



HAL
open science

Filtrage, réduction de dimension, classification et segmentation morphologique hyperspectrale

Guillaume Noyel

► **To cite this version:**

Guillaume Noyel. Filtrage, réduction de dimension, classification et segmentation morphologique hyperspectrale. Mathématiques [math]. École Nationale Supérieure des Mines de Paris, 2008. Français. NNT : 2008ENMP1558 . pastel-00004473

HAL Id: pastel-00004473

<https://pastel.hal.science/pastel-00004473>

Submitted on 15 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A mes parents

A mon frère

Remerciements

Choisir de faire une thèse est un engagement important dans une vie professionnelle puisqu'il initie le parcours du chercheur. C'est aussi croire que la recherche est chemin une quête vers quelque chose qui nous dépasse et que l'on essaye de comprendre.

Durant cette thèse plusieurs personnes m'ont accompagné et guidé et je tiens tout particulièrement à leur signifier toute ma gratitude :

Dominique Jeulin qui m'a fait découvrir le Centre de Morphologie Mathématique (CMM) lorsque j'étais en master recherche à l'Université Jean Monnet de Saint Etienne. Il a toujours été de bon conseil, sachant diriger mon travail de manière efficace et pleine d'humanité. Il a su me conduire dans des domaines de recherche de haut niveau.

Jesús Angulo a également dirigé mon travail de thèse. Sa disponibilité, ses nombreuses idées et sa gentillesse ont constitué une aide précieuse pour l'accomplissement de ce travail. Il a toujours su m'aider à aller vers des solutions innovantes.

Fernand Meyer qui m'a accueilli dans l'équipe du CMM et qui a toujours su me prodiguer de très bons conseils pour l'avancement de mon travail.

Michel Jourlin que j'ai eu comme professeur à l'école d'ingénieurs CPE-Lyon. C'est lui qui m'a donné le goût du traitement d'images et qui m'a conseillé pour choisir les différentes équipes de recherches internationales avec lesquelles j'ai collaboré.

Fred Hamprecht qui a accepté d'être rapporteur de cette thèse. Je le remercie également pour son accueil lors de ma visite dans son équipe de l'Université d'Heidelberg.

Charles-André Cuenod qui a été le référent dans le domaine médical de ce travail. Il a toujours su se mettre à notre portée pour discuter des points techniques. Catherine Moysan qui a toujours su faire avancer avec brio les difficultés administratives. Sa gentillesse et sa disponibilité ont été une source de bonheur durant ces trois ans. Jean Serra, Beatriz Marcotegui, Etienne Decencière, Michel Bilodeau, Petr Dokladal, Serge Beucher et Jean-Claude Klein qui ont toujours pris le temps de discuter de points techniques. Leurs conseils ont toujours été très utiles.

Je ne saurai oublier tous les autres collègues du CMM qui ont su créer une très bonne ambiance de travail.

Jean Stawiaski avec qui je faisais déjà du traitement d'image à CPE-Lyon. Ses idées, sa gentillesse et son entrain m'ont toujours stimulé.

Mes amis de l'Ecole des Mines qui ont créé une ambiance sympathique en dehors du travail.

Enfin mes amis de Fontainebleau qui par leurs encouragements permanents et leur présence m'ont beaucoup apporté pendant ces trois ans.

Finalement je remercie mes parents et mon frère qui m'ont toujours soutenu et aidé au cours de

mes études tout autant qu'ils ont pu. Je leur en suis très reconnaissant et les en remercie de tout cœur.
Ce mémoire et mon travail de thèse leur sont dédiés.

Résumé

Le traitement d'images hyperspectrales est la généralisation de l'analyse des images couleurs, à trois composantes rouge, vert et bleu, aux images multivariées à plusieurs dizaines ou plusieurs centaines de composantes. Dans un sens général, les images hyperspectrales ne sont pas uniquement acquises dans le domaine des longueurs d'ondes mais correspondent à une description d'un pixel par un ensemble de valeurs : c'est à dire un vecteur. Chacune des composantes d'une image hyperspectrale constitue un canal spectral, et le vecteur associé à chaque pixel est appelé spectre.

Pour valider la généralité de nos méthodes de traitement, nous les avons appliquées à plusieurs types d'imagerie correspondant aux images hyperspectrales les plus variées : des photos avec quelques dizaines de composantes acquises dans le domaine des longueurs d'ondes, des images satellites de télédétection, des séries temporelles d'imagerie par résonance dynamique (DCE-MRI) et des séries temporelles d'imagerie thermique.

Durant cette thèse, nous avons développé une chaîne complète de segmentation automatique des images hyperspectrales par des techniques morphologiques. Pour ce faire, nous avons mis au point une méthode efficace de débruitage spectral, par Analyse Factorielle des Correspondances (AFC), qui permet de conserver les contours spatiaux des objets, ce qui est très utile pour la segmentation morphologique. Puis nous avons fait de la réduction de dimension, par des méthodes d'analyse de données ou par modélisation des spectres, afin d'obtenir une autre représentation de l'image avec un nombre restreint de canaux.

A partir de cette image de plus faible dimension, nous avons effectué une classification (supervisée ou non) pour grouper les pixels en classes spectralement homogènes. Cependant, les classes obtenues n'étant pas homogènes spatialement, i.e. connexes, une étape de segmentation s'est donc avérée nécessaire. Nous avons démontré que la méthode récente de la Ligne de Partage des Eaux Probabiliste était particulièrement adaptée à la segmentation des images hyperspectrales. Elle utilise différentes réalisations de marqueurs aléatoires, conditionnés par la classification spectrale, pour obtenir des réalisations de contours par Ligne de Partage des Eaux (LPE). Ces réalisations de contours permettent d'estimer une fonction de densité de probabilité de contours (pdf) qui est très facile à segmenter par une LPE classique. En définitive, la LPE probabiliste est conditionnée par la classification spectrale et produit donc des segmentations spatio-spectrales dont les contours sont très lisses.

Cette chaîne de traitement a été mise en œuvre sur des séquences d'imagerie par résonance magnétique dynamique (DCE-MRI) et a permis d'établir une méthode automatique d'aide au diagnostic pour la détection de tumeurs cancéreuses.

En outre, d'autres techniques de segmentation spatio-spectrales ont été développées pour les images hyperspectrales : les régions η -bornées et les boules μ -géodésiques. Grâce à l'introduction

d'information régionale, elles améliorent les segmentations par zones quasi-plates qui n'utilisent quant à elles que de l'information locale.

Enfin, nous avons mis au point une méthode très efficace de calcul de toutes les paires de distances géodésiques d'une image, puisqu'elle permet de réduire jusqu'à 50 % le nombre d'opérations par rapport à une approche naïve et jusqu'à 30 % par rapport aux autres méthodes. Le calcul efficace de ce tableau de distances offre des perspectives très prometteuses pour la réduction de dimension spatio-spectrale.

Mots clés :

traitement d'images hyperspectrales, images multivariées, segmentation morphologique, Ligne de Partage des Eaux Probabiliste, réduction de dimension, débruitage, Analyse Factorielle des Correspondances, paires de distances géodésiques, régions η -bornées, boules μ -géodésiques, morphologie mathématique

Abstract

Hyperspectral image processing is the generalization of analysis of color images, having three components red, green and blue, to multivariate images having several tens or hundreds of components. In a general way, hyperspectral images are not only acquired in the domain of wavelengths but correspond to a pixel description by a set of values : i.e. a vector. Each one of the components of an hyperspectral image is a spectral channel, and the vector which is associated to each pixel is called spectrum.

In order to validate the generality of our processing methods, we have applied them to several kinds of imageries corresponding to the most various hyperspectral images : some pictures with a few tens of components acquired in the domain of wavelengths, some satellite images of remote sensing, some temporal series of Dynamic Contrast Enhanced Magnetic Resonance Imagery (DCE-MRI) and some temporal series of thermal imagery.

During this PhD, we have developed a complete chain for automatic segmentation of hyperspectral images by morphological technics. In order to realize it, we have perfected an efficient method of spectral denoising using Factor Correspondence Analysis (FCA). This method preserves spatial contours, which is very useful for morphological segmentation. Then, we have reduced the dimension of the image using data analysis methods, or using spectral modelling, in order to get another representation of the image with a reduced number of channels.

Starting from this image of smaller dimension, we have made a classification (supervised or not) in order to group pixels into homogeneous spectral classes. However, the obtained classes being not spatially homogeneous, i.e. connected, a segmentation stage is necessary. We have demonstrated that the recent method of the Probabilistic Watershed is particularly adapted to the segmentation of hyperspectral images. It is based on different realizations of markers, conditioned by the spectral classification, to get some realizations of contours by Watershed. These contours realizations ensure the estimate of a probability density function (pdf) which is very easy to segment by a standard Watershed. Finally, the probabilistic Watershed is conditioned by the spectral classification. Therefore, it produces spatio-spectral segmentations with very smooth contours.

This treatment chain has been applied on series of DCE-MRI and gives us the possibility to build an automatic method for computer aided detection of cancerous tumors.

Moreover, some other segmentation approaches have been developed for hyperspectral images : η -bounded regions and μ -geodesic balls. Thanks to the introduction of regional information, they improve the segmentations by quasi-flat zones which are only based on local information.

Finally, we have perfected a very efficient method to compute all the pixels pairs of geodesic distances in an image. It reduces until 50 % the number of operations compared to a naïve approach and

until 30 % compared to other methods. The efficient computation of this table of distances opens very promising perspectives for spatio-spectral dimensionality reduction.

Keywords :

hyperspectral image processing, multivariate images, morphological segmentation, Probabilistic Watershed, dimensionality reduction, denoising, Factor Correspondence Analysis, all pairs of geodesic distances, η -bounded regions, μ -geodesic balls, mathematical morphology

Table des matières

1	Introduction	1
1.1	Contexte	1
1.2	Objectifs de la thèse	2
1.3	Applications	2
1.4	Plan et contenu de la thèse	3
2	Filtrage, réduction de dimension et classification	5
2.1	Notations et hypothèses	6
2.2	État de l’art de la réduction de dimension	7
2.2.1	Les méthodes linéaires	9
2.2.2	Analyse en Composantes Principales	9
2.2.3	Analyse Factorielle des Correspondances	12
2.2.4	Analyse en composantes indépendantes	16
2.2.5	Maximum Noise Fraction	17
2.2.6	Poursuite de projection	18
2.2.7	Les méthodes non linéaires	18
2.3	Images hyperspectrales	19
2.4	Filtrage et réduction de dimension	22
2.4.1	Réduction de dimension par AFC	23
2.4.2	Filtrage spectral du bruit par AFC	32
2.4.3	Filtrage spectral du bruit par ACP	35
2.4.4	Composition d’AFC et de reconstitutions	35
2.4.5	Application de la composition d’AFC-reconstitutions	41
2.4.6	Réduction de dimension par modélisation du spectre	46
2.4.7	D’autres méthodes pour la réduction de dimension	48
2.4.8	Conclusion du filtrage et de la réduction de dimension	52
2.5	État de l’art de la classification	52
2.5.1	Kmeans	52
2.5.2	k-medoids	53
2.5.3	Clara	55
2.5.4	LDA	55
2.5.5	D’autres classificateurs	56
2.6	Classification	57

2.6.1	Kmeans	57
2.6.2	Clara	58
2.6.3	Approche modèle	61
2.6.4	Analyse discriminante linéaire (LDA)	63
2.6.5	Classification sur des images similaires	68
2.6.6	Regroupement des canaux	73
2.7	Conclusion du chapitre	75
3	Segmentation hyperspectrale par Ligne de Partage des Eaux	77
3.1	État de l'art en segmentation des images hyperspectrales	78
3.1.1	La Ligne de Partage des Eaux	78
3.1.2	Classification des profils morphologiques	78
3.1.3	D'autres techniques de segmentation hyperspectrales non morphologiques	79
3.2	Comment segmenter une image hyperspectrale avec la LPE ?	79
3.2.1	Introduction à la LPE	80
3.2.2	La segmentation des images hyperspectrales par LPE	81
3.3	Marqueurs provenant d'une classification	82
3.3.1	Le filtrage morphologique de la classification	83
3.3.2	Exemple A : Marqueurs de l'image « bouchon en plastique »	84
3.3.3	Exemple B : Marqueurs de l'image de souris	85
3.3.4	Exemple C : Marqueurs de l'image satellite	86
3.4	Gradients hyperspectraux et segmentation	89
3.4.1	Gradients multivariés	90
3.4.2	Distances entre spectres	91
3.4.3	Choix d'une métrique adaptée à l'espace de représentation de l'image	93
3.4.4	Exemples comparatifs : Application à la segmentation en imagerie thermique	93
3.5	Introduction de la LPE probabiliste	99
3.5.1	Motivations de la création de la LPE probabiliste	101
3.5.2	Quelques notions introductives	102
3.5.3	Segmentation de germes uniformément aléatoires	104
3.5.4	Influence des paramètres N et M	106
3.5.5	Gradient probabiliste	107
3.5.6	Segmentation à partir de germes ponctuels aléatoires régionalisés	107
3.5.7	Considérations sur l'implémentation informatique	109
3.6	Densité de probabilité de contours pour les images hyperspectrales	109
3.6.1	Densité de probabilité de contours marginale $mpdf$	109
3.6.2	Densité de probabilité de contours vectorielle $vpdf$	111
3.6.3	Comparaison entre les densités de probabilités de contours	112
3.7	Conditionnement des germes de la pdf par la classification	116
3.7.1	LPE déterministe	116
3.7.2	LPE probabiliste avec des germes uniformément aléatoires	118
3.7.3	Germes ponctuels aléatoires régionalisés par une classification	119
3.7.4	Germes boules aléatoires régionalisés par une classification	121

3.8	Application à la détection de tumeurs cancéreuses	128
3.9	Méthode générale de segmentation des images hyperspectrales	133
3.10	Autres pistes de travail et perspectives	138
3.10.1	Estimation des densités de probabilités de contours par noyaux spatio-spectraux	138
3.10.2	Morphologie mathématique sur les facteurs pixels	138
3.10.3	Morphologie mathématique multivariée	140
3.11	Conclusion du chapitre	143
4	Géodésie, connexions et images multivariées	145
4.1	Notions introductives	146
4.1.1	La segmentation par critère	146
4.1.2	Les segmentations définies par des connexions	147
4.1.3	Les zones plates et quasi-plates	148
4.1.4	Notions géodésiques	150
4.1.5	Médiane vectorielle	153
4.2	Segmentation par connexions sur des images hyperspectrales	153
4.2.1	Motivations pour élaborer de nouvelles connexions	154
4.2.2	Régions η -bornées	155
4.2.3	Boules μ -géodésiques	157
4.2.4	Résultats de segmentation	160
4.2.5	Conclusion et perspectives sur la segmentation hyperspectrale par connexions	165
4.3	Calcul de toutes les paires de distances géodésiques d'une image	166
4.3.1	Motivations	166
4.3.2	Introduction du problème	166
4.3.3	Méthode spirale	168
4.3.4	Méthode des extrema géodésiques	174
4.3.5	Méthode taux de remplissage	178
4.3.6	Méthode spirale avec répulsion	179
4.3.7	Comparaison des méthodes	182
4.4	Conclusion du chapitre	183
5	Conclusion	185
5.1	Apports de cette thèse	185
5.2	Perspectives	187
A	Annexe A : Filtrage, réduction de dimension et classification	189
A.1	Introduction à l'analyse de données	189
A.2	Analyse factorielle sur un espace euclidien	190
A.2.1	Notations. L'espace euclidien \mathbb{R}^L	191
A.2.2	Choix du sous-espace	192
A.2.3	Inertie expliquée, inertie résiduelle	192
A.2.4	Expression matricielle de l'inertie	193
A.2.5	Recherche des sous-espaces de projection	194

A.2.6	Facteurs principaux	196
A.3	L'Analyse en Composantes Principales	197
A.3.1	Distance entre individus	197
A.3.2	Transformation des données	198
A.3.3	Justification de la transformation des données	198
A.3.4	Analyse du nuage des individus	199
A.3.5	Analyse du nuage des variables	200
A.3.6	Relation de transition entre les facteurs « individus » et les facteurs « variables »	201
A.3.7	Reconstitution des données	201
A.3.8	Choix du nombre d'axes	201
A.3.9	Équivalence entre distance inverse des variances dans l'espace image et distance euclidienne dans l'espace factoriel	203
A.4	L'Analyse Factorielle des Correspondances	204
A.4.1	Le tableau de données	204
A.4.2	Les nuages de points	207
A.4.3	Analyse du nuage des modalités de \mathcal{X} :	210
A.4.4	Analyse du nuage des modalités de \mathcal{Y} :	212
A.4.5	Liens entre les deux analyses :	213
A.4.6	Reconstitution des données :	215
A.4.7	Quelques démonstrations concernant les distances	217
A.5	L'AFC et le bruit	218
A.6	Composition d'AFC et de reconstitutions pour le filtrage spectral	219
A.6.1	Rapport [S]/[B] pour un nombre d'axes variable	219
A.6.2	Comparaison entre une composition d'AFC sur quelques canaux et sur tous les canaux	220
A.7	Influence de la translation des données sur l'AFC	224
A.7.1	Image sans bruit	224
A.7.2	Image bruitée	227
A.8	Covariance et rapports signaux à bruits	229
B	Annexe B : Segmentation morphologique	233
B.1	Notions élémentaires	233
B.1.1	Treillis et morphologie mathématique	233
B.1.2	Propriétés de base des transformations morphologiques	234
B.2	Théorie générale de la segmentation	235
B.2.1	Critère et segmentation	235
B.2.2	Critères connectifs et théorème de la segmentation	237
B.2.3	Demi-treillis des critères connectifs	238
B.3	Connexion	238
B.4	Nivellements	240
B.5	La segmentation morphologique : du SKIZ à la LPE	241
B.5.1	Le squelette par zones d'influence : SKIZ	242
B.5.2	La Ligne de Partage des Eaux (LPE)	243

B.5.3	La LPE contrôlée par marqueurs	248
B.5.4	LPE hiérarchique : algorithme des cascades	252

Bibliographie		255
----------------------	--	------------

Chapitre 1

Introduction

1.1 Contexte

Ce travail se situe dans l'activité du Centre de Morphologie Mathématique de L'Ecole Nationale Supérieure des Mines de Paris. En effet, un travail de recherche développé depuis plusieurs années a conduit à la création d'outils pour la morphologie mathématique couleur. Les différents travaux menés dans l'équipe ont montré que la morphologie mathématique pouvait constituer un moyen puissant d'analyse des images couleurs. Les conclusions tirées de ces images couleurs à trois bandes, ont permis d'envisager l'extension de ces techniques à des images plus complexes, multi bandes. C'est dans ce contexte que se situe cette étude dont l'objectif est d'améliorer et de développer de nouvelles techniques de filtrage, de réduction de données et de segmentation morphologique, adaptées au domaine hyperspectral. Comme le fixe le laboratoire, les travaux sont menés sous un double aspect théorique et applicatif. En effet, l'imagerie est par essence applicative mais nécessite un cadre théorique poussé qui facilite la mise œuvre d'outils mathématiques et justifie de leur pertinence.

Les images hyperspectrales sont des fonctions discrètes, multivariées avec typiquement plusieurs dizaines ou centaines de bandes spectrales. D'une manière générale, chaque pixel d'une image hyperspectrale est un vecteur dont les valeurs correspondent à des longueurs d'ondes, des instants, ou sont associées à un index. Par la suite, nous appellerons « spectre » l'ensemble des valeurs du vecteur pixel.

Dans un sens général, les images hyperspectrales sont des images multivariées ou multi-sources pas uniquement acquises dans le domaine spectral des longueurs d'ondes. Elles peuvent par exemple correspondre à plusieurs modalités en imagerie médicale. La dimension spectrale correspond alors à la modalité acquise. Il peut également s'agir d'images microscopiques acquises après l'utilisation de différents réactifs chimiques pour faire apparaître telle ou telle partie d'un objet. La dimension spectrale est alors le réactif associé. On peut également considérer des séquences temporelles d'imagerie de perfusion. La dimension spectrale correspond dans ce cas au temps. En imagerie satellite, pour la télédétection, les canaux sont acquis à différentes longueurs d'onde.

De manière plus formelle, pour chaque longueur d'onde, instant ou index correspond une image appelée canal. Nous parlerons, dans la suite de ce mémoire, de canal spectral pour décrire la dimension correspondant aux canaux. Le nombre de canaux dépend de la nature de l'image considérée : images satellites, images microscopiques, images médicales, séries temporelles.

On peut se demander à quoi servent les nombreux canaux des images hyperspectrales ? Les canaux permettent par exemple de discriminer un constituant chimique d'un autre à l'intérieur d'une image de télédétection, à suivre de manière précise une cinétique d'un produit de contraste en imagerie de perfusion. Comme les spectres des différents constituants peuvent parfois être très proches, il est nécessaire d'avoir un échantillonnage spectral très fin si on espère pouvoir différencier deux constituants proches spectralement.

1.2 Objectifs de la thèse

Comme présenté, cette thèse est la poursuite du travail déjà engagé sur des images couleurs à trois bandes à des images hyperspectrales de plus grandes dimensions : i.e. jusqu'à quelques centaines ou quelques milliers de canaux. C'est donc l'initiation d'un nouvel axe de recherche du Centre de Morphologie Mathématique : le développement de la Morphologie Mathématique Hyperspectrale.

En conséquence, l'objectif de cette thèse est la conception et le développement de nouvelles méthodes et algorithmes morphologiques de segmentation performants qui soient adaptés aux images hyperspectrales. Il s'agit donc de construire un corpus méthodologique de la segmentation morphologique des images hyperspectrales. Ce travail n'est donc pas axé sur la construction d'opérateurs morphologiques hyperspectraux (i.e. érosion, dilatations, ...) qui nécessitent l'établissement d'une relation d'ordre entre les pixels-vecteurs des images.

De manière plus précise les objectifs assignés sont les suivants :

- exploration de techniques de réduction des données hyperspectrales adaptées aux traitements spatiaux ultérieurs à réaliser. Le but est d'extraire uniquement l'information utile pour la segmentation afin de réduire les temps de calculs ;
- évaluation de ces techniques par des méthodes de classification non supervisées ou supervisées. Les techniques supervisées permettront l'introduction de connaissances expertes par l'utilisation de données d'apprentissage ;
- filtrage des images hyperspectrales afin de réduire leur bruit et de conserver les propriétés de formes des objets ;
- segmentation afin d'extraire les objets d'intérêts. A partir de la segmentation on pourra détecter des zones d'intérêt ou les classifier selon un critère de ressemblance à une base de données de référence. Pour ce faire on étendra les techniques de segmentation morphologiques aux images multivariées. On essaiera de combiner l'information spatiale avec l'information spectrale ;
- développer des algorithmes associés efficaces pour une chaîne de traitement d'images performante des images hyperspectrales.

Ainsi ce travail de thèse est à la fois méthodologique puisqu'il s'agit d'établir les bases de la segmentation morphologique et applicatif, puisqu'il faut démontrer l'efficacité des algorithmes pour différents types d'applications.

1.3 Applications

Le traitement de l'image nécessite de présenter les résultats méthodologiques sur plusieurs applications afin de justifier de leur pertinence. C'est la raison pour laquelle, nous n'avons pas voulu nous

limiter aux images de télédétection mais élargir le champs applicatif à d'autres types d'imagerie.

Pour chaque application la technologie d'acquisition et la finalité des utilisateurs changent, cependant la chaîne de traitement reste la même, seules les spécifications applicatives finales varient.

Cette thèse sera illustrée à partir : d'images prises dans le domaine du visible, de séquences temporelles d'imagerie de perfusion par résonance magnétique (IRM), d'images satellites, de séries temporelles acquises par imagerie thermique.

L'image prise dans le domaine du visible dont les contours sont bien définis permettra de vérifier la qualité de la réduction de dimension et de la segmentation. Les autres images permettront de tester nos algorithmes sur des cas réels plus difficiles à traiter et de vérifier les résultats auprès d'experts du domaine.

Nous avons ainsi eu une collaboration très suivie avec l'équipe du Professeur *C.-A. Cuenod* du service imagerie de l'Hôpital Européen Georges Pompidou à Paris. Les images fournies sont des séquences d'imagerie par résonance magnétique dynamique (DCE-MRI) de petit animal. L'objectif consistait à détecter des tumeurs cancéreuses présentes sur ces images ce qui permettra ensuite aux médecins de mesurer l'efficacité de produits anti-angiogéniques pour empêcher la vascularisation des tumeurs. Des références pour la détection de tumeurs nous ont ainsi été fournies ainsi que des segmentations anatomiques manuelles qui ont été très utiles pour la validation des résultats.

Une autre collaboration a été établie avec l'équipe du Pr. *G. Flouzat* du Laboratoire de Télédétection à Haute Résolution, de l'Université Paul Sabatier à Toulouse. Des images simulées de télédétection PLEIADES nous ont en effet été fournies grâce à la collaboration de son équipe au programme ORFEO du Centre National d'Études Spatiales (CNES).

1.4 Plan et contenu de la thèse

Les images hyperspectrales contiennent jusqu'à plusieurs centaines ou milliers de canaux. Il est donc nécessaire de compresser l'information spectrale afin d'éliminer la redondance entre les canaux. Pour cela on cherche à réduire le volume de données tout en conservant l'information utile. Plusieurs méthodes ont été testées : des approches par analyse de données et des approches par modélisation du spectre. Cette réduction de dimension permet d'améliorer la précision de la classification et de réduire les temps de calcul. L'analyse de données permet également de filtrer le bruit spectral de l'image en conservant l'information spatiale ce qui est très utile pour effectuer des traitements morphologiques.

Dans l'espace de dimension réduite on effectue une classification spectrale des données non supervisée et supervisée. La classification supervisée repose sur une connaissance experte a priori qui permet d'entraîner le classificateur, ce qui permet d'améliorer la qualité des résultats. Les pixels vecteurs sont ainsi répartis en classes ayant des spectres similaires. Une première répartition des spectres en classes est ainsi obtenue.

Dans le chapitre 2 sont ainsi présentés, la réduction de dimension, le filtrage spectral et la classification sur des images hyperspectrales.

Les classes de la classification obtenues sont homogènes dans le domaine spectral mais pas dans le domaine spatial, c'est à dire qu'elles ne sont pas connexes. Afin d'introduire la dimension spatiale dans le traitement, la segmentation morphologique des images hyperspectrales est présentée dans le chapitre 3. Pour cela des gradients sont définis pour des images hyperspectrales. Ces gradients peuvent

être déterministes ou probabilistes et doivent être adaptés aux espaces de représentation des images hyperspectrales. Ils sont ensuite utilisés comme fonction à inonder pour la ligne de Partage des Eaux. Il est à noter que les gradients probabilistes font partie d'une nouvelle méthode de segmentation : la Ligne de Partage des Eaux (LPE) Probabiliste créée pour les images couleurs par *J. Angulo et D. Jeulin* [4] et étendue aux images hyperspectrales dans cette thèse par *G. Noyel, J. Angulo et D. Jeulin* [137].

Le chapitre 4 sera axé sur la géodésie et l'établissement de nouvelles méthodes de segmentation par connexions, pour améliorer les zones quasi-plates grâce à l'introduction d'information régionale. Ces méthodes constituent une autre partie de la segmentation morphologique que la Ligne de Partage des Eaux. Elles permettent également une segmentation spatio-spectrale. En outre nous avons également conçu une méthode de calcul efficace de tableaux de distances géodésiques qui permet de réduire de 50 % le nombre de calculs par rapport à une approche naïve.

Chapitre 2

Filtrage, réduction de dimension et classification

Les images hyperspectrales peuvent contenir jusqu'à plusieurs centaines ou milliers de canaux. La réduction de la dimension spectrale est possible car les canaux ne sont pas statistiquement indépendants les uns des autres. En outre, cette réduction est nécessaire pour lutter contre le « phénomène de Hughes » [81] encore appelé « malédiction de la dimension ».

Le phénomène de Hughes met en évidence deux aspects avec l'accroissement de la dimension (i.e. du nombre de canaux) pour une image hyperspectrale :

- La séparabilité augmente avec la dimension de l'espace : i.e on peut plus facilement séparer les spectres lorsque le nombre de bandes augmente.
- La précision de l'estimation statistique diminue avec la dimension de l'espace. En effet plus on augmente le nombre de bandes, plus l'espace est vide. L'estimation statistique est alors mauvaise.

Hughes montre qu'il existe un optimum pour la précision en fonction de la complexité de la mesure c'est à dire du nombre de canaux. Ce phénomène a notamment été étudié dans le cadre des images hyperspectrales dans [96, 104].

La réduction de la dimension permet également de réduire la quantité de données manipulées et donc d'améliorer les temps de calcul.

Dans cette partie, après une définition des notations, nous ferons un état de l'art de différentes techniques de réduction de données pour la dimension spectrale, puis nous présenterons les méthodes que nous avons utilisées : l'Analyse Factorielle des Correspondances (AFC), l'Analyse en Composantes Principales (ACP) et les approches modèles. Nous verrons que les approches par analyse de données permettent, de réduire non seulement la dimension spectrale, mais aussi le bruit des spectres de l'image. Enfin nous étudierons différents types de classifications spectrales non supervisées et semi-supervisées.

2.1 Notations et hypothèses

Dans cette partie, nous introduisons les notations et les hypothèses qui seront utilisées dans tout ce mémoire.

Une image numérique est un ensemble de points répartis sur une grille régulière appelée le bitmap. Ces points sont appelés des pixels. Ces pixels sont à valeurs dans $\mathcal{T} \subset \mathbb{R}$ pour des images à niveaux de gris (eq. 2.1.1).

$$f : \begin{cases} E & \rightarrow \mathcal{T} \\ x & \rightarrow f(x) \end{cases} \quad (2.1.1)$$

avec $E \subset \mathbb{R}^2$ pour une image bidimensionnelle (2D) ou $E \subset \mathbb{R}^3$ pour une image tridimensionnelle (3D).

Pour des images couleurs à trois canaux, Rouge Vert et Bleu (RVB), les pixels sont à valeurs dans $\mathcal{T} \times \mathcal{T} \times \mathcal{T} \subset \mathbb{R}^3$ (avec $\mathcal{T} \times \mathcal{T} \times \mathcal{T} = \mathcal{T}^3$). Chaque pixel est ainsi représenté par 3 valeurs (eq. 2.1.2) :

$$\mathbf{f} : \begin{cases} E & \rightarrow \mathcal{T}^3 \\ x & \rightarrow \mathbf{f}(x) = (f_r(x), f_v(x), f_b(x)) \end{cases} \quad (2.1.2)$$

avec $f_r(x)$ la valeur du rouge, $f_v(x)$ la valeur du vert et $f_b(x)$ la valeur du bleu pour le pixel x .

Les images hyperspectrales dans un sens général sont des images pour lesquelles chaque pixel contient plusieurs valeurs : i.e un vecteur de valeurs. L'intérêt de ces images est d'avoir une information plus fine pour chaque pixel (eq. 2.1.3).

$$\mathbf{f}_\lambda : \begin{cases} E & \rightarrow \mathcal{T}^L \\ x & \rightarrow \mathbf{f}_\lambda(x) = (f_{\lambda_1}(x), f_{\lambda_2}(x), \dots, f_{\lambda_L}(x)) \end{cases} \quad (2.1.3)$$

est une image hyperspectrale, avec :

- $E \subset \mathbb{R}^2$, $\mathcal{T} \subset \mathbb{R}$ et $\mathcal{T}^L = \mathcal{T} \times \mathcal{T} \times \dots \times \mathcal{T}$
- $x = x_i \setminus i \in \{1, 2, \dots, P\}$ est la coordonnée spatiale du pixel vecteur $\mathbf{f}_\lambda(x_i)$ (P est le nombre de pixels de E)
- $f_{\lambda_j} \setminus j \in \{1, 2, \dots, L\}$ est un canal (L est le nombre de canaux)
- $f_{\lambda_j}(x_i)$ est la valeur du pixel vecteur $\mathbf{f}_\lambda(x_i)$ sur le canal f_{λ_j} .

Sur la figure 2.1 sont représentés un canal f_{λ_j} et pixel vecteur $\mathbf{f}_\lambda(x_i)$ sur une image hyperspectrale en longueurs d'ondes.

Les images hyperspectrales sont ainsi des fonctions discrètes, multivariées avec typiquement plusieurs dizaines ou centaines de bandes spectrales. D'une manière générale, chaque pixel d'une image hyperspectrale est un vecteur dont les valeurs d'intensité correspondent à des longueurs d'ondes, des instants, ou sont associées à un index j . Nous appellerons « spectre » l'ensemble des valeurs du vecteur pixel.

Pour chaque longueur d'onde, instant index correspond une image appelée canal. Dans cette thèse, pour décrire la dimension correspondant aux canaux nous parlerons de dimension spectrale.

Le nombre de canaux dépend de la nature de l'image considérée : image satellite, image microscopique, image médicale, série temporelle. Les images hyperspectrales existent ainsi en imagerie

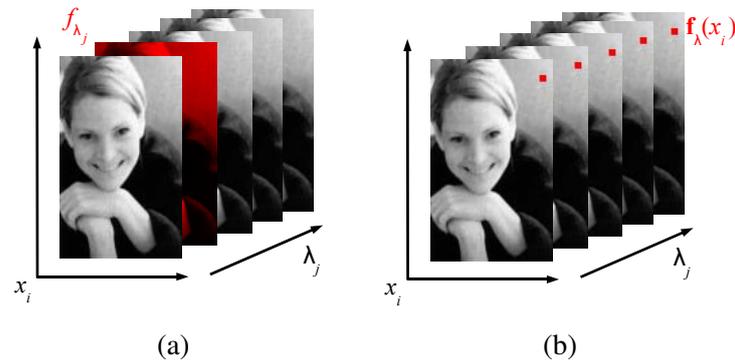


FIG. 2.1 – (a) Canal f_{λ_j} et (b) pixel vecteur $\mathbf{f}_{\lambda}(x_i)$ sur une image hyperspectrale en longueurs d'ondes ($152 \times 91 \times 61$ pixels) à 61 canaux acquis entre 400 nm et 700 nm avec un pas de 5 nm. (Source de l'image hyperspectrale : Spectral Database, University of Joensuu Color Group, <http://spectral.joensuu.fi/>).

satellite, ou aéroportée, et peuvent être issues de satellites comme « SPOT 5 », en imagerie médicale sur des séquences d'Imagerie par Résonance Magnétique (IRM). Il existe également des séquences temporelles où l'on a des acquisitions à plusieurs instants. Pour des canaux correspondant à un index j , on peut avoir l'image d'un constituant soumis successivement à plusieurs réactifs chimiques [65].

En outre, on effectue une distinction entre imagerie multispectrale et imagerie hyperspectrale : la première se contente d'une dizaine de canaux, $L \simeq 10$, tandis que la seconde dépasse la centaine de canaux, $L \geq 100$.

Deux hypothèses principales sont à la base de ce document.

Hypothèse 1. Cohérence spectrale On suppose que les canaux sont cohérents spectralement, c'est à dire que le même objet a été acquis sur chaque canal spectral.

Ainsi une image à niveaux de gris d'une pomme correspondant à un premier canal puis une image à niveaux de gris d'un cheval correspondant à un deuxième canal, ne constituent pas une image hyperspectrale.

Hypothèse 2. Recalage On suppose, que les canaux sont recalés, c'est à dire que l'information portée par chaque composante d'un pixel vecteur $\mathbf{f}_{\lambda}(x_i)$ correspond à la même position spatiale x_i .

L'hypothèse 2 implique la propriété suivante :

Propriété 2.1.1. Même résolution spatiale On suppose que les canaux ont tous la même résolution spatiale.

Un vecteur pixel $\mathbf{f}_{\lambda}(x_i)$ est ainsi la réponse d'un même point pour différentes valeur de la dimension spectrale λ_j .

2.2 État de l'art de la réduction de dimension

De nombreuses méthodes linéaires et non linéaires de réduction de données existent. La réduction de données pour la dimension spectrale, consiste à appliquer une transformation ζ sur l'espace image

à L dimensions (i.e. canaux) pour obtenir un autre espace de représentation à K dimensions avec $K < L$ et souvent $K \ll L$. L'image hyperspectrale \mathbf{f}_λ représentée dans cet espace de dimension réduite est notée $\mathbf{c}_\alpha^{\mathbf{f}}$.

$$\zeta : \begin{cases} \mathcal{T}^L & \rightarrow \mathcal{T}^K / K < L \\ \mathbf{f}_\lambda(x) & \rightarrow \mathbf{c}_\alpha^{\mathbf{f}}(x) = (c_{\alpha_1}^{\mathbf{f}}(x), \dots, c_{\alpha_K}^{\mathbf{f}}(x)) \end{cases} \quad (2.2.1)$$

Notons \mathbf{F} la matrice $P \times L$ de représentation de l'image \mathbf{f}_λ . \mathbf{F} est composée de P lignes représentant les pixels vecteurs (i.e. les individus) et de L colonnes représentant les canaux (i.e. les variables) (eq. 2.2.6). En termes mathématiques, la dimension intrinsèque signifie que les individus dans le jeu de données \mathbf{F} sont situés, ou presque situés, dans un sous espace de dimension K inclus dans l'espace de dimension L .

Pour alléger la notation, on s'autorise pour les calculs matriciels à utiliser la notation :

$$f_{ij} = f_{\lambda_j}(x_i) \quad f_j = f_{\lambda_j} \quad f_i = \mathbf{f}_\lambda(x_i) \quad (2.2.2)$$

$$\mathbf{F} = \begin{bmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1L} \\ \vdots & & \vdots & & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{iL} \\ \vdots & & \vdots & & \vdots \\ f_{P1} & \cdots & f_{Pj} & \cdots & f_{PL} \end{bmatrix} \quad (2.2.3)$$

Les vecteurs-lignes (i.e. vecteurs pixels) de \mathbf{F} sont les éléments de \mathcal{T}^L et sont assimilés à des vecteurs de \mathbb{R}^L :

$$f_i = (f_{i1}, \dots, f_{ij}, \dots, f_{iL}) \quad (2.2.4)$$

avec $f_i = \mathbf{f}_\lambda(x_i)$.

Les vecteurs-colonnes (i.e. canaux) de \mathbf{F} sont les éléments de \mathbb{R}^P et sont assimilés à des vecteurs de \mathbb{R}^P :

$$f_j = (f_{1j}, \dots, f_{ij}, \dots, f_{Pj}) \quad (2.2.5)$$

avec $f_j = f_{\lambda_j}$.

On définit de la même manière la matrice des données transformées \mathbf{C} :

$$\mathbf{C} = \begin{bmatrix} c_{11} & \cdots & c_{1\alpha} & \cdots & c_{1K} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & c_{i\alpha} & \cdots & c_{iK} \\ \vdots & & \vdots & & \vdots \\ c_{P1} & \cdots & c_{P\alpha} & \cdots & c_{PK} \end{bmatrix} \quad (2.2.6)$$

avec : $c_{i\alpha} = c_{\alpha_k}(x_i)$. On note parfois $c_{i\alpha} = c_i^\alpha$.

Voici quelques références qui recensent et qui comparent différentes méthodes de réduction de données [106, 184, 158].

2.2.1 Les méthodes linéaires

Les méthodes linéaires sont très utilisés en réduction de données. Dans une étude comparative approfondie entre des méthodes linéaires et non linéaires *Maaten et al.* [184] montre que les méthodes non linéaires ne surpassent pas les méthodes linéaires sur des données réelles, c'est à dire non simulées. En outre elles sont moins coûteuses en temps de calcul.

Dans les annexes, l'Analyse en Composantes Principales (ACP) (annexes A.3, p : 197 et suivantes) et l'Analyse Factorielle des Correspondances (AFC) (annexes A.4, p : 204 et suivantes) sont expliquées plus en détail. On montre ainsi que l'ACP et l'AFC se déduisent de l'Analyse Factorielle sur un espace euclidien grâce à des poids affectées à chacun des points de la matrice \mathbf{F} qui correspondent au passage de la distance « inverse des variances » en ACP, ou de la métrique du χ^2 en AFC, à la métrique euclidienne.

2.2.2 Analyse en Composantes Principales

L'Analyse en Composantes Principales (ACP) [79] construit une représentation de l'image hyperspectrale f_λ de plus faible dimension $K < L$ qui représente au mieux la variance des canaux. Ceci s'effectue par la recherche d'une transformation linéaire qui conduise à une base de dimension réduite dans lequel la variance des données (variables) est maximale [184].

2.2.2.1 Présentation de la méthode

La motivation qui pousse à appliquer l'ACP repose souvent sur l'hypothèse que les directions de forte variance contiennent plus d'information que les directions de faible variance. La raison derrière cette affirmation est que le bruit est souvent supposé uniformément réparti. Donc, les directions de forte variance ont un rapport signal à bruit plus élevé [14].

On note $\mathbf{X} = \mathbf{F} - \bar{\mathbf{F}}$ la matrice des variables centrées, i.e. on remplace les colonnes f_j de \mathbf{F} par $f_j - m_j$, avec m_j la moyenne de f_j .

L'ACP recherche une application linéaire \mathbf{U} , qui transforme les données \mathbf{X} en une matrice \mathbf{C} dont les colonnes (les variables) sont décorrélées :

$$\mathbf{C} = \mathbf{X}\mathbf{U} \quad (2.2.7)$$

Les matrice \mathbf{X} et \mathbf{C} peuvent être caractérisées par leurs moments d'ordre 1, $E[\mathbf{X}] = \bar{\mathbf{X}} = 0$, et $E[\mathbf{C}] = \bar{\mathbf{C}} = 0$, et par leurs moments d'ordres 2 regroupés dans leurs matrices de covariances, $\Sigma_{\mathbf{X}}$, $\Sigma_{\mathbf{C}}$. La matrice de covariance de \mathbf{C} , $\Sigma_{\mathbf{C}}$, peut se décomposer de la façon suivante :

$$\begin{aligned} \Sigma_{\mathbf{C}} &= E[(\mathbf{C} - \bar{\mathbf{C}})(\mathbf{C} - \bar{\mathbf{C}})^t] \\ &= E[\mathbf{C}\mathbf{C}^t] \\ &= E[(\mathbf{X}\mathbf{U})(\mathbf{X}\mathbf{U})^t] \\ &= \mathbf{U}^t E[\mathbf{X}\mathbf{X}^t] \mathbf{U} \\ &= \mathbf{U}^t \Sigma_{\mathbf{X}} \mathbf{U} \end{aligned}$$

En terme mathématique, l'ACP tente de trouver une application linéaire \mathbf{U} qui maximise $\mathbf{U}^t \Sigma_{\mathbf{X}} \mathbf{U}$. C'est à dire que si les composantes (colonnes) de \mathbf{C} sont décorréllées, la matrice $\Sigma_{\mathbf{C}}$ doit être diagonale. L'application linéaire \mathbf{U} est donc formée des K vecteurs propres de la matrice de covariance des données centrées $\Sigma_{\mathbf{X}}$. Par conséquent l'ACP résout le problème de recherche des K principales valeurs propres et des vecteurs propres :

$$\Sigma_{\mathbf{X}} \mathbf{U} = \mu \mathbf{U} \quad (2.2.8)$$

La représentation en faible dimension des pixels vecteurs $\mathbf{f}_\lambda(x_i)$ est obtenue en les projetant sur la base linéaire engendrée par les colonnes de \mathbf{U} : $\mathbf{C} = \mathbf{X}\mathbf{U}$. Les vecteurs colonnes de \mathbf{C} sont appelés les facteurs principaux (pour plus de détails cf. annexes A.3, p : 197 et suivantes).

Remarque 2.2.1. Si chacune des colonnes f_j de \mathbf{F} est centrée et réduite avec $\mathbf{X} = \frac{\mathbf{F} - \bar{\mathbf{F}}}{\sigma_{\mathbf{F}}}$ (i.e. $x_i^j = \frac{f_j - m_j}{\sigma_j}$, avec σ_j la variance de f_j) au lieu d'être centrée, les vecteurs propres de la matrice de corrélation sont alors recherchés : $cor_{\mathbf{X}} \mathbf{U} = \mu \mathbf{U}$. On préfère souvent centrer et réduire les variables afin de les normaliser par rapport à leur écart type. En outre, les propriétés mathématiques sont plus concises avec des variables centrées réduites. Afin de faire le lien avec la partie A.3 des annexes (p : 197 et suivantes), \mathbf{U} est la matrice des vecteurs propres u_α de la matrice d'inertie $\mathbf{V} = cor_{\mathbf{X}}$:

$$\mathbf{U} = \begin{bmatrix} u_1^1 & \cdots & u_1^\alpha & \cdots & u_1^K \\ \vdots & & \vdots & & \vdots \\ u_j^1 & \cdots & u_j^\alpha & \cdots & u_j^K \\ \vdots & & \vdots & & \vdots \\ u_L^1 & \cdots & u_L^\alpha & \cdots & u_L^K \end{bmatrix} \quad (2.2.9)$$

avec $K = L$ si la matrice $\Sigma_{\mathbf{X}}$, ou $cor_{\mathbf{X}}$, est de rang L (ce qui est en général le cas) et si l'on conserve tous les axes factoriels. Quant aux vecteurs u_α , ils engendrent les axes principaux d'inertie.

Remarque 2.2.2. Lorsque l'on utilise des variables centrées réduites, la métrique entre les individus (i.e. les pixels vecteurs), sous-jacente à l'ACP, est la métrique inverse des variances :

$$d_{1/\sigma^2}^2(f_i, f_{i'}) = \sum_{j=1}^L \frac{1}{\sigma_j^2} (f_{ij} - f_{i'j})^2 = \sum_{j=1}^L \left(\frac{f_{ij} - m_j}{\sigma_j} - \frac{f_{i'j} - m_j}{\sigma_j} \right)^2 \quad (2.2.10)$$

Les poids associées aux individus sont tous égaux à $1/P$ avec P le nombre d'individus.

2.2.2.2 Reconstitution des données

Il est possible d'obtenir une reconstitution approchée des données à partir d'un nombre limité de facteurs :

$$\frac{\widehat{f_j - m_j}}{\sigma_j} = \widehat{x^j} = \sum_{\alpha=1}^L c_i^\alpha u_\alpha^j = \sum_{\alpha=1}^L \mu_\alpha^{-1/2} c_i^\alpha d_\alpha^j \quad (2.2.11)$$

avec d_{α}^j les facteurs associés aux variables (cf. annexes A.3, p : 197 et suivantes). En effet, on effectue simultanément une analyse sur les variables et les individus.

Les vecteurs u_{α} sont orthogonaux deux à deux dans \mathbb{R}^L . Par conséquent \mathbf{U} est une matrice orthogonale, donc $\mathbf{U}^t = \mathbf{U}^{-1}$. On peut donc écrire la reconstitution des données comme :

$$\widehat{\mathbf{X}} = \mathbf{C}\mathbf{U}^t \quad (2.2.12)$$

L'ACP est donc une transformation ζ de l'espace image de départ \mathbf{f}_{λ} à L dimensions en un espace $\mathbf{c}_{\alpha}^{\mathbf{f}}$ de dimension réduite $K \leq L$ ainsi qu'un ensemble de paramètres :

$$\zeta : \begin{cases} \mathcal{T}^L & \rightarrow \mathcal{T}^K / K \leq L \\ \mathbf{f}_{\lambda}(x) & \rightarrow \begin{pmatrix} \mathbf{c}_{\alpha}^{\mathbf{f}}(x) = (c_{\alpha_1}^{\mathbf{f}}(x), \dots, c_{\alpha_K}^{\mathbf{f}}(x)) \\ \mathbf{d}_{\alpha}^{\mathbf{f}} = (d_{\alpha_1}^{\mathbf{f}}, \dots, d_{\alpha_K}^{\mathbf{f}}) \\ \{\mu_{\alpha}\}_{\alpha=1\dots K} \\ \{m_j\}_{j=1\dots L} \\ \{\sigma_j\}_{j=1\dots L} \end{pmatrix} \end{cases} \quad (2.2.13)$$

La reconstitution est une transformation pseudo-inverse $\widehat{\zeta}^{-1}$, ou une transformation inverse lorsque tous les axes sont conservés $K = L$:

$$\widehat{\zeta}^{-1} : \begin{cases} \mathcal{T}^K & \rightarrow \mathcal{T}^L / K < L \\ \begin{pmatrix} \mathbf{c}_{\alpha}^{\mathbf{f}}(x) \\ \mathbf{d}_{\alpha}^{\mathbf{f}} = (d_{\alpha_1}^{\mathbf{f}}, \dots, d_{\alpha_K}^{\mathbf{f}}) \\ \{\mu_{\alpha}\}_{\alpha=1\dots K} \\ \{m_j\}_{j=1\dots L} \\ \{\sigma_j\}_{j=1\dots L} \end{pmatrix} & \rightarrow \widehat{\mathbf{f}}_{\lambda}(x) = (\widehat{f}_{\lambda_1}(x), \dots, \widehat{f}_{\lambda_L}(x)) \end{cases} \quad (2.2.14)$$

2.2.2.3 Filtrage du bruit

Si le bruit est gaussien, indépendant du signal et de variance plus faible que le signal, il est possible de séparer le bruit en appliquant un seuil sur les valeurs propres.

L'image bruitée est notée $\mathbf{f}_{\lambda}^{br}$. On la bruite par un bruit blanc gaussien centré et d'écart type σ : $n_{\lambda_j} = \mathcal{N}(0, \sigma)$:

$$f_{\lambda_j}^{br}(x_i) = f_{\lambda_j}(x_i) + n$$

En notation matricielle, on a : $\mathbf{F}^{br} = \mathbf{F} + \mathbf{N}$. Comme le bruit est déjà centré ($\overline{\mathbf{N}} = 0$), la matrice des variables centrées s'écrit : $\mathbf{X}^{br} = \mathbf{F}^{br} - \overline{\mathbf{F}^{br}} = \mathbf{F} + \mathbf{N} - \overline{\mathbf{F}} = \mathbf{X} + \mathbf{N}$, avec $\mathbf{X} = \mathbf{F} - \overline{\mathbf{F}}$. La matrice de covariance $\Sigma_{\mathbf{X}^{br}}$ se note :

$$\begin{aligned} \Sigma_{\mathbf{X}^{br}} &= E[\mathbf{X}^{br}(\mathbf{X}^{br})^t] \\ &= E[(\overline{\mathbf{X}} + \mathbf{N})(\overline{\mathbf{X}} + \mathbf{N})^t] \\ &= E[\overline{\mathbf{X}}\overline{\mathbf{X}}^t + \overline{\mathbf{X}}\mathbf{N}^t + \mathbf{N}\overline{\mathbf{X}}^t + \mathbf{N}\mathbf{N}^t] \\ &= E[\overline{\mathbf{X}}\overline{\mathbf{X}}^t] + E[\mathbf{N}\mathbf{N}^t] \\ &= \Sigma_{\mathbf{X}} + \Sigma_{\mathbf{N}} \end{aligned}$$

En effet comme le bruit est décorrélié du signal, on a : $E[\overline{\mathbf{X}}\mathbf{N}^t] = 0$ et $E[\overline{\mathbf{X}}^t\mathbf{N}] = 0$.

Comme le bruit est gaussien, de même variance et que ses réalisations sont indépendantes les unes des autres, sa matrice de covariance est diagonale et s'écrit $\Sigma_{\mathbf{N}} = \sigma^2\mathbf{I}$, avec \mathbf{I} la matrice identité. La recherche des valeurs propres de la matrice de covariance $\Sigma_{\mathbf{X}^{br}}$ donne donc :

$$\begin{aligned}\Sigma_{\mathbf{X}^{br}}u_{\alpha}^{br} &= \mu_{\alpha}^{br}u_{\alpha}^{br} \\ \Leftrightarrow (\Sigma_{\mathbf{X}} + \sigma^2\mathbf{I})u_{\alpha}^{br} &= \mu_{\alpha}^{br}u_{\alpha}^{br} \\ \Leftrightarrow \Sigma_{\mathbf{X}}u_{\alpha}^{br} &= (\mu_{\alpha}^{br} - \sigma^2)u_{\alpha}^{br}\end{aligned}$$

Or les valeurs propres de $\Sigma_{\mathbf{X}}$ s'écrivent $\Sigma_{\mathbf{X}}u_{\alpha} = \mu_{\alpha}u_{\alpha}$. Par identification avec l'équation 2.2.15 les vecteurs propres sont conservés $u_{\alpha}^{br} = u_{\alpha}$ alors que les valeurs propres sont modifiées de la manière suivante $\mu_{\alpha}^{br} = \mu_{\alpha} - \sigma^2$. La matrice de covariance $\Sigma_{\mathbf{X}}$ s'écrit donc :

$$\Sigma_{\mathbf{X}}u_{\alpha} = (\mu_{\alpha} - \sigma^2)u_{\alpha} \quad (2.2.15)$$

Comme $\mu_{\alpha} = \text{var}(c_{\alpha}^f)$ (cf. annexes A.3.4, p : 199), c'est à dire la variance du facteur pixel c_{α}^f , et $\mu_{\alpha}^{br} = \mu_{\alpha} - \sigma^2$, les facteurs pixels sans bruit sont ceux pour lesquels $\mu_{\alpha} > \sigma^2$. Les axes étant rangés par inertie décroissante, les facteurs pixels qui contiennent le moins de bruit sont donc les facteurs pixels des premiers axes. On peut donc séparer le bruit par seuillage sur l'inertie des axes.

2.2.3 Analyse Factorielle des Correspondances

L'Analyse Factorielle des Correspondances (AFC) a été introduite, développée puis publiée par *J.P. Benzécri* en 1973 [21]. C'est une méthode similaire à l'ACP, la métrique sous-jacente à l'AFC étant la métrique du chi-deux χ^2 (cf. annexes partie A.4 pour plus de détails).

2.2.3.1 Le tableau de données

On considère la matrice \mathbf{F} associée à l'image hyperspectrale \mathbf{f}_{λ} comme un tableau de contingence (ou tableau croisé) ventilant une population d'individus $\Omega = \{1 \dots \omega \dots n\}$, avec $n = P \times L$, selon deux variables \mathcal{X} et \mathcal{Y} . L'ensemble des modalités de \mathcal{X} est noté $I = \{1 \dots i \dots P\}$, l'ensemble des modalités de \mathcal{Y} est noté $J = \{1 \dots j \dots L\}$.

Au lieu de centrer et réduire les variables f_j comme en ACP, on va appliquer une autre normalisation.

Pour ce faire, le tableau de fréquence $\Phi = \nu_{\mathbf{IJ}}$ se déduit du tableau \mathbf{F} en divisant chaque effectif f_{ij} par l'effectif total $f = \sum_{i=1}^P \sum_{j=1}^L f_{ij}$.

$$\nu_{\mathbf{IJ}} = \Phi = \begin{bmatrix} \nu_{11} & \cdots & \nu_{1j} & \cdots & \nu_{1L} \\ \vdots & & \vdots & & \vdots \\ \nu_{i1} & \cdots & \nu_{ij} & \cdots & \nu_{iL} \\ \vdots & & \vdots & & \vdots \\ \nu_{P1} & \cdots & \nu_{Pj} & \cdots & \nu_{PL} \end{bmatrix} \nu_{i.}, \quad \text{où } \nu_{ij} = \frac{f_{ij}}{f} \quad (2.2.16)$$

Remarque 2.2.3. Les valeurs f_{ij} en AFC doivent être positives car l'AFC repose sur le calcul de fréquences et de probabilités.

On note les fréquences marginales :

$$\nu_{i.} = \sum_{j=1}^L \nu_{ij} \quad , \quad \nu_{.j} = \sum_{i=1}^P \nu_{ij} \quad (2.2.17)$$

On note les lois marginales du tableau :

$$\nu_I = \{\nu_{i.} | i \in I\} \quad , \quad \nu_J = \{\nu_{.j} | j \in J\} \quad (2.2.18)$$

$\nu_I = \{\nu_{i.} | i \in I\}$ est la loi de probabilité marginale ou distribution marginale de \mathcal{X} définie sur I .

$\nu_J = \{\nu_{.j} | j \in J\}$ est la loi de probabilité marginale ou distribution marginale de \mathcal{Y} définie sur J .

Par division de chaque ligne $\{\nu_{ij} | j \in J\}$ du tableau de fréquence par sa masse totale $\nu_{i.}$ (total de la ligne) on définit le profil de la ligne i :

$$\nu_J^i = \{\nu_{.j}^i | j \in J\} = \left\{ \frac{\nu_{ij}}{\nu_{i.}} | j \in J \right\} \quad (2.2.19)$$

Ce profil ν_J^i est appelé distribution conditionnelle, de j pour i donné ou encore i -ème profil ligne.

De même, par division de chaque colonne $\{\nu_{ij} | i \in I\}$ du tableau de fréquence par sa masse totale $\nu_{.j}$ (total de la colonne) on définit le profil de la colonne j :

$$\nu_I^j = \{\nu_{i.}^j | i \in I\} = \left\{ \frac{\nu_{ij}}{\nu_{.j}} | i \in I \right\} \quad (2.2.20)$$

Ce profil ν_I^j est appelé distribution conditionnelle, de i pour j donné ou encore j -ème profil colonne.

On montre que les variables \mathcal{X} et \mathcal{Y} sont indépendantes (dans la population Ω) si et seulement si l'une des 3 conditions (équivalentes entre elles) est vérifiée :

1. tous les profils-lignes sont égaux

$$\nu_J^i = \nu_J \quad \forall i \in I \quad (2.2.21)$$

2. ou tous les profils-colonnes sont égaux

$$\nu_I^j = \nu_I \quad \forall j \in J \quad (2.2.22)$$

3. ou la loi de probabilité sur l'ensemble produit est égale au produit des lois de probabilités marginales :

$$\nu_{ij} = \nu_{i.} \nu_{.j} \quad \forall i \in I \quad \forall j \in J \quad (2.2.23)$$

On retrouve là la notion classique d'indépendance au sens de la théorie des probabilités.

2.2.3.2 Les objectifs de l'AFC :

L'AFC est une technique de description d'un tableau de contingence, qui va permettre de repérer les cases du tableau où les effectifs observés ν_{ij} sont nettement différents des quantités $\nu_{i.}\nu_{.j}$, appelées effectifs « théoriques », qui seraient ceux que l'on observerait s'il y avait indépendance entre \mathcal{X} et \mathcal{Y} . D'une façon générale, l'AFC peut être présentée comme une méthode faisant apparaître, par ordre d'importance décroissant, les écarts à la situation d'indépendance, tant au niveau des lignes, que des colonnes, ou des cases du tableau de contingence.

A chaque modalité i de \mathcal{X} on associe le vecteur de \mathbb{R}^L représentant le profil-ligne de j sachant i :

$$\nu_j^i = (\nu_1^i \dots \nu_j^i) \in \mathbb{R}^L$$

Pour mesurer les proximités entre les modalités i , on doit définir une distance dans \mathbb{R}^L . Pour donner le même poids aux colonnes de $\nu_{i\mathbf{J}}$, on pondère les quantités $(\nu_j^i - \nu_j^{i'})^2$ par les inverses des poids $\nu_{.j}$ des modalités j : on augmente ainsi l'importance des petits écarts dans les colonnes ν_j de poids faible.

La distance ainsi définie s'appelle la distance du chi-deux entre profils-lignes :

$$d_{\chi^2}^2(\nu_j^i, \nu_j^{i'}) = \sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{i.}} - \frac{\nu_{i'j}}{\nu_{i'.}} \right)^2 = \sum_{j=1}^L \frac{1}{\nu_{.j}} (\nu_j^i - \nu_j^{i'})^2 \quad (2.2.24)$$

Propriété 2.2.1. « Équivalence distributionnelle » : si l'on agrège deux colonnes du tableau de contingence \mathbf{F} ayant même profil, i.e. si on remplace deux colonnes proportionnelles par une seule colonne somme des deux précédentes, alors les distances du chi-deux entre profils-lignes restent inchangées, (de même bien sûr que les distances du chi-deux entre les profils-colonnes).

2.2.3.3 Normalisation des données :

Pour faire le lien avec l'Analyse Factorielle sur un espace euclidien (annexes A.2, p : 190 et suivantes), on se ramène à la distance euclidienne canonique en représentant chaque individu (i.e. pixel vecteur) par le vecteur :

$$x_i = \begin{bmatrix} x_i^1 \\ \vdots \\ x_i^j \\ \vdots \\ x_i^L \end{bmatrix} \quad \text{où} \quad x_i^j = \frac{\nu_j^i}{\sqrt{\nu_{.j}}} = \frac{\nu_{ij}}{\sqrt{\nu_{.j}\nu_{i.}}} \quad (2.2.25)$$

On a alors comme distance du chi-deux entre les profils-lignes :

$$d^2(x_i, x_{i'}) = d_{\chi^2}^2(\nu_j^i, \nu_j^{i'}) \quad (2.2.26)$$

On pose :

$$\mathbf{X}_{(P, L)} = \begin{bmatrix} \vdots & & \\ \dots & x_i^j & \dots \\ \vdots & & \end{bmatrix} \quad \mathbf{D}_{\mathbf{X}}_{(P, P)} = \begin{bmatrix} \nu_{i.} & & 0 \\ & \ddots & \\ 0 & & \nu_{P.} \end{bmatrix} \quad \mathbf{D}_{\mathbf{Y}}_{(L, L)} = \begin{bmatrix} \nu_{.1} & & 0 \\ & \ddots & \\ 0 & & \nu_{.L} \end{bmatrix}$$

Alors :

$$\mathbf{X} = \mathbf{D}_X^{-1} \Phi \mathbf{D}_Y^{-1/2} \quad (2.2.27)$$

où \mathbf{D}_X^{-1} est la matrice diagonale de terme général $1/\nu_i$, $\mathbf{D}_Y^{-1/2}$ la matrice diagonale de terme général $1/\sqrt{\nu_j}$ et Φ la matrice de fréquence de terme général ν_{ij} .

On munit chaque point x_i d'un poids égal à la fréquence ν_i de la modalité i dans la population : \mathbf{D}_X est la matrice des poids des lignes.

On note N_I le nuage représentant les modalités de \mathcal{X} ainsi défini :

$$N_I = \{(x_i, \nu_i)\}_{i=1 \dots P} \quad (2.2.28)$$

Le centre de gravité du nuage N_I est noté g_X .

L'analyse du nuage N_J , associé aux canaux (i.e. à la variable \mathcal{Y}), est analogue à celle de N_I ; il suffit de permuter les indices i et j , et les notations X et Y . Le centre de gravité du nuage N_J est noté g_Y .

2.2.3.4 Liens entre les deux analyses :

Les matrices d'inertie (par rapport à l'origine) des deux nuages ont pour expression :

$$\mathbf{V}_0^X = \mathbf{H}^t \mathbf{H} \quad \mathbf{V}_0^Y = \mathbf{H} \mathbf{H}^t \quad \text{avec } \mathbf{H} = \mathbf{D}_X^{-1/2} \Phi \mathbf{D}_Y^{-1/2} \quad (2.2.29)$$

On montre qu'elles ont mêmes valeurs propres. Les vecteurs propres de \mathbf{V}_0^X sont notés u_α , et ceux de \mathbf{V}_0^Y sont notés v_α . Ils engendrent les mêmes axes factoriels principaux.

2.2.3.5 Axes principaux et facteurs principaux

On note c^α les facteurs principaux du nuage N_I des pixels vecteurs, i.e. la projection des profils lignes ν_j^i dans l'espace factoriel réduit, et d^α les facteurs principaux du nuage N_J des canaux : $c^\alpha = \mathbf{X} u_\alpha \forall \alpha = 1 \dots L-1$. Il y a au maximum $K = L-1$ axes principaux car on ne prend pas en compte le vecteur propre correspondant au centre de gravité g_X du nuage N_I associé à la valeur propre égale à 1.

Sous forme matricielle, les facteurs associés aux pixels c^α sont les colonnes de la matrice \mathbf{C} et les vecteurs propres u_α correspondent aux colonnes de la matrice \mathbf{U} .

$$\mathbf{C} = \mathbf{X} \mathbf{U} \quad (2.2.30)$$

$$\mathbf{U} = \begin{bmatrix} u_1^1 & \cdots & u_1^\alpha & \cdots & u_1^{L-1} \\ \vdots & & \vdots & & \vdots \\ u_j^1 & \cdots & u_j^\alpha & \cdots & u_j^{L-1} \\ \vdots & & \vdots & & \vdots \\ u_L^1 & \cdots & u_L^\alpha & \cdots & u_L^{L-1} \end{bmatrix} \quad (2.2.31)$$

Les facteurs principaux d'un nuage se déduisent des axes principaux de l'autre nuage :

$$c^\alpha = \sqrt{\mu_\alpha} \mathbf{D}_X^{-1/2} v_\alpha \quad (2.2.32)$$

De même :

$$d^\alpha = \sqrt{\mu_\alpha} \mathbf{D}_Y^{-1/2} u_\alpha \quad (2.2.33)$$

La formule de transition entre les facteurs associés à chacune des analyses s'écrit :

$$\begin{aligned} \sqrt{\mu_\alpha} c_i^\alpha &= \sum_{j=1}^L \frac{\nu_{ij}}{\nu_i} d_j^\alpha = \sum_{j=1}^L \nu_j^i d_j^\alpha \\ \sqrt{\mu_\alpha} d_j^\alpha &= \sum_{i=1}^P \frac{\nu_{ij}}{\nu_j} c_i^\alpha = \sum_{i=1}^P \nu_i^j c_i^\alpha \end{aligned} \quad (2.2.34)$$

Ces formules permettent donc de « transiter » de \mathbb{R}^P dans \mathbb{R}^L , et réciproquement, pour représenter les facteurs principaux des deux analyses. Elles justifient la représentation graphique simultanée des résultats des deux analyses que l'on utilise habituellement en AFC (cf. annexes A.4, p : 204 et suivantes). Ces formules de transition sont aussi appelées relations « barycentriques ».

La part d'inertie expliquée par l'axe factoriel engendré par u_α vaut :

$$\tau_\alpha = \frac{\mu_\alpha}{\sum_{\alpha=1}^{L-1} \mu_\alpha} \quad (2.2.35)$$

2.2.3.6 Reconstitution des données

La formule de reconstitution à partir des k premiers axes s'écrit :

$$\widehat{\nu}_{ij}^{(k)} = \nu_i \nu_j \left(1 + \sum_{\alpha=1}^k \mu_\alpha^{-1/2} c_i^\alpha d_j^\alpha \right) \quad (2.2.36)$$

L'AFC apparaît ainsi comme une méthode de reconstitution d'un tableau de fréquences à partir de ses fréquences marginales ν_i et ν_j .

2.2.3.7 Méthodes connexes à l'AFC

Récemment a été développé une méthode connexe à l'AFC appelée Analyse des Correspondances d'un tableau de contingence dont les lignes et les colonnes sont munies d'une structure de graphe bistochastique [131]. Cette méthode peut permettre l'introduction de graphe de voisinage et donc l'ajout de la dimension spatiale dans l'AFC. Cette technique qui semble intéressante reste à explorer pour l'imagerie hyperspectrale.

2.2.4 Analyse en composantes indépendantes

L'analyse en composantes indépendantes (ACI) est une méthode plus récente qui permet d'obtenir une représentation de l'image hyperspectrale dans laquelle les canaux : $\mathbf{c}_\alpha^f(x) = (c_{\alpha_1}^f(x), \dots, c_{\alpha_K}^f(x))$ sont indépendants au lieu d'être décorrélés comme en ACP.

On a donc la même équation que pour l'ACP en notant $\mathbf{A} = \mathbf{U}^t$ (eq. 2.2.12) : $\mathbf{X}^t = \mathbf{A}\mathbf{C}^t$

Les colonnes de la matrice \mathbf{C} correspondent aux canaux (les sources que l'on recherche) et \mathbf{A} est la matrice de l'application linéaire permettant de passer des canaux de l'image (les lignes de \mathbf{X}^t), après normalisation, à ses sources \mathbf{C} .

$$\begin{aligned} x^1 &= a_{11}c_{\alpha_1} + \dots + a_{1P}c_{\alpha_L} \\ &\vdots \\ x^j &= a_{j1}c_{\alpha_1} + \dots + a_{jP}c_{\alpha_L} \\ &\vdots \\ x^L &= a_{L1}c_{\alpha_1} + \dots + a_{LP}c_{\alpha_L} \end{aligned} \quad (2.2.37)$$

L'idée majeure est de trouver les axes portant une distribution la moins gaussienne possible. En effet, d'après le théorème central limite, la distribution gaussienne est la distribution limite quand on mélange linéairement une grande quantité de variables aléatoires indépendantes quelles que soient leurs densités de distributions propres. Ceci indique que plus un axe porte une distribution non-gaussienne, plus il a de chances de s'approcher d'une source « pure » indépendante [158].

L'ACI et l'ACP ont toutes deux pour but de retrouver les variables sous-jacentes c_{α}^f aux problèmes (eq. 2.2.37). L'ACI cherche à rendre les variables indépendantes alors que l'ACP cherche à les décorréler (l'indépendance implique la décorrélation mais la réciproque n'est pas vraie). De manière intuitive, la non corrélation détermine les moments statistiques du deuxième ordre (i.e. les covariances) d'une distribution multivariée, tandis que l'indépendance détermine tous les moments statistiques [76].

Hyvärinen et Oja [82] proposent une méthode fondée sur la « néguentropie » qui conduit à l'algorithme « Fast ICA ». Il existe d'autres méthodes appelées JADE ou « Joint Approximate Diagonalisation of Eigenmatrices » [37], IFA ou « Independent Factor Analysis » [12].

Pour les images hyperspectrales l'ACI a été très employée. Citons entre autres [106, 148, 158].

Notons que l'ACI n'a pas été utilisée dans cette thèse. Il pourrait être intéressant, dans de prochains travaux, de comparer l'ACI par rapport à l'AFC et l'ACP en ayant comme objectif la segmentation morphologique.

2.2.5 Maximum Noise Fraction

La transformée « Maximum Noise Fraction » (MNF) [23, 67] permet de réduire le bruit dans le domaine spectral pour une image hyperspectrale. C'est une méthode dérivée de l'ACP, où le critère pour la génération des composantes principales est qu'elles maximisent le bruit représentée par chacune des composantes, plutôt que la variance des données. Dans l'ordre inverse que dans l'ACP, ces composantes maximisent le rapport signal à bruit (RSB) de chaque composante. La MNF est équivalente à une transformation des données dans un système de coordonnées dans lequel la matrice de covariance du bruit est la matrice identité (cette transformation est obtenue par ACP), suivie par une ACP. La succession des deux ACP est la Noise Adjusted Principal Components (NAPC) [102]. La NAPC possède toutes les propriétés de l'ACP, y compris la caractéristique principale de concentrer la partie correspondant au signal sur un faible nombre de composantes.

Dans la pratique on procède de la manière suivante :

1. la matrice de covariance du bruit doit être estimée. Dans [67] une méthode spatiale par différences entre pixels voisins est présentée ;

2. une première ACP permet de diagonaliser la matrice de covariance du bruit, ce qui conduit à « sphériser » le bruit (matrice identité) indépendamment des données ;
3. une seconde ACP diagonalise ces données « ajustées au bruit », conduisant à trouver les directions de variance maximale de la composante indépendamment du bruit.

Du fait de cette reformulation, la transformation est également appelée « analyse en composantes principales ajustées au bruit » (« Noise-Adjusted Principal Components Analysis », NAPCA). La méthode permet ainsi d'ordonner les composantes dans l'espace transformé par ordre décroissant de rapport signal sur bruit et donne en outre l'accès à une estimation de ce rapport signal sur bruit pour chaque composante [104].

Remarque 2.2.4. *Green et al.* ont démontrés que l'ACP et la MNF sont équivalentes dans le cas d'un bruit blanc décorrélé de même variance sur tous les canaux. Ceci explique que l'ACP soit très utilisée pour réduire le bruit sur des images hyperspectrales sur lesquelles le bruit est approximativement de même amplitude et décorrélé sur tous les canaux.

2.2.6 Poursuite de projection

La poursuite de projection ou « projection poursuit » (PP) a été développée par *Friedman et Tuckey* [58] en 1974. Ils proposent une technique d'exploration graphique pour visualiser les données de grandes dimensions. Leur idée est que la plupart des projections en faible dimension (1D ou 2D) de données de grandes dimensions ont une distribution gaussienne. Les structures d'intérêts, comme les « clusters » ou les « longues traînes », peuvent donc être mises en évidence par des projections non-gaussiennes [76].

Il s'agit de déterminer un sous espace de dimension réduite, en projetant les données sur des axes choisis itérativement. Le choix se fait par maximisation d'une fonction de coût appelée « indice de projection ». L'axe est choisi s'il porte un extremum de l'indice de projection. *Friedman et Tuckey* proposent plusieurs indices de projection pour l'optimisation, chacun étant orienté vers un éloignement de la normalité. Depuis leur proposition initiale plusieurs indices ont été suggérés. *Arzuaga-Cruz et al.* proposent, entre autres, la « distance » de Kullback-Liebler symétrisée [10] et *Jimenez* propose la distance de Bhattacharyya [85].

L'indice de projection est déterminé a priori par l'utilisateur (variance élevée, non-normalité,...). La poursuite de projection permet ainsi de généraliser la méthode d'ACP (variance élevée), ACI (non-normalité élevée), à n'importe quel indice de projection [104]. L'autre point majeur qui diffère des ACP/MNF/ACI est le fait que la PP cherche chaque axe itérativement. Il s'agit de trouver le meilleur premier axe, le retirer et recommencer la procédure entière. La limitation majeure de cette méthodologie réside dans l'itérativité : si un certain axe est mal estimé, tous les suivants sont pollués par la contribution résiduelle [158].

Pour les images hyperspectrales la poursuite de projection a été notamment utilisée dans [149, 85, 83].

2.2.7 Les méthodes non linéaires

Plus récemment que les méthodes linéaires, sont apparues des méthodes non linéaires de réduction de données. Certaines de ces techniques préservent des propriétés globales des données :

- Multidimensional scaling (MDS) [45, 94];
- Isomap [177, 178];
- Kernel PCA [157] qui est la reformulation de l'ACP dans un espace de grande dimension à l'aide d'une fonction noyau ;
- Diffusion maps [95, 133];
- Multilayer auto-encoders [50, 78].

D'autres méthodes préservent des propriétés locales (i.e. sur un petit voisinage) des données :

- Local Linear Embedding (LLE) [153]. Une adaptation pour la réduction de dimension spectrale en tenant compte de la dimension spatiale est faite dans [130];
- Laplacian Eigenmaps [17];
- Hessian LLE (HLLE) [52];
- Local Tangent Space Analysis (LTSA) [192].

Enfin d'autres techniques consistent en des alignements globaux de modèles linéaires. C'est à dire que plusieurs modèles linéaires sont calculés puis une représentation des données en plus faible dimension est effectuée par alignement de ces modèles linéaires : Locally Linear Coordination (LLC) [176]. La thèse de *Journaux* [86] présente un état de l'art détaillé des méthodes de réduction de données et notamment des méthodes non-linéaires.

Dans une étude comparative approfondie entre des méthodes linéaires et non linéaires *Maaten et al.* [184] montrent que les méthodes non linéaires ne surpassent pas les méthodes linéaires sur des données réelles, c'est à dire non simulées. En outre elles sont plus coûteuses en temps de calcul. En conséquence, ces méthodes n'ont pas été utilisées dans cette thèse. Néanmoins pour un problème spécifique se résolvant mieux par des méthodes non linéaires, il pourrait être intéressant, dans de futurs travaux, d'étudier l'influence de ces méthodes en vue de la segmentation morphologique.

Il existe aussi des méthodes tensorielles, encore appelées multilinéaires, qui permettent de réduire les données [150, 186, 187].

En outre des méthodes à partir d'ondelettes ont été développées. Citons notamment « wawanglet » [158]. La compression d'images hyperspectrales et l'impact sur la qualité des données ont également été étudiés, e.g. [41].

2.3 Images hyperspectrales

Les résultats vont être présentés sur différentes images. La première série concerne l'image « woman face » acquise dans le domaine spectral des longueurs d'ondes (fig. 2.2).

La deuxième image hyperspectrale est une série temporelle enregistrée sur un bouchon après un temps de chauffage de 2 secondes (fig. 2.3).

Les images de la figure 2.5 présentent deux vues du bouchon en plastique. Des occlusions de colle sont visibles au centre du bouchon sur la figure 2.5. Différentes méthodes de contrôle non destructif ont été étudiées par *Legrand et al.* [103]. La solution proposée est un contrôle par thermographie active [72] : la surface supérieure du bouchon est soumise à une impulsion thermique d'une durée définie. A l'aide d'une caméra thermique, l'évolution temporelle de la répartition en température de la surface chauffée du bouchon est ensuite enregistrée. L'analyse de cette évolution permet de détecter la présence ou non de colle à l'intérieur du bouchon. Celle-ci se manifeste par des variations de la

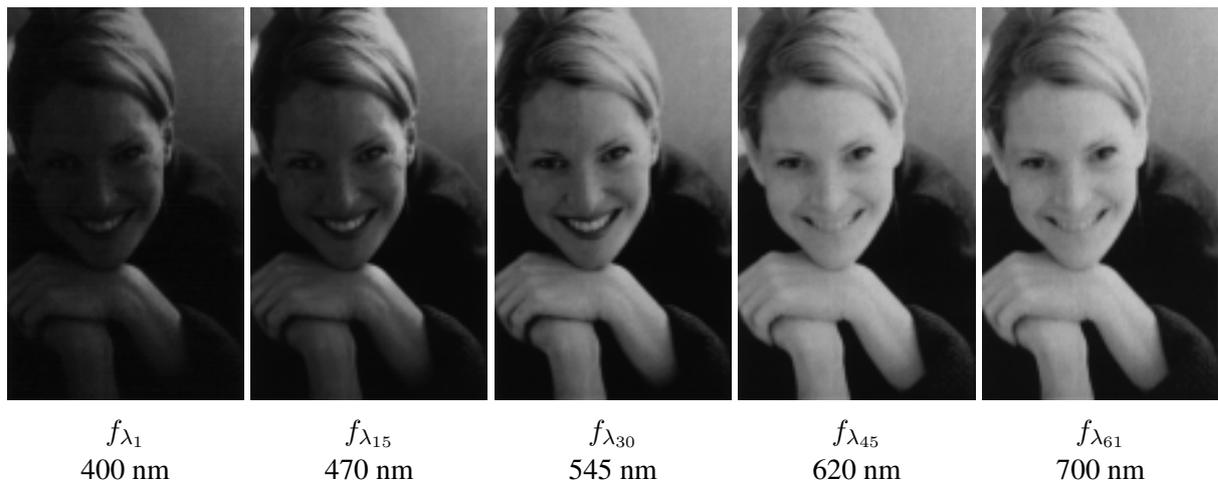


FIG. 2.2 – Cinq canaux de l'image hyperspectrale « woman face » acquise dans le domaine spectral des longueur d'ondes ($152 \times 91 \times 61$ pixels) à 61 canaux acquis dans le spectre entre 400 nm et 700 nm avec un pas de 5 nm. (Source de l'image : Spectral Database, University of Joensuu Color Group, <http://spectral.joensuu.fi/>).

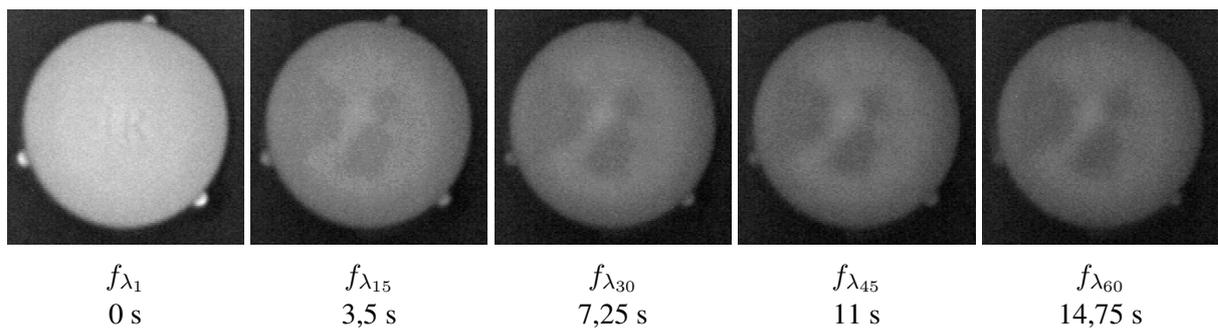


FIG. 2.3 – Cinq canaux de l'image hyperspectrale « bouchon en plastique » qui est une série temporelle ($256 \times 256 \times 60$ pixels) à 60 canaux acquises toutes les 250 ms. (Source de l'image hyperspectrale : Laboratoire LE2i, Le Creusot, France).

transmission du front thermique et se visualise sur les images par des zones sombres froides. Le but est de quantifier le plus tôt possible ces empreintes thermiques de colle.

Il n'existe pas de vérité « terrain », mais on dispose de résultats sur la figure 2.4 obtenus sur la séquence 2.

La méthode pour obtenir ce résultat est décrite dans [103]. Elle consiste, principalement, à extraire les zones de colles par différence d'images entre images sans colle et avec colle. Comme nous ne possédons pas l'image sans la colle il nous est impossible d'utiliser cette méthode.

La troisième image hyperspectrale est une série d'IRM de contraste (Imagerie par Résonance Magnétique) « DCE-MRI : Dynamic Contrast Enhanced Magnetic Resonance Imaging » 2.6. L'image est une série d'IRM temporelle de 512 canaux 128×128 acquise, avec un pas temporel régulier d'une seconde, sur une souris présentant une tumeur [13]. Les canaux sont recalés car la souris est

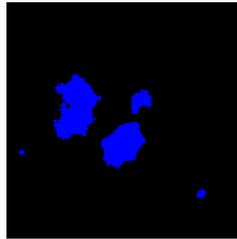


FIG. 2.4 – Référence pour la zone de colle du bouchon en plastique.

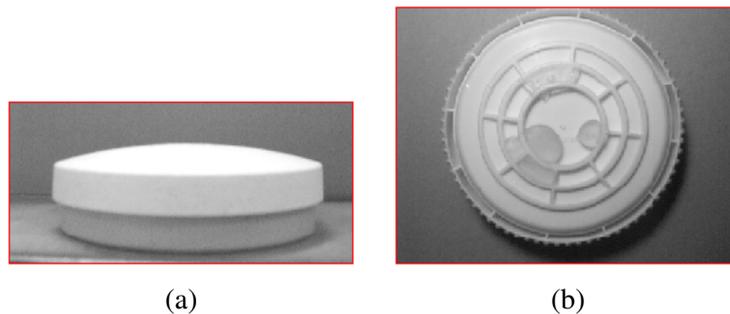


FIG. 2.5 – Vues de côté (a) et de dessus (b) du bouchon en plastique.

maintenue immobile par un système mécanique. L'image provient de l'équipe du Pr. C.A. Cuenod du service imagerie de l'Hôpital Européen Georges Pompidou à Paris.

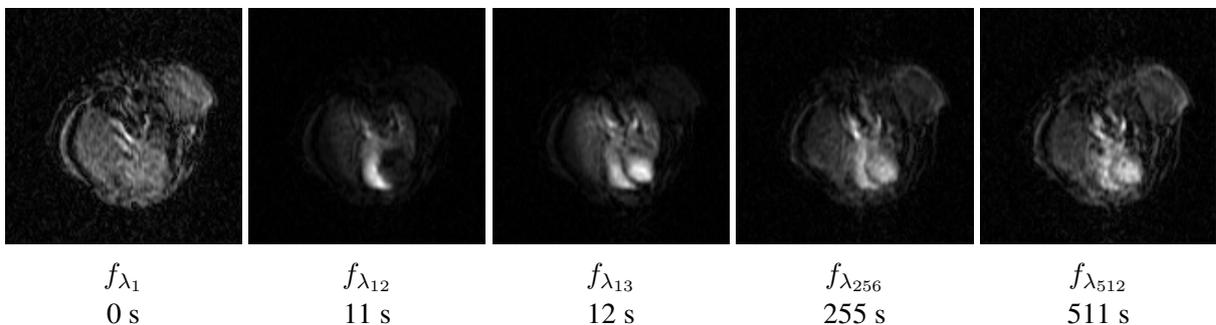


FIG. 2.6 – Cinq canaux de l'image hyperspectrale de souris « serim447 » qui est une série temporelle ($128 \times 128 \times 512$ pixels) à 512 canaux acquis toutes les secondes. (Source de l'image hyperspectrale : Pr. Cuenod Service Imagerie - Hôpital Européen Georges Pompidou, Paris).

Une quatrième série d'image correspond à des images de télédétection à 5 composantes (fig. 2.7) :

- f_{λ_1} canal bleu de taille 365×365 pixels avec une résolution de 2,80 m et dont les longueurs d'ondes sont acquises de 0,45 à 0,53 μm ;
- f_{λ_2} canal vert de taille 365×365 pixels avec une résolution de 2,80 m et dont les longueurs d'ondes sont acquises de 0,51 à 0,59 μm ;
- f_{λ_3} canal rouge de taille 365×365 pixels avec une résolution de 2,80 m et dont les longueurs d'ondes sont acquises de 0,62 à 0,70 μm ;
- f_{λ_4} canal proche-infrarouge de taille 365×365 pixels avec une résolution de 2,80 m et dont les

longueurs d'ondes sont acquises de 0,78 à 0,92 μm ;

- f_{λ_5} canal panchromatique de taille 1460×1460 pixels avec une résolution de 0,70 m.

Ces images sont des images satellites PLEIADES simulées à partir d'images PELICAN. Elles ont été acquises par le CNES (Centre National d'Étude Spatiale) et nous ont été fournies par *G. Flouzat* [101].

Pour des questions de temps de calculs, nous avons choisi de sous-échantillonner le canal panchromatique f_{λ_5} pour en faire un canal de taille 365×365 pixels. En effet, notre but n'est pas de détecter le plus précisément possible une cible dans l'image mais de comparer des algorithmes entre eux. Il est évident que pour de la détection de cible il faudrait travailler sur une image hyperspectrale, de taille $1460 \times 1460 \times 5$ pixels, issue de la fusion des canaux f_{λ_1} (bleu), f_{λ_2} (vert), f_{λ_3} (rouge) et f_{λ_4} (proche Infra-Rouge) avec le canal panchromatique f_{λ_5} . La fusion de ces images est présentée dans [101]. De nombreux travaux existent sur la fusion d'images de résolution spatiales différentes, citons entre autres [179].

Pour représenter cette image, nous avons créé une image de synthèse RVB composée du canal rouge f_{λ_3} , du canal vert f_{λ_2} et du canal bleu f_{λ_1} .

