedecor, 202
- flou, 44
- fonction
 - d'autocorrélation, 215
- fonction de transfert, 43
- fonctions colorimétriques, 252
- fonctions de mélange, 253
- forme développée, 87
- fovéa, 12
- fréquence de Nyquist, 41

- fractal, 46, 47
- Freeman, 228
- Fuzzy-C means, 236

- Gabor, 217
- Gauss-Seidel, 89
- Gestalt Theorie, 17, 210
- gradient, 128
- gradient , 178
- gradient conjugué, 89
- gradient morphologique, 105
- Graduated Non Convexity, 208
- Gram-Schmidt, 156
- granularité, 24, 246, 248
- granulométrie, 108
- graphe
 - d'adjacence, 202
- groupement perceptuel, 210
- growing, 200
- Guzman, 226

- Hilbert, 265
- homogénéité, 161
- homothétie interne, 46, 48
- Hough, 240
- Hueckel, 182
- hystérésis, 179

- ICM, 221
- illusion d'optique, 17
- image d'étiquettes, 223
- incohérence, 24
- indice d'allongement, 225
- interactions spatiales, 14
- interpolation, 171
- invariance affine, 143
- invariance par changement de contraste, 130
- inverse, 148, 150

- Jacobi, 88
- Jordan, 62
- JPEG2000, 151
- Julesz, 214

- Karhunen Loeve, 197
- Kirsch, 181
- kurtosis, 214

- labels, 223
- lagrangien, 276
- Lambert, 248

- laplacien, 128
- laplacien , 180
 - d'une gaussienne, 181
- LBG, 266
- level sets, 189
- ligne de paratage des eaux, 109
- lignes
 - de crête, 180
- lignes de niveau, 131
- linéarité, 161
- LMS, 12
- LOG, 180
- log-vraisemblance, 204
- loi de Lambert, 248
- look-up table, 161
- lumière cohérente, 24

- M-estimateurs, 236
- mélange, 236
- métamères, 250
- méthode de projection, 90
- méthodes itératives, 87
- Mahalanobis, 197
- maillage, 56
- Marr, 181
- Martelli, 191
- masques adaptés, 181
- matching pursuit, 155
- matrice circulante, 81
- matrice de Töplitz, 81
- matrice hessienne, 133
- matrice Töplitz-blocs-Töplitz, 87
- Maximum A Posteriori, 198
- maximums du gradient, 176
- mean curvature, 167
- membrane, 206
- merging, 201
- mesure, 46
- mesures morphologiques, 104
- modèles
 - géodésiques actifs, 190
- moindres carrés, 272
- moment
 - invariant, 226
 - normé, 226
- Montanari, 191
- morphologie mathématique, 93
- mouvement par courbure moyenne, 135, 167
- moyenne frustrée, 265
- multi-échelles, 138

- Mumford et Shah, 205
- Netto, 265
- Nevatia et Babu, 181
- niveau de résolution, 153
- nombre d'Euler, 63
- non-linéarité, 161
- non-redondance, 151
- NTSC, 260
- nuées dynamiques, 236
- Nyquist, 41
- ondelette
 - mère, 146, 147
- ondelettes, 145, 218
- ordre, 263
- ordre canonique, 264
- ordre partiel, 263
- ordre total, 263
- orientation, 178
- orthogonalisation, 156
- ouverture à niveaux de gris, 103
- ouverture binaire, 102
- PAL, 260
- partage de région, 200
- partition, 54, 199
- pavages, 54
- Peano, 48, 265
- photopique, 12
- plaques minces, 206
- plus grande pente, 89
- POCS, 90
- Potts, 206
- poursuite, 180
- poursuite adaptative, 154
- pré-ordre, 263
- prédicat, 199
- prétraitements, 161
- Prewitt, 181
- primal sketch, 180
- probabilité de détection, 196
- processus
 - ARMA, 219
 - de bords, 199, 221
- produit scalaire, 149
- programmation dynamique, 191
- propriété de la corde, 67
- pseudo-inverse, 85, 150, 271
- pyramide, 202
- QMF, 218
- quad-tree, 202
- quantification adaptée, 214
- quantification vectorielle, 266
- quaternions, 268
- quinconce, 45
- réflexion bidirectionnelle, 248
- régularisation, 85, 198
- réponse impulsionnelle, 43
- résidu, 88
- RanSac, 236
- rapport de vraisemblance, 196
- rapport iso-périmétrique, 225
- reconstruction, 41
- recuit simulé, 221
- redondance, 34, 42
- repliement de spectre, 44
- représentation
 - complète, 147
 - redondante, 149
- ridgelets, 147
- Rudin, 138
- Saint Marc, 168
- saturation, 254
- scotopique, 12
- SECAM, 260
- Shannon, 41
- Shen et Castan, 184
- singularité, 84
- Smirnov-Kolmogorov, 221
- snakes, 186
- solution principale, 84
- sous-dictionnaire, 156
- speckle, 24, 246, 248
- spectre de densité de puissance, 32
- spline, 147, 156, 238
- splines, 171
- splitting, 200
- stationnarité, 35
- statistiques du premier ordre, 214
- structure locale des images, 133
- Student, 202
- suite de Farey, 69
- Suk, 202
- sur-segmentations, 202
- suréchantillonnage, 42
- SVD, 86, 269
- synthèse d'ouverture, 90

table de transcodage, 161
tache d'Airy, 43
tavelure, 24, 246
teinte, 254, 256
template matching, 181
tests statistiques, 202
texture, 209
théorème
 de Netto, 265
théorie bayésienne de la décision, 196
topologie, 56
Torrance-Sparrow, 248
trame, 149
 ajustée, 149
 d'ondelettes, 153
 de Gabor modifiées, 154
 de ridgelets, 154
 de Weyl-Heisenberg, 153
 duale, 150, 151
trame hexagonale, 45
trames
 algorithme des , 152
transformée
 de Fourier, 145
transformée
 en Z, 172
transformation
 linéaire, 145
transformation
 de Hough, 239
tri planaire, 223
triangle de Maxwell, 252
triangulation de Delaunay, 72
trivariance, 250
TV, 167

variation totale, 137
voisinage, 58
von Koch, 48
Voronoi, 72

Wiener, 84
Wiener-Khinchine, 32
Wiener-Kinchine, 215
Wilcoxon, 202

Yule-Walker, 219

zéro du laplacien, 180
zigzag, 265