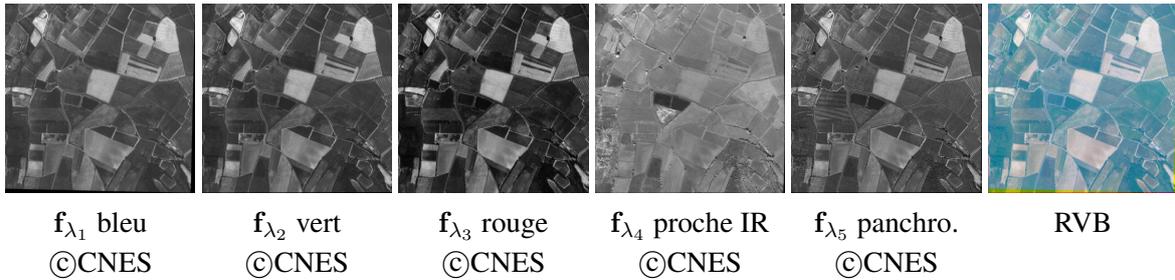


FIG. 2.7 – Les 5 canaux de l'image hyperspectrales « Roujan » et l'image de synthèse RVB.

2.4 Filtrage et réduction de dimension

Comme expliqué au début du chapitre 2, la réduction de dimension des images hyperspectrales est nécessaire pour lutter contre le phénomène de Hughes et pour accélérer les temps de traitements.

La réduction de la dimension spectrale a été effectuée à l'aide de l'Analyse Factorielle des Correspondances. On obtient un espace factoriel de dimension réduite par rapport à l'espace image de départ. Cet espace réduit est composé des facteurs pixels c_{α}^f de l'image f_{λ} .

On peut également reconstituer l'image f_{λ} à partir d'un nombre limité de facteurs et ainsi obtenir une bonne approximation \hat{f}_{λ} de l'image de départ. Cette image reconstituée contient, sous certaines conditions, un bruit spectral plus faible que sur l'image originale. Nous présenterons une méthode permettant de réduire le bruit des facteurs pixels c_{α}^f .

En outre, réduire la dimension spectrale de l'image, par ajustement d'un modèle de spectre en chaque pixel vecteur, a été étudié. Il est intéressant d'appliquer la réduction de dimension par approche modèle sur l'image filtrée, par Analyse de Données, afin de profiter du filtrage spectral déjà effectué. On obtient ainsi des cartes de paramètres du modèle. Ces cartes constituent également un espace réduit que l'on pourra utiliser pour la classification et la segmentation.

2.4.1 Réduction de dimension par AFC

Nous avons privilégié l'AFC pour réduire la dimension de l'image hyperspectrale car notre objectif est la segmentation d'images hyperspectrales. En outre l'AFC présente l'avantage par rapport à l'ACP de traiter les pixels vecteurs et les canaux de la même manière. On peut ainsi les représenter de manière simultanée dans l'espace factoriel. D'autres méthodes peuvent néanmoins être utilisées comme l'ACI, cependant elles nécessitent un choix d'algorithme, i.e. Fast ICA, JADE (cf. partie 2.2.4, p : 16) pour rendre indépendant les canaux.

L'AFC peut être vue comme une transformation ζ de l'espace image de départ \mathbf{f}_λ , à L dimensions, dans un espace \mathbf{c}_α^f , de dimension réduite $K < L$, ainsi qu'un ensemble de paramètres (eq. 2.4.1). Dans l'espace factoriel, les coordonnées des vecteurs pixels sur chacun des axes factoriels sont appelées facteurs pixels \mathbf{c}_α^f . On représente ces facteurs pixels sur la grille d'échantillonnage de l'image : le bitmap.

$$\zeta : \begin{cases} \mathcal{T}^L & \rightarrow \mathcal{T}^K / K < L \\ \mathbf{f}_\lambda(x) & \rightarrow \begin{pmatrix} \mathbf{c}_\alpha^f(x) = (c_{\alpha_1}^f(x), \dots, c_{\alpha_K}^f(x)) \\ \mathbf{d}_{\alpha\lambda}^f = (d_{\alpha_1\lambda}^f, \dots, d_{\alpha_K\lambda}^f) \\ \{\mu_\alpha\}_{\alpha=1\dots K} \\ \{\nu_i\}_{i=1\dots P} \\ \{\nu_j\}_{j=1\dots L} \\ f = \sum_i \sum_j f_{ij} \end{pmatrix} \end{cases} \quad (2.4.1)$$

avec $\mathbf{d}_{\alpha\lambda}^f$ les facteurs-canaux. Dans la formule de reconstitution des données (eq. 2.2.36) on utilise la notation d_j^α qui correspond à $d_j^\alpha = d_{\alpha_k\lambda_j}^f$.

Un nombre limité K d'axes factoriels est généralement choisi. Ainsi, l'AFC peut être vue comme une projection des pixels initiaux de l'espace image dans l'espace factoriel de plus faible dimension que l'espace image ($K < L$).

La reconstitution est une transformation pseudo-inverse $\hat{\zeta}^{-1}$, ou une transformation inverse lorsque tous les axes sont conservés $K = L - 1$ (eq. 2.4.2). Elle consiste à reconstituer l'image $\hat{\mathbf{f}}_\lambda$ depuis les facteurs \mathbf{c}_α^f et les paramètres. L'image reconstituée est une approximation de l'image originale si l'on reconstitue à partir d'une partie des axes factoriels :

$$\hat{\zeta}^{-1} : \begin{cases} \mathcal{T}^K & \rightarrow \mathcal{T}^L / K < L \\ \begin{pmatrix} \mathbf{c}_\alpha^f(x) \\ \mathbf{d}_{\alpha\lambda}^f = (d_{\alpha_1\lambda}^f, \dots, d_{\alpha_K\lambda}^f) \\ \{\mu_\alpha\}_{\alpha=1\dots K} \\ \{\nu_i\}_{i=1\dots P} \\ \{\nu_j\}_{j=1\dots L} \\ f = \sum_i \sum_j f_{ij} \end{pmatrix} & \rightarrow \hat{\mathbf{f}}_\lambda(x) = (\hat{f}_{\lambda_1}(x), \dots, \hat{f}_{\lambda_L}(x)) \end{cases} \quad (2.4.2)$$

2.4.1.1 Choix du nombre d'axes factoriels

Afin de comprendre les effets de l'AFC, on va étudier l'image « woman face ». On doit tout d'abord choisir le nombre d'axes factoriels à conserver (fig. 2.8) qui dépend de :

- la part d'inertie qu'ils représentent (fig. 2.9). On peut utiliser un critère sur l'inertie comme le critère de *Kaiser* ou le « scree test » (cf. annexes A.3)
- la quantité d'information portée par les facteurs pixels. On peut observer de visu l'information portée par les facteurs pixels. Nous proposerons également un critère utilisant le rapport signal à bruit sur les facteurs pixels.

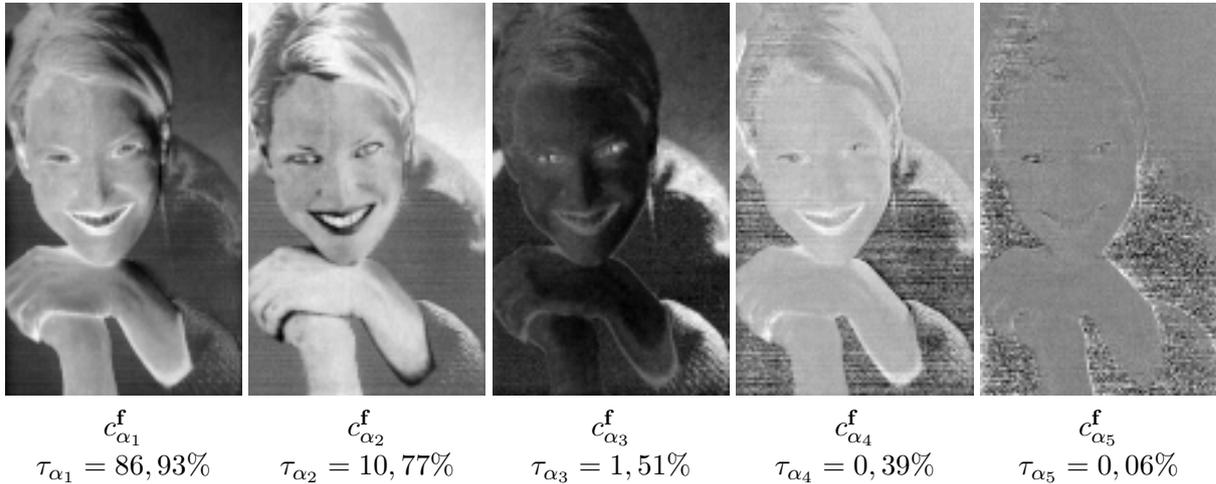


FIG. 2.8 – Facteurs pixels $c_{\alpha_k}^f$ correspondant aux 5 premiers axes factoriels de l'image hyperspectrale « woman face ». Les parts d'inertie des axes sont notées τ_{α_k} .

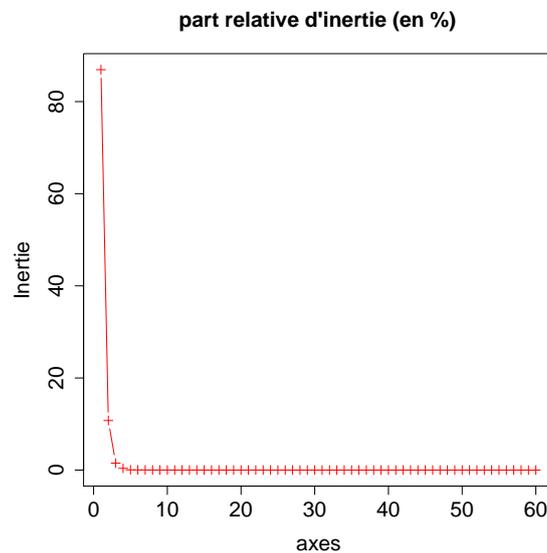


FIG. 2.9 – Part d'inertie des axes factoriels de l'image « woman face ».

Par observation de l'inertie des axes factoriels (fig. 2.9) on constate que jusqu'au troisième axe Δu_{α_3} , les inerties sont significatives, ensuite elles sont négligeables. La part d'inertie expliquée par le sous espace F_3 engendré par les trois premiers axes factoriels vaut : $\tau_{\alpha_1} + \tau_{\alpha_2} + \tau_{\alpha_3} = 99,21\%$.

En outre, sur la figure 2.8 les facteurs pixels $c_{\alpha_1}^f$, $c_{\alpha_2}^f$ et $c_{\alpha_3}^f$ sont, de visu, moins bruités que les suivants.

Par conséquent on choisit de ne retenir que les trois premiers axes factoriels et les facteurs pixels associés : $\mathbf{c}_\alpha^f = (c_{\alpha_1}^f, c_{\alpha_2}^f, c_{\alpha_3}^f)$.

Le choix des axes, tel que proposé, nécessite une observation qualitative des facteurs pixels. Afin de quantifier la quantité d'information portée par les facteurs pixels on va mesurer leur rapport signal à bruit par une nouvelle méthode que nous avons développée [138].

On considère ici le canal $c_{\alpha_k}^f$ (ou f_{λ_j}) comme un ensemble de réalisations $c_{\alpha_k}^f(x)$ (ou $f_{\lambda_j}(x)$) d'une variable aléatoire sur le compact E , qui correspond à la grille d'échantillonnage de l'image.

Pour chaque axe factoriel Δu_{α_k} , on estime la covariance spatiale centrée, sur les facteurs pixels, par une Transformée de Fourier Rapide : « Fast Fourier Transform » FFT, en supposant que $\bar{c}_{\alpha_k}^f(x)$ est une fonction stationnaire :

$$\bar{g}_{\alpha_k}(h) = E[\bar{c}_{\alpha_k}^f(x)\bar{c}_{\alpha_k}^f(x+h)] \quad (2.4.3)$$

avec $\bar{c}_{\alpha_k}^f$ le canal α_k centré $\bar{c}_{\alpha_k}^f(x) = c_{\alpha_k}^f(x) - E[c_{\alpha_k}^f(x)]$ et $E[Y]$ l'espérance de la variable aléatoire Y . $E[c_{\alpha_k}^f(x)]$ correspond à la moyenne du canal $c_{\alpha_k}^f(x)$.

Le pic de la covariance contient la somme de la variance du signal et de la variance du bruit : effet pépite (cf. § « effet pépite » ci-après). Donc la variance du signal est estimée par le maximum (i.e. la valeur à l'origine) de la covariance \bar{g} après une ouverture morphologique γ : $Var(signal) = \gamma\bar{g}_{\alpha_k}(0)$. L'élément structurant est le plus petit possible : on choisit donc un élément structurant de taille 3×3 pixels.

Comme une des propriétés de l'ouverture morphologique est de supprimer les pics blancs une fonction, la variance du bruit est égale au résidu de l'ouverture sur la covariance à l'origine : $Var(bruit) = \bar{g}_{\alpha_k}(0) - \gamma\bar{g}_{\alpha_k}(0)$. On constate, sur la figure 2.10, que la covariance présente un pic plus resserré et plus haut pour les facteurs pixels $c_{\alpha_5}^f$ qui sont plus bruités, que pour les facteurs pixels $c_{\alpha_5}^f$ qui sont très peu bruités.

On peut donc définir un rapport signal à bruit.

Définition 1. Le rapport signal à bruit est défini par :

$$RSB_{\alpha_k} = \frac{Var(signal)}{Var(bruit)} = \frac{\gamma\bar{g}_{\alpha_k}(0)}{\bar{g}_{\alpha_k}(0) - \gamma\bar{g}_{\alpha_k}(0)} \quad (2.4.4)$$

avec \bar{g} la covariance centrée et γ l'ouverture morphologique.

Par observation des rapports signaux sur bruits, RSB, (fig. 2.11), les facteurs pixels sur les axes 1, 2 et 3 sont les moins bruités. On pourrait envisager de conserver l'axe 4 dont les facteurs pixels ont un RSB supérieurs aux facteurs pixels des axes suivants. Cependant comme sa part d'inertie est très faible ($\tau_{\alpha_4} = 0,39\%$), par rapport aux autres on décide de ne pas le conserver. En outre, on verra plus loin que l'écart de reconstitution par rapport à l'image de départ avec trois axes ou quatre axes est quasiment le même.

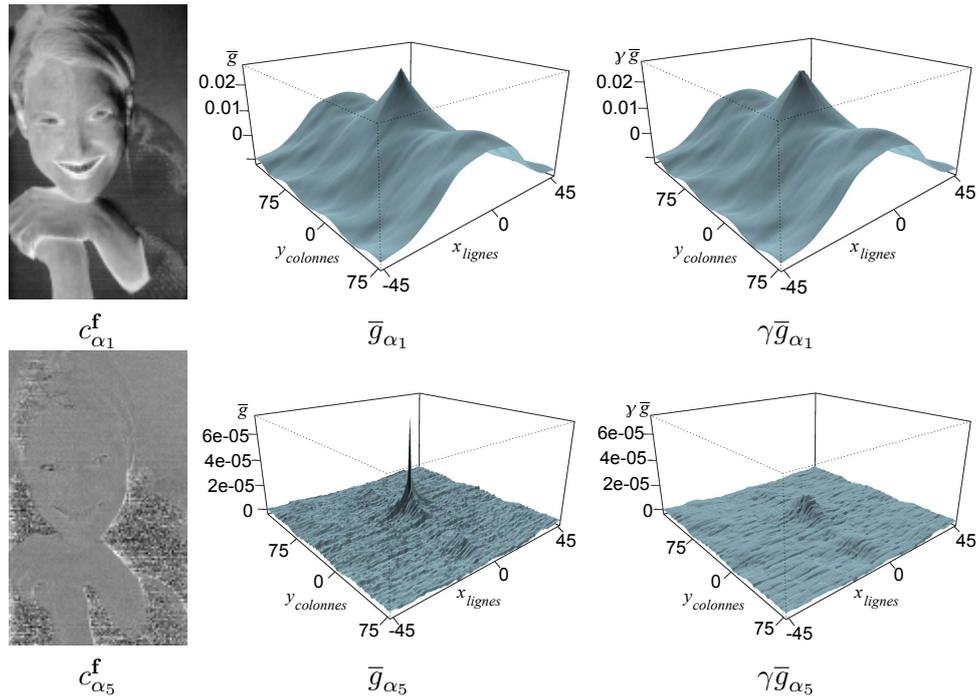


FIG. 2.10 – Covariance avant (\bar{g}) et après ouverture ($\gamma \bar{g}$) sur les facteurs pixels c_{α}^f de l'image « woman face ».

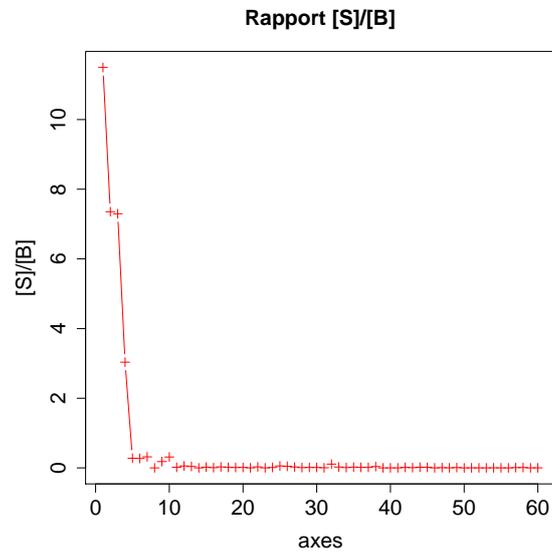


FIG. 2.11 – Rapport signal à bruit sur les facteurs pixels c_{α}^f de l'image « woman face ».

2.4.1.2 Effet pépité

L'effet pépité [114] consiste en une discontinuité de la covariance $\bar{\gamma}_{\lambda_j}$ à l'origine en présence de bruit sur la fonction considérée (i.e. le canal). On se propose de démontrer cet effet.

On utilise la notation pour l'image centrée $\overline{f_{\lambda_j}}(x) = f_{\lambda_j}(x) - E[f_{\lambda_j}(x)]$. On modélise un point de l'image hyperspectrale $f_{\lambda_j}(x)$ par la somme d'un signal $Z_{\lambda_j}(x)$ et d'un bruit $\varepsilon_{\lambda_j}(x)$ dont les réalisations sont indépendantes $\forall x \in E$; ce qui donne : $f_{\lambda_j}(x) = Z_{\lambda_j}(x) + \varepsilon_{\lambda_j}(x)$.

La covariance entre une fonction stationnaire en un point x_i et la même fonction en un point $x_{i'}$ s'exprime sous la forme d'une fonction g dépendant de la différence $x_i - x_{i'} = h$ avec $h \in E$: $g(h) = E[\overline{f_{\lambda_j}}(x_i)\overline{f_{\lambda_j}}(x_{i'})] = E[\overline{f_{\lambda_j}}(x_i)\overline{f_{\lambda_j}}(x_i + h)]$.

Dans le domaine spatial, pour le canal f_{λ_j} supposé stationnaire, la covariance s'écrit :

$$\begin{aligned} \overline{g_{\lambda_j}}(h) &= E[\overline{f_{\lambda_j}}(x)\overline{f_{\lambda_j}}(x+h)] \\ &= E[(\overline{Z_{\lambda_j}}(x) + \overline{\varepsilon_{\lambda_j}}(x))(\overline{Z_{\lambda_j}}(x+h) + \overline{\varepsilon_{\lambda_j}}(x+h))] \\ &= E[\overline{Z_{\lambda_j}}(x)\overline{Z_{\lambda_j}}(x+h)] + E[\overline{\varepsilon_{\lambda_j}}(x)\overline{\varepsilon_{\lambda_j}}(x+h)] \\ &= E[\overline{Z_{\lambda_j}}(x)\overline{Z_{\lambda_j}}(x+h)] + E[\overline{\varepsilon_{\lambda_j}}(x)^2]\delta_0(h) \end{aligned} \quad (2.4.5)$$

En effet si $h \neq 0$ $E[\overline{\varepsilon_{\lambda_j}}(x)\overline{\varepsilon_{\lambda_j}}(x+h)] = 0$ car les réalisations du bruits sont indépendantes.

L'effet pépité à l'origine correspond à la distribution de Dirac $\delta_0(h)$ multipliée par la variance du bruit $E[\overline{\varepsilon_{\lambda_j}}(x)^2]$.

Pour une variable aléatoire, Y on a $Var[Y] = E[(Y - E[Y])^2] = E[\overline{Y}^2] = Var[\overline{Y}]$.

Exprimons maintenant la covariance à l'origine (i.e. pour $h = 0$) :

$$\begin{aligned} \overline{g_{\lambda_j}}(0) &= E[\overline{Z_{\lambda_j}}(x)^2] + E[\overline{\varepsilon_{\lambda_j}}(x)^2] \\ &= Var[\overline{Z_{\lambda_j}}(x)] + Var[\overline{\varepsilon_{\lambda_j}}(x)] \\ &= Var[Z_{\lambda_j}(x)] + Var[\varepsilon_{\lambda_j}(x)] \end{aligned} \quad (2.4.6)$$

L'origine de la covariance $\overline{g_{\lambda_j}}(0)$ est donc la somme de la variance du signal et de la variance du bruit.

Remarque 2.4.1. L'hypothèse de stationnarité est vérifiée dans le domaine spatial à petite échelle. Comme on observe la covariance autour de l'origine cela ne pose pas de problème [114].

2.4.1.3 Covariance spectrale

Dans le domaine spectral, on ne fait pas d'hypothèse de stationnarité. La matrice de covariance centrée s'écrit $\overline{\Sigma}(\lambda_j, \lambda_{j'})_{j,j'=1\dots L}$:

$$\begin{aligned} \overline{\Sigma}(\lambda_j, \lambda_{j'}) &= E[\overline{f}(x, \lambda_j)\overline{f}(x, \lambda_{j'})] \\ &= E[(\overline{Z}(x, \lambda_j) + \overline{\varepsilon}(x, \lambda_j))(\overline{Z}(x, \lambda_{j'}) + \overline{\varepsilon}(x, \lambda_{j'}))] \\ &= E[\overline{Z}(x, \lambda_j)\overline{Z}(x, \lambda_{j'})] + E[\overline{\varepsilon}(x, \lambda_j)\overline{\varepsilon}(x, \lambda_{j'})] \\ &= E[\overline{Z}(x, \lambda_j)\overline{Z}(x, \lambda_{j'})] + E[\overline{\varepsilon}(x, \lambda_j)^2]\delta_{\lambda_j=\lambda_{j'}} \end{aligned}$$

A l'origine pour $\lambda_j = \lambda_{j'}$ on a :

$$\begin{aligned}\overline{\Sigma}(\lambda_j, \lambda_j) &= E[\overline{Z}(x, \lambda_j)^2] + E[\overline{\varepsilon}(x, \lambda_j)^2] \\ &= \text{Var}[\overline{Z}(x, \lambda_j)] + \text{Var}[\overline{\varepsilon}(x, \lambda_j)] \\ &= \text{Var}[Z_{\lambda_j}(x)] + \text{Var}[\varepsilon_{\lambda_j}(x)]\end{aligned}\tag{2.4.7}$$

D'après les équations 2.4.6 et 2.4.7, on a $\forall \lambda_j, j = 1, \dots, L$:

$$\overline{\Sigma}(\lambda_j, \lambda_j) = \overline{g}_{\lambda_j}(0)\tag{2.4.8}$$

On vérifie expérimentalement, cette relation sur l'image de souris « serim447 » sur les canaux $f_{\lambda_{13}}$ et $f_{\lambda_{50}}$. On mesure la covariance spatiale dans l'espace image et on observe les valeurs à l'origine $\overline{g}_{\alpha_{13}}(0)$ et $\overline{g}_{\alpha_{50}}(0)$ pour les deux canaux. Dans le domaine spectral, on calcule la matrice de covariance centrée entre les canaux f_{λ_j} , et on mesure les valeurs $\overline{\Sigma}(\lambda_{13}, \lambda_{13})$ et $\overline{\Sigma}(\lambda_{50}, \lambda_{50})$.

Pour le canal $f_{\lambda_{50}}$, on constate que :

- $\overline{g}_{\lambda_{50}}(0) = 2\,207\,890$
- $\overline{\Sigma}(\lambda_{50}, \lambda_{50}) = 2\,208\,140$

donc $\overline{g}_{\lambda_{50}}(0) \simeq \overline{\Sigma}(\lambda_{50}, \lambda_{50})$. L'écart relatif de $\overline{g}_{\lambda_{50}}(0)$ par rapport à $\overline{\Sigma}(\lambda_{50}, \lambda_{50})$ vaut 0,01 %.

Pour le canal $f_{\lambda_{13}}$, on constate que :

- $\overline{g}_{\lambda_{13}}(0) = 4\,080\,225$
- $\overline{\Sigma}(\lambda_{13}, \lambda_{13}) = 4\,080\,688$

donc $\overline{g}_{\lambda_{13}}(0) \simeq \overline{\Sigma}(\lambda_{13}, \lambda_{13})$. L'écart relatif de $\overline{g}_{\lambda_{13}}(0)$ par rapport à $\overline{\Sigma}(\lambda_{13}, \lambda_{13})$ vaut 0,01 %.

Par conséquent, l'équation 2.4.8 est vérifiée expérimentalement.

Remarque 2.4.2. S'il est possible de mesurer le RSB sur les canaux dans l'espace image par ouverture morphologique sur la covariance \overline{g}_{λ_j} , il n'est pas toujours possible de le mesurer par ouverture unidimensionnelle sur la covariance spectral $\overline{\Sigma}(\lambda_j, \lambda_{j'})$. Ceci est dû à l'échantillonnage spectral qui ne respecte pas le théorème de Shannon pour l'image de souris. En fait entre deux canaux il peut y avoir plusieurs battements cardiaques. En outre, même sur une autre séquence d'images, l'échantillonnage spatial et l'échantillonnage spectral n'ont pas le même pas. On ne peut donc pas comparer les RSB par ouverture dans le domaine spatial et dans le domaine spectral (cf. annexes pour plus de détails).

2.4.1.4 Reconstitution de l'image de départ

Après avoir réduit le nombre d'axes factoriels à 3 axes, il est intéressant de reconstituer l'image de départ (eq. 2.4.2 et 2.2.36). Pour rappel : la reconstitution donne une image à 61 canaux $\widehat{\mathbf{f}}_{\lambda}$, comme l'image de départ, et qui est une bonne approximation de l'image de départ \mathbf{f}_{λ} .

Observons les reconstitutions de l'image de départ avec un nombre différents d'axes (fig. 2.12). On n'observe pas de changement visuel sur le canal $f_{\lambda_{61}}$.

Afin de quantifier les changements pouvant apparaître sur toute l'image on mesure la distance du chi-deux entre les pixels vecteurs de l'image de départ \mathbf{f}_{λ} et ceux de l'image reconstituée $\widehat{\mathbf{f}}_{\lambda}$:

$$d_{\chi^2}^2(\mathbf{f}_{\lambda}(x_i), \widehat{\mathbf{f}}_{\lambda}(x_i)) = \sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{.i}} - \frac{\widehat{\nu}_{ij}}{\nu_{.i}} \right)^2\tag{2.4.9}$$



FIG. 2.12 – Canal 61 de l'image « woman face » avant $f_{\lambda_{61}}$ et après AFC et reconstitution $\hat{f}_{\lambda_{61}}$ à partir de 3 axes factoriels.

En effet d'après les annexes les fréquences marginales de l'image de départ et de l'image reconstituée sont égales : $\nu_i = \hat{\nu}_i$ et $\nu_j = \hat{\nu}_j$ (cf. annexes eq. A.4.50).

On obtient ainsi une carte de distances dont on calcule l'histogramme normalisé à 255 classes (fig. 2.13). On constate que les histogrammes présentent les mêmes modes même si ceux-ci n'ont pas la même amplitude. En outre, les plus fortes distances du chi deux (i.e. les niveaux de gris les plus proches du blanc) sont réparties sur les vêtements de la personne. Ainsi l'erreur de reconstitution est dispersée sur de nombreux pixels. On en conclut que les facteurs pixels des trois premiers axes factoriels sont suffisants pour décrire l'image hyperspectrale. On est ainsi passé de 61 à 3 canaux d'où une réduction de dimension conséquente.

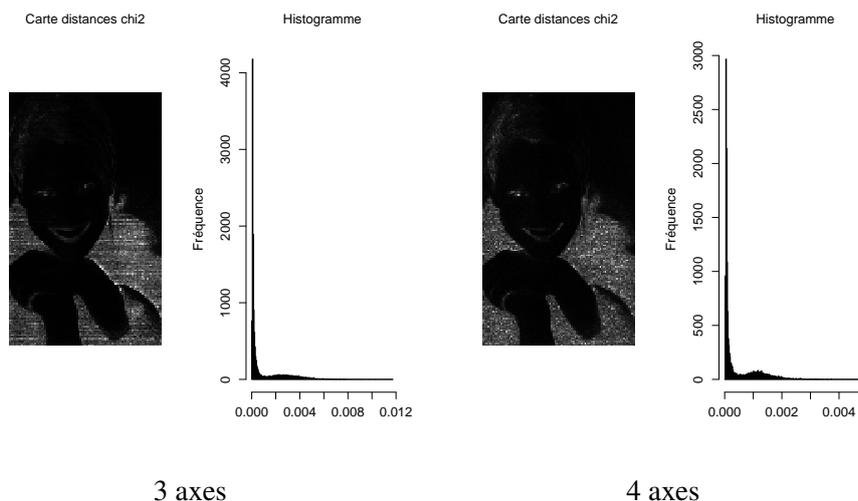


FIG. 2.13 – Cartes des distances du chi deux entre l'image de départ f_{λ} et l'image reconstituée \hat{f}_{λ} avec 3 axes ou avec 4 axes factoriels. Histogrammes de ces cartes.

2.4.1.5 Exemple de la souris

On applique le même procédé sur l'image de souris « serim447 ». Les axes sont sélectionnés par un seuil sur leur rapport signal à bruit. Le seuil est fixé à $s = 0,3$ après observation des facteurs pixels avec l'aide des médecins. Ce seuil sera réutilisé pour d'autres images hyperspectrales de souris acquises avec le même protocole expérimental.

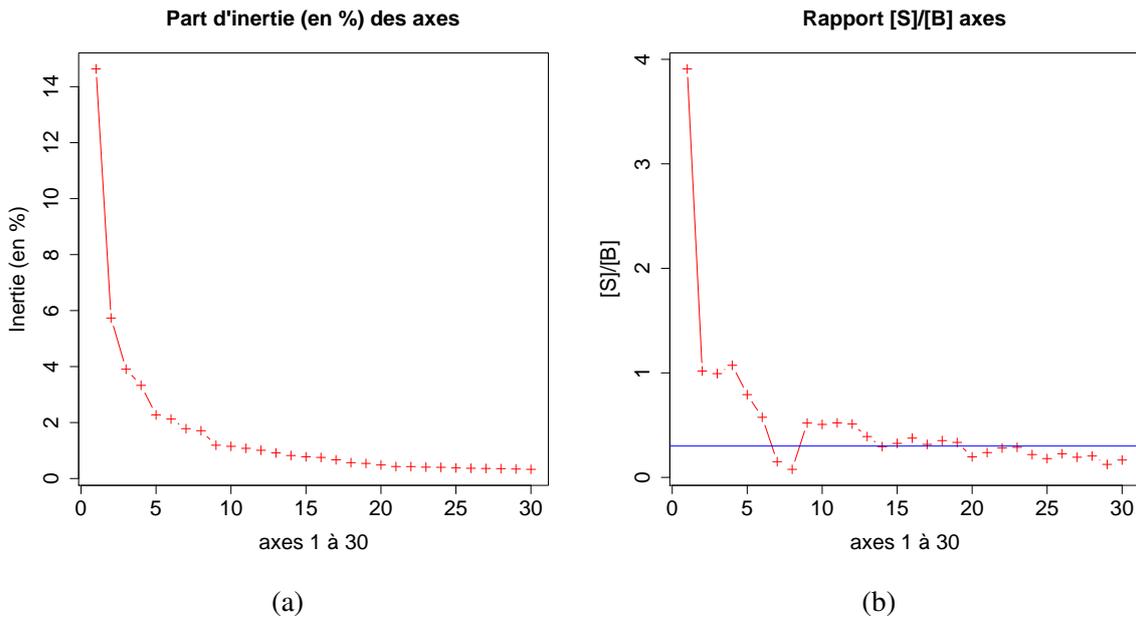


FIG. 2.14 – Part d'inertie (a) et rapport signal à bruit sur les facteurs pixels c_α^f sur les 30 premiers axes de l'image « serim447 ». Le seuil à $s = 0,3$ sur le rapport signal à bruit est indiqué en bleu.

Les facteurs pixels marqués en noir dans le tableau 2.1 et sur la figure 2.15 sont conservés alors que ceux en rouge sont rejetés. On constate que les axes sélectionnés ne sont pas contigus. Ainsi 16 axes, dont les facteurs pixels ont un RSB supérieur à 0,3, sont retenus.

Observons maintenant quelques canaux avant et après reconstitution (fig. 2.16). Sur le canal $f_{\lambda_{12}}$, on note bien l'alternance « sombre » « clair » entre les deux ventricules du cœur. Cette alternance est bien conservée après reconstitution du canal $\hat{f}_{\lambda_{12}}$. De plus, on constate, de visu, que l'image \hat{f}_λ est bien reconstituée et qu'une partie du bruit présent sur l'image de départ f_λ a été éliminé par l'AFC. Le bruit a donc été éliminé par un filtrage spectral dû à l'AFC. Les structures spatiales de l'image ne sont donc pas touchées ce qui est très important pour faire de la segmentation par morphologie mathématique. Nous allons étudier la réduction de bruit dans la partie 2.4.2 (p : 32).

Axes	1	2	3	4	5	6	7	8	9	10
RSB	3,91	1,02	0,99	1,07	0,79	0,58	0,15	0,08	0,52	0,51
Inertie en %	14,63	5,73	3,91	3,33	2,27	2,12	1,78	1,71	1,19	1,15
Axes	11	12	13	14	15	16	17	18	19	
RSB	0,52	0,51	0,39	0,29	0,33	0,38	0,32	0,35	0,33	
Inertie en %	1,08	1,01	0,92	0,82	0,78	0,75	0,67	0,57	0,54	

TAB. 2.1 – Les 16 axes conservés (en noir) et les 3 axes supprimés (en rouge), le RSB de leurs facteurs pixels (>0,3 pour les axes conservés) et leurs parts d'inertie.

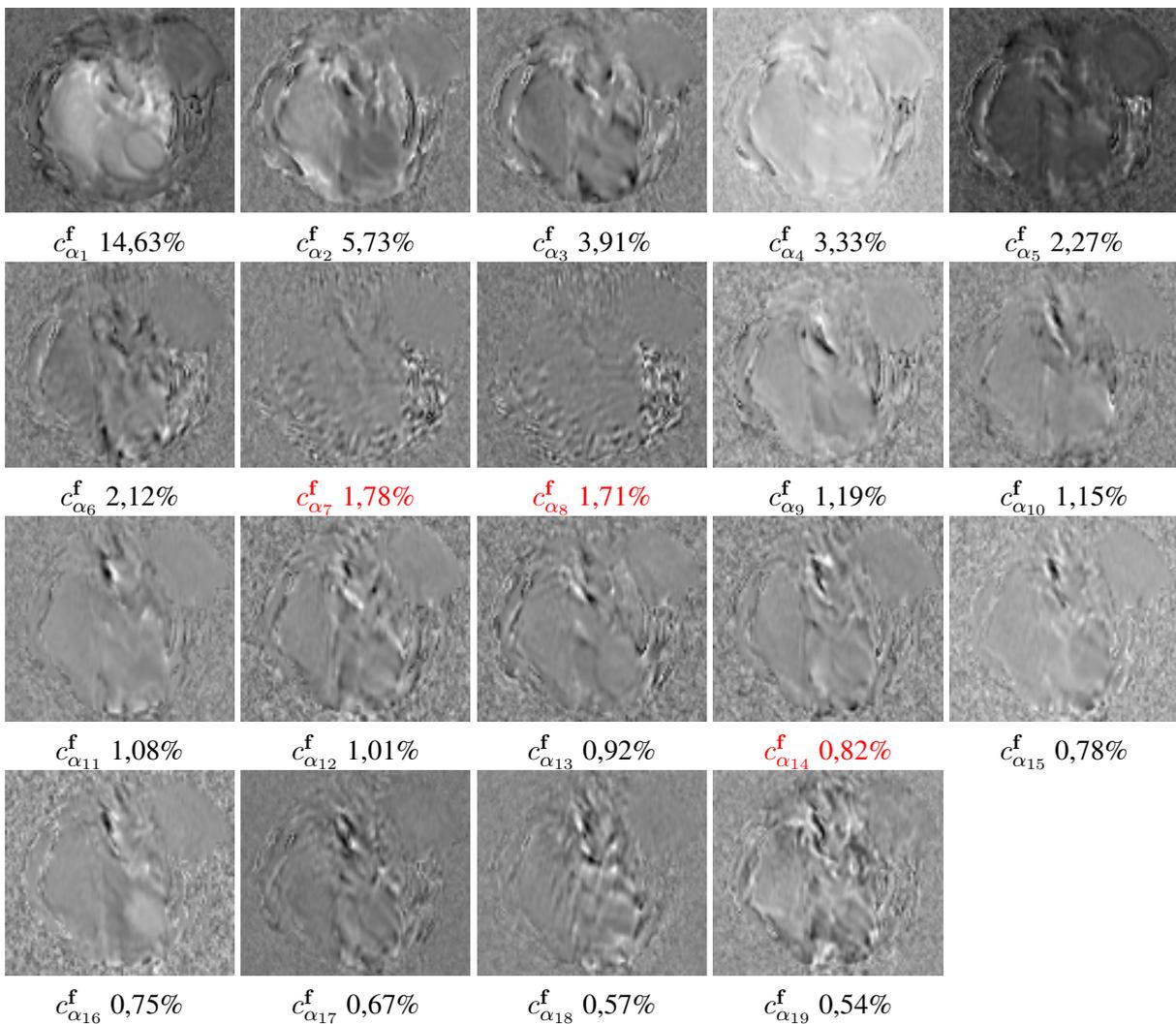


FIG. 2.15 – Les facteurs pixels sur les 19 premiers axes et leurs inerties. Les axes marqués en noirs sont conservés et ceux en rouge sont rejetés.

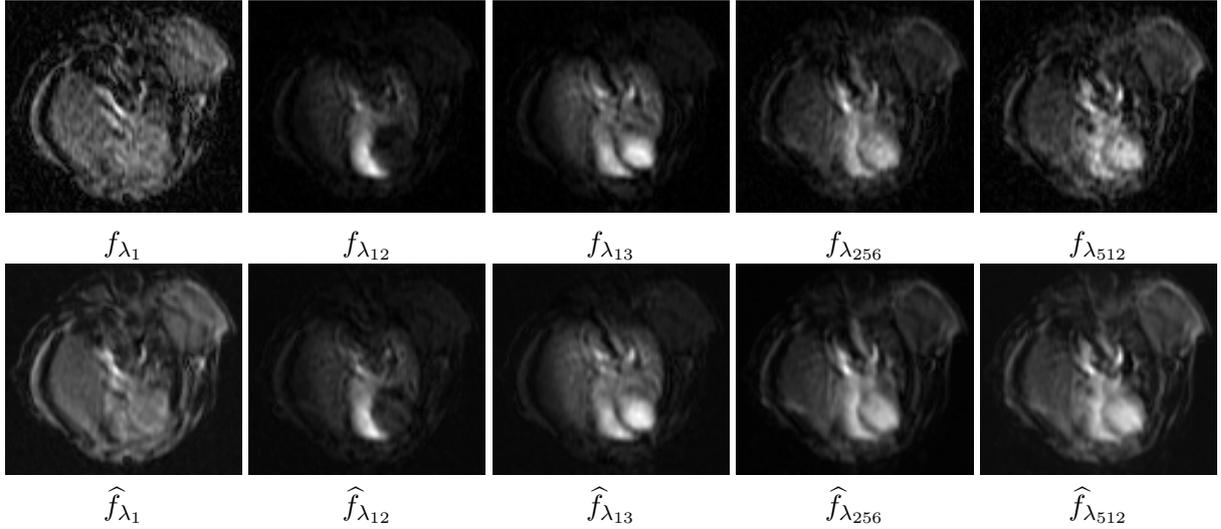


FIG. 2.16 – Cinq canaux de l’image hyperspectrale de souris « serim447 » avant et après reconstitution suite à une AFC à 16 axes. (Un zoom a été fait sur la souris)

2.4.2 Filtrage spectral du bruit par AFC

L’AFC permet d’effectuer un filtrage spectral des données. *Benzécri* [21, 20] et *Orfeuill* [140] ont utilisé cette méthode de filtrage sur des tableaux de « données euclidiennes entachées d’erreur ».

Pour étudier la réduction du bruit spectral, nous avons choisi de bruiteur l’image woman face et de la reconstituer à partir de 3 axes factoriels.

L’image bruitée est notée $\mathbf{f}_{\lambda}^{br}$. On la bruite par un bruit blanc gaussien centré et d’écart type σ multiplié par l’écart type du canal considéré $\sigma(f_{\lambda_j})$: $n_{\lambda_j} = \mathcal{N}(0, \sigma \times \sigma(f_{\lambda_j}))$:

$$f_{\lambda_j}^{br}(x_i) = f_{\lambda_j}(x_i) + \mathcal{N}(0, \sigma \times \sigma(f_{\lambda_j})) \quad (2.4.10)$$

Le rapport signal à bruit théorique pour un canal f_{λ_j} vaut donc :

$$RSB_{\lambda_j} = \frac{\text{var}(\text{signal})}{\text{var}(\text{bruit})} = \frac{\text{var}(f_{\lambda_j})}{\text{var}(n_{\lambda_j})} = \frac{\text{var}(f_{\lambda_j})}{(\sigma \times \sigma(f_{\lambda_j}))^2} = \frac{1}{\sigma^2} \quad (2.4.11)$$

Le rapport signal à bruit RSB_{λ_j} est donc le même pour tous les canaux.

Le rapport signal à bruit hyperspectral entre l’image bruitée $\mathbf{f}_{\lambda}^{br}$ et l’image initiale \mathbf{f}_{λ} est défini comme :

$$RSB_{hyper}(\mathbf{f}_{\lambda}^{br}) = \frac{\sum_{j=1}^L \text{var}(f_{\lambda_j})}{\sum_{j=1}^L \text{var}(f_{\lambda_j}^{br} - f_{\lambda_j})} \quad (2.4.12)$$

Dans le cadre du bruit $n_{\lambda_j} = \mathcal{N}(0, \sigma \times \sigma(f_{\lambda_j}))$, ce RSB hyperspectral vaut $RSB_{hyper}(\mathbf{f}_{\lambda}^{br}) = 1/\sigma^2$.

De même, on définit un rapport signal à bruit hyperspectral de l’image bruitée puis reconstituée $\hat{\mathbf{f}}_{\lambda}^{br}$ par rapport à l’image initiale \mathbf{f}_{λ} comme :

$$RSB_{hyper}(\hat{\mathbf{f}}_{\lambda}^{br}) = \frac{\sum_{j=1}^L \text{var}(f_{\lambda_j})}{\sum_{j=1}^L \text{var}(\hat{f}_{\lambda_j}^{br} - f_{\lambda_j})} \quad (2.4.13)$$

Remarque 2.4.3. L'AFC nécessite un tableau de contingence avec des valeurs positives. Or l'ajout de bruit peut créer des valeurs négatives. On force donc à 0 les valeurs négatives et à 1 les valeurs supérieures à l'unité. Ce seuillage par hystérésis correspond à la dynamique maximale d'un capteur : i.e. sur un capteur 8 bits les niveaux de gris sont nécessairement compris entre 0 et 255. Ce seuillage modifie légèrement les rapports signaux à bruits f_{λ}^{br} de l'image bruitée. Ces RSB restent néanmoins quasiment identiques sur tous les canaux.

Choisissons un bruit tel que $\sigma = 0,5$. Le rapport signal à bruit sur chaque canal vaut en théorie $RSB_{\lambda_j} = \frac{1}{\sigma^2} = 4$, mais du fait du seuillage il vaut en moyenne 4,68. De même le rapport signal à bruit hyperspectral de l'image bruitée vaut en théorie $RSB_{hyperspectral}(f_{\lambda}^{br}) = 4$ mais en fait 4,63.

Appliquons maintenant l'AFC sur l'image bruitée f_{λ}^{br} et reconstituons l'image $\widehat{f_{\lambda}^{br}}$.

Sur la figure 2.17, les facteurs pixels sur les 3 premiers axes sont bruités. Les facteurs pixels du troisième axe ne contiennent quasiment que du bruit. Les canaux 38 et 61 de l'image bruitée puis reconstituée $\widehat{f_{\lambda_{38}}^{br}}$, $\widehat{f_{\lambda_{61}}^{br}}$ sont, de visu, moins bruités que les canaux 38 et 61 de l'image bruitée $f_{\lambda_{38}}^{br}$, $f_{\lambda_{61}}^{br}$. Il y a donc bien eu un débruitage des spectres par l'AFC.

Le rapport signal à bruit hyperspectral de l'image bruitée puis reconstituée vaut $RSB_{hyperspectral}(\widehat{f_{\lambda}^{br}}) = 60$. D'où une amélioration d'un facteur 13 du RSB hyperspectral (i.e. 1200 %). Cependant même si globalement tous les axes ont été améliorés, certains ont un meilleur rapport signal à bruit que d'autres après l'AFC (fig. 2.18).

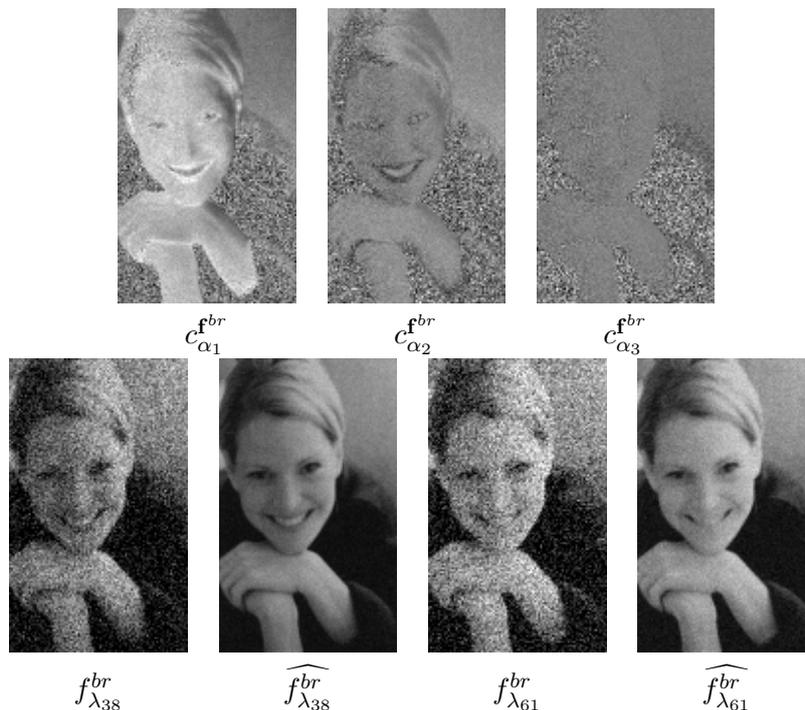


FIG. 2.17 – Facteurs pixels c_{α}^{br} de l'image bruitée f_{λ}^{br} , canaux 38 et 61 de l'image bruitée $f_{\lambda_{38}}^{br}$, $f_{\lambda_{61}}^{br}$ et canaux 38 et 61 de l'image bruitée et reconstituée $\widehat{f_{\lambda_{38}}^{br}}$, $\widehat{f_{\lambda_{61}}^{br}}$.

L'influence de l'AFC sur les rapports signaux à bruits hyperspectraux, de l'image bruitée $RSB_{hyperspectral}(f_{\lambda}^{br})$ puis reconstituée par AFC $RSB_{hyperspectral}(\widehat{f_{\lambda}^{br}})$, est étudiée en fonction du coefficient multiplicateur de

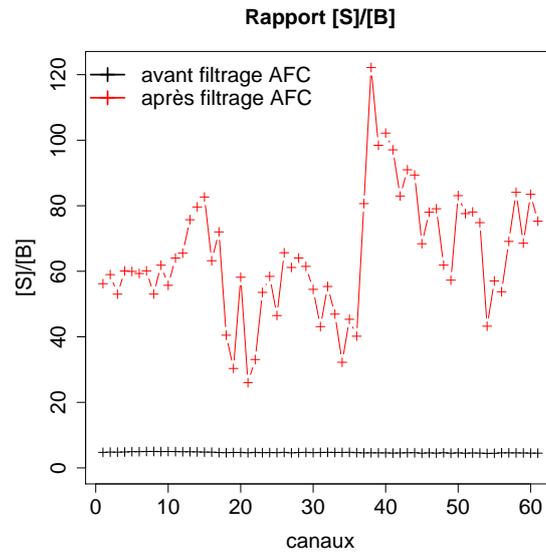


FIG. 2.18 – RSB sur les canaux de l'image bruitée \mathbf{f}_λ^{br} (en noir) et de l'image bruitée puis reconstituée $\widehat{\mathbf{f}}_\lambda^{br}$ (en rouge) par AFC à 3 axes.

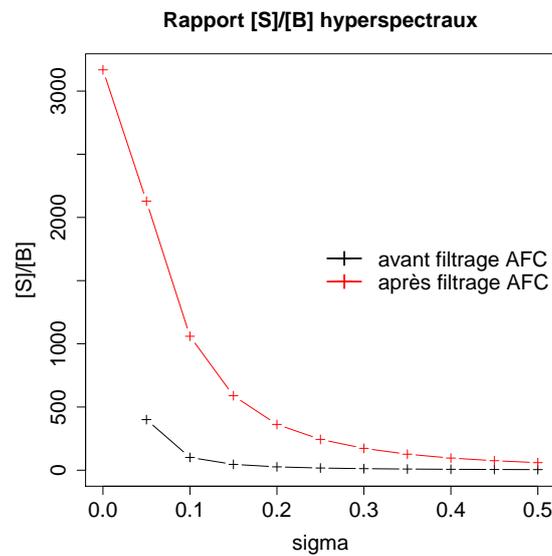


FIG. 2.19 – RSB hyperspectraux de l'image bruitée $RSB_{hyper}(\mathbf{f}_\lambda^{br})$ (en noir) puis reconstituée par AFC à 3 axes $RSB_{hyper}(\widehat{\mathbf{f}}_\lambda^{br})$ (en rouge) en fonction du coefficient multiplicateur de l'écart type du bruit σ (sigma). Pour $\sigma = 0$, $RSB_{hyper}(\mathbf{f}_\lambda^{br}) \rightarrow \infty$ (en noir) est égale : l'image n'est pas bruitée.

l'écart type du bruit σ (fig. 2.19). On constate que les RSB augmentent après reconstitution de l'image par AFC. L'AFC a donc bien une fonction de filtrage spectral des données. Pour $\sigma = 0$, $RSB_{hyper}(\mathbf{f}_\lambda^{br}) \rightarrow \infty$ car l'image n'est pas bruitée. On mesure donc le RSB entre l'image non bruitée

\mathbf{f}_λ et elle même.

2.4.3 Filtrage spectral du bruit par ACP

On veut vérifier comment se comporte l'ACP vis-à-vis de la réduction de bruit avec le même protocole que présenté pour l'AFC dans la partie 2.4.2 (p : 32 et suivantes).

Comme pour l'AFC, l'influence de l'ACP sur les RSB hyperspectraux, de l'image bruitée $RSB_{hyper}(\mathbf{f}_\lambda^{br})$ puis reconstituée par ACP $RSB_{hyper}(\widehat{\mathbf{f}}_\lambda^{br})$, est étudiée en fonction du coefficient multiplicateur de l'écart type du bruit σ (fig. 2.20). On constate que les RSB augmentent après reconstitution de l'image par ACP. L'ACP a donc, comme l'AFC, une fonction de filtrage spectral des données.

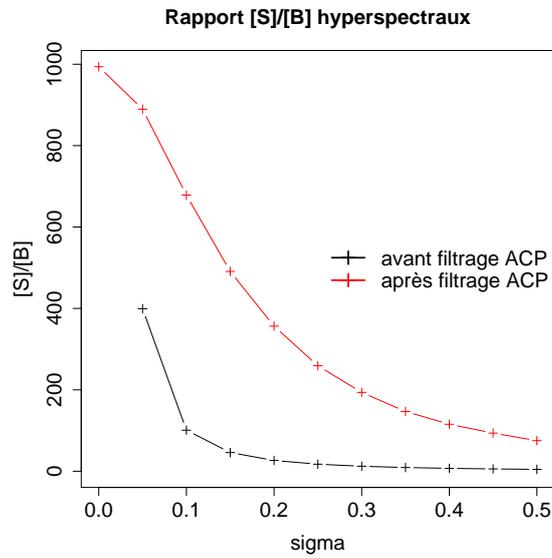


FIG. 2.20 – RSB hyperspectraux de l'image bruitée $RSB_{hyper}(\mathbf{f}_\lambda^{br})$ (en noir) puis reconstituée par ACP à 3 axes $RSB_{hyper}(\widehat{\mathbf{f}}_\lambda^{br})$ (en rouge) en fonction du coefficient multiplicateur de l'écart type du bruit σ (sigma). Pour $\sigma = 0$, $RSB_{hyper}(\mathbf{f}_\lambda^{br}) \rightarrow \infty$ (en noir) est égale : l'image n'est pas bruitée.

2.4.4 Composition d'AFC et de reconstitutions

Dans cette partie, nous démontrons l'impact de deux AFC et reconstitutions $\widehat{\zeta}^{-1} \circ \zeta$ successives avec une translation des données t_ε (i.e. ajout d'une constante) pour la deuxième AFC et sa reconstitution et enfin une soustraction de cette constante $t_{-\varepsilon}$:

$$\widehat{\mathbf{f}}_\lambda^{(2)} = t_\varepsilon \circ \widehat{\zeta}^{-1} \circ \zeta \circ t_{-\varepsilon} \circ \widehat{\zeta}^{-1} \circ \zeta(\mathbf{f}_\lambda) \quad (2.4.14)$$

La translation t_ε s'exprime sous la forme :

$$t_\varepsilon : \begin{cases} \mathcal{T}^L & \rightarrow \mathcal{T}^L \\ \mathbf{f}_\lambda & \rightarrow t_\varepsilon(\mathbf{f}_\lambda) \text{ tel que } \forall i = 1 \dots P, \forall j = 1 \dots L \quad t_\varepsilon(\mathbf{f}_{\lambda_j})(x_i) = \mathbf{f}_{\lambda_j}(x_i) + \varepsilon \end{cases} \quad \varepsilon \in \mathbb{R} \quad (2.4.15)$$

La seconde AFC, avec reconstitution des données, n'est donc pas identique à la première car les données sont légèrement modifiées par la translation.

2.4.4.1 Mise en évidence du phénomène

D'où vient cette idée un peu étrange d'ajouter une constante. Lorsque nous avons fait la reconstitution des données avec la souris « serim447 », nous avons essayé de classifier les données dans l'espace factoriel de l'AFC 1 $\mathbf{c}_\alpha^{\mathbf{f}^{(1)}}$ par *kmeans* (cf. partie 2.6, p : 57). Le résultat n'étant pas satisfaisant, l'idée d'essayer un deuxième filtrage par AFC est apparue. Comme l'image reconstituée $\hat{\mathbf{f}}_\lambda^{(1)}$ après $\hat{\zeta}^{-1} \circ \zeta$ présentait des valeurs négatives, il nous est venu l'idée d'ajouter une constante qui correspondait à l'opposé du minimum de l'image. Ceci ayant pour but d'obtenir des valeurs positives de l'image afin d'appliquer une AFC et sa reconstitution. Enfin, la constante est retranchée à l'image reconstituée pour donner l'image $\hat{\mathbf{f}}_\lambda^{(2)}$. Finalement la classification dans le nouvel espace factoriel, de l'AFC 2, $\mathbf{c}_\alpha^{\mathbf{f}^{(2)}}$ s'est avérée meilleure (fig. 2.21).

Par conséquent, la composition de deux AFC avec reconstitution semble faire diminuer le bruit dans l'espace factoriel. Il reste maintenant à le montrer.

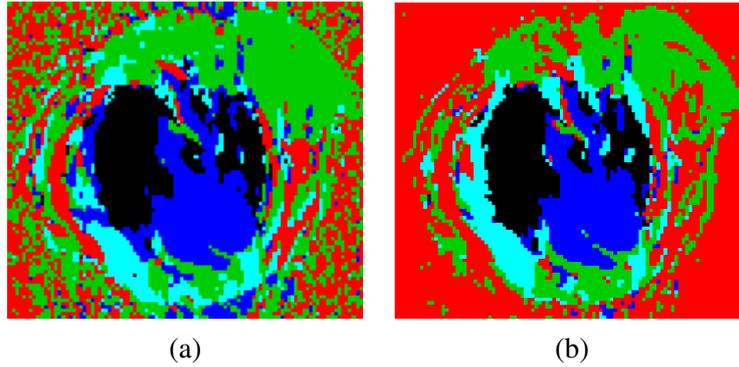


FIG. 2.21 – Classifications *kmeans* en 5 classes dans l'espace factoriel (a) de l'AFC 1 $\mathbf{c}_\alpha^{\mathbf{f}^{(1)}}$ et (b) de l'AFC 2 $\mathbf{c}_\alpha^{\mathbf{f}^{(2)}}$.

2.4.4.2 Étude de la composition de deux AFC-reconstitutions

Effectuons maintenant la composition de deux AFC. On choisit comme valeur pour la translation :

$$\varepsilon = \min_{i,j} f_{\lambda_j}(x_i) \quad i = 1 \dots P \quad j = 1 \dots L \quad (2.4.16)$$

Comme on a défini l'image reconstituée après 2 AFC-reconstitutions $\hat{\mathbf{f}}_\lambda^{(2)}$, on définit l'image reconstituée après 3 AFC-reconstitutions $\hat{\mathbf{f}}_\lambda^{(3)}$:

$$\hat{\mathbf{f}}_\lambda^{(3)} = t_\varepsilon \circ \hat{\zeta}^{-1} \circ \zeta \circ t_{-\varepsilon}(\hat{\mathbf{f}}_\lambda^{(2)})$$

Le phénomène va être étudié sur l'image « woman face » que l'on a bruitée avec un bruit d'écart type σ constant $n_{\lambda_j} = \mathcal{N}(0, \sigma)$ (eq. 2.4.10). Les canaux n'ont donc par un RSB constant, à la différence de la partie 2.4.2 (p : 32).

2.4.4.2.1 Inertie des axes Sur la figure 2.22, l'inertie expliquée par le sous espace F_3 engendré par les 3 premiers axes a tendance à diminuer avec l'augmentation du bruit (i.e. l'augmentation de σ). Or le bruit fait aussi partie de l'inertie des données. Lorsqu'il augmente, la part exprimée par les 3 premiers axes, qui correspondent surtout au signal, a tendance à diminuer. Par conséquent, le bruit a tendance à être rejeté sur les autres axes et donc influence l'inertie du sous-espace engendré par ces derniers axes.

A partir de la deuxième itération l'inertie expliquée est proche de 100 %. En effet comme la reconstitution après l'AFC 1 s'est faite avec 3 axes, l'image hyperspectrale $\widehat{\mathbf{f}}_\lambda^{br(1)}$ est dans un espace dégénéré à 3 dimensions.

Ainsi seule la première AFC fixe la quantité d'inertie expliquée par les axes sélectionnés.

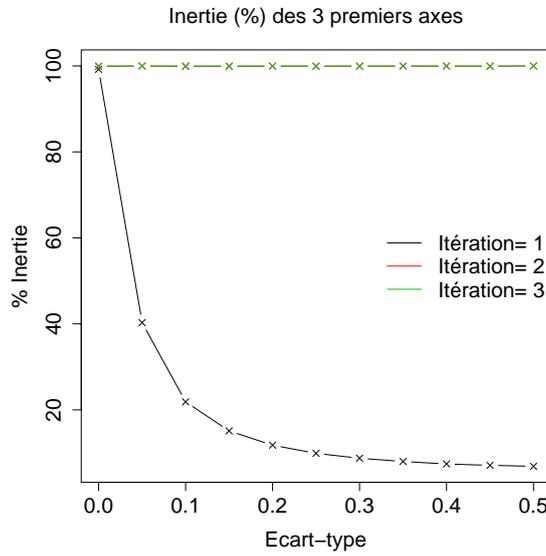


FIG. 2.22 – Part d'inertie (en %) expliquée par les 3 premiers axes de chaque AFC en fonction de l'écart type du bruit σ pour différentes itérations d'AFC (3 axes).

2.4.4.2.2 RSB des facteurs pixels Mesurons les RSB des facteurs pixels des images bruitées au cours des itérations d'AFC $\mathbf{c}_\alpha^{f^{br(n)}}$ par rapport aux facteurs pixels de l'image non bruitée \mathbf{c}_α^f :

$$RSB_{\alpha_k} = \frac{\text{var}(\text{signal})}{\text{var}(\text{bruit})} = \frac{\text{var}(\mathbf{c}_{\alpha_k}^f)}{\text{var}(n_{\alpha_k})} = \frac{\text{var}(\mathbf{c}_{\alpha_k}^f)}{\text{var}(\mathbf{c}_{\alpha_k}^{f^{br(n)}} - \mathbf{c}_{\alpha_k}^f)} \quad (2.4.17)$$

Remarque 2.4.4. Attention il y a un décalage d'indice entre les facteurs pixels et le numéro de l'AFC : pour la première AFC les facteurs pixels sont notés $\mathbf{c}_\alpha^{f^{br(0)}}$ car ils sont associés à l'image de départ $\mathbf{f}_\lambda^{br(0)}$. Après la première reconstitution on obtient l'image $\widehat{\mathbf{f}}_\lambda^{br(1)}$ dont les facteurs pixels sont obtenus par la deuxième AFC $\mathbf{c}_\alpha^{f^{br(1)}}$.

Observations : Sur la figure 2.25 pour l'image bruitée ($\sigma > 0$) le rapport signal sur bruit augmente lors de la deuxième AFC puis se stabilise. Pour l'image non bruitée ($\sigma = 0$) le rapport signal sur bruit

diminue avec les itérations. A la première itération et pour l'image non bruitée $\sigma = 0$, le RSB des facteurs pixels tend vers l'infini car on calcule leurs RSB par rapport à eux-mêmes : i.e. $c_{\alpha}^{f^{br(0)}} = c_{\alpha}^f$ pour $\sigma = 0$. Pour l'image bruitée $\sigma > 0$, les RSB convergent à la deuxième itération d'AFC.

Interprétations : Itérer une AFC sur une image sans bruit f_{λ} réduit l'information et fait augmenter l'erreur sur ses facteurs pixels c_{α}^f . Itérer une AFC sur une image bruitée f_{λ}^{br} fait diminuer le bruit sur ses facteurs pixels. Les RSB atteignent un point fixe à la deuxième itération d'AFC-reconstitution.

Conclusions : Pour obtenir des facteurs pixels avec le moins de bruit possible :

- sur une image sans bruit : prendre les facteurs c_{α}^f pixels après une AFC ;
- sur une image bruitée : faire deux AFC-reconstitutions, pour obtenir un point fixe du RSB sur les facteurs pixels, et prendre les facteurs pixels $c_{\alpha}^{f^{br(2)}}$.

Sur la figure 2.24, on peut vérifier de visu ces observations. Le bruit sur les facteurs pixels de l'image non bruitée $c_{\alpha}^{f^{(n)}}$ augmente avec le nombre d'itérations alors que le bruit sur les facteurs pixels de l'image bruitée $c_{\alpha}^{f^{br(n)}}$ diminue avec le nombre d'itérations.

Remarque 2.4.5. On peut se demander pourquoi pour filtrer l'image bruitée une seule AFC suffit alors que pour filtrer les facteurs pixels de cette même image bruitée il faut faire plusieurs AFC. En fait le bruit sur les facteurs pixels a tendance à avoir moins d'importance lors de la reconstitution de l'image puisque chaque facteur pixel est pondéré par sa valeur propre. En revanche pour éliminer plus de bruit sur les facteurs pixels il faut itérer l'AFC.

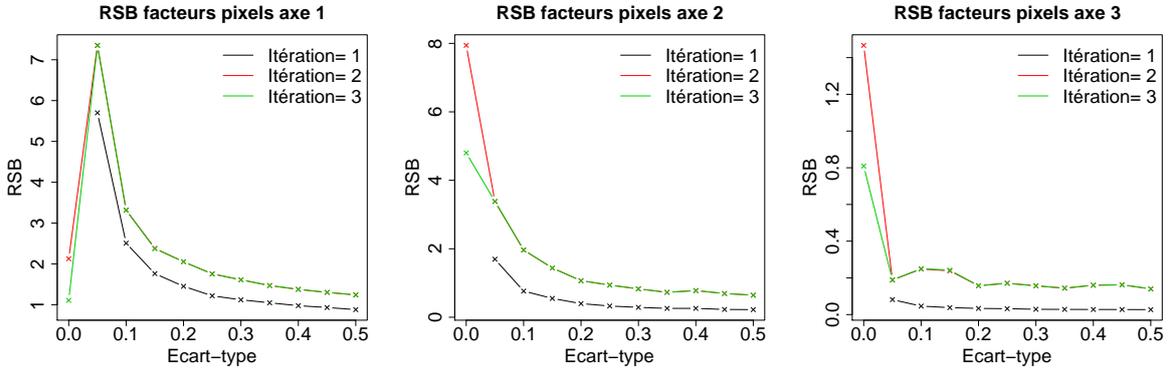


FIG. 2.23 – RSB sur les facteurs pixels de l'image bruitée $c_{\alpha}^{f^{br(n)}}$ en fonction de l'écart type σ du bruit pour différentes itérations d'AFC (3 axes).

2.4.4.2.3 Influence dans l'espace image On mesure le RSB hyperspectral dans l'espace image $\widehat{f_{\lambda}^{br(n)}}$ pour différentes itérations d'AFC-reconstitution par rapport à l'image sans bruit f_{λ} (eq. 2.4.13).

Sur la figure 2.25, on constate que le RSB augmente entre l'image sans bruit (itération 0) et l'image bruitée puis reconstituée par une AFC $\widehat{f_{\lambda}^{br(1)}}$ (itération 1). Il est presque constant dès la deuxième itération sauf pour l'image non bruitée f_{λ} ($\sigma = 0$) où il décroît.

Remarque 2.4.6. Pour un écart type de 0, à l'itération 0, le RSB est égal à l'infini car il est calculé entre l'image non bruitée f_{λ} et elle-même.

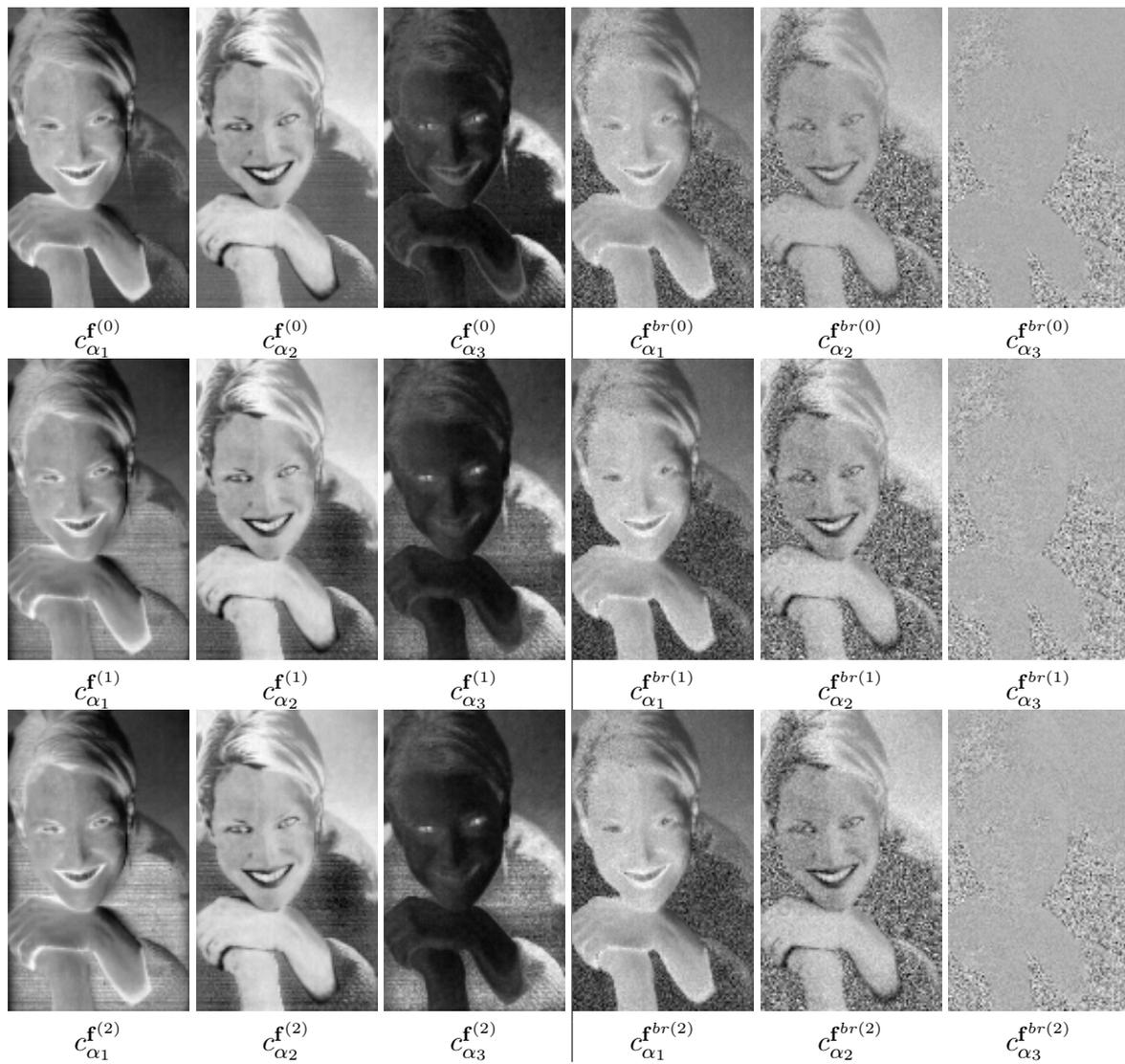


FIG. 2.24 – Facteurs pixels pour différentes itérations de l'AFC sur l'image sans bruit f_λ et sur l'image bruitée f_λ^{br} ($\sigma = 0,05$).

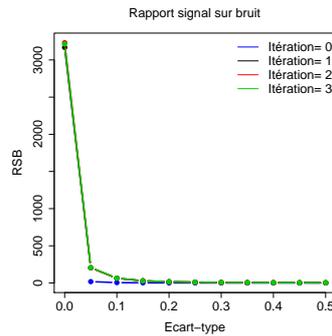


FIG. 2.25 – RSB hyperspectral de l'image bruitée $\widehat{\mathbf{f}}_{\lambda}^{br(n)}$ en fonction de l'écart type σ du bruit pour différentes itérations d'AFC-reconstitution (3 axes).

Seule la première AFC-reconstitution permet de réduire le bruit. Au delà il n'y a aucun apport des AFC-reconstitutions pour réduire le bruit. Pour une image sans bruit \mathbf{f}_{λ} , l'AFC introduit une erreur du fait que l'on ne conserve pas tous les axes. A chaque itération de l'information est ainsi perdue.

Par conséquent une seule AFC-reconstitution suffit à débruiter l'image bruitée $\mathbf{f}_{\lambda}^{br}$ (cf. remarque 2.4.5, p : 38).

2.4.4.3 Conclusion sur la composition d'AFC et de reconstitutions

Il apparaît ainsi que la composition d'AFC-reconstitutions permet de réduire le bruit sur une image hyperspectrale. L'étude approfondie sur une image que l'on a bruitée a permis la mise en évidence de plusieurs résultats intéressants :

- Pour une image hyperspectrale bruitée : faire une AFC afin de réduire la dimension et le bruit spectral.
- Pour une image hyperspectrale non bruitée : faire une AFC pour réduire la dimension.
- Pour obtenir des facteurs pixels avec le moins de bruit possible :
 - sur une image sans bruit : prendre les facteurs pixels après une AFC ;
 - sur une image bruitée : faire une composition d'une AFC avec une AFC-reconstitution et prendre les facteurs pixels $\mathbf{c}_{\alpha}^{f(2)}$.

Pourquoi pour filtrer l'image bruitée une seule AFC suffit alors que pour filtrer les facteurs pixels de cette même image bruitée il faut faire deux AFC ? En fait le bruit sur les facteurs pixels a tendance à avoir moins d'importance lors de la reconstitution de l'image puisque chacun des facteurs pixels est pondéré par sa valeur propre. En outre la reconstitution de l'image dépend du produit des fréquences marginales ν_i, ν_j qui correspond à la reconstitution du centre de gravité ; ce centre donne l'apparence générale de l'image. En revanche pour éliminer plus de bruit sur les facteurs pixels il faut itérer l'AFC.

En outre pour une image bruitée si le nombre d'axes décroît selon les nombre d'itérations, jusqu'à ne garder que les axes contenant de l'information, le RSB s'améliore. Le bruit est ainsi mieux filtré. Le cas idéal est donc de filtrer une image bruitée par composition d'AFC-reconstitutions avec un nombre d'axes optimal dès la première itération. Si ce nombre d'axe est trop important lors la première itération, il faut le réduire à l'itération suivante.

2.4.5 Application de la composition d'AFC-reconstitutions

Afin de débruiter l'image IRM de souris « serim447 » et l'image « bouchon en plastique », on utilise la composition de deux AFC-reconstitutions.

2.4.5.1 Image de souris

A partir des 16 axes factoriels retenus lors de la première AFC (cf. partie 2.4.1.5, p : 30), on reconstitue l'image et on effectue une deuxième AFC-reconstitutions avec translation (eq. 2.4.14 p : 35).

Afin de vérifier l'utilité de 2 AFC-reconstitutions, les RSB sont estimés par ouverture sur les canaux de l'image initiale f_λ , reconstituée après 1 AFC $\hat{f}_\lambda^{(1)}$ et reconstituée après 2 AFC $\hat{f}_\lambda^{(2)}$. Les RSB sont également estimés sur les facteurs pixels de la première AFC $c_\alpha^{f(0)}$ et de la deuxième AFC $c_\alpha^{f(1)}$ (fig. 2.26). On observe bien une amélioration du RSB dans l'espace image après la première AFC. En revanche la deuxième AFC, n'apporte pas d'amélioration dans l'espace image. Dans l'espace factoriel, le RSB sur les facteur pixels s'améliore à la deuxième itération d'AFC-reconstitution, par rapport à la première. On retrouve bien le résultat de la partie 2.4.4.3 (p : 40).

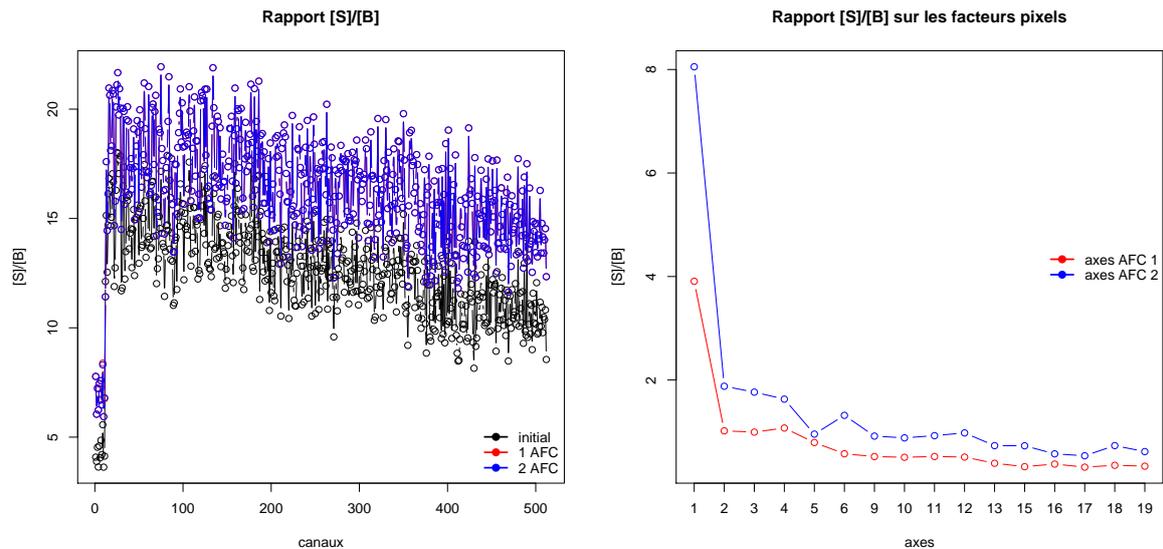


FIG. 2.26 – RSB sur les canaux de l'image et les facteurs pixels pour les différentes itérations d'AFC-reconstitutions à 16 axes.

Remarque 2.4.7. Lors de la deuxième AFC, l'espace image est un espace dégénéré à 16 dimensions, il y a donc 16 axes factoriels. Cependant, afin d'avoir des numéros d'axes de la deuxième AFC comparables à ceux la première AFC, on conserve les numéros des axes de la première AFC. Comme on a supprimé, à la première AFC, les facteurs pixels associés aux axes 7 et 8, $c_{\alpha_7}^{f(0)}$ et $c_{\alpha_8}^{f(0)}$, les facteurs pixels associés au septième axe de la deuxième AFC sont notés $c_{\alpha_9}^{f(0)}$. Lorsque l'on parle de l'inertie des axes, on parle de l'inertie des axes de la première AFC qui est la seule à avoir un sens par rapport

à l'ensemble de l'espace image. En effet, lors de la deuxième AFC la somme des inerties des axes, conservés à la première AFC, est égale à 100% de la part d'inertie.

Observons les facteurs pixels lors de la deuxième AFC (fig. 2.27).

Après 2 AFC avec sélection du nombre d'axes par seuillage du RSB sur les facteurs pixels, on calcule la valeur absolue du résidu, canal par canal, entre l'image de départ \mathbf{f}_λ et l'image reconstituée après 2 AFC-reconstitutions $\widehat{\mathbf{f}}_\lambda = \widehat{\mathbf{f}}_\lambda^{(2)}$:

$$\forall i \in [1; 2; \dots; P], \forall j \in [1; 2; \dots; L] \quad r_{\lambda_j}(x_i) = |f_{\lambda_j}(x_i) - \widehat{f}_{\lambda_j}(x_i)| \quad (2.4.18)$$

On peut donc définir une image hyperspectrale de résidus \mathbf{r}_λ :

$$\mathbf{r}_\lambda = |\mathbf{f}_\lambda - \widehat{\mathbf{f}}_\lambda| \quad (2.4.19)$$

dont les canaux r_{λ_j} sont définis par :

$$\forall j \in [1; 2; \dots; L] \quad r_{\lambda_j} = |f_{\lambda_j} - \widehat{f}_{\lambda_j}|. \quad (2.4.20)$$

Observons quelques canaux reconstitués après 2 AFC-reconstitutions (fig. 2.28) ainsi que les images de résidus associés. On constate que du bruit a bien été retiré. Une discussion avec les médecins, a montré qu'il s'agit d'un bruit lié à l'acquisition et qui est surtout présent au centre de l'image.

Les covariances spatiales centrées, $\bar{g}_{r_{\lambda_j}}$, sur les canaux de ces résidus sont aussi mesurées :

$$\forall j \in [1; 2; \dots; L] \quad \bar{g}_{r_{\lambda_j}} = E[\bar{r}_{\lambda_j} \bar{r}_{\lambda_j}] \quad (2.4.21)$$

avec $\bar{r}_{\lambda_j} = \frac{1}{P} \sum_{i=1}^P r_{\lambda_j}(x_i)$.

On remarque que les covariances centrées $\bar{g}_{r_{\lambda_j}}$ présentent bien un pic à l'origine ce qui signifie que les résidus correspondent à du bruit et non à du signal (fig. 2.28).

Afin d'observer, l'influence du débruitage dans le domaine spectral, on trace les spectres $\mathbf{f}_\alpha(x_i)$ en quelques points (fig. 2.29). Les spectres filtrés par deux AFC-reconstitutions $\widehat{\mathbf{f}}_\lambda^{(2)}(x_i)$ ont une variabilité plus faible que les spectres non filtrés, en outre, ils permettent de mettre plus facilement en évidence la tendance générale des spectres. Ainsi pour les spectres correspondant à la tumeur, i.e. les points (22,81) et (30,90), la cinétique du signal est à l'augmentation. Ceci correspond à l'accumulation du produit radioactif dans la tumeur.

Remarque 2.4.8. Dans la suite de ce travail, nous utiliserons la notation $\widehat{\mathbf{f}}_\lambda$ pour les images reconstituées après composition de deux AFC-reconstitutions : $\widehat{\mathbf{f}}_\lambda = \widehat{\mathbf{f}}_\lambda^{(2)}$.

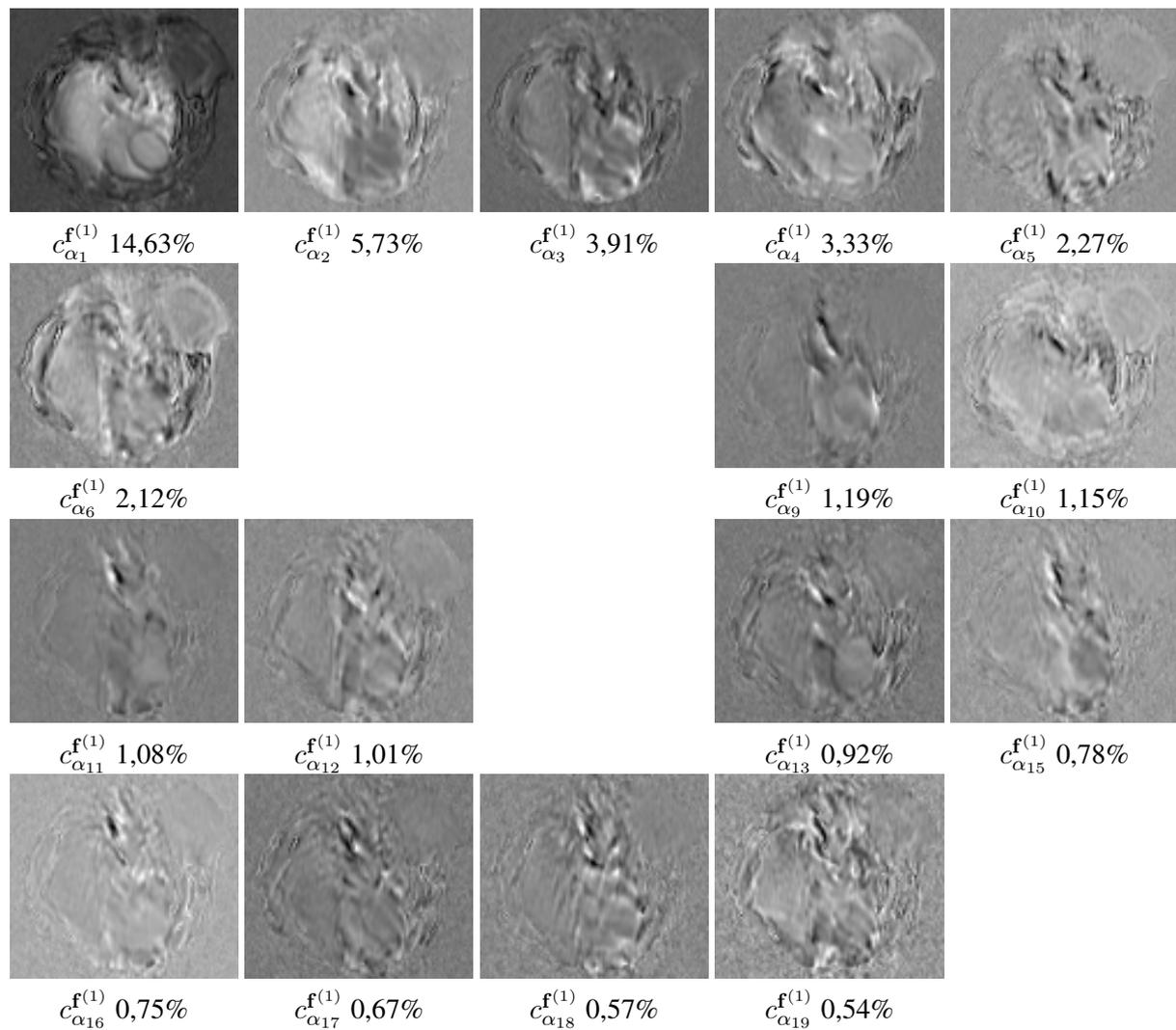


FIG. 2.27 – Les facteurs pixels sur les 16 axes de la deuxième AFC et les inerties des axes.

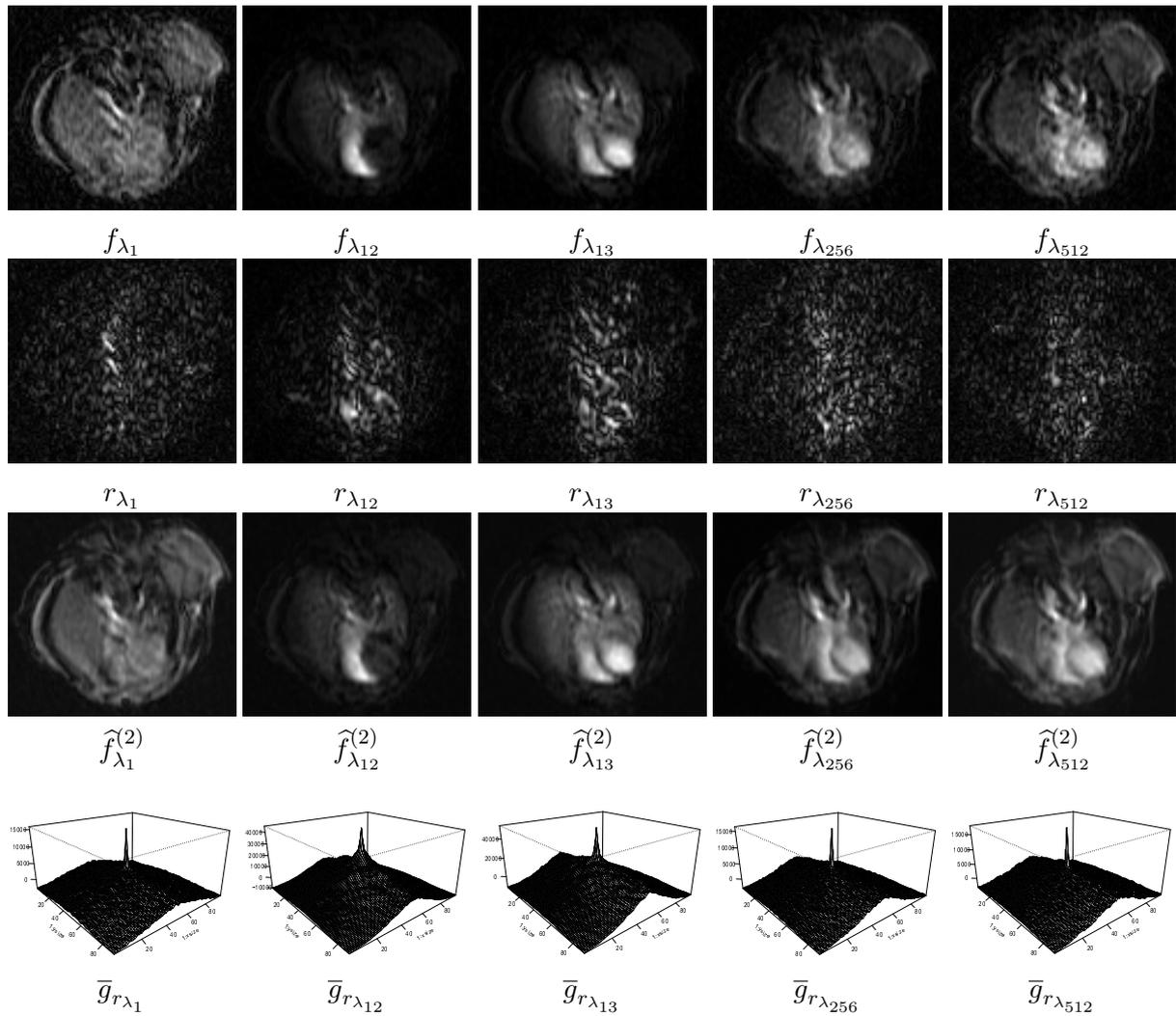


FIG. 2.28 – Cinq canaux de l'image hyperspectrale de souris « serim447 » avant \mathbf{F} et après 2 AFC-reconstitutions à 16 axes $\hat{\mathbf{f}}_{\lambda}^{(2)}$, leurs résidus \mathbf{r}_{λ} et leurs covariances $\bar{g}_{r_{\lambda}}$. Pour des questions de visualisation la dynamique de niveaux de gris des résidus à été étirée entre 0 et 255.

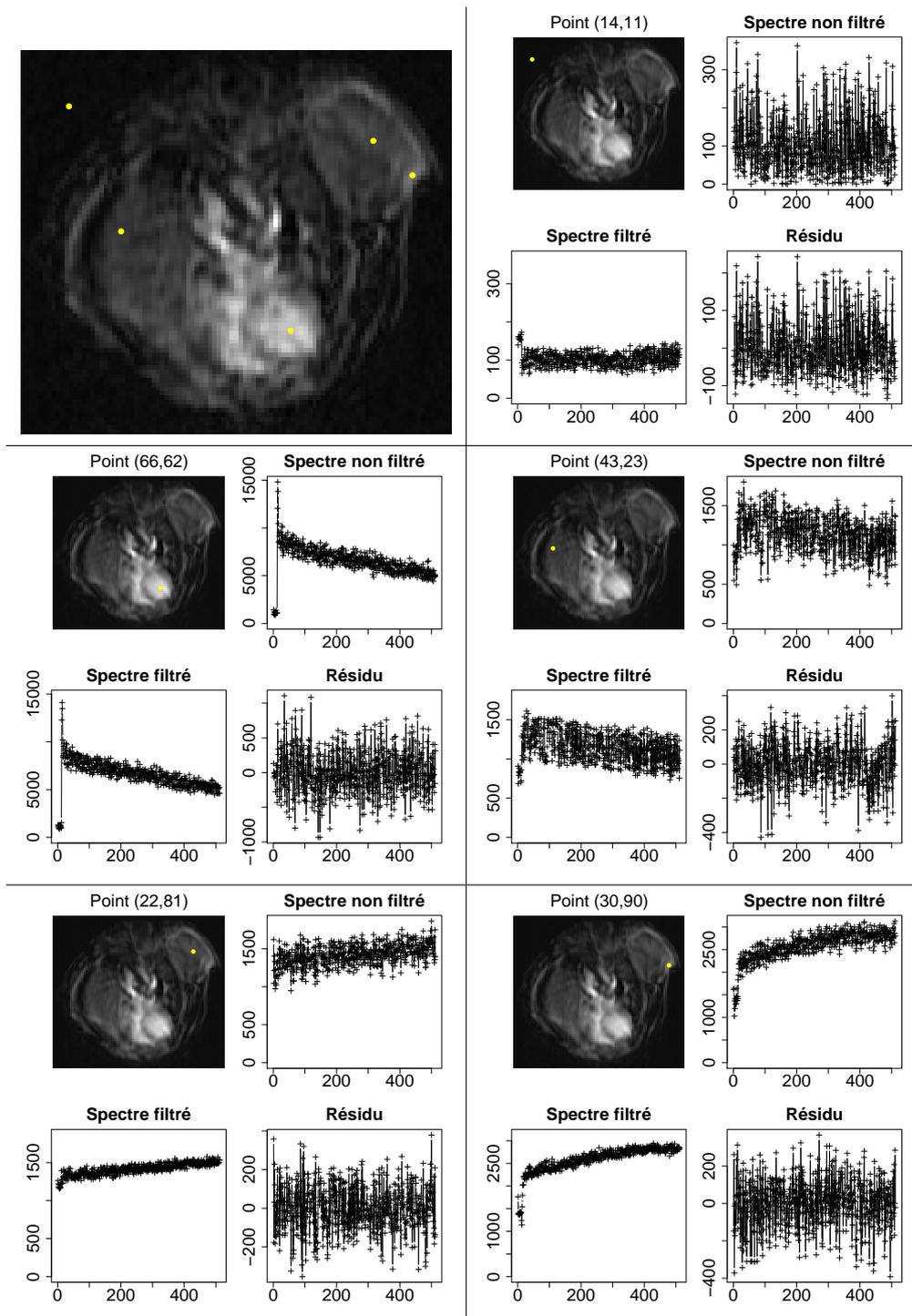


FIG. 2.29 – Mesures de 5 spectres avant et après 2 AFC-reconstructions à 16 axes ainsi que des résidus associés.

2.4.5.2 Image du bouchon en plastique

Sur l'image « bouchon en plastique », on effectue la composition deux AFC-reconstitutions. Par observation des facteurs pixels et de leurs RSB et inerties on ne retient que deux axes factoriels (fig. 2.30).

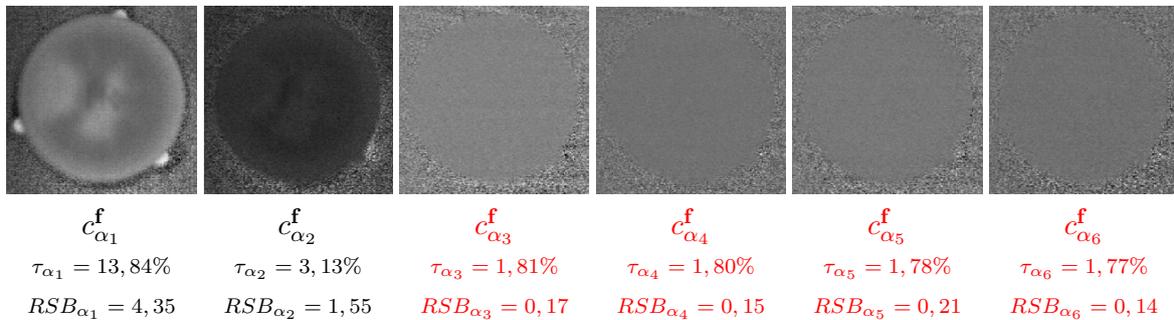


FIG. 2.30 – Facteurs pixels sur les 6 premiers axes factoriels de l'image bouchon en plastique, leurs inerties τ_{α} et leurs RSB RSB_{α} . Les facteurs pixels associés aux axes rejetés sont notés en rouge.

On observe l'effet du débruitage spectral dans le domaine spatial sur un canal et dans le domaine spectral. On observe que le bruit a été réduit dans le domaine spectral (fig. 2.31).

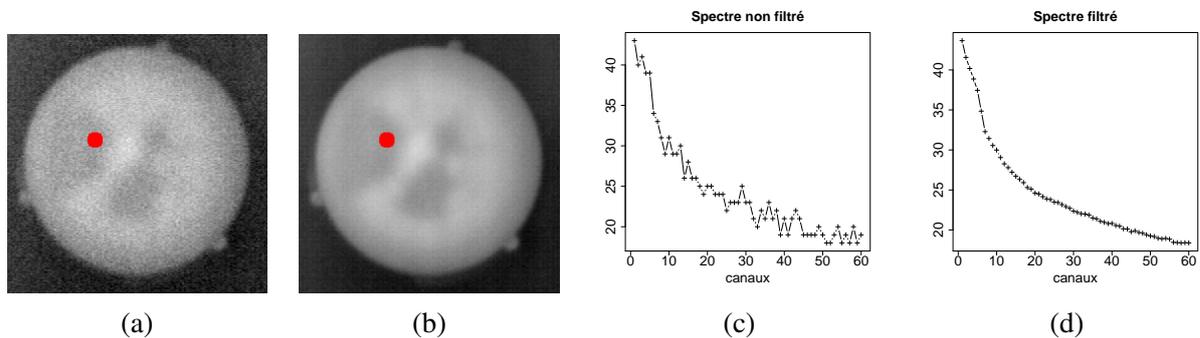


FIG. 2.31 – (a) Canal $f_{\lambda_{50}}$. (b) Canal $\hat{f}_{\lambda_{50}}$ après 2 AFC-reconstitutions à 2 axes, le point de mesure du spectre est indiqué en rouge. (c) Spectre avant filtrage. (d) Spectre filtré.

2.4.6 Réduction de dimension par modélisation du spectre

La dernière méthode utilisée dans ce travail pour réduire les données est la modélisation du spectre qui permet d'obtenir des cartes de paramètres utilisées ensuite, pour la classification et la segmentation morphologique. L'ensemble des cartes de paramètres forme donc une image hyperspectrale :

$$\mathbf{p}(x) = (p_1(x), \dots, p_M(x)) \quad (2.4.22)$$

Un modèle mathématique de droite, rapide à calculer, a été ajusté sur une partie de la courbe des spectres. On peut envisager d'ajuster des modèles plus complexes, mais ceux-ci qui tiennent compte

de la physique du phénomène [13] seront plus longs à ajuster. Cependant, comme un des objectifs de ce travail est de présenter des méthodes suffisamment générales pour être utilisées dans différents types d'imagerie, nous avons uniquement étudié ce modèle mathématique.

Sur chaque spectre, de l'image filtrés \hat{f}_λ par deux AFC-reconstitutions, dont on a retiré les 20 premières valeurs correspondant au régime transitoire, on choisit d'ajuster le modèle de droite (fig. 2.32) :

$$\hat{f}_\lambda(x_i) \sim a(x_i)\lambda + b(x_i) \quad \text{avec } \lambda = \lambda_{21} \dots \lambda_{512} \quad \forall i = 1 \dots P \quad (2.4.23)$$

Ce modèle est composé de deux paramètres :

- la pente $p_1 = a$
- l'ordonnée à l'origine $p_2 = b$

Afin de tenir compte de l'information contenue dans le régime transitoire, qui correspond à l'injection du produit de contraste, on définit un troisième paramètre qui correspond à l'amplitude maximale sur les 20 premières valeurs de chaque spectre. Ce paramètre, appelée « montée », $p_3 = m$, est défini par :

$$m(x_i) = \max_{j \in [1:20]} (\hat{f}_{\lambda_j}(x_i)) - \min_{j \in [1:20]} (\hat{f}_{\lambda_j}(x_i)) \quad \forall i = 1 \dots P \quad (2.4.24)$$

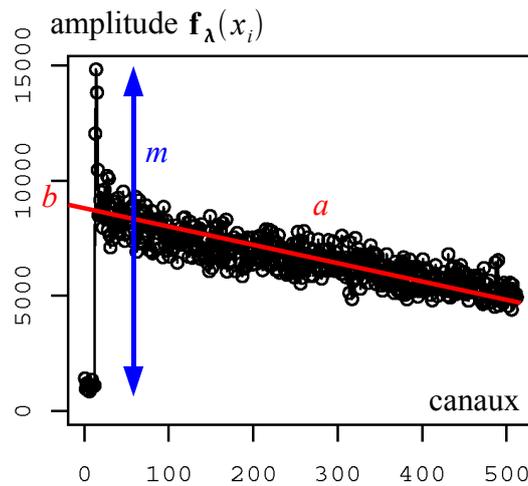


FIG. 2.32 – Ajustement du modèle sur un spectre $f_\lambda(x_i)$.

Après ajustement du modèle pour chaque spectre $f_\lambda(x_i)$, on obtient des cartes de paramètres (fig. 2.33). On observe que les principales structures morphologiques de la souris, i.e. le cœur, la tumeur et les poumons, ressortent bien sur ces cartes de paramètres.

En outre, la réduction de données est forte puisque l'on passe d'une image hyperspectrale à 512 canaux à une image à 3 canaux.

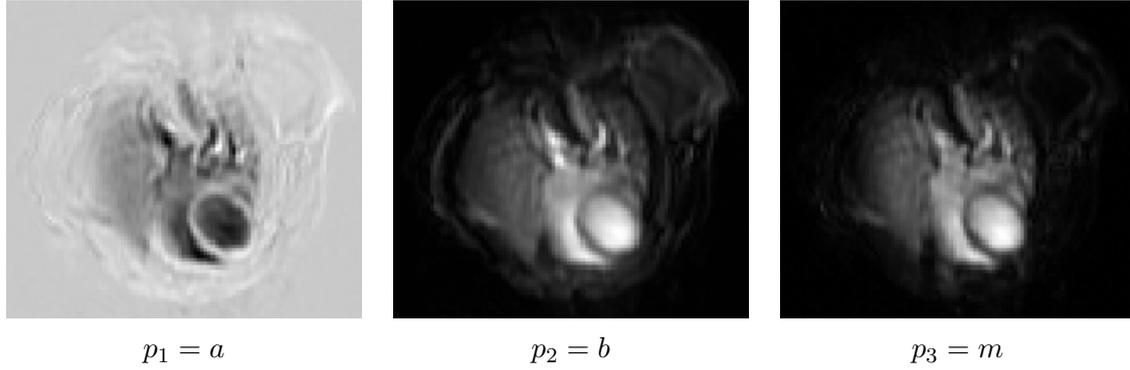


FIG. 2.33 – Cartes des paramètres de pente $p_1 = a$, d'ordonnée à l'origine $p_2 = b$ et de montée $p_3 = m$.

2.4.7 D'autres méthodes pour la réduction de dimension

2.4.7.1 Combinaison de plusieurs AFC avec information spatio-spectrale

Afin d'introduire l'information spatiale dans l'AFC, on choisit de créer, à partir des facteurs pixels $\mathbf{c}_\alpha^f = (c_{\alpha_1}^f, c_{\alpha_2}^f, c_{\alpha_3}^f)$, de nouvelles variables qui tiennent compte de l'information spatiale. Ces nouvelles variables sont ensuite filtrées par AFC et reconstituées. Puis on effectue une dernière reconstitution pour récupérer l'image de départ à partir de ces variables (fig. 2.34).

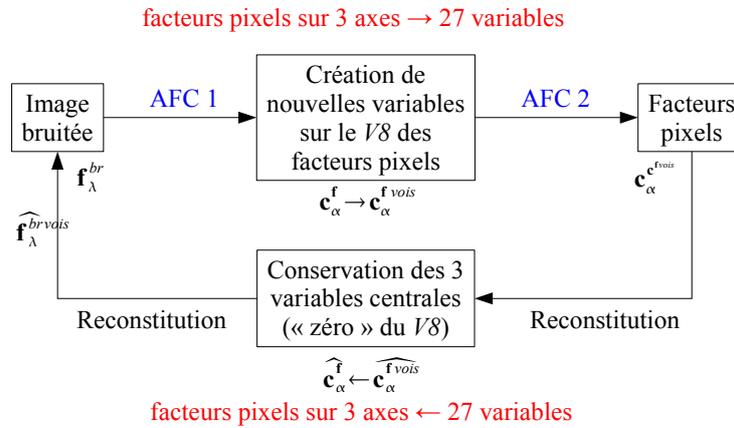


FIG. 2.34 – Schéma de la méthode de combinaison d'AFC avec introduction d'information spatiale.

Les nouveaux facteurs pixels sont obtenus par remplacement de chaque pixel par un de ses voisins sur un voisinage V_8 (fig. 2.35) à partir d'un facteur pixel $c_{\alpha_k}^f : \{c_{\alpha_k}^{fvois_0}, c_{\alpha_k}^{fvois_1}, \dots, c_{\alpha_k}^{fvois_9}\}$. On remarque que $c_{\alpha_k}^f = c_{\alpha_k}^{fvois_0}$.

Les nouvelles variables sont donc :

$$\mathbf{c}_\alpha^{fvois} = \{c_{\alpha_1}^{fvois_0}, c_{\alpha_1}^{fvois_1}, \dots, c_{\alpha_1}^{fvois_9}, \dots, c_{\alpha_k}^{fvois_0}, c_{\alpha_k}^{fvois_1}, \dots, c_{\alpha_k}^{fvois_9}, \dots, c_{\alpha_K}^{fvois_0}, c_{\alpha_K}^{fvois_1}, \dots, c_{\alpha_K}^{fvois_9}\} \quad (2.4.25)$$

Sur les nouvelles variables $\mathbf{c}_\alpha^{fvois}$, on applique une AFC et l'on conserve les facteurs pixels des deux premiers axes factoriels $\mathbf{c}_\alpha^{c_{fvois}} = (c_{\alpha_1}^{c_{fvois}}, c_{\alpha_2}^{c_{fvois}})$.

4	3	2
5	0	1
6	7	8

FIG. 2.35 – Voisinage V8.

A partir de ces deux axes, on reconstitue les 27 variables $\widehat{c_\alpha^{fvois}}$. On sélectionne les 3 variables correspondant à la position 0 : $\widehat{c_{\alpha_1}^{fvois_0}}$, $\widehat{c_{\alpha_2}^{fvois_0}}$, $\widehat{c_{\alpha_3}^{fvois_0}}$. A partir des ces 3 variables correspondant aux facteurs pixels $c_\alpha^f = (c_{\alpha_1}^f, c_{\alpha_2}^f, c_{\alpha_3}^f)$, on reconstitue l'image de départ $\widehat{f_\lambda^{br}^{vois}}$.

Remarque 2.4.9. Dans la pratique, les nouvelles variables c_α^{fvois} sont obtenues par translation sur un voisinage V8. Il est à noter que pour éviter les problèmes de bord on réduit l'image des nouvelles variables. Pour ce faire, on leur enlève un cadre extérieur d'un pixel d'épaisseur. On garde en mémoire chacun des cadres, associés à chacun des axes, pour la reconstitution de l'image $\widehat{f_\lambda^{vois}}$. En effet, on rajoute alors ce cadre sur les variables $\widehat{c_{\alpha_1}^{fvois_0}}$, $\widehat{c_{\alpha_2}^{fvois_0}}$, $\widehat{c_{\alpha_3}^{fvois_0}}$.

Pour faire la manipulation, on bruite une image par un bruit de Poisson de paramètre égal à l'amplitude de chaque point $f_{\lambda_j}(x_i)$ (fig. 2.36).

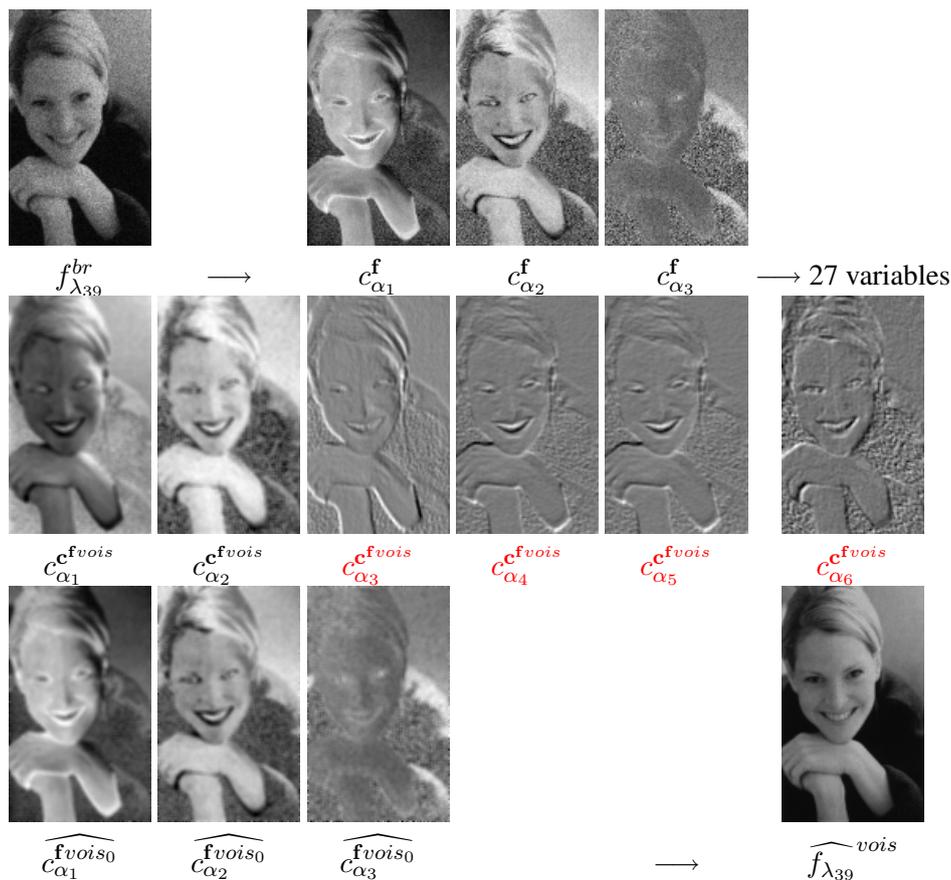


FIG. 2.36 – Facteurs pixels dans les différents espaces de composition d’AFC avec information spatiale sur un voisinage V8. Les facteurs pixels des axes non retenus sont indiqués en rouge.

On constate que cette analyse introduit sur l'image reconstituée un flou à cause de l'introduction de la dimension spatiale. En outre les facteurs pixels de la deuxième AFC $c_{\alpha_3}^{c^{fvois}}$, $c_{\alpha_4}^{c^{fvois}}$, $c_{\alpha_5}^{c^{fvois}}$ et $c_{\alpha_6}^{c^{fvois}}$ correspondent à la translation spatiale des facteurs c_{α}^f . Dans notre analyse, ils ne sont pas conservés, cependant ils pourraient être utilisés pour détecter le mouvement.

Afin de comprendre les effets de ces deux AFC avec introduction de l'information spatiale. On décide de filtrer, par une moyenne sur un voisinage V_8 , les facteurs pixels après la première AFC c_{α}^f , puis on reconstitue l'image de départ $\widehat{f}_{\lambda}^{f_{\lambda}^{filtre}}$. On trace la carte des distance du chi-deux, et l'histogramme, entre l'image de départ f_{λ} sans bruit et les images reconstituées après 2 AFC spatiales $\widehat{f}_{\lambda}^{br^{vois}}$, et 1 AFC avec filtrage par moyenne $\widehat{f}_{\lambda}^{br^{filtre}}$ (fig. 2.37). On constate que l'on obtient des résultats similaires. Ainsi au niveau du rendu, la composition d'AFC spatiale revient au filtrage spatial des facteurs pixels par moyenne.

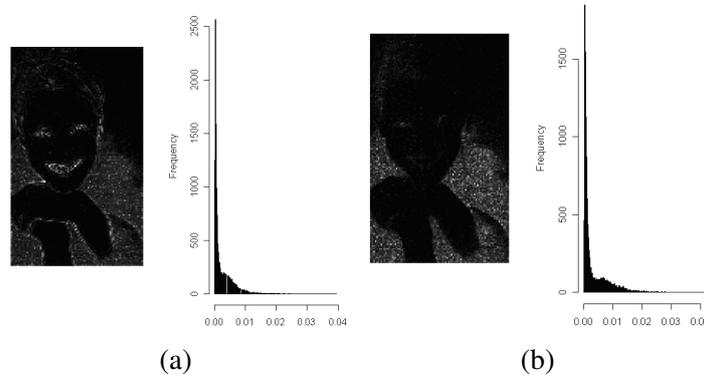


FIG. 2.37 – Cartes de distance du chi-deux, et histogrammes associés, (a) entre l'image sans bruit f_{λ} et l'image reconstituée après 2 AFC spatiales $\widehat{f}_{\lambda}^{br^{vois}}$, (b) entre l'image sans bruit et l'image reconstituée après 1 AFC avec filtrage par moyenne $\widehat{f}_{\lambda}^{br^{filtre}}$.

2.4.7.2 Méthodes tensorielles

Les méthodes tensorielles [186], qui n'ont pas été utilisées dans cette thèse, pourraient servir à analyser les images hyperspectrales, si l'on disposait de séries de transformations sur les canaux. Par exemple des ouvertures γ et fermetures ϕ de taille croissante : $\gamma_1, \gamma_2, \gamma_3, \dots$ ou $\phi_1, \phi_2, \phi_3, \dots$ (fig. 2.38). Dans [150], appliquent directement la méthode tensorielle sur le cube de données hyperspectrales, sans avoir d'autres descripteurs, comme les ouvertures ou les fermetures, ce qui est un peu limitant. Les méthodes tensorielles sont aussi utilisées en imagerie médicale comme l'Imagerie par Résonance Magnétique Fonctionnelle (fMRI) [15].

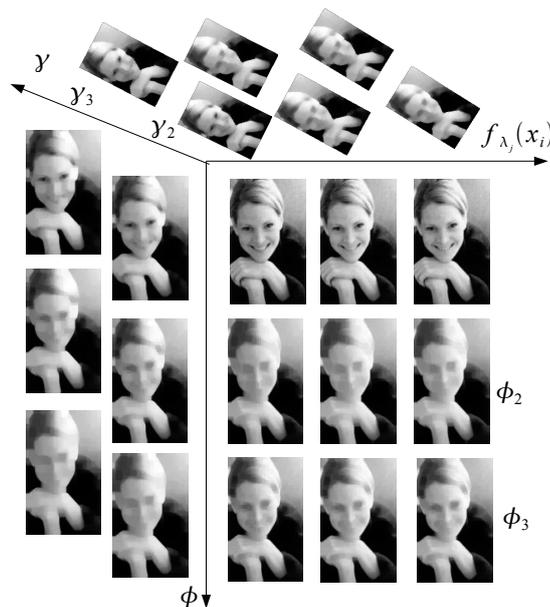


FIG. 2.38 – Tenseur des fermetures suivies d’ouvertures $\gamma_l \circ \phi_k$ de tailles croissantes pour différents canaux de l’image « woman face ».

2.4.7.3 Sélection des bandes spectrales d’intérêt

Il existe de nombreuses méthodes pour sélectionner les bandes spectrales d’intérêt à partir de paramètres physiques des constituants d’une image. La thèse de *F. Schmidt* [158] présentent ces méthodes.

Lors d’une AFC ou d’une ACP, on peut également sélectionner les canaux qui sont les plus représentés par un axe factoriel à l’aide du cercle de corrélation. La corrélation d’un canal par rapport à un axe factoriel est l’angle, normalisé entre -1 et 1, entre le vecteur associé au canal et l’axe factoriel considéré (fig. 2.39). Par conséquent plus un canal est proche d’une axe extrémité d’un axe, plus il est représenté par cet axe. En AFC, on ne parle pas de corrélation mais de qualité de représentation sur un axe.

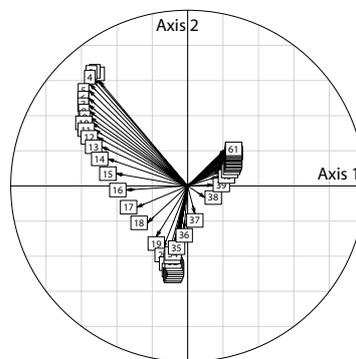


FIG. 2.39 – Cercle de corrélation des canaux de l’image « woman face » dans le plan factoriel AFC.

2.4.8 Conclusion du filtrage et de la réduction de dimension

Les diverses méthodes de réduction de dimension présentées dans ce mémoire, permettent d'obtenir des espaces de dimension réduite et ainsi de lutter contre le phénomène de *Hughes* [81]. Dans les nouveaux espaces ainsi créés, on va pouvoir classifier correctement les spectres, évitant ce phénomène qui fait que les espaces de grande dimension sont très « vides ». On effectue ainsi une comparaison entre les pixels vecteurs afin de les regrouper selon une similarité spectrale. Ainsi chaque pixel vecteur est comparé à l'ensemble des autres. Il s'agit donc bien d'une comparaison globale sur l'image qui est effectuée.

2.5 État de l'art de la classification

Dans cette partie nous présentons les méthodes utilisées dans ce travail et les méthodes non-supervisées et supervisées les plus courantes. Les méthodes supervisées nécessitent un échantillon d'apprentissage défini par un opérateur, sur lequel le classificateur est entraîné. Une fois que le classificateur est entraîné, on classe les données sans intervention d'un utilisateur. Quant aux méthodes non-supervisées, elles ne nécessitent pas d'ensemble d'entraînement. Par contre, il est nécessaire d'avoir un a-priori comme le nombre de classes.

2.5.1 Kmeans

K-means est une méthode des plus connues pour la classification non-supervisée créée par *Hartigan et Wong* en 1979 [75]. Elle nécessite comme unique paramètre un nombre de classes. Elle correspond à la méthode des « nuées dynamiques » créée en France par *Diday* en 1971 [51]. Cet algorithme nécessite un nombre de classes pré-défini.

L'algorithme *kmeans* consiste à tirer aléatoirement des centres de K classes, puis de manière itérée jusqu'à convergence :

- recherche de la classe la plus proche de chaque point ;
- échange de points entre deux classes. pour minimiser la somme des distances intra classe $W(C)$ (eq. 2.5.1)

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(i)=k} \sum_{C(i')=k} \|x_i - x_{i'}\|^2 \quad (2.5.1)$$

avec $C(i)$ la classe du point x_i , $\|x_i - x_{i'}\|^2$ la distance euclidienne carrée entre deux points x_i et $x_{i'}$: $\|x_i - x_{i'}\|^2 = \sum_{j=1}^L (x_{ij} - x_{i'j})^2$.

La somme des distances intra-classe peut s'écrire [76] :

$$W(C) = \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2,$$

où $\bar{x}_k = (\bar{x}_{1k}, \dots, \bar{x}_{Lk})$ est le vecteur moyen associé à la k^e classe, et $N_k = \sum_{i=1}^P I(C(i) = k)$. $I(a)$ est l'indicatrice de a qui vaut 1 si a est vérifiée et 0 sinon. Donc, le critère est minimisé en assignant

les P observations aux K classes telles qu'à l'intérieur de chaque classe la différence des observations à la moyenne de la classe, définie pas les points de la classe, soit minimale.

Un algorithme itératif, descendant, pour résoudre

$$C^* = \min_C \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2$$

peut être obtenu en remarquant que pour chaque ensemble d'observations S :

$$\bar{x}_S = \operatorname{argmin}_m \sum_{i \in S} \|x_i - m\|^2. \quad (2.5.2)$$

Ainsi, on obtient C^* en résolvant le problème d'optimisation :

$$\min_{C, \{m_k\}_1^K} \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2. \quad (2.5.3)$$

C^* peut être minimisé par la procédure d'optimisation alternée décrite dans l'algorithme 1.

Algorithme 1 Classification *kmeans*

- 1: Pour une affectation donnée C des classes, la variance totale de la classe (eq. 2.5.3) est minimisée par rapport à (m_1, \dots, m_K) ce qui modifie les moyennes des classes actuellement définies (eq. 2.5.2).
- 2: Étant donné l'ensemble courant des moyennes (m_1, \dots, m_K) , l'équation 2.5.3 est minimisée par affectation de chaque observation à la moyenne de la classe (courante) la plus proche, i.e. :

$$C(i) = \operatorname{argmin}_{1 \leq k \leq K} \|x_i - m_k\|^2$$

- 3: Les étapes 1 et 2 sont itérées jusqu'à ce que les assignation ne changent plus.
-

Remarque 2.5.1. Il est important d'exécuter l'algorithme de nombreuses fois avec des centres de classes aléatoires, et de choisir la solution pour laquelle la fonction C^* est minimale.

Remarque 2.5.2. La métrique associée à *kmeans* est la métrique euclidienne. Il est donc important d'utiliser un espace où les données sont euclidiennes. Les espaces factoriels de l'AFC ou de l'ACP s'y prêtent donc naturellement puisque leur métrique est la métrique euclidienne. En outre ces espaces sont de dimension réduites ce qui améliore les temps de calcul et réduit le phénomène de Hughes.

2.5.2 k-medoids

L'algorithme des *k-medoids* est non supervisé et nécessite comme unique paramètre un nombre de classes.

L'algorithme *kmeans* nécessite une mesure de dissemblance qui est la distance euclidienne carrée $D(x_i, x_{i'})$ [76]. La seule partie de *kmeans* qui suppose une distance euclidienne carrée est l'étape de minimisation (eq. 2.5.2). Les représentants des classes $\{m_1, \dots, m_K\}$ (eq. 2.5.3) sont supposés être

les moyennes des classes courantes. L'algorithme peut être généralisé avec une mesure de dissemblance quelconque $D(x_i, x_{i'})$ en remplaçant cette étape par une optimisation explicite par rapport à $\{m_1, \dots, m_K\}$ dans l'équation 2.5.3. Dans les méthodes les plus communes, les centres de chaque classe sont restreints à être une des observations affectées à la classe, comme expliqué dans l'algorithme 2. On peut également utiliser cette méthode sur des données décrites par des matrices de dissemblance. Il n'est pas nécessaire de calculer de manière explicite le centre des classes ; il suffit juste des indices i_k^* .

Remarque 2.5.3. Les mesures de dissemblance, à la différence des distances, ne vérifient pas l'inégalité triangulaire.

Résoudre l'équation 2.5.2 pour chaque classe provisoire k requiert un nombre d'opérations proportionnel au nombre d'observations qui lui sont affectées, tandis que pour résoudre l'équation 2.5.5 la complexité est en $O(N_k^2)$. Étant donné un ensemble de centres de classes, $\{i_1, \dots, i_K\}$, l'obtention des nouvelles affectations :

$$C(i) = \operatorname{argmin}_{1 \leq k \leq K} D_{x_i, x_{i_k^*}} \quad (2.5.4)$$

requiert un nombre d'opérations proportionnel à $K \times N$. Par conséquent, l'algorithme des k -medoids a un nombre de calculs beaucoup plus élevé que k means.

Algorithme 2 Classification k -medoids

- 1: Pour une affectation donnée C des classes, trouver l'observation de la classe qui minimise la distance totale aux autres points dans cette classe (i.e. la médiane vectorielle) :

$$i_k^* = \operatorname{argmin}_{\{i: C(i)=k\}} \sum_{C(i')=k} D(x_i, x_{i'}) \quad (2.5.5)$$

Alors les $m_k = x_{i_k^*}$, $k = 1, 2, \dots, K$ sont les estimations courantes des centres des classes.

- 2: Étant donné un ensemble courant des centres de classes (m_1, \dots, m_K) , minimiser l'erreur totale en assignant chaque observation au centre de classe (courant) le plus proche :

$$C(i) = \operatorname{argmin}_{1 \leq k \leq K} D(x_i, m_k)$$

- 3: Les étapes 1 et 2 sont itérées jusqu'à ce que les assignation ne changent plus.
-

L'alternance entre les équations 2.5.5 et 2.5.4 constitue une stratégie particulière de recherche d'heuristique pour essayer de résoudre :

$$\min_{C, \{i_k\}_1^K} \sum_{k=1}^K \sum_{C(i)=k} D_{x_i, x_{i_k}} \quad (2.5.6)$$

Kaufman et Rousseuw [90], proposent une stratégie alternative qui résout directement l'équation 2.5.6 en échangeant provisoirement chaque centre i_k avec une observation qui n'est pas un centre courant ; ceci en sélectionnant les échanges qui donnent la plus grande réduction du critère (eq. 2.5.6). Ce procédé est répété jusqu'à ce qu'aucun autre échange avantageux ne puisse t'être trouvé.

Remarque 2.5.4. Comme l'algorithme des *k-medoids* est très coûteux en temps de calcul, *Kaufman et Rousseeuw* ont développé l'algorithme Clara (Clustering LArge Applications)

2.5.3 Clara

L'algorithme CLARA (Clustering Large Applications) a été créée par *Kaufman et Rousseeuw* en 1986 [90] pour classifier des ensembles ayant beaucoup de données avec des temps de calcul acceptables. *Clara* est un algorithme non supervisé qui nécessite comme unique paramètre un nombre de classes.

Sur un sous échantillon de l'ensemble des points à classer, une classification par *k-medoids* est effectuée. *k* objets représentatifs sont ainsi déterminés. Les autres points sont classifiés selon le centre de la classe qui leur est le plus proche. La distance moyenne obtenue pour l'affectation est utilisée comme une mesure de qualité de la classification. Le procédé est répété 5 fois en tenant compte, à chaque itération, des centres des classes de l'itération précédente. La classification retenue est celle qui a la distance moyenne la plus faible.

Remarque 2.5.5. Cet algorithme a été très utilisé pour cette thèse en raison de son efficacité pour les ensemble contenant beaucoup de points ; ce qui est le cas pour des images hyperspectrales.

2.5.4 LDA

L'Analyse Discriminante Linéaire, Linear Discriminant Analysis (LDA) en anglais, consiste à séparer par des hyperplans linéaires les classes d'une classification. Pour dire si un point appartient à une classe où à une autre, on utilise une fonction discriminante $\delta_k(x)$ pour chaque classe *k* et on classifie le point *x* dans la classe qui a la plus grande valeur pour cette fonction discriminante.

Classifier revient à connaître les probabilités conjointes des classes $G = k$ connaissant les données $X = x$ $Pr(G|X)$ [76]. Supposons que $f_k(x)$ est la distribution conditionnelle de *X* dans la classe $G = k$, et soit π_k la probabilité a priori de la classe *k*, avec $\sum_{k=1}^K \pi_k = 1$. L'application simple du théorème de Bayes donne :

$$Pr(G = k|X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^K f_l(x)\pi_l}. \quad (2.5.7)$$

On remarque qu'en terme d'aptitude à classifier, connaître les $f_k(x)$ est pratiquement équivalent à connaître $Pr(G = k|X = x)$. La LDA est un modèle qui s'appuie sur la distribution des classes. La LDA utilise des distributions gaussiennes multivariées :

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1} (x-\mu_k)}. \quad (2.5.8)$$

La LDA est utilisée dans le cas particulier où l'on suppose que les classes *k* ont la même matrice de covariance $\Sigma_k = \Sigma \forall k$. Par comparaison de deux classes *k* et *l*, en prenant le logarithme du rapport, on obtient une équation linéaire en *x* :

$$\begin{aligned} \ln \frac{Pr(G = k|X = x)}{Pr(G = l|X = x)} &= \ln \frac{f_k(x)}{f_k(l)} + \ln \frac{\pi_k}{\pi_l} \\ &= \ln \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^t \Sigma^{-1} (\mu_k - \mu_l) + x^t \Sigma^{-1} (\mu_k - \mu_l). \end{aligned} \quad (2.5.9)$$

Cette équation est celle de la frontière entre les classes k et l . C'est l'équation d'un hyperplan en dimension p . Si l'on divise \mathbb{R}^p en régions qui sont classifiées comme : classe 1, classe 2, ..., ces régions seront séparées par des hyperplans.

De l'équation 2.5.9, on en déduit les fonctions discriminantes linéaires qui sont équivalentes à la règle de décision, avec $G(x) = \operatorname{argmax}_k \delta_k(x)$.

$$\delta_k(x) = x^t \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^t \Sigma^{-1} \mu_k + \ln \pi_k \quad (2.5.10)$$

Dans la pratique les paramètres des distributions gaussiennes ne sont pas connus. Il est donc nécessaire de les estimer sur l'ensemble d'entraînement :

- $\hat{\pi}_k = N_k/N$, où N_k est le cardinal de la classe k ;
- $\hat{\mu}_k = \sum_{g_i=k} x_i / N_k$;
- $\hat{\Sigma} = \sum_{k=1}^K \sum_{g_i=k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^t / (N - k)$.

Remarque 2.5.6. La LDA nécessite un ensemble d'entraînement pour apprendre les paramètres du classificateur, un ensemble de validation pour mesurer les erreurs d'entraînement et de test et enfin un ensemble de test : c'est à dire l'ensemble des données que l'on veut classifier. Souvent on parle uniquement d'ensemble d'entraînement et d'ensemble de test. Dans ce cas l'ensemble de validation est un sous-ensemble de l'ensemble d'entraînement. Lorsque l'on a peu de données, une technique pour diviser l'ensemble d'entraînement en « train » et « validation » est la validation croisée ou « cross-validation » [76].

La LDA est utilisée dans la littérature pour des données multivariées ; citons entre autres [180].

2.5.5 D'autres classificateurs

Il existe de nombreuses autres techniques de classification dont :

- les « Support Vector Machines » (SVM) [36]. Cette méthode n'est pas sensible au phénomène de Hughes [104]. Une présentation concise de la méthode est faite dans [76]. Les SVM ont notamment été appliqués en imagerie hyperspectrale dans [71, 105, 9] ;
- K-nearest neighbors : c'est à dire les k plus proches voisins [76, 97] classification par approche spatio-spectrale [111, 148] ;
- de nombreuses méthodes de classification pour les images hyperspectrales sont recensées dans la thèse de *F. Schmidt* [158].

Comme classification spatio-spectrale, citons entre autre *Paclik* [143] qui obtient, grâce à des classificateurs statistiques les probabilités d'appartenance des pixels à des clusters dans le domaine spectral et des probabilités d'appartenance des pixels dans le domaine spatial. Les probabilités d'appartenance spatio-spectrales des pixels sont obtenues par multiplication des deux probabilités. En effet on suppose que l'information spatiale et l'information spectrale sont indépendantes. Les pixels sont ainsi classifiés et le procédé est répété jusqu'à convergence.

2.6 Classification

Après avoir fait un état de l'art des différents classificateurs, nous présentons dans cette partie, nos résultats de classification. Les classificateurs employés sont les suivants : *kmeans*, *clara* et l'Analyse Discriminante Linéaire (LDA). Ils permettent de regrouper les pixels vecteurs en classes spectralement homogènes. Chaque pixel est ainsi comparé à tous les autres présents dans l'image. Il s'agit donc bien d'une analyse globale des spectres de l'image.

2.6.1 Kmeans

La métrique sous-jacente à l'algorithme *kmeans* est la métrique euclidienne. Il est donc important d'appliquer cet algorithme dans un espace dont la métrique associée est la métrique euclidienne, comme les espaces factoriels de l'AFC ou de l'ACP.

Appliquer *kmeans* directement dans l'espace image \mathbf{f}_λ le rendrait très dépendant aux points aberrants. Les normalisations effectuées sur chacune des variables (i.e. les canaux) par l'AFC ou l'ACP, afin de créer des espaces euclidien, sont essentielles pour palier à ce problème.

Comme expliqué dans la partie 2.4.4 (p : 35) nous appliquons la composition de deux AFC-reconstitutions sur les images bruitées comme celle de souris « serim447 » ou « bouchon en plastique » afin de réduire le bruit dans l'espace factoriel. Pour une image quasiment sans bruit, comme l'image « woman face », une seule AFC suffit.

Classifions les spectres dans l'espace factoriel en 5 classes. Les points de départ sont choisis aléatoirement (fig. 2.40). On relance *kmeans* 1000 fois avec 20 itérations d'échange de classes à chaque fois. On observe que la classification $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,5}$ uniquement spectrale respecte les contours principaux de l'image. Pour comparaison, on effectue une classification en 4 classes $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,4}$. On constate que les classes « cyan » et « verte » de la classification en 5 classes ont été regroupées dans la classe « verte » de la classification en 4 classes. On préférera, par la suite, conserver la classification en 5 classes $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,5}$.

Pour l'image de souris « serim447 », on effectue une classification en 5 classes après deux AFC-reconstitutions $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,5}$ (2.41). On constate que l'anatomie de la souris ressort bien pour ce nombre de classes :

- classe verte : correspond pour la grosse partie en haut à droite à la tumeur ;
- classe rouge : le fond même si certaines parties de la souris ont une réponse spectrale similaire à celle du fond ;
- classe bleue : le cœur ;
- classe noire : poumon ;
- classe cyan : classe intermédiaire entre les poumons et l'extérieur du corps de la souris.

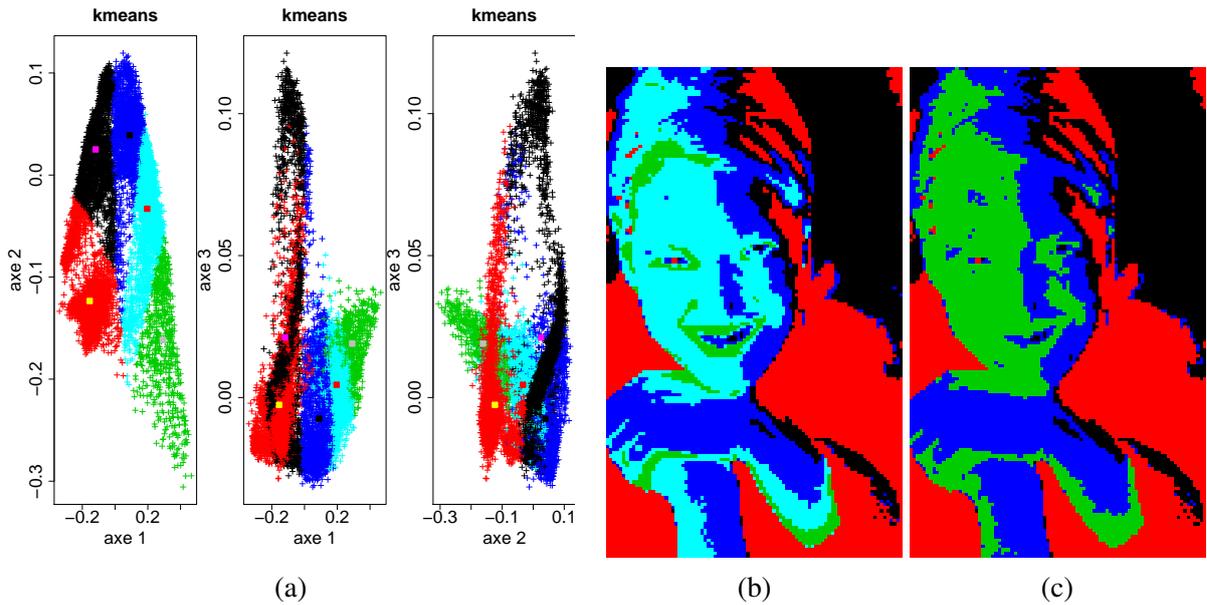


FIG. 2.40 – Classification *kmeans* (5 classes) dans l'espace factoriel \mathbf{c}_α^f : (a) représentation des facteurs pixels classifiés et des centres des classes dans l'espace factoriel formé des 3 axes factoriels et (b) classification associée $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,5}$. (c) Classification *kmeans* (4 classes) $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,4}$.

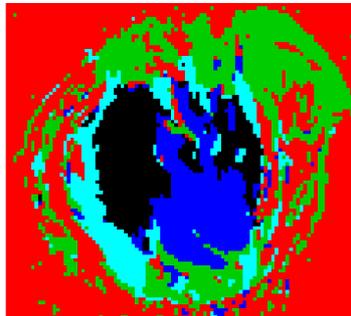


FIG. 2.41 – Classification *kmeans* (5 classes) $\kappa_{\mathbf{c}_\alpha^f}^{kmeans,5}$ dans l'espace factoriel \mathbf{c}_α^f .

2.6.2 Clara

En ce qui concerne l'image « bouchon en plastique », on cherche à détecter les tâches de colle. Pour ce faire, on effectue une classification *clara* $\kappa_{\mathbf{c}_\alpha^f}^{clara,3}$ en 3 classes dans l'espace factoriel \mathbf{c}_α^f après deux AFC-reconstitutions. On constate que la classification fait apparaître les 3 tâches de colles. Cependant la classe verte correspondant à la colle regroupe également d'autres parties du bouchon et du fond (fig. 2.42).

Pour améliorer les résultats, on décide d'ajuster un modèle de droite sur l'image filtrée \hat{f}_λ sur les spectres après avoir retiré les 10 premières valeurs correspondant au décalage : $\hat{f}_\lambda(x_{11}), \hat{f}_\lambda(x_{12}), \dots, \hat{f}_\lambda(x_{60})$. Les 10 premières valeurs sont utilisées pour ajuster le paramètre de montée du modèle (cf. partie 2.4.6, p : 46). Des cartes de paramètres $\mathbf{p} = (a, b, m)$ sont ainsi obtenues.

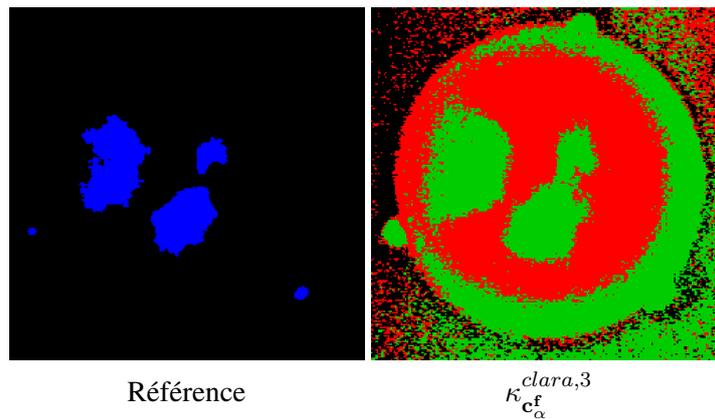


FIG. 2.42 – Référence et classification clara (3 classes) $\kappa_{c_\alpha^f}^{clara,3}$ dans l'espace factoriel c_α^f pour l'image « bouchon en plastique ».

nues (fig. 2.43).

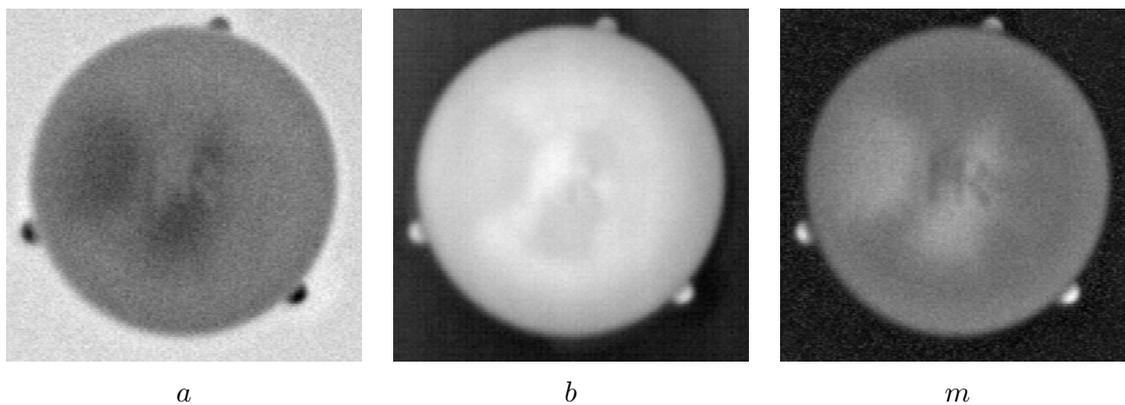


FIG. 2.43 – Cartes des paramètres du modèle ajusté sur l'image « bouchon en plastique » filtrée \hat{f}_λ : pente a , ordonnée à l'origine b et montée m .

Une classification clara en 9 classes dans l'espace des paramètres \mathbf{p} est ensuite effectuée $\kappa_{\mathbf{p}}^{clara,9}$. On voit bien ressortir les tâches de colles sur la classe de couleur magenta (fig. 2.44 (a)). Les résultats sont donc meilleurs que ceux obtenus avec la classification sur les axes factoriels $\kappa_{c_\alpha^f}^{clara,3}$.

Comme on voudrait avoir une classification avec deux classes : la colle et le reste de l'image, on utilise un autre procédé. En outre la métrique euclidienne dans l'espace des paramètres n'est pas forcément adaptée. Il est donc nécessaire d'orthogonaliser les paramètres par une ACP. On ne conserve que les deux premiers axes factoriels sur 3 axes factoriels après observation des RSB et des inerties (fig. 2.45) :

$$c_\beta^p = (c_{\beta_1}^p, c_{\beta_2}^p, c_{\beta_3}^p) \quad (2.6.1)$$

On effectue une classification clara 3 classes sur les facteurs pixels du l'axe 1 de l'ACP $\kappa_{c_{\beta_1}^p}^{clara,3}$ (fig. 2.46 (a)). On sélectionne la classe verte $\kappa_{c_{\beta_1}^p}^{clara,3}(\text{verte})$ qui correspond au centre du bouchon, c'est à dire là où se trouve la colle. On effectue ensuite une classification clara en deux classes,

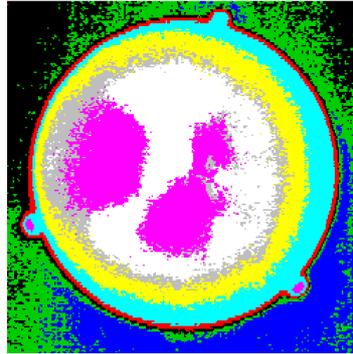
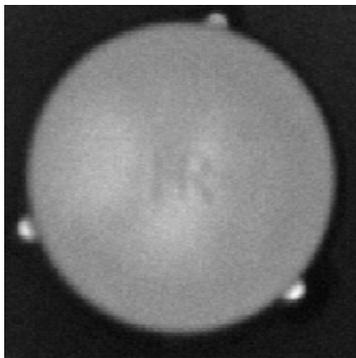
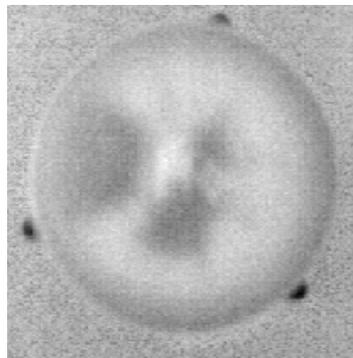


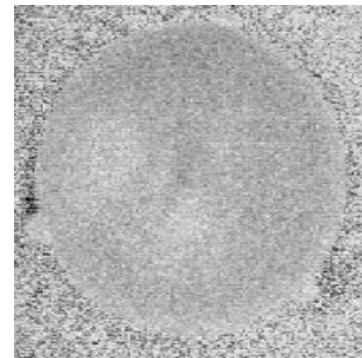
FIG. 2.44 – Classification clara pour l'image « bouchon en plastique » en 9 classes dans l'espace des 3 paramètres $\mathbf{p} = (a, b, m) \kappa_{\mathbf{p}}^{clara,9}$.



$$\begin{aligned} c_{\beta_1}^{\mathbf{p}} \\ \tau_{\beta_1} = 97,24 \% \\ RSB_{\beta_1} = 115 \end{aligned}$$



$$\begin{aligned} \text{(b) } c_{\beta_2}^{\mathbf{p}} \\ \tau_{\beta_2} = 2,10 \% \\ RSB_{\beta_2} = 6 \end{aligned}$$



$$\begin{aligned} \text{(c) } c_{\beta_3}^{\mathbf{p}} \\ \tau_{\beta_3} = 0,66 \% \\ RSB_{\beta_3} = 0,4 \end{aligned}$$

FIG. 2.45 – Facteurs pixels $c_{\beta}^{\mathbf{p}}$ de l'ACP sur les paramètres $\mathbf{p} = (a, b, m)$, leurs parts d'inerties, τ_{β} , et leur RSB, RSB_{β} .

$\kappa_{c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}}}^{clara,2} | x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)$, sur les facteurs pixels des deux axes, $c_{\beta_1}^{\mathbf{p}}$ et $c_{\beta_2}^{\mathbf{p}}$, pour la restriction de leurs pixels à la classe verte précédente $x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)$. Les autres pixels sont mis dans la classe correspondant au fond (fig. 2.46). On constate que la classification obtenue est très similaire à la référence.

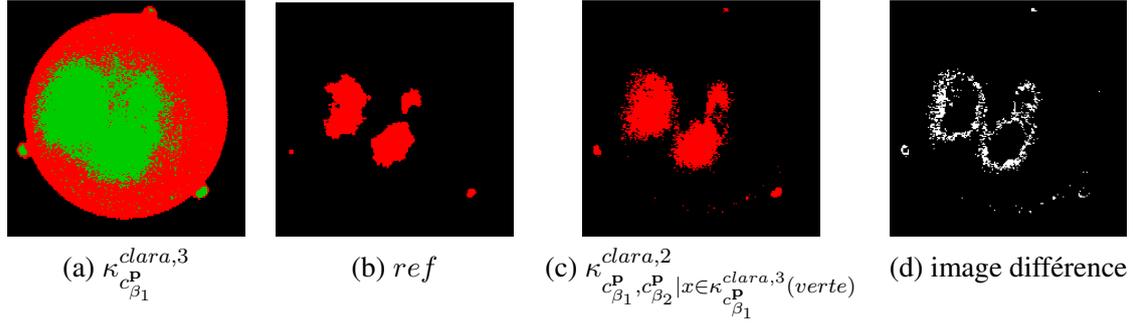


FIG. 2.46 – (a) Classification *clara* 3 classes sur les facteurs pixels de l'axe 1 de l'ACP $\kappa_{c_{\beta_1}^P}^{clara,3}$, (b) référence *ref*, (c) classification $\kappa_{a,m|x \in \kappa_{c_{\beta_1}^P}^{clara,3}(verte)}^{clara,2}$ *clara* en 2 classes sur les facteurs pixels $c_{\beta_1}^P$ et $c_{\beta_2}^P$ des deux axes pour les pixels sélectionnés $x \in \kappa_{c_{\beta_1}^P}^{clara,3}(verte)$ et (d) image différence entre *ref* et $\kappa_{c_{\beta_1}^P, c_{\beta_2}^P | x \in \kappa_{c_{\beta_1}^P}^{clara,3}(verte)}^{clara,2}$.

2.6.3 Approche modèle

Nous avons introduit une nouvelle méthode de classification qui classe les spectres par rapport à des spectres de référence obtenus par classification *kmeans* (fig. 2.47). Plutôt que de travailler directement sur les spectres avec une métrique du chi-deux ce qui redonnerait la classification *kmeans*, on utilise le modèle de droite ajusté sur le spectre et la métrique associée à la norme 1.

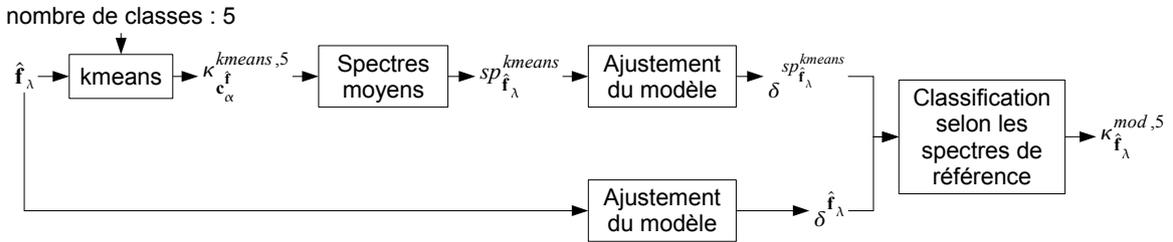


FIG. 2.47 – Schéma de la classification par approche modèle.

Observons les spectres moyens de l'image filtrée \hat{f}_λ pour la classification *kmeans* (fig. 2.48). On constate que les tendances des spectres moyens $sp_{\hat{f}_\lambda}^{kmeans}$ correspondent à la cinétique du produit de contraste injecté dans la souris. C'est à dire que :

- le produit s'accumule dans la tumeur (classe verte) ;
- le produit est expulsé des cavités cardiaques (classe bleue) car il est injecté dans l'aorte ;
- le spectre du fond est quasiment constant.

Sur ces spectres moyens, un modèle de droite est ajusté $\delta^{sp_{\hat{f}_\lambda}^{kmeans}} = \delta_1^{sp_{\hat{f}_\lambda}^{kmeans}}, \dots, \delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}, \dots, \delta_5^{sp_{\hat{f}_\lambda}^{kmeans}}$.

On fait de même pour chaque pixel vecteur $f_\lambda(x_i)$ et l'on note le modèle $\delta^{\hat{f}_\lambda} = \delta_{x_1}^{\hat{f}_\lambda}, \dots, \delta_{x_i}^{\hat{f}_\lambda}, \dots, \delta_{x_P}^{\hat{f}_\lambda}$.

On classe chaque point x_i par minimisation de la distance d_1 entre modèle des spectres moyens

$\delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}$ et modèle des spectres en chaque pixel $\delta_{x_i}^{\hat{f}_\lambda}$. La classe du point x_i est notée $C(x_i)$ (eq. 2.6.2, fig. 2.49).

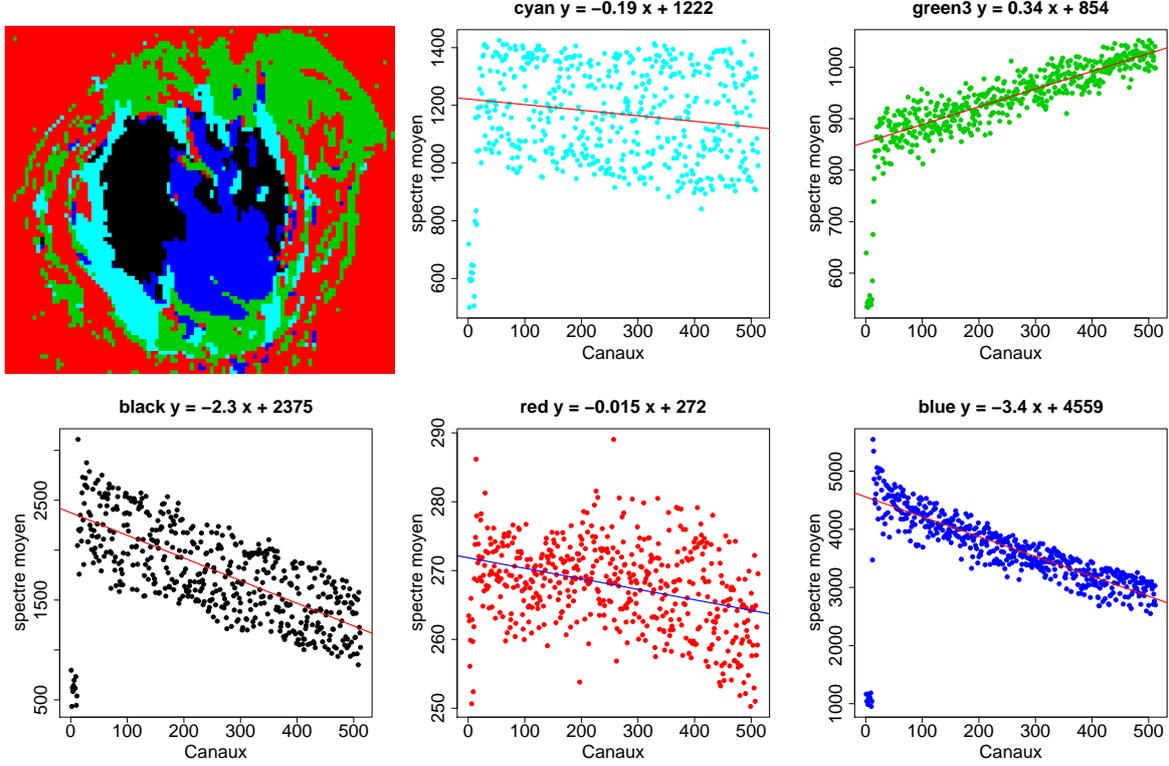


FIG. 2.48 – Spectres moyens de la classification $kmeans_{\kappa_{\mathbf{c}_\alpha}^{kmeans,5}}$.

$$C(x_i) = \operatorname{argmin}_k d_1(\delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}, \delta_{x_i}^{\hat{f}_\lambda}) = \operatorname{argmin}_k \left(\sum_{j=1}^L \left| \delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}(\lambda_j) - \delta_{x_i}^{\hat{f}_\lambda}(\lambda_j) \right| \right) \quad (2.6.2)$$

La classification obtenue $\kappa_{\hat{f}_\lambda}^{mod,5}$, par approche modèle, donne un filtrage plus robuste qu'avec $kmeans$ car un filtrage statistique des données est fait par le modèle (fig. 2.50). La classification $kmeans$ est néanmoins nécessaire pour obtenir les spectres moyens $\delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}$.

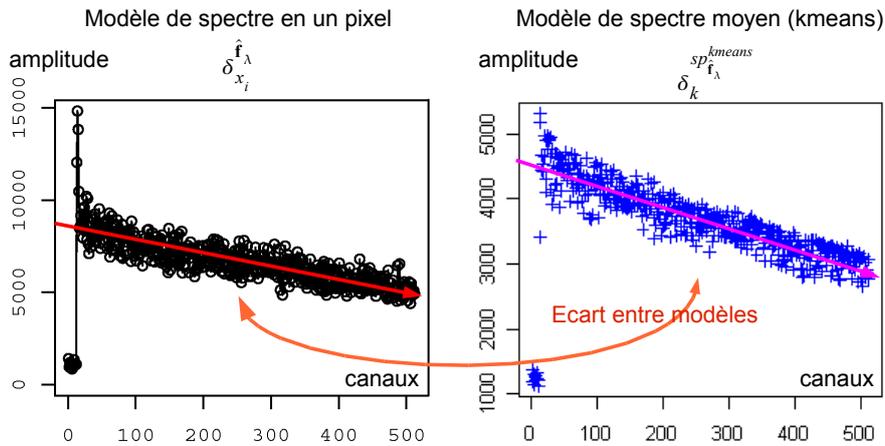


FIG. 2.49 – Minimisation de la distance d_1 entre modèle des spectres moyens $\delta_k^{sp_{\hat{f}_\lambda}^{kmeans}}$ et modèle des spectres en chaque pixel $\delta_{x_i}^{\hat{f}_\lambda}$.

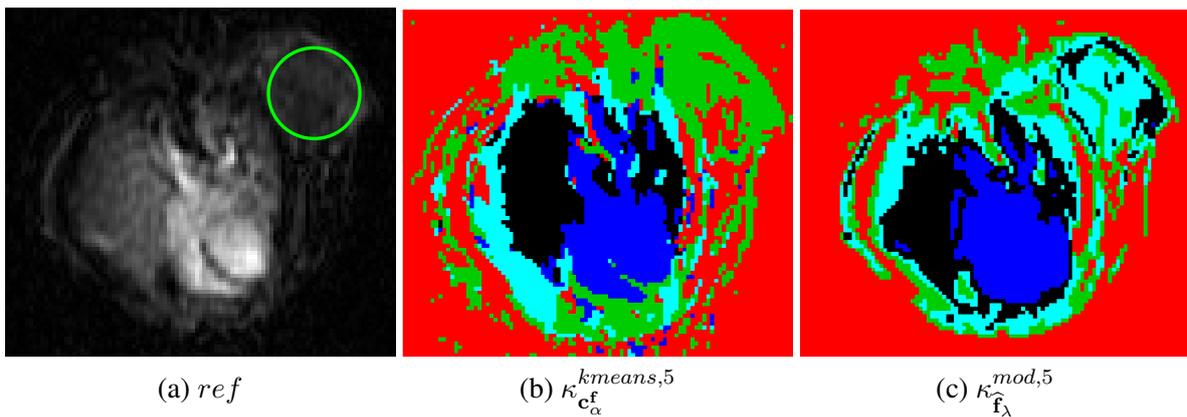


FIG. 2.50 – (a) Référence *ref* des médecins, (b) classification *kmeans* en 5 classes $K_{c_\alpha}^{kmeans,5}$, et (c) classification par approche modèle $K_{\hat{f}_\lambda}^{mod,5}$ en 5 classes sur l’image filtrée par AFC \hat{f}_λ .

2.6.4 Analyse discriminante linéaire (LDA)

Comme expliqué dans la partie 2.5.4 (p : 55), la LDA requiert un ensemble d’entraînement. Pour ce faire des pixels appartenant de manière certaine à quatre parties distinctes de l’anatomie de la souris, sont sélectionnées :

- la tumeur en vert $train(vert) = t_1$;
- les cavités cardiaques, encore appelées cœur, en bleu $train(bleu) = t_2$;
- le fond en rouge $train(rouge) = t_3$;
- les poumons en noir ou en blanc $train(noir) = t_4$.

Chaque classe de l’ensemble d’entraînement $T = train = (t_1, t_2, t_3, t_4)$ contient 80 pixels. On mesure également les spectres moyens de l’image filtrée \hat{f}_λ sur ces classes $sp_{\hat{f}_\lambda}^{train}$ (fig. 2.51). On constate que les classes de l’ensemble d’entraînement, *train*, ont des cinétiques différentes.

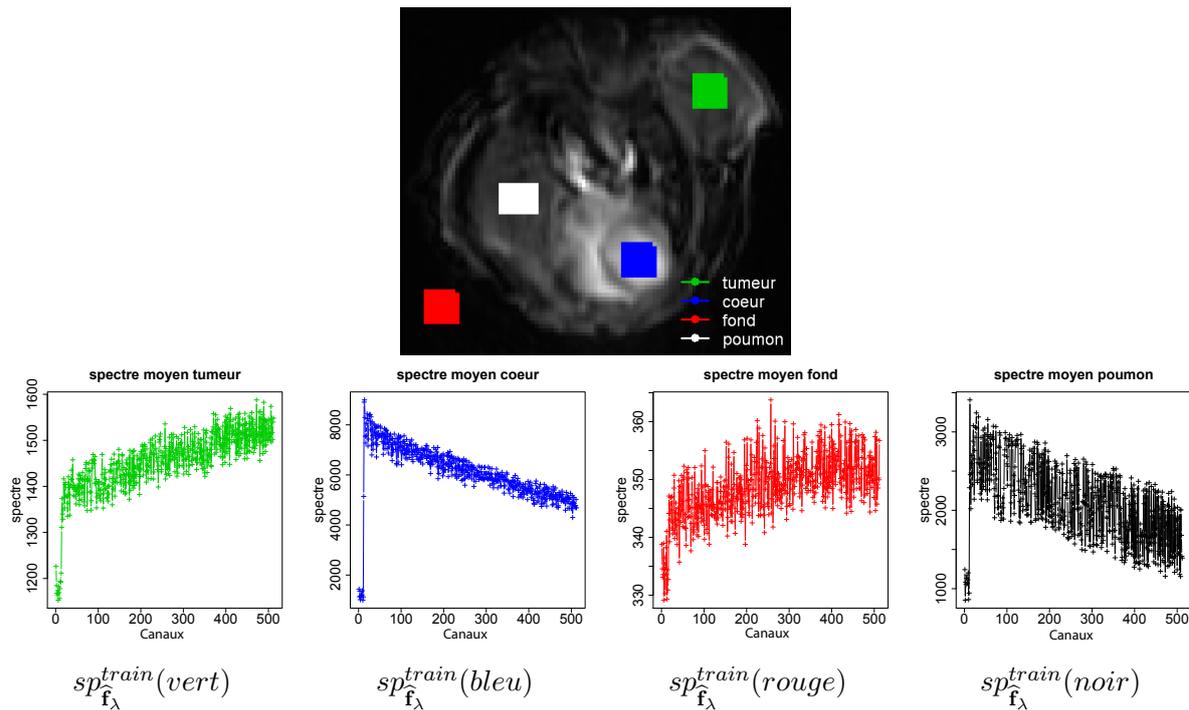


FIG. 2.51 – Classes de l'ensemble d'entraînement, $train$, et spectres moyens $sp_{\hat{f}_\lambda}^{train}$, sur l'image filtrée \hat{f}_λ , associés à ces classes, pour l'image « serim447 ».

On va appliquer une classification LDA dans différents espaces (fig. 2.52) :

- les facteurs pixels ACP des spectres d'entraînement : $\zeta(\{\hat{f}_\lambda(x)|x \in T\})$
- sur les spectres d'entraînement : $\{\hat{f}_\lambda(x)|x \in T\}$
- sur les paramètres des spectres d'entraînement : $\{\mathbf{p}(x)|x \in T\}$

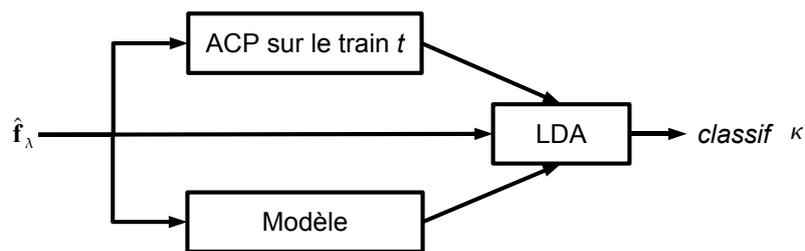


FIG. 2.52 – Schéma général de la classification supervisée par LDA.

Afin de connaître l'erreur d'entraînement et l'erreur de test, on effectue une validation croisée à cinq sous-jeux de données « 5-fold cross-validation » [76]. Cette validation croisée consiste à séparer l'ensemble d'entraînement en $n = 5$ sous-jeux contenant environ autant de pixels de chaque classe t_k après un mélange aléatoire dans chacune des classes. Ensuite un apprentissage du classificateur est effectué sur $n - 1$ sous-jeux $train^{CV}$ et une validation sur le dernier sous-jeu. On effectue ensuite toutes les combinaisons possibles entre les n sous-jeux pour en avoir $n - 1$ pour l'entraînement et

un pour la validation. L'erreur d'entraînement est estimée sur l'ensemble d'entraînement, formé des $n - 1$ sous-jeux, et l'erreur de test est estimée sur l'ensemble de validation. Comme on connaît la vérité donnée par l'utilisateur on peut mesurer les erreurs.

Pour le cas étudié, on considère 4 classes $T = \{t_1, t_2, t_3, t_4\}$ de 80 points, c'est à dire 320 spectres. On divise ces 320 spectres en $k = 5$ sous-jeux de 64 spectres chacun, pour la validation croisée, $\mathcal{G} = \{g_1, g_2, g_3, g_4, g_5\}$. Chaque partie contient autant de spectres de chacune des 4 classes $T = \{t_1, t_2, t_3, t_4\}$.

On effectue une classification sur l'ensemble d'entraînement $train^{CV}$. La validation est obtenue par projection des points de l'ensemble de validation $valid^{CV}$ sur la classification.

L'erreur d'entraînement err^{train} est définie par :

$$err^{train} = \frac{1}{card(train^{CV})} \sum_{i=1}^{card(train^{CV})} L(y_i, \kappa(\hat{f}_\lambda(x_i)|x_i \in train^{CV})) \quad (2.6.3)$$

avec y_i la classe du point $x_i \in train^{CV}$ définie par l'opérateur, $\kappa(\hat{f}_\lambda(x_i)|x_i \in train^{CV})$ la classe du point x_i donnée par le classificateur et $L(p, q)$ la fonction coût qui vaut 1 si $p = q$ et 0 sinon.

L'erreur de test err^{test} est définie par :

$$err^{test} = \frac{1}{card(valid^{CV})} \sum_{i=1}^{card(valid^{CV})} L(y_i, \kappa(\hat{f}_\lambda(x_i)|x_i \in valid^{CV})) \quad (2.6.4)$$

2.6.4.1 LDA sur les facteurs pixels ACP

On effectue une ACP sur l'ensemble d'entraînement de la validation croisée $train^{CV}$, par exemple $train^{CV} = \{g_n\}_{n=1,2,3,4}$, et on fait apprendre le classificateur LDA. Des classes sont obtenues sur cet ensemble. On projette ensuite les points de l'ensemble de validation $valid^{CV} = g_5$ sur ces classes.

Pour l'ACP, on ne conserve que les 3 premiers axes avec les parts d'inerties suivantes : 97,7% ; 1,51 % et 0,68 %.

Les résultats obtenus par validation croisée sont présentés dans le tableau 2.2. On observe que les erreurs d'entraînement err^{train} et de test err^{test} sont toujours nulles pour les différentes validation croisées. Donc, le classificateur sépare très bien les classes.

Validation croisée	CV_1	CV_2	CV_3	CV_4	CV_5
$train^{CV}$	$\{g_n\}_{n=1,2,3,4}$	$\{g_n\}_{n=1,3,4,5}$	$\{g_n\}_{n=1,2,4,5}$	$\{g_n\}_{n=1,2,3,5}$	$\{g_n\}_{n=2,3,4,5}$
$valid^{CV}$	g_5	g_2	g_3	g_4	g_1
err^{train}	0	0	0	0	0
err^{test}	0	0	0	0	0

TAB. 2.2 – Ensembles d'entraînement $train^{CV}$ et de validation $valid^{CV}$ ainsi que les erreurs d'entraînement et de test par validation croisée à 5 sous-jeux de données sur les facteurs pixels ACP de l'ensemble d'entraînement $\zeta(\{\hat{f}_\lambda(x)|x \in T\})$.

Après avoir mesuré les erreurs d'entraînement err^{train} et de test err^{test} , on effectue une classification LDA.

Pour ce faire, le classificateur LDA est entraîné sur les facteurs pixels ACP des spectres d'entraînement : $\zeta(\{\hat{\mathbf{f}}_\lambda(x)|x \in T\})$ noté aussi \mathbf{c}_β^{train} (fig. 2.53). On constate qu'effectivement les classes sont assez bien séparées.

On projette ensuite les autres spectres dans cet espace ACP comme points supplémentaires. On prédit ensuite les classes de ces spectres grâce au classificateur (fig. 2.54). La classification obtenue semble assez bonne puisque l'on aperçoit nettement les cavités cardiaques (en bleu) et la tumeur qui correspond à la grosse zone verte.

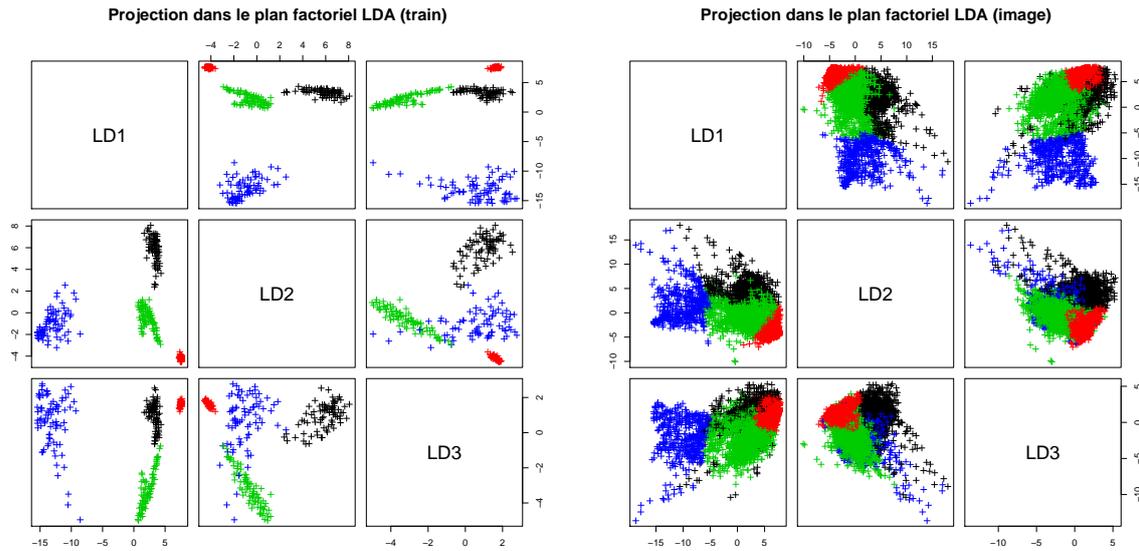


FIG. 2.53 – Espace LDA des facteurs pixels ACP des spectres de l'ensemble d'entraînement $T = \{t_1, t_2, t_3, t_4\}$. Les 4 classes sont les suivantes : tumeur (vert), cœur (bleu), fond (rouge) et poumon (noir). (a) Classification sur l'ensemble d'entraînement et (b) classification sur l'ensemble des pixels.

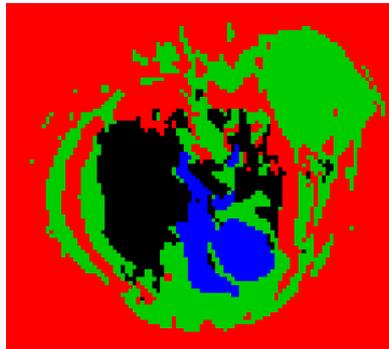


FIG. 2.54 – Classification LDA $\kappa_{\zeta(\{\hat{\mathbf{f}}_\lambda(x)|x \in T\})}^{LDA,4}$ en 4 classes des spectres de l'image filtrée $\hat{\mathbf{f}}_\lambda$ projetés dans l'espace factoriel ACP des spectres de l'ensemble d'entraînement $\zeta(\{\hat{\mathbf{f}}_\lambda(x)|x \in T\})$.

2.6.4.2 LDA sur l'image

On essaye également une LDA directement sur l'image filtrée \hat{f}_λ sans appliquer d'AFC (fig. 2.55). La classification se fait donc dans un espace à 512 dimensions ; ce qui n'est pas optimal pour la précision du classificateur et pour les calculs. Les erreurs d'entraînement et de classification sont toutes les deux nulles. La classification obtenue est quasiment identique à la précédente, ce qui montre que la réduction de dimension par ACP à 3 axes, sur l'ensemble des spectres d'entraînement, est correcte.

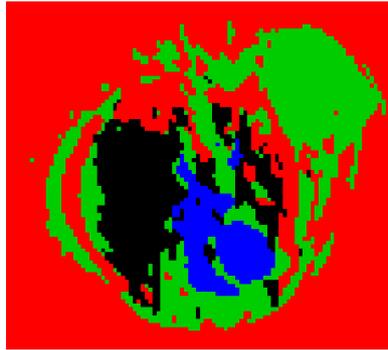


FIG. 2.55 – Classification LDA $\kappa_{\hat{f}_\lambda}^{LDA,4}$ sur toute l'image après entraînement sur les spectres de l'image filtrée : $\{\hat{f}_\lambda(x)|x \in T\}$.

2.6.4.3 LDA sur les cartes de paramètres

On effectue, en outre, une LDA directement sur les paramètres du modèle : pente a , ordonnée à l'origine b et montée m . Dans le tableau 2.3, sont présentées les erreurs d'entraînements et de test obtenues par validation croisée. On constate que ces erreurs sont quasiment nulles.

Validation croisée	CV_1	CV_2	CV_3	CV_4	CV_5
$train^{CV}$	$\{g_n\}_{n=1,2,3,4}$	$\{g_n\}_{n=1,3,4,5}$	$\{g_n\}_{n=1,2,4,5}$	$\{g_n\}_{n=1,2,3,5}$	$\{g_n\}_{n=2,3,4,5}$
$valid^{CV}$	g_5	g_2	g_3	g_4	g_1
$err^{train} (\%)$	0	0,39	0,39	0	0
$err^{test} (\%)$	0	1,56	1,56	0	0

TAB. 2.3 – Ensembles d'entraînement $train^{CV}$ et de validation $valid^{CV}$ ainsi que les erreurs d'entraînement et de test par validation croisée à 5 sous-jeux de données sur les paramètres de l'ensemble d'entraînement $\{\mathbf{p}(x)|x \in T\}$.

Les résultats de la classification sur l'ensemble de l'image formée par les paramètres sont visibles sur la figure 2.56. On constate que la classification donne des résultats similaires aux classifications LDA précédentes, même si les résultats sur l'espace des paramètres \mathbf{p} semblent un peu moins bons que dans l'espace image et dans l'espace ACP des spectres d'entraînement.

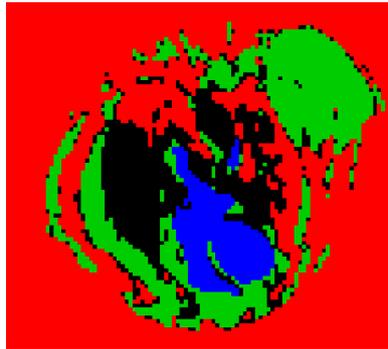


FIG. 2.56 – Classification LDA $\kappa_{\mathbf{p}}^{LDA,4}$ sur toute l'image formée par les paramètres après entraînement sur les paramètres des spectres : $\{\mathbf{p}(x)|x \in T\}$.

2.6.5 Classification sur des images similaires

La classification d'images similaires à une image ne pose aucun problème en classification non supervisée, i.e. *kmeans* ou approche modèle. En classification supervisée, i.e. LDA, le niveau de gris moyen n'étant pas le même d'une image à l'autre, il est nécessaire d'effectuer une normalisation d'histogramme si l'on veut entraîner le classificateur sur un sous-ensemble d'une image et tester sur les autres.

Nous disposons de 25 séquences d'images IRM de souris « serimxxx ». Après filtrage par composition de deux AFC-reconstitutions avec le même seuil de rapport signal à bruit sur les facteurs pixels, nous avons effectué une classification.

Pour les classifications non supervisées, on applique directement les classificateurs *kmeans* ou par approche modèle en 5 classes (fig. 2.57). Les classifications sont correctes par rapport à la référence et donnent des résultats comparable aux classifications non supervisées de l'image de souris « serim447 ».

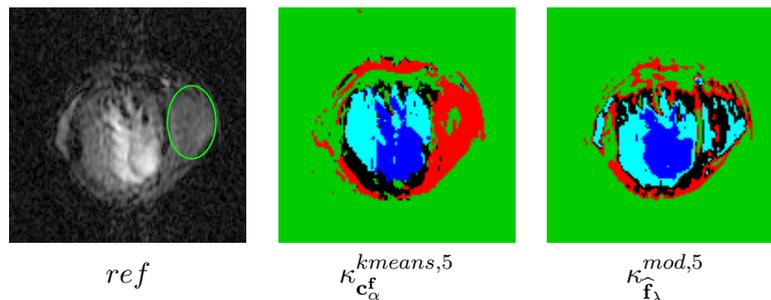


FIG. 2.57 – Classifications *kmeans* $\kappa_{c_{f_{\alpha}}}^{kmeans,5}$ et par approche modèle $\kappa_{f_{\lambda}}^{mod,5}$ sur une autre images de souris « serim460 ».

On essaye directement la classification supervisée par LDA avec entraînement sur les facteurs ACP de l'ensemble des pixels d'entraînement de l'image « serim447 » $\kappa_{\zeta(\hat{f}_{\lambda}(x)|x \in T_{s447})}^{LDA,4}$ (cf. partie 2.6.4.1, p : 65). On teste ensuite la LDA sur l'image « serim460 » en projetant ses spectres filtrés \hat{f}_{λ} dans l'espace ACP de l'ensemble d'entraînement $\kappa_{\mathbf{p}(x)|x \in T_{s447}}^{LDA,4}$. Les classifications obtenues sont aberrantes (fig. 2.58). Il en est de même sur l'image des paramètres.

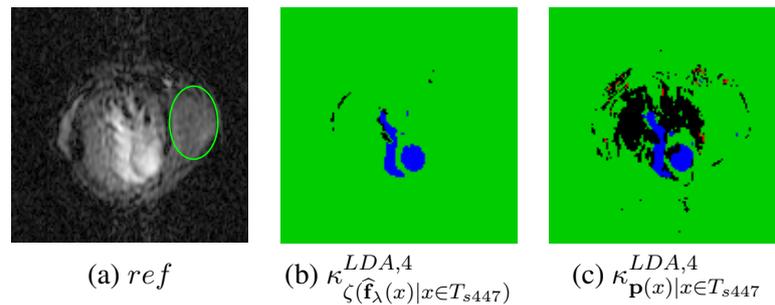


FIG. 2.58 – (a) Référence (b) Classification par LDA $\kappa_{\zeta(\hat{\mathbf{f}}_\lambda(x)|x \in T_{s447})}^{LDA,4}$ sur l'image « serim460 » projetée dans l'espace factoriel ACP des spectres d'entraînement de l'image « serim447 » (c) Classification par LDA sur les paramètres de l'image « serim460 » avec entraînement sur l'image « serim447 ».

En comparant les spectres moyens de zones homogènes, i.e. tumeur, cœur, fond et poumon, entre les images « serim447 » et « serim460 », on constate que les dynamiques des spectres ne sont pas les mêmes (fig. 2.59). Ceci pose des problèmes pour la segmentation supervisée.

Afin d'améliorer les segmentations on souhaite utiliser les cartes de paramètres qui semblent plus robustes. Pour cela, on observe les fonctions de répartition à partir des histogrammes de chaque paramètre (fig. 2.60). Les histogrammes sont constitués de 255 classes.

On constate qu'il faut normaliser les fonctions de répartition des niveaux de gris. Pour chacune des classes en abscisse de la fonction de répartition d'un paramètre, les niveaux de gris des pixels sont mis à la valeur moyenne des niveaux de gris des pixels de la classe correspondante de la fonction de répartition de l'image référence.

On choisit comme référence l'image « serim447 » sur laquelle l'apprentissage du classificateur a été effectué. Après normalisation, les fonctions de répartition se superposent bien (fig. 2.60). L'image des paramètres normalisés est notée $\tilde{\mathbf{p}}$.

La classification obtenue par LDA $\kappa_{\tilde{\mathbf{p}}}^{LDA,4}$ (fig. 2.61) est bien meilleure qu'avant la normalisation des paramètres.

On compare également les classifications *kmeans*, par approche modèle et LDA sur d'autres images (fig. 2.62). On constate que les classifications par approche modèle sont plus robustes que celles obtenues par *kmeans*. La LDA fournit également de bonnes classifications qui présentent bien les limites des cavités cardiaques (i.e. cœur). En outre sur les classifications LDA, les tumeurs ressortent comme étant les gros « amas » de la classe verte.

Remarque 2.6.1. Sur la figure 2.62, les classes de classifications non supervisées, *kmeans* et approche modèle, n'ont pas les mêmes couleurs que la classification LDA car comme ce sont des classifications non supervisées on ne sait pas à priori à quelle partie anatomique (par exemple le cœur) correspond une classe. Ceci n'est pas le cas pour la LDA, puisqu'en raison de l'apprentissage, on connaît a priori à quoi correspondent chacune des classes. Pour les classifications non supervisées, on pourrait déterminer à quelle partie anatomique correspond une classe par comparaison des spectres moyens à des spectres de référence.

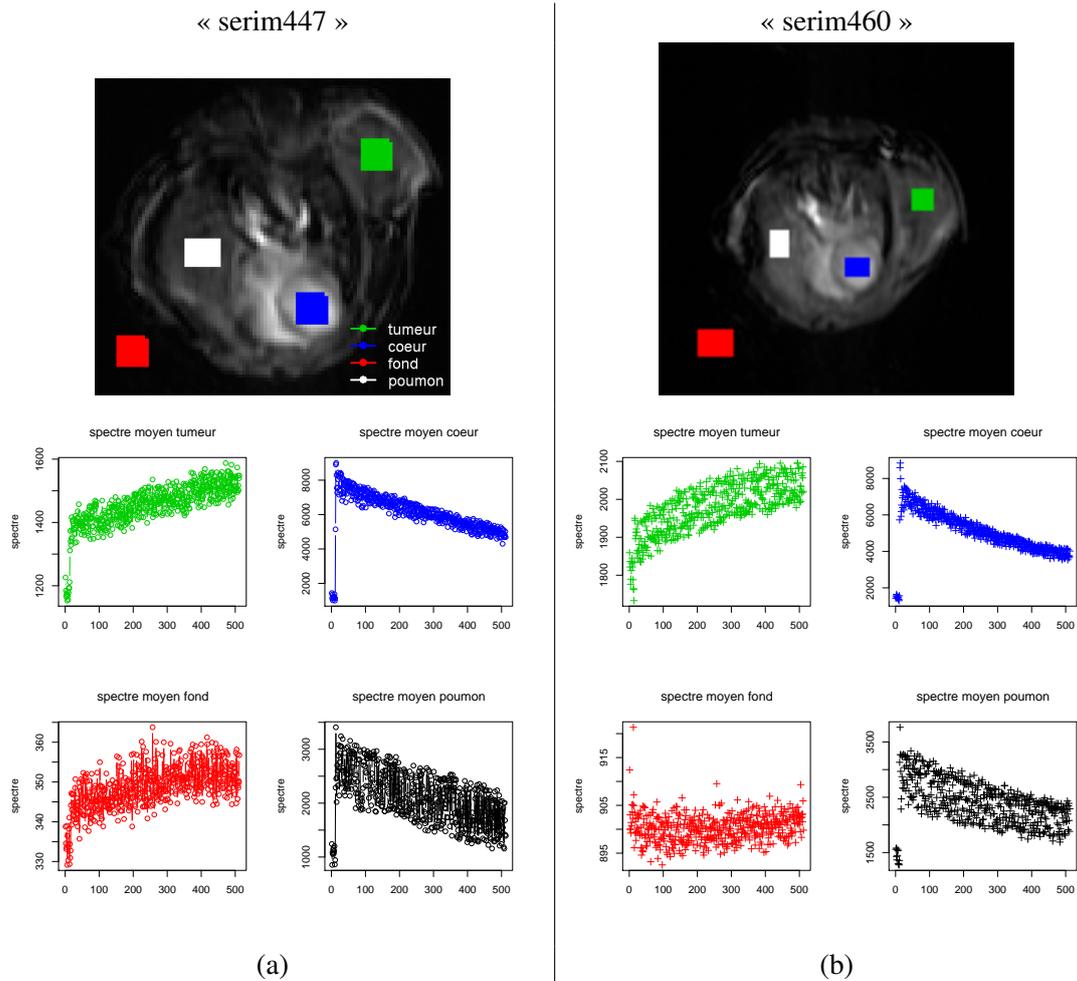
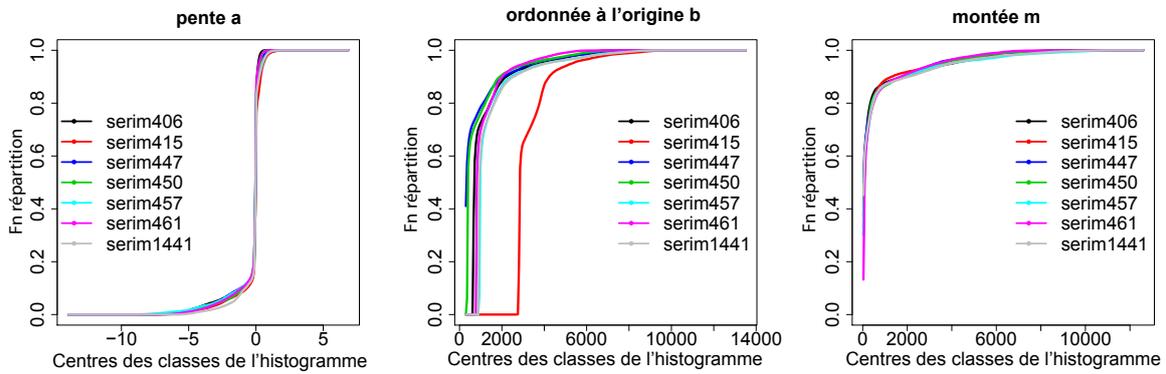
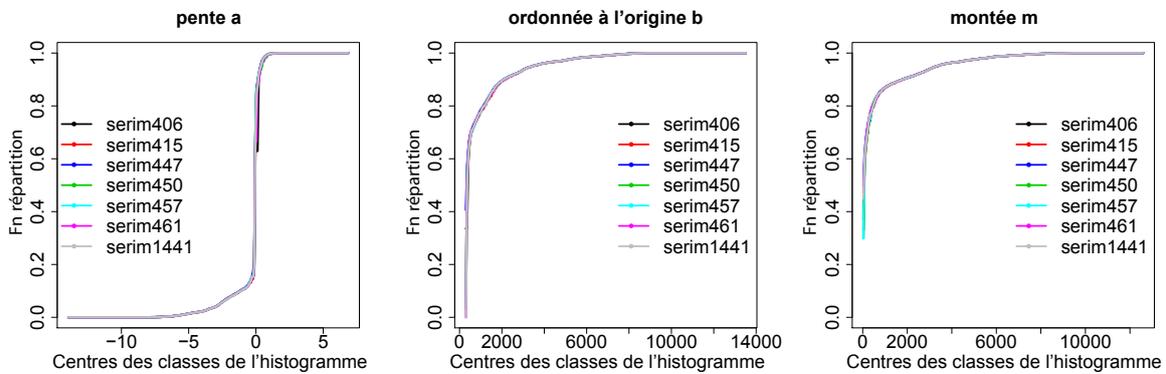


FIG. 2.59 – Zones sélectionnées et spectres moyens associés pour les images « serim447 » et « serim460 » après deux AFC-reconstitutions \hat{f}_λ (un zoom a été effectué sur l'image « serim447 »).



Fonctions de répartition non recalées



Fonctions de répartition recalées

FIG. 2.60 – Fonctions de répartition, des paramètres, non normalisées (en haut) et normalisées (en bas) avec comme référence l’image « serim447 ».

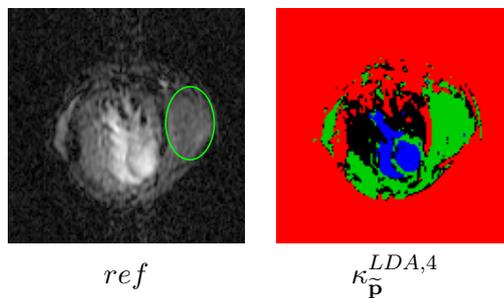


FIG. 2.61 – Classifications LDA en 4 classes sur les cartes de paramètres de l’image « serim460 » après normalisation des niveaux de gris et apprentissage sur les cartes de paramètres de l’image « serim447 » $\kappa_{\tilde{\mathbf{p}}}^{LDA,4}$.

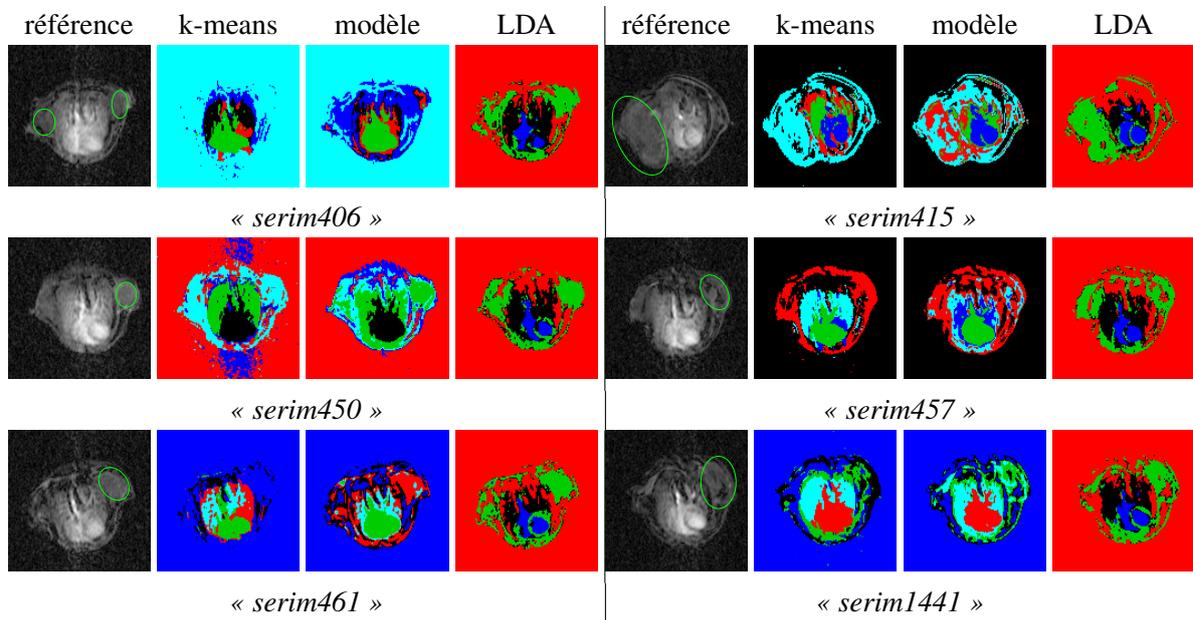


FIG. 2.62 – Classifications $kmeans$ $\kappa_{\mathbf{c}_\alpha}^{kmeans,5}$, par approche modèle $\kappa_{\mathbf{f}_\lambda}^{mod,5}$ et LDA $\kappa_{\mathbf{p}}^{LDA,4}$ sur d'autres images de souris « serimxxx ».

2.6.6 Regroupement des canaux

Afin de réduire la dimension spectrale, il peut être intéressant de regrouper les canaux. L'AFC, contrairement à l'ACP, se prête à ce regroupement de canaux grâce au principe d'équivalence distributionnelle (cf. propriété A.4.1 en annexes, p : 207). On peut ainsi remplacer deux canaux quasiment proportionnels par un canal somme des deux canaux précédents, sans que les résultats de l'AFC soient changés.

Pour ce faire, on classe les canaux de l'image « woman face » dans l'espace factoriel AFC en 5 classes grâce à k-means (fig. 2.63). On somme ensuite les canaux de chacune des classes et on obtient une image à 5 canaux f_λ^{5can} . (fig. 2.64). On compare ensuite les facteurs pixels associés aux trois premiers axes sur l'image à 61 canaux f_λ^{61can} et sur l'image à 5 canaux f_λ^{5can} (fig. 2.65). Pour ce faire, on mesure la somme des carrés des écarts moyens $SCEM$ entre les facteurs pixels (eq. 2.6.5).

$$SCEM_{\alpha_k} = \frac{\sum_{i=1}^P \left(c_{\alpha_k}^{61can}(x_i) - c_{\alpha_k}^{5can}(x_i) \right)^2}{P} \quad (2.6.5)$$

On constate que les $SCEM$ sont toutes quasiment nulles (fig. 2.65). Par conséquent, on peut effectuer une AFC sur l'image à 5 canaux regroupés f_λ^{5can} sans changement notable par rapport à l'image à 61 canaux f_λ^{61can} .

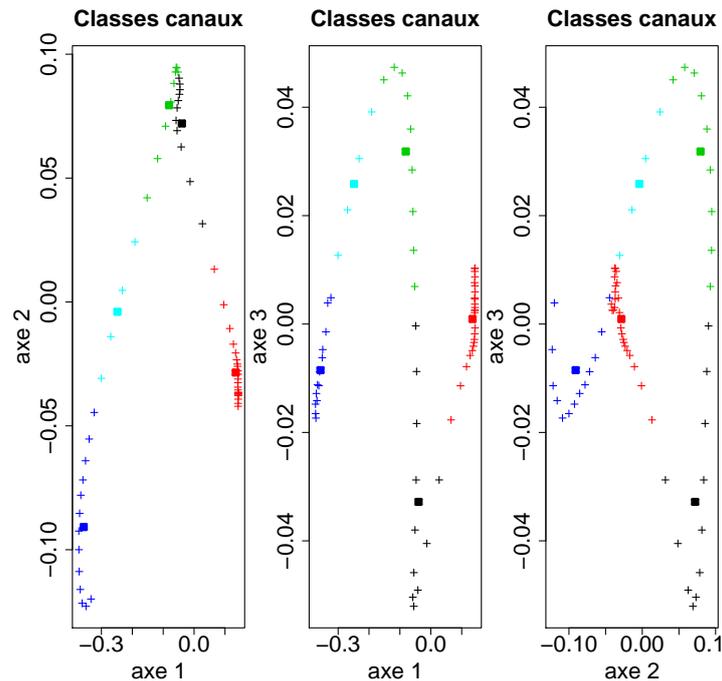


FIG. 2.63 – Classification *kmeans* en 5 classes des facteurs canaux d^α dans le plan factoriel AFC.

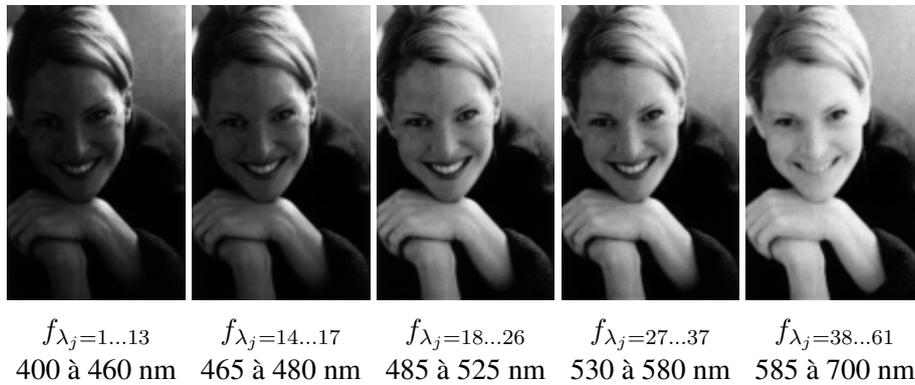


FIG. 2.64 – Image « woman face » après réduction à 5 canaux ($195 \times 91 \times 5$). Les nouveaux canaux sont la somme des anciens.

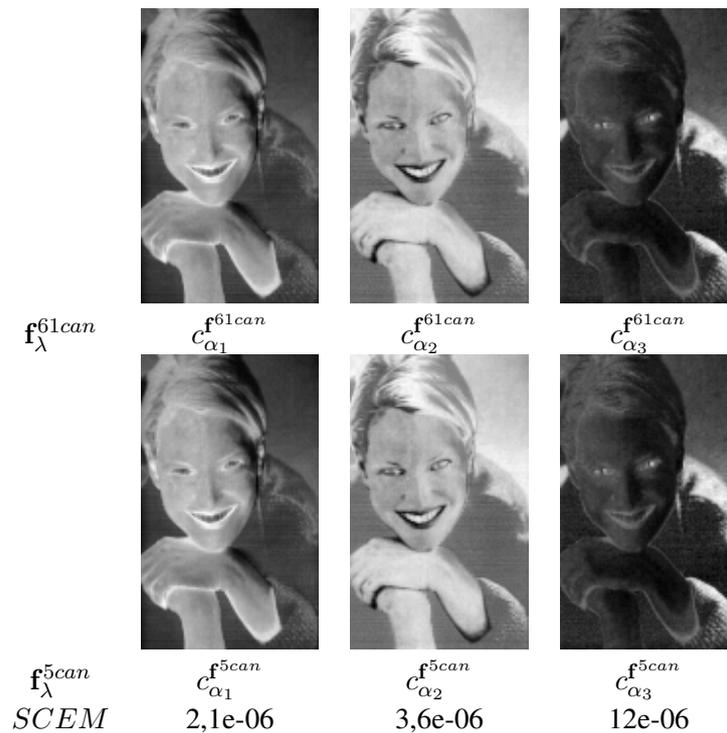


FIG. 2.65 – Facteurs pixels sur l'image « woman face » à 61 canaux f_{λ}^{61can} et à 5 canaux f_{λ}^{5can} et somme des carrés des écarts moyens (SCEM) entre les facteurs pixels.

2.7 Conclusion du chapitre

Dans ce chapitre, nous avons relevé plusieurs points importants concernant le début d'une chaîne de traitement pour les images hyperspectrales. Il est ainsi nécessaire de réduire la dimension de l'espace afin de lutter contre le phénomène de Hughes [81] et d'améliorer les temps de traitement. Pour ce faire on utilise une approche par composition d'Analyse Factorielle des Correspondances et de reconstitutions qui permet de réduire le bruit spectral sur les images bruitées. L'espace de dimension réduite peut être constitué des facteurs pixels de l'espace factoriel ou des cartes de paramètres d'un modèle ajusté sur l'image filtrée par AFC. On effectue ensuite une classification spectrale par une méthode supervisée ou non supervisée. L'avantage de la méthode supervisée est qu'elle introduit une connaissance supplémentaire qui améliore la classification et son inconvénient est qu'elle n'est pas complètement automatique.

On pourrait se contenter des classifications spectrales obtenues, cependant leurs classes ne sont pas connexes et en outre elles tiennent compte uniquement de l'information spectrale. Afin de résoudre ces deux problèmes nous allons ajouter une étape de segmentation qui permet d'introduire l'information spatiale et qui garantit la connexité des régions segmentées.

Chapitre 3

Segmentation hyperspectrale par Ligne de Partage des Eaux

Après avoir filtré le bruit, réduit la dimension et classifié les pixels des images hyperspectrales, on cherche à avoir des classes connexes. En effet la classification fournissant des classes non connexes, il est nécessaire d'introduire une étape de segmentation afin d'obtenir une partition connexe. Quelques notions générales sur la théorie de la segmentation sont rappelées en annexe B. En outre, dans la classification seule la dimension spectrale a été considérée, il est donc intéressant d'introduire la dimension spatiale grâce à la segmentation.

Après un état de l'art sur la segmentation des images hyperspectrales, nous présenterons la problématique de la segmentation morphologique d'images hyperspectrales, puis l'obtention de marqueurs provenant de la classification spectrale et des gradient hyperspectraux pour la segmentation par Ligne de Partage des Eaux.

Nous introduirons ensuite un algorithme innovant de segmentation des images hyperspectrales : la Ligne de Partage des Eaux Probabiliste. Nous montrerons que cet algorithme est particulièrement adapté pour renforcer et lisser les contours des régions les plus importantes, même faiblement contrastées. Nous montrerons également qu'il est possible, voire très performant, de conditionner cette segmentation spatiale par la classification spectrale, ce qui conduit à une méthode spatio-spectrale de segmentation.

Puis, nous présenterons une application de cet algorithme pour la détection de tumeurs sur des images DCE-MRI bruitées de souris, avec une faible résolution spatiale et un faible contraste au niveau de la tumeur.

Finalement, nous montrerons un schéma général de segmentation des images hyperspectrales par LPE probabiliste à partir d'un procédé dont le seul paramètre est le nombre de classes spectrales de la classification, qui est un paramètre robuste.

3.1 État de l'art en segmentation des images hyperspectrales

Dans cet état de l'art, à l'exception de la classification par profil morphologique, on ne tient pas compte des méthodes par classification qui ont été présentées dans le chapitre 2. En effet, nous considérons la segmentation comme une partition de classes connexes. La classification fournit quant à elle des classes non-connexes. Nous allons ainsi succinctement présenter des méthodes pouvant être utilisées pour la segmentation des images hyperspectrales ou multivariées.

3.1.1 La Ligne de Partage des Eaux

En morphologie mathématique, la méthode la plus classique et puissante de segmentation est la Ligne de Partage des Eaux (LPE) (« Watershed » en anglais), calculée sur l'image gradient à partir des marqueurs des régions à extraire. Dans cette partie nous rappelons brièvement le principe de cet outil de segmentation. Plus de détails sont donnés en annexe (sec. B, p : 233).

La Ligne de Partage des Eaux, comme son nom le laisse imaginer, trouve son origine dans une interprétation topographique de l'image. La LPE a été introduite et formalisée par *Lantuejoul et Beucher* [100, 28, 25, 29], avec des contributions très importantes de *Meyer et Beucher* sur la notion de marqueurs [126] ainsi que sur son implémentation par files d'attente hiérarchiques [117], puis des approches hiérarchiques multi-échelles à base de graphes [118, 121, 122]. L'algorithme original a été amélioré par *Vincent et Soille* [188].

Plusieurs auteurs ont essayé des méthodes de segmentation des images hyperspectrales par LPE. *Flouzat et al.* [57] utilisent une segmentation spatio-spectrale à partir d'un graphe d'adjacence de régions construit sur l'image. Dans [109], les auteurs calculent un gradient morphologique multi-échelles par sommation de gradients morphologiques avec des tailles croissantes d'éléments structurants. Comme la LPE nécessite la norme d'un gradient (i.e. une image scalaire), les gradients sont sommés entre eux avec les mêmes poids. Après un filtrage morphologique sur le gradient, les minima locaux sont utilisés pour la LPE.

De la même manière *Scheunders* [142] calcule un gradient par sommation des gradients marginaux (i.e. sur chacun des canaux) avant d'effectuer une LPE. *Soille* [1] combine la classification spectrale sur des histogrammes et la segmentation spatiale. L'histogramme multivarié est segmenté par la LPE pour obtenir une image classifiée. Sur l'image classifiée les minima du gradient de l'image hyperspectrale sont fixés par une connaissance a priori. La classification est ensuite filtrée pour obtenir des marqueurs. Avec le gradient et les marqueurs, il effectue une segmentation par LPE.

3.1.2 Classification des profils morphologiques

Pour analyser les structures d'une image, *Benediktsson et al.* [19] ont développé le Profil Morphologique (PM) [146] qui provient du principe de la granulométrie [159, 172]. Le profil morphologique est composé du Profil d'Ouverture (PO), qui consiste en un ensemble d'ouvertures par reconstruction de taille croissante, et du Profil de Fermeture (PF), obtenu par l'opération duale [172]. Les ouvertures et fermetures par reconstruction sont des opérateurs connexes qui satisfont la propriété suivante : si la structure de l'image ne peut contenir l'élément structurant, alors elle est totalement supprimée, sinon elle est entièrement conservée. Pour un élément structurant donné, l'ouverture et la fermeture géodé-

siques permettent de connaître la taille ou la forme des objets présents dans l'image. Ceux qui sont supprimés sont plus petits que l'élément structurant, tandis que ceux qui sont plus grands sont conservés. Pour déterminer la taille ou la forme de tous les éléments présents dans l'image, il est nécessaire d'utiliser différentes tailles d'éléments structurants. Cette hypothèse conduit au PM :

$$PM(x) = \{PF_k(x), \dots, f(x), \dots, PO_k(x)\} \quad (3.1.1)$$

Les informations spatiales (taille, orientation, contraste local) sont incluses dans le PM. Cependant l'information spectrale n'est pas prise en compte. Une méthode simple pour résoudre ce problème est de construire le PM sur chacun des canaux d'une image hyperspectrale. Cette approche est appelée Profil Morphologique Étendu (PME).

Afin de développer une approche morphologique étendue, *Benediktsson et al.* [18] proposent d'utiliser les premières composantes principales (c_{α}^f) de l'image hyperspectrale. Les composantes principales $c_{\alpha_k}^f$ sont sélectionnées par leurs inerties. Les PM sont construits sur chacune des composantes principales. Le profil morphologique étendu résultant est ainsi la pile des différents profils morphologiques, qui peut être utilisée comme une collection de paramètres pour la classification. D'après la notation précédente (eq. 3.1.1), le PME à la position x peut être simplement représenté par :

$$PM_{\text{étendu}}(x) = \{PF_{c_{\alpha_1}^f}(x), \dots, f(x), \dots, PF_{c_{\alpha_k}^f}(x)\} \quad (3.1.2)$$

Il est intéressant de remarquer que le profil morphologique étendu prend en compte l'information spatio-spectrale. Un classificateur, comme les SVM ou un réseau de neurones [18, 55], est ensuite appliqué sur le PME. Les classes obtenues ne sont pas connexes, il ne s'agit donc pas d'une segmentation comme définie dans cette thèse.

3.1.3 D'autres techniques de segmentation hyperspectrales non morphologiques

Il existe de nombreuses autres méthodes de segmentations qui peuvent être utilisées pour des images hyperspectrales.

- le seuillage [141, 62] ;
- la détection de contour (comme la LPE) ;
- la segmentation fondée sur la croissance de régions (comme la LPE) ;
- la segmentation par « Mean Shift », e.g. [43, 61, 144] ;
- la segmentation par coupe de graphes, e.g. [31] ;
- la segmentation à partir de modèles, e.g. [129, 16] ;
- les méthodes de segmentation à partir de techniques probabilistes comme les « Markov Chain Monte Carlo » (MCMC), e.g. [181], les champs de Markov ou les arbres de Markov, e.g. [56, 144], ou encore les champs aléatoires, e.g. [91, 70].

Pour un état de l'art plus exhaustif sur la segmentation on peut consulter *M. Sonka* [174].

3.2 Comment segmenter une image hyperspectrale avec la LPE ?

Ayant choisi d'utiliser dans ce travail de thèse une approche morphologique à partir de la Ligne de Partage des Eaux (LPE), nous allons présenter cette méthode et une méthodologie d'application aux images hyperspectrales.

3.2.1 Introduction à la LPE

Dans cette partie après avoir rappelé la définition de la LPE, nous expliquons les différentes manières de choisir les marqueurs : par imposition des minima ou par l'utilisation des fonctions d'extinction. De plus amples précisions sont données dans l'annexe B.5.

3.2.1.1 Définition de la LPE

La notion de LPE est étroitement liée à celle de minimum régional et de bassin versant. Considérons, le relief de la fonction numérique f sous la pluie. Soit M un minimum régional de f . Le bassin versant associé à M , noté $BV(M)$, est l'ensemble des points x tels qu'une goutte d'eau tombant en x rejoint finalement le fond de la vallée associée à M . La notion de bassin versant permet d'associer à chaque minimum régional une portion de l'image : la vallée qui lui correspond. L'ensemble des bassins versants associés à chaque minimum régional définit une partition de l'image. Il existe une autre manière d'interpréter la LPE qui permet de régler les indéterminations des écoulements, e.g. où va la goutte d'eau sur les plateaux de pixels ou bien sur les pixels ayant plus d'un voisin avec le niveau de gris le plus faible. Imaginons que de l'eau jaillisse de chaque minimum régional et que la surface soit inondée à partir de ces sources. Progressivement, le niveau de l'eau s'élève. Pour empêcher le mélange des eaux venants de minimum différents, on crée un barrage élémentaire en chaque point de contact. A la fin, ne restent que les digues achevées, entourées de l'eau : c'est la LPE. Dans ce qui suit, on notera par $LPE(f)$ la LPE de la fonction f .

Il faut noter que la LPE est un ensemble de zones disjointes, partitionnant l'espace de définition de l'image. C'est à dire que la LPE fournit une partition.

La LPE est généralement calculée non pas sur l'image originale mais sur son gradient. De cette manière, la LPE correspond aux points crêtes du gradient autour des minima, c'est à dire aux lieux de forte transition d'intensité de la fonction initiale. Évidemment, les zones de fortes transition sont associées aux frontières ou contours des régions homogènes de l'image. L'image sur laquelle est appliquée la LPE, en général le gradient, est appelée fonction à segmenter ou fonction à inonder.

Meyer [118] a proposé des méthodes fondées sur des graphes qui ont abouti à des algorithmes de segmentation hiérarchiques multi-échelles très rapides et efficaces, voir les thèses de *Marcotegui* [112], *Gomilla* [63] et *Zanoguera* [190].

Il est à noter que le nombre de parties de la segmentation est fixé par le nombre de sources. Pour choisir ce nombre de parties, on peut choisir a priori ou par filtrage le nombre de minima ou bien sélectionner les minima sur des critères morphologiques grâce aux fonctions d'extinction. On parle alors de LPE contrôlée par les marqueurs.

3.2.1.2 Choix des minima de la fonction à segmenter

Par définition, le nombre de régions de la LPE (sans utilisation de marqueurs) est égal au nombre de minima régionaux de l'image gradient (i.e. la fonction à segmenter), et ceci correspond au nombre de plateaux de l'image originale. Étant donné que pour les images réelles il y a un grand nombre de minima présents sur leur gradient, l'effet est une sur-segmentation. Pour régler ce problème, l'idée consiste à calculer la LPE à partir d'un nombre plus faible de minima, associés aux régions les plus importantes. En pratique, on entend par marqueurs la fonction $mrk(x)$, où $mrk(x) = 1$ si x appartient

aux marqueurs et $mrk(x) = 0$ sinon, c'est à dire un ensemble de composantes connexes (les minima de mrk), permettant de localiser les régions qui doivent être segmentées. Une manière de contrôler le résultat de la segmentation est de modifier les minima de l'image sur laquelle on calcule la LPE, c'est à dire d'imposer les minima de mrk comme seuls minima de l'image g . Pour effectuer ce marquage, l'algorithme utilisé consiste en une fermeture par reconstruction, ou reconstruction géodésique, de mrk selon $(f + 1) \wedge mrk$ (cf. annexe B.5.3.2, p : 250) :

$$f' = \varphi^{rec}(mrk, (f + 1) \wedge mrk) .$$

On peut calculer alors la LPE sur la nouvelle fonction g' dans laquelle il y aura un bassin versant associé à chaque marqueur-minimum. Le nombre de régions obtenu est exactement égal au nombre de marqueur-minimum. On peut noter cette transformation par $LPE(g, mrk)$.

Les minima peuvent être obtenus par filtrage de l'image d'entrée ou bien par interaction avec l'utilisateur. Il est donc possible d'introduire une connaissance a priori. La LPE est alors appelée LPE contrôlée par marqueurs. En annexe, plus de détails sont donnés concernant la LPE avec imposition des marqueurs (cf. annexe B.5.3, p : 248).

3.2.1.3 Extraction des marqueurs à l'aide des fonctions d'extinction

Il existe une manière « automatique » de choisir des marqueurs pertinents pour la LPE. On dispose généralement d'informations permettant de caractériser les régions d'intérêt : leur contraste, taille, forme, couleur, ... Extraire automatiquement des marqueurs pour les régions consiste donc à utiliser l'ensemble de ces informations.

Les marqueurs sont des minima, par conséquent, le calcul des fonctions d'extinction des minima permet de sélectionner les minima les plus significatifs selon un critère morphologique donné, comme la dynamique maximale des niveaux de gris d'un bassin versant (BV), la surface de ce BV ou son volume. La sélection des BV se fait par simple seuillage des fonctions d'extinction. La valeur du seuil correspond à la dynamique, à la surface ou au volume du BV. Les valeurs d'extinction sont donc très utiles pour la segmentation automatique des fonctions numériques. Pour une étude détaillée de ces fonction, consulter les thèses de *Grimaud* [69] et *Vachier* [183]).

Donnons maintenant quelques propriétés des fonctions d'extinction. La dynamique est très sensible lorsqu'elle calculée sur une image de gradient, aux variations locales d'intensité sur les lignes de crête du gradient qui peuvent modifier radicalement les valeurs de la dynamique. Les valeurs d'extinction surfaciques calculées sur l'image gradient sont également une mesure de la surface des régions de l'image : contrairement à la dynamique, il n'y a pas de différence d'interprétation entre la mesure de surface calculée sur l'image originale et celle calculée sur l'image gradient, même si ces mesures peuvent être différentes. Les fonctions d'extinction volumiques, qui tiennent compte à la fois de la taille et du contraste des régions, sont aussi très utiles pour la segmentation.

3.2.2 La segmentation des images hyperspectrales par LPE

Segmenter une image par LPE nécessite deux éléments :

1. des marqueurs mrk ;
2. une fonction à inonder g .

Les marqueurs peuvent être déterminés par l'utilisateur, selon un critère morphologique (e.g. volume du bassin versant) ou selon les classes de la classification. Les marqueurs issus de la classification sur les spectres permettent d'introduire l'information spectrale dans la segmentation. Si on utilise une classification spatio-spectrale, comme celle par profils morphologiques, on peut introduire une information sur l'espace et sur le spectre. Dans ce travail de thèse nous avons utilisé uniquement des classifications spectrales, car l'information spatiale a été introduite dans la fonction inonder.

La fonction à inonder est une fonction $g : E \rightarrow \mathcal{T}$, c'est à dire une image numérique (i.e. scalaire) 2D pour une image hyperspectrale 2D, i.e. $E \subset \mathbb{R}^2$. Nous avons ainsi développé une méthode déterministe à partir d'un gradient, ou plutôt la norme d'un gradient. Nous avons aussi introduit une nouvelle méthode probabiliste pour déterminer une fonction de densité de probabilité de contours pdf . Cette pdf est conditionnée par la classification spectrale κ . La pdf est ensuite utilisée comme fonction à inonder.

Une fonction à inonder comme le gradient contient l'information spatiale. Afin d'ajouter de l'information, nous avons introduit une nouvelle fonction à inonder g spatio-spectrale.

Après obtention des marqueurs mrk et de la fonction à inonder g , on applique l'algorithme de LPE qui utilise l'imposition des minima sur la fonction inonder. La segmentation fournie $LPE(f, mrk)$, est la Ligne de Partage des Eaux contrôlée par marqueur de l'image f . Il est également possible d'utiliser les fonctions d'extinction, pour sélectionner les minima sur un critère morphologique (e.g. le volume du bassin versant), à la place des marqueurs. La segmentation résultante est notée $LPE(f, \xi_N)$ où ξ est le critère retenu pour le calcul des valeurs d'extinction ($\xi = C$ pour le contraste, $\xi = S$ pour la surface et $\xi = V$ pour le volume) et N est le nombre de régions retenues selon leur valeur d'extinction.

Le paradigme de segmentation par LPE est présenté sur la figure 3.1.

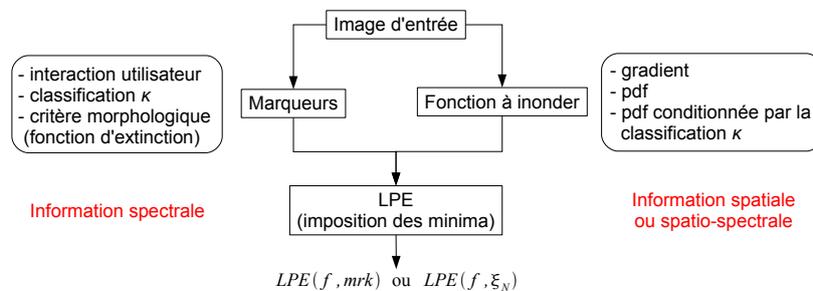


FIG. 3.1 – Schéma général de segmentation des images hyperspectrales par LPE.

3.3 Marqueurs provenant d'une classification

Comme vu précédemment, la LPE requiert des marqueurs qui seront des sources d'inondations de la fonction à inonder. Nous introduirons tout d'abord une méthode de filtrage morphologique des classes de la classification, puis nous présenterons des applications de cette méthode sur des exemples d'image thermique, médicale et satellite.

3.3.1 Le filtrage morphologique de la classification

Dans cette partie, nous présentons la méthode générale de filtrage morphologique de la classification.

La classification spectrale κ fournit des marqueurs qui sont imposés comme sources de la fonction à inonder par LPE. Les classes de la classification peuvent également servir de connaissance a priori pour imposer une cohérence spectrale dans la construction de la fonction de densité de contours pdf utilisée comme fonction à inonder probabiliste (section 3.7, p : 116).

Quelle que soit la finalité des classes, marqueurs ou connaissance a priori, il est souhaitable qu'elles se rapprochent le plus possible des contours des objets à segmenter.

La classification fournit des classes non connexes. Or la segmentation doit conduire à une partition connexe. On cherche donc à déterminer les composantes connexes pertinentes de la classification. Pour ce faire, un filtrage spatial est appliqué afin de séparer les classes les unes des autres.

Les classes de la classification κ sont notées :

$$cl(\kappa) = \{\kappa(1), \dots, \kappa(n), \dots, \kappa(N)\}. \quad (3.3.1)$$

On a bien sûr : $\kappa = \cup_n \kappa(n)$.

Les classes après filtrage spatial et un nouvel étiquetage sont notées :

$$\mathcal{C} = \{C_1, \dots, C_k, \dots, C_K\} \quad \text{avec } K > N. \quad (3.3.2)$$

Pourquoi séparer spatialement les classes $\kappa(n)$? Ces classes partitionnent complètement l'espace E puisque $\kappa = \cup_n \kappa(n)$. Si on utilise directement ces classes comme marqueurs $mark$, après un nouvel étiquetage pour avoir des classes connexes, la LPE correspondra exactement à leurs frontières. L'information apportée par la fonction à inonder ne sera donc pas prise en compte. Pour donner un certain degré de liberté aux lignes de partage des eaux, il est nécessaire de séparer spatialement ces classes les unes des autres.

Pour ce faire, on applique un filtrage morphologique sur l'image de la classification. Ce filtrage s'effectue classe par classe c'est à dire que l'on considère une image binaire $h_{\kappa(n)}$ associée à la classe $\kappa(n)$:

$$h_{\kappa(n)}(x) = \begin{cases} 1, & \text{si } x \in \kappa(n); \\ 0, & \text{sinon.} \end{cases} \quad (3.3.3)$$

On applique un filtrage morphologique Υ sur $h_{\kappa(n)}$, $\forall n = 1 \dots N$, et on obtient $h'_{\kappa(n)}$. Ce filtrage Υ recouvre deux objectifs :

- filtrer les composantes connexes associées à la classification avec un filtre anti-extensif (comme une ouverture ou une érosion), afin d'éliminer les classes les plus petites et d'introduire une nouvelle classe « vide », suivi d'un filtre extensif (comme une fermeture par reconstruction qui bouche les trous) ;
- introduire des degrés de liberté sous forme d'une nouvelle classe dénommée classe « vide ».

Cela implique qu'il y aura des composantes connexes qui vont disparaître et que toutes les autres seront plus petites.

Après le filtrage Υ un nouvel étiquetage Λ des parties connexes est effectué et l'on obtient :

$$\mathcal{C} = \Lambda[\{h'_{\kappa(n)}\}_{n=1}^N] = \{C_1, \dots, C_k, \dots, C_K\} \quad \text{avec } K > N. \quad (3.3.4)$$

L'image marque mrk est donc définie par :

$$mrk = \Lambda[\bigcup_{n=1}^N h'_{\kappa(n)}]. \quad (3.3.5)$$

La figure 3.2 présente un schéma du filtrage morphologique Υ des classes de la classification κ .

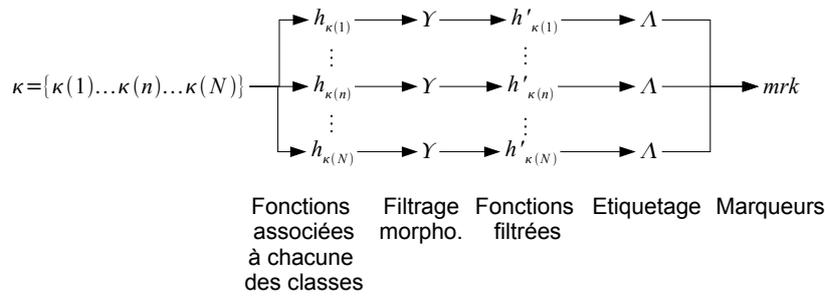


FIG. 3.2 – Schéma général de filtrage des marqueurs.

Nous allons maintenant présenter différentes applications pour lesquelles le filtrage de classe de la classification a été nécessaire pour obtenir des marqueurs

3.3.2 Exemple A : Marqueurs de l'image « bouchon en plastique »

On cherche à détecter des zones de colles sur le bouchon. Le filtrage des marqueurs a été présenté dans [134, 135]. On utilise la classification *clara* dans l'espace ACP des paramètres $\kappa_{c_{\beta_1}^P, c_{\beta_2}^P}^{clara,2} | x \in \kappa_{c_{\beta_1}^P}^{clara,3}(verte)$, notée plus simplement κ , et présentée dans la section 2.6.2 (p : 58).

On filtre les classes de la classification par ouverture morphologique avec un élément structurant hexagonal dilaté cinq fois par lui-même (i.e. de taille 5). On détecte également le fond de l'image bouchon. Pour ce faire on applique une LPE sur le gradient morphologique des facteurs pixels ACP de l'axe 1 $c_{\beta_1}^P$ avec un marqueur externe sur le contour de l'image et un marqueur ponctuel au centre du bouchon. On conserve la zone correspondant au fond que l'on érode (e.s. hexagonal de taille 12). On fait ensuite l'union des marqueurs des taches de colle provenant de la classification mrk_{κ} avec le marqueur correspondant au fond mrk_{fond} . Après étiquetage des classes connexes, on obtient l'image des marqueurs mrk qui servira pour la LPE sur l'image hyperspectrale (fig. 3.3).

Remarque 3.3.1. Il est à noter que la partie en noir de l'image des marqueurs mrk sur la figure 3.3 (d) correspond à la classe « vide » et n'est donc pas considérée comme un marqueur.

Il est important de choisir un espace adapté à la classification pour obtenir de bons marqueurs. Nous avons appliqué la classification *clara* dans plusieurs espaces (fig. 3.4) :

- l'espace image filtrée $\widehat{f}_{\lambda}, \kappa_{\widehat{f}_{\lambda}}^{clara,5}$, où l'on conserve la classe correspondant à la colle. La classification résultante n'est pas correcte comparée à la référence ;
- l'espace factoriel AFC $c_{\alpha}^f, \kappa_{c_{\alpha}^f}^{clara,3}$, où l'on conserve la classe correspondant à la colle. La classification résultante n'est pas correcte comparée à la référence ;

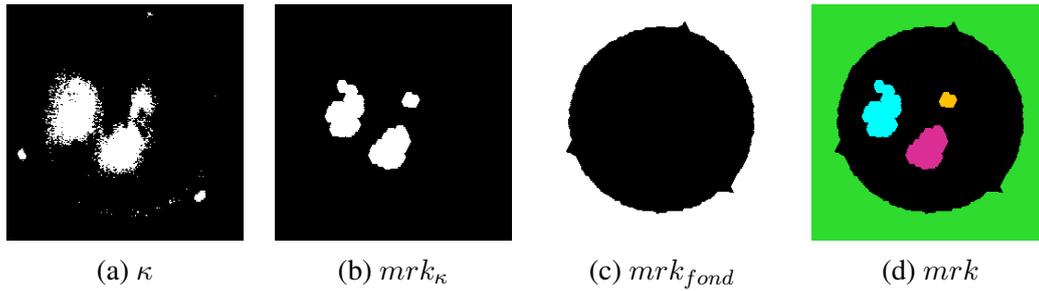


FIG. 3.3 – (a) Classification spectrale $\kappa = \kappa_{c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}} | x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)}$, (b) marqueurs de la colle obtenus par filtrage des classes la classification mrk_{κ} par ouverture morphologique avec un élément structurant hexagonal de taille 2, (c) marqueur du fond mrk_{fond} et (d) marqueurs pour la segmentation morphologique hyperspectrale (image « bouchon en plastique »).

- l'espace du paramètre d'ordonnée à l'origine b , $\kappa_b^{clara,4}$, où l'on sélectionne la classe du centre du bouchon $\kappa_b^{clara,4}(centre)$. On effectue ensuite une deuxième classification $clara$ sur les paramètres de pente a et de montée m restreints à la classe centrale précédente $\kappa_{a,m | x \in \kappa_b^{clara,4}(centre)}^{clara,2}$. La classification obtenue est assez proche de la référence, mais moins bonne que la classification dans l'espace ACP des paramètres $\kappa_{c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}} | x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)}$ (fig. 3.3 (a)). Cependant, même si la classification $\kappa_{a,m | x \in \kappa_b^{clara,4}(centre)}^{clara,2}$ est correcte, il n'en reste pas moins que la métrique euclidienne, sous-jacente à l'algorithme $clara$, n'est pas adaptée à l'espace des paramètres \mathbf{p} .

Suite à ces observations, on préférera utiliser la classification dans l'espace ACP des paramètres $\kappa_{c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}} | x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)}$, notée plus simplement κ .

3.3.3 Exemple B : Marqueurs de l'image de souris

Pour cette image on cherche à détecter la zone cancéreuse présente sur la souris. Pour ce faire, on effectue un filtrage morphologique sur la classification obtenue par approche modèle $\kappa_{\hat{f}_{\lambda}}^{mod,5}$ et celle par LDA sur l'image des paramètres $\kappa_{\mathbf{p}}^{LDA,4}$.

Pour ce faire, on effectue une érosion, avec un élément structurant (e.s.) carré de taille 5×5 pixels, suivie d'une fermeture par reconstruction, avec un e.s. de taille 3×3 pixels. Une nouvelle classe, i.e. la classe « vide », apparaît sur les images des marqueurs, par réduction de la surface des classes de la classification (fig. 3.5). La classe « vide » est étiquetée en noir sur les images marqueurs $mrk_{\kappa_{\hat{f}_{\lambda}}^{mod,5}}$ et $mrk_{\kappa_{\mathbf{p}}^{LDA,4}}$, et ne sera pas considérée comme un marqueur pour la LPE. Elle correspond donc au degré de liberté laissé aux contours de la LPE.

Les éléments structurants choisis ne doivent être de taille ni trop importante, pour ne pas faire disparaître des marqueurs d'intérêt, ni trop petite pour laisser un espace suffisamment large lorsque la LPE détermine les contours des régions.

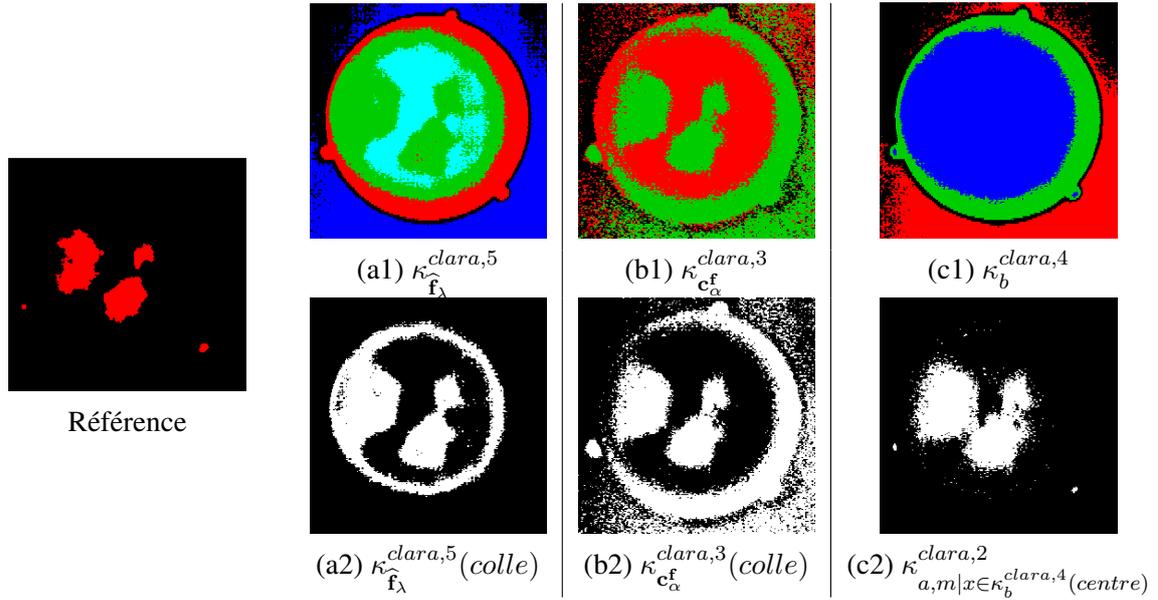


FIG. 3.4 – (a1) Classification dans l’espace image filtrée $\kappa_{\hat{f}_\lambda}^{clara,5}$ (a2) et la classe de la colle $\kappa_{\hat{f}_\lambda}^{clara,5}(colle)$. (b1) Classification dans l’espace factoriel AFC $\kappa_{c_\alpha^f}^{clara,3}$ (b2) et la classe de la colle $\kappa_{c_\alpha^f}^{clara,3}(colle)$. (c1) Classification dans l’espace du paramètre d’ordonnée à l’origine b , $\kappa_b^{clara,4}$ (c2) et deuxième classification sur les paramètres de pente a et de montée m restreints à la classe centrale précédente $\kappa_{a,m|x \in \kappa_b^{clara,4}}^{clara,2}(colle)$.

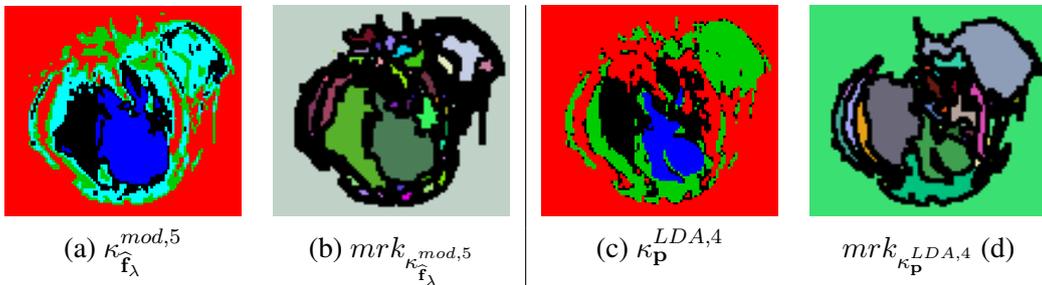


FIG. 3.5 – (a) Classification spectrale par approche modèle $\kappa_{\hat{f}_\lambda}^{mod,5}$ (b) et marqueurs associés. (c) Classification par LDA sur l’image des paramètres $\kappa_{\mathbf{p}}^{LDA,4}$ (d) et marqueurs associés. (image souris « serim447 »).

3.3.4 Exemple C : Marqueurs de l’image satellite

On utilise l’image satellite « Roujan » présentée dans la section 2.3 (p : 19). Sur la figure 3.6, on peut voir l’image de synthèse couleur créée à partir de trois canaux de l’image hyperspectrale « Roujan ».

On effectue une AFC en conservant les facteurs pixels associés aux axes 1 et 3 après seuillage du rapport signal à bruit calculé sur ces facteurs (fig. 3.7).

On effectue une classification *clara* $\kappa = \kappa_{c_\alpha^f, k=1;3}^{clara,3}$ en 3 classes dans l’espace factoriel formé des



FIG. 3.6 – Image de synthèse RVB créée à partir des canaux f_{λ_3} , f_{λ_2} et f_{λ_1} de l'image satellite « Roujan ».

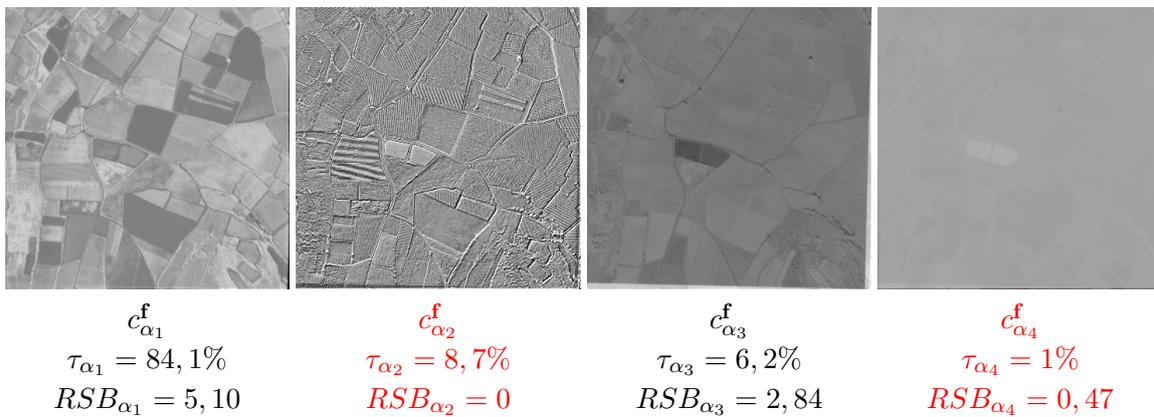


FIG. 3.7 – Facteurs pixels AFC c_{α}^f sur les 4 axes factoriels de l' image « Roujan » ainsi que leurs inerties τ_{α} et leurs RSB RSB_{α} associés. Les facteurs pixels $c_{\alpha_k}^f$ des axes non retenus sont indiqués en rouge.

facteurs pixels : $c_{\alpha_1}^f$ et $c_{\alpha_3}^f$. Pour obtenir l'image des marqueurs, mrk_{κ} , on applique un filtrage morphologique sur la classification. Ce filtrage est composé d'une érosion, avec un e.s. carré de taille 5×5 pixels, suivie d'une fermeture par reconstruction, avec un e.s. de taille 3×3 pixels (fig. 3.8). Il permet de faire apparaître la classe « vide », en noir sur l'image marqueur mrk_{κ} ; c'est à dire la classe dans laquelle passera la Ligne de Partage des Eaux.

Remarque 3.3.2. Il est à noter que l'on a utilisé les mêmes e.s. que pour le filtrage morphologique de la classification associée à l'image souris (cf. section 3.3.3). Ainsi la méthode de filtrage proposée semble être assez robuste pour différents types d'images.

Sur la figure 3.9 on peut voir d'autres classifications sur des images satellites similaires de champs « Roujan 0 2 », « Roujan 1 9 » et une image d'un port « Port de Bouc » et les marqueurs associés.

En conclusion de cette section, il est nécessaire de faire un filtrage morphologique Υ sur les classes de la classification κ pour obtenir une image de marqueurs mrk_{κ} pour la LPE. Ce filtrage assure un certain degré de liberté à la LPE qui va passer dans la classe correspondant au fond.

En outre, au lieu d'utiliser des marqueurs provenant de la classification spectrale, on pourrait choisir un nombre de régions par un critère morphologique comme le volume des bassins versants. Cependant, cette dernière méthode ne prend pas en compte l'information spectrale contenue dans la

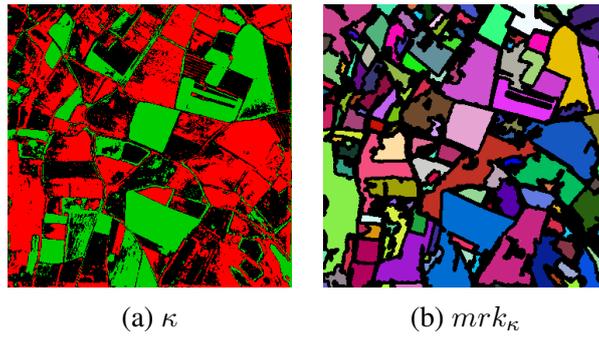


FIG. 3.8 – (a) Classification *clara* $\kappa_{c_{\alpha_k=1;3}}^{clara,3}$ et (b) mrk_{κ} marqueurs associés obtenus par filtrage morphologique sur l'image satellite « Roujan ».

classification.

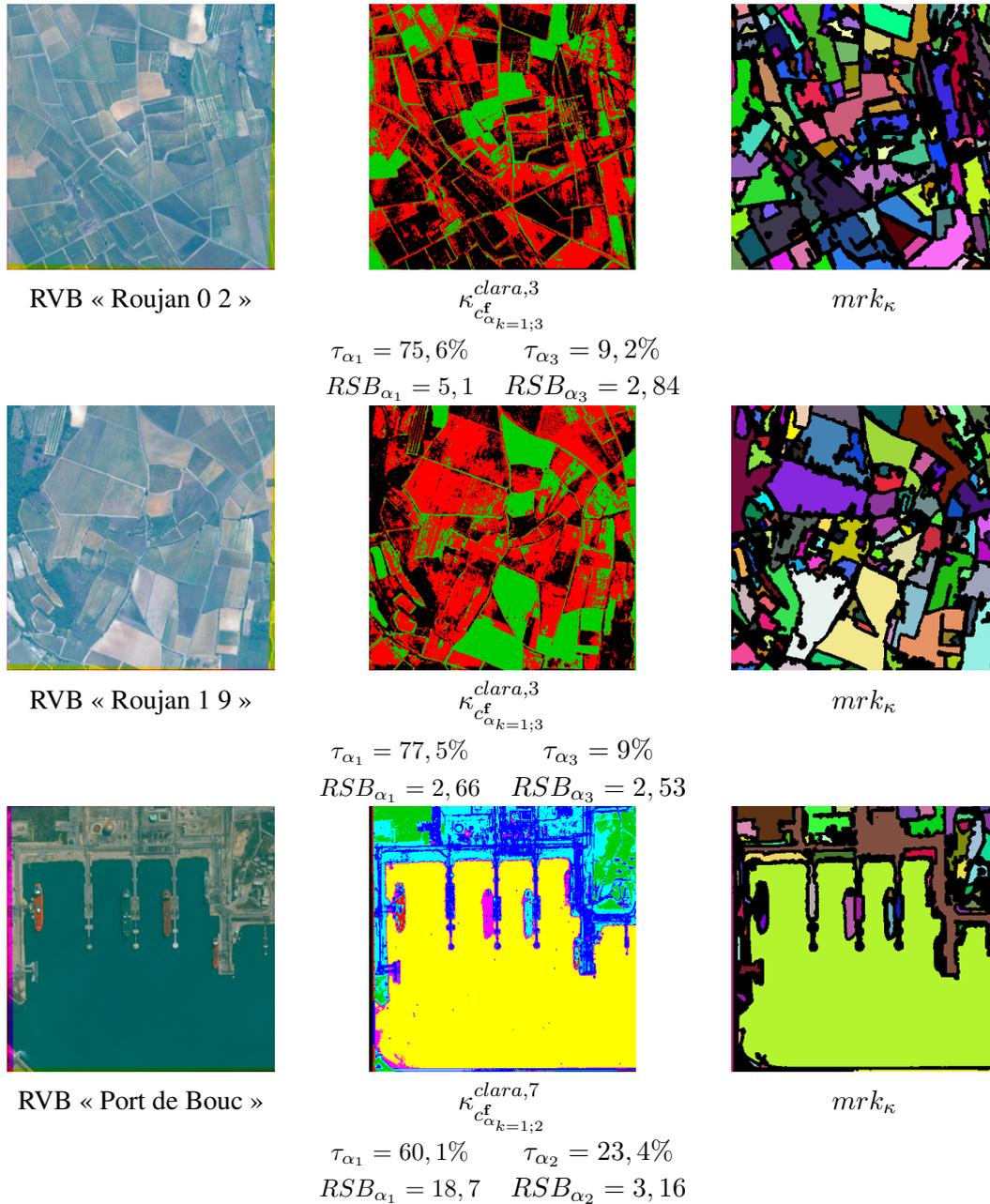


FIG. 3.9 – Image de synthèse RVB « Roujan 0 2 », « Roujan 1 9 » et « Port de Bouc » ainsi que les classification clara κ et marqueurs associés mrk_{κ} obtenus par filtrage morphologique de la classification. L'inertie τ_{α} et les rapports signaux à bruits RSB_{α} sont donnés pour les facteurs pixels utilisés lors de la classification.

3.4 Gradients hyperspectraux et segmentation

Dans cette section nous présentons une méthodologie de création de fonction à inonder à partir de gradients adaptés aux différents espaces de représentation des images multivariées. Ce cadre de traitement a été publié dans [134, 135]. Les fonctions à inonder sont ensuite utilisées dans la LPE.

Une image hyperspectrale peut être traitée :

- indépendamment sur chaque canal f_{λ_j} : dans ce cas de traitement marginal seule l'information spatiale est prise en compte ;
- indépendamment en chaque pixel vecteur $\mathbf{f}_\lambda(x_i)$: seule l'information spatiale est considérée pour ce traitement spectral ;
- en considérant \mathbf{f}_λ comme une fonction spectrale et spatiale : on effectue alors un traitement spatio-spectral.

Dans cette thèse nous avons privilégié la dernière approche qui tient compte de l'information spatio-spectrale. Nous présenterons tout d'abord les gradient multivariés, puis les distances spectrales entre pixels vecteurs et enfin nous appliquerons ces différents gradients afin de les comparer.

3.4.1 Gradients multivariés

Une image gradient ϱ , en fait la norme, est une fonction scalaire. Pour pouvoir comparer ou sommer des gradients entre eux, on normalise leurs valeurs dans l'espace réduit $[0, 1]$:

$$\varrho : E \rightarrow [0; 1] . \quad (3.4.1)$$

Pour définir un gradient morphologique sur une image hyperspectrale, quatre méthodes sont considérées :

- un gradient morphologique sur chaque canal $\varrho(f_{\lambda_j})$ (approche scalaire) ;
- un gradient supremum des gradients morphologiques sur chaque canal $\varrho_{\vee} \mathbf{f}_\lambda$ (approche marginale) ;
- un gradient somme pondérée des gradients morphologiques sur chaque canal $\varrho_{+} \mathbf{f}_\lambda$ (approche marginale) ;
- un gradient construit sur une distance $\varrho_d \mathbf{f}_\lambda$ (approche vectorielle).

On note $\varrho(f_{\lambda_j})$ le gradient morphologique symétrique, ou gradient de *Beucher*, associé au résidu d'une dilatation δ_B et d'une érosion ε_B sur un voisinage B . $B(x)$ est ce voisinage centré au point $x \in E$. Le gradient morphologique est défini comme :

$$\begin{aligned} \varrho(f_{\lambda_j}) &= \delta_B(f_{\lambda_j}) - \varepsilon_B(f_{\lambda_j}) \\ &= \vee[f_{\lambda_j}(y), y \in B(x)] - \wedge[f_{\lambda_j}(y), y \in B(x)] . \end{aligned} \quad (3.4.2)$$

Le gradient supremum des gradients morphologiques sur chaque canal est un gradient marginal défini comme :

$$\varrho_{\vee} \mathbf{f}_\lambda(x) = \vee[\varrho(f_{\lambda_j}(x)), j \in \{1, \dots, L\}] . \quad (3.4.3)$$

Il est à noter que les valeurs de chaque gradient morphologique sont normalisées dans l'intervalle $[0, 1]$ avant de prendre la somme ; afin d'avoir des gradients comparables.

Au lieu de prendre le supremum des gradients morphologiques, on peut définir un gradient somme pondérée des gradients morphologiques :

$$\varrho_{+} \mathbf{f}_\lambda(x) = \sum_{j=1}^L w_{\lambda_j} \varrho(f_{\lambda_j}(x)) \quad \text{avec} \quad \sum_j w_{\lambda_j} = 1 . \quad (3.4.4)$$

w_{λ_j} est le poids du gradient du canal f_{λ_j} . Ce gradient est un gradient marginal.

Le gradient construit sur une distance est un gradient vectoriel défini comme la différence entre le supremum et l'infimum d'une distance d entre pixels sur un voisinage B :

$$\varrho_d \mathbf{f}_\lambda(x) = \vee[d(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)), y \in B(x)] - \wedge[d(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)), y \in B(x)]. \quad (3.4.5)$$

Ce gradient est une extension aux images multivariées du gradient morphologique défini sur des images scalaires. Démontrons ce lien entre le gradient construit sur une distance ϱ_d et le gradient morphologique ϱ .

Démonstration 3.4.1 (D'après *Hanbury et Serra* [74]). Soit f une fonction scalaire, i.e. à niveaux de gris. Le gradient morphologique est défini par :

$$\varrho(f_{\lambda_j}) = \vee[f_{\lambda_j}(y), y \in B(x)] - \wedge[f_{\lambda_j}(y), y \in B(x)].$$

En utilisant la relation suivante :

$$\{f(x) - \vee[f(y), y \in B(x)]\} = \wedge\{f(x) - f(y), y \in B(x)\}$$

et la relation obtenue par inversion des suprema et des infima, on peut écrire le gradient morphologique sous une forme qui ne contient que des accroissements :

$$\varrho(f)(x) = \vee[f(x) - f(y), y \in B(x)] - \wedge[f(x) - f(y), y \in B(x)].$$

Pour passer des fonctions scalaires à des fonctions multivariées, il suffit de remplacer l'accroissement $[f(x) - f(y)]$ par une distance entre pixels vecteurs $d(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y))$ pour obtenir :

$$\varrho_d \mathbf{f}_\lambda(x) = \vee[d(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)), y \in B(x)] - \wedge[d(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)), y \in B(x)].$$

Cette dernière équation est celle du gradient construit sur une distance, ϱ_d .

Comme son nom l'indique, le gradient construit sur une distance nécessite une distance entre pixels vecteurs.

3.4.2 Distances entre spectres

Nous allons maintenant introduire quelques distances ou métriques entre spectres, ou pixels vecteurs. Nous avons utilisé dans nos travaux trois distances :

- la distance euclidienne d_E ;
- la distance inverse des variances d_{1/σ^2} ;
- la distance du chi-deux d_{χ^2} .

La distance euclidienne est définie par :

$$d_E(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)) = \sqrt{\sum_{j=1}^L (f_{\lambda_j}(x) - f_{\lambda_j}(y))^2}. \quad (3.4.6)$$

La distance inverse des variances d_{1/σ^2} vaut :

$$d_{1/\sigma^2}(\mathbf{f}_\lambda(x), \mathbf{f}_\lambda(y)) = \sqrt{\sum_{j=1}^L \left(\frac{f_{\lambda_j}(x) - f_{\lambda_j}(y)}{\sigma_{\lambda_j}} \right)^2}, \quad (3.4.7)$$

avec $\sigma_j^2 = E[(f_{\lambda_j} - m_j)^2]$ la variance du canal f_{λ_j} et $m_j = E[f_{\lambda_j}]$ sa moyenne.

Remarque 3.4.1. La distance Mahalanobis s'exprime sous la forme :

$$d_M(f(x), f(y)) = \sqrt{(f(x) - f(y))^t \Sigma^{-1} (f(x) - f(y))} \quad (3.4.8)$$

où Σ est la matrice de covariance entre les variables. La distance inverse des variances correspond à la distance de Mahalanobis si les variables (canaux) sont décorrélées. En effet la matrice de covariance est alors diagonale et les valeurs de sa diagonale sont égales à la variance des canaux : σ_j^2 avec $j = 1, \dots, L$. Cependant, on a fait l'hypothèse que les canaux sont corrélés pour une image hyperspectrale (dans l'espace image de départ).

La distance du chi-deux se définit comme :

$$d_{\chi^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = \sqrt{\sum_{j=1}^L \frac{N}{f_{\cdot \lambda_j}} \left(\frac{f_{\lambda_j}(x_i)}{f_{x_i \cdot}} - \frac{f_{\lambda_j}(x_{i'})}{f_{x_{i'} \cdot}} \right)^2} \quad (3.4.9)$$

avec $f_{\cdot \lambda_j} = \sum_{i=1}^P f_{\lambda_j}(x_i)$, $f_{x_i \cdot} = \sum_{j=1}^J f_{\lambda_j}(x_i)$ et $N = \sum_{j=1}^L \sum_{i=1}^P f_{\lambda_j}(x_i)$.

Le lien entre l'expression de la distance du chi-deux que nous venons de donner et son expression dans l'équation 2.2.24 (p : 14) est expliqué dans l'annexe A.4.7.1 (p : 217) :

$$d_{\chi^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = d_E(c_\alpha^{\mathbf{f}, AFC}(x_i), c_\alpha^{\mathbf{f}, AFC}(x_{i'})) . \quad (3.4.10)$$

Pour une image hyperspectrale dont les valeurs sont positives, on utilise une AFC plutôt qu'une ACP parce que les canaux peuvent être considérés comme des distributions de probabilités. La métrique du chi-deux, utilisée en AFC, est adaptée aux lois de probabilité. En outre l'AFC présente l'avantage de pouvoir représenter simultanément les facteurs associés aux pixels et les facteurs associés aux variables dans un même espace.

La métrique du chi-deux, entre spectres, dans l'espace image est équivalente à la métrique euclidienne dans l'espace factoriel si tous les axes factoriels sont conservés. Si seule une partie des axes factoriels est conservée alors la métrique euclidienne dans l'espace factoriel est une bonne approximation de la métrique du chi-deux entre spectres dans l'espace image [21].

Cette propriété est démontrée dans l'annexe A.4.7.2 (p : 217).

De même, dans le cas de l'ACP, la métrique inverse des variances entre spectres de l'image est égale à la métrique euclidienne entre facteurs pixels dans l'espace factoriel ACP. La distance inverse des variances sera donc employée pour des images dont les valeurs peuvent être négatives.

$$d_{1/\sigma^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = d_E(c_\alpha^{\mathbf{f}, ACP}(x_i), c_\alpha^{\mathbf{f}, ACP}(x_{i'})) . \quad (3.4.11)$$

Cette propriété est démontrée dans l'annexe A.3.9 (p : 203).

Il existe d'autres distances utilisées en hyperspectral. Rappelons qu'une distance est définie de la manière suivante.

Définition 2. Soit \mathcal{T} un ensemble. Une distance $d(u, v)$ est une application $\mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}^+$ qui vérifie les propriétés suivantes :

- la positivité $\forall u, v \in \mathcal{T}, \quad d(u, v) \geq 0$
- la séparation ou identité $\forall u, v \in \mathcal{T}, \quad u = v \Leftrightarrow d(u, v) = 0$
- la symétrie $\forall u, v \in \mathcal{T}, \quad d(u, v) = d(v, u)$
- l'inégalité triangulaire $\forall u, v, w \in \mathcal{T}, \quad d(u, v) \leq d(u, w) + d(w, v)$

Le coefficient de corrélation indique la ressemblance entre deux vecteurs :

$$\text{cor}(f_\lambda(x), f_\lambda(y)) = \frac{E[(f_\lambda(x) - m_{f_\lambda(x)})(f_\lambda(y) - m_{f_\lambda(y)})]}{\sigma_{f_\lambda(x)}\sigma_{f_\lambda(y)}}$$

avec $m_{f_\lambda(x)}$ la moyenne du pixel vecteur $f_\lambda(x)$ et $\sigma_{f_\lambda(x)}$ son écart type.

Un coefficient de corrélation de valeur 0 indique une décorrélation, +1 une corrélation parfaite et -1 une anticorrélation. Le coefficient de corrélation n'est donc pas une distance au sens mathématique puisque qu'il ne vérifie pas les propriétés de positivité et d'inégalité triangulaire. Par contre le « Spectral Angle Mapper » (MAP) qui correspond à l'angle associé au coefficient de corrélation est une distance entre spectres :

$$d_{SAM}(f_\lambda(x), f_\lambda(y)) = \arccos(\text{cor}(f_\lambda(x), f_\lambda(y))) .$$

Comme nous n'avons pas utilisé cette méthode, nous invitons le lecteur à se référer à la thèse de F. Schmidt [158] pour une démonstration des propriétés de la distance.

De nombreux articles traitent des distances utilisées en imagerie hyperspectrale, e.g. [92, 185, 53].

3.4.3 Choix d'une métrique adaptée à l'espace de représentation de l'image

Il est nécessaire de choisir une distance adaptée à l'espace de représentation. Nous avons effectué les choix suivants :

- dans l'espace de l'image filtrée $\hat{\mathbf{f}}_\lambda$: la distance du chi-deux d_{χ^2} est utilisée car les canaux peuvent être considérés comme des distributions de probabilités. En outre c'est la distance associée à l'AFC ;
- dans l'espace factoriel AFC \mathbf{c}_α^f : on utilise la distance euclidienne d_E qui est équivalente à la distance du chi-deux calculée dans l'espace image ;
- dans l'espace formé par les cartes de paramètres \mathbf{p} : on utilise la distance inverse des variances qui permet de traiter les valeurs négatives des paramètres. Cette distance est celle associée à l'ACP ;
- dans l'espace factoriel ACP des paramètres \mathbf{c}_β^p : la distance euclidienne est employée car elle est équivalente à la distance inverse des variances calculée dans l'espace image.

Il serait bien sûr possible d'utiliser la distance inverse des variances dans l'espace image $\hat{\mathbf{f}}_\lambda$. Cependant comme nous avons choisi d'appliquer l'AFC sur l'espace de l'image $\hat{\mathbf{f}}_\lambda$ dont les valeurs sont positives, nous avons employé la distance du chi-deux.

3.4.4 Exemples comparatifs : Application à la segmentation en imagerie thermique

Après avoir défini les gradient multivariés et les distances entre spectres, intéressons nous maintenant à leur application sur l'image thermique « bouchon en plastique ».

Nous présenterons tout d'abord pourquoi la segmentation sur un canal, même bien choisi, ne peut pas fonctionner. Puis nous verrons divers gradients et segmentations dans l'espace de l'image, l'espace factoriel AFC de l'image c_α^f , l'espace des paramètres p et l'espace factoriel ACP des paramètres c_β^p .

3.4.4.1 Segmentation sur un canal

La segmentation hyperspectrale est nécessaire. En effet, une méthode trop directe pourrait consister à choisir un canal bien contrasté et à appliquer une méthode de segmentation sur ce canal. Nous avons testé cette méthode avec la LPE.

Le canal $f_{\lambda_{60}}$ qui est un des mieux contrasté a été retenu. Afin de réduire le nombre de minima du gradient un filtrage morphologique a été effectué. Ce filtrage se compose :

- d'un filtre gaussien de taille 11×11 pixels destiné à améliorer le contraste entre la colle et le bouchon ;
- et enfin d'un nivellement morphologique avec un e.s. hexagonal.

Son gradient morphologique a été calculé. La LPE est ensuite calculée à partir des minima du gradient et d'un marqueur sur les bords de l'image $LPE_{\varrho(f_{\lambda_{60}})}$. Cette méthode conduit à une sur-segmentation (fig. 3.10), qui ne permet pas d'extraire les taches de colles.

Il est donc nécessaire de faire un traitement vectoriel des données et non un traitement marginal, i.e. exclusivement sur un canal.

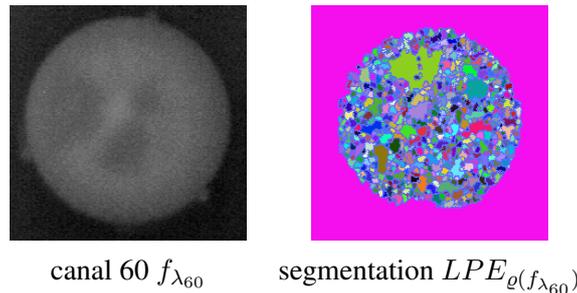


FIG. 3.10 – Segmentation par LPE sur le gradient morphologique du canal $f_{\lambda_{60}}$. Un marqueur sur les bords de l'image a été ajouté.

Comparons maintenant différents résultats de segmentation obtenus dans les espaces différents.

3.4.4.2 Segmentation dans l'espace image filtrée

Pour la segmentation dans les différents espaces et notamment l'espace image filtrée \hat{f}_λ , nous avons procédé comme expliqué à la section 3.2.2 (p : 81) :

- classification κ dans l'espace considéré, filtrage morphologique Υ de cette classification et ajout d'un marqueur correspondant au contour du bouchon suivi d'un étiquetage pour obtenir les marqueurs mrk ;
- calcul d'un ou plusieurs gradients associés à l'image ;
- calcul de la LPE sur le gradient à partir des marqueurs mrk .

L'ajout du marqueur correspondant au contour du bouchon se fait comme expliqué à la section 3.3.2 (p : 84) si ce n'est qu'au lieu d'utiliser un gradient morphologique on utilise le gradient considéré

(marginal ou vectoriel).

Dans l'espace image filtrée \widehat{f}_λ , nous avons calculé un gradient construit sur la distance du chi-deux ϱ_{χ^2} . La classe correspondant à la colle extraite par classification $\kappa_{\widehat{f}_\lambda}^{clara,5}(colle)$ est ensuite filtrée par une ouverture morphologique de taille 5 puis étiquetée pour donner l'image marqueur mrk (fig. 3.11). Comme la classification *clara* n'est pas adaptée à l'espace image (car la métrique utilisée est euclidienne) les marqueurs ne sont pas bons. C'est ce qui explique les mauvais résultats de la segmentation $LPE^{\widehat{f}_\lambda}(\varrho_{\chi^2}, mrk)$ dans cet espace.

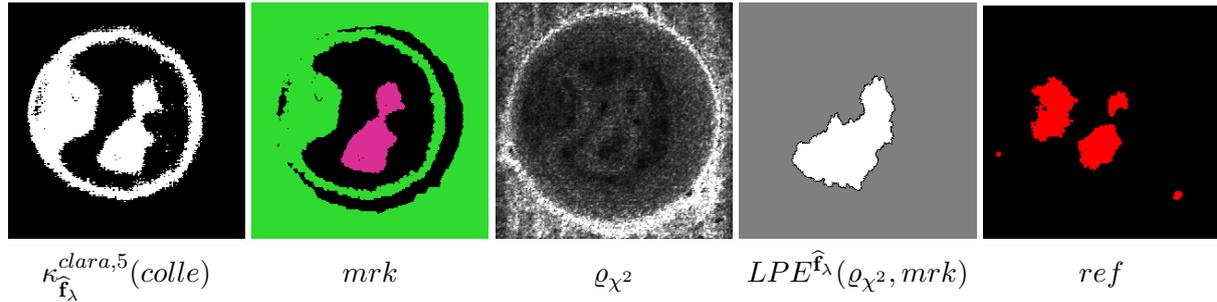


FIG. 3.11 – Classe « colle » de la classification $\kappa_{\widehat{f}_\lambda}^{clara,5}(colle)$, marqueurs mrk , gradient construit sur la distance du chi-deux ϱ_{χ^2} , segmentation $LPE^{\widehat{f}_\lambda}(\varrho_{\chi^2}, mrk)$ dans l'espace image filtrée \widehat{f}_λ et référence ref . L'image du gradient a été multipliée par un facteur pour la visualisation.

3.4.4.3 Segmentation dans l'espace factoriel AFC de l'image

Dans l'espace factoriel AFC de l'image \mathbf{c}_α^f , la classe correspondant à la colle extraite par classification $\kappa_{\mathbf{c}_\alpha^f}^{clara,3}(colle)$ est ensuite filtrée par une ouverture morphologique de taille 5 puis étiquetée pour donner l'image marqueur mrk . Un gradient construit sur la métrique euclidienne $\varrho_E \mathbf{c}_\alpha^f$ a ensuite été calculée avant de le segmenter à l'aide des marqueurs mrk (fig. 3.12).

On constate que la segmentation résultante se rapproche de la référence. Ceci est dû aux marqueurs et au gradient qui sont de meilleure qualité dans cet espace.

On peut également vérifier de visu que le gradient euclidien dans l'espace factoriel $\varrho_E \mathbf{c}_\alpha^f$ est très proche du gradient du chi-deux dans l'espace image $\varrho_{\chi^2} \widehat{f}_\lambda$ (fig. 3.11).

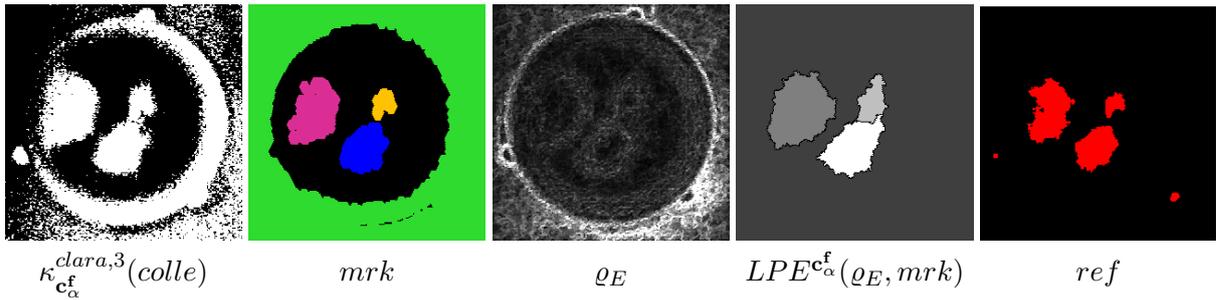


FIG. 3.12 – Classe « colle » de la classification $\kappa_{c_\alpha}^{clara,3}(colle)$, marqueurs mrk , gradient construit sur la distance euclidienne ρ_E , segmentation $LPE_{c_\alpha}^f(\rho_E, mrk)$ dans l'espace factoriel AFC de l'image c_α^f et référence ref . L'image du gradient a été multipliée par un facteur pour la visualisation.

3.4.4.4 Segmentation dans l'espace des paramètres

Dans l'espace des paramètres \mathbf{p} , on applique une ouverture morphologique de taille 5 (e.s. hexagonal) sur la classe « colle » de la classification $\kappa_{a,m|x \in \kappa_b^{clara,4}(centre)}^{clara,2}(colle)$, notée plus simplement $\kappa_{a,m|x}^{clara,2}(colle)$.

Deux gradients sont ensuite construits :

- un gradient construit sur la distance inverse des variances ρ_{1/σ^2} (gradient vectoriel) ;
- un gradient supremum des gradients morphologiques sur chaque canal ρ_V (gradient marginal).

La LPE est ensuite appliquée sur les divers gradient et les marqueurs mrk (fig. 3.13). Les deux segmentations sont bonnes par rapport à la référence.

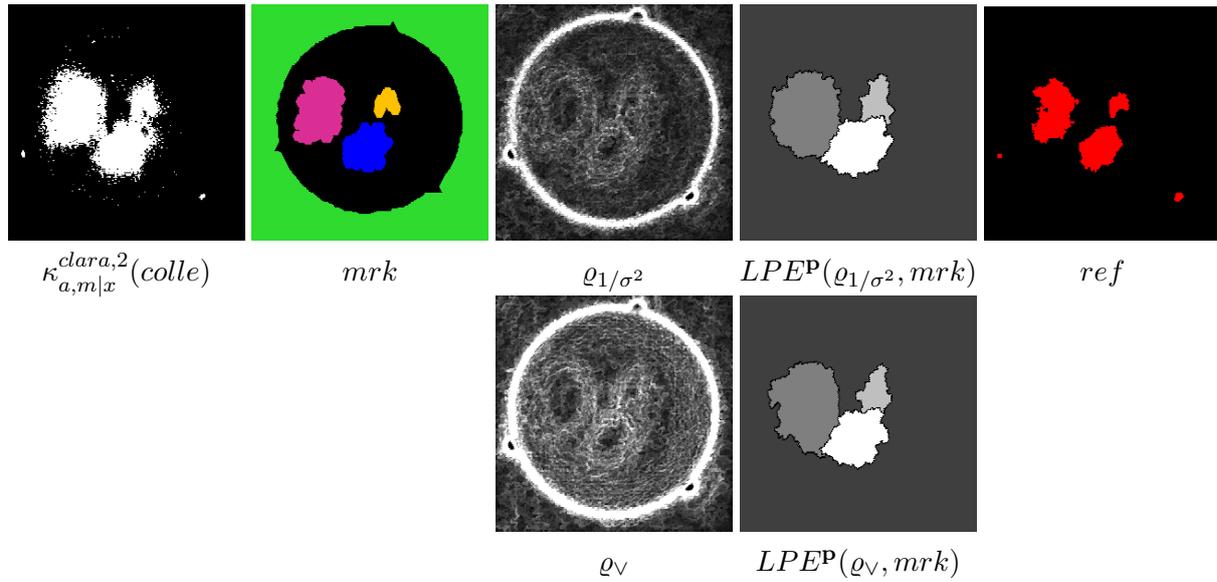


FIG. 3.13 – Classe « colle » de la classification $\kappa_{a,m|x}^{clara,2}(colle)$, marqueurs mrk , gradient construit sur la distance inverse des variances ρ_{1/σ^2} et gradient supremum ρ_V , segmentations associées $LPEP(\rho_{1/\sigma^2}, mrk)$ et $LPEP(\rho_V, mrk)$ dans l'espace des paramètres \mathbf{p} et référence ref . Les images des gradients ont été multipliées par un facteur pour la visualisation.

3.4.4.5 Segmentation dans l'espace factoriel ACP des paramètres

Dans l'espace factoriel ACP des paramètres formé des deux premiers axes factoriels $\mathbf{c}_\beta^{\mathbf{p}} = (c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}})$, on applique une ouverture morphologique de taille 5 (e.s. hexagonal) sur la classe correspondant à la colle de la classification $\kappa_{c_{\beta_1}^{\mathbf{p}}, c_{\beta_2}^{\mathbf{p}} | x \in \kappa_{c_{\beta_1}^{\mathbf{p}}}^{clara,3}(verte)}(colle)$, notée plus simplement $\kappa_{c_{\beta_{k=1;2}}^{\mathbf{p}} | x}^{clara,2}(colle)$.

Plusieurs gradients ont été essayés :

- un gradient construit sur la distance euclidienne ρ_E ;
- un gradient supremum des gradients morphologiques sur chaque canal ρ_V ;
- un gradient somme pondérée des gradients morphologiques sur chaque canal ρ_+ . Pour ce dernier gradient les poids correspondent aux inerties des axes factoriels $w_1 = 0,97$, $w_2 = 0,021$ et $w_3 = 0,0066$.

La LPE est ensuite appliquée sur les divers gradients et les marqueurs mrk (fig. 3.14). Les segmentations sur le gradient euclidien ρ_E et le gradient supremum ρ_V sont bonnes comparées à la référence. En revanche la segmentation sur le gradient somme pondérée ρ_+ est moins bonne à cause de fuites présentes sur le gradient au niveau des occlusions de colles.

On peut également remarquer la ressemblance entre le gradient construit sur la distance euclidienne dans l'espace factoriel ACP des paramètres $\rho_E \mathbf{c}_\beta^{\mathbf{p}}$ et le gradient construit sur la distance inverse des variances $\rho_{1/\sigma^2} \mathbf{p}$ dans l'espace des paramètres.

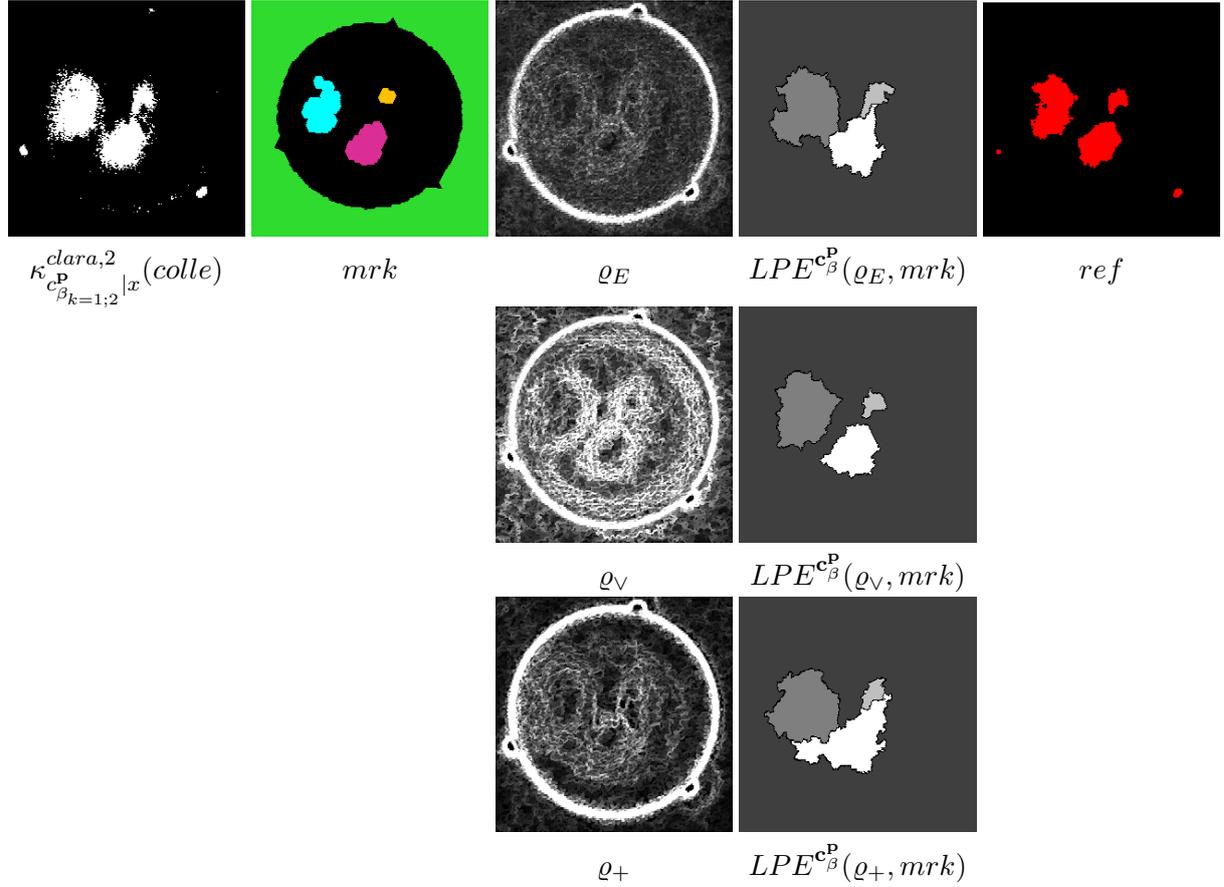


FIG. 3.14 – Classe « colle » de la classification $\kappa_{c_{\beta k=1;2}^p}^{clara,2}(colle)$, marqueurs mrk , gradient construit sur la distance euclidienne ϱ_E , gradient supremum ϱ_V et gradient somme pondérée ϱ_+ (de poids $w_1 = 0,97$, $w_2 = 0,021$ et $w_3 = 0,0066$), segmentations associées $LPE^{cb}(\varrho_E, mrk)$, $LPE^{c^p_\beta}(\varrho_V, mrk)$ et $LPE^{c^p_\beta}(\varrho_+, mrk)$ dans l'espace ACP des paramètres c^p_β et référence ref . Les images des gradients ont été multipliées par un facteur pour la visualisation.

3.4.4.6 Comparaison des différentes segmentations

Afin d'établir une comparaison entre les différentes segmentations obtenues par LPE sur les gradients multivariés à l'aide de marqueurs provenant de la classification, l'ensemble des résultats de segmentation sont présentés sur la figure 3.15. On constate que les meilleures segmentations sont obtenues dans l'espace des paramètres \mathbf{p} ou dans l'espace ACP des paramètres c^p_β . En effet le modèle ajusté a tendance à réduire l'entropie de l'image. Les résultats sont donc plus robustes. Dans l'espace AFC c^f_α on obtient également une bonne segmentation. Par contre dans l'espace image filtrée \hat{f}_λ la segmentation est éloignée de la référence à cause des marqueurs qui ne sont pas corrects dans cet espace. En fait l'espace image n'est pas adapté à la métrique euclidienne sous-jacente à la classification. En outre le grand nombre de canaux dans cet espace augmente considérablement les temps de calculs et pose des problèmes de séparabilité des classes à cause du phénomène de Hughes. Par conséquent il est nécessaire d'effectuer la classification et la segmentation dans un espace de dimension réduite.

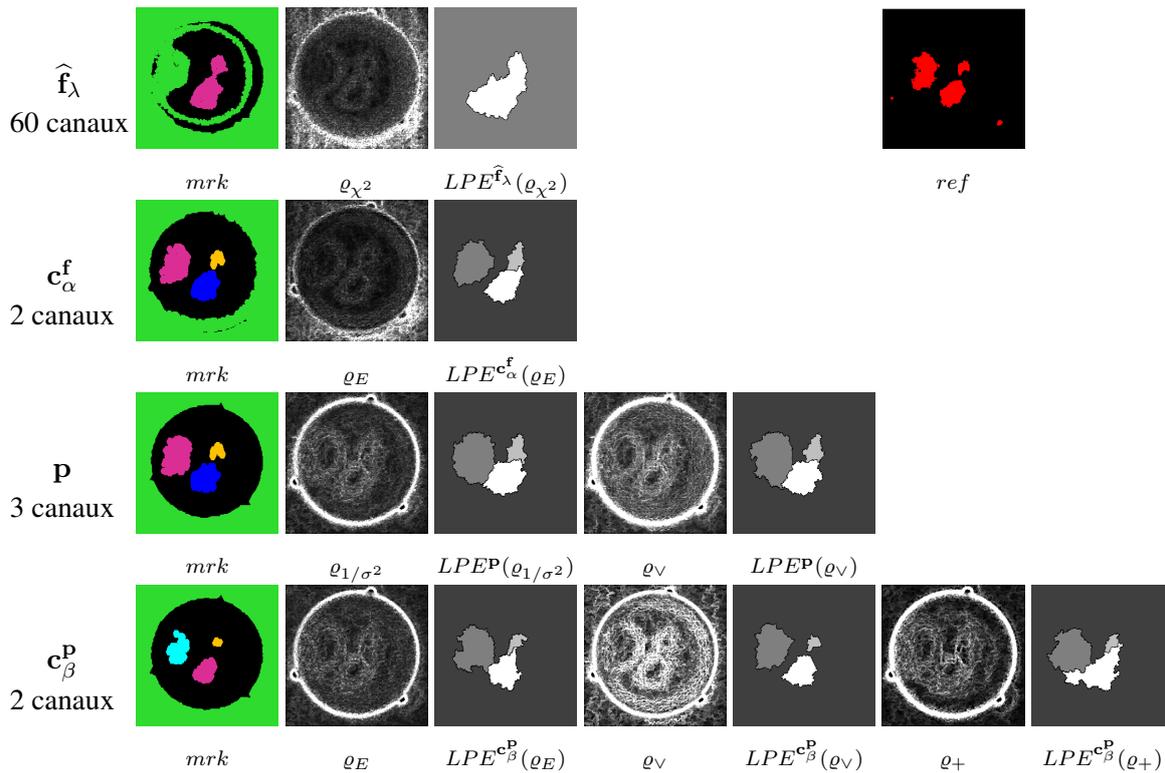


FIG. 3.15 – Résultats de segmentation par LPE sur les gradients multivariés avec des marqueurs provenant de la classification (image « bouchon en plastique »).

Dans cette section, nous avons présenté des méthodes de segmentation des images hyperspectrales par LPE. Des marqueurs sont obtenus par classification spectrale et un gradient vectoriel ou marginal permet d'introduire la dimension spatiale. Cette approche est complètement déterministe.

Une des limites de la LPE réside dans le fait qu'il est important de trouver des marqueurs pertinents. En outre une fuite sur le gradient, peut créer une segmentation aberrante. Afin de rendre plus robustes les contours du gradient, on va utiliser une approche probabiliste qui consiste à tirer des marqueurs aléatoires pour obtenir une fonction de densité de probabilités de contours qui est ensuite segmentée. Nous allons détailler cette approche dans la section suivante.

3.5 Introduction de la LPE probabiliste

La segmentation par LPE probabiliste a été créée par J. Angulo et D. Jeulin pour des images à niveaux de gris et des images couleurs [4]. Un des apports majeurs de cette thèse est son extension aux images multivariées ou hyperspectrales [137, 139, 138]. Plus récemment *Debeir et al.* ont commencé à développer une version plus basique appelée « Weak Watershed » [48].

Un des problèmes de la LPE est que les petites régions, ou les régions faiblement contrastées, dépendent fortement de la position des marqueurs ou du critère morphologique associé au minima. Un critère fréquemment utilisé est le volume qui est calculé comme l'intégrale des niveaux de gris du bassin versant. En fait il existe deux types de contours associés à la LPE d'un gradient :

- les contours du premier ordre qui correspondent aux régions significatives et qui sont relativement indépendantes des marqueurs ;
- les contours du deuxième ordre associés aux régions de petite taille, faiblement contrastée ou texturée et qui dépendent fortement de la position des marqueurs.

Le but de la LPE probabiliste est de rehausser les régions du premier ordre et de les rendre plus robustes faces aux effets d'échantillonnage dus aux marqueurs.

Après avoir présenté la LPE probabiliste pour les images couleurs [4], nous montrons comment tirer des marqueurs aléatoires pour des images hyperspectrales. A partir de ces marqueurs des fonctions de densité de probabilité (pdf) de contours sont construites puis segmentées. Ces différentes étapes sont illustrées par plusieurs applications dans les sections suivantes.

Nous présentons dans cette partie la LPE probabiliste telle que proposée par *Angulo et Jeulin* [4].

Nous expliquons tout d'abord les raisons de l'introduction de la LPE probabiliste, puis nous présentons quelques notions nécessaires à sa mise en pratique. Nous verrons ensuite la segmentation de germes uniformément aléatoires, l'influence réelle des paramètres, la segmentation à partir de germes aléatoires régionalisés. Enfin quelques considérations sur l'implémentation informatique seront apportées.

3.5.1 Motivations de la création de la LPE probabiliste

Il existe principalement deux approches hiérarchiques lorsque des marqueurs ne peuvent être définis :

1. l'algorithme des cascades (cf. annexe B.5.4, p : 252) ;
2. les hiérarchies de partition construites à partir des valeurs d'extinction qui permettent de sélectionner les minima utilisés dans la LPE selon des critères morphologiques comme la dynamique, la surface ou le volume des bassins versants.

La segmentation par critère volumique est très utilisée dans de nombreuses applications puisque ce critère qui combine la dynamique et la surface des bassins versants sélectionne les régions les plus significatives d'un point de vue perceptuel. Cependant les performances de cette approche s'amenuisent lorsque l'image est segmentée en très peu de régions ; ce qui est justement le but de plusieurs applications comme l'indexation d'image à partir de la segmentation en régions significatives.

La figure 3.16 montre quatre images couleurs segmentées par LPE en $R = 10, 20$ et 50 régions. Nous pouvons observer que les régions les plus importantes ne sont pas bien déterminées (même lorsque $R = 50$). La solution habituelle consiste à filtrer l'image afin de simplifier les détails et de rehausser les principales régions, avec par exemple des filtres morphologiques comme les nivellements [124].

Le problème se pose pour la détermination du critère volumique, calculé pour chacun des minima de la fonction à inonder. Ce critère dépend de l'information locale sur l'image, alors que le processus d'inondation est une compétition entre les différents minima pour déterminer la partition optimale. En fait, c'est la recherche du plus court chemin lorsque le « coût » du chemin est le maximum des poids des arcs du graphe du chemin [123].

Le but de la LPE probabiliste est de détecter les contours qui sont les plus robustes par rapport aux variations des conditions de segmentation. Plus précisément, nous allons explorer une approche probabiliste à partir de marqueurs aléatoires utilisés pour construire une fonction de densité de probabilité de contours *pdf* qui est ensuite segmentée, par exemple par LPE volumique [4], afin de définir les régions les plus significatives. Il est important de noter que le but de la version originale de la LPE probabiliste était la segmentation non supervisée des images naturelles en très peu de régions.

Auparavant la segmentation probabiliste avait déjà été étudiée dans la littérature, par exemple en utilisant les probabilités de cooccurrence sur les graphes [40], les méthodes bayésiennes [189], les champs de Markov aléatoires [60] (combinés avec une segmentation par LPE [66]), les méthodes « Markov Chain Monte Carlo » (MCMC) [181].

La LPE probabiliste est la première étude de segmentation probabiliste à partir de marqueurs aléatoires pour la transformation par LPE. Les travaux antérieurs les plus proches de cette méthode sont ceux de *Serra* [168], où la somme de LPE provenant de séries d'images polarimétriques est utilisée pour définir une distribution de contours.

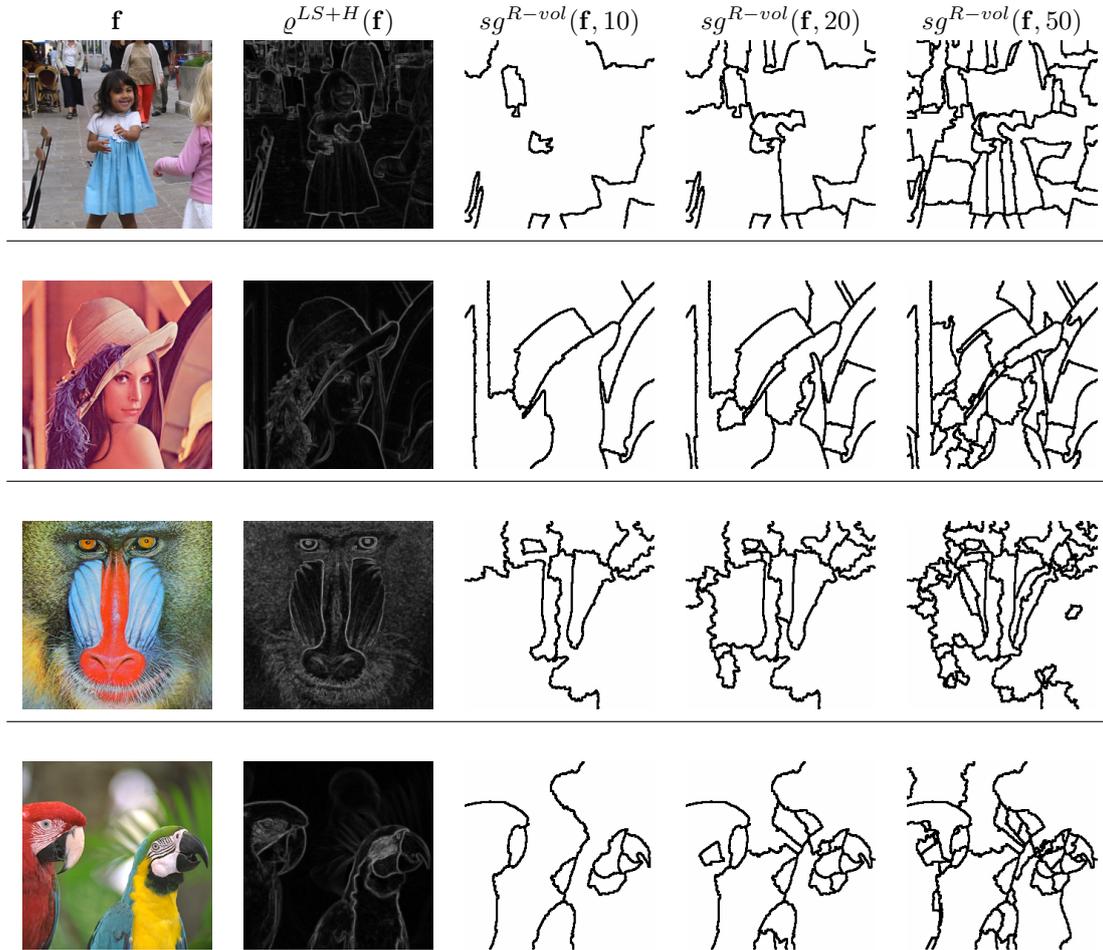


FIG. 3.16 – Exemples d’images couleurs segmentées par LPE volumique en R régions. Première colonne : images originales \mathbf{f} , deuxième colonne : gradients couleurs $\varrho^{LS+H}(\mathbf{f})$, troisième colonne : segmentation volumique $sg^{R-vol}(\varrho^{LS+H}(\mathbf{f}), 10)$, quatrième colonne : $sg^{R-vol}(\varrho^{LS+H}(\mathbf{f}), 20)$ et dernière colonne : $sg^{R-vol}(\varrho^{LS+H}(\mathbf{f}), 50)$ (Source *J. Angulo* [4]).

3.5.2 Quelques notions introductives

Introduisons tout d’abord quelques notions nécessaires à la suite de la présentation de la LPE probabiliste.

3.5.2.1 Le gradient couleur

La fonction à inonder utilisée pour la transformation par LPE est en général une image de gradient. Pour des images couleurs, selon les travaux d’*Angulo et Serra* [5], il est particulièrement intéressant d’utiliser en gradient couleur dans l’espace de représentation couleur LSH : Luminance, Saturation et Hue (teinte). En effet, il présente de meilleures performances que les autres gradients couleurs. Cependant, n’importe quel autre gradient couleur pourrait être utilisé, y compris un gradient marginal en RVB. Soit $\mathbf{f}(x) = (f_L(x), f_S(x), f_H(x))$ une image couleur en représentation LSH. Le gradient

couleur est défini par :

$$\varrho^{LS+H}(\mathbf{f}(x)) = f_S(x) \times \varrho^\circ(f_H(x)) + (1 - f_S(x)) \times \varrho(f_L(x)) + \varrho(f_S(x)) \quad (3.5.1)$$

où $\varrho(g(x))$ est le gradient morphologique de la fonction scalaire $g(x)$ et $\varrho^\circ(a(x))$ est le gradient circulaire centré de la fonction angulaire $a(x)$.

3.5.2.2 Notation pour la segmentation par LPE

On note $mrk(x)$, l'image des marqueurs. L'image binaire des contours issus de la segmentation associés à ces marqueurs, selon le gradient $\varrho^{LS+H}(\mathbf{f}(x))$, est, quant à elle, notée $sg^{mrk}(\varrho^{LS+H}(\mathbf{f}(x)))$. En utilisant le même gradient, la segmentation par LPE volumique est dénommée $sg^{R-vol}(\varrho^{LS+H}(\mathbf{f}(x)))$.

3.5.2.3 Génération des germes aléatoires

Le paradigme de la segmentation par LPE repose sur le choix approprié des marqueurs, qui sont les sources pour générer les bassins d'attractions [29, 26]. Il est notoire que la partie la plus « intelligente » de cette technique est l'utilisation d'un critère utilisé pour sélectionner les marqueurs. Dans la présente approche, nous suivons la direction opposée en sélectionnant des germes aléatoires comme marqueurs. Ce choix arbitraire sera compensé par l'utilisation d'un nombre donnée M de réalisations, afin de filtrer les fluctuations insignifiantes.

Un moyen naturel d'introduire des germes aléatoires [84], consiste à générer des réalisations d'un processus de points de Poisson avec une intensité constante θ (à savoir le nombre de points moyens par unité de surface). Il est bien connu que le nombre de points aléatoires N tombant dans un domaine D d'aire $|D|$ suit une distribution de Poisson de paramètre $\theta|D|$. En outre, conditionnellement au fait que $N = n$, les n points sont distribués de manière indépendante et uniforme sur D . Dans ce qui suit, nous fixerons la valeur N du nombre de germes aléatoires (au lieu d'utiliser un nombre aléatoire comme pour le processus de points de Poisson), et nous générerons des réalisations indépendantes des lieux des germes dans D . Dans certains cas, comme illustré ci-après, il peut être intéressant de générer une distribution non uniforme de germes, avec une intensité régionale (ou mesure) $\theta(x)$. Dans les cas de Poisson, N suit une distribution de Poisson de paramètre $\theta(D)$, et conditionnellement au fait que $N = n$, les n points sont distribués indépendamment sur D avec la densité de probabilité $\theta(x)/\theta(D)$. Dans ce qui suit, l'intensité $\theta(x)$ sera générée à partir de l'image, et nous utiliserons un nombre de germes fixe, comme pour le cas homogène.

3.5.2.4 La méthode de Parzen pour calculer une fonction de densité de probabilité

L'estimation de densité par une méthode à noyaux, ou méthode de la fenêtre de Parzen [54], est une manière d'estimer la fonction de densité de probabilité (pdf) d'une variable aléatoire. Soient $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M \in \mathbb{R}^n$ M réalisations d'une variable aléatoire, l'approximation par noyau de la densité de sa pdf est :

$$\widehat{pdf}_h(x) = \frac{1}{Nh} \sum_{i=1}^M K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right), \quad (3.5.2)$$

où $K(\mathbf{x})$ est un noyau et sa bande passante h un paramètre de lissage. Habituellement, on choisit pour $K(x)$ une fonction gaussienne de moyenne 0 et de variance σ^2 , qui détermine l'effet de lissage.

3.5.3 Segmentation de germes uniformément aléatoires

Soit $\{mrk_i(x)\}_{i=1}^M$, une série de M réalisations de N marqueurs uniformément aléatoires. Chacune de ces images binaires de points est considérée comme les marqueurs d'une segmentation par LPE du gradient couleur $sg^{mrk}(\rho^{LS+H}(x))$ et par conséquent, une série de segmentation est obtenue, i.e. $\{sg_i^{mrk}(x)\}_{i=1}^M$ (fig. 3.17). Notons, que le nombre de points détermine le nombre de régions obtenu, i.e. une propriété essentielle de la LPE. Comme nous pouvons le constater dans l'exemple, les contours principaux apparaissent quelle que soit la position des germes.

A partir de M réalisations de contours, la densité de contours est calculée par la méthode de la fenêtre de Parzen. L'effet de lissage du noyau gaussien (typiquement $\sigma = 3$ pixels pour ces exemples) est important pour obtenir une fonction pour laquelle les contours fermés (e.g. dans les régions texturées ou pour les petites régions) sont ajoutés. La densité de contours $pdf(x)$ pourrait ensuite être seuillée pour obtenir les contours les plus proéminents, cependant le résultat serait composé de morceaux de contours qui n'entoureraient pas les régions. De plus, nous avons étudié les histogrammes pour divers exemples et il n'existe pas de seuil optimal pour séparer les classes de contours.

Un schéma présente les différentes étapes nécessaires à l'élaboration de la pdf (fig. 3.19).

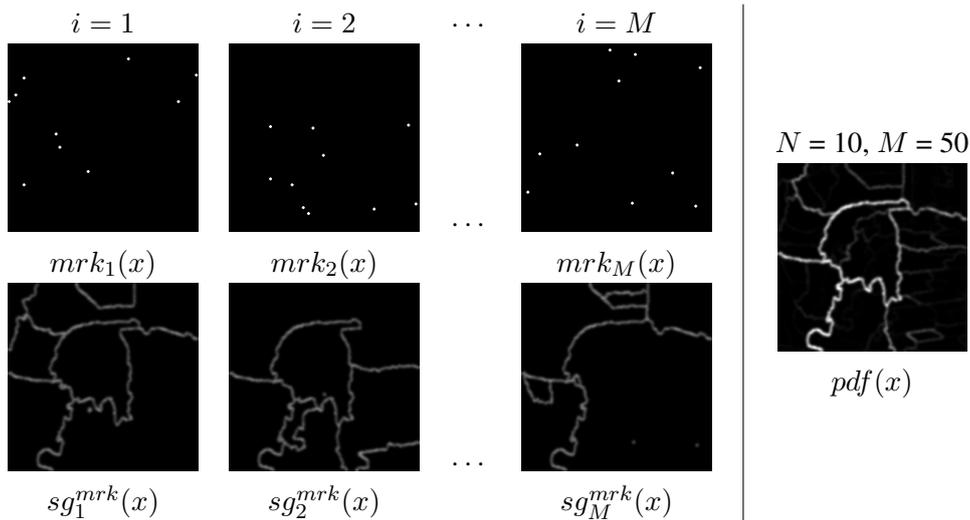


FIG. 3.17 – A gauche, M réalisations de N germes aléatoires uniformes, $mrk_i(x)$, et contours associés de la LPE contrôlée par marqueurs, $sg_i^{mrk}(x)$. A droite, la fonction de densité de probabilité de contours calculée par la méthode de la fenêtre de Parzen pour $N = 10$ et $M = 50$.

L'inconvénient majeur lors de l'utilisation d'une distribution uniforme de marqueurs aléatoires est d'induire une sur-segmentation des régions les plus grandes puisque le nombre moyen de germes tombant dans une région donnée est proportionnel à sa surface. Ce phénomène est évité par l'emploi d'une segmentation par LPE volumique, ou par l'utilisation d'une densité régionalisée de germes. Nous proposons de partitionner la $pdf(x)$ de contours avec la LPE volumique pour obtenir les R régions les plus significatives, i.e. $sg^{R-vol}(pdf(x), R)$. Chaque bassin versant (chaque minimum) de $pdf(x)$ correspond à l'une des régions de la somme des différentes réalisations de segmentation $sg_i^{mrk}(x)$. L'intégrale de chaque bassin versant correspond à la probabilité d'être une région de la segmentation. Par conséquent, la LPE volumique de $pdf(x)$ donne des régions selon leur probabilité. La figure

3.18 montre une comparaison de segmentation en $R = 10, 20$ et 50 régions pour deux $pdf(x)$ différentes. Les résultats peuvent être comparés avec ceux de la segmentation volumique (déterministe) $sg^{R-vol}(q^{LS+H}(x), R)$ (fig. 3.18). Une propriété du filtre gaussien, observée dans les exemples, est la régularisation de la $pdf(x)$ qui se compose de contours de LPE relativement arrondis.

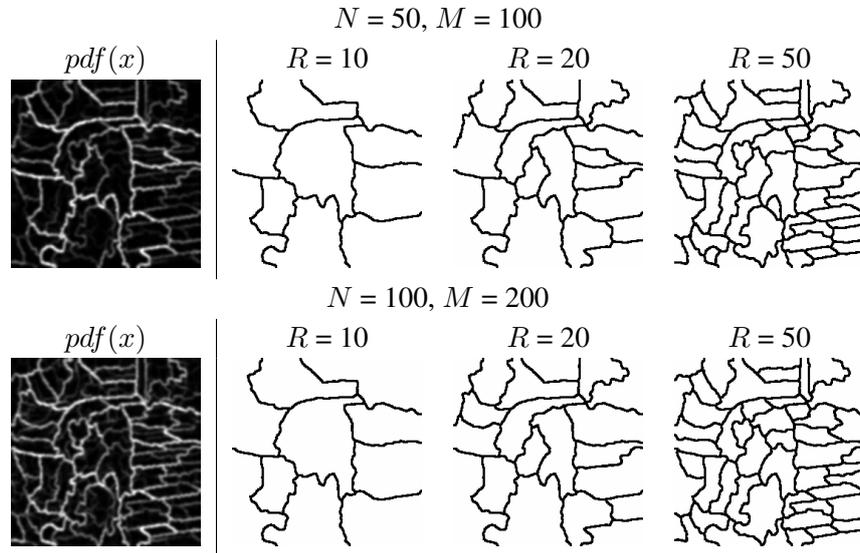


FIG. 3.18 – A gauche, fonctions de densité de probabilités de contours, $pdf(x)$ pour N germes uniformément aléatoires et M réalisations. A droite, segmentation par LPE volumique de $pdf(x)$ en R régions les plus significatives, $sg^{R-vol}(pdf, R)$.

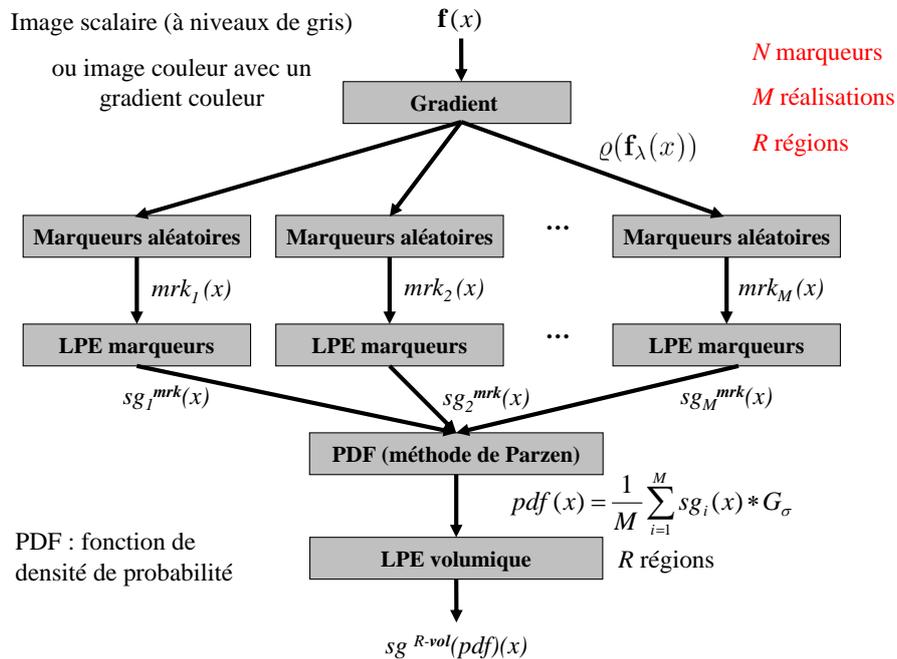


FIG. 3.19 – Schéma général de calcul de la pdf.

3.5.4 Influence des paramètres N et M

A partir des exemples de la figure 3.18, on constate que la méthode dépend très faiblement du nombre de réalisations M , ce qui est caractéristique de sa robustesse. Il a été vérifié empiriquement que la $pdf(x)$ converge vers une distribution stable de contours même pour des faibles valeurs de M , 20 ou 50. Nous proposons de toute manière, de prendre une plus grande valeur, par exemple $M = 100$ ou 200 afin d'obtenir des contours plus réguliers.

Les points aléatoires explorent uniformément l'image et le choix de M est important pour fixer le degré d'échantillonnage probabiliste (notons que la probabilité dépend du ratio entre N et la taille de l'image, c'est à dire du nombre de pixels). En outre, si la valeur de N est faible, une segmentation en larges régions est privilégiée ; tandis qu'une forte valeur de N produira de plus petites régions. Si N est trop grand, la sur-segmentation de sg_i^{mrk} conduit à une $pdf(x)$ très lisse. Elle perd alors ces propriétés pour sélectionner les R contours. En fait, nous pouvons conclure que la segmentation par germes uniformes dépend principalement du paramètre N qui est relié à R , i.e. le nombre de régions à déterminer. Il est donc logique de prendre $N > R$. De plus, la méthode est assez robuste pour le choix de N . Sur la figure 3.18, les images, de taille 256×256 pixels, à segmenter en $R = 10, 20$ ou 50 régions, le choix de $N = 50$ ou 100 fournit exactement le même résultat.

3.5.5 Gradient probabiliste

La fonction $pdf(x)$ peut être combinée avec le gradient initial afin de renforcer les contours du gradient qui ont une forte probabilité :

$$\varrho(x) = \omega_1 \varrho^{LS+H}(\mathbf{f}(x)) + \omega_2 pdf(x) \quad (3.5.3)$$

avec l'utilisation d'une combinaison barycentrique : les deux fonctions définies sur $[0; 1]$ et $\omega_1 = (1 - \iota)$ et $\omega_2 = \iota$.

Nous avons étudié le comportement de $\varrho(x)$ pour la segmentation volumique, $sg^{R-vol}(\rho(x))$, selon la valeur ι . Remarquons que pour $\iota = 0$ le gradient est obtenu alors que pour $\iota = 1$, la pdf est exclusivement prise en compte. Sur la figure 3.20, on peut observer un exemple de segmentation en 20 régions pour différents ι . On constate que même pour de faibles valeurs de ι , les résultats de la segmentation sont notablement améliorés. Ceci est cohérent avec le fait que $pdf(x)$, dérivée du gradient, contient toutes les informations essentielles pour la segmentation. De toute façon, nous avons la confirmation que lorsque $\iota = 0,5$ (combinaison moyenne), les résultats sont en général plus satisfaisants.

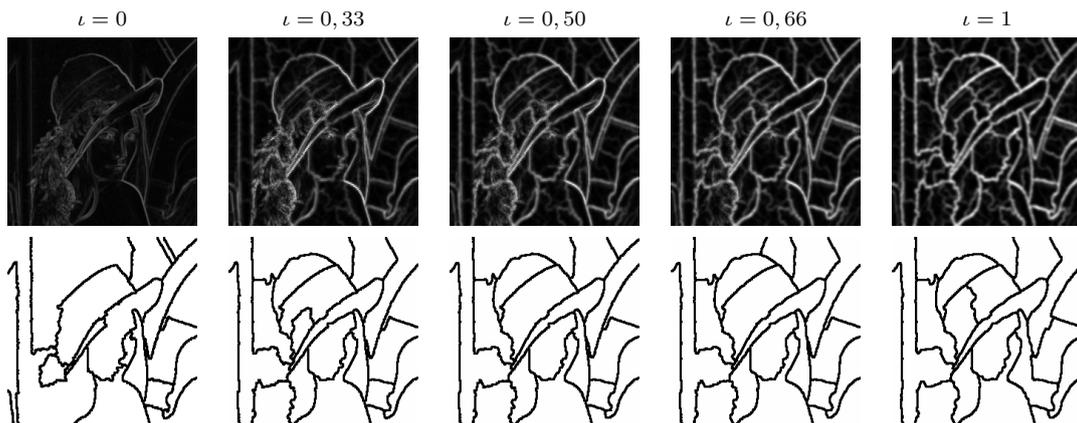


FIG. 3.20 – Première ligne, gradients probabilistes $\varrho(x)$ (i.e., combinaison linéaire de gradient couleur et de pdf) pour différentes valeurs de λ et deuxième ligne, segmentations associées par LPE volumique en $R = 20$ régions.

3.5.6 Segmentation à partir de germes ponctuels aléatoires régionalisés

Dans l'algorithme précédent, les germes aléatoires étaient uniformément distribués dans le domaine de l'image. Une autre approche, pour construire la pdf, par régionalisation des germes a été proposée par *Angulo et Jeulin* [4].

La première question est de savoir quelle fonction de régionalisation $\theta(x)$ utiliser. Plusieurs alternatives sont possibles. Une première alternative consiste à travailler, pour les images couleur, sur la composante de luminance $\theta(x) = f_L(x)$ (respectivement, le négatif de la luminance $f_L^c(x)$) de telle manière que les régions les plus illuminées (respectivement, les régions sombres) produiront des germes aléatoires. Cette forme de régionalisation n'est pas très utile pour la segmentation. Il semble

plus naturel de travailler sur le gradient couleur, $\theta(x) = \rho^{LS+H}(x)$. Dans ce cas, les germes de $mrk_i^\theta(x)$ sont localisés autour des zones de fort gradient, c'est à dire les zones proches des contours. Une fois que les séries de M contours $sg_i^{mrk^\theta}(x)$ sont calculées, la densité de probabilité de contours associée est obtenue par la méthode de la fenêtre de Parzen. Comme précédemment, cette fonction est ensuite segmentée par une LPE volumique (fig. 3.21). On constate que les segmentations régionalisées dépendent des propriétés de la dynamique du gradient couleur. En outre, les différentes réalisations de points aléatoires, utilisant le même $\theta(x)$, sont assez similaires et par conséquent les réalisations de contours aussi. A cause de cet échantillonnage régionalisé, une autre caractéristique de $pdf^\theta(x)$ est que la distribution est très similaire au gradient, uniquement là où les contours sont les plus importants. Les résultats de segmentation de $sg^{R-vol}(pdf^\theta, R)$ sont de toute manière meilleurs que la segmentation volumique du gradient $sg^{R-vol}(\rho^{LS+H}, R)$. Le fait de prendre $\theta(x)$ égale au négatif du gradient (i.e. localisés les germes sur les faibles zones du gradient) a aussi été étudiée. Cependant, dans ce cas, trop de germes sont introduits dans chaque réalisation et la sur-segmentation fourni des pdf inutiles.

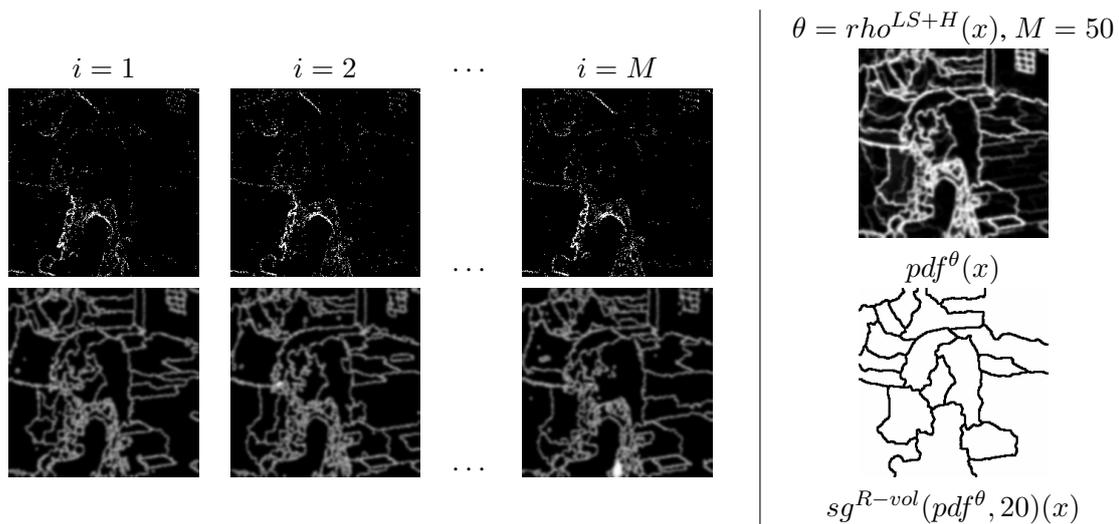


FIG. 3.21 – A gauche, M réalisations de germes aléatoires régionalisés (la fonction de régionalisation est le gradient couleur $\theta = \rho^{LS+H}$), $mrk_i(x)$, et les contours associés de la LPE par marqueurs, $sg_i^{mrk^\theta}(x)$. A droite, la pdf calculée par la méthode de la fenêtre de Parzen pour $M = 50$, et la segmentation volumique de $pdf^\theta(x)$ en $R = 20$ régions.

Remarque 3.5.1. Angulo et Jeulin ont également développé des nivellements morphologiques à partir de germes aléatoires. L'ensemble des réalisations des nivellements est ensuite utilisée pour faire une segmentation par LPE probabiliste. L'intérêt d'utiliser des nivellements est de réduire le nombre de minima du gradient d'une image. Cette approche donne de bons résultats. Nous invitons le lecteur intéressé à se reporter à [4]. Cette méthode n'étant pas utilisée dans la suite de ce mémoire, il ne sera pas donné plus de détails.

3.5.7 Considérations sur l'implémentation informatique

Les M réalisations de contours pour des germes aléatoires, uniformes ou régionalisés, sont obtenues à partir de la même fonction (i.e. le gradient couleur) en utilisant différents marqueurs. Par conséquent, en utilisant le graphe de voisinage des bassins versants et son arbre de poids minimum, « Minimum Spanning Tree » (MST), [125], les N marqueurs aléatoires peuvent être considérés comme N nœuds aléatoires du MST au lieu de N points de l'image. L'implémentation par graphe apporte deux avantages principaux : primo, un calcul rapide de M segmentations avec des marqueurs différents sur le même MST ; et secundo le contrôle du biais de la LPE pouvant être dû à la position aléatoire des marqueurs [27].

Les implémentations rapides de la LPE (100 *ms* pour une image 256×256 pixels sur un ordinateur standard Pentium 4, 3 GHz, 2 Go de RAM) permettent de segmenter une image par la LPE probabiliste en 10 secondes.

Après cette présentation de la LPE probabiliste pour les images couleurs, présentons maintenant les extensions de cette technique aux images hyperspectrales.

3.6 Densité de probabilité de contours pour les images hyperspectrales

Dans cette partie nous présentons tout d'abord plusieurs méthodes pour élaborer des fonctions de densité de probabilité pour les images hyperspectrales à partir de marqueurs uniformément aléatoires.

3.6.1 Densité de probabilité de contours marginale $mpdf$

Comment étendre les fonctions de densité de probabilités pdf de contours aux images hyperspectrales ? Pour ce faire, trois fonctions à inonder ont été étudiées :

- une approche marginale conduisant à une pdf marginale notée $mpdf$;
- une approche vectorielle conduisant à une pdf vectorielle notée $vpdf$;
- un gradient hyperspectral probabiliste ϱ_{prob} .

La pdf marginale pondérée $mpdf$ est calculée selon l'algorithme 3. Un schéma représente également les différentes étapes nécessaires à son calcul (fig. 3.22).

Le gradient hyperspectral probabiliste est défini comme la somme du gradient vectoriel construit sur une distance et de la pdf marginale $mpdf$:

$$\varrho_{prob} = mpdf + \varrho^d. \quad (3.6.3)$$

Il est à noter que les valeurs de la $mpdf$ et du gradient distance ϱ^d ont été normalisés dans $[0; 1]$.

Algorithme 3 *mpdf*

- 1: Pour le gradient morphologique de chaque canal, $\varrho(f_{\lambda_j})$ avec $j = 1 \dots L$, faire M réalisations de N germes uniformément aléatoires, i.e. les marqueurs $\{mrk_i^j\}_{i=1 \dots M}^{j=1 \dots L}$. Comme on génère M réalisations de marqueurs pour chaque canal, il y a en tout $M \times L$ réalisations. Obtenir la série de contours par segmentation LPE à partir des marqueurs sur les gradients morphologiques des canaux, $\{sg_i^{mrk_i^j}(x)\}_{i=1 \dots M}^{j=1 \dots L}$.
- 2: Calculer par la méthode de la fenêtre de Parzen les pdf marginales sur chacun des canaux :

$$pdf_j(x) = \frac{1}{M} \sum_{i=1}^M sg_i^j(x) * G_\sigma . \quad (3.6.1)$$

- 3: Calculer la pdf marginale pondérée :

$$mpdf(x) = \sum_{j=1}^L w_j pdf_j(x) \quad (3.6.2)$$

avec des poids w_j . Dans l'espace image de dimension L , on peut prendre $w_j = 1/L$ et dans l'espace factoriel prendre les poids w_j égaux aux inerties des axes.

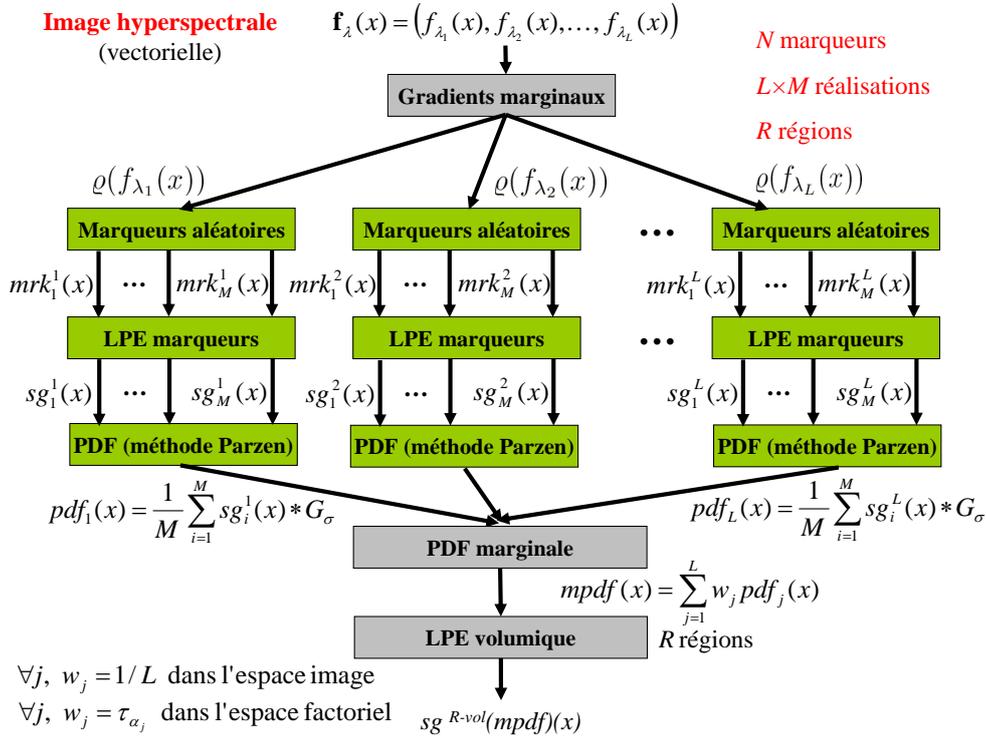


FIG. 3.22 – Schéma général de calcul de la pdf marginale $mpdf$.

3.6.2 Densité de probabilité de contours vectorielle $vpdf$

La pdf vectorielle $vpdf$ est calculée selon l'algorithme 4. Les étapes pour l'obtention de la $vpdf$ sont présentées sur un schéma (fig. 3.23).

Algorithme 4 $vpdf$

- 1: Pour le gradient vectoriel $\varrho^d(\mathbf{f}_\lambda)$, faire $M \times L$ réalisations de N germes uniformément aléatoires, i.e. les marqueurs $\{mrk_i\}_{i=1\dots M \times L}$, avec L le nombre de canaux. Obtenir la segmentation, $\{sg_i(x)\}_{i=1\dots M \times L}$, par LPE sur le gradient vectoriel construit sur une distance $\varrho^d(\mathbf{f}_\lambda)$, avec d la distance adaptée à l'espace de représentation : $d = d_{\chi^2}$ dans l'espace image ou $d = d_E$ dans l'espace factoriel.
- 2: Calculer la pdf vectorielle :

$$vpdf(x) = \frac{1}{M \times L} \sum_{i=1}^{M \times L} sg_i(x) * G_\sigma. \quad (3.6.4)$$

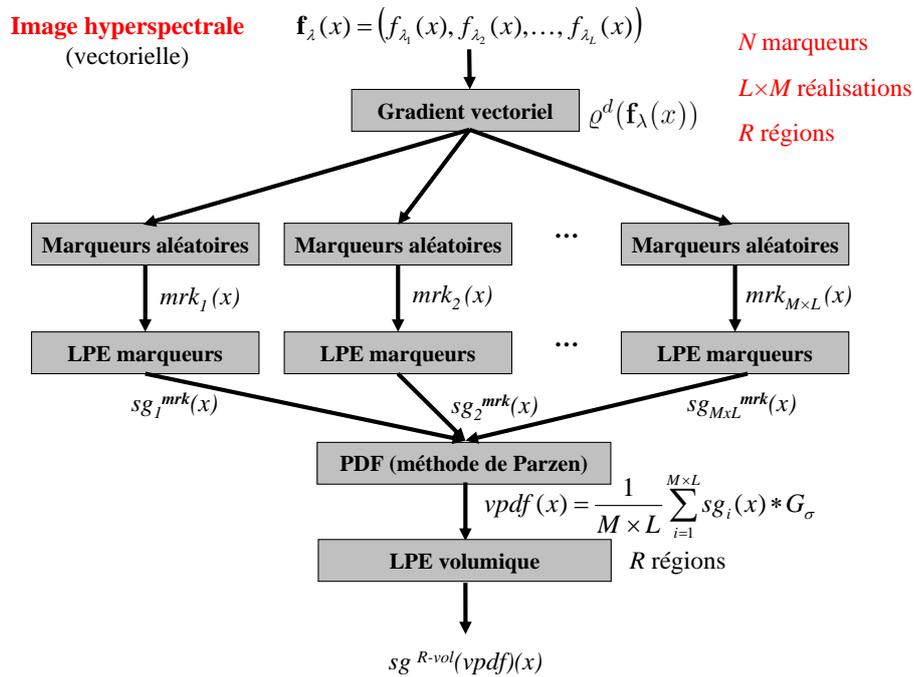


FIG. 3.23 – Schéma général de calcul de la pdf marginale $vpdf$.

3.6.3 Comparaison entre les densités de probabilités de contours

Pour comparer ces trois fonctions à inonder, on les segmente par une LPE volumique avec une contrainte de volume R . La comparaison s'effectue sur des images satellites dans l'espace image \mathbf{f}_λ et dans l'espace factoriel \mathbf{c}_α^f . Les marqueurs sont des marqueurs uniformément aléatoires : $mrk_i(x)$.

Dans l'espace image (fig. 3.24) on effectue 4 segmentations par LPE volumique à partir :

- du gradient construit sur la distance du chi-deux : $sg^{R-vol}(\varrho^{\chi^2}(\mathbf{f}_\lambda))$;
- de la $mpdf$: $sg^{R-vol}(mpdf(\mathbf{f}_\lambda))$;
- de la $vpdf$: $sg^{R-vol}(vpdf(\mathbf{f}_\lambda))$;
- du gradient probabiliste : $sg^{R-vol}(\varrho_{prob}(\mathbf{f}_\lambda))$.

On constate que les résultats de segmentations sont bien meilleurs (i.e. les contours sont plus homogènes dans l'espace) pour les LPE volumique sur les fonctions d'inondation probabilistes : $mpdf(\mathbf{f}_\lambda)$, $vpdf(\mathbf{f}_\lambda)$ et $\varrho_{prob}(\mathbf{f}_\lambda)$ que sur le gradient vectoriel $\varrho^{\chi^2}(\mathbf{f}_\lambda)$ qui correspond à la LPE classique, i.e. déterministe. Par conséquent la LPE probabiliste est meilleure que la LPE déterministe sur les images hyperspectrales. En outre le gradient probabiliste ϱ_{prob} , qui combine un gradient vectoriel construit sur une distance ϱ^d et pdf marginale $mpdf$, améliore encore les contours puisqu'il permet de segmenter à la fois les contours du premier ordre (les plus grandes régions) et les contours du second ordre liés à des petites régions mais qui sont bien contrastées dans le gradient.

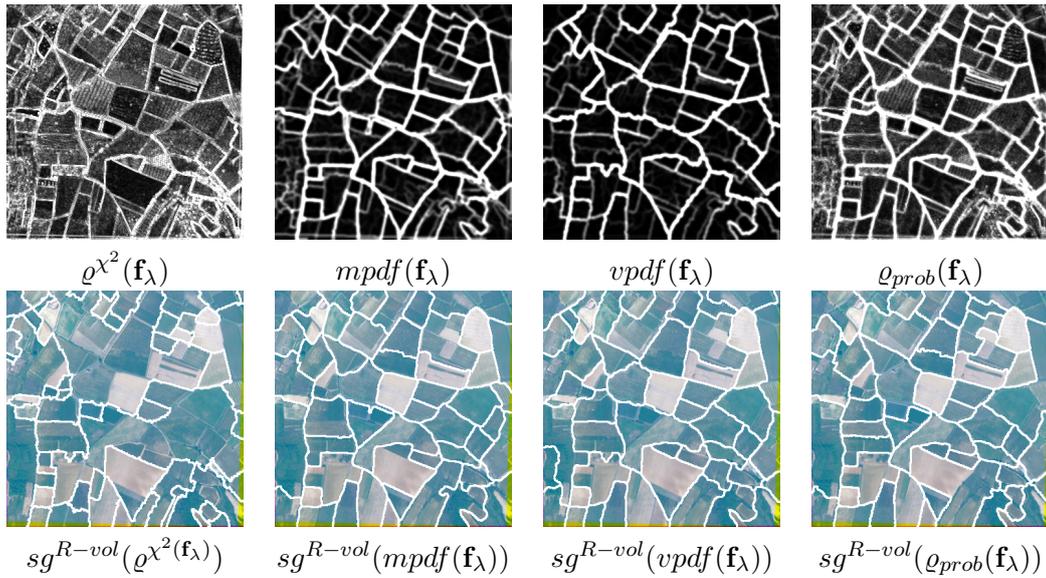


FIG. 3.24 – En haut : gradients et pdf ; en bas : segmentations LPE associés sur l'image « Roujan » dans l'espace image \mathbf{f}_λ avec $N = 50$ points , $M = 100$ réalisations et $R = 50$ régions.

Nous avons fait une étude similaire dans l'espace factoriel AFC à une dimension $\mathbf{c}_{\alpha_1}^f = (c_{\alpha_1}^f)$. Comme seul le premier axe factoriel est conservé, le gradient construit sur la distance euclidienne ϱ^E est très proche du gradient morphologique ϱ sur les facteurs pixels de l'axe $c_{\alpha_1}^f$. On utilisera donc le gradient morphologique dans cet espace factoriel à une dimension. Le gradient probabiliste sera donc défini comme la somme du gradient morphologique et de la $mpdf$: $\varrho_{prob}(c_{\alpha_1}^f) = \varrho(c_{\alpha_1}^f) + mpdf(c_{\alpha_1}^f)$.

Nous avons effectué les segmentation par LPE volumique associées (fig. 3.25) :

- au gradient construit sur la distance euclidienne : $sg^{R-vol}(\varrho^E(c_\alpha^f))$;

- à la *mpdf* : $sg^{R-vol}(mpdf(\mathbf{c}_{\alpha}^f))$;
- au gradient probabiliste : $sg^{R-vol}(\varrho_{prob}(\mathbf{c}_{\alpha}^f))$.

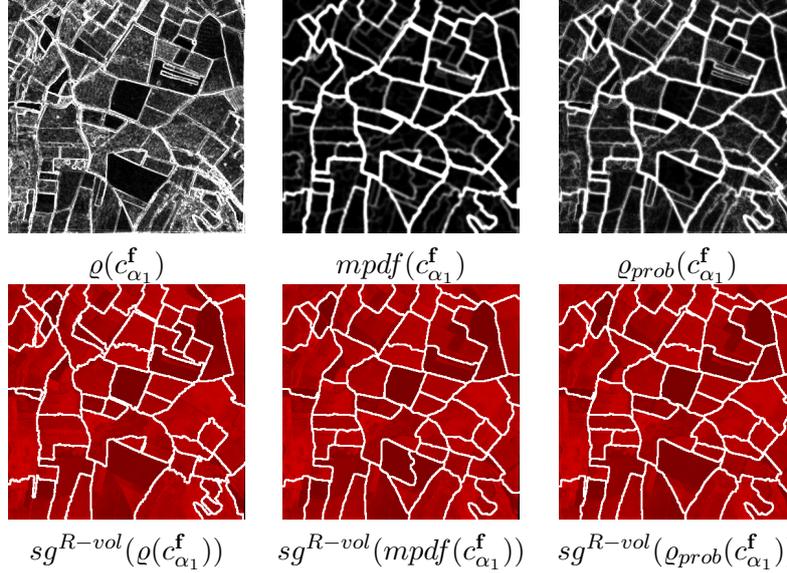


FIG. 3.25 – Segmentations LPE dans l'espace factoriel AFC \mathbf{c}_{α}^f $N = 50$ points, $M = 100$ réalisations et $R = 50$ régions.

On constate que les contours de la LPE du gradient morphologique $\varrho(\mathbf{c}_{\alpha}^f)$ sont comparables à ceux de la LPE sur les fonctions à inonder probabilistes : $mpdf(\mathbf{c}_{\alpha}^f)$ et $\varrho_{prob}(\mathbf{c}_{\alpha}^f)$. En effet, grâce à l'AFC, le bruit a été réduit dans l'espace factoriel \mathbf{c}_{α}^f (cf. section 2.4.2, p : 32 et suivantes). Travailler dans cet espace permet donc l'amélioration de la segmentation. Les contours du premier ordre (i.e. associés aux régions les plus importantes) semblent meilleurs pour les fonctions à inonder probabilistes $mpdf(\mathbf{c}_{\alpha}^f)$ et $\varrho_{prob}(\mathbf{c}_{\alpha}^f)$ que pour le gradient morphologique (i.e. déterministe) $\varrho(\mathbf{c}_{\alpha}^f)$.

Nous avons également effectué des comparaisons entre les fonctions à inonder probabilistes, $mpdf$ et $vpdf$, et le gradient construit à partir d'une distance ϱ^d (fig. 3.26 et 3.27). On constate que les segmentations sur la pdf marginale $mpdf$ sont aussi bonnes que celles sur la pdf vectorielle $vpdf$. Par ailleurs la pdf vectorielle nécessite le choix d'une distance, alors que ce n'est pas le cas pour la pdf marginale. Par conséquent, on utilisera par la suite la pdf marginale $mpdf$. Néanmoins, il peut être intéressant d'utiliser la pdf vectorielle $vpdf$ avec moins de réalisations : i.e. M au lieu de $M \times L$. Ainsi les temps de calculs sont accélérés par rapport à la $mpdf$ où il faut faire M réalisations sur chacun des L canaux.

Après avoir introduit des pdf pouvant être utilisées pour les images hyperspectrales, nous introduisons de nouveaux marqueurs aléatoires régionalisés par la classification.

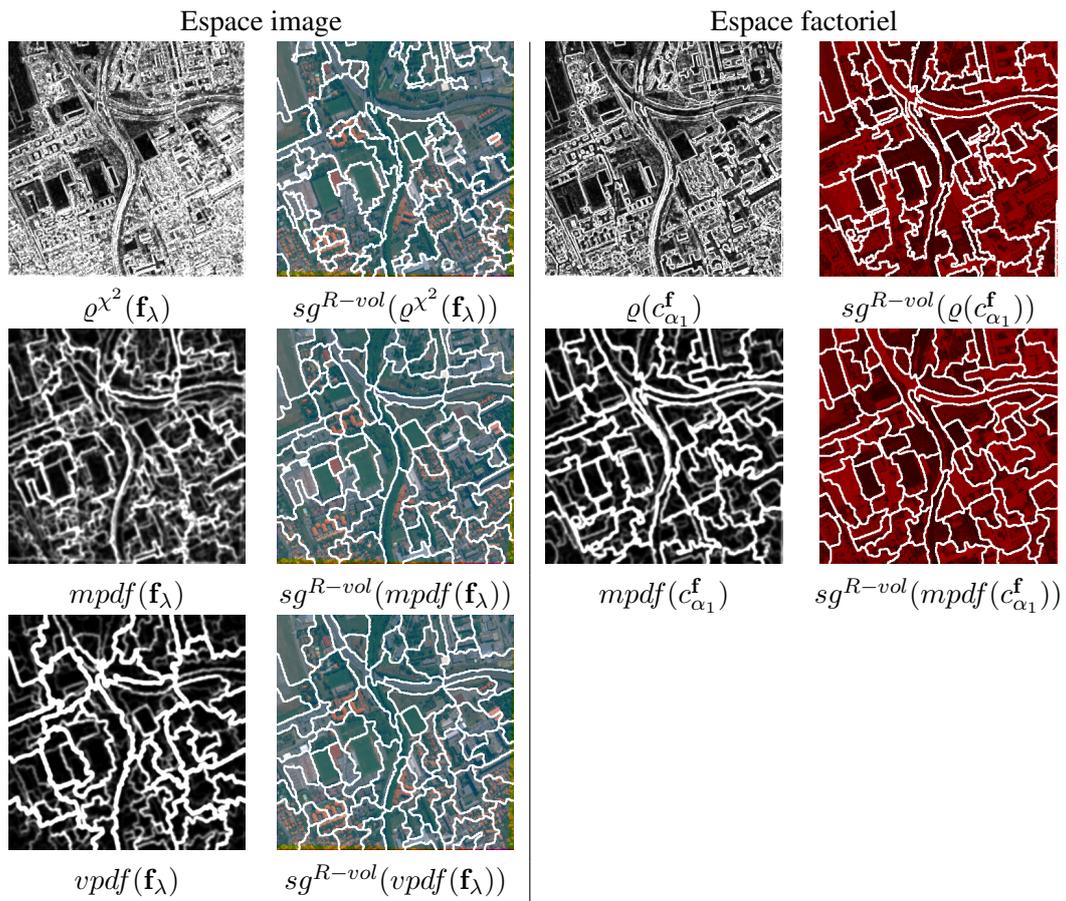


FIG. 3.26 – Segmentations par LPE volumique et fonctions à inonder associées dans l'espace image \mathbf{f}_λ et dans l'espace factoriel $\mathbf{c}_\alpha^{\mathbf{f}} = (c_{\alpha_1}^{\mathbf{f}})$ (inertie 77,4 %) pour l'image « Toulouse » (5 canaux) avec $N = 100$, $M = 100$, $R = 75$.

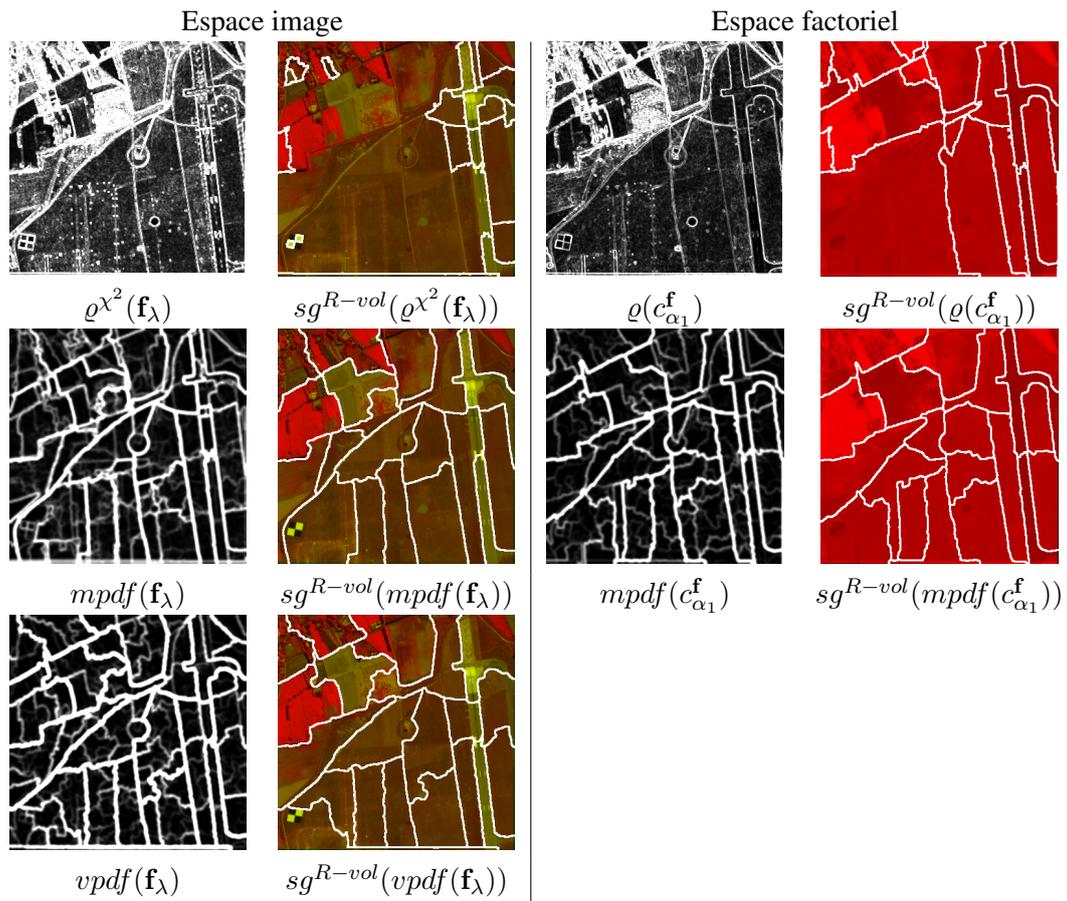


FIG. 3.27 – Segmentations par LPE volumique et fonctions à inonder associées dans l'espace image \mathbf{f}_λ et dans l'espace factoriel $\mathbf{c}_\alpha^{\mathbf{f}} = (c_{\alpha_1}^{\mathbf{f}})$ (inertie 82,2%) pour l'image « Salon de Provence » (4 canaux) avec $N = 50$, $M = 100$, $R = 20$.

3.7 Conditionnement des germes de la pdf par la classification

La pdf avec des germes (ou marqueurs) uniformément aléatoires ne contient que de l'information spatiale. Le conditionnement des germes aléatoires par la classification spectrale permet d'introduire la dimension spectrale dans la pdf, ce qui conduit à une pdf spatio-spectrale.

Dans cette partie nous introduisons de nouveaux marqueurs aléatoires qui sont régionalisés par une classification. Puis nous présentons des résultats de segmentation à partir de ces marqueurs.

Voici tout d'abord les différents types de marqueurs étudiés dans cette partie de la thèse :

1. des marqueurs uniformément aléatoires $mrk_i(x)$;
2. des marqueurs aléatoires régionalisés $mrk_i^\theta(x)$;
3. des marqueurs aléatoires régionalisés par une classification :
 - a) sous forme ponctuelle $mrk_i^{\kappa-pt}(x)$;
 - b) sous forme de boules où :
 - une classe connexe peut être touchée au plus une fois $mrk_i^{\kappa-b}(x)$;
 - une classe connexe peut être touchée plusieurs fois et l'on fait l'union des marqueurs boules tombés dans une classe connexe $mrk_i^{\kappa-\cup b}(x)$;
 - une classe connexe peut être touchée plusieurs fois et l'on fait l'union des marqueurs boules connexes tombés dans une classe connexe $mrk_i^{\kappa-\cup b-connex}(x)$.

Les marqueurs uniformément aléatoires $mrk_i(x)$ et les marqueurs régionalisés $mrk_i^\theta(x)$ ont été présentées dans la section 3.5 (p : 99). L'obtention des marqueurs uniformément aléatoires $mrk_i(x)$ consiste à tirer des N germes de manière uniforme dans l'espace image. Quant aux marqueurs aléatoires régionalisés $mrk_i^\theta(x)$, leur tirage est effectué selon la densité de probabilité θ . On choisit souvent un gradient normalisé comme densité θ .

Nous allons présenter des résultats sur l'image de souris « serim447 ». Plusieurs espaces de représentations sont considérés :

- l'espace factoriel AFC \mathbf{c}_α^f ;
- l'espace des paramètres \mathbf{p} ;
- l'espace factoriel ACP de l'ensemble d'entraînement utilisé lors de la classification LDA \mathbf{c}_β^{train} .

Ce dernier espace peut sembler « étrange », cependant il est adapté aux classes de la classification par LDA et il a permis de bien les différencier.

3.7.1 LPE déterministe

Tout d'abord, les segmentations par LPE déterministes par marqueurs issus de la classification ou sur critère volumique ne permettent pas de segmenter correctement l'image de la souris.

Pour les segmentations par méthode déterministe, on utilise des gradients construits sur une distance dont la distance est adaptée à l'espace image (fig. 3.28 et 3.29). On observe que même si un certain nombre de régions sont détectées, les segmentations ne sont pas très bonnes. La segmentation par marqueurs dans l'espace ACP des paramètres d'entraînement $sg(\varrho_E(\mathbf{c}_\beta^{train}), mrk_\kappa)$ (fig. 3.29 c) semble la meilleure mais reste limitée.

Remarque 3.7.1. Le gradient construit sur distance a été calculé après filtrage par nivellement sur chacun des canaux de l'espace image considéré. La référence pour le nivellement est obtenue par un filtre gaussien de taille 11×11 pixels.

Remarque 3.7.2. Les marqueurs mrk_κ (fig. 3.28) ont été obtenus par filtrage morphologique de la classification à partir d'une ouverture avec un e.s. hexagonal de taille 5 (cf. section 3.3, p : 82).

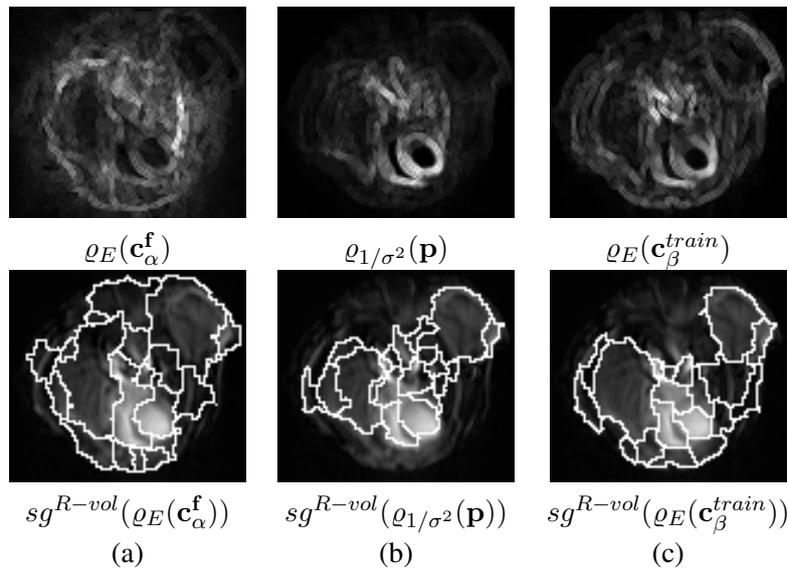


FIG. 3.28 – Gradients construits sur des distances et segmentations par LPE volumique associées en $R = 20$ régions : (a) dans l'espace factoriel \mathbf{c}_α^f , (b) dans l'espace des paramètres \mathbf{p} et (c) dans l'espace ACP du train sur les paramètres \mathbf{c}_β^{train} .

Les segmentations par LPE déterministes n'étant pas satisfaisantes, intéressons nous maintenant à la LPE probabiliste.

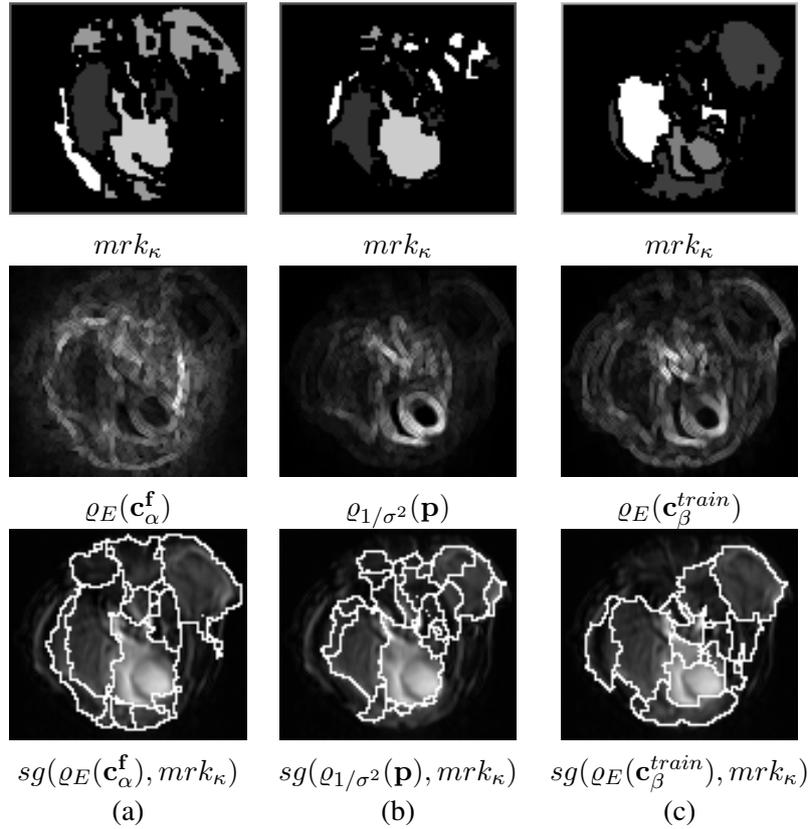


FIG. 3.29 – Segmentations LPE par marqueurs dans différents espaces : (a) dans l’espace factoriel \mathbf{c}_α^f , (b) dans l’espace des paramètres \mathbf{p} et (c) dans l’espace ACP du train sur les paramètres \mathbf{c}_β^{train} .

3.7.2 LPE probabiliste avec des germes uniformément aléatoires

Afin de renforcer les contours de la fonction à inonder, on emploie la LPE probabiliste avec des marqueurs uniformément aléatoires $mrk_i(x)$ comme expliqué à la section 3.6 (p : 109). Pour ce faire, on utilise la pdf marginale $mpdf$ avec $N = 20$ points et $M = 100$ réalisations. On effectue ensuite une segmentation par LPE volumique en $R = 20$ régions (fig. 3.30). Les segmentations de la $mpdf$ par LPE volumique sont présentées sur un canal de l’espace de représentation considéré : espace factoriel \mathbf{c}_α^f , espace des paramètres \mathbf{p} et espace ACP d’entraînement \mathbf{c}_β^{train} .

Les résultats de segmentation sont meilleurs (au niveau de la tumeur) dans l’espace ACP de l’ensemble d’entraînement \mathbf{c}_β^{train} . Cependant ils ne sont pas parfaits notamment au niveau des autres organes.

Remarque 3.7.3. Pour éviter la segmentation du fond en plusieurs parties, il serait intéressant d’ajouter un marqueur sur le contour de l’image pour chacune des réalisations des marqueurs aléatoires $mrk_i(x)$, avec $i = 1 \dots M$. Ce marqueur n’aurait bien sûr aucun impact sur la tumeur qui est située à l’intérieur de l’image et non sur son bord.

Pour améliorer ces résultats nous allons conditionner les marqueurs par la classification spectrale précédemment obtenue.

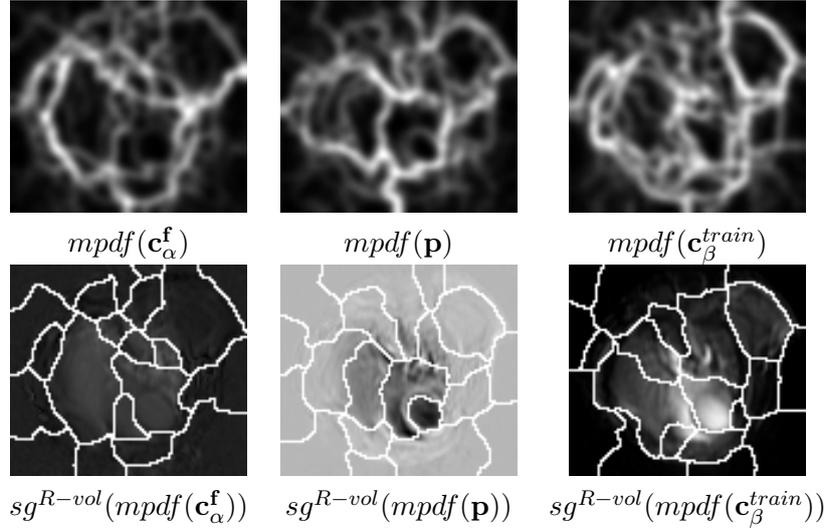


FIG. 3.30 – Segmentation par LPE probabiliste sur la $mpdf$ par critère volumique ($N = 20$ points, $M = 100$ réalisations, $R = 20$ régions).

3.7.3 Germes ponctuels aléatoires régionalisés par une classification

Afin de tenir compte de l'information spectrale dans la construction de la pdf, on va tirer des marqueurs conditionnellement à la classification κ .

Pour ce faire, les classes de la classification κ sont ré-étiquetées (cf. section 3.3.1, p : 83) pour obtenir une partition $\hat{\kappa}$ en composantes connexes de l'espace support de l'image E :

$$E = \cup_k C_k \text{ avec } \cap C_k = \emptyset .$$

Il est à noter qu'il n'est pas nécessaire de faire un filtrage morphologique sur les classes. En effet comme les classes connexes de la classification ne sont pas utilisées comme des marqueurs pour la LPE sur la $mpdf$, la LPE ne requiert pas de degré de liberté entre les classes (cf. section 3.3.1, p : 83).

Des marqueurs ponctuels aléatoires régionalisés $mrk_i^{\kappa-pt}(x)$ sont tirés sur les classes connexe C_k de la classification spectrale. Pour cela on effectue un tirage de N marqueurs aléatoires par acceptation rejet. Les marqueurs sont tirés uniformément sur les classes C_k . Si un marqueur tombe dans une composante connexe C_k non encore marquée alors ce marqueur est conservé sinon il est rejeté (cf. algorithme 5). Afin de réduire l'influence des trop petites classes on ajoute un paramètre d'aire minimum S : i.e. $aire(C_k) > S$.

Il est à noter que, pour cette application, on ajoute un marqueur externe lors des réalisations des marqueurs et que les marqueurs ne peuvent toucher la classe correspondant au fond, afin d'éviter d'avoir des contours séparant le fond sur la pdf.

Remarque 3.7.4. Tous les N marqueurs ne sont pas conservés dans cet algorithme. Il faut donc tirer un nombre de points N suffisant pour que les principales classes puissent être marquées. Les marqueurs conservés sont appelés marqueurs implantés.

Les segmentations de la $mpdf$ par LPE volumique dans les différents espaces sont présentées sur un canal de l'espace de représentation considéré : espace factoriel \mathbf{c}_α^f , espace des paramètres \mathbf{p} et espace ACP d'entraînement \mathbf{c}_β^{train} (fig. 3.31).

Algorithme 5 Marqueurs ponctuels aléatoires régionalisés par une classification $mrk_i^{\kappa-pt}(x)$

- 1: On fixe N le nombre de germes m tirés et S l'aire minimale d'une composante connexe.
 - 2: indiquer éventuellement que la classe connexe du fond est marquée
 - 3: ajouter éventuellement un marqueur contour sur la réalisation $mrk_i^{\kappa-pt}(x)$
 - 4: **Pour tout** marqueurs m de 1 à N **Faire**
 - 5: Déterminer la composante connexe C_k telle que $m \in C_k$
 - 6: **Si** C_k est non marquée et $aire(C_k) \geq S$ **Alors**
 - 7: conserver m
 - 8: indiquer que C_k est marquée
 - 9: **Fin Si**
 - 10: **Fin Pour**
-

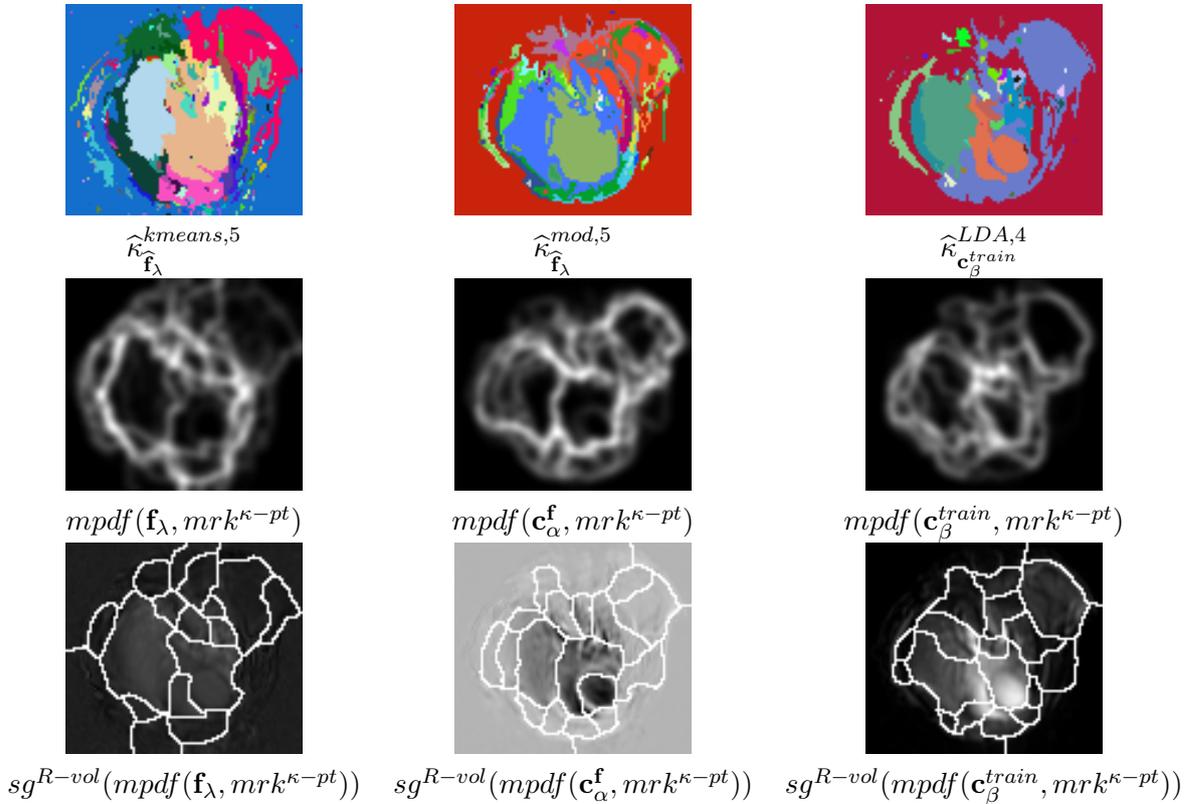


FIG. 3.31 – Segmentations par LPE probabiliste par critère volumique sur la $mpdf$ calculée avec des marqueurs ponctuels aléatoires régionalisés par la classification κ ($N = 100$ points, $M = 100$ réalisations, $R = 20$ régions, aire $S = 10$ pixels).

Dans l'espace ACP d'entraînement, on utilise la classification par LDA sur cet espace $\kappa^{LDA,4}_{\mathbf{c}_\beta^{train}}$ et non sur les paramètres $\kappa^{LDA,4}_p$. On pourrait néanmoins utiliser celle sur les paramètres qui est très proche de celle dans l'espace ACP d'entraînement (cf. section 2.6.4, p : 63).

Pour avoir des segmentations comparables aux autres figures il faut prendre un nombre de points $N = 100$ pour une aire minimale $S = 10$ pixels, ce qui fait que l'on a en moyenne 15 marqueurs

implantés sur chacune des réalisations $mrk_i^{\kappa-pt}(x)$, avec $i = 1 \dots M$.

On observe que les fonctions à inonder $mpdf$ calculées avec des marqueurs ponctuels régionalisés $mrk_i^{\kappa-pt}$ sont bien meilleures que les $mpdf$ calculées avec des marqueurs ponctuels uniformément aléatoires mrk_i . La segmentation dans l'espace des paramètres $sg^{R-vol}(mpdf(\mathbf{c}_\alpha^f, mrk^{\kappa-pt}))$, qui nécessite une classification non supervisée, et la segmentation dans l'espace ACP de l'ensemble d'entraînement $sg^{R-vol}(mpdf(\mathbf{c}_\beta^{train}, mrk^{\kappa-pt}))$ (classification semi-supervisée) sont les meilleures.

Remarque 3.7.5. Même si un marqueur contour sur les réalisations de marqueurs $mrk_i^{\kappa-pt}$ est ajouté, pour le calcul de la pdf, et même si les marqueurs aléatoires ne peuvent toucher la classe correspondant au fond de l'image, la segmentation de la pdf sépare la classe correspondant au fond en plusieurs parties. En effet la pdf touche le bord de cette image, car l'espace support E de l'image est trop petit. Lorsque l'on prend un espace support plus large le problème disparaît. Par la suite on n'utilisera pas de marqueur externe.

Néanmoins, afin d'améliorer les pdf et donc les segmentations nous avons testés des marqueurs boules aléatoires régionalisés par la classification.

3.7.4 Germes boules aléatoires régionalisés par une classification

Les marqueurs ponctuels présentent l'inconvénient d'accentuer les faibles contours. Avec des marqueurs non ponctuels on diminue la probabilité des contours peu importants. Plusieurs cas ont été étudiés pour les marqueurs boules aléatoires régionalisés par une classification.

3.7.4.1 Une classe connexe peut être touchée au plus une fois

Les marqueurs boules aléatoires régionalisés sont un ensemble de boules de rayon aléatoire compris dans $[1; Rmax]$. On tire les centres m de ces boules comme des marqueurs ponctuels aléatoires régionalisés $mrk^{\kappa-pt}(x)$. Si le centre m est conservé on tire uniformément un rayon aléatoire r dans $[1; Rmax]$. On ne conserve comme marqueur que l'intersection entre la boule, $B(m, r)$ de centre m et de rayon r , et la classe connexe C_k (cf. algorithme 6). Chaque classe connexe C_k ne peut donc être touchée au plus qu'une fois. On note ces marqueurs $mrk_i^{\kappa-b}(x)$.

Pour cette application, les marqueurs ne peuvent toucher la classe correspondant au fond, afin d'éviter d'avoir des contours séparant le fond sur la pdf.

Les segmentations dans les différents espaces sont présentées sur la figure 3.32. Du fait des marqueurs boules $mrk^{\kappa-b}$ plus larges que des marqueurs ponctuels $mrk^{\kappa-pt}$ les segmentations sont meilleures qu'avec des marqueurs ponctuels aléatoires régionalisés par une classification.

La segmentation dans l'espace des paramètres, $sg^{R-vol}(mpdf(\mathbf{c}_\alpha^f, mrk^{\kappa-b}))$, (avec une classification non supervisée), et la segmentation dans l'espace ACP de l'ensemble d'entraînement, $sg^{R-vol}(mpdf(\mathbf{c}_\beta^{train}, mrk^{\kappa-b}))$, (avec une classification semi-supervisée) restent les meilleures.

L'inconvénient de cette méthode est que seul un petit nombre de classes connexes C_k sont marquées à chaque fois : en moyenne 15 classes sont marquées pour chaque réalisation de marqueurs $mrk^{\kappa-b}$, avec $i = 1 \dots M$, sur 41 classes connexes de la classification. Pour palier à cette limitation, on va permettre aux marqueurs de toucher plusieurs fois une classe connexe.

Algorithme 6 Marqueurs boules aléatoires régionalisés par une classification (chaque classe pouvant être touché au plus une fois) $mrk_i^{\kappa-b}(x)$

- 1: On fixe N le nombre de marqueurs m tirés et S l'aire minimale d'une composante connexe.
- 2: **Pour tout** marqueurs m de 1 à N **Faire**
- 3: Déterminer la composante connexe C_k telle que $m \in C_k$
- 4: **Si** C_k est non marquée et $aire(C_k) \geq S$ **Alors**
- 5: $r = \mathcal{U}[1, Rmax]$
- 6: conserver comme marqueur $B(m, r) \cap C_k$
- 7: indiquer que C_k est marquée
- 8: **Fin Si**
- 9: **Fin Pour**

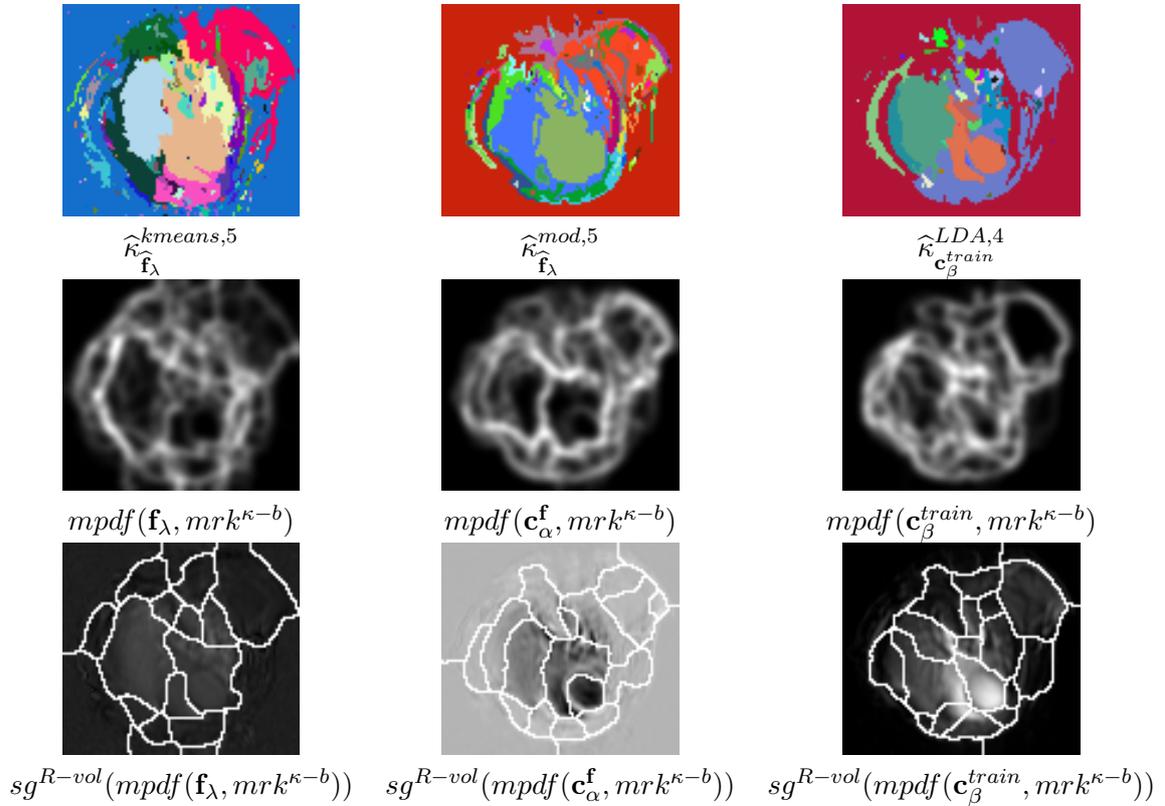


FIG. 3.32 – Segmentation par LPE probabiliste par critère volumique sur la $mpdf$ calculée avec des marqueurs-boules aléatoires régionalisés ($N = 100$ points, $M = 100$ réalisations, $R = 20$ régions, aire $S = 10$ pixels, $Rmax = 30$ pixels).

3.7.4.2 Une classe connexe peut être touchée plusieurs fois

Nous avons étudié deux possibilités pour ces marqueurs : l'union des marqueurs boules tombés dans une classe connexe ou l'union des marqueurs boules connexes tombés dans une classe connexe. A la différence des marqueurs boules aléatoires régionalisés par une classification $mrk_i^{\kappa-b}(x)$, plusieurs boules peuvent tomber dans une classe connexe C_k de la classification.

3.7.4.2.1 Union des germes boules tombés dans une classe connexe

On tire les centres des boules aléatoirement. On ne conserve que l'intersection entre la boule, $B(m_n, r)$ de centre m_n et de rayon r , et la classe connexe C_k . Cette intersection est notée $B_n^{C_k} = B(m_n, r) \cap C_k$. Lorsque plusieurs boules tombent dans une même classe C_k on fait l'union des marqueurs boules $B_n^{C_k}$ de cette classe connexe : $B_0^{C_k} \cup B_1^{C_k} \cup \dots \cup B_N^{C_k}$ (cf. algorithme 7). On note ces marqueurs $mrk_i^{\kappa-\cup b}(x)$.

Algorithme 7 Marqueurs-boules aléatoires régionalisés avec union des boules dans chaque classe connexe de la classification $mrk_i^{\kappa-\cup b}(x)$

- 1: On fixe N le nombre de points m_n tirés et S l'aire minimale d'une composante connexe.
 - 2: **Pour tout** points m_n de 1 à N **Faire**
 - 3: Déterminer la composante connexe C_k telle que $m_n \in C_k$
 - 4: **Si** $aire(C_k) \geq S$ **Alors**
 - 5: $r = \mathcal{U}[1, Rmax]$
 - 6: $B_n^{C_k} = (B(m_n, r) \cap C_k) \cup B_{n-1}^{C_k}$
 - 7: **Fin Si**
 - 8: **Fin Pour**
-

Comme on a vu précédemment que, pour l'exemple de la souris, la segmentation dans l'espace des paramètres \mathbf{p} donne de bons résultats, les segmentations suivantes sont faites uniquement dans cet espace. Pour conditionner les marqueurs on utilise des classifications :

- par approche modèle, effectuée dans l'espace image filtré, $\kappa_{\hat{\mathbf{f}}_\lambda}^{mod,5}$;
- LDA, effectuée dans l'espace des paramètres, $\kappa_{\mathbf{p}}^{LDA,4}$.

En outre, un filtrage morphologique des classes de la classification est appliqué (cf. section 3.3.3, p : 85). Ce filtrage conduit aux images qui vont servir à conditionner les marqueurs aléatoires $mrk_i^{\kappa-\cup b}$: $\hat{\kappa}_{\hat{\mathbf{f}}_\lambda}^{mod,5}$ et $\hat{\kappa}_{\mathbf{p}}^{LDA,4}$. Comme ces images peuvent aussi être utilisées comme marqueurs pour la LPE sur la *mpdf*, elles sont également notées : $mrk_{\hat{\kappa}_{\hat{\mathbf{f}}_\lambda}^{mod,5}}$ et $mrk_{\hat{\kappa}_{\mathbf{p}}^{LDA,4}}$.

Ce filtrage a pour but :

- d'éviter qu'un marqueur aléatoire ne tombe sur la limite entre deux bassins versants et crée ainsi une fuite. En réduisant la surface des classes connexes, de la classification, on diminue le risque qu'une classe ne recouvre les limites du bassin versant auquel elle appartient ;
- d'augmenter les « degrés de libertés » de la LPE lors de la segmentation de la pdf.

Dans l'espace des paramètres, on utilise la classification par LDA sur cet espace $\kappa_{\mathbf{p}}^{LDA,4}$ et non sur les paramètres $\kappa_{\mathbf{p}}^{LDA,4}$. On pourrait néanmoins utiliser celle sur les paramètres qui est très proche de celle dans l'espace ACP d'entraînement (cf. section 2.6.4, p : 63).

Remarque 3.7.6. Un marqueur contour est également ajouté lors des réalisations des marqueurs, $\{mrk_i^{\kappa-\cup b}\}_{i=1\dots M}$, pour le calcul de la pdf. De plus les marqueurs aléatoires ne peuvent toucher la classe correspondant au fond de l'image.

Sur la figure 3.33, les segmentations de la *mpdf* par critère volumique ou bien par marqueurs donnent de bons résultats.

Il est important de noter que l'on introduit ici le concept important de segmentation de la *mpdf* par une LPE à partir de marqueurs provenant de la classification. Nous reviendrons sur ces conséquences

par la suite.

Un dernier test a été effectué en choisissant comme germes de la *mpdf* non plus des boules tombées dans une classe connexe, de la classification, mais des boules connexes tombées dans une classe connexe.

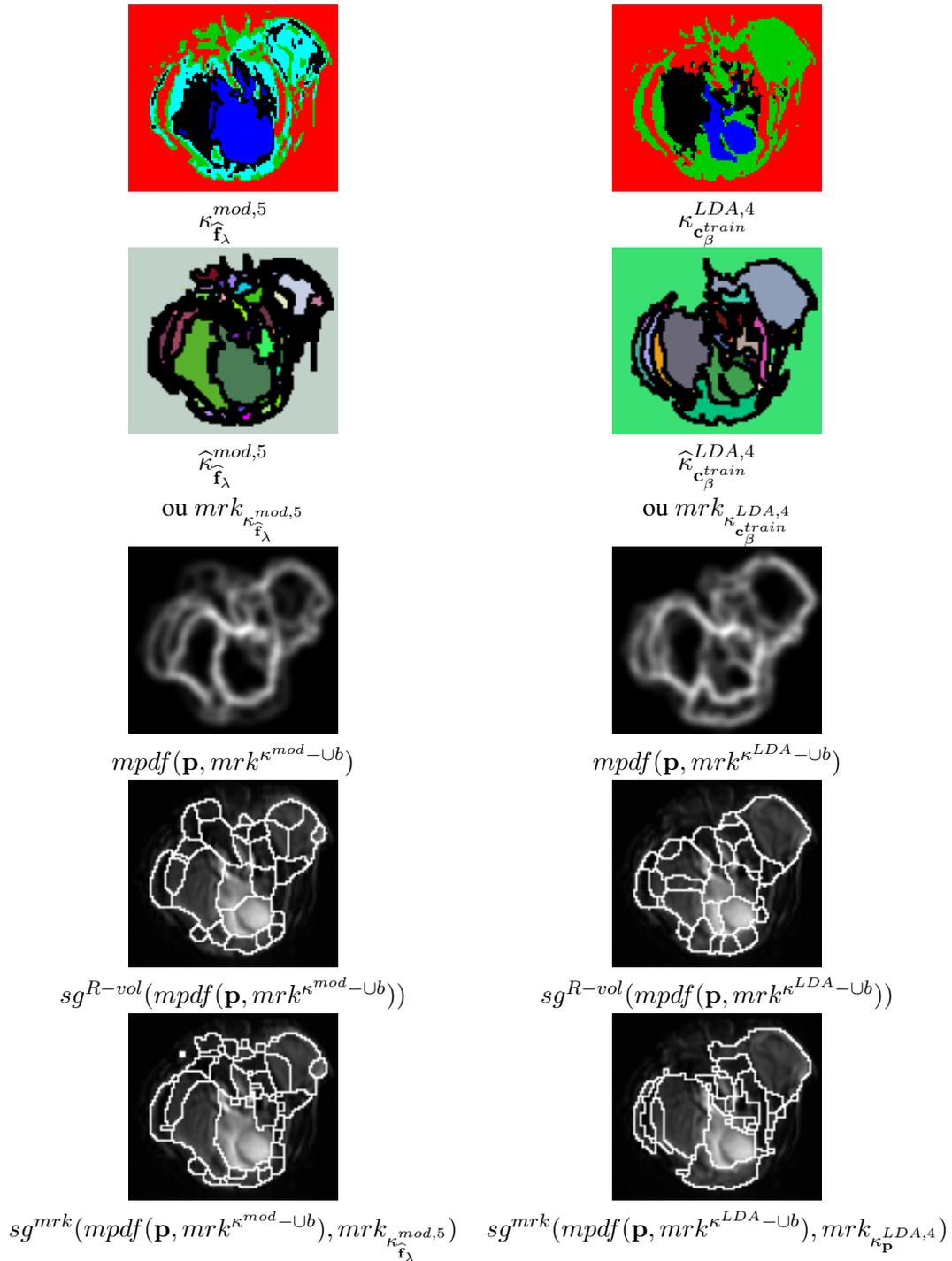


FIG. 3.33 – Segmentations par LPE probabiliste par critère volumique ou par marqueurs issus de la classification sur la $mpdf$ calculée dans l'espace des paramètres \mathbf{p} . On utilise comme germes de la $mpdf$ des marqueurs-boules aléatoires régionalisés avec union des boules dans une même classe connexe de la classification ($N = 100$ points, $M = 100$ réalisations, $R = 30$ régions, aire $S = 2$ pixels, $Rmax = 30$ pixels).

3.7.4.2.2 Union des germes boules connexes tombés dans une classe connexe On tire les centres des boules aléatoirement. On ne conserve que l'intersection entre la boule, $B(m_n, r)$ de centre m_n et de rayon r , et la classe connexe C_k . Cette intersection est notée $B_n^{C_k} = B(m_n, r) \cap C_k$. Lorsque plusieurs boules tombent dans une même classe C_k on fait l'union des marqueurs boules connexes $B_n^{C_k}$ de cette classe connexe $\cup_{n=1}^N B_n^{C_k}$ si $\cap_{n=1}^N B_n^{C_k} = \emptyset$. Pour ce faire il suffit de ré-étiqueter les classes connexes A_l de l'image marqueurs pour obtenir comme marqueurs $\cup_l A_l$ avec $\cap_l A_l \neq \emptyset$ (cf. algorithme 8). On note ces marqueurs $mrk_i^{\kappa-\cup b-connex}(x)$.

Algorithme 8 Marqueurs-boules aléatoires régionalisés avec union des boules connexes dans chaque classe connexe de la classification $mrk_i^{\kappa-\cup b-connex}(x)$

- 1: On fixe N le nombre de points m tirés et S l'aire minimale d'une composante connexe.
 - 2: **Pour tout** points m_n de 1 à N **Faire**
 - 3: Déterminer la composante connexe C_k telle que $m_n \in C_k$
 - 4: **Si** $aire(C_k) \geq S$ **Alors**
 - 5: $r = \mathcal{U}[1, Rmax]$
 - 6: $B_n^{C_k} = (B(m_n, r) \cap C_k) \cup B_{n-1}^{C_k}$
 - 7: **Fin Si**
 - 8: **Fin Pour**
 - 9: Étiqueter les classes connexes $\{A_l\}_l$ dans l'image marqueur
-

Comme précédemment, un marqueur contour est également ajouté lors des réalisations des marqueurs, $\{mrk_i^{\kappa-\cup b-connex}\}_{i=1\dots M}$, pour le calcul de la pdf. De plus les marqueurs aléatoires ne peuvent toucher la classe correspondant au fond de l'image.

En outre on réduit la taille des classes de la classification par filtrage morphologique (cf. paragraphe 3.7.4.2.1 et section 3.3.3, p : 85)

Après calcul des *mpdf* dans l'espace des paramètres \mathbf{p} à partir des germes $mrk_i^{\kappa-\cup b-connex}$, on les segmente par une LPE volumique (fig. 3.34). La segmentation dans l'espace des paramètres avec $R = 20$ régions à partir de la classification LDA semble être la meilleure.

Le calcul de la *mpdf* à partir de ces derniers germes, $mrk_i^{\kappa-\cup b-connex}(x)$, nous a semblé donner les meilleurs résultats.

Afin de bien comprendre comment sont utilisés les marqueurs-boules aléatoires régionalisés avec union des boules connexes dans chaque classe connexe de la classification $mrk_i^{\kappa-\cup b-connex}(x)$, on observe sur la figure 3.35 quelques réalisations de marqueurs et de contours nécessaires à la construction de la *mpdf*. Les marqueurs aléatoires sont conditionnés par la classification LDA filtrée $\hat{\kappa}_{c\beta}^{LDA,4}$ avec $N = 100$ points, $M = 100$ réalisations et $Rmax = 30$ pixels.

Après cette étude sur les marqueurs, nous allons segmenter d'autres séries temporelles de DCE-MRI de souris en vue d'une application d'aide au diagnostic pour la détection de tumeurs cancéreuses.

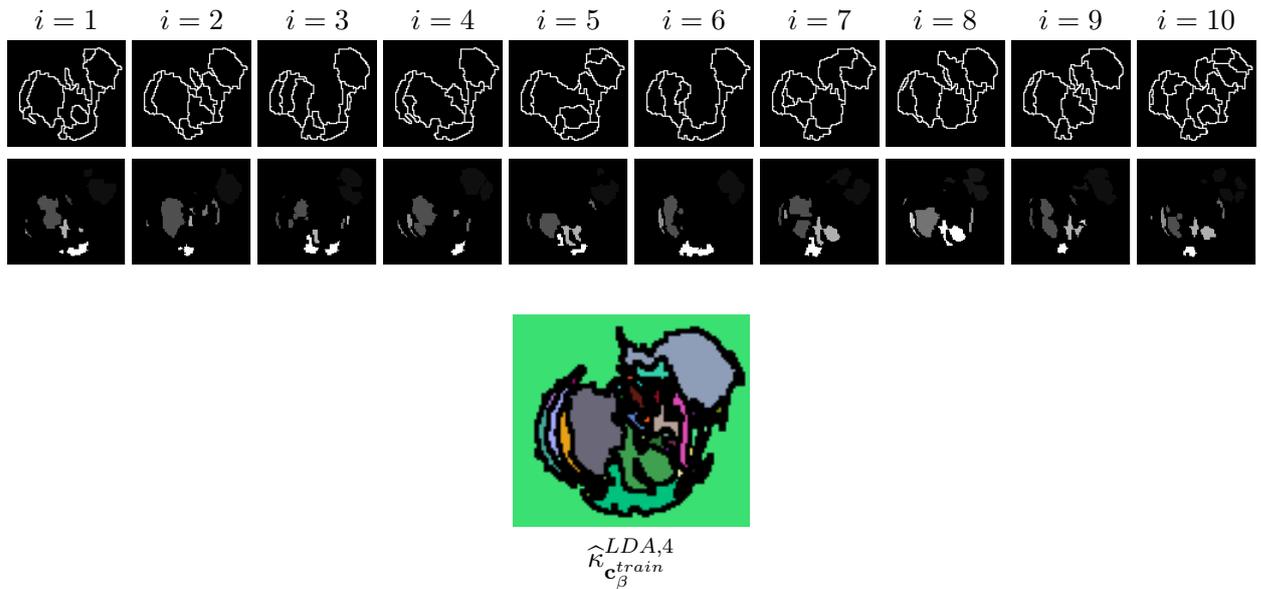


FIG. 3.35 – Première ligne, quelques réalisations des contours nécessaires (haut) à la construction de la $mpdf$ et, sur la deuxième ligne, les marqueurs $mrk_i^{\kappa-\cup b-\text{connex}}(x)$ régionalisés par classification LDA filtrée $\hat{\kappa}_{\beta}^{LDA,4}$ (avec $N = 100$ points, $M = 100$ réalisations et $Rmax = 30$ pixels).

3.8 Application à la détection de tumeurs cancéreuses

Une des applications très importantes de cette thèse, est la détection de tumeurs cancéreuses sur des séries temporelles DCE-MRI de souris. Nous disposons de 25 séries d'images fournies par l'équipe du Pr. C.-A. Cuenod du service imagerie de l'Hôpital Européen Georges Pompidou. Ces séries d'images de souris permettent de détecter des tumeurs par une cinétique croissante des spectres correspondant à l'accumulation du produit de contraste dans la zone cancéreuse (cf. fig. 2.27, p : 43). Cette application permet de montrer le potentiel de notre méthode de segmentation. Elle a été publiée dans [139]. Elle ne constitue néanmoins pas une étude exhaustive pour la détection de tumeurs. Pour ce faire, il faudrait une étude de validation complète sur des cas plus nombreux.

Pour pouvoir détecter des zones potentiellement tumorales nous avons choisi de segmenter ces images par LPE probabiliste, puis, de classifier les régions de la segmentation en zones potentiellement tumorales ou non tumorales. Le schéma complet de traitement des images est présenté en figure 3.36. Plusieurs parties de ce schéma de traitement ont été abordées précédemment :

- le pré-traitement : réduction de bruit par AFC (cf. section 2.4.5.1 , p : 41) et ajustement d'un modèle (cf. section 2.4.6 , p : 46) ;
- la classification par LDA après normalisation de l'histogramme des cartes de paramètres du modèle ajusté (cf. section 2.6.5 , p : 68) ;
- la segmentation par LPE probabiliste par des germes-boules aléatoires régionalisés avec union des boules connexes dans chaque classe connexe de la classification, $mrk_i^{\kappa-\cup b-\text{connex}}(x)$ (cf. section 3.7.4.2.2 , p : 126).

Dans cette partie, nous allons présenter la détection de tumeur et les cartes de confiance des zones

tumorales permettant au médecin de juger sur le fait que la région détectée soit ou non une tumeur.

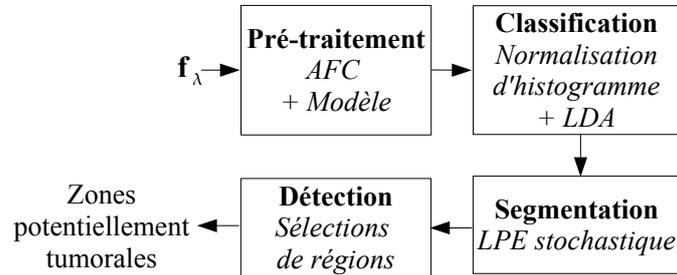


FIG. 3.36 – Schéma de traitement pour la détection de tumeurs.

On choisit de travailler à partir d'une segmentation par LPE probabiliste sur les paramètres des germes-boules aléatoires régionalisés $mrk_i^{\kappa_i - \cup b - connex}(x)$. Pour conditionner les marqueurs, on utilise la classification LDA dans l'espace des paramètres, $\kappa_p^{LDA,4}$. Cette classification nécessite un apprentissage sur une seule série. Après cet apprentissage, notre méthode est entièrement automatique.

La détection des régions potentiellement tumorales dépend de deux critères définis empiriquement. Ces régions doivent avoir :

- un paramètre moyen de pente a positif ;
- une ordonnée à l'origine moyenne b supérieure à 800 pour éliminer les zones correspondant au fond et qui peuvent avoir un paramètre ajusté de pente très légèrement positif. Cette valeur de seuil est la même pour toutes les séries d'images car les histogrammes des paramètres ont été normalisés.

Afin de permettre au médecin de juger de la pertinence des zones sélectionnées : on fournit des cartes de confiance sur les paramètres. Pour ce faire, sur chaque zone on calcule le coefficient de variation β_a de la pente et d'ordonnée à l'origine β_b :

$$\beta_a = \frac{\sigma_a}{E\{a\}} \quad \text{et} \quad \beta_b = \frac{\sigma_b}{E\{b\}} \quad (3.8.1)$$

avec σ l'écart type et E la moyenne du paramètre sur la zone considérée. Ces coefficients sont ensuite seuillés :

- pour la pente a à 5 ;
- pour l'ordonnée à l'origine b à 1.

Plus les coefficients de variation sont proche de zéro plus la confiance dans le paramètre est grande, plus il s'en éloigne plus la confiance est faible. Une échelle de visualisation est ensuite appliquée : le bleu correspond à la plus grande confiance $\beta = 0$ et le rouge à la plus faible confiance $\beta_a \geq 5$ ou $\beta_b \geq 1$.

La détection $det(x)$ et les cartes de confiance obtenues pour l'image de souris « serim447 » sont présentées sur la figure 3.37. Les résultats de segmentation $sg^{R-vol}(mpdf)$ et de détection $det(x)$ sont affichés sur le canal $\hat{f}_{\lambda_{512}}$ de l'image filtrée.

On constate que la grosse zone potentiellement tumorale pour laquelle on a une confiance élevée (en bleu) correspond bien à la tumeur sur l'image référence ref donnée par les médecins. En revanche,

la petite zone potentiellement tumorale pour laquelle on a une confiance faible (en rouge) n'est pas une tumeur sur l'image de référence. La méthode de détection fonctionne donc.

Nous avons appliqué cette méthode sur les 25 séries d'images. Nous présentons les résultats pour 6 images (fig. 3.38). On constate que les détections de zones correspondent bien aux références *ref* des médecins. Sur les images « serim450 » et « serim457 », des zones potentiellement tumorales avec un degré élevé de confiance sont détectées, alors qu'elle ne sont pas indiquées sur la référence.

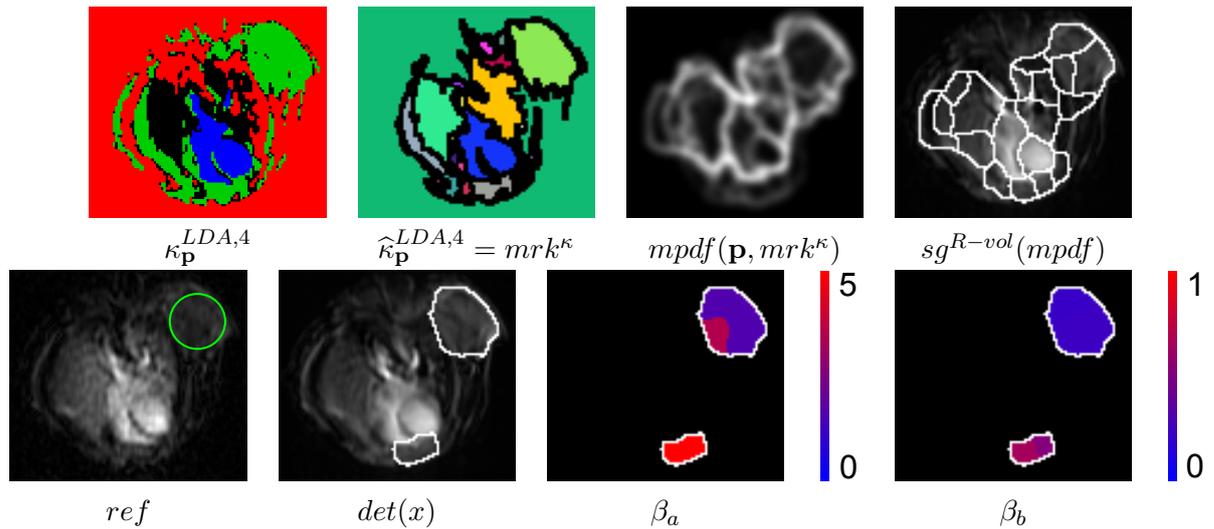


FIG. 3.37 – Classification LDA dans l'espace des paramètres $\hat{\kappa}_{\mathbf{p}}^{LDA,4}$, filtrage morphologique de la classification pour obtenir les marqueurs mrk^{κ} , $mpdf$ avec des germes-boules aléatoires régionalisés par la classification $mpdf(\mathbf{p}, mrk^{\kappa})$ ($N = 100$ points, $M = 100$ réalisations, aire $S = 10$ pixels, $Rmax = 30$ pixels), segmentation par LPE volumique associée en $R = 20$ régions $sg^{R-vol}(mpdf)$, référence *ref*, détection $det(x)$ des zones potentiellement tumorales ($moy(a) > 0$ et $moy(b) > 800$), cartes de confiance sur la pente β_a et sur l'ordonnée à l'origine β_b .

Pour quelques séries d'images, des segmentations anatomiques manuelles ont été réalisées par le Pr. C.A. Cuenod. Elles sont présentées sur la figure 3.39. Les couleurs des classes correspondent aux parties anatomiques de la façon suivante :

- classe verte : tumeurs ;
- classe bleue : cavité cardiaques (appelées parfois cœur) ;
- classe rouge : fond ;
- classe cyan : poumons.

On constate que notre méthode automatique a détecté, pour ces images, toutes les tumeurs. Afin d'améliorer les contours des zones potentiellement tumorales, il pourrait être intéressant de faire un apprentissage de la pdf sur les contours des classes de la segmentation manuelle afin de construire les pdf des autres images.

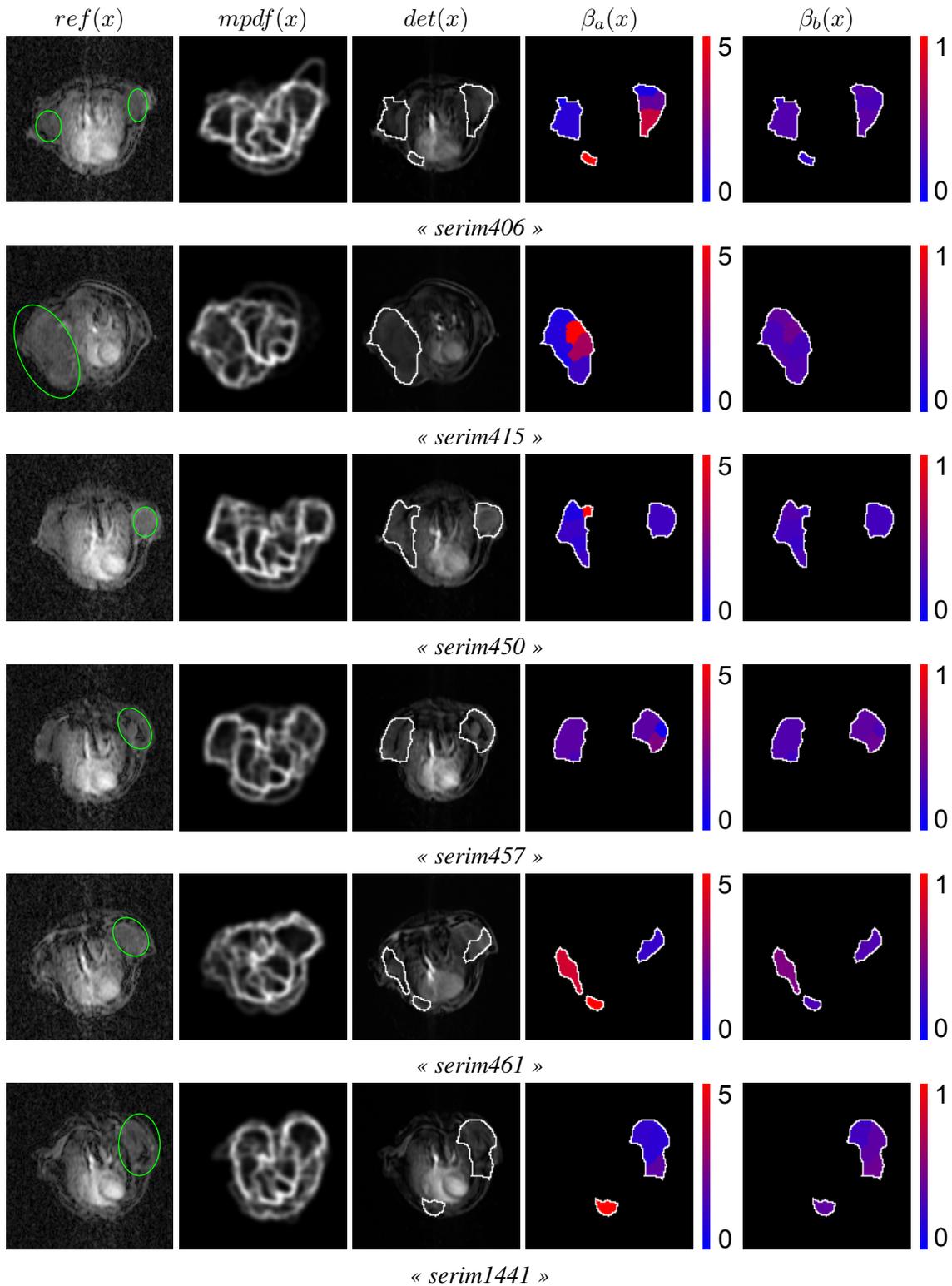


FIG. 3.38 – Référence des médecins ref , pdf marginale $mpdf(\mathbf{p}, mrk^{\kappa^{LDA}})(x)$ (germes-boules aléatoires régionalisés avec union des boules connexes, $N = 100$ points, $M = 100$ réalisations, aire $S = 10$ pixels, $Rmax = 30$ pixels), détection $det(x)$ des zones potentiellement tumorales dans l'espace des paramètres (recalés) ($moy(a) > 0$ et $moy(b) > 800$) et cartes de confiance sur la pente β_a et sur l'ordonnée à l'origine β_b .

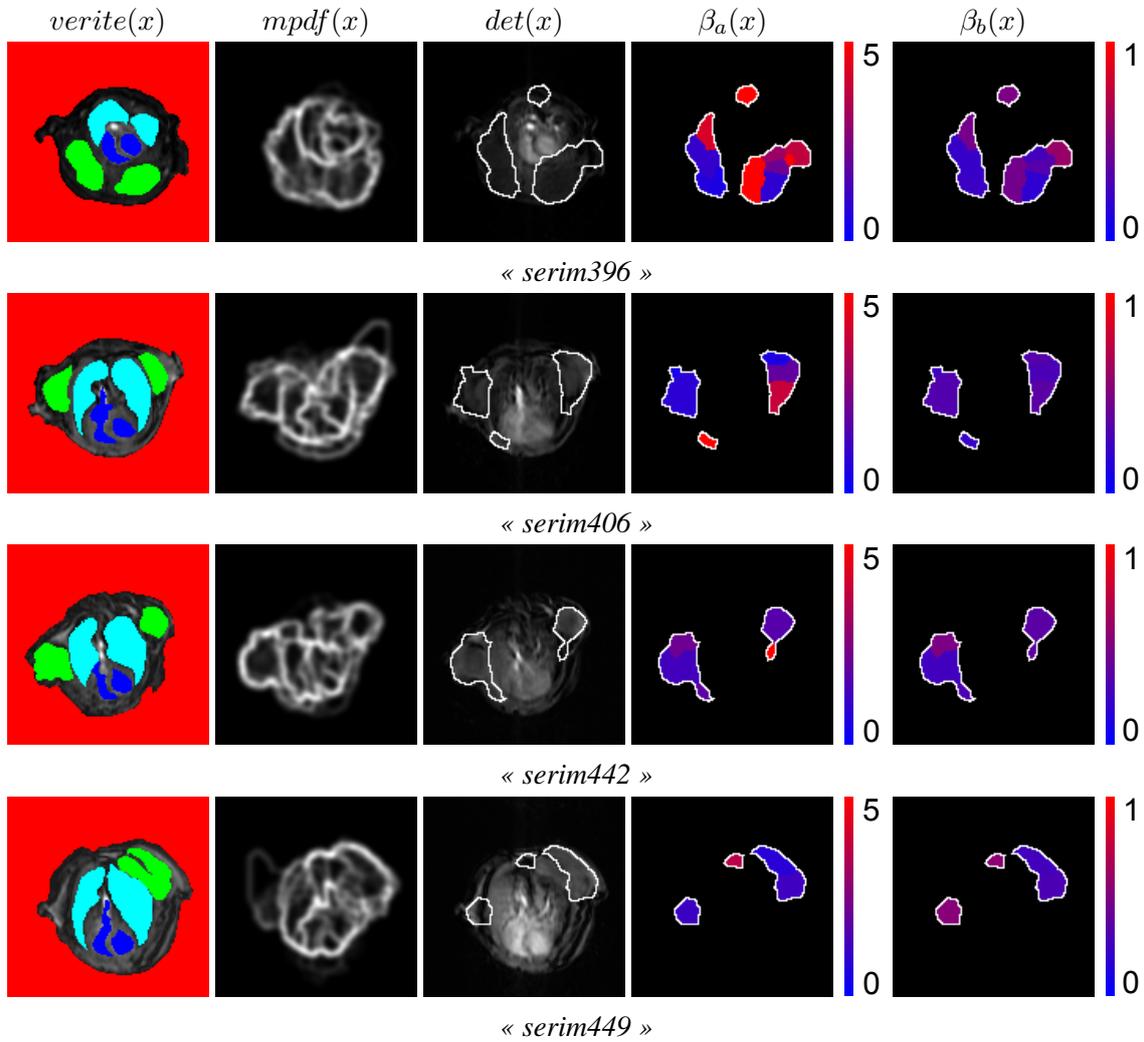


FIG. 3.39 – Segmentation anatomique manuelle des médecins *verite*, pdf marginale $mpdf(\mathbf{p}, mrk^{\kappa^{LDA}})(x)$ (germes-boules aléatoires régionalisés avec union des boules connexes, $N = 100$ points, $M = 100$ réalisations, aire $S = 10$ pixels, $Rmax = 30$ pixels), détection $det(x)$ des zones potentiellement tumorales dans l'espace des paramètres (recalés) ($moy(a) > 0$ et $moy(b) > 800$) et cartes de confiance sur la pente β_a et sur l'ordonnée à l'origine β_b .

3.9 Méthode générale de segmentation des images hyperspectrales

Introduisons maintenant une méthode générale de segmentation des images hyperspectrales. Les résultats sont présentés sur des images satellites. Cette méthode a été publiée dans [138].

Le diagramme de la méthode est présenté sur la figure 3.40. Les différentes étapes sont les suivantes :

- le filtrage spectral par AFC avec éventuellement une composition d’AFC-reconstitutions (cf. section 2.4.4, p : 35 et section 2.4.5, p : 41). L’AFC permet également d’effectuer une réduction de dimension (cf. section 2.4.1, p : 23) ;
- la précédente étape s’accompagne de la sélection des axes factoriels par le rapport signal à bruit des facteurs pixels (cf. section 2.4.1.1, p : 23) ;
- un modèle spectral peut éventuellement être ajusté sur l’image reconstituée $\hat{\mathbf{f}}_\lambda$ (cf. section 2.4.6, p : 46) ;
- une classification spectrale (cf. section 2.6, p : 57) ;
- un filtrage morphologique Υ de chacune des classes (cf. section 3.3, p : 82) ;
- le calcul d’une fonction de densité de probabilité (pdf) de contours marginale, ou vectorielle, $mpdf$, ou $vpdf$ respectivement, (cf. section 3.6, p : 109) à partir d’un gradient sur l’image et de germes-boules aléatoires régionalisés par la classification spectrale (cf. section 3.7.4.2.2, p : 126) ;
- une segmentation de la pdf par une LPE contrôlée par des marqueurs provenant des classes de la classification.

Nous allons illustrer notre démarche sur une image hyperspectrale satellite. Pour cet exemple on travaille dans l’espace image \mathbf{f}_λ à 5 dimension pour montrer la pertinence de notre méthode sur de nombreux canaux. Il pourrait être intéressant de travailler dans l’espace factoriel à 2 canaux ($c_{\alpha_1}^{\mathbf{f}}$, $c_{\alpha_3}^{\mathbf{f}}$).

Pour ce faire, on compare la LPE contrôlée par les marqueurs provenant de la classification κ sur une pdf probabiliste $mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}-\cup b-\text{connex}})$, calculée avec des germes-boules aléatoires avec union des boules connexes dans chaque classe connexe de la classification $mrk_i^{\kappa-\cup b-\text{connex}}(x)$, avec une LPE contrôlée par les mêmes marqueurs sur un gradient construit sur la distance du chi-deux d_{χ^2} (fig. 3.41). Les paramètres pour la LPE probabiliste avec des paramètres provenant de la classification filtrée, $\hat{\kappa}$, sont :

- le nombre de classe de la classification $\kappa : Q = 3$. C’est le seul paramètre important ;
- le nombre maximum de germes-boules aléatoires : $N = 50$ qui doit être suffisamment élevé et du même ordre de grandeur que le nombre de régions dans l’image ;
- le nombre de réalisations pour chaque canal : $M = 100$ qui a toujours la même valeur ;
- l’aire minimum pour les classes connexes de la classification $S = 10$ pixels. Ce paramètre est fonction de la taille de l’image ;
- du rayon maximal des boules $Rmax = 30$, qui est généralement le même.

On constate que les contours de l’approche probabiliste, $sg^{mrk}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}}), \hat{\kappa})$, sont plus lisses et suivent mieux les limites des régions que ceux de l’approche déterministe, $sg^{mrk}(\varrho_{\chi^2}(\mathbf{f}_\lambda), \hat{\kappa})$. Pour l’image « Port de Bouc » classifiés en $Q = 7$ classes (cf. section 3.3.4, p : 86), on observe les mêmes résultats (fig. 3.42).

Afin de comparer la segmentation par marqueurs avec la segmentation avec un nombre de régions ayant le plus gros volume, on choisit d’autres images « Roujan 0 2 » et « Roujan 1 9 » similaires à

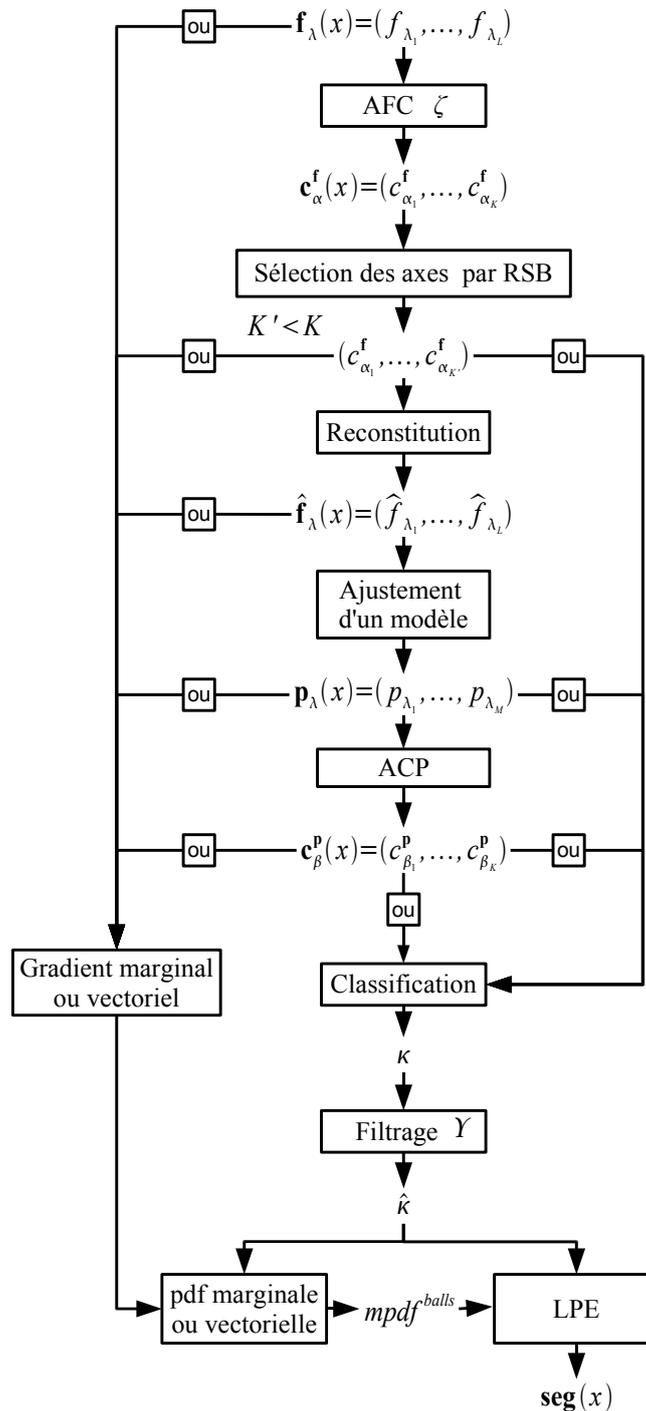


FIG. 3.40 – Schéma général de la LPE probabiliste contrôlée par une classification, avec des germes-boules aléatoires régionalisés.

l'image « Roujan », mais qui ont des nombres de régions différents. Comme pour l'image « Roujan », les facteurs pixels des axes factoriels 1 et 3 sont conservés ($c_{\alpha_1}^f, c_{\alpha_3}^f$). On effectue une classification

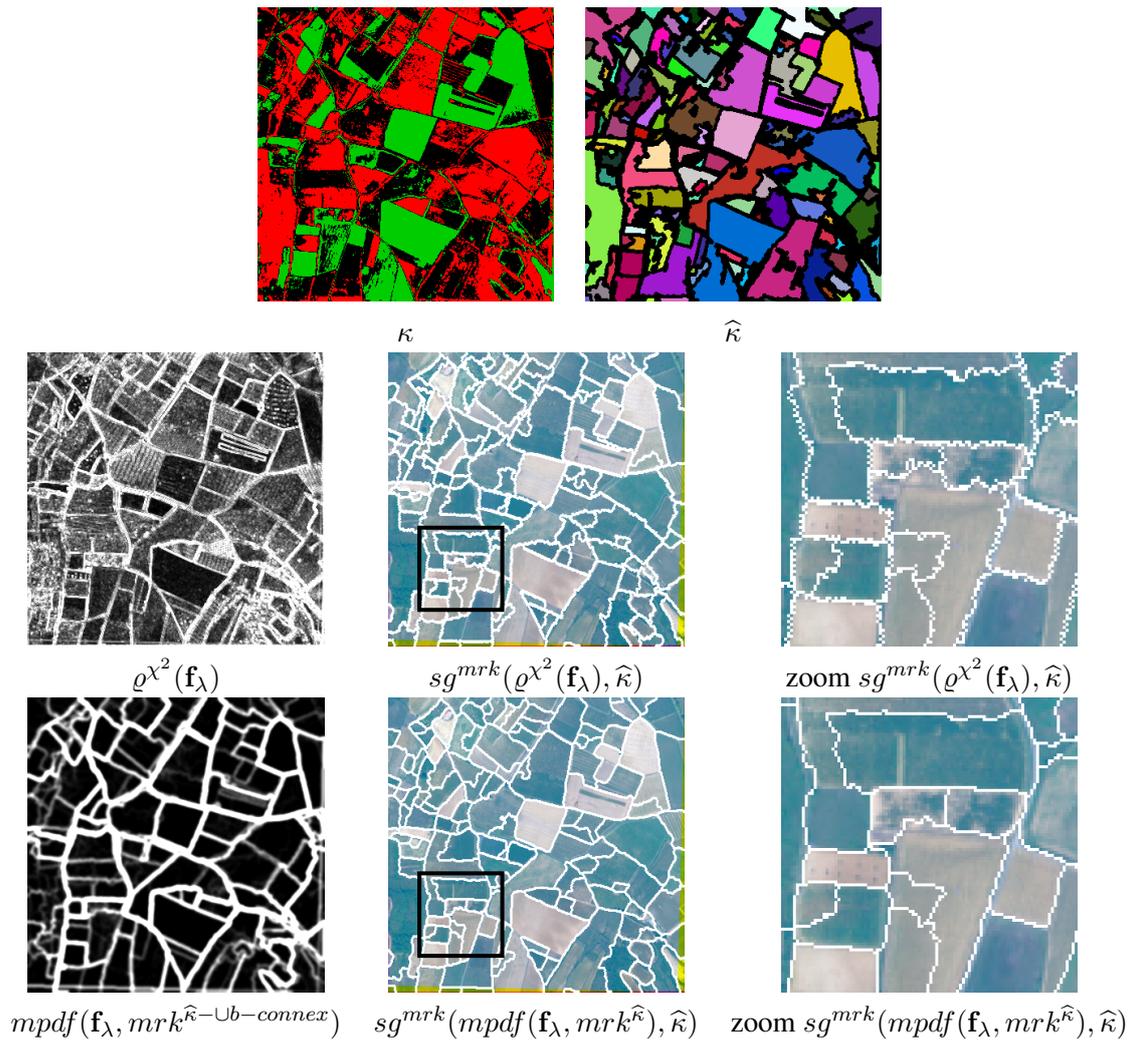


FIG. 3.41 – Classification en 3 classes et marqueurs provenant de la classification $\hat{\kappa}$ (première ligne), approche déterministe $sg^{mrk}(\rho^{\chi^2}(\mathbf{f}_\lambda), \hat{\kappa})$ sur le gradient construit sur la distance du chi-deux $\rho^{\chi^2}(\mathbf{f}_\lambda)$ (deuxième ligne) et approche probabiliste $sg^{mrk}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}}), \hat{\kappa})$ sur la $mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}-\cup b-connex})$ (troisième ligne) pour l'image « Roujan ». Dans les deux cas, les marqueurs de la LPE sur le gradient ou la $mpdf$ sont les classes de la classification filtrée $\hat{\kappa}$ ($Q = 3$ classes, $N = 50$ germes, $M = 100$ réalisations aire minimale $S = 10$ pixels, rayon maximum $Rmax = 30$ pixels). Pour la présentation des images sans zoom, les contours sont dilatés par un e.s. 3×3 pixels.

dans l'espace factoriel que l'on filtre ensuite (cf. section 3.3.4, p : 86).

On constate, sur la figure 3.43, que le nombre de régions résultant de la segmentation, dépend fortement de l'image, tandis que le nombre de classes est le même pour des images similaires. En fait le nombre de régions pour la LPE volumique, $sg^{R-vol}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}}))$, doit être choisi en fonction de l'image considérée, et ne peut être fixé a priori comme dans cet exemple où l'on a choisi $R = 50$ régions. En effet, le nombre de régions dépend fortement de la taille et de la complexité de l'image, tandis que le nombre de classes dépend du contenu spectral. En conséquence, la LPE par marqueurs

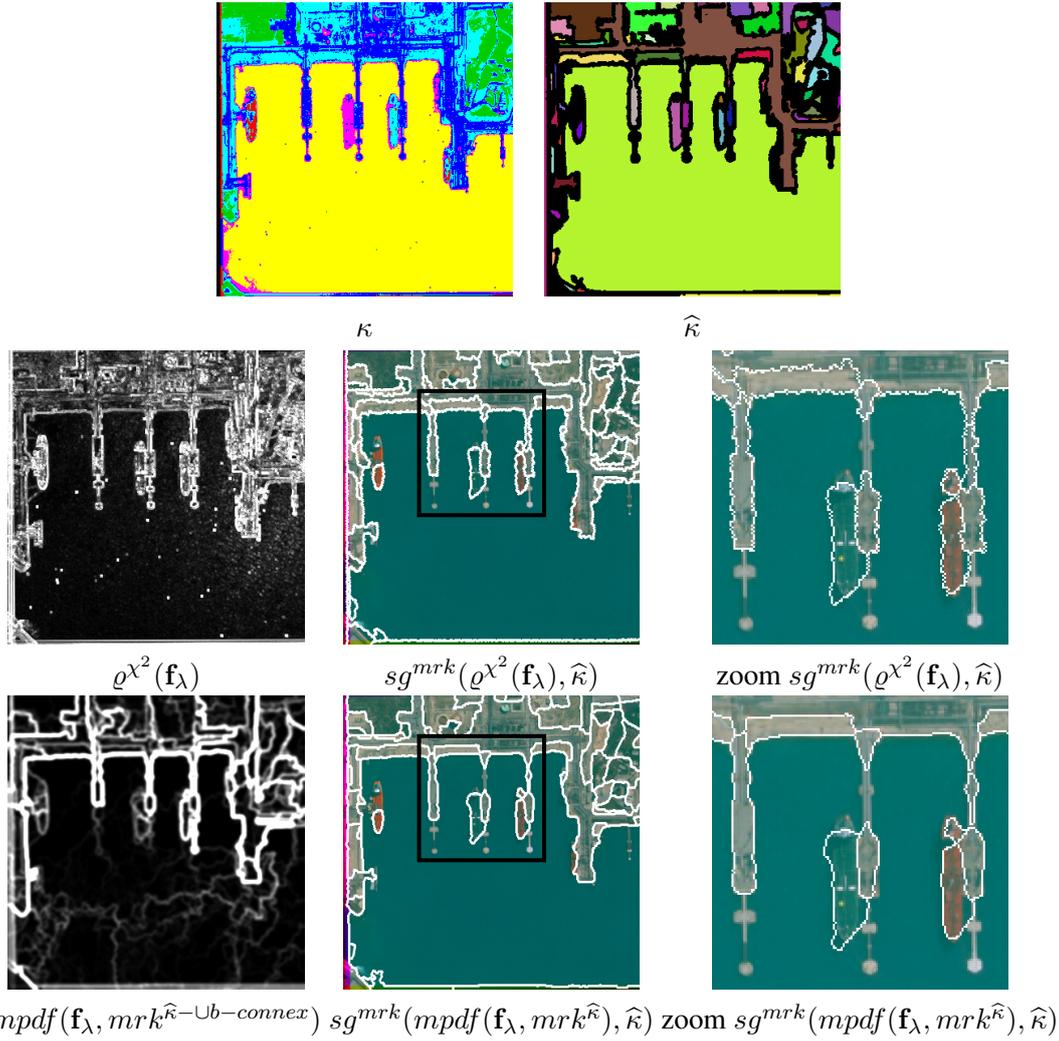


FIG. 3.42 – Classification en 7 classes et marqueurs provenant de la classification $\hat{\kappa}$ (première ligne), approche déterministe $sg^{mrk}(\varrho^{\chi^2}(\mathbf{f}_\lambda), \hat{\kappa})$ sur le gradient du chi-deux $\varrho^{\chi^2}(\mathbf{f}_\lambda)$ (deuxième ligne) et approche probabiliste $sg^{mrk}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}-\cup b-connex}), \hat{\kappa})$ sur la $mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}-\cup b-connex})$ (troisième ligne) pour l'image « Port de Bouc ». Dans les deux cas, les marqueurs de la LPE sur le gradient ou la $mpdf$ sont les classes de la classification filtrée $\hat{\kappa}$ ($Q = 7$ classes, $N = 50$ germes, $M = 100$ réalisations aire minimale $S = 10$ pixels, rayon maximum $Rmax = 30$ pixels). Pour la présentation des images sans zoom, les contours sont dilatés par un e.s. 3×3 pixels.

issus de la classification est plus pertinente que la LPE volumique, pour les images hyperspectrales. Un seul paramètre est requis pour la LPE par marqueurs : le nombre de classes de la classification. Ce paramètre est plus robuste que le nombre de régions puisqu'il est le même pour des images similaires. La LPE par marqueurs produit donc des segmentations plus robustes que la LPE volumique.



FIG. 3.43 – Comparaison entre les segmentations de $mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}-Ub-connex})$ par LPE probabiliste contrôlée par marqueurs issus de la classification $sg^{mrk}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}}), \hat{\kappa})$ ou par LPE probabiliste avec un nombre de régions, $R = 50$, défini a priori $sg^{R-vol}(mpdf(\mathbf{f}_\lambda, mrk^{\hat{\kappa}}))$. Les résultats sont présentés sur les images : « Roujan », « Roujan 0 2 », « Roujan 1 9 ».

3.10 Autres pistes de travail et perspectives

Dans cette partie, nous présentons des pistes de travail étudiées et les perspectives à venir concernant l'estimation spatio-spectrale des pdf de contours et les développements futurs de la morphologie mathématique multivariée.

3.10.1 Estimation des densités de probabilités de contours par noyaux spatio-spectraux

Dans le futur, il serait particulièrement intéressant de construire des pdf à partir de noyaux spatio-spectraux. On peut étendre l'expression de la pdf pour un canal (eq. 3.6.1, p : 110) à l'ensemble de l'image :

$$pdf(x) = \frac{1}{ML} \sum_{j=1}^L \sum_{i=1}^M (sg_i^{\lambda_j}(x) * G_{\sigma_x}) \times (sg_i^{\lambda_j}(x) * G_{\sigma_{\lambda_j}}) \quad (3.10.1)$$

avec $j = 1 \dots L$ l'indice des canaux, $i = 1 \dots M$ l'indice des réalisations, G_{σ_x} un noyau gaussien spatial d'écart type σ_x et $G_{\sigma_{\lambda_j}}$ un noyau gaussien spectral d'écart type σ_{λ_j} .

Les marqueurs aléatoires tirés pour une réalisation doivent être les mêmes pour tous les canaux afin d'obtenir des contours cohérents dans la dimension spectrale.

3.10.2 Morphologie mathématique sur les facteurs pixels

Nous avons effectué plusieurs essais concernant des opérations morphologiques sur les facteurs pixels de l'espace factoriel. Cette approche peut sembler intéressante car elle pourrait permettre de faire un filtrage morphologiques (spatial) sur les facteurs pixels en plus du filtrage spectral de l'AFC. L'application de filtres spatiaux sur les facteurs pixels peut être intéressante si l'on ne reconstitue pas l'image à partir des facteurs modifiés. Sinon, cette approche introduit de fausses couleurs lors de la reconstitution.

Afin de représenter l'impact des transformations dans l'espace image avant et après reconstitution avec modification des facteurs pixels, on crée une image composite couleur. Pour ce faire, les canaux retenus sont les suivants : $f_{\lambda_{61}}$ (700 nm) pour le rouge, $f_{\lambda_{30}}$ (545 nm) pour le vert et f_{λ_8} (435 nm) pour le bleu. Ces canaux sont affichés individuellement et sous forme d'une image composite couleur (fig. 3.44).

Observons les effets d'une ouverture ϕ_B , par une e.s carré 3×3 pixels dilaté 5 fois par lui même (i.e. de taille 5), sur les facteurs pixels c_α^f de l'image « woman face » (fig. 3.45). On constate que les facteurs pixels c_α^f sont très modifiés par cette opération morphologique.

Reconstituons l'image à partir de ces facteurs pixels (fig. 3.46). On constate que les canaux reconstitués $\widehat{f}^{\gamma_{\lambda_i}}$ sont peu modifiés par rapport aux facteurs pixels $\gamma_B c_{\alpha_i}^f$. La reconstitution par AFC nécessite en fait le produit des fréquences marginales qui elles restent inchangées par le filtrage. Ces fréquences marginales correspondent aux fréquences de l'image si les canaux étaient indépendants des variables. Elles constituent donc une sorte d'« image moyenne » (cf. annexe A.4.6, p : 215). On constate néanmoins l'apparition de fausses couleurs sur l'image synthétique après reconstitution RVB^γ .

Le filtrage des facteurs pixels c_α^f introduit de fausses couleurs dans l'image de synthèse couleur. Plusieurs raisons peuvent expliquer ces effets. Les facteurs pixels vérifient certaines propriétés (cf.



FIG. 3.44 – Facteurs pixels des canaux $f_{\lambda_{61}}$: 700 nm (rouge), $f_{\lambda_{30}}$: 545 nm (vert) et f_{λ_8} : 435 nm (bleu) et image couleur RVB associée.

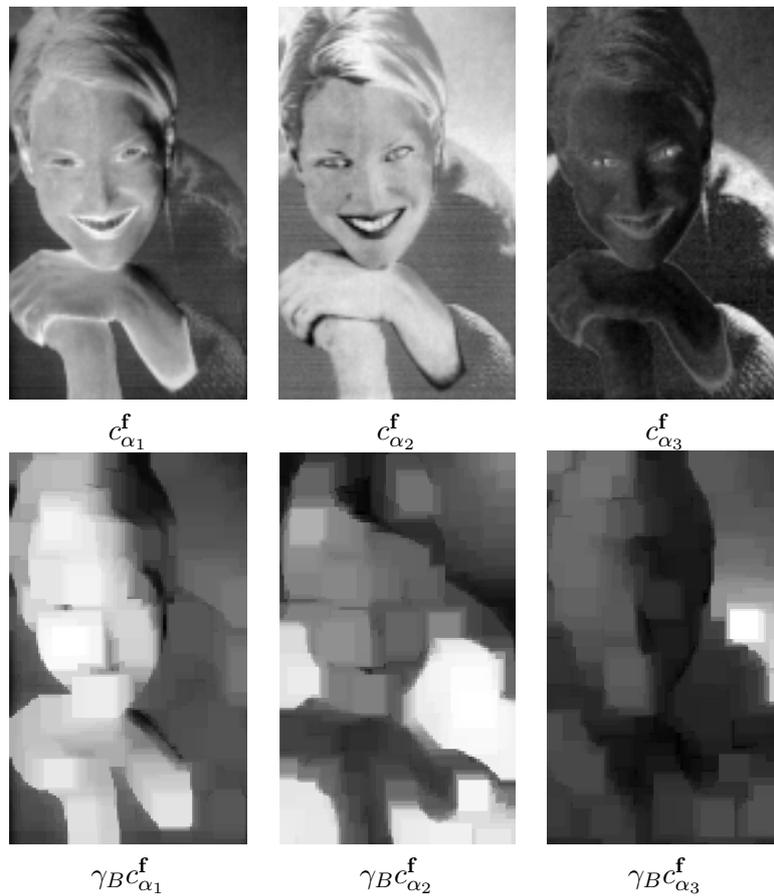


FIG. 3.45 – Facteurs pixels avant c_{α}^f et après ouverture $\gamma_B c_{\alpha}^f$ par un e.s. carré 3×3 pixels de taille 5.

section A.4.3.4, p : 211) qui ne sont plus respectées lors du filtrage morphologique. A savoir :



FIG. 3.46 – Canaux après reconstitution des facteurs pixels sur lesquels une ouverture morphologique avec e.s carré 3×3 pixels de taille 5 $\widehat{f}^{\gamma}_{\lambda_{61}}$, $\widehat{f}^{\gamma}_{\lambda_{30}}$ et $\widehat{f}^{\gamma}_{\lambda_8}$. Image couleur après reconstitution associée RVB^{γ} .

- Les codages sont centrés :

$$\frac{1}{P} \sum_{k=1}^P C^{\alpha}(k) = \frac{1}{P} \sum_{i=1}^P f_i \cdot c_i^{\alpha} = \sum_{i=1}^P \nu_i \cdot c_i^{\alpha} = 0$$

- Les variances des codages sont les valeurs propres μ_{α}

$$\frac{1}{P} \sum_{k=1}^P [C^{\alpha}(k)]^2 = \frac{1}{P} \sum_{i=1}^P f_i \cdot (c_i^{\alpha})^2 = \sum_{i=1}^P \nu_i \cdot (c_i^{\alpha})^2 = \mu_{\alpha}$$

- Les covariances entre codages différents sont nulles

$$\frac{1}{P} \sum_{k=1}^P C^{\alpha}(k) C^{\beta}(k) = \frac{1}{P} \sum_{i=1}^P f_i \cdot c_i^{\alpha} c_i^{\beta} = \sum_{i=1}^P \nu_i \cdot c_i^{\alpha} c_i^{\beta} = 0$$

En outre les formules de transition, liant les facteurs pixels aux facteurs canaux, ne sont pas vérifiées après modification des facteurs pixels (eq. 2.2.34, p : 16).

Comme ces différentes propriétés ne sont plus valables après filtrage morphologique, de fausses couleurs sont donc introduites. Nous avons également essayé de forcer les facteurs pixels modifiés à respecter certaines de ces propriétés. Cependant elles ne peuvent pas être toutes respectées à la fois. De fausses couleurs apparaissent donc toujours.

Si la morphologie mathématique sur les facteurs pixels ne fonctionne pas, la morphologie mathématique pour les images multivariées n'est pas encore établie.

3.10.3 Morphologie mathématique multivariée

Plusieurs essais ont été effectués pour développer une morphologie mathématique multivariée ou hyperspectrale. *Angulo* [3] définit la morphologie couleur sur des images multivariées à 3 composantes. *Aptoula et Lefèvre* [6] présentent un état de l'art exhaustif des méthodes pour faire de la

morphologie mathématique vectorielle, i.e. hyperspectrale. La difficulté consiste à trouver un ordre pour les vecteurs [64]. *Chanussot* [39] a également proposé un ordre de vecteurs par entrelacement de bits. *Li* ordonne les vecteurs selon les composantes principales [108]. Citons aussi des travaux très récents de *Köppen et Franke* sur la morphologie couleur [93]. *A. Plaza et al.* [147], quant à eux, proposent d'ordonner les vecteurs en utilisant leurs distances cumulées à tous les autres. Ils utilisent le fait que la morphologie mathématique puisse être définie en terme de distance.

En effet l'érosion ε_B et la dilatation δ_B sur un élément structurant B sont définies comme :

$$\begin{aligned}\varepsilon_B(f(x)) &= \{f(y) | f(y) = \wedge [f(z)], z \in B_x\} \\ \delta_B(f(x)) &= \{f(y) | f(y) = \vee [f(z)], z \in B_x\}\end{aligned}$$

En général, pour les images à niveaux de gris, le premier plan correspond aux pixels les plus clairs (proches de t_{max}) et le fond aux pixels sombres (proches de t_{min}).

Après avoir défini comme référence le niveau de gris maximum $g_0 = t_{max}$, la morphologie mathématique à niveau de gris peut être interprétée en terme de distance à cette référence [3] :

- la dilatation δ a tendance à rapprocher les niveaux de gris de la référence : c'est à dire que δ , au point x , est le niveau de gris qui a la distance minimale à la référence g_0 dans l'élément structurant B centré en x :

$$\delta_B(f(x)) = \{f(y) | y = \underset{z}{\operatorname{arg\,inf}}(f(z) - t_{max}), z \in B_x\};$$

- l'érosion ε a tendance à éloigner les niveaux de gris, de la référence g_0 : c'est à dire que ε est la valeur ayant la distance maximale à g_0 :

$$\varepsilon_B(f(x)) = \{f(y) | y = \underset{z}{\operatorname{arg\,sup}}(f(z) - t_{max}), z \in B_x\};$$

C'est l'unique convention que nous avons choisie. En effet, il est aussi possible de définir la dilatation (l'érosion) comme l'opération qui a tendance à éloigner (à rapprocher) de la référence g_0 , qui est définie dans ce cas comme $g_0 = t_{min}$. Dans ce cas la dilatation est définie comme :

$$\delta_B(f(x)) = \{f(y) | y = \underset{z}{\operatorname{arg\,sup}}(f(z) - t_{in}), z \in B_x\}$$

et l'érosion est définie par la relation duale.

La figure 3.47 (b) donne un exemple de dilatation et d'érosion à niveaux de gris en terme de distance à une valeur de référence $g_0 = t_{max}$. On remarque qu'il est nécessaire d'avoir une référence.

En fait, il serait intéressant de définir deux opérateurs à partir de la distance entre les niveaux de gris et qui soient aussi indépendants de la référence. Pour ce faire, il faut d'abord calculer la distance cumulée de chaque point à tous les autres à l'intérieur d'un élément structurant B_x :

$$D(f(x)) = \sum_k d_E(f(x), f(k)) \quad k \in B_x,$$

avec $d_E(f(x), f(k))$ la distance euclidienne $f(x)$ et $f(k)$.

Ensuite, la médiane calculée à partir d'une distance est définie par :

$$\mu_\varepsilon(f(x)) = f(y) | y = \underset{z}{\operatorname{arg\,inf}}[D(f(z))], z \in B_x\}.$$

Par dualité, l'anti-médiane calculée à partir d'une distance se définit comme :

$$\mu_\delta(f(x)) = \{f(y) | y = \operatorname{argsup}_z [D(f(z))], z \in B_x\}.$$

La figure 3.47 (c) montre un exemple de ces opérateurs « excentriques » qui ne donnent pas les mêmes valeurs qu'une dilatation δ ou une érosion ε classiques. Une généralisation de ces opérateurs, calculés à partir des médianes ou anti-médianes, aux images hyperspectrales conduit aux filtres par médiane vectorielle [11, 145]. Les travaux de A. Plaza *et al.* [147], proposent d'utiliser ces deux filtres pour définir des opérateurs morphologiques hyperspectraux. Cependant, il est facile de remarquer que les opérateurs obtenus ne sont pas des érosions, ou dilatations, par adjonction. En particulier, μ_ε et μ_δ ne commutent pas l'infimum et le supremum, respectivement ; c'est à dire une propriété sous-jacente à la définition de la dilatation et de l'érosion. Tout au plus, μ_ε et μ_δ peuvent être considérés comme des opérateurs pseudo-morphologiques. Nous pensons, en effet, que toute généralisation des opérateurs morphologiques aux fonctions multivariées (i.e. images hyperspectrales) doit être compatible avec la définition de ces opérateurs pour les fonctions univariées (i.e. les images à niveau de gris).

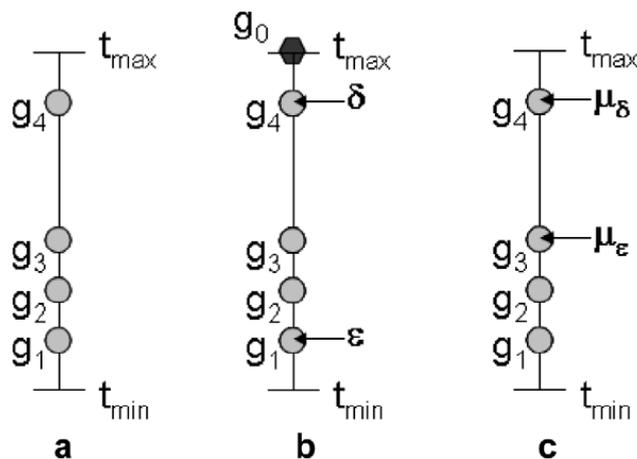


FIG. 3.47 – Exemple de dilatation et d'érosion à niveaux de gris en terme de distance à la valeur de référence : (a) quatre points scalaires de départ, $g_i \in \mathcal{T}$, (b) les valeurs correspondantes de dilatation δ et d'érosion ε quand la référence est $g_0 = t_{\max}$. En (c) sont données les valeurs de la médiane μ_ε et de l'anti-médiane μ_δ définies en terme de distance cumulée. (Source : J. Angulo [3]).

Même s'il existe un certain nombre d'opérateurs morphologiques multivariés, ceux-ci ne sont pas encore pertinents pour le traitement d'image (introduction de fausses couleurs, compatibilité avec les opérateurs à niveaux de gris, ...). Des opérateurs morphologiques multivariés pertinents restent donc encore à définir.

3.11 Conclusion du chapitre

Ce chapitre eu pour but de présenter les travaux effectués en segmentation d'images hyperspectrales par des méthodes morphologiques. Pour cela nous avons montré qu'une approche par Ligne de Partage des Eaux contrôlée par marqueurs provenant d'une classification spectrale était la méthode la plus adaptée pour segmenter ces images. En outre nous avons étendu aux images hyperspectrales une méthode récemment introduite : la Ligne de Partage des Eaux Probabiliste. Cette méthode consiste à construire une carte de probabilité de contours, pdf, par des simulations de Monte Carlo. Des germes aléatoires utilisés comme marqueurs pour une LPE sont tirés sur l'espace image. A l'aide d'un gradient sur l'image, ou sur ses canaux, et des germes aléatoires, la LPE permet d'obtenir des contours. La répétition des réalisations de contours permet d'obtenir la carte de probabilité de contour grâce à une estimation de la densité de probabilité de contours par une méthode à noyau. Les germes de la pdf peuvent être conditionnés par la classification spectrale, ce qui permet la construction d'une pdf contenant à la fois l'information spatiale et l'information spectrale. En outre, si les marqueurs sont des boules aléatoires les contours de la pdf sont encore plus précis.

Cette pdf est très pertinente pour la segmentation des images en un faible nombre de régions puisqu'elle privilégie les grandes régions fortement contrastées aux petites régions faiblement contrastées ou texturées mais apparaissant dans les gradients déterministes.

Une fois cette pdf obtenue, elle est segmentée par une LPE volumique ou contrôlée par des marqueurs provenant de la classification spectrale. Une méthode complète de segmentation des images hyperspectrales par LPE probabiliste a ainsi été introduite. Cette méthode repose sur un nombre de classes spectrales de la classification, qui est un critère plus robuste que le nombre de régions de l'image.

Au niveau des perspectives, deux points semblent particulièrement intéressants : d'une part faire de l'apprentissage de la pdf à partir de segmentations manuelles et d'autre part estimer la pdf à partir d'un noyau spatio-spectral.

Chapitre 4

Géodésie, connexions et images multivariées

La géodésie est une notion très utilisée en morphologie mathématique. Elle conduit à la définition d'une classe d'opérateurs géodésiques.

La connexion, quant à elle, est une notion de nature topologique qui conduit à des segmentations. Les connexions considérées dans le cadre de la théorie des treillis appliquée aux structures spatiales, sans prendre en compte de la topologie, constituent un cadre plus général mis en évidence par *Serra* [166] pour la définition des segmentations. La Ligne de Partage des Eaux qui est la méthode morphologique par excellence est elle même une connexion. Néanmoins il existe d'autres méthodes de segmentation morphologiques comme les zones plates ou quasi plates qui sont également des connexions. Après avoir présenté des notions introductives, nous proposons tout d'abord d'améliorer les zones quasi-plates qui ne prennent en compte que l'information locale par introduction d'information régionale. Ceci conduit à la définition de nouvelles zones : les régions η -bornées et les boules μ -géodésiques. Ces nouvelles segmentations sont extensibles aux images hyperspectrales par l'utilisation de distances entre spectres. On aboutit ainsi à la création d'autres méthodes de segmentation spatio-spectrales.

Les boules μ -géodésiques comme leur nom l'indique nécessitent l'utilisation de distances géodésiques.

Un autre point très important est le calcul de toutes les paires de distances géodésiques entre les pixels d'une image. Ce calcul est très coûteux en temps de calcul et en terme d'algorithme. Néanmoins un tel tableau de paires de distances géodésique pourrait être très utile pour la réduction de données des images hyperspectrales par tableau de masses et distances [21] ou par des méthodes comme ISOMAP [177, 178] utilisant le calcul de paires de distances géodésiques. Ce tableau est également utilisé pour la segmentation par distance géodésique à partir de coupes de graphes, e.g. [175].

Il est tout d'abord nécessaire de préciser quelques notions introductives.

4.1 Notions introductives

Des définitions et des propriétés plus complètes sur les notions de segmentation et de connexité selon *Serra* [166] sont détaillées en annexes B.2 (p : 235) et B.3 (p : 238).

4.1.1 La segmentation par critère

Définition 3 (Critère). Soient E et T deux ensembles arbitraires et soit \mathcal{F} une famille de fonctions de E dans T . Un critère σ sur la classe \mathcal{F} est une fonction binaire de $\mathcal{F} \otimes \mathcal{P}(E)$ dans $\{0, 1\}$ telle que, pour chaque fonction $f \in \mathcal{F}$, et pour chaque ensemble $A \in \mathcal{P}(E)$, l'on ait :

ou bien $\sigma[f, A] = 1$ (le critère est satisfait sur A)

ou bien $\sigma[f, A] = 0$ (le critère est refusé sur A).

De plus, de manière conventionnelle, on décide que, pour toutes les fonctions, tous les critères sont satisfaits sur l'ensemble vide, i.e.

$$\sigma[f, \emptyset] = 1 \quad \forall f \in \mathcal{F}$$

Définition 4 (Critère connectif). Un critère $\sigma : \mathcal{F} \otimes \mathcal{P}(E) \rightarrow [0, 1]$ est connectif :

- quand il est vérifié sur la classe \mathcal{S} des singletons

$$\forall f \in \mathcal{F}, \{x\} \in \mathcal{S} \Rightarrow \sigma[f, \{x\}] = 1$$

- et quand pour toute fonction $f \in \mathcal{F}$ et pour toute famille $\{A_i\}$ dans $\mathcal{P}(E)$ on a

$$\cap A_i \neq \emptyset \text{ et } \wedge \sigma[f, A_i] = 1 \Rightarrow \sigma[f, \cup A_i] = 1$$

Autrement dit, si le critère est vérifié par la fonction f sur deux ensembles A et B de l'espace, et que A et B ont un point en commun, alors il est aussi vérifié sur réunion $A \cup B$. Les critères connectifs caractérisent les processus de segmentation comme le montre le théorème suivant :

Théorème 4.1.1 (Segmentation - *J. Serra*). *Considérons deux ensembles arbitraires E et \mathcal{T} , et une famille \mathcal{F} de fonctions $f : E \rightarrow \mathcal{T}$. Soit d'autre part σ un critère sur la classe \mathcal{F} . Alors les trois propositions suivantes sont équivalentes :*

1. le critère σ est connectif;
2. pour chaque fonction $f \in \mathcal{F}$, les ensembles sur lesquels le critère σ est satisfait génèrent une connexion \mathcal{C} :

$$\mathcal{C} = \{A | A \in \mathcal{P}(E) \text{ et } \sigma[f, A] = 1\};$$

3. le critère σ segmente toutes les fonctions de la famille \mathcal{F} .

Le théorème 4.1.1 relie la segmentation à la propriété de critère connectif qui est bien plus manipulable. Il est remarquable, que l'on ait pu identifier la notion de segmentation à certaines familles de connexions sans avoir eu à doter l'espace de départ E , ni celui d'arrivé \mathcal{T} d'aucune propriété. Aussi le théorème ouvre la voie à toutes les applications où des variables hétérogènes sont définies sur un même espace. Ces circonstances se produisent par exemple :

- en imagerie couleur, avec la teinte,

- en géographie, où des données radiométriques (satellites) cohabitent avec des variables physiques (altitude, pente du sol, ensoleillement, distance à la mer, ...) et avec des données statistiques (démographies, fortunes, épidémies, ...) [167]. Parmi cet ensemble de données se trouvent les images multivariées dont on a supposé les canaux définis sur le même espace support E .

Intéressons nous maintenant aux segmentations définies par des connexions.

4.1.2 Les segmentations définies par des connexions

Définissons tout d'abord ce qu'est une connexion.

4.1.2.1 Connexion et partitions

Définition 5 (Connexion). Soit E un ensemble arbitraire non vide. On appelle classe connexe ou connexion \mathcal{C} toute famille de $\mathcal{P}(E)$ telle que :

- (0) $\emptyset \in \mathcal{C}$
- (i) quel que soit $x \in E$, $\{x\} \in \mathcal{C}$,
- (ii) pour toute famille $\{C_i, i \in I\}$ de \mathcal{C} , $\cap_i C_i \neq \emptyset$ entraîne $\cup_i C_i \in \mathcal{C}$.

où $\{x\}$ désigne un singleton, i.e. un élément ponctuel de $\mathcal{P}(E)$. Tout ensemble C d'une classe connexe \mathcal{C} est dit connexe

Les connexions induisent des partitions et donc des segmentations.

Définissons ce qu'est une partition

Définition 6 (Partition). Soit E un ensemble arbitraire. On appelle partition de l'espace E toute application $D : E \rightarrow \mathcal{P}(E)$ telle que :

- (i) $\forall x \in E : x \in D(x)$ et
- (ii) $\forall x, y \in E : D(x) = D(y)$ ou $D(x) \cap D(y) = \emptyset$.

$D(x)$ se nomme la classe de la partition qui contient x .

L'ensemble des partitions d'un ensemble arbitraire E est ordonné comme suit.

Définition 7 (Ordre de partitions). Une partition \mathcal{A} est dite plus fine (respectivement plus grossière) qu'une partition \mathcal{B} , $\mathcal{A} \leq \mathcal{B}$ (respectivement $\mathcal{A} \geq \mathcal{B}$), quand chaque classe de \mathcal{A} est incluse dans une classe de \mathcal{B} .

Ceci conduit à la notion de hiérarchie ordonnée de partitions $\Pi_{i=1}^N D_i$, telle que $D_i \leq D_{i+1}$ et même de treillis complet (cf. annexe B.3, p : 238) [166].

Après avoir défini la notion générale de connexion, intéressons nous à un cas particulier, celui des connexions par sauts qui conduisent à des segmentations.

4.1.2.2 Segmentation par connexion par sauts

La connexion par sauts est une méthode de segmentation par croissance de régions proposée par *J. Serra* [161, 163], et qui consiste en des variations bornées d'amplitude, en fixant les origines des variations aux minima ou aux maxima de la fonction f . On remarque ici l'intérêt des extrema pour marquer les structures remarquables de l'image.

Définition 8 (Connexion par sauts). Soit m un minimum de la fonction f . Pour une variation d'amplitude k donnée, considérons la composante connexe qui contient m et dont les points z accusent un dénivelé $f(z) - m$ au dessus de m tel que $0 < f(z) - m \leq k$,

$$A(m) = \{z : z \in E, 0 < f(z) - m \leq k\}.$$

Soit $D_{k,m}$ la partition composée de $A(m)$ plus les classes ponctuelles sur $E \setminus A(m)$. On prend le supremum D_k de toutes les partitions $D_{k,m}$ associées à tous les minima de la fonction. Puis on peut itérer le processus sur l'ensemble $E \setminus A(m)$, i.e. zones de dénivelé $k < f(z) - m \leq 2k$, puis $2k < f(z) - m \leq 3k, \dots$

La segmentation selon ce critère connectif, noté σ_k^{JC} , engendre la connexion par sauts. La procédure alternative à partir des maxima se construit par dualité, et l'on peut combiner les deux dans une démarche symétrique, où la remontée s'arrête quand elle rencontre une descente, et où les zones à cheval entre montée et descente sont subdivisées.

Étant donné qu'elle dépend d'un paramètre croissant k , la connexion par sauts engendre une pyramide. Par ailleurs, elle est aussi une granulométrie lorsque l'on restreint les valeurs du paramètre aux progressions géométriques [165].

En pratique la segmentation par sauts s'avère une bonne méthode de partition : peu de zones ponctuelles et de classes visuellement significatives (fig. 4.1).

La segmentation par sauts fait partie des méthodes plus générales de segmentation à partir de germes comme notamment :

- les méthodes de croissance de régions comme l'algorithme des Nuées Dynamiques [51] encore appelé *kmeans* [75] ;
- la Ligne de Partage des Eaux et les segmentations par « snakes » [89].

4.1.3 Les zones plates et quasi-plates

Les zones plates d'une fonction f sont les composantes connexes de f ayant une valeur constante.

La notion de zones plates d'une image est due à *Salembier et Serra* [155, 156] et a été très utilisée par *Crespo* [46]. Elle constitue un critère connectif de segmentation.

La définition des zones plates nécessite la notion de chemin.

Définition 9 (Chemin). Un chemin C de longueur $l(C) = n + 1$ ayant comme extrémités les points p et q (ou $C_{p,q}$) dans la trame V_G (de voisinage associé N_G) est un $(n + 1)$ -uplet (p_0, p_1, \dots, p_n) de pixels tels que :

$$\begin{aligned} p_0 &= p \text{ et } p_n = q \\ \forall i \in [1, n], p_i &\in N_G(p_{i-1}). \end{aligned}$$

Les zones plates sont définies de la manière suivante :

Définition 10 (Zones plates). Deux points $x, y \in E$ appartiennent à la même zone plate d'une fonction f si et seulement s'il existe un chemin de points (p_0, p_1, \dots, p_n) tel que $p_0 = x$ et $p_n = y$ et, $\forall i, (p_i, p_{i+1}) \in E^2$ sont voisins et $f(p_i) = f(p_{i+1})$.

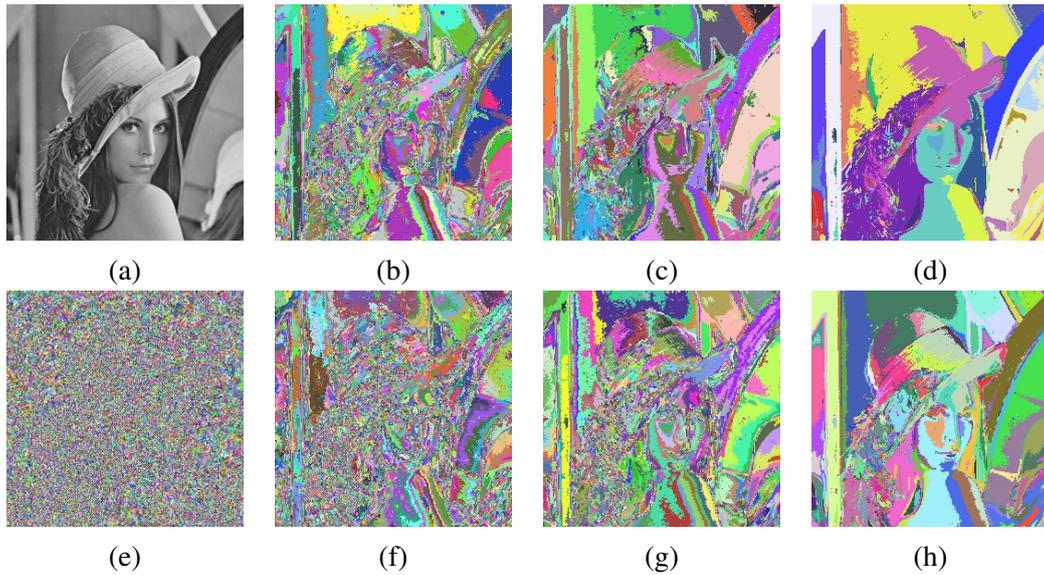


FIG. 4.1 – Pyramides de segmentation paramétrique de l’image « lenna » : (a) image initiale, (b) à (d) connexion par sauts de paramètres $k = 5, 10$ et 30 respectivement, (e) zones plates, (f) à (h) zones quasi-plates de paramètre $\lambda = 5, 10$ et 30 respectivement.

Si on obtient les zones plates d’une image brute, la partition obtenue n’est guère intéressante à cause d’une sur-segmentation. Il existe quelques alternatives à suivre. Par exemple, filtrer préalablement l’image f pour obtenir une simplification des détails. Pour ce faire, on doit utiliser un filtre connexe, typiquement un nivellement dont l’une des propriétés principales est l’élargissement des zones plates. Si on considère maintenant une séquence de nivellements avec des marqueurs de plus en plus forts (une pyramide de nivellements), les zones plates associées à la famille de fonctions successives constituent une hiérarchie ou pyramide. Pour de plus amples détails sur les nivellements, se référer à l’annexe B.4 (p : 240).

Une autre possibilité, étudiée par *Angulo* [2], est la fusion de zones plates selon un critère donné. En outre on peut sélectionner des sous-ensembles de zones plates qui sont des marqueurs pour la LPE [47].

La notion de zones quasi-plates a été introduite par *Meyer* [119, 120] en élargissant la définition des zones plates, ce qui conduit à des partitions beaucoup plus significatives.

Définition 11 (Zones quasi-plates ou zones λ -plates). Deux points $x, y \in E$ appartiennent à la même zone quasi-plate d’une image hyperspectrale f_λ si et seulement s’il existe un chemin $(p_0, p_1, \dots, p_n) \in E^n$ tel que $p_0 = x$ and $p_n = y$ et, si, pour tout i , $(p_i, p_{i+1}) \in E^2$ sont voisins et $f(p_i) \approx f(p_{i+1})$.

Des zones quasi-plates avec des caractéristiques différentes sont obtenues selon la définition de la relation \approx . Nous allons utiliser la plus directe qui pour les images à niveaux de gris s’écrit : $f(p) \approx f(q) \Leftrightarrow f(q) - \lambda < f(p) < f(q) + \lambda$, avec $\lambda \in \mathbb{R}^+$. C’est à dire qu’un ensemble de pixels est considéré comme appartenant à la même zone quasi-plate de paramètre λ (ou zone λ plate) si pour toute paire de pixels il existe un chemin tel que l’écart entre deux pixels adjacents est inférieur à λ . Pour $\lambda = 0$, on retrouve les zones plates. Cette connexion est notée σ_λ^{QFZ} .

A nouveau, le paramètre croissant λ engendre une hiérarchie ou pyramide de partitions. La figure 4.1 illustre quatre niveaux d'une telle hiérarchie. On trouve une certaine correspondance dans les objets mis en évidence par la segmentation par sauts et les zones quasi-plates.

Les zones plates, et leur généralisation, les zones quasi-plates ou zones λ -plates, initialement introduites pour des fonctions scalaires (i.e des images à niveau de gris), ont été généralisées aux images couleurs, et multivariées, par *Zanoguera et Meyer* [191] grâce à la distance spectrale :

Définition 12 (Distance spectrale). Une distance spectrale est une fonction $d : \mathcal{T}^L \times \mathcal{T}^L \rightarrow \mathbb{R}^+$ qui vérifie, pour $s_i, s_j, s_k \in \mathcal{T}^L$, les propriétés suivantes :

1. $d(s_i, s_j) \geq 0$ (positivité) ;
2. $d(s_i, s_j) = 0 \Leftrightarrow s_i = s_j$ (séparation ou identité) ;
3. $d(s_i, s_j) = d(s_j, s_i)$ (symétrie) ;
4. $d(s_i, s_j) \leq d(s_i, s_k) + d(s_k, s_j)$ (inégalité triangulaire).

La relation \approx s'écrit alors pour les images multivariées : $\mathbf{f}_\lambda(p) \approx \mathbf{f}_\lambda(q) \Leftrightarrow d(\mathbf{f}_\lambda(p), \mathbf{f}_\lambda(q)) < \lambda$, avec $\lambda \in \mathbb{R}^+$.

Pour le choix d'une distance spectrale adaptée à l'espace de représentation de l'image, on peut se reporter à la section 3.4.2 (p : 91). De nouveau, le paramètre croissant λ engendre une hiérarchie ou pyramide de partitions (cf. annexe B.2, p : 235).

4.1.4 Notions géodésiques

La notion de géodésie est d'une importance capitale en morphologie mathématique, puisqu'actuellement tous les opérateurs les plus évolués et les plus puissants se fondent sur la géodésie.

Nous verrons comment la distance géodésique est définie pour des images binaires et pour des images à niveau de gris. Nous montrerons ensuite comment l'étendre aux images hyperspectrales.

4.1.4.1 Géodésie pour des images binaires

Le point d'entrée de la géodésie est la notion de distance géodésique restreinte à un ensemble de référence X .

Définition 13 (Distance géodésique binaire). Soit X un ensemble de \mathbb{Z}^2 (binaire). La distance géodésique entre deux points p et q à l'intérieur de l'ensemble X ($p, q \in X$) est la longueur du plus court chemin $C_{p,q}$ entre p et q restant à l'intérieur de X , i.e.

$$d_X^g(p, q) = \inf\{l(C_{p,q})\}.$$

On considère par définition que cette distance vaut $+\infty$ s'il n'existe aucun chemin entre p et q à l'intérieur de X : i.e. si X n'est pas simplement connexe.

L'ensemble X est appelé masque géodésique (binaire). Il peut exister plusieurs chemins reliant p à q dans X et qui peuvent avoir la longueur minimale $d_X^g(p, q)$. Les chemins les plus courts sont appelés géodésiques. Un chemin géodésique est illustré sur la figure 4.2.

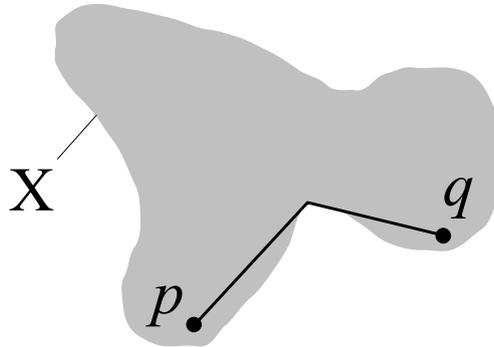


FIG. 4.2 – Géodésique entre p et q à l'intérieur du masque X .

La distance géodésique satisfait aux quatre axiomes d'une distance (cf. définition 12), mais elle a ses valeurs dans $\mathbb{R}^+ \cup \infty$ plutôt que dans \mathbb{R}^+ . On l'appelle donc une distance généralisée.

La boule géodésique de rayon $n \in \mathbb{Z}^+$ et centrée au point $p \in X$ est l'ensemble $B_X(p, n)$ défini par : $B_X(p, n) = \{p' \in X, d_X^g(p', p) \leq n\}$.

On peut interpréter cette définition de manière intuitive : lorsque n augmente, les boules progressent comme le front d'une onde émise depuis p dans l'environnement de X .

Remarque 4.1.1. Il existe également la notion de distance géodésique d'un point p à un ensemble Y , avec $p \in X$ et $Y \subset X$, nous invitons le lecteur intéressé à se reporter au livre de *Soille* [172].

Plusieurs opérateurs ont été établis à partir de cette distance grâce aux travaux de *Lantuejoul et Beucher* [98, 99].

4.1.4.2 Géodésie généralisée aux images à niveau de gris

Nous utilisons ici l'approche présentée par *Soille* dans [172].

Les opérateurs géodésiques avec des masques géodésiques binaires peuvent être généralisés aux opérateurs avec des masques géodésiques à niveau de gris par *Soille* [173]. Les propagations géodésiques sont alors contrôlées par les valeurs du masque à niveau de gris.

Plusieurs algorithmes de calcul des chemins géodésiques existent, parmi eux citons celui de *Soille* [171, 172] et celui des « Fast Marching » initié par *Sethian* [169, 42, 170].

Considérons un chemin \mathcal{P} défini sur le domaine de définition E d'une image à niveau de gris ou d'une fonction positive intégrable f . Le temps nécessaire pour parcourir \mathcal{P} est défini comme l'intégrale des valeurs de f le long de \mathcal{P} et noté $\tau_f(\mathcal{P})$:

$$\tau_f(\mathcal{P}) = \int_{\mathcal{P}} f(s) ds. \quad (4.1.1)$$

Remarquons que l'appellation « temps » se justifie en considérant les unités comme l'inverse d'une vitesse qui serait mesurée par les niveaux de gris d'une image f : plus le niveau de gris est

faible, plus la vitesse est élevée. Une définition adaptée aux images à niveau de gris est maintenant donnée :

Définition 14 (Temps géodésique pour les images à niveau de gris). Le temps $\tau_f(\mathcal{P})$ nécessaire pour parcourir un chemin discret \mathcal{P} de longueur l défini sur une image discrète à niveaux de gris (positifs) est égal à la moyenne des valeurs de f prises à un instant donné le long de \mathcal{P} :

$$\tau_f(\mathcal{P}) = \sum_{i=1}^l \frac{f(p_{i-1}) + f(p_i)}{2} = \frac{f(p_0)}{2} + \frac{f(p_l)}{2} + \sum_{i=1}^{l-1} f(p_i). \quad (4.1.2)$$

Le temps géodésique $\tau_f(p, q)$ séparant deux points p et q dans une image à niveaux de gris f est le temps le plus court permettant de relier p à q dans f :

$$\tau_f(p, q) = \inf\{\mathcal{T}_f(\mathcal{P}) | \mathcal{P} \text{ reliant } p \text{ et } q\}. \quad (4.1.3)$$

Les géodésiques pour une image à niveaux de gris sont les chemins qui relient les points d'une image en un minimum de temps. Il peut y avoir plus d'une géodésique entre deux points.

Contrairement à la distance géodésique pour les images binaires, le temps géodésique $\tau_f(p, q)$ ne définit pas une métrique mais un écart puisque la propriété de séparation n'est pas vérifiée, i.e. $\tau_f(p, q) = 0 \not\Rightarrow p = q$. En effet d'après l'équation 4.1.1, on a $\tau_f(p, q) = 0 \Rightarrow f(x) = 0$ pour $x \in \mathcal{P}$. Par abus de langage on parle quand même de distance géodésique que l'on note :

$$d_f^g(p, q) = \tau_f(p, q). \quad (4.1.4)$$

Attention tout de même, car $\tau_f(\mathcal{P})$ est indépendant de la direction prise pour parcourir le chemin ce qui en trame carrée peut provoquer des écarts de 41 %. Pour calculer de manière efficace une distance sur trame carrée on peut se référer aux travaux de *Borgefors* [30]. Cependant si on ne veut pas pondérer la distance sur trame carrée par la distance entre pixels, il est possible de travailler directement avec les temps géodésiques. Dans ce mémoire c'est cette dernière option que nous avons retenue.

4.1.4.3 Extension de la géodésie aux images hyperspectrales

De la même manière que pour les images à niveaux de gris, un chemin (p_0, p_1, \dots, p_n) sur une image hyperspectrale peut être vu comme un graphe pour lequel les nœuds correspondent aux pixels vecteurs reliés entre eux par des arêtes. Pour tout $i = 0 \dots n$, les arêtes entre les nœuds p_i et p_{i+1} sont pondérées par la distance $d(\mathbf{f}_\lambda(p_i), \mathbf{f}_\lambda(p_{i+1}))$.

On peut donc définir le chemin géodésique pour une image hyperspectrale.

Définition 15 (Chemin géodésique dans un graphe valué). Étant donné le graphe associé à une image dont les nœuds correspondent aux pixels et les poids des arêtes aux distances entre pixels, le chemin géodésique entre deux points x et y de E est le chemin de poids minimum.

La distance géodésique entre deux points p et q d'une image hyperspectrale (en fait un écart) est définie comme étant le poids du chemin de poids minimum :

$$d_{\mathbf{f}_\lambda}^g(p, q) = \inf_{\mathcal{P}=(p_0, \dots, p_n)} \sum_{i=0}^n d(\mathbf{f}_\lambda(p_i), \mathbf{f}_\lambda(p_{i+1})) \text{ avec } p_0 = p \text{ et } p_n = q. \quad (4.1.5)$$

Différents travaux concernent la segmentation d'images couleur à partir de métriques géodésiques, e.g. [8, 7].

Remarque 4.1.2. Pour une image hyperspectrale la distance entre deux points voisins p_i et p_{i+1} sur un chemin vaut $d(\mathbf{f}_\lambda(p_i), \mathbf{f}_\lambda(p_{i+1}))$. Sur une image à niveaux de gris cette distance correspondrait à la norme L_1 entre les niveaux de gris : $|f_\lambda(p_i) - f_\lambda(p_{i+1})|$. Dans l'équation 4.1.2 du temps géodésique on utilise comme distance entre pixels voisins la moyenne de leurs niveaux de gris $(f_\lambda(p_i) + f_\lambda(p_{i+1}))/2$ qui permet d'incrémenter le temps géodésique pour deux points voisins appartenant à une zone plate, i.e. ayant le même niveau de gris. Par contre, si on utilise leur différence de niveaux de gris, la distance géodésique sur les zones plates de l'image n'augmente pas, i.e. $|f_\lambda(p_i) - f_\lambda(p_{i+1})| = 0$ si $f_\lambda(p_i) = f_\lambda(p_{i+1})$. Néanmoins pour une extension aux images hyperspectrales, il est nécessaire de passer par une distance entre spectres. Il pourrait être intéressant de définir une métrique qui tiendrait compte à la fois de la distance spectrale et de la distance spatiale pour ensuite calculer une distance géodésique.

4.1.5 Médiane vectorielle

Pour les images multivariées, les extrema (i.e. minimum ou maximum) n'étant pas définis, la médiane vectorielle est une notion très intéressante pour trier et sélectionner des pixels vecteurs [11].

Définition 16 (Médiane vectorielle). Une médiane vectorielle d'un ensemble $R \subset E$ est toute valeur $\mathbf{f}_\lambda(k)$ de l'ensemble au point $k \in R$ telle que :

$$k = \operatorname{argmin}_{p \in R} \sum_{i/x_i \in R} d(\mathbf{f}_\lambda(p), \mathbf{f}_\lambda(x_i)) = \operatorname{argmin}_{p \in R} \Delta_R(\mathbf{f}_\lambda(p)) \quad (4.1.6)$$

La médiane vectorielle est donc un des points qui minimisent la somme des distances à tous les autres. Afin de calculer la médiane vectorielle $\mathbf{f}_\lambda(k)$, $k \in R$, la distance cumulée $\Delta_R(\mathbf{f}_\lambda(p))$ doit être calculée pour tous les points $p \in R$. L'ensemble de ces distances cumulées est ensuite trié par ordre croissant de distance et dénommé « liste par ordre croissant de distances cumulées », Δ_R . Cet ensemble se note :

$$\Delta_R = \{\Delta_R(\mathbf{f}_\lambda(p_0)), \Delta_R(\mathbf{f}_\lambda(p_1)), \dots, \Delta_R(\mathbf{f}_\lambda(p_n))\} \\ \text{tel que } \Delta_R(\mathbf{f}_\lambda(p_0)) < \Delta_R(\mathbf{f}_\lambda(p_1)) < \dots < \Delta_R(\mathbf{f}_\lambda(p_n)). \quad (4.1.7)$$

Le premier élément de la liste $\Delta_R(\mathbf{f}_\lambda(p_0))$ est une médiane vectorielle $\mathbf{f}_\lambda(k)$. Le dernier élément $\Delta_R(\mathbf{f}_\lambda(p_n))$ est considéré comme l'anti-médiane.

4.2 Segmentation par connexions sur des images hyperspectrales

Le but de cette partie est d'introduire les connexions η et μ afin d'ajouter de l'information régionale aux zones λ -plates, qui ne prennent en compte que l'information locale. Une approche « top-down » est utilisée. Tout d'abord les zones λ -plates sont calculées de manière à produire une sous-segmentation. Puis une segmentation plus fine est obtenue en calculant les régions η -bornées et les boules μ -géodésiques à l'intérieur des zones λ -plates. Ces nouvelles zones ont été développées au cours de cette thèse et ont fait l'objet d'une publication [136].

4.2.1 Motivations pour élaborer de nouvelles connexions

Habituellement, les zones λ -plates permettent d'obtenir une première partition de l'image, i.e. une partition fine. L'inconvénient est que cette partition présente des petites régions dans les zones de fort gradient (e.g. régions proches des contours, texturées, etc.). Plusieurs travaux précédents présentent des méthodes pour le résoudre. *Brunner et Soille* [35] proposent une méthode pour réduire la sur-segmentation des zones quasi plates. Ils utilisent un algorithme itératif, sur les images hyperspectrales à partir de germes de plus grande surface qu'un seuil donné. Leur approche consiste à calculer une sur-segmentation puis à fusionner les petites régions. En outre, *Salembier et al.* ont développé une autre méthode pour supprimer les régions plates d'aire inférieure à un seuil donné [154]. Ils travaillent sur le graphe d'adjacence des zones plates et ils définissent un ordre de fusion qui utilise des critères entre régions. De plus, une zone, en dessous d'une certaine taille, ne peut pas être partagée par plusieurs zones plates. *Crespo et al.* ont également introduit une approche similaire [47].

L'inconvénient principal des zones λ -plates, pour la partition de l'espace, est qu'elles sont très sensibles à de petites variations du paramètre λ . L'exemple pédagogique de la figure 4.3 montre que l'on a soit 21 régions soit 1 région selon une très faible variation du paramètre λ . En fait, le problème des zones λ -plates provient de leur définition exclusivement locale : l'information régionale n'est pas prise en compte. Dans l'exemple, pour $\lambda = 10$, on obtient une seule région connexe, qui présente une faible variation locale mais qui induit une forte variation régionale. Ce problème est plus difficile à résoudre que la suppression des petites régions.

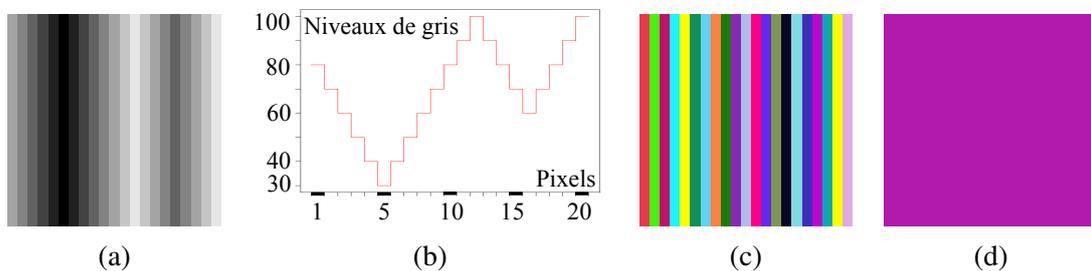


FIG. 4.3 – Zones λ -plates de l'image « dents de scie » (21×21 pixels). (a) Image « dents de scie », (b) profil ligne, (c) $\lambda = 9.9$ (21 zones), (d) $\lambda = 10$ (1 zone).

Le but de ces nouvelles connexions est de traiter ce problème. On débute par une partition initiale en zones λ -plates, avec une valeur suffisamment élevée de λ , de telle manière que son choix ne soit pas critique, et qui conduit à une sous-segmentation, i.e. de larges classes dans la partition. Puis, pour chaque région, on souhaite introduire une deuxième segmentation selon un critère régional. En fait, deux nouvelles connexions sont introduites et qui correspondent aux :

1. régions η -bornées ;
2. boules μ -géodésiques.

Leurs algorithmes utilisent la croissance de régions à partir de germes à l'intérieur des zones λ -plates. On montre que les segmentations obtenues sont moins critiques par rapport au choix du paramètre λ et qu'elles sont plus appropriées pour les images hyperspectrales.

Nous allons maintenant exprimer ces régions d'un point de vue théorique grâce à la théorie de la segmentation de *Serra* [166] (cf. section 4.1, p : 146).

4.2.2 Régions η -bornées

Un des principes majeurs est de comprendre que, pour les zones quasi-plates, la distance, ou pente, entre deux pixels voisins doit être inférieure au paramètre λ pour qu'ils fassent partie de la même zone. On peut établir une comparaison avec un marcheur, partant d'un point et qui ne se préoccuperait que de la pente locale et non du dénivelé global qu'il parcourerait à l'intérieur d'une zone λ -plate. Afin de prendre en compte cette limitation, on définit une nouvelle sorte de boules-régionales selon la connexion suivante.

Définition 17 (Connexion η -bornée). Soit une image hyperspectrale \mathbf{f}_λ et sa partition initiale en zones λ -plates, λFZ , où λFZ_i est la classe connexe i et $R_i \subseteq E$ (de cardinal K) est l'ensemble des points p_k , $k = 0, 1, 2, \dots, K - 1$, qui appartiennent à la classe i . Soit p_0 un point de R_i , appelé le centre de la classe i , et soit $\eta \in \mathbb{R}^+$ un réel positif. Un point p_k appartient à la composante η -connexe centrée en p_0 , et notée $\eta BR_i^{p_0}$, si et seulement si $d(\mathbf{f}_\lambda(p_0), \mathbf{f}_\lambda(p_k)) \leq \eta$ et p_0 et p_k sont connexes.

Pour chaque λFZ_i la méthode est itérée en utilisant différents centres p_j ($j = 0, 1, \dots, J$) jusqu'à ce que tout l'espace de la composante λ -connexe soit segmenté : $\cup_{j=0}^J \eta BR_i^{p_j} = \lambda FZ_i$, $\cap_{j=0}^J \eta BR_i^{p_j} = \emptyset$, où les régions η -bornées sont connexes. Chaque centre p_j appartient à $\lambda FZ_i \setminus \cup_{l=0}^{j-1} \eta BR_i^{p_l}$.

La nouvelle partition de l'image, associée aux connexions η -bornées, est notée ηBR . Il est évident que cette connexion de deuxième ordre est incluse dans la connexion initiale λFZ , i.e. $\eta BR(x) \leq \lambda FZ(x)$. Comme montré par *Serra* [166], un centre ou graine p_j est nécessaire pour garantir la connexité. Pour être précis, la région est bornée selon son centre. La connexion η -bornée est une généralisation de la connexion par saut, à la différence que sur des images hyperspectrales, les germes p_j ne peuvent pas être des minima ou des maxima. On propose de calculer la valeur médiane comme germe initial.

Remarque 4.2.1. La méthode peut aussi être appliquée directement sur l'espace E de l'image initiale \mathbf{f}_λ sans considérer la segmentation en zones λ -plates, λFZ , initiale ; ce qui équivaut à prendre une valeur λ égale à la dynamique maximale de l'image. L'avantage de notre approche est que nous avons maintenant un contrôle de la variation locale, par le paramètre λ , et de la variation régionale, bornée

par η . En outre le calcul des germes est plus cohérent lorsqu'il est fait dans des régions relativement homogènes.

En pratique, la méthode de partition fonctionne de la manière suivante. Pour obtenir les germes, la liste par ordre croissant des distances cumulées Δ_i est calculée pour chacun des K points de chaque zone λ -plate λFZ_i . Puis le premier élément de la liste, i.e. la médiane vectorielle, est utilisé comme premier germe k de la région ηBR_i^1 . On mesure la distance depuis le germe à chaque point $p \in \lambda FZ_i$, telle que ce point p soit voisin d'un point de la région ηBR_i^1 , i.e. $p \in vois(q)$, sachant que $q \in \eta BR_i^1$. Si la distance est inférieure à η alors p est ajouté à la région η -bornée ηBR_i^1 et retiré de la liste Δ_i . Pour tous les autres points $q \in \lambda FZ_i \setminus \eta BR_i^1$, le premier point de la liste Δ_i est le germe de la deuxième région η -bornée ηBR_i^2 . Ensuite, le procédé est itéré jusqu'à ce que tous les points de la zone λ -plate λFZ_i soient affectés à une région η -bornée. L'algorithme 9 explique en détail comment les régions η -bornées sont calculées.

Algorithme 9 Régions η -bornées

```

1: Soient une distance  $d$ , la segmentation en zones  $\lambda$ -plates,  $\lambda FZ$ , d'une image  $f_\lambda$ ,  $\Delta$  une liste
   croissante de distances cumulées,  $Q$  une file d'attente et  $imOut$  une image scalaire de sortie
2: Initialiser la valeur de  $\eta$ 
3:  $currentlabel \leftarrow 0$ 
4: Pour tout  $\lambda FZ \in f_\lambda$  Faire
5:   Pour tout point  $p \in \lambda FZ$  Faire
6:      $distance \leftarrow \sum_{q \in \lambda FZ} d(p, q)$ 
7:      $\Delta \leftarrow$  ajouter la paire  $(p, distance)$ 
8:   Fin Pour
9:   Tri par ordre croissant sur le paramètre  $distance$  de  $\Delta$ 
10:  Tant que  $\Delta$  n'est pas vide Faire
11:     $k \leftarrow$  premier point de  $\Delta$ 
12:     $Q \leftarrow push(k)$ 
13:    Tant que  $Q$  n'est pas vide Faire
14:       $p \leftarrow pop(Q)$ 
15:       $imOut(p) \leftarrow currentlabel$ 
16:      Supprimer  $q$  et sa  $distance$  dans  $\Delta$ 
17:      Pour tout  $q \in vois(p)$  et  $q \in \Delta$  Faire
18:        Si  $d(k, q) \leq \eta$  Alors
19:           $Q \leftarrow push(q)$ 
20:        Fin Si
21:      Fin Pour
22:    Fin Tant que
23:     $currentlabel \leftarrow currentlabel + 1$ 
24:  Fin Tant que
25: Fin Pour

```

Pour comprendre de manière plus intuitive, on peut utiliser la comparaison de la « montagne » : un marcheur qui part d'un point selon une marche restreinte en différence d'altitude à une boule de

diamètre 2η centrée au point de départ. Il ne peut aller au dessus ou au dessous de cette frontière. Les points qui peuvent être atteints par le marcheur, étant données les conditions précédentes, sont connexes et constituent une région η -bornée.

Pour comprendre les effets de ces régions, elles sont appliquées à l'image « dents de scies » avec une distance euclidienne d_E (fig. 4.4). Pour la clarté, on considère une image à un seul canal dont les niveaux de gris ont une forme de « dents de scie » (fig. 4.3 b). On pourrait néanmoins ajouter d'autres canaux de valeur constante ce qui ne changerait pas le résultat.

Sur la figure 4.4, on constate que plus le paramètre η est petit, plus la surface des régions η -bornées est faible. En outre, les régions η -bornées sont très sensibles aux pics. En effet, si η est inférieur à la différence d'altitude entre le germe et le pic, les points entre le germe et le pic sont dans la même région η -bornée. Cependant, si η est plus grand que cette différence d'altitude entre le germe et le pic, les points situés derrière le pic, pour lesquels la différence d'altitude avec le germe est inférieure à η , sont dans la même région η -bornée (fig. 4.5).

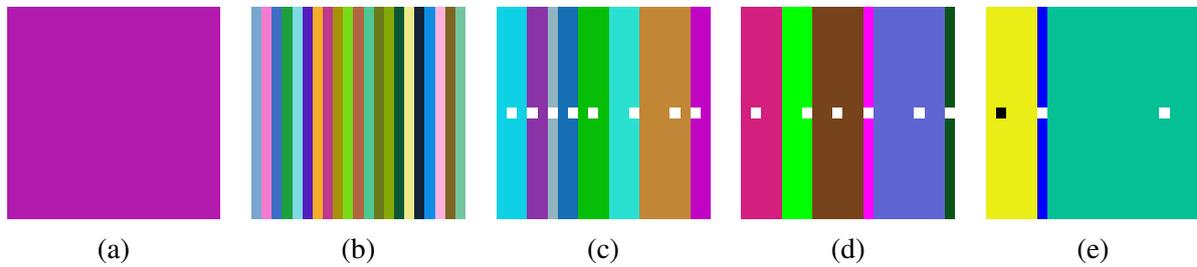


FIG. 4.4 – Régions η -bornées de l'image « dents de scie » (21×21 pixels) pour $\lambda = 10$. Les germes sont indiqués grâce à un point blanc (ou noir), lorsqu'ils ne sont pas triviaux. (a) λFZ pour $\lambda = 10$. ηBR pour $\lambda = 10$ et (b) $\eta = 0$, (c) $\eta = 10$, (d) $\eta = 20$, (e) $\eta = 30$.

4.2.3 Boules μ -géodésiques

De la même manière que pour les régions η -bornées, nous avons créé les boules μ -géodésiques, μGB , afin d'améliorer les zones λ -plates, mais en introduisant maintenant un contrôle sur la « dimension géodésique » de la zone. Tout d'abord, les zones λ -plates sont construites. Puis le dénivelé cumulé est mesuré depuis le point de départ (i.e. le germe) dans chaque zone λ -plate. A partir de ce germe, une boule géodésique de rayon μ est calculée. Cette zone est dénommée boule μ -géodésique. Elle correspond au nombre maximum de pas (i.e. de points) qui peuvent être atteints par un marcheur partant du germe à l'intérieur de la zone λ -plate pour un dénivelé cumulé donné. Les boules μ -géodésiques sont définies en utilisant la connexion suivante :

Définition 18 (Connexion μ -géodésique). Soient une image hyperspectrale $\mathbf{f}_\lambda(x)$ et sa partition initiale en zones λ -plates, λFZ , où λFZ_i est la classe connexe i et $R_i \subseteq E$ (de cardinal K) est l'ensemble des points p_k , $k = 0, 1, 2, \dots, K - 1$, qui appartiennent à la classe i . Soit p_0 un point de R_i , appelé le centre de la classe i , et soit $\mu \in \mathbb{R}^+$ un réel positif. Un point p_k appartient à la composante μ -géodésique centrée en p_0 , notée $\mu GB_i^{p_0}$, si et seulement si $d_{geo}(\mathbf{f}_\lambda(p_0), \mathbf{f}_\lambda(p_k)) \leq \mu$.

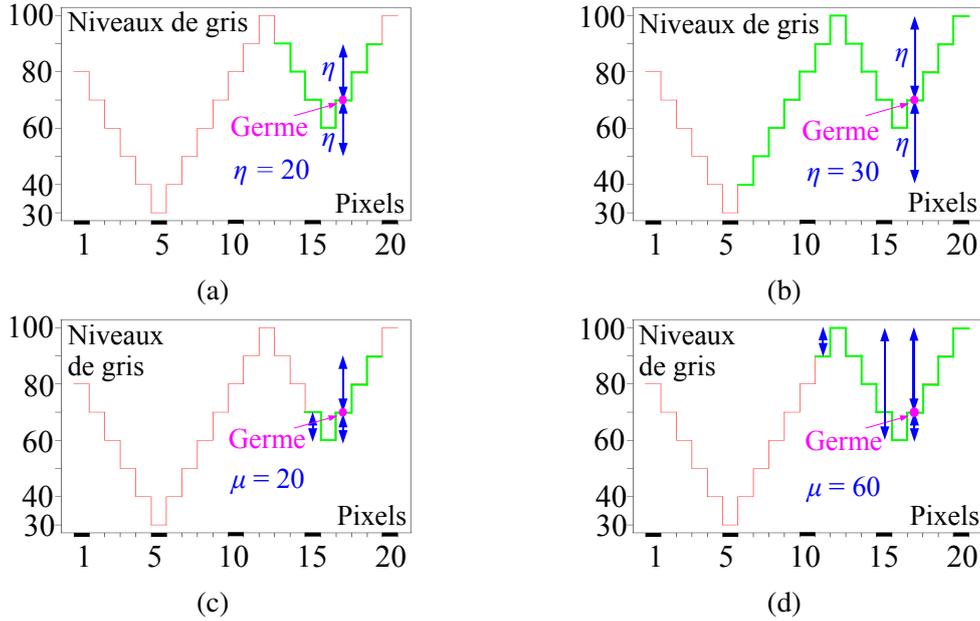


FIG. 4.5 – Régions η -bornées et boules μ -géodésiques (en vert) sur le profil de l'image « dents de scie » pour $\lambda = 10$ et (a) $\eta = 20$, (b) $\eta = 30$, (c) $\mu = 20$, (d) $\mu = 60$.

Il est important de noter, que les chemins géodésiques imposent la connexité aux boules μ -géodésiques.

En pratique, pour chaque zone λ -plate λFZ_i la liste par ordre croissant des distances cumulées Δ_i est calculée. Pour la première boule μ -géodésique μGB_i^1 , d'une zone λ -plate, le premier élément de la liste Δ_i (i.e. la médiane vectorielle de λFZ_i) est utilisé comme germe p_0 . Puis une propagation géodésique depuis le germe à chaque point $p \in \lambda FZ_i$ est calculée. Si $d_{geo}(f_\lambda(p_0), f_\lambda(p)) \leq \mu$ alors p est affecté à la boule μ -géodésique μGB_i^1 et retiré de la liste Δ_i . Pour tous les autres points $q \in \lambda FZ_i \setminus \mu GB_i^1$, le premier point de la liste Δ_i est le germe de la deuxième boule μ -géodésique μGB_i^2 . Ensuite, le procédé est itéré jusqu'à ce que tous les points de la zone λ -plate λFZ_i soient affectés à une boule μ -géodésique. L'algorithme 10 explique en détail comment les boules μ -géodésiques sont calculées.

La nouvelle partition de l'image associée à la connexion μ -géodésique est notée μGB . Comme pour les connexions η -bornées, cette connexion de second ordre est contenue dans la connexion initiale λFZ , i.e. $\mu GB(x) \leq \lambda FZ(x)$. L'avantage de cette approche est que l'on a maintenant un contrôle régional de la « taille géodésique » des classes en mesurant leur distance géodésique, limitée par μ à l'intérieur de la variation locale limitée, elle, par λ . En pratique, les boules μ -géodésiques sont construites de la même manière que les régions η -bornées, sauf qu'à partir de chaque germe une boule géodésique est calculée dans la zone λ -plate.

Les boules μ géodésiques sont calculées avec une distance euclidienne d_E pour l'image « dents de scie » (fig. 4.6). On constate que plus μ est petit, plus petite est la surface des boules μ -géodésiques. En outre, ces boules sont peu sensibles aux pics. En effet, si μ est inférieur à la différence d'altitude entre le germe et le pic, les points situés entre ce germe et ce pic sont dans la même boule μ -géodésique. Cependant, si μ est plus grand que cette différence d'altitude entre le germe et le pic, seuls les points

Algorithme 10 Boules μ -géodésiques

```

1: Soient une distance  $d$ , la segmentation en zones  $\lambda$ -plates,  $\lambda FZ$ , d'une image  $f_\lambda$ ,  $\Delta$  une liste
   croissante de distances cumulées,  $Q$  une file d'attente et  $imOut$  une image scalaire de sortie
2: Initialiser la valeur de  $\mu$ 
3:  $currentlabel \leftarrow 0$ 
4: Pour tout  $\lambda FZ \in f_\lambda$  Faire
5:   Pour tout point  $p \in \lambda FZ$  Faire
6:      $distance \leftarrow \sum_{q \in \lambda FZ} d(p, q)$ 
7:      $\Delta \leftarrow$  ajouter la paire  $(p, distance)$ 
8:   Fin Pour
9: Tri par ordre croissant sur le paramètre  $distance$  de  $\Delta$ 
10: Tant que  $\Delta$  n'est pas vide Faire
11:    $k \leftarrow$  premier point de  $\Delta$ 
12:   Pour tout point  $p$  à l'intérieur de la boule  $\mu$ -géodésique de centre  $k$  et de rayon  $\mu$  à
   l'intérieur de la  $\lambda FZ$  Faire
13:      $imOut(p) \leftarrow currentlabel$ 
14:     Retirer  $p$  et sa  $distance$  dans  $\Delta$ 
15:   Fin Pour
16:    $currentlabel \leftarrow currentlabel + 1$ 
17: Fin Tant que
18: Fin Pour

```

situés derrière le pic, pour lesquels le dénivelé cumulé depuis la graine est inférieur à μ , sont dans la même boule μ -géodésique (fig. 4.5).

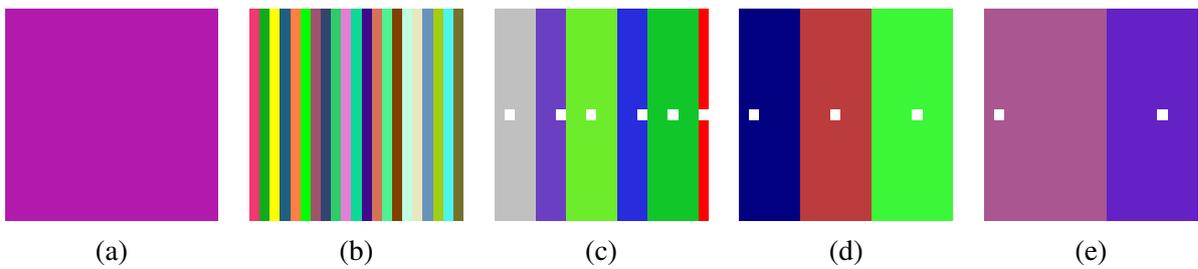


FIG. 4.6 – Boules μ -géodésiques de l'image « dents de scie » (21×21 pixels) pour $\lambda = 10$. Les germes sont indiqués grâce à un point blanc (ou noir), lorsqu'ils ne sont pas triviaux. (a) λFZ pour $\lambda = 10$. μGB pour $\lambda = 10$ et (b) $\mu = 0$, (c) $\mu = 20$, (d) $\mu = 40$, (e) $\mu = 100$.

4.2.4 Résultats de segmentation

Illustrons maintenant ces connexions sur des images réelles : « woman face » et l'image satellite « Roujan ». Auparavant, afin de réduire le nombre de zones plates dans les images, un nivellement morphologique est appliqué sur chaque canal, avec des marqueurs obtenus par un filtre alterné séquentiel (ouverture puis fermeture) par un élément structurant (e.s.) 3×3 pixels, i.e. $\psi_1\phi_1$. Ensuite les zones spectrales λ -plates sont calculées.

Les temps de calculs sont très rapides : avec notre implémentation Python le calcul est inférieur à une minute pour les images satellites. Une implémentation C++ serait beaucoup plus rapide.

4.2.4.1 Segmentations dans l'espace image avec des germes issus de la « médiane »

Nous allons tout d'abord réaliser des segmentations par les connexions η et μ en utilisant la liste par ordre croissant des distances cumulées dont le premier élément est la médiane vectorielle de la zones λ -plate considérée.

Pour segmenter par ces connexions, il est important de choisir une distance appropriée à l'espace de représentation. Dans l'espace image, f_λ , on choisit la distance du chi-deux et dans l'espace factoriel, c_α^f , la distance euclidienne (cf. section 3.4.2, p : 91). Sur la figure 4.7 plusieurs segmentations par zones λ -plates sont produites à partir de la médiane vectorielle utilisée comme germe et de la distance du chi-deux. On constate que les segmentations sont formées soit de très petites zones (souvent de taille 1 pixel) soit de zones λ -plates très étendues.

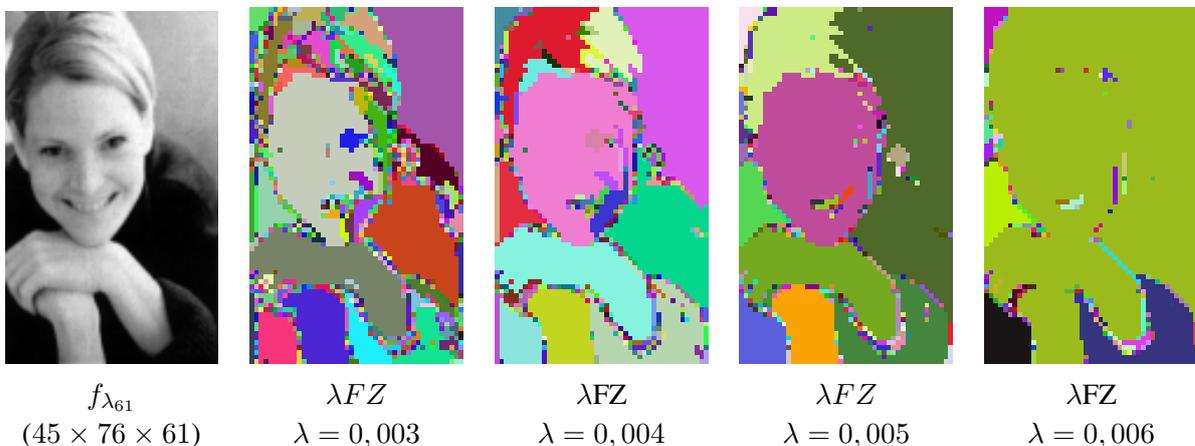


FIG. 4.7 – Canal $f_{\lambda_{61}}$ de l'image « woman face » et zones λ -plates (de l'image à 61 canaux) pour $\lambda = 0,003$; $\lambda = 0,004$; $\lambda = 0,005$ et $\lambda = 0,006$.

Sur la figure 4.8, on constate que les régions η -bornées et les boules μ -géodésiques permettent de palier aux limitations des zones λ -plates, puisque même des petites régions comme les yeux sont bien détectées. On part d'une sous-segmentation en zones λ -plates (avec $\lambda = 0,005$). Puis on applique les connexions η et μ . On peut également vérifier que les régions η -bornées sont moins sensibles que les boules μ -géodésiques aux faibles variations de distance entre les points. Cependant, l'aire des boules μ -géodésiques est plus contrôlée que l'aire des régions η -bornées. Par conséquent, les régions

η -bornées sont meilleures pour trouver les détails et les boules μ -géodésiques sont meilleures pour obtenir des zones plus lisses et de taille plus homogènes.

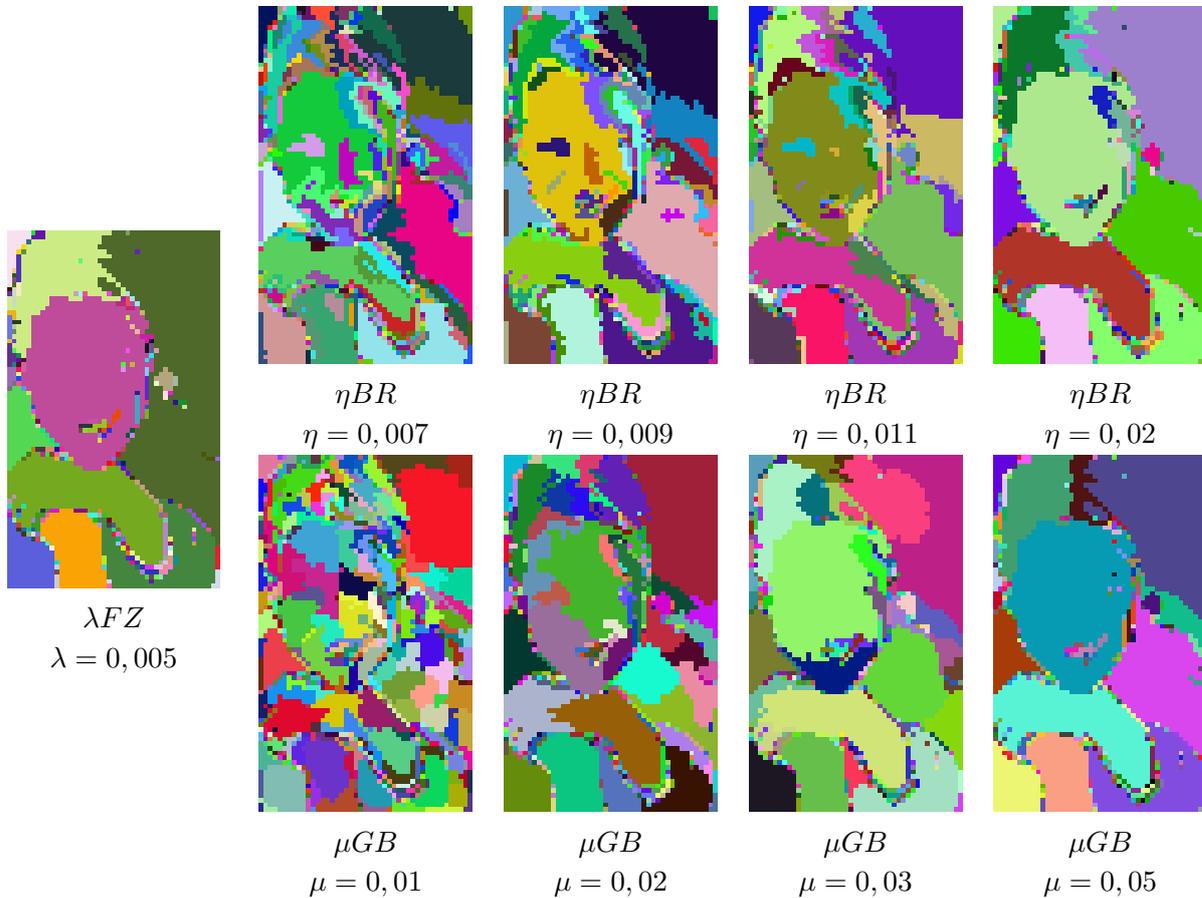


FIG. 4.8 – Segmentations par des régions η -bornées, ηBR , ou des boules μ -géodésiques, μGB , à partir des zones λ -plates de l'image « woman face » avec $\lambda = 0,005$. La distance du chi-deux est ici utilisée dans l'espace image f_λ .

On mesure également pour chaque segmentation le nombre de zones η -bornées ou boules μ -géodésiques moins le nombre de zones λ -plates. Ce nombre de zones décroît avec les paramètres η ou μ . L'allure de la courbe est une exponentielle décroissante caractéristique d'un processus de file d'attente (fig. 4.9).

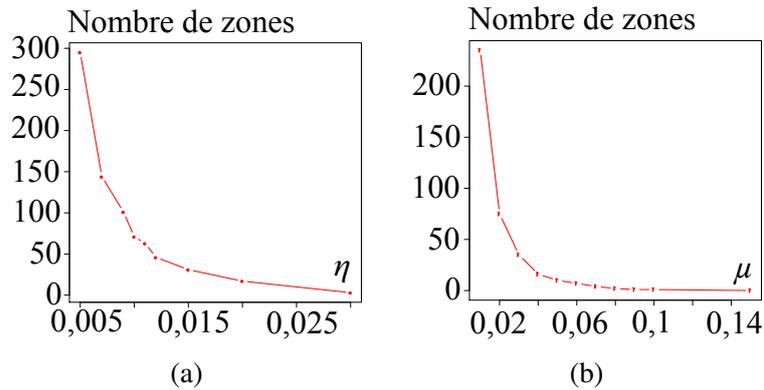


FIG. 4.9 – Variations du nombre de (a) régions η -bornées, ηBR , ou (b) μ -géodésiques, μGB , moins le nombre de zones λ -plates, λFZ , en fonction du paramètre η ou μ dans l'image « woman face » pour $\lambda = 0,005$ avec la distance du chi-deux d_{χ^2} dans l'espace image f_{λ} .

4.2.4.2 Segmentations dans l'espace image avec des germes issus de l'« anti-médiane »

De plus, afin d'évaluer l'importance du choix de germe, on décide d'utiliser le dernier élément de la liste par ordre croissant des distances cumulées qui correspond à l'anti-médiane vectorielle. Sur la figure 4.10 et par comparaison avec la figure 4.8, on constate que les segmentations par connexions η ou μ à partir de germes issus de l'anti-médiane sont similaires à celles obtenues par des germes issus de la médiane. En conséquence, les régions η -bornées et les boules μ -géodésiques ont une faible dépendance aux germes choisis. Ces méthodes sont donc robustes par rapport aux germes.

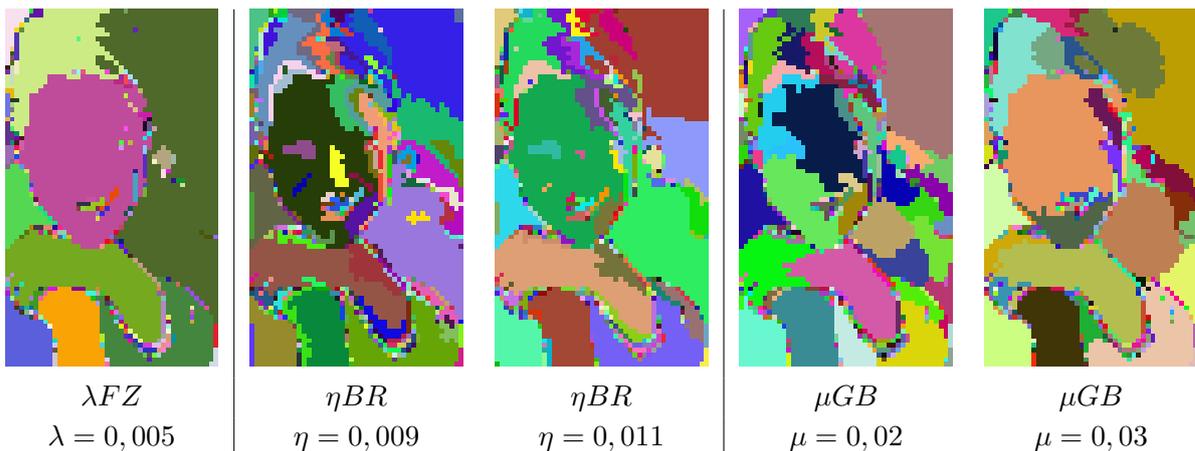


FIG. 4.10 – Régions η -bornées, ηBR , et boules μ -géodésiques, μGB , calculées avec des germes issus de l'anti-médiane dans les zones λ -plates pour $\lambda = 0,005$ sur l'image « woman face » (avec la distance du chi-deux d_{χ^2}).

4.2.4.3 Segmentations dans l'espace factoriel

D'autres segmentations peuvent être obtenues dans l'espace factoriel \mathbf{c}_α^f avec une distance euclidienne parce qu'elle est équivalente à la distance du chi-deux dans l'espace image \mathbf{f}_λ (fig. 4.11). En outre, à cause de la réduction du volume de données, le temps de calcul est meilleur. On travaille dans l'espace factoriel à 3 axes d'inerties respectives : $\tau_{\alpha_1} = 87,8\%$, $\tau_{\alpha_2} = 10,2\%$ et $\tau_{\alpha_3} = 1,5\%$. Les germes utilisés sont issus de la « médiane ».

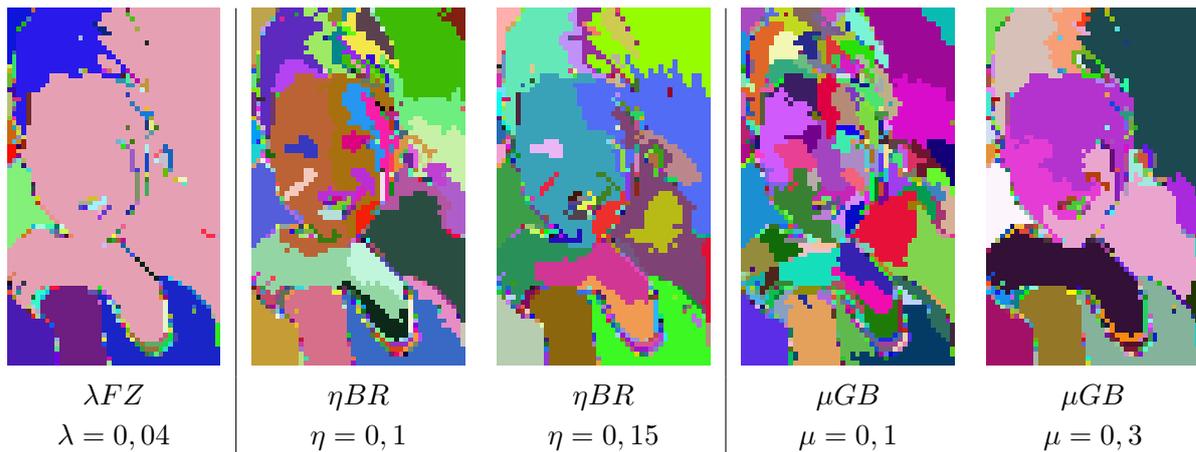


FIG. 4.11 – Régions η -bornées, ηBR , et boules μ -géodésiques, μGB , calculées avec des germes, issus de la médiane, dans les zones λ -plates pour $\lambda = 0,04$. Les segmentations sont effectuées dans l'espace factoriel AFC \mathbf{c}_α^f de l'image « woman face » (avec la distance euclidienne d_E).

4.2.4.4 D'autres résultats sur une image satellite

Afin de tester la pertinence de la méthode sur des images plus complexes, on réalise des segmentations par connexions η et μ sur l'image satellite « Roujan » à 5 canaux (cf. section 2.3, p : 19).

On calcule tout d'abord les zones λ -plates et on retient la sous-segmentation pour le paramètre $\lambda = 0,015$ (fig. 4.12). Sur cette segmentation, on définit les segmentations du second ordre : les régions η -bornées et les boules μ -géodésiques (fig. 4.13). Par rapport à la segmentation par connexion λ , ces nouvelles segmentations conduisent à un meilleur découpage des zones de l'image en introduisant de l'information régionale qui permet de contrôler la taille des régions.

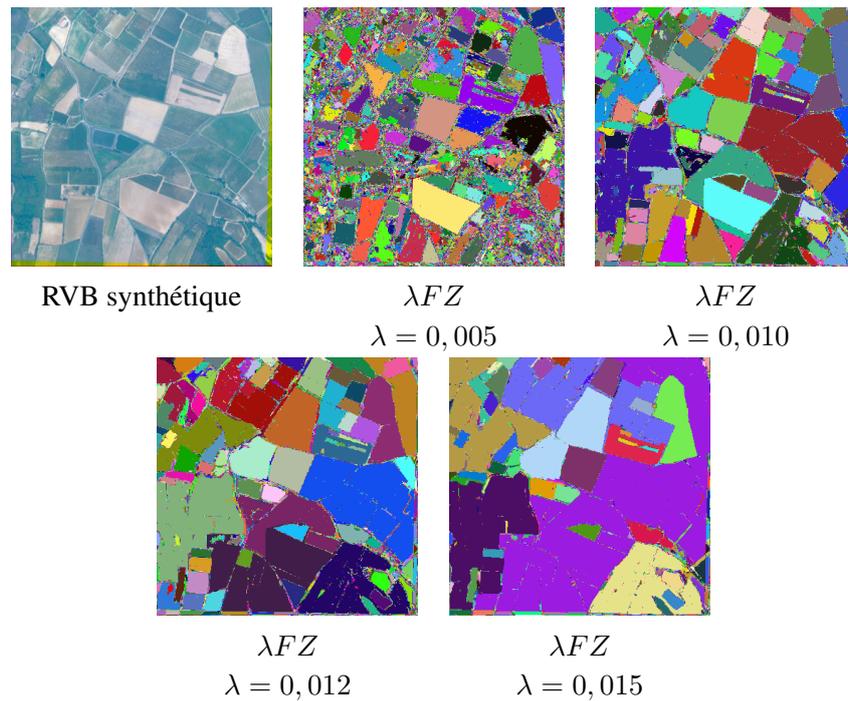


FIG. 4.12 – Image de synthèse RVB et zones λ -plates, λFZ , pour l'image « Roujan ».

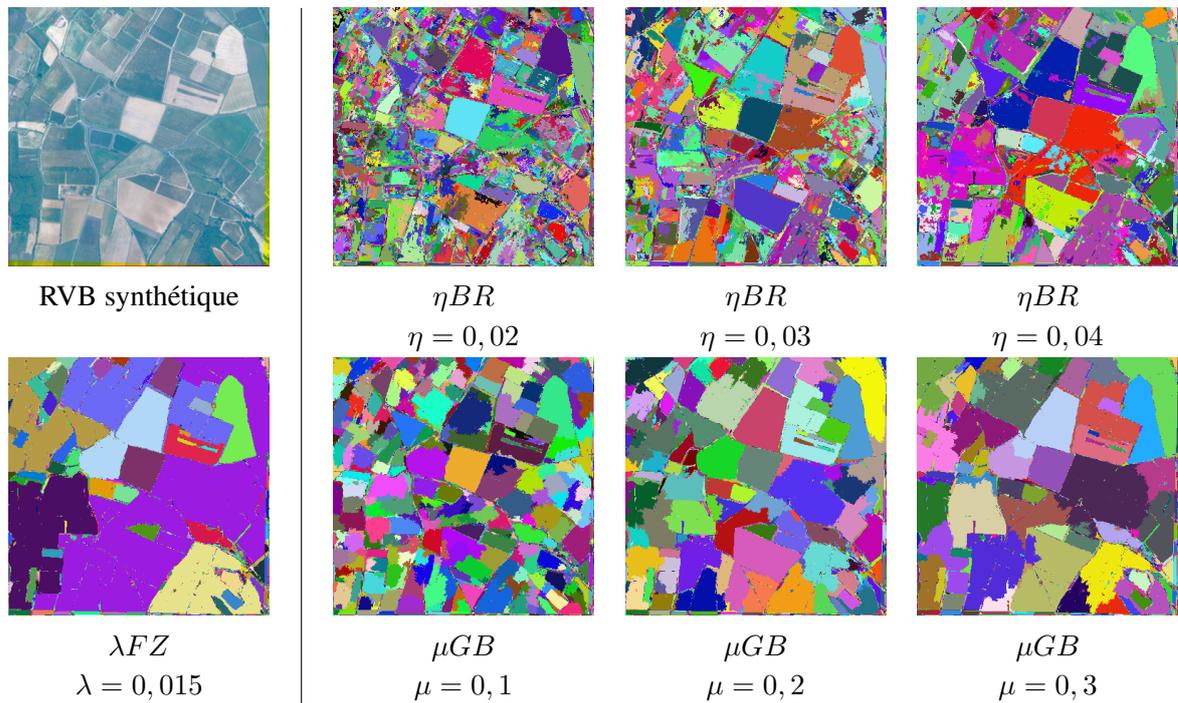


FIG. 4.13 – Régions η -bornées, ηBR , et boules μ -géodésiques, μGB , calculées dans les zones λ -plates, λFZ , pour $\lambda = 0,015$ sur l'image satellite « Roujan » (avec la distance du chi-deux d_{χ^2}).

4.2.5 Conclusion et perspectives sur la segmentation hyperspectrale par connexions

Les connexions η et μ améliorent la segmentation par zones λ -plates, qui ne prennent en compte que l'information locale, grâce à l'introduction de l'information régionale. En outre, ces nouvelles connexions sont de second ordre car elles sont incluses dans les zones connexes λ -plates. Afin de construire ces nouvelles zones, la méthode consiste à sélectionner un paramètre λ suffisamment élevé pour obtenir une sous-segmentation. Puis, les régions η -bornées et les boules μ -géodésiques sont construites, conduisant à une segmentation par agrégation « top-down ». Les régions η -bornées introduisent un paramètre qui contrôle les variations spectrales (i.e. le dénivelé spectral) dans les zones λ -plates, tandis que les boules μ -géodésiques introduisent un paramètre de contrôle de la taille des boules par la mesure de la distance spectrale cumulée dans les zones λ -plates.

En outre, ces connexions du second ordre produisent des pyramides de partition avec un nombre décroissant de régions quand la valeur η ou μ croît, jusqu'à atteindre la partition associée au niveau qui correspond à celle des zones λ -plates. Cependant, il est intéressant de remarquer que cette pyramide ne constitue pas une hiérarchie ordonnée en raison des classes qui ne sont pas ordonnées par niveau croissant.

De plus, nous avons proposé des algorithmes pour la construction des partitions associées à ces deux connexions η et μ . Ils utilisent des files d'attentes avec un tri des germes selon leur distance cumulée à tous les autres points d'une zone λ -plate.

Ces nouvelles segmentations pourraient constituer des pré-segmentations utilisables comme marqueurs pour la Ligne de Partage des Eaux. Une autre application pourrait concerner la détection et la caractérisation des textures.

Au niveau des perspectives, on peut remarquer que notre méthode ne résout pas le problème des petites classes de la partition initiale en zones λ -plates. Il pourrait être intéressant de combiner notre méthode avec les approches d'agrégation des petites régions aux plus importantes [35, 47, 162]. Une autre piste intéressante consisterait à sélectionner les germes de manière plus avancée en déterminant localement pour chaque zone λ -plate la valeur appropriée de η ou de μ .

4.3 Calcul de toutes les paires de distances géodésiques d'une image

Dans cette section, une nouvelle méthode, très efficace, de calcul du tableau de paires de distances géodésiques entre les pixels d'une image est présentée.

4.3.1 Motivations

Un tableau de paires de distances géodésiques entre les pixels d'une image pourrait être très utile pour faire de la réduction de dimension spatio-spectrale grâce à l'analyse par tableau de masses et distances introduite par *Benzécri* [21]. En effet, on peut combiner linéairement le carré d'une distance spectrale avec le carré de la distance géodésique entre pixels pour créer une distance au carrée spatio-spectrale. Les tableaux de paires de distances sont également utilisés dans la segmentation par coupes de graphes [175]. La méthode de réduction de données *Isomap* [177, 178] utilise également les distances géodésiques calculées sur un voisinage. Avec un tel tableau on pourrait envisager de travailler sur toute l'image ou du moins sur un voisinage plus grand. Présentons tout d'abord la problématique de ce calcul.

Calculer toutes les paires de distances géodésiques entre les pixels d'une image est très coûteux en temps de calcul. L'approche naïve consiste en effet à relancer P fois, avec P le nombre de pixels, l'algorithme de propagation géodésique utilisé qui est de complexité M . La complexité totale est donc en $O(P \times M)$.

Comme algorithme de propagation géodésique il existe notamment l'algorithme des *Fast Marching* [169], ou encore l'algorithme proposé par *Soille* [171]. Afin de comparer nos méthodes de calculs des paires de distances géodésiques nous avons choisi d'utiliser ce dernier algorithme qui s'applique très facilement sur les trames carrées.

4.3.2 Introduction du problème

Il existe une propriété intéressante pour les chemins géodésiques utilisée par *Bertelli et al.* [24] pour réduire le nombre de chemins géodésiques à calculer.

Propriété 4.3.1. *Soit un chemin géodésique (p_0, p_1, \dots, p_n) . La distance géodésique, $d_{geo}(p_i, p_j)$, entre deux point p_i et p_j , $i < j$, du chemin est égale à la différence $d_{geo}(p_0, p_j) - d_{geo}(p_0, p_i)$.*

Vérifions cette propriété par l'absurde.

Démonstration 4.3.1. Considérons un chemin géodésique reliant un point A à un point B et supposons que les points C et D appartiennent à ce chemin géodésique (fig. 4.14) qui est celui de plus courte distance entre A et B . On a donc :

$$d_{geo}(C, D) = d_{geo}(A, D) - d_{geo}(A, C) \quad (4.3.1)$$

Supposons qu'il y ait un autre chemin reliant C à D qui soit plus court :

$$d_{geo}(C, D)_{path2} < d_{geo}(C, D)_{path1}$$

On peut alors écrire :

$$d_{geo}(A, C)_{path1} + d_{geo}(C, D)_{path2} + d_{geo}(D, B)_{path2} < d_{geo}(A, B)_{path1},$$

ce qui est une contradiction et ce qui prouve que $d_{geo}(C, D)_{path1}$ correspond au chemin minimum.

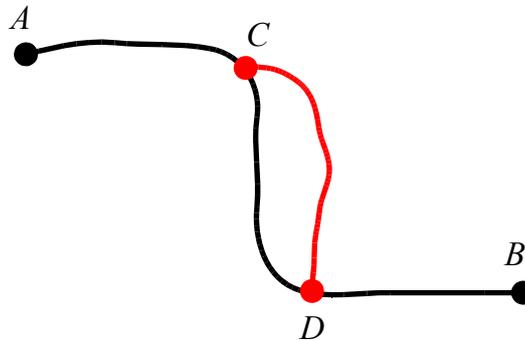


FIG. 4.14 – Le chemin noir est le plus court chemin entre A et B , qui est aussi le plus court chemin entre toutes les paires de points le long de ce chemin (dans cet exemple C et D). Tout autre chemin entre C et D donnerait une distance supérieure (ou égale).

D'après cette propriété, lorsque la distance géodésique entre deux points de l'image est calculée, on connaît donc la distance entre toutes les paires de points situées sur le chemin géodésique reliant ces deux points. Le remplissage du tableau des paires de distance Tab s'effectue donc plus vite en utilisant cette redondance.

Pour calculer facilement les distances entre les points le long des chemins géodésiques, il est intéressant d'allouer un arbre pour chaque propagation géodésique [24]. Dans cet arbre, on distingue trois types de nœuds :

1. la racine qui est le point de départ de la propagation géodésique et qui a une distance nulle et qui n'a pas de parents ;
2. les nœuds qui sont des points avec à la fois des parents et des enfants ;
3. les feuilles qui sont des points sans enfants.

En partant des feuilles et en remontant à la racine (ou bien en partant de la racine vers les feuilles), les distances entre les points appartenant au même chemin minimum peuvent être estimées. La distance géodésique entre deux points d'un chemin le plus court (i.e. géodésique) est la différence entre leurs distances depuis la racine. Sur la figure 4.15, la distance géodésique entre p_0 et p_5 est simplement $d_5 - d_1$.

Une autre propriété des chemins géodésiques est particulièrement intéressante pour calculer un tableau de paires de distances.

Propriété 4.3.2. *Plus les chemins géodésiques sont longs, plus on peut calculer de paires de distances pour une propagation géodésique.*

Afin de tirer parti de cette propriété, Bertelli et al. [24] choisissent comme point de départ des propagations géodésiques (points sources) des points tirés aléatoirement en spirale depuis les bords de l'image vers le centre de l'image. En effet les points des bords ont tendance à avoir des chemins géodésiques plus longs que les points situés au centre. Cette approche qui peut être intéressante d'un

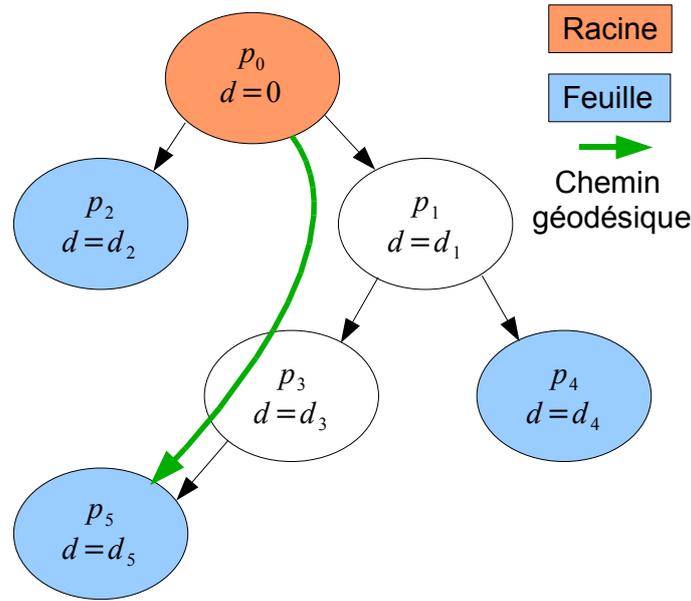


FIG. 4.15 – Arbre géodésique : en chaque nœud la position du point et sa distance d à la racine sont conservées en mémoire.

point de vue pratique ne garantit néanmoins pas que l'on choisisse des points sources pour lesquels les chemins géodésiques sont les plus longs. Cette méthode sera dénommée ci-après : « méthode spirale ».

Plusieurs mesures de dissimilarités peuvent être considérées entre deux pixels de position x_i et $x_{i'}$ et de valeur de niveau de gris $f(x_i)$ et $f(x_{i'})$:

– Écart L_1 :

$$d(x_i, x_{i'}) = |f(x_i) - f(x_{i'})| \quad (4.3.2)$$

– Écart somme des niveaux de gris :

$$d(x_i, x_{i'}) = f(x_i) + f(x_{i'}) \quad (4.3.3)$$

– Écart moyenne des niveaux de gris (similaire à l'écart précédent) :

$$d(x_i, x_{i'}) = \frac{f(x_i) + f(x_{i'})}{2} \quad (4.3.4)$$

Par la suite, nous préférons l'écart somme des niveaux de gris par rapport à la métrique L_1 car elle permet de faire croître la distance géodésique pour un chemin passant par une zone plate (cf. remarque 4.1.2, p : 153).

4.3.3 Méthode spirale

La méthode spirale [24] consiste à choisir des points de manière aléatoire sur les spires concentriques de l'image. Une spire concentrique est un cadre d'épaisseur un pixel dont le coin supérieur gauche est à la position (1, 1) ou (2, 2) ou (3, 3) ou ... La figure 4.16 en donne un exemple. Tant que

tous les pixels de la spire ne sont pas pleins, on tire aléatoirement un point de la spire sinon on passe à la spire suivante (algorithme 11).

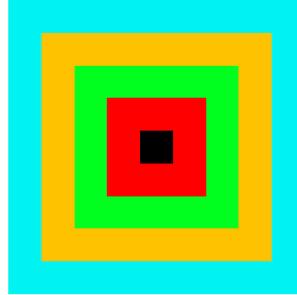


FIG. 4.16 – Les spirales sur une image 9×9 pixels.

Algorithme 11 Méthode spirale

- 1: Soit D le tableau de distances de taille $P \times P$
 - 2: **Tant que** D n'est pas plein **Faire**
 - 3: Prendre la spire la plus extérieure non remplie
 - 4: Déterminer S la liste des points non remplis de la spire
 - 5: **Tant que** S n'est pas vide **Faire**
 - 6: choisir un point source s aléatoirement dans S
 - 7: calculer l'arbre de propagation géodésique depuis s
 - 8: remplir le tableau D
 - 9: retirer les points de S qui sont remplis
 - 10: **Fin Tant que**
 - 11: **Fin Tant que**
-

Comme signalé précédemment, pour le calcul de la distance géodésique, on préfère utiliser l'écart somme des niveaux de gris puisqu'il permet d'augmenter la distance géodésique sur les zones plates (cf. section 4.1.4.2, p : 151).

On va tester et comparer la méthode spirale avec d'autres méthodes que nous avons conçues sur des images de test. Ces images ont de fortes dynamiques de niveaux de gris compris dans l'intervalle $[1; 255]$. La valeur 0 est exclue de la dynamique car elle conduit à une distance géodésique nulle entre deux pixels voisins. Ces images ont une taille de 25×25 pixels (fig. 4.17).

Afin de juger de l'efficacité d'une méthode par rapport à une autre, on mesure le taux de remplissage T^{ux} du tableau des paires de distances géodésiques D de taille $P \times P$, avec P le nombre de pixels de l'image. Ce taux est défini comme le rapport entre le nombre de chemins calculés $NbChCalc$ et le nombre de chemin initiaux $NbChIni$. Le nombre de chemins initiaux $NbChIni$ est égal au nombre d'éléments du tableau auquel on retranche le nombre d'éléments de la diagonale car la distance d'un point à lui-même est nulle. Comme la distance (ou l'écart) utilisée est symétrique, ces deux nombres sont en fait divisés par deux.

Pour des images de taille 25×25 pixels, le nombre total de pixels est donc $P = 25 \times 25 = 625$ pixels. Le tableau des paires de distance a donc une taille de $625 \times 625 = P^2$ éléments. On comprend

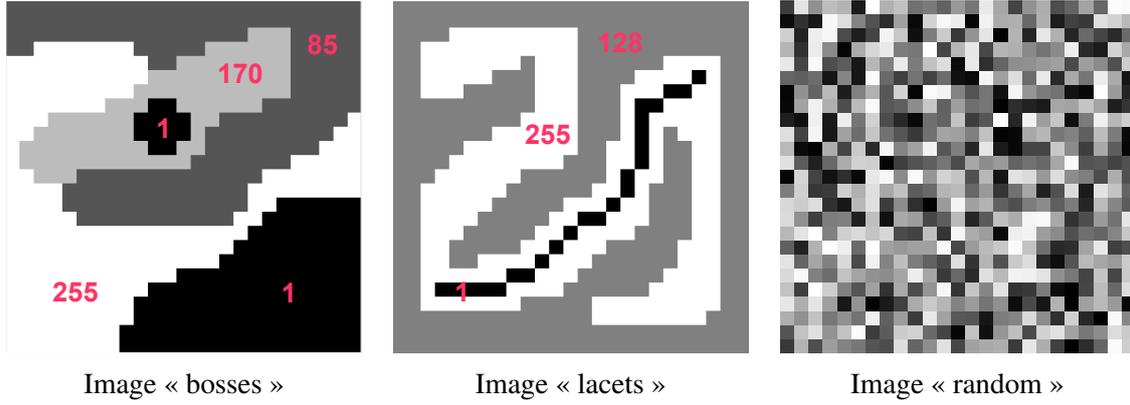


FIG. 4.17 – Images de test 25×25 pixels « bosses », « lacets » et « random » dont les niveaux de gris sont compris entre 1 et 255.

donc que le calcul de l'ensemble des paires de distances géodésiques est très coûteux en temps de calcul et en espace mémoire pour des images de taille beaucoup plus grande.

Le nombre de chemin initiaux $NbChIni$ à calculer vaut :

$$NbChIni = \frac{P^2 - P}{2}. \quad (4.3.5)$$

Le nombre de chemins calculés $NbChCalc$ est déterminé par un comptage des éléments remplis du tableau D . En pratique, il est intéressant d'avoir un tableau de booléens D_{mrk} de taille $P \times P$ et dont les éléments valent 1, si la distance entre les deux pixels correspondant au numéro de la ligne et au numéro de la colonne a été calculée, et 0 sinon. Il ne reste ensuite plus qu'à faire la somme des éléments de cette matrice, à lui retrancher sa trace et à diviser le tout par deux :

$$NbChCalc = \frac{1}{2} \left[\left(\sum_{k=1}^P \sum_{l=1}^P D_{mrk}(k, l) \right) - trace(D_{mrk}) \right]. \quad (4.3.6)$$

Le taux de remplissage vaut donc :

$$T^{ux} = \frac{NbChCalc}{NbChIni}. \quad (4.3.7)$$

On est également capable de mesurer la proportion de chemins restant à calculer pour un pixel x_i donné. On parlera de remplissage d'un point :

$$T^{ux}(x_i) = \frac{\left(\sum_{k=1}^P D_{mrk}(k, i) \right) - D_{mrk}(i, i)}{P - 1}. \quad (4.3.8)$$

Lorsque toutes les distances d'un point x_i à tous les autres ont été calculées, i.e. $T^{ux}(x_i) = \frac{P-1}{P-1} = 1$, on dit que ce point est « rempli ».

Observons maintenant le remplissage des points de l'image selon les itérations. Lorsqu'un point est rempli il est marqué en vert, lorsqu'il s'agit du point source de la propagation on le marque en rouge, sinon il garde son niveau de gris. Sur la figure 4.19 le remplissage des points de l'image

« bosses » pour les 100 premiers points sources déterminés par la méthode spirale est représenté. On constate que même si les points sont remplis en priorité par couches concentriques en partant du bord de l'image, d'autres points sont également remplis, notamment ceux situés dans le coin inférieur droit. La méthode de remplissage en spirale permet donc d'accélérer le remplissage du tableau de distances. Observons également le remplissage pour les images « lacets » (fig. 4.19) et « random » (fig. 4.20).

Pour chacune des images, on trace le taux de remplissage en fonction du nombre de propagations géodésiques effectuées (fig. 4.21). Par la méthode naïve, le taux de remplissage T^x a une variation linéaire en fonction du nombre de propagations. La méthode spirale est donc bien meilleure que la méthode naïve puisque le tableau de distances est bien plus vite rempli (tableau 4.1).

	Méthode spirale	Méthode naïve	Écart
« Bosses »	465	625	25,6 %
« Lacets »	541	625	13,4 %
« Random »	524	625	16,6 %

TAB. 4.1 – Comparaison du nombre de propagation nécessaires pour remplir le tableau des distances par la méthode naïve et par la méthode spirale.

Nous avons cherché à améliorer la méthode spirale en cherchant les extrema géodésiques.



FIG. 4.18 – Remplissage (pixels verts) de l'image « bosses » pour les 100 premiers points sources (en rouge) par la méthode spirale.

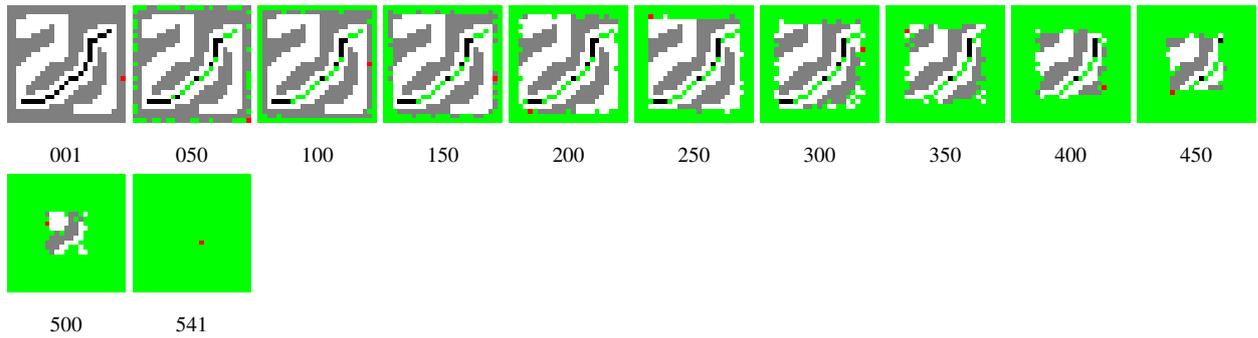


FIG. 4.19 – Quelques images du remplissage (pixels verts) de l'image « lacets » pour différents points sources (en rouge) par la méthode spirale.

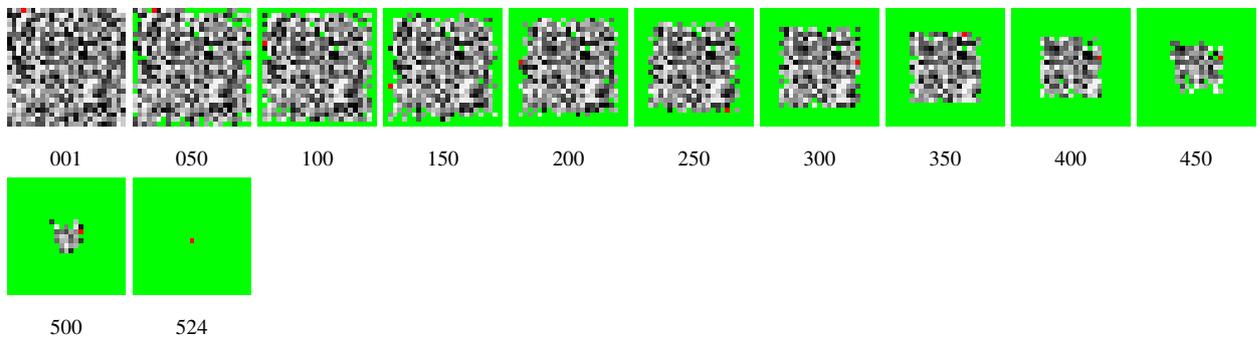


FIG. 4.20 – Quelques images du remplissage (pixels verts) de l'image « random » pour différents points sources (en rouge) par la méthode spirale.

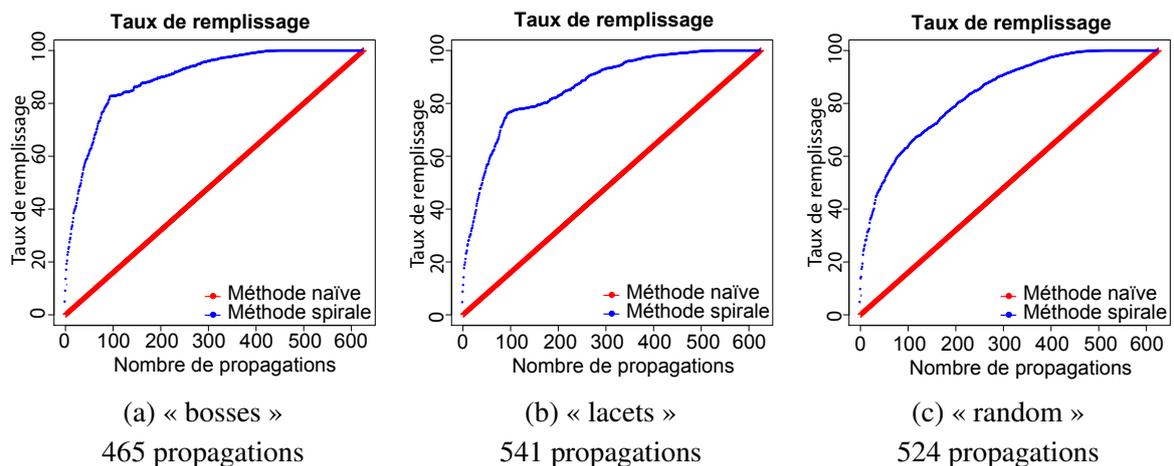


FIG. 4.21 – Comparaison des taux de remplissage, du tableau de paires de distances, entre la méthode naïve (en rouge) et la méthode spirale (en bleu) pour les images « bosses », « lacets » et « random ». Le nombre de propagations nécessaires pour remplir le tableau par la méthode spirale est indiqué sur la ligne du bas.

4.3.4 Méthode des extrema géodésiques

Comme les chemins géodésiques les plus longs sont ceux qui remplissent le plus vite le tableau de distance, on cherche les extrema géodésiques de l'image. Pour calculer les extrema géodésiques, on utilise deux propagations géodésiques :

- une propagation depuis les pixels du bord de l'image. A partir de cette propagation on choisit un des points c ayant la plus grande distance aux bords. Ce point est appelé le centroïde géodésique ;
- une deuxième propagation depuis le centroïde géodésique qui permet d'obtenir les points les plus éloignés du centroïde, i.e. les extrema géodésiques de l'image.

La figure 4.22 montre les résultats de ces deux propagations. On constate qu'effectivement les extrema géodésiques se trouvent plutôt sur les bords de l'image.

On trie ensuite les pixels de l'image par distance géodésique décroissante au centroïde dans la liste Ext , i.e. la liste des extrema géodésiques. Cette liste est utilisée pour choisir les points sources s des propagations géodésiques. Si plusieurs points ont la même distance géodésique par rapport au centroïde, alors on choisit le moins rempli (algorithme 12).

Algorithme 12 Méthode des extrema

- 1: Soient D le tableau de distances de taille $P \times P$
 - 2: Calculer la liste décroissante des extrema géodésiques Ext dont le premier élément $Ext[1]$ est le plus grand extrema géodésique non rempli
 - 3: Initialiser à zéro la liste, de taille P , des taux de remplissage des points T^{ux}
 - 4: **Tant que** D n'est pas plein **Faire**
 - 5: $A \leftarrow \{x \in Ext \mid d_{geo}(x, c) = d_{geo}(Ext[1], c)\}$
 - 6: **Si** $Card\{A\} > 1$ **Alors**
 - 7: $s \leftarrow \operatorname{argmin}_{x \in A} T^{ux}(x)$
 - 8: **Sinon**
 - 9: $s \leftarrow x$
 - 10: **Fin Si**
 - 11: calculer l'arbre de propagation géodésique depuis s
 - 12: remplir le tableau D
 - 13: retirer les points de Ext qui sont remplis
 - 14: mettre à jour la liste des taux de remplissage T^{ux}
 - 15: **Fin Tant que**
-

Observons le remplissage de l'image « lacets » pour quelques propagations (fig. 4.23). On constate que la méthode par extrema permet d'accélérer le remplissage du tableau de distances.

Afin de quantifier cette amélioration, traçons les taux de remplissage en fonction du nombre de propagations (fig. 4.24). Le nombre de propagations nécessaires au remplissage du tableau est également calculé (tableau 4.2). Notons qu'il faut rajouter une propagation géodésique afin de déterminer le centroïde géodésique. On observe que le taux de remplissage de la méthode par extrema est inférieur au taux de remplissage de la méthode spirale pour les propagations du début. Ensuite les tendances s'inversent. Comme ce qui nous importe est d'avoir un tableau entièrement rempli, la méthode par

extrema est donc légèrement meilleure que la méthode spirale. D'ailleurs, dans le tableau 4.2, l'écart relatif entre la méthode extrema et la méthode spirale est de l'ordre de 5%.

	Méthode extrema	Méthode naïve	Écart (extrema-naïve)	Méthode spirale	Écart (extrema-spirale)
« Bosses »	437	625	30,1 %	465	6,0 %
« Lacets »	513	625	17,9 %	541	5,2 %
« Random »	506	625	19,0 %	524	3,4 %

TAB. 4.2 – Comparaison du nombre de propagation nécessaires pour remplir le tableau des distances par la méthode extrema, par la méthode naïve et par la méthode spirale.

Remarque 4.3.1. D'autres essais consistant à recalculer la liste des extrema géodésiques à différentes étapes ont été effectués. Cependant on ne note pas d'amélioration puisqu'en fait à part pour les premières propagations, c'est surtout le taux de remplissage qui est intéressant pour remplir plus rapidement le tableau.

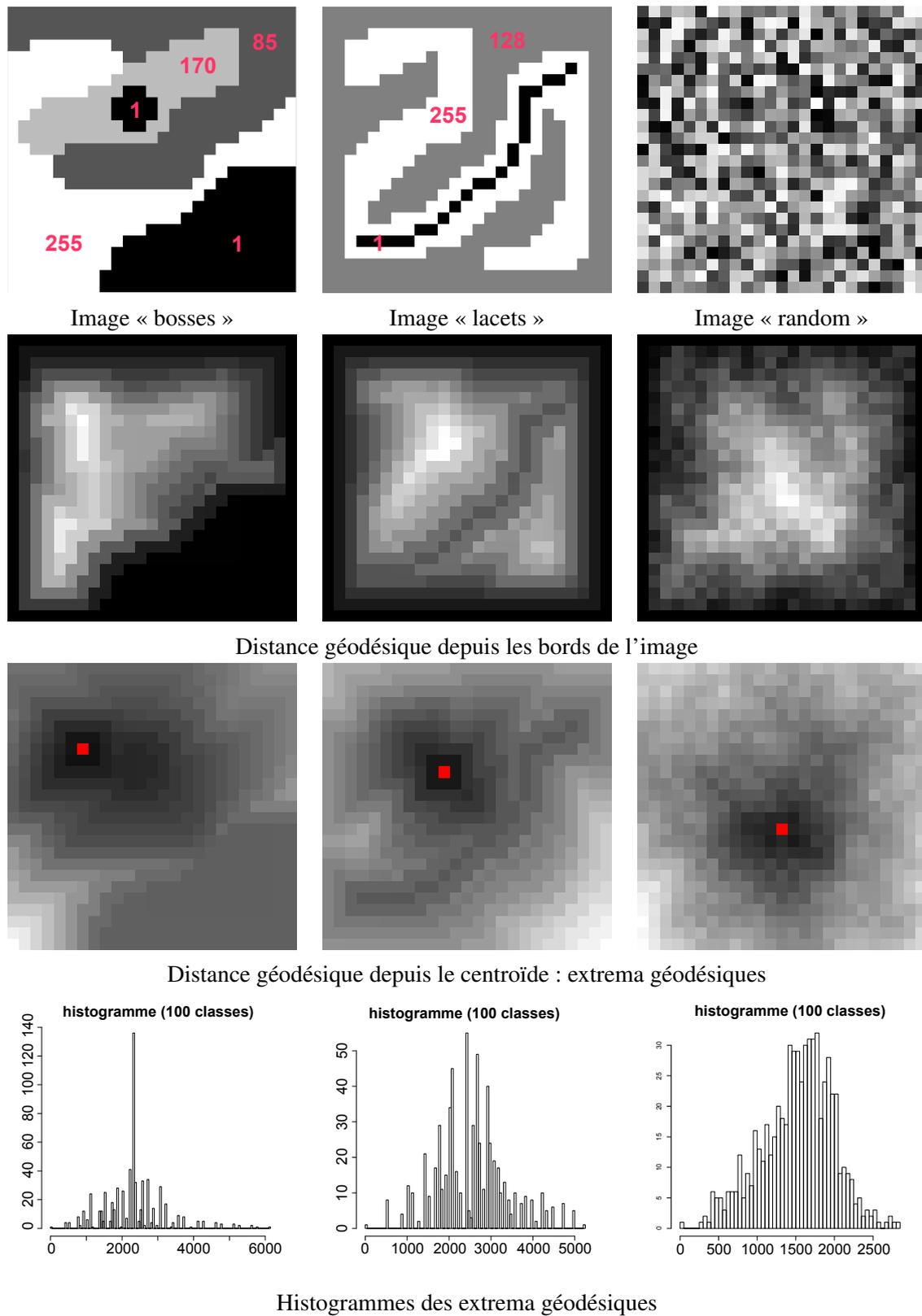


FIG. 4.22 – Distances géodésiques depuis les bords de l'image (deuxième ligne) et depuis le centroïde en rouge (troisième ligne) pour les images « bosses », « lacets » et « random ». Histogrammes des extrema géodésiques (quatrième ligne).



FIG. 4.23 – Quelques images du remplissage (pixels verts) de l'image « lacets » pour différents points sources (en rouge) par la méthode des extrema géodésiques.

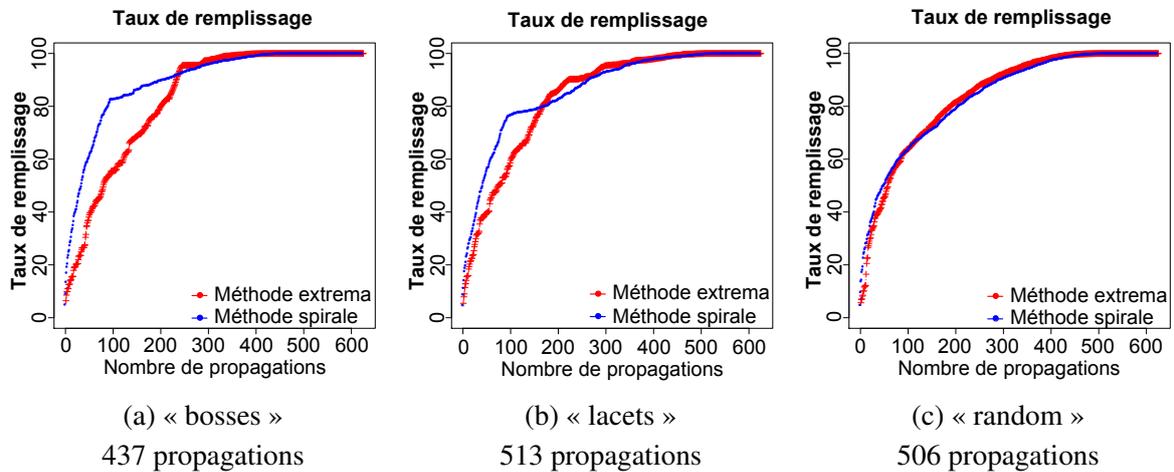


FIG. 4.24 – Comparaison des taux de remplissage, du tableau de paires de distances, entre la méthode spirale (en rouge) et la méthode extrema (en bleu) pour les images « bosses », « lacets » et « random ». Le nombre de propagations nécessaires pour remplir le tableau par la méthode extrema est indiqué sur la ligne du bas.

4.3.5 Méthode taux de remplissage

Plutôt que de choisir les points selon leur distance géodésique au centroïde géodésique, on sélectionne en priorité les points les moins remplis. Pour ce faire, on calcule après chaque propagation géodésique le taux de remplissage. On sélectionne ensuite le point le moins rempli comme source de la propagation. S'il y a plusieurs points parmi les moins remplis, on prend le plus grand extremum géodésique parmi ces points. La méthode de choix des sources propagations géodésiques est décrite dans l'algorithme 13.

Algorithme 13 Méthode du taux de remplissage

- 1: Soient D le tableau de distances de taille $P \times P$
 - 2: Calculer la liste décroissante des extrema géodésiques Ext
 - 3: Initialiser à zéro la liste, de taille P , des taux de remplissage des points T^{ux}
 - 4: **Tant que** D n'est pas plein **Faire**
 - 5: $B \leftarrow \{\operatorname{argmin}_{x \in E} Rempl[x]\}$
 - 6: **Si** $Card\{B\} > 1$ **Alors**
 - 7: $s \leftarrow \operatorname{argmin}_{x \in B} Ext[x]$
 - 8: **Sinon**
 - 9: $s \leftarrow \operatorname{argmin}_{x \in E} T^{ux}[x]$
 - 10: **Fin Si**
 - 11: calculer l'arbre de propagation géodésique depuis s
 - 12: remplir le tableau D
 - 13: retirer les points de Ext qui sont remplis
 - 14: mettre à jour la liste des taux de remplissage T^{ux}
 - 15: **Fin Tant que**
-

Le remplissage pour l'image « lacets » est présenté en superposition de cette image sur la figure 4.25. On compare également les taux de remplissage de cette méthode dite « taux de remplissage » à ceux de la méthode « spirale » (fig. 4.25). On fait de même pour le nombre de propagations nécessaires pour remplir le tableau de distances D (tableau 4.26). On constate que la méthode par taux de remplissage est bien meilleure que la méthode spirale sur toutes les images. Il est important de remarquer que même pour l'image « random », pour laquelle en pratique on ne calcule pas de propagation géodésique, notre méthode est meilleure.

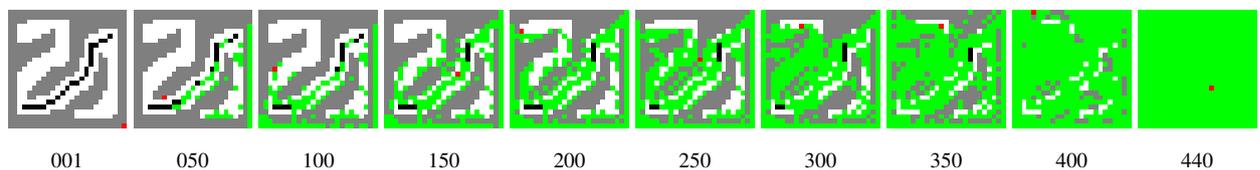


FIG. 4.25 – Quelques images du remplissage (pixels verts) de l'image « lacets » pour différents points sources (en rouge) par la méthode des taux de remplissage.

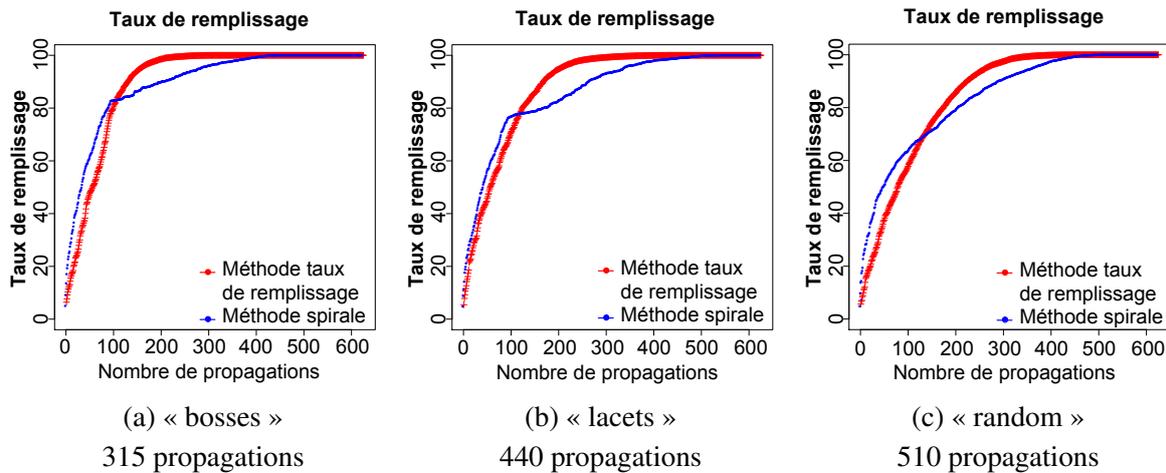


FIG. 4.26 – Comparaison des taux de remplissage, du tableau de paires de distances, entre la méthode spirale (en rouge) et la méthode par taux de remplissage (en bleu) pour les images « bosses », « lacets » et « random ». Le nombre de propagations nécessaires pour remplir le tableau par la méthode par taux de remplissage est indiqué sur la ligne du bas.

	Méthode extrema	Méthode naïve	Écart (Taux remplissage-naïve)	Méthode spirale	Écart (Taux remplissage-spirale)
« Bosses »	315	625	49,6 %	465	32,3 %
« Lacets »	440	625	29,6 %	541	18,7 %
« Random »	510	625	18,4 %	524	2,7 %

TAB. 4.3 – Comparaison du nombre de propagation nécessaires pour remplir le tableau des distances par la méthode par taux de remplissage, par la méthode naïve et par la méthode spirale.

4.3.6 Méthode spirale avec répulsion

La méthode spirale donne quand même de bons résultats. En outre, deux points voisins ont une forte probabilité d'avoir des arbres géodésiques très proches l'un de l'autre. Par conséquent, nous avons proposé une méthode avec un phénomène de répulsion entre les points tirés aléatoirement en spirales (algorithme 14). Divers essais nous ont montré qu'une distance de répulsion de trois pixels de part et d'autre de chaque point donne les meilleurs taux de remplissage.

Comme pour les autres méthodes on présente les remplissages pour quelques propagations géodésiques sur l'image « lacets » (fig. 4.27). On mesure également le taux de remplissage de la méthode spirale avec répulsion par rapport à celui de la méthode spirale (fig. 4.28). On fait de même pour le nombre de propagations nécessaires pour remplir la tableau de distances D (tableau 4.4). A partir de ces différents résultats, on constate que la répulsion entre les sources de propagation améliore le taux de remplissage.

Algorithme 14 Méthode spirale avec répulsion

-
- 1: Soit D le tableau de distances de taille $P \times P$
 - 2: Soit h la distance de répulsion : $h \leftarrow 3$ pixels
 - 3: **Tant que** D n'est pas plein **Faire**
 - 4: Prendre la spire la plus extérieure non remplie
 - 5: Déterminer S la liste des points non remplis de la spire
 - 6: **Tant que** S n'est pas vide **Faire**
 - 7: choisir un point source s aléatoirement dans S
 - 8: retirer de la liste S les deux points à gauche de s et les deux points à droite de s s'il sont toujours dans S
 - 9: calculer l'arbre de propagation géodésique depuis s
 - 10: remplir le tableau D
 - 11: retirer les points de S qui sont remplis
 - 12: **Fin Tant que**
 - 13: **Fin Tant que**
-

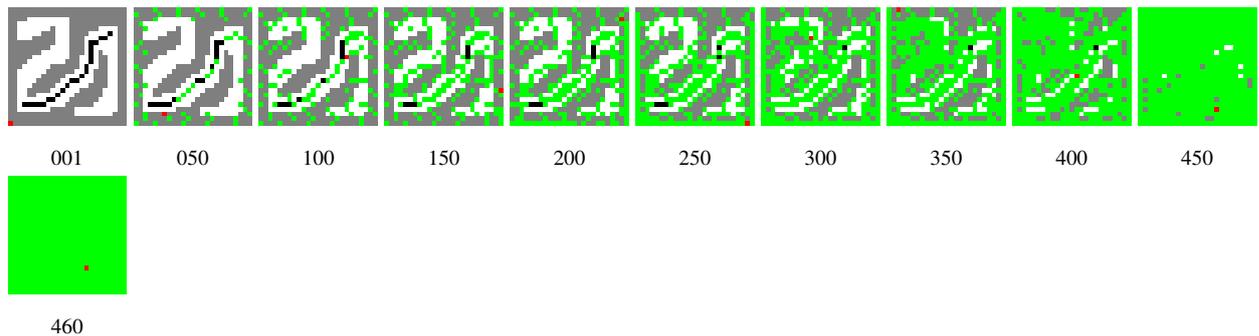


FIG. 4.27 – Quelques images du remplissage (pixels verts) de l'image « lacets » pour différents points sources (en rouge) par la méthode spirale avec répulsion de trois pixels.

	Méthode extrema	Méthode naïve	Écart (spirale répulsion-naïve)	Méthode spirale	Écart (spirale répulsion-spirale)
« Bosses »	384	625	38,6 %	465	17,4 %
« Lacets »	460	625	26,4 %	541	15,0 %
« Random »	496	625	20,6 %	524	5,3 %

TAB. 4.4 – Comparaison du nombre de propagation nécessaires pour remplir le tableau des distances par la méthode spirale avec répulsion de trois pixels, par la méthode naïve et par la méthode spirale.

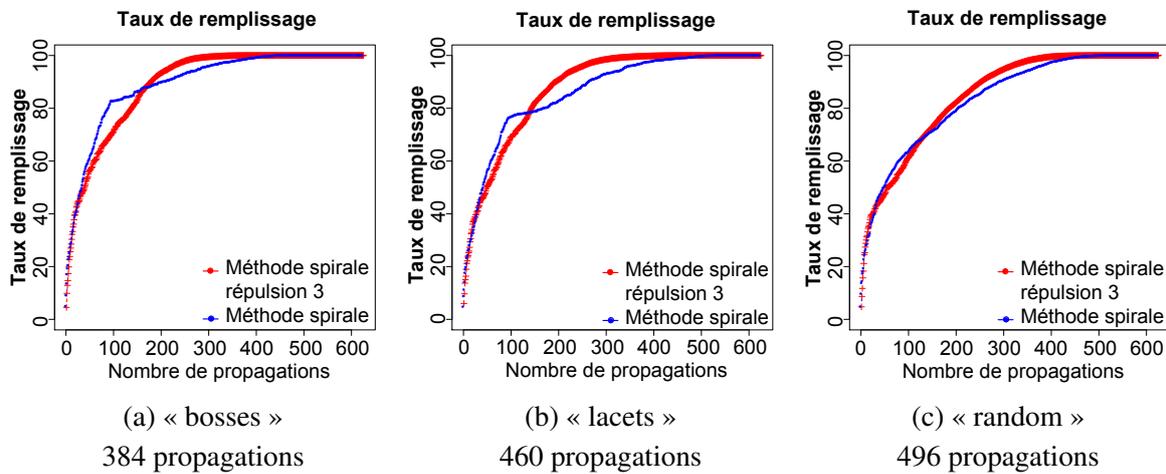


FIG. 4.28 – Comparaison des taux de remplissage, du tableau de paires de distances, entre la méthode spirale (en rouge) et la méthode spirale avec répulsion de trois pixels (en bleu) pour les images « bosses », « lacets » et « random ». Le nombre de propagations nécessaires pour remplir le tableau par la méthode spirale avec répulsion est indiqué sur la ligne du bas.

4.3.7 Comparaison des méthodes

Après avoir présenté et testé différentes méthodes de remplissage du tableau de distances, comparons les entre elles. Pour ce faire, on compare les taux de remplissage sur les images « Bosses », « Lacets » et « random » (fig. 4.29) pour les méthodes :

- spirale ;
- extrema ;
- taux de remplissage ;
- spirale avec répulsion de 3 pixels.

On en conclut que la méthode par taux de remplissage est la meilleure, sur les différentes images, pour remplir le tableau des paires de distances géodésiques. En effet, selon les images on diminue de 18 % à 50 % le nombre de propagations géodésiques par rapport à la méthode naïve (cf. tableau 4.3, p : 179). En outre on améliore nettement la méthode spirale de *Bertelli et al.* [24] puisqu'on arrive à gagner jusqu'à 32 % de propagations géodésiques par rapport à leur méthode.

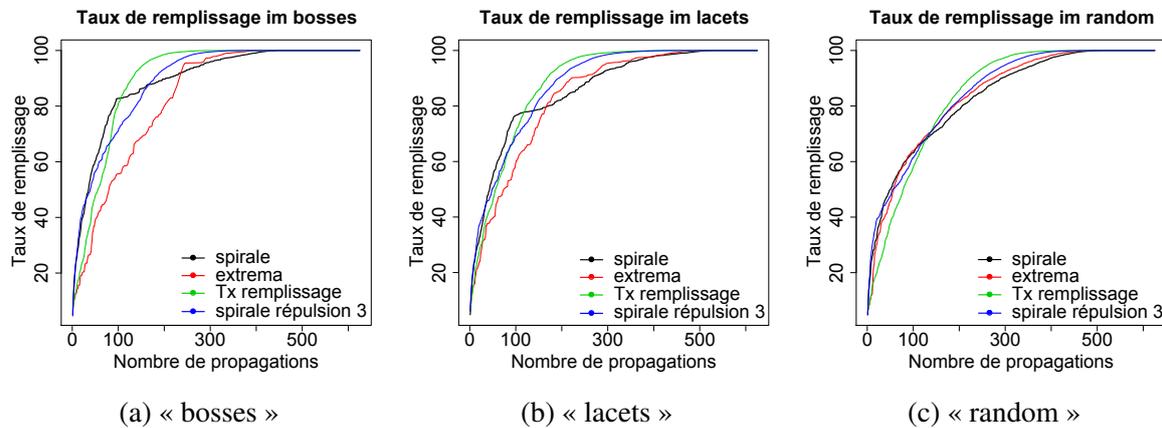


FIG. 4.29 – Comparaison des taux de remplissage, du tableau de paires de distances, pour les méthodes : spirale, extrema, taux de remplissage et spirale avec répulsion de 3 pixels pour les images « bosses », « lacets » et « random ».

4.4 Conclusion du chapitre

Dans ce chapitre deux points ont été abordés : l'introduction de nouvelles connexions η et μ et la conception d'une méthode de calcul efficace des paires de distances géodésiques.

Les nouvelles connexions η et μ permettent de palier le problème de segmentation des zones quasi-plates ; à savoir de faibles variations d'amplitudes locales conduisent à des fortes variations régionales. À partir d'un paramètre λ , non critique, des zones quasi-plates, choisi suffisamment grand, on introduit un paramètre de contrôle régional de la taille des régions. Ce paramètre renforce considérablement les zones quasi-plates. Ces nouvelles segmentations en régions η -bornées ou en boules μ -géodésiques, peuvent servir de pré-segmentation pour obtenir des marqueurs pertinents pour la Ligne de Partage des Eaux. Dans le cas des images hyperspectrales, ces segmentations tiennent compte de l'information spatiale, par les paramètres locaux et régionaux, et de l'information spectrale par les distances entre spectres.

En outre, un autre point particulièrement intéressant a été traité : la création d'une méthode de calcul efficace du tableau de paires de distances géodésiques. Sur des petites images de taille 25×25 pixels, on arrive à gagner jusqu'à 50% de propagations géodésiques par rapport à une méthode naïve. Pour ce faire, on exploite la redondance des propagations géodésiques grâce à l'utilisation d'un arbre géodésique. Pour des images plus grandes, Bertelli *et al.* [24] ont montré que les résultats s'améliorent avec les méthodes d'exploitation de la redondance des arbres géodésiques. Notre méthode qui est meilleure que la leur, dite « spirale » (jusqu'à 32 % de calculs en moins), pourra donc permettre le calcul efficace de tableaux de plus grandes dimensions.

Les tableaux de paires de distances géodésiques pourraient en effet être utilisés pour des méthodes d'analyse de données par tableau de masses et distances [21]. On pourrait ainsi envisager de faire de la réduction de données spatio-spectrale en combinant des tableaux de paires de distances géodésiques et de distances spectrales. En outre, les tableaux de paires de distances géodésiques sont également utilisées dans la segmentation par coupes de graphes, e.g. [175].

Chapitre 5

Conclusion

Ce dernier chapitre constitue la conclusion de ce travail et montre les perspectives de développements futurs. Dans ce mémoire, nous avons présenté les avancées les plus récentes en segmentation morphologiques hyperspectrales. Nous allons maintenant faire la synthèse des différentes méthodes et présenter leurs apports pour la construction de l'analyse des images hyperspectrales par morphologie mathématique.

5.1 Apports de cette thèse

Cette thèse constitue la mise en place des bases du traitement morphologiques des images hyperspectrales et plus particulièrement de leur segmentation. Une chaîne complète de segmentation morphologique hyperspectrale a ainsi été introduite. Cette chaîne se compose de plusieurs étapes.

La première partie consiste à filtrer le bruit spectral de l'image, à réduire sa dimension et à effectuer une classification. De manière plus détaillée nous avons procédé de la manière suivante :

- après avoir introduit les notations et hypothèses de calcul indispensables, nous avons fait l'état de l'art sur la réduction de dimension ;
- afin de nous assurer de la robustesse de nos algorithmes d'analyse et de traitement, nous avons fait un choix d'images hyperspectrales diversifiées : des images acquises dans domaine des longueurs d'ondes, des séquences temporelles d'imagerie résonance magnétique dynamique (DCE-MRI), des images thermiques et des images satellites ;
- la première étape du traitement est la réduction de dimension de l'image et le filtrage spectral du bruit. Nous avons ainsi développé une méthode de débruitage, à la fois pertinente et rapide, par composition de deux AFC-reconstitutions. Cette méthode de débruitage spectral, qui conserve les contours spatiaux tout en réduisant le bruit, offre l'avantage d'avoir des contours les plus nets possibles en vue du traitement morphologique ;
- après avoir débruité l'image, nous avons présenté des méthodes de réduction de dimension par Analyse Factorielle des Correspondances et par Analyse en Composantes principales afin de normaliser les données ainsi qu'une méthode de modélisation du spectre. Les canaux de l'image sont ainsi transformés pour obtenir un espace de dimension réduite composé de facteurs pixels (i.e. vecteurs propres de l'AFC ou de l'ACP) ou des cartes des paramètres du modèle qui constituent une nouvelle représentation de l'image hyperspectrale de départ. Par ailleurs, une

nouvelle méthode de sélection des axes factoriels par leur rapport signal à bruit a été développée. Elle s'avère être particulièrement intéressante car elle prend en compte l'information spatiale portée par les axes et non plus un critère statistique.

- à partir des canaux de ces images hyperspectrales de dimension réduite on effectue une classification spectrale non supervisée (*k-means* ou *clara*) ou supervisée (LDA) à partir d'un ensemble d'apprentissage. La classification supervisée donnant de meilleurs résultats, pour l'appliquer à différentes images à partir d'un apprentissage effectué sur une seule image hyperspectrale, nous avons proposé une méthode par normalisation d'histogrammes des cartes de paramètres du modèle ajusté. Cette méthode s'est révélée être la plus robuste. Néanmoins le traitement morphologique ne peut se contenter de classes de spectres non connexes. En outre, seule l'information spectrale a, jusqu'à présent, été prise en compte ;
- nous avons également proposé une méthode de réduction spatio-spectrale de bruit par AFC sur le voisinage spatial des facteurs pixels.

Après ces étapes de filtrage spectral, de réduction de dimension et de classification spectrale, nous avons développé la segmentation morphologique des images hyperspectrales par Ligne de Partage des Eaux (LPE) afin d'obtenir des régions connexes. Plusieurs études ont été effectuées :

- tout d'abord nous présentons une méthodologie de segmentation des images hyperspectrales par LPE. L'approche par LPE nécessite en effet une fonction à inonder scalaire (i.e. une image à niveaux de gris) et des marqueurs. Pour obtenir ces marqueurs, nous avons montré qu'il est possible d'utiliser les classes de la classification spectrale précédemment obtenues après quelques traitements mineurs (réduction de leur surface et ré-étiquetage) ;
- nous avons également montré qu'une fonction à inonder représentée par un gradient marginal ou vectoriel nécessite une distance spectrale adaptée à l'espace image de représentation ;
- nous avons fait clairement apparaître que pour les images hyperspectrales, la LPE probabiliste est bien adaptée si l'on construit des densités de probabilités de contours (pdf) pour ces images à partir de germes tirés aléatoirement ; ceci pour obtenir plusieurs réalisations de contours par LPE. Ces réalisations permettent ensuite d'estimer la densité de contours par des méthodes à noyau. L'estimation effectuée est donc de type Monte-Carlo ;
- nous avons également introduit un tirage de germes boules aléatoires conditionnés par la classification spectrale. Dans ces conditions, la densité de probabilité de contours (pdf) obtenue prend donc à la fois en compte l'information spatiale et l'information spectrale ;
- la pdf présente les contours de régions selon leur importance. Les plus importantes et les plus contrastées ont des probabilités élevées (i.e. niveaux de gris proches du blanc) et les plus faiblement contrastées ont des probabilités beaucoup plus faibles (i.e. niveaux de gris proches du noir). La segmentation de cette fonction est ensuite effectuée par LPE sur un critère morphologique permettant de sélectionner un nombre a priori de régions selon le volume du bassin versant associé. Nous avons également développé une approche plus robuste que le critère du nombre de régions. Il en effet particulièrement intéressant d'utiliser des marqueurs issus de la classification spectrale, pour choisir les régions, puisque seul le nombre de classes est requis. Ce critère est plus robuste que le nombre de régions car il est le même pour des images similaires présentant le même nombre de classes spectrales alors que ce n'est pas le cas pour le nombre de régions. Cette méthode de segmentation nous semble donc la plus aboutie pour la segmenta-

tion hyperspectrale car elle permet d'obtenir de manière automatique une segmentation sur des images similaires ;

- cette méthode a été illustrée de manière très détaillée sur les séries temporelles DCE-MRI afin de fournir une segmentation en vue de la détection de tumeurs cancéreuses sur des souris. Les régions de la segmentation sont en effet classifiées pour détecter des zones potentiellement tumorales.
- finalement l'approche globale de segmentation des images hyperspectrales est résumée depuis les étapes de filtrage et de réduction de dimension.

Dans un autre étape, nous avons introduit de nouvelles méthodes de segmentations, par connexions, qui sont étendues aux images hyperspectrales :

- afin d'améliorer la segmentation des zones quasi-plates qui ne prennent en compte que l'information locale, nous avons introduit deux nouvelles connexions utilisant l'information régionale. Ces connexions conduisent aux régions η -bornées et aux boules μ -géodésiques qui sont des méthodes spatio-spectrales de segmentation. Ces segmentations ne sont pas aussi régulières que celles obtenues par la LPE probabiliste. C'est pourquoi ces connexions pourraient être utilisées pour obtenir des pré-segmentations pouvant servir comme marqueurs pour la LPE.
- enfin nous avons présenté une méthode de calcul très efficace des paires de distances géodésiques entre les pixels d'une image puisqu'elle permet de réduire de 50 % le nombre d'opérations par rapport à une approche naïve et qu'elle permet une diminution d'opérations jusqu'à 30 % par rapport à une méthode précédemment existante.

5.2 Perspectives

Intéressons nous maintenant aux perspectives créées par ce travail.

Tout d'abord nous avons répondu aux objectifs fixés puisqu'une méthodologie complète de segmentation a été proposée depuis le débruitage et la réduction de dimension en passant par la classification spectrale jusqu'à la segmentation morphologique par LPE probabiliste. Cette segmentation utilise une pdf dont le tirage des marqueurs peut être conditionné par la classification spectrale. La pdf est calculée pour le moment en utilisant un noyau spatial. Il serait intéressant d'essayer un noyau spatio-spectral sur ces images. En outre à partir de segmentation manuelle, il pourrait être envisageable de développer une méthode d'apprentissage de la densité de probabilité de contours *pdf*.

Au niveau de la classification, des classificateurs spatio-spectraux pourraient être utilisés en veillant bien à l'équilibre entre l'amélioration des résultats et l'augmentation inévitable du temps de calcul.

D'autres perspectives sont à explorer du côté de la segmentation par connexions η et μ puisqu'on pourrait imaginer de sélectionner les paramètres η et μ de manière locale et non plus globale.

Pour la réduction de dimensions d'autres méthodes non linéaires pourraient être testées en prenant en compte l'amélioration réelle apportée par rapport aux méthodes linéaires en vue de la segmentation morphologique.

Grâce à notre méthode de calcul du tableau de paires de distances géodésiques, de nouvelles perspectives semblent maintenant envisageables. Il sera en effet possible d'effectuer une réduction de dimension spatio-spectrale en utilisant une combinaison linéaire de la distance géodésique et d'une distance spectrale dans une analyse par tableau de masses et distances [21]. Cette analyse semble très

prometteuse à cause de l'utilisation de ces deux informations et notamment de l'information spatiale sur toute l'image et non pas sur un voisinage local.

Une autre méthode d'analyses de données semble également très intéressante : l'Analyse des Correspondances d'un tableau de contingence dont les lignes et les colonnes sont munies d'une structure de graphe bistochastique [131]. On peut en effet introduire un graphe sur le voisinage de chacun des pixels et donc ajouter de l'information spatiale de manière locale.

La perspective à venir la plus importante est certainement le développement d'opérateurs de morphologie mathématique multivariés (érosion, dilatation, ...) qui soient réellement pertinents. Pour cela il sera nécessaire de définir une relation d'ordre entre les pixels vecteurs dans l'idée de ce qui a été fait par *Angulo* [3] pour les images couleurs. La réduction de dimension aura là toute son importance et pourra permettre de d'obtenir des espaces bien adaptés à la définition d'un ordre total entre les vecteurs.

Comme on l'aura compris, le champ à explorer pour la création de la morphologie mathématique multivariée reste important et prometteur. Nous espérons que cette thèse aura permis de poser les bases de la morphologie mathématique sur les images hyperspectrales et plus particulièrement de la segmentation morphologique.

Annexe A

Filtrage, réduction de dimension et classification

Cette annexe se veut un complément détailler du chapitre 2. Y sont ainsi présentés : une introduction à l'analyse de données, puis l'Analyse Factorielle sur un espace euclidien qui est une formulation plus générale de l'Analyse en Composantes Principales (ACP) et de l'Analyse Factorielle des Correspondances (AFC). Ensuite nous présentons des compléments sur l'AFC par rapport à la réduction de bruit et notamment la méthode que nous avons développée : sur la composition d'AFC et de reconstitution pour le filtrage spectral. Nous donnons ensuite des précisions sur les rapports signaux à bruits calculés à partir de la covariance.

A.1 Introduction à l'analyse de données

Nous introduirons l'Analyse Factorielle (AF) comme une analyse sur un tableau muni de masses et de distances dans un espace euclidien. Nous utiliserons les résultats de l'AF pour en déduire ceux de l'Analyse en Composantes Principales (ACP) et ceux de l'Analyse Factorielle des Correspondances (AFC) en spécifiant la distance et la masse utilisée en chaque point.

La réduction de données pour la dimension spectrale, consiste à appliquer une transformation sur l'espace image à L dimensions (i.e. canaux) pour obtenir un autre espace de représentation à K dimensions avec $K < L$ et souvent $K \ll L$.

Notons \mathbf{F} la matrice $P \times L$ de représentation de l'image \mathbf{f}_λ . \mathbf{F} est composée de P lignes représentant les pixels vecteurs (i.e. les individus) et de L colonnes représentant les canaux (i.e. les variables). En termes mathématiques, la dimension intrinsèque signifie que les individus dans le jeu de données \mathbf{F} sont situés, ou presque situés, dans un sous espace de dimension K inclus dans l'espace de dimension L .

Pour alléger la notation, on s'autorise pour les calculs matriciels à utiliser la notation :

$$f_{ij} = f_{\lambda_j}(x_i) \quad f_j = f_{\lambda_j} \quad f_i = \mathbf{f}_\lambda(x_i) \quad (\text{A.1.1})$$

On considère un tableau \mathbf{F} à P lignes et L colonnes, dont les éléments sont constitués des valeurs (i.e. des nombres réels) de l'image \mathbf{f}_λ (eq. A.1.2). Les colonnes de ce tableaux correspondent aux

canaux de l'image qui ont été transformés de matrices 2D en vecteurs 1D.

$$\mathbf{F} = \begin{bmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1L} \\ \vdots & & \vdots & & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{iL} \\ \vdots & & \vdots & & \vdots \\ f_{P1} & \cdots & f_{Pj} & \cdots & f_{PL} \end{bmatrix} \quad (\text{A.1.2})$$

où $f_{ij} \in \mathbb{R}, \forall i = 1 \dots P, \forall j = 1 \dots L$.

Les vecteurs-lignes (i.e. vecteurs pixels) de \mathbf{F} sont les éléments de \mathcal{T}^L et sont assimilés à des vecteurs de \mathbb{R}^L :

$$f_i = (f_{i1}, \dots, f_{ij}, \dots, f_{iL}) \quad (\text{A.1.3})$$

avec $f_i = \mathbf{f}_\lambda(x_i)$.

Les vecteurs-colonnes (i.e canaux) de \mathbf{F} sont les éléments de \mathbb{R}^P et sont assimilés à des vecteurs de \mathbb{R}^P :

$$f_j = (f_{1j}, \dots, f_{ij}, \dots, f_{Pj}) \quad (\text{A.1.4})$$

avec $f_j = f_{\lambda_j}$.

Dans cette partie nous présenterons l'Analyse Factorielle sur un espace euclidien. Puis nous montrerons que l'Analyse en Composantes Principales et l'Analyse Factorielle des Correspondances s'en déduisent grâce à des poids affectées à chacun des points de la matrice \mathbf{F} qui correspondent au passage de la distance « inverse des variances » en ACP, ou de la métrique du χ^2 en AFC, à la métrique euclidienne.

Ainsi dans [22], *Benzécri*, indique que l'ACP et l'AFC se déduisent de l'Analyse Factorielle sur un espace euclidien. Mais alors qu'en AFC on représente la colonne f_{λ_j} de la matrice \mathbf{F} par son profil (en divisant chaque nombre de la colonne par le total de la colonne), l'usage en ACP est d'amener chaque colonne à avoir moyenne nulle et variance 1. C'est à dire que les colonnes f_{λ_j} (i.e. les canaux) sont centrées réduites.

A.2 Analyse factorielle sur un espace euclidien

Cette analyse est décrite théoriquement dans [132]. Nous résumons ici cette approche.

Les méthodes d'analyse factorielle présentée dans ces annexes : Analyse en Composantes Principales (ACP), Analyse Factorielle des Correspondances (AFC) sont toutes fondées sur la même démarche : « l'Analyse Factorielle d'un nuage de points ». Étant donné un ensemble de P points dans une espace vectoriel \mathbb{R}^L , munis chacun d'un poids, on cherche les « axes d'inertie » du nuage de points, qui sont tels que les projections des points sur les premiers axes factoriels vont fournir les « meilleures » représentations du nuage dans des sous-espaces de faible dimension.

L'ACP et l'AFC se différencieront par :

- les poids attribués aux points des nuages. Dans cette partie, les poids seront à priori quelconques ;

- la métrique utilisée dans \mathbb{R}^L pour calculer les distances entre points. Toutefois on montrera que, moyennant une transformation préalable des données en ACP et en AFC, on peut toujours se ramener à l'utilisation de la métrique euclidienne canonique.

A.2.1 Notations. L'espace euclidien \mathbb{R}^L

Dans un espace vectoriel \mathbb{R}^L muni du produit scalaire canonique défini par :

$$\forall x = \begin{bmatrix} x^1 \\ \vdots \\ x^L \end{bmatrix} \in \mathbb{R}^L, \forall y = \begin{bmatrix} y^1 \\ \vdots \\ y^L \end{bmatrix} \in \mathbb{R}^L : (x|y) = \sum_{j=1}^L x^j y^j = x^t y \quad (\text{A.2.1})$$

Ce produit scalaire fait de \mathbb{R}^L un espace vectoriel euclidien.

$\|x\| = \sqrt{(x|x)}$ définit la norme euclidienne canonique.

$d(x, y) = \|x - y\| = \sqrt{\sum_{j=1}^L (x^j - y^j)^2}$ définit la distance euclidienne canonique qui fait de \mathbb{R}^L un espace métrique.

On munit \mathbb{R}^L , i.e. l'espace dans lequel sont représentés les individus, de sa base canonique $(e_1 \dots e_j \dots e_L)$. On munit également \mathbb{R}^P , i.e. l'espace dans lequel sont représentés les variables, de sa base canonique $(e'_1 \dots e'_i \dots e'_P)$.

On considère un tableau \mathbf{X} à P lignes et L colonnes, dont les éléments sont constitués de nombres réels (eq. A.2.2).

$$\mathbf{X} = \begin{bmatrix} x_1^1 & \cdots & x_1^j & \cdots & x_1^L \\ \vdots & & \vdots & & \vdots \\ x_i^1 & \cdots & x_i^j & \cdots & x_i^L \\ \vdots & & \vdots & & \vdots \\ x_P^1 & \cdots & x_P^j & \cdots & x_P^L \end{bmatrix} \quad (\text{A.2.2})$$

où $x_i^j \in \mathbb{R}, \forall i = 1 \dots P, \forall j = 1 \dots L$.

Les vecteurs-lignes (i.e. vecteurs pixels) de \mathbf{X} sont les éléments de \mathbb{R}^L :

$$x_i = (x_i^1, \dots, x_i^j, \dots, x_i^L) \quad (\text{A.2.3})$$

Les vecteurs-colonnes (i.e. canaux) de \mathbf{X} sont les éléments de \mathbb{R}^P :

$$x^j = (x_1^j, \dots, x_i^j, \dots, x_P^j) \quad (\text{A.2.4})$$

On munit les P vecteurs lignes $x_1, \dots, x_i, \dots, x_P$ de \mathcal{T}^L de poids $p_1, \dots, p_i, \dots, p_P$ avec $\forall i = 1 \dots P, p_i > 0$ et $\sum_{i=1}^P p_i = 1$.

L'ensemble de ces points affectés de poids constitue un nuage de points N_I :

$$N_I = \{(x_i, p_i)\}_{i=1 \dots P}. \quad (\text{A.2.5})$$

Le centre de gravité est noté :

$$g = \sum_{i=1}^P p_i x_i \quad (\text{A.2.6})$$

A.2.2 Choix du sous-espace

La réduction de dimension consiste à chercher à représenter le nuage N_I dans un espace de plus faible dimension $K < L$.

On note H , de dimension $K < L$, le sous-espace sur lequel on va projeter le nuage. On note $\hat{x}_i = P_H(x_i)$ la projection orthogonale de x_i sur H : \hat{x}_i est l'unique point de H tel que $\hat{x}_i - x_i$ soit orthogonal à H , et vérifie : $d(x_i, \hat{x}_i) = \ll \text{distance} \gg$ du point x au sous-espace H .

On va chercher le sous-espace H « le plus proche du nuage » (critère de proximité). Pour cela, on souhaite que les distances $d(x_i, \hat{x}_i)$ entre les points du nuage et leurs projections soient les plus petites possibles. On va choisir comme critère la minimisation du moment d'inertie du nuage N_I par rapport au sous-espace H :

$$M^{\text{ent}}(H) = \sum_{i=1}^P p_i d^2(x_i, \hat{x}_i) \quad (\text{A.2.7})$$

On montre, en utilisant le théorème de Huygens [21], que H passe nécessairement par le centre de gravité g du nuage N_I . On montre également que le critère de proximité est équivalent à un critère de fidélité des distances. Ce dernier critère fait que H est tel que les distances entre les points projetés sur H sont les plus proches possibles des distances entre les points du nuage N_I .

A.2.3 Inertie expliquée, inertie résiduelle

On supposera dorénavant que le centre de gravité du nuage est confondu avec l'origine : $g = 0$ (ce qui revient à faire un changement d'origine dans \mathbb{R}^P) ; le nuage est dit centré.

Pour $K (\leq L)$ fixé, on va chercher le sous-espace affine F_K de dimension K passant par $g = 0$ (assimilé à un sous-espace vectoriel) minimisant $M^{\text{ent}}(H)$:

$$\forall H \text{ sous-espace affine de dimension } K, M^{\text{ent}}(H) \geq M^{\text{ent}}(F_K)$$

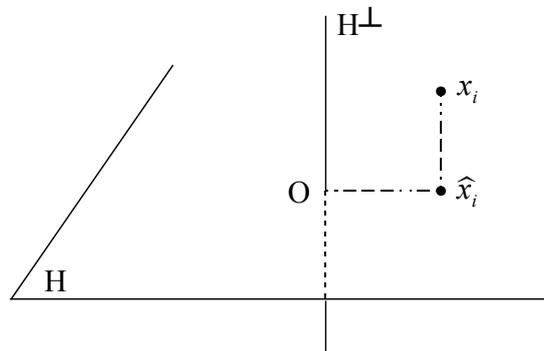


FIG. A.1 – Projection d'un point x_i sur le sous-espace H . Le projeté de x_i sur H est noté \hat{x}_i . $H^\perp =$ est le sous-espace orthogonal à H .

Définition 19. On appelle inertie expliquée par H la quantité :

$$I(H) = \sum_{i=1}^P p_i \|\hat{x}_i\|^2 \quad (\text{A.2.8})$$

L'inertie résiduelle autour de H est la quantité :

$$\sum_{i=1}^P p_i \|x_i - \hat{x}_i\|^2 \quad (\text{A.2.9})$$

L'inertie totale du nuage par rapport à son centre de gravité, est la quantité :

$$I_0 = I_g = \sum_{i=1}^P p_i \|x_i\|^2 \quad (\text{A.2.10})$$

D'après le théorème de Pythagore, on a :

$$\forall i = 1 \dots P \quad \|x_i\|^2 = \|\hat{x}_i\|^2 + \|x_i - \hat{x}_i\|^2 \quad (\text{A.2.11})$$

D'où la relation :

$$\text{Inertie totale} = \text{Inertie expliquée} + \text{Inertie résiduelle}$$

Si on note $\hat{N}(P) = \{(\hat{x}_i, p_i)\}_{1 \leq i \leq P}$ le nuage projeté sur H, qui a aussi l'origine pour centre de gravité, alors :

- l'inertie expliquée par H est l'inertie totale de \hat{N} par rapport à O ; c'est aussi le moment d'inertie de N par rapport à H^\perp ;
- l'inertie résiduelle est l'inertie par rapport à O du nuage projeté sur H^\perp ; c'est aussi le moment d'inertie de N par rapport à H.

Le critère choisi pour la détermination de F_K est donc la minimisation de l'inertie résiduelle, ou de façon équivalente, la maximisation de l'inertie expliquée : on cherche à ce que le nuage projeté sur F_K ait une dispersion maximale autour de l'origine O.

A.2.4 Expression matricielle de l'inertie

Soit Δu une droite passant par l'origine (qui est le centre de gravité du nuage) engendrés par le vecteur normé u ($\|u\| = 1$).

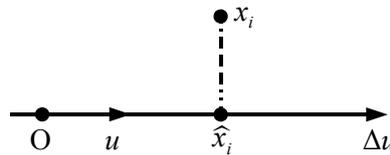


FIG. A.2 – Projection \hat{x}_i d'un point x_i sur la droite Δu

La projection orthogonale d'un point x_i du nuage sur la droite Δ_u vaut :

$$\hat{x}_i = (x_i|u)u = (x_i^t u)u$$

– pour déterminer le plan F_2 , le lemme indique que :

a) F_2 contient F_1 : on peut donc écrire $F_2 = \Delta u_1 \oplus \Delta u_2$, où le vecteur normé u_2 est orthogonal à u_1

b) u_2 est solution du problème :

$$\max_{u \in \mathbb{R}^L, u \perp u_1} I(\Delta u)$$

– etc ...

Les droites $\Delta u_1, \Delta u_2, \dots, \Delta u_L$ sont appelées 1^{er}, 2^{ème}, ..., $p^{\text{ème}}$ axe principal d'inertie du nuage.

A.2.5.2 Détermination du premier axe principal d'inertie

On cherche un vecteur u de \mathbb{R}^L , normé, tel que l'inertie expliquée par la droite Δ_u soit maximale. Cette inertie s'écrit :

$$I(\Delta u) = u^t \mathbf{V} u$$

D'après une propriété classique démontrée dans tout bon ouvrage d'algèbre linéaire, la matrice \mathbf{V} de \mathbb{R}^L , symétrique, est diagonalisable, et il existe une base orthonormée $(u_1 \dots u_k \dots u_L)$ de \mathbb{R}^L formée de vecteurs propres de \mathbf{V} :

$$\forall k = 1 \dots p \quad \mathbf{V} u_k = \mu_k u_k \quad (\text{A.2.18})$$

On numérote les u_k de façon que : $\mu_1 \geq \dots \geq \mu_k \geq \dots \geq \mu_N$.

Alors, $I(\Delta u)$ est maximale pour $u = u_1$, et on a :

$$I(\Delta u_1) = \mu_1 \quad (\text{A.2.19})$$

Remarque A.2.1. u_1 est un vecteur normé arbitraire dans le sous-espace propre associé à la plus grande valeur propre μ_1 de \mathbf{V} ; en particulier u_1 est défini au signe près si μ_1 est d'ordre de multiplicité égal à 1.

A.2.5.3 Détermination du $k^{\text{ème}}$ axe principal d'inertie

On raisonne par récurrence : on suppose que les $k - 1$ premiers axes principaux sont engendrés respectivement par $u_1 \dots u_{k-1}$. On cherche un vecteur u de \mathbb{R}^L , normé, orthogonal à $u_1 \dots u_{k-1}$, tel que $I(\Delta u)$ soit maximale.

Alors, $I(\Delta u)$ est maximale pour $u = u_k$, et on a :

$$\forall k = 1 \dots L \quad I(\Delta u_k) = \mu_k \quad (\text{A.2.20})$$

Remarque A.2.2. Toute inertie étant positive ou nulle, on en déduit que $\mu_k \geq 0 \forall k = 1 \dots L$ (cela vient de ce que \mathbf{V} est une matrice positive : $\forall u \in \mathbb{R}^L \quad u^t \mathbf{V} u \geq 0$).

A.2.5.4 Bilan

Le sous-espace vectoriel de dimension k qui explique le mieux l'inertie du nuage, appelé « sous-espace principal de dimension k », est :

$$F_k = \Delta u_1 \oplus \dots \oplus \Delta u_k \quad (\text{A.2.21})$$

avec $I(F_k) = \mu_1 + \dots + \mu_k$

Si $k = L$, on obtient : $\mathbb{R}^L = \Delta u_1 \oplus \dots \oplus \Delta u_L$.

L'inertie totale du nuage par rapport à l'origine est l'inertie expliquée par \mathbb{R}^L :

$$I_0 = \sum_{\alpha=1}^L \mu_\alpha = \text{Tr}(\mathbf{V}) \quad (\text{A.2.22})$$

On appelle part d'inertie expliquée par l'axe principal Δu_α la quantité :

$$\tau_\alpha = \frac{I(\Delta u_\alpha)}{I_0} = \frac{\mu_\alpha}{\sum_{j=1}^L \mu_j} \quad (\text{A.2.23})$$

La part d'inertie expliquée par le sous-espace principal F_k est égale à :

$$\frac{I(F_k)}{I_0} = \frac{\sum_{\alpha=1}^k \mu_\alpha}{\sum_{j=1}^L \mu_j} \quad (\text{A.2.24})$$

A.2.6 Facteurs principaux

Définition 21. On appelle α -ème facteur principal (ou facteur pixel) le vecteur c^α de \mathbb{R}^P dont les composantes sont les coordonnées c_i^α des points x_i de \mathbb{R}^L sur le α -ème axe principal Δu_α .

$$c^\alpha = \begin{bmatrix} c_1^\alpha \\ \vdots \\ c_i^\alpha \\ \vdots \\ c_n^\alpha \end{bmatrix} = \begin{bmatrix} x_1^t u_\alpha \\ \vdots \\ x_i^t u_\alpha \\ \vdots \\ x_P^t u_\alpha \end{bmatrix} = \begin{bmatrix} x_1^t \\ \vdots \\ x_i^t \\ \vdots \\ x_P^t \end{bmatrix} u_\alpha = \mathbf{X} u_\alpha \in \mathbb{R}^P \quad (\text{A.2.25})$$

avec $c_i^\alpha = (x_i | u_\alpha) = x_i^t u_\alpha$ les coordonnées de x_i sur Δu_α .

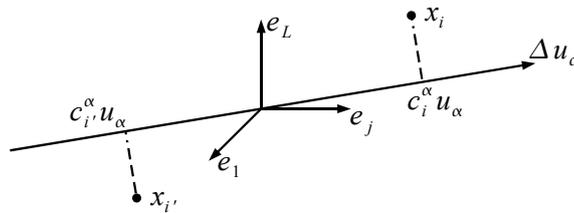


FIG. A.3 – Projection $c_i^\alpha u_\alpha$ d'un point x_i sur l'axe factoriel Δu_α .

Remarque A.2.3. L'« analyse factorielle » d'un nuage de points doit son nom à ce qu'elle peut être présentée comme étant la recherche des « facteurs principaux », associés aux axes principaux (appelés aussi « axes factoriels »), du nuage.

On montre [132] que les facteurs principaux ont les propriétés suivantes :

$$\begin{aligned} \sum_{i=1}^P p_i c_i^\alpha &= 0 \\ \sum_{i=1}^P p_i (c_i^\alpha)^2 &= \mu_\alpha \\ \sum_{i=1}^P p_i c_i^\alpha c_i^\beta &= 0 \quad \text{si } \alpha \neq \beta \end{aligned} \quad (\text{A.2.26})$$

A.3 L'Analyse en Composantes Principales

L'Analyse en Composantes Principales a été développée par *H. Hotelling* en 1933 [79, 80].

Les résultats de la partie précédente A.2 vont être appliqués à l'étude du tableau correspondant à la matrice \mathbf{F} après une transformation des données. Les vecteurs pixels correspondent aux individus i et les canaux aux variables j .

A.3.1 Distance entre individus

Un des objectifs de l'ACP est de décrire les ressemblances et les oppositions entre individus (pixels), en essayant de mettre en évidence des groupes homogènes d'individus. Pour évaluer les ressemblances, ou proximités, on munit \mathbb{R}^L de la distance « inverse des variances » :

$$d_{1/\sigma^2}^2(f_i, f_{i'}) = \sum_{j=1}^L \frac{1}{\sigma_j^2} (f_{ij} - f_{i'j})^2 = \sum_{j=1}^L \left(\frac{f_{ij} - m_j}{\sigma_j} - \frac{f_{i'j} - m_j}{\sigma_j} \right)^2 \quad (\text{A.3.1})$$

avec $m_j = E[f_j]$ la moyenne de la variable f_j et $\sigma_j^2 = E[(f_j - m_j)^2]$ sa variance.

Cette distance, qui pondère les quantités $(f_{ij} - f_{i'j})^2$ relatives à une variable f_j par l'inverse de sa variance (d'où sa dénomination), revient à utiliser les variables f_j/σ_j , dites « variables réduites », plutôt que les variables initiales. Elle offre les avantages suivants :

- les variables f_j/σ_j étant sans dimension, les distances ne dépendent pas des unités de mesure
- on utilise de fait comme unité de mesure pour la variable f_j son écart type σ_j ; on donne ainsi la même importance à toutes les variables (la variance de chaque variable réduite vaut 1).

Remarque A.3.1. La distance Mahalanobis s'exprime sous la forme :

$$d_M^2(f_i, f_{i'}) = (f_i - f_{i'})^t \Sigma^{-1} (f_i - f_{i'}) \quad (\text{A.3.2})$$

où Σ est la matrice de covariance entre les variables. La distance « inverse des variances » correspond à la distance de Mahalanobis si les variables (canaux) sont décorrélés. En effet la matrice de covariance est alors diagonale et les valeurs de sa diagonale sont égales à la variance des canaux : σ_j^2 avec $j = 1, \dots, L$. Cependant si l'on veut appliquer une ACP les canaux sont en général corrélés puisque l'un des objectifs de l'ACP est de décorréler des variables.

A.3.2 Transformation des données

Afin de se ramener au cas de l'Analyse Factorielle sur un espace euclidien, et ainsi utiliser la métrique euclidienne canonique, on transforme les données de la manière suivante :

$$\forall i = 1 \dots P \quad \forall j = 1 \dots L \quad x_i^j = \frac{f_{ij} - m_j}{\sigma_j} \quad (\text{A.3.3})$$

Cette opération consiste à centrer et réduire les variables f_j , i.e. à remplacer la variable f_j par la variable centrée (i.e. de moyenne nulle) et réduite (i.e. de variance 1) $\frac{f_j - m_j}{\sigma_j}$.

Chaque individu est maintenant représenté par le vecteur :

$$x_i = \begin{bmatrix} x_i^1 \\ \vdots \\ x_i^j \\ \vdots \\ x_i^L \end{bmatrix} \in \mathbb{R}^L \quad (\text{A.3.4})$$

Le nuage des individus est le nuage de point de \mathbb{R}^L :

$$N_I = \left\{ \left(x_i, \frac{1}{P} \right) \right\}_{i=1 \dots P} \quad (\text{A.3.5})$$

A.3.3 Justification de la transformation des données

Deux raisons justifient la transformation des données :

1. Centrer les variables permet de faire coïncider le centre de gravité du nuage des individus avec l'origine :

$$g = \sum_{i=1}^P \frac{1}{P} x_i = \sum_{i=1}^P \frac{1}{P} \left(\sum_{j=1}^L x_i^j e_j \right) = \sum_{j=1}^L \left(\frac{1}{P} \sum_{i=1}^P x_i^j \right) e_j = \sum_{j=1}^L \bar{x}^j e_j = 0$$

où $\bar{x}^j = \frac{1}{P} \sum_{i=1}^P x_i^j$ est la moyenne de la variable centrée réduite $(f_j - m_j)/\sigma_j$ qui vaut 0.

On appelle individu moyen de la population, un individu qui prendrait, pour chaque variable f_j , une valeur égale à la moyenne m_j de cette variable. L'individu moyen est représenté par le centre de gravité, donc l'origine de \mathbb{R}^L .

2. Réduire les variables permet d'avoir l'équivalence entre la distance euclidienne entre les individus x_i dans un espace euclidien et la distance « inverse des variances » entre les individus dans l'espace image (i.e. les vecteurs pixels) :

$$\begin{aligned} d_E^2(x_i, x_{i'}) &= \sum_{j=1}^L (x_i^j - x_{i'}^j)^2 = \sum_{j=1}^L \left(\frac{f_{ij} - m_j}{\sigma_j} - \frac{f_{i'j} - m_j}{\sigma_j} \right)^2 \\ &= \sum_{j=1}^L \frac{1}{\sigma_j^2} (f_{ij} - f_{i'j})^2 = d_{1/\sigma^2}^2(f_i, f_{i'}) \end{aligned} \quad (\text{A.3.6})$$

A.3.4 Analyse du nuage des individus

La matrice d'inertie s'écrit ici

$$\mathbf{V} = \frac{1}{P} \mathbf{X}^t \mathbf{X} \quad (\text{A.3.7})$$

puisque la matrice diagonale des poids est $D = \frac{1}{P} I$. Elle a pour terme général

$$v_{jj'} = \frac{1}{P} x^{jt} x^{j't} = \frac{1}{\sigma_j \sigma_{j'}} \frac{1}{P} (f_j - m_j)^t (f_{j'} - m_{j'}) = \frac{\text{Cov}(f_j, f_{j'})}{\sigma_j \sigma_{j'}} = \text{Cor}(f_j, f_{j'})$$

avec $\text{Cor}(f_j, f_{j'})$ le coefficient de corrélation et $\text{Cov}(f_j, f_{j'})$ la covariance centrée entre les variables f_j et $f_{j'}$.

\mathbf{V} est donc la matrice des corrélations des variables (canaux) $f_j, j = 1 \dots L$.

L'inertie totale du nuage vaut :

$$I_0 = L \quad (\text{A.3.8})$$

Démonstration A.3.1.

$$I_0 = \text{Tr}(\mathbf{V}) = \sum_{j=1}^L \text{Cor}(f_j, f_j) = \sum_{j=1}^L 1 = L$$

Comme dans l'équation A.2.18, les axes principaux sont les vecteurs propres normés de \mathbf{V} :

$$\mathbf{V} u_\alpha = \mu_\alpha u_\alpha \quad , \quad \alpha = 1 \dots p \quad , \quad \|u_\alpha\| = 1 \quad (\text{A.3.9})$$

De même que dans l'équation A.2.25, les facteurs principaux sont des vecteurs de \mathbb{R}^P :

$$c^\alpha = \begin{bmatrix} c_1^\alpha \\ \vdots \\ c_i^\alpha \\ \vdots \\ c_P^\alpha \end{bmatrix} = \mathbf{X} u_\alpha \quad , \quad \alpha = 1 \dots L \quad (\text{A.3.10})$$

où $c_i^\alpha = x_i^t u_\alpha$ est la coordonnée de l'individu i sur l'axe Δu_α .

Remarque A.3.2. $c^\alpha = 0$ si $\alpha > \text{rang}(\mathbf{X})$

Les composantes principales ont les propriétés suivantes :

Propriété A.3.1. Les facteurs principaux c^α sont des combinaisons linéaires des variables centrés $f_j - m_j$.

$$c^\alpha = \sum_{j=1}^L \frac{u_\alpha^j}{\sigma_j} (f_j - m_j) \quad (\text{A.3.11})$$

Donc c^α est elle même une variable centrée : $c^\alpha = 0$.

Propriété A.3.2. Les composantes principales sont centrées, on a :

$$\forall \alpha, \beta = 1 \dots L \quad (\alpha \neq \beta) \quad \text{Var}(c^\alpha) = \mu_\alpha \quad , \quad \text{Cov}(c^\alpha, c^\beta) = 0 \quad (\text{A.3.12})$$

A.3.5 Analyse du nuage des variables

De même que l'on fait une analyse du nuage des individus, une analyse du nuage des variables est effectuée. Les démonstrations liées à cette analyse sont faites dans [132]. Nous résumons ici les principaux résultats.

On associe à chaque variable f_j le vecteur z^j de \mathbb{R}^P représentant la variable centrée réduite $(f_j - m_j)/\sigma_j$:

$$z^j = \frac{1}{\sqrt{P}} \sum_{i=1}^P \frac{f_{ij} - m_j}{\sigma_j} = \frac{1}{\sqrt{P}} \sum_{i=1}^P x_i^j e_i' = \frac{1}{\sqrt{P}} x^j \quad (\text{A.3.13})$$

où x^j est le j -ème vecteur-colonne de \mathbf{X} . Le facteur $1/\sqrt{P}$ permet une interprétation statistique du produit scalaire et de la norme :

$$\begin{aligned} \|z^j\|^2 &= \frac{1}{P} \sum_{i=1}^P (x_i^j)^2 = \text{Var}(x^j) \\ (z^j | z^{j'}) &= \frac{1}{P} \sum_{i=1}^P x_i^j x_i^{j'} = \text{Cov}(x^j, x^{j'}) \\ \cos(z^j, z^{j'}) &= \frac{(z^j | z^{j'})}{\|z^j\| \|z^{j'}\|} = \frac{\text{Cov}(x^j, x^{j'})}{\sqrt{\text{Var}(x^j) \text{Var}(x^{j'})}} = \text{Cor}(x^j, x^{j'}) \end{aligned} \quad (\text{A.3.14})$$

Le nuage des points associés aux variables est défini de la manière suivante :

$$N_V = \left\{ \left(z^j, \frac{1}{L} \right) \right\}_{j=1 \dots L} \quad (\text{A.3.15})$$

Remarque A.3.3. Ce nuage n'est pas centré : il n'y a aucune raison pour que l'on ait $\frac{1}{L} \sum_{j=1}^L z^j = 0$. Ce nuage est situé sur la sphère-unité de \mathbb{R}^P puisque $\forall j \|z^j\| = 1$.

Le tableau associé à ce nuage de points est :

$$Z = \begin{bmatrix} z^{1t} \\ \vdots \\ z^{jt} \\ \vdots \\ z^{Lt} \end{bmatrix} = \frac{1}{\sqrt{P}} \begin{bmatrix} x^{1t} \\ \vdots \\ x^{jt} \\ \vdots \\ x^{Lt} \end{bmatrix} = \frac{1}{\sqrt{P}} \mathbf{X}^t \quad (\text{A.3.16})$$

On va effectuer l'analyse du nuage N_V sans le centrer.

La matrice d'inertie \mathbf{W} associée au nuage des variables est égale à :

$$\mathbf{W} = \frac{1}{L} \mathbf{Z}^t \mathbf{Z} = \frac{1}{PL} \mathbf{X} \mathbf{X}^t \quad (\text{A.3.17})$$

On peut montrer [132] que les axes principaux associés au nuage N_V sont égaux à $\frac{c^\alpha}{\sqrt{P\mu_\alpha}}$. Les axes principaux sont ainsi engendrés par les facteurs principaux c^α du nuage N_I . Les facteurs principaux de N_V sont les vecteurs de \mathbb{R}^L :

$$d_\alpha = \mathbf{Z} \frac{c^\alpha}{\sqrt{P\mu_\alpha}} = \frac{1}{P\sqrt{\mu_\alpha}} \mathbf{X}^t c^\alpha = \frac{1}{\sqrt{\mu_\alpha}} \left(\frac{1}{P} \mathbf{X}^t \mathbf{X} \right) u_\alpha = \frac{1}{\sqrt{\mu_\alpha}} \mathbf{V} u_\alpha = \sqrt{\mu_\alpha} u_\alpha \quad (\text{A.3.18})$$

Par conséquent, les facteurs principaux de N_V engendrent les axes principaux du nuage des individus N_I .

A.3.6 Relation de transition entre les facteurs « individus » et les facteurs « variables »

De la relation $c_i^\alpha = (x_i|u_\alpha) = \frac{1}{\sqrt{\mu_\alpha}}(x_i|d_\alpha)$, on en déduit la relation de transition entre les facteurs des individus et les facteurs des variables :

$$c_i^\alpha = \frac{1}{\sqrt{\mu_\alpha}} \sum_{j=1}^L \frac{f_{ij} - m_j}{\sigma_j} d_\alpha^j \quad (\text{A.3.19})$$

La coordonnée c_i^α de l'individu (i.e. pixel vecteur) sur le α -ème axe principal (dans \mathbb{R}^L) est une combinaison linéaire des coordonnées d_α^j des variables (i.e. canaux) f_j , $j = 1 \dots L$, sur le α -ème axe principal (dans \mathbb{R}^P).

A.3.7 Reconstitution des données

Il est possible de reconstituer les données à partir des axes factoriels. La formule de reconstitution des données, s'écrit pour le nuage des individus :

$$x_i^j = \sum_{\alpha=1}^L c_i^\alpha u_\alpha^j \quad (\text{A.3.20})$$

puisque c_i^α est la projection de x_i sur u_α . En introduisant les variables f_j , cette formule devient :

$$\frac{f_{ij} - m_j}{\sigma_j} = \sum_{\alpha=1}^L \frac{c_i^\alpha d_\alpha^j}{\sqrt{\mu_\alpha}} \quad (\text{A.3.21})$$

En se limitant aux K premiers axes factoriels, avec $K < L$, on obtient une approximation \tilde{x}_i^j de $(f_{ij} - m_j)/\sigma_j$.

Propriété A.3.3. *Il est intéressant de noter que l'erreur moyenne de reconstitution est égale à la somme des valeurs propres des axes rejetés [59] :*

$$\overline{err}^2 = \|\mathbf{X} - \widehat{\mathbf{X}}\|^2 = \sum_{j=1}^L \|x^j - \widehat{x}^j\|^2 = \sum_{j=K+1}^L \mu_{\alpha_j} \quad K < L \quad (\text{A.3.22})$$

avec $\widehat{\mathbf{X}}$ la matrice \mathbf{X} reconstituée.

A.3.8 Choix du nombre d'axes

Comme expliqué dans [132, 107], deux critères sont principalement utilisés : le critère de *Kaiser* [87] et le « scree-test » de *Cattell* [88].

La α -ème valeur propre μ_α est l'inertie expliquée par le α -ème axe principal du nuage, et c'est aussi la variance de la composante principal c^α . Plus la valeur propre est grande, plus la composante principale est intéressante, car la forte variabilité fait d'elle une variable (synthétique) pertinente pour étudier, et différencier, les individus.

Les deux valeurs extrêmes que peut prendre la première valeur propre sont les suivantes :

- $\mu_1 = L$ d'où $\mu_2 = \dots = \mu_L = 0$ (car $I_0 = \sum_{\alpha=1}^L \mu_\alpha = L$) Dans ce cas, les matrices \mathbf{V} et \mathbf{X} sont de rang 1, i.e. toutes les variables centrées réduites sont proportionnelles. Donc les variables (i.e. canaux) f_j se déduisent l'une de l'autre par une transformation affine.
- $\mu_1 = 1$ d'où $\mu_2 = \dots = \mu_L = 1$ Dans ce cas là, le nuage des individus N_I , de même que le nuage des variables N_V , est « sphérique » ; i.e. il n'y a pas de direction privilégiée. Ceci se produit lorsque les variables de départ sont toutes non corrélées.

L'inertie expliquée par un axe quelconque de la base canonique vaut 1 (ou, de façon équivalente, les variables de départ $(f_j - m_j)/\sigma_j$ ont pour variance 1). Seuls sont intéressants les axes principaux qui apportent une information supérieure à celle d'un axe de la base canonique, i.e. dont l'inertie expliquée est supérieure à 1 (ou de façon équivalente, seules présentent un intérêt les composantes principales de variance supérieure à 1).

Ceci conduit à un premier critère possible pour choisir le nombre d'axes que l'on va chercher à interpréter, le critère de *Kaiser* : retenir les axes tels que $\mu_\alpha > 1$.

Définition 22 (Critère de *Kaiser* [87]). On retient un axe si sa part d'inertie expliquée τ_α est supérieure à la part d'inertie expliquée « moyenne », égale à $1/L$, où L est le nombre maximum d'axes :

$$\tau_\alpha > \frac{1}{L} \Leftrightarrow \frac{\mu_\alpha}{I_0} > \frac{1}{L} \Leftrightarrow \mu_\alpha > 1 \quad \text{car } I_0 = L \quad (\text{A.3.23})$$

Définition 23. Le « Scree-test » de *Cattell* [88] consiste à repérer s'il existe un point d'inflexion dans la courbe des valeurs propres, et à retenir les axes dont le numéro est inférieur ou égal à celui correspondant au point d'inflexion.

A la droite de ce point, on suppose que se trouvent les facteurs « scree ». « scree » en anglais se traduisant par le terme de géologie : éboulis, c'est à dire les morceaux de rochers qui sont tombés sur la partie inférieures d'une pente rocheuse.

Les deux critères ont été étudiés en détail dans [34, 38, 73, 110, 182]. On peut évaluer ces critères, de manière théorique, en générant des données aléatoires à partir d'un nombre de facteurs donné. On peut alors savoir si le nombre de facteurs est précisément détecté par ces critères. En utilisant cette technique, la première méthode, « critère de *Kaiser* », retient parfois trop de facteurs, tandis que la seconde technique, « scree test », en retient parfois trop peu. Cependant, les deux fonctionnent assez bien dans des conditions normales, c'est à dire lorsqu'il y a relativement peu de facteurs et beaucoup de réalisations.

Une variante graphique du « Scree-test » est la règle du « coude » : chercher le « coude » dans la courbe des valeurs propres, i.e. l'axe à partir duquel la décroissance des valeurs propres devient plus lente, et ne retenir que les axes qui précèdent le coude.

Le simple examen des parts d'inertie expliquée par les axes (τ_α) ne permet pas de juger si un axe est significatif ou non. Il faut en effet tenir compte du nombre de variables : $\tau_\alpha = 10\%$ est une valeur faible si le tableau à 8 variables, forte si le tableau a 100 variables.

La règle qui est peut-être la plus utile dans la pratique est la suivante : ne retenir parmi les axes principaux (ou les composantes principales) que ceux que l'on parvient à interpréter. C'est à dire, dans le traitement de l'image, ceux qui contribuent le plus à l'objet que l'on considère. C'est à dire que l'on observe la forme que peuvent avoir les facteurs principaux représentés sur la grille bitmap de l'image.

Les règles précédentes ont néanmoins leur intérêt, comme indications a priori du nombre d'axes que l'on a de bonnes chances de conserver, ce qui permet d'avoir une estimation du nombre d'axes à conserver. Mais ne pas les appliquer aveuglément permet de ne pas rejeter des axes parfaitement interprétables et qui n'auraient pas été retenus, ou au contraire de ne pas chercher des interprétations par trop hasardeuses pour des axes sélectionnés.

A.3.9 Équivalence entre distance inverse des variances dans l'espace image et distance euclidienne dans l'espace factoriel

Dans le cas de l'ACP, la métrique inverse des variances entre spectres de l'image est égale à la métrique euclidienne entre facteurs pixels dans l'espace factoriel ACP lorsque tous les axes factoriels sont conservés. Si seule une partie des axes est conservée la métrique euclidienne dans l'espace factoriel est une bonne approximation de la métrique inverse des variances dans l'espace image. On se propose de démontrer cette propriété.

Démonstration A.3.2 (Distance inverse des variances et distance euclidienne).

$$\begin{aligned} d_{1/\sigma^2}^2(f_i, f_{i'}) &= \sum_{j=1}^L \left(\frac{f_{ij} - m_j}{\sigma_j} - \frac{f_{i'j} - m_j}{\sigma_j} \right)^2 && \text{(cf. équation A.3.1, p : 197)} \\ &= d_E^2(x_i, x_{i'}) && \text{(cf. équation A.3.6, p : 198)} \end{aligned}$$

D'après l'équation A.3.10 (p : 199), la projection des données s'écrit $c^\alpha = \mathbf{X}u_\alpha$. On en déduit que le vecteur ligne x_i de \mathbf{X} s'écrit :

$$x_i \simeq \sum_{j=1}^L x_i^j e_j \simeq \sum_{j=1}^L \left(\sum_{\alpha=1}^K u_j^\alpha c_i^\alpha \right) e_j$$

avec $\{e_j\}_{j=1\dots L}$ la base canonique de \mathbb{R}^L . Si $K = (L - 1)$ l'équation de reconstitution est alors exacte.

La distance euclidienne entre vecteurs lignes x_i et $x_{i'}$ vaut alors :

$$\begin{aligned}
d_E^2(x_i, x_{i'}) &\simeq \sum_{j=1}^L \left[\left(\sum_{\alpha=1}^K u_j^\alpha c_i^\alpha \right) - \left(\sum_{\alpha=1}^K u_j^\alpha c_{i'}^\alpha \right) \right]^2 \\
&\simeq \sum_{j=1}^L \left[\sum_{\alpha=1}^K u_j^\alpha (c_i^\alpha - c_{i'}^\alpha) \right]^2 \\
&\simeq \sum_{j=1}^L \left[\sum_{\alpha=1}^K (u_j^\alpha (c_i^\alpha - c_{i'}^\alpha))^2 \right] \quad \text{car } \sum_{j=1}^L u_j^{al} u_j^{al'} = 0 \\
&\simeq \sum_{\alpha=1}^K \left[(c_i^\alpha - c_{i'}^\alpha)^2 \sum_{j=1}^L (u_j^\alpha)^2 \right] \\
&\simeq \sum_{\alpha=1}^K [(c_i^\alpha - c_{i'}^\alpha)^2] \\
&\simeq d_E^2(c_i, c_{i'}) \\
&\simeq d_E^2(c_\alpha^f(x_i), c_\alpha^f(x_{i'}))
\end{aligned}$$

D'après les équations A.3.24 et A.3.24, on a l'équivalence entre distance inverse des variances entre spectres dans l'espace image et distance euclidienne entre spectres des facteurs pixels dans l'espace factoriel ACP :

$$d_{1/\sigma^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = d_E(c_\alpha^f(x_i), c_\alpha^f(x_{i'})) . \quad (\text{A.3.24})$$

A.4 L'Analyse Factorielle des Correspondances

L'Analyse Factorielle des Correspondances a été introduite, développée puis publiée par *J.P. Benzécri* en 1973 [21]. Une référence en langue anglaise pour cette analyse est *M.J. Greenacre* [68]. Cette partie s'appuie sur les références [21, 22, 132].

A.4.1 Le tableau de données

A.4.1.1 Introduction des notations de l'AFC

On considère la matrice \mathbf{F} associée à l'image hyperspectrale \mathbf{f}_λ comme un tableau de contingence (ou tableau croisé) ventilant une population d'individus $\Omega = \{1 \dots \omega \dots n\}$, avec $n = P \times L$, selon deux variables \mathcal{X} et \mathcal{Y} . L'ensemble des modalités de \mathcal{X} est noté $I = \{1 \dots i \dots P\}$, l'ensemble des

modalités de \mathcal{Y} est noté $J = \{1 \dots j \dots L\}$.

$$\mathbf{F} = \begin{matrix} \begin{bmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1L} \\ \vdots & & \vdots & & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{iL} \\ \vdots & & \vdots & & \vdots \\ f_{P1} & \cdots & f_{Pj} & \cdots & f_{PL} \end{bmatrix} & \begin{matrix} f_{i.} \leftarrow \text{modalité } i \text{ de } \mathcal{X} \\ \\ \\ \\ f \end{matrix} \\ \begin{matrix} f_{.j} \\ \uparrow \\ \text{modalité } j \text{ de } \mathcal{Y} \end{matrix} \end{matrix} \quad (\text{A.4.1})$$

où f_{ij} est la valeur de l'individu de modalité i pour la variable de modalité j .

Remarque A.4.1. Les valeurs f_{ij} en AFC doivent être positives car l'AFC repose sur le calcul de fréquences et de probabilités.

On note les effectifs marginaux :

$$f_{i.} = \sum_{j=1}^L f_{ij} \quad , \quad f_{.j} = \sum_{i=1}^P f_{ij} \quad (\text{A.4.2})$$

L'effectif total de la population est :

$$f = \sum_{i=1}^P f_{i.} = \sum_{j=1}^L f_{.j} = \sum_{i=1}^P \sum_{j=1}^L f_{ij} \quad (\text{A.4.3})$$

Le tableau de fréquence $\Phi = \nu_{\mathbf{IJ}}$ se déduit du tableau \mathbf{F} en divisant chaque effectif f_{ij} par l'effectif total f .

$$\nu_{\mathbf{IJ}} = \Phi = \begin{matrix} \begin{bmatrix} \vdots \\ \cdots & \nu_{ij} & \cdots \\ \vdots \\ \nu_{.j} \end{bmatrix} & \nu_{i.} \quad , \quad \text{où } \nu_{ij} = \frac{f_{ij}}{f} \end{matrix} \quad (\text{A.4.4})$$

On note les fréquences marginales :

$$\nu_{i.} = \sum_{j=1}^L \nu_{ij} \quad , \quad \nu_{.j} = \sum_{i=1}^P \nu_{ij} \quad (\text{A.4.5})$$

On a :

$$1 = \sum_{i=1}^P \nu_{i.} = \sum_{j=1}^L \nu_{.j} = \sum_{i=1}^P \sum_{j=1}^L \nu_{ij} \quad (\text{A.4.6})$$

On peut considérer que l'on a ainsi défini une loi de probabilité sur l'ensemble produit $I \times J$.

On note :

$$\nu_I = \{\nu_{i.} | i \in I\} \quad , \quad \nu_J = \{\nu_{.j} | j \in J\} \quad (\text{A.4.7})$$

Ces lois sont dites lois marginales du tableau.

$\nu_I = \{\nu_i | i \in I\}$ est la loi de probabilité marginale ou distribution marginale de \mathcal{X} définie sur I .

$\nu_J = \{\nu_{.j} | j \in J\}$ est la loi de probabilité marginale ou distribution marginale de \mathcal{Y} définie sur J .

Par division de chaque ligne $\{\nu_{ij} | j \in J\}$ du tableau de fréquence par sa masse totale ν_i . (total de la ligne) on définit le profil de la ligne i :

$$\nu_J^i = \{\nu_j^i | j \in J\} = \left\{ \frac{\nu_{ij}}{\nu_i} | j \in J \right\} \quad (\text{A.4.8})$$

Ce profil ν_J^i est appelé loi conditionnelle, ou distribution conditionnelle, de j pour i donné ou encore i -ème profil ligne.

De même, par division de chaque colonne $\{\nu_{ij} | i \in I\}$ du tableau de fréquence par sa masse totale $\nu_{.j}$ (total de la colonne) on définit le profil de la colonne j :

$$\nu_I^j = \{\nu_i^j | i \in I\} = \left\{ \frac{\nu_{ij}}{\nu_{.j}} | i \in I \right\} \quad (\text{A.4.9})$$

Ce profil ν_I^j est appelé loi conditionnelle, ou distribution conditionnelle, de i pour j donné ou encore j -ème profil colonne.

On montre [132] que les variables \mathcal{X} et \mathcal{Y} sont indépendantes (dans la population Ω) si et seulement si l'une des 3 conditions (équivalentes entre elles) est vérifiée :

1. tous les profils-lignes sont égaux

$$\nu_J^i = \nu_J \quad \forall i \in I \quad (\text{A.4.10})$$

2. ou tous les profils-colonnes sont égaux

$$\nu_I^j = \nu_I \quad \forall j \in J \quad (\text{A.4.11})$$

A partir des deux premières conditions, on retrouve la propriété connue suivante : la proportionnalité des lignes d'un tableau entraîne celle des colonnes (et réciproquement). Ou de façon plus savante : le rang d'une matrice, qui vaut ici 1, est égal au rang de sa transposée.

3. ou la loi de probabilité sur l'ensemble produit est égale au produit des lois de probabilités marginales :

$$\nu_{ij} = \nu_i \cdot \nu_{.j} \quad \forall i \in I \quad \forall j \in J \quad (\text{A.4.12})$$

On retrouve là la notion classique d'indépendance au sens de la théorie des probabilités.

A.4.1.2 Les objectifs de l'AFC :

L'AFC est une technique de description d'un tableau de contingence, qui va permettre de :

- comparer les profils lignes entre eux, i.e. les distributions de la variable \mathcal{Y} dans les P sous-populations de Ω définies par les modalités I de \mathcal{X} , en faisant apparaître les modalités de \mathcal{X} qui se ressemblent, celles qui s'opposent, celles où la distribution de \mathcal{Y} est très différente de celle observée dans l'ensemble de la population ;

- comparer les profils-colonnes entre eux, i.e. les distributions de la variables \mathcal{X} dans les L sous-populations de Ω définies par les modalités de J de \mathcal{Y} ;
- repérer les cases du tableau où les effectifs observés f_{ij} sont nettement différents des quantités $\frac{f_{i.}f_{.j}}{f}$, appelées effectifs « théoriques », qui seraient ceux que l'on observerait s'il y avait indépendance entre \mathcal{X} et \mathcal{Y} .

D'une façon générale, l'AFC peut être présentée comme une méthode faisant apparaître, par ordre d'importance décroissant, les écarts à la situation d'indépendance, tant au niveau des lignes, que des colonnes, ou des cases du tableau de contingence.

A.4.2 Les nuages de points

A.4.2.1 Le nuage représentant les modalités de \mathcal{X} :

A chaque modalité i de \mathcal{X} on associe le vecteur de \mathbb{R}^L représentant le profil-ligne de j sachant i :

$$\nu_j^i = (\nu_1^i \dots \nu_L^i) \in \mathbb{R}^L$$

Pour mesurer les proximités entre les modalités i , on doit définir une distance dans \mathbb{R}^L . Si l'on choisit la distance euclidienne canonique :

$$d_E^2(\nu_j^i, \nu_j^{i'}) = \sum_{j=1}^L \left(\frac{\nu_{ij}}{\nu_{i.}} - \frac{\nu_{i'j}}{\nu_{i'.}} \right)^2 = \sum_{j=1}^L (\nu_j^i - \nu_j^{i'})^2$$

on risque de donner plus d'importance aux différences $\nu_j^i - \nu_j^{i'}$, concernant les modalités j de \mathcal{Y} les plus fréquentes, pour lesquelles, en général, les ν_j^i vont le plus varier (en valeur absolue) selon i .

Par exemple, pour une modalité j rare, telle que $\nu_{.j} = 5\%$, les ν_j^i varieront, selon les modalités i , de 1% à 5%, alors qu'une modalité j' fréquente ($\nu_{.j'} = 40\%$) variera de 30% à 50% selon les i .

Pour éviter cet inconvénient, on pondère les quantités $(\nu_j^i - \nu_j^{i'})^2$ par les inverses des poids $\nu_{.j}$ des modalités j : on augmente ainsi l'importance des petits écarts dans les colonnes ν_j de poids faible.

La distance ainsi définie s'appelle la distance du chi-deux entre profils-lignes :

$$d_{\chi^2}^2(\nu_j^i, \nu_j^{i'}) = \sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{i.}} - \frac{\nu_{i'j}}{\nu_{i'.}} \right)^2 = \sum_{j=1}^L \frac{1}{\nu_{.j}} (\nu_j^i - \nu_j^{i'})^2 \quad (\text{A.4.13})$$

Remarque A.4.2. Si p, q, r sont trois lois de probabilité définies sur une population $\mathcal{P} = \{1 \dots i \dots P\}$, on appelle « distance du chi-deux entre p et q centrée sur r » la quantité :

$$\sum_{i=1}^P \frac{(p_i - q_i)^2}{r_i}$$

La distance définie auparavant apparaît donc comme la distance du chi-deux entre les distributions conditionnelles de j sachant i et sachant i' , centrée sur la distribution marginale de \mathcal{Y} .

Propriété A.4.1. « *Équivalence distributionnelle* » : si l'on agrège deux colonnes du tableau de contingence \mathbf{F} ayant même profil, i.e. si on remplace deux colonnes proportionnelles par une seule colonne somme des deux précédentes, alors les distances du chi-deux entre profils-lignes restent inchangées, (de même bien sûr que les distances du chi-deux entre les profils-colonnes).

Remarque A.4.3. En statistique pour éprouver l'hypothèse qu'un échantillon de profil ν_j^i et d'effectif f_i est issu d'une loi de probabilité p_J , on doit apprécier si l'écart entre ν_j^i et p_J peut n'être dû qu'aux fluctuations d'échantillonnage; ce qu'on fait depuis *Pearson* en calculant sous le nom de χ^2 la distance $d_{\chi^2}^2(p_J, \nu_j^i)$ (avec pour coefficients les $1/p_j$, inverses des composantes de p_J); et l'on compare $d_{\chi^2}^2(p_J, \nu_j^i) f_i$ à des valeurs numériques lues sur une table.

Pour pouvoir utiliser les résultats de la partie A.2, on se ramène à la distance euclidienne canonique en représentant chaque modalité i de \mathcal{X} par le vecteur :

$$x_i = \begin{bmatrix} x_i^1 \\ \vdots \\ x_i^j \\ \vdots \\ x_i^L \end{bmatrix} \quad \text{où} \quad x_i^j = \frac{\nu_j^i}{\sqrt{\nu_{\cdot j}}} = \frac{\nu_{ij}}{\sqrt{\nu_{\cdot j} \nu_{i \cdot}}} \quad (\text{A.4.14})$$

On a alors l'égalité entre la distance euclidienne entre x_i et $x_{i'}$ et la distance du chi-deux entre les profils-lignes ν_j^i et $\nu_j^{i'}$:

$$d_E^2(x_i, x_{i'}) = d_{\chi^2}^2(\nu_j^i, \nu_j^{i'}) \quad (\text{A.4.15})$$

On pose :

$$\begin{matrix} \mathbf{X} \\ (P, L) \end{matrix} = \begin{bmatrix} \vdots & & \\ \cdots & x_i^j & \cdots \\ \vdots & & \end{bmatrix} \quad \begin{matrix} \mathbf{D}_\mathbf{X} \\ (P, P) \end{matrix} = \begin{bmatrix} \nu_{i \cdot} & & 0 \\ & \ddots & \\ 0 & & \nu_{P \cdot} \end{bmatrix} \quad \begin{matrix} \mathbf{D}_\mathbf{Y} \\ (L, L) \end{matrix} = \begin{bmatrix} \nu_{\cdot 1} & & 0 \\ & \ddots & \\ 0 & & \nu_{\cdot L} \end{bmatrix}$$

Alors :

$$\mathbf{X} = \mathbf{D}_\mathbf{X}^{-1} \Phi \mathbf{D}_\mathbf{Y}^{-1/2} \quad (\text{A.4.16})$$

où $\mathbf{D}_\mathbf{X}^{-1}$ est la matrice diagonale de terme général $1/\nu_{i \cdot}$, $\mathbf{D}_\mathbf{Y}^{-1/2}$ la matrice diagonale de terme général $1/\sqrt{\nu_{\cdot j}}$ et Φ la matrice de fréquence de terme général ν_{ij} .

On munit chaque point x_i d'un poids égal à la fréquence $\nu_{i \cdot}$ de la modalité i dans la population : $\mathbf{D}_\mathbf{X}$ est la matrice des poids des lignes.

On note N_I le nuage représentant les modalités de \mathcal{X} ainsi défini :

$$N_I = \{(x_i, \nu_{i \cdot})\}_{i=1 \dots P} \quad (\text{A.4.17})$$

Le centre de gravité du nuage N_I est :

$$g_X = \sum_{j=1}^L \sqrt{\nu_{\cdot j}} e_j \quad (\text{A.4.18})$$

Remarque A.4.4. N_I est inclus dans l'hyperplan affine S de \mathbb{R}^L défini par :

$$S = \left\{ z = \sum_{j=1}^L z_j e_j \right\} \quad \text{tel que} \quad \sum_{j=1}^L \sqrt{\nu_{\cdot j}} z_j = 1 \quad (\text{A.4.19})$$

En effet : $\forall i = 1 \dots P \quad \sum_{j=1}^L \sqrt{\nu_{.j}} x_i^j = \sum_{j=1}^L \frac{\nu_{ij}}{\nu_{i.}} = \frac{\nu_{i.}}{\nu_{i.}} = 1.$

Propriété A.4.2. On montre que [132] le vecteur g_X est normé et orthogonal à l'hyperplan affine S .

L'inertie totale du nuage N_I par rapport à son centre de gravité vaut :

$$I(N_I) = \sum_{i=1}^P \sum_{j=1}^L \frac{(\nu_{ij} - \nu_{i.}\nu_{.j})^2}{\nu_{i.}\nu_{.j}} = \frac{\chi^2}{f} = \varphi^2 \quad (\text{A.4.20})$$

où χ^2 désigne la statistique du chi-deux associée au tableau de contingence \mathbf{F} , et φ^2 le « phi-deux de contingence ». Ces quantités mesurent l'écart entre le tableau observé \mathbf{F} et le tableau théorique correspondant à la situation d'indépendance entre les deux variables.

Contrairement au cas de l'ACP, l'inertie du nuage a ici une interprétation statistique : elle mesure l'intensité de la liaison entre les variables \mathcal{X} et \mathcal{Y}

Démonstration A.4.1.

$$I(N_I) = \sum_i \nu_{i.} \|x_i - g_X\|^2 = \sum_i \sum_j \nu_{i.} \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{i.}} - \nu_{.j} \right)^2 = \sum_i \sum_j \frac{(\nu_{ij} - \nu_{i.}\nu_{.j})^2}{\nu_{i.}\nu_{.j}} = \varphi^2$$

A.4.2.2 Le nuage représentant les modalités de \mathcal{Y} :

La variable en ligne \mathcal{X} et la variable en colonne \mathcal{Y} jouent en AFC des rôles parfaitement symétriques. Le nuage représentant les modalités de \mathcal{Y} est donc défini de façon analogue au nuage N_I .

On représente chaque modalité j de \mathcal{Y} par le vecteur :

$$y_j = \begin{bmatrix} y_j^1 \\ \vdots \\ y_j^i \\ \vdots \\ y_j^P \end{bmatrix} \quad \text{où} \quad y_j^i = \frac{\nu_{ij}^j}{\sqrt{\nu_{i.}}} = \frac{\nu_{ij}}{\sqrt{\nu_{i.}\nu_{.j}}} \quad (\text{A.4.21})$$

On a alors comme distance du chi-deux entre le profils-colonnes :

$$d^2(y_j, y_{j'}) = d_{\chi^2}^2(\nu_I^j, \nu_I^{j'}) = \sum_{i=1}^P \frac{1}{\nu_{i.}} (\nu_{ij}^j - \nu_{ij}^{j'})^2 \quad (\text{A.4.22})$$

On pose :

$$\mathbf{Y} = \begin{bmatrix} \vdots \\ \dots & y_j^i & \dots \\ \vdots \end{bmatrix} \quad (\text{A.4.23})$$

(L, P)

(les y_j sont les lignes du tableau \mathbf{Y}).

Alors :

$$\mathbf{Y} = \mathbf{D}_Y^{-1} {}^t \Phi \mathbf{D}_X^{-1/2} \quad (\text{A.4.24})$$

On munit chaque point y_j d'un poids égal à la fréquence ν_j de la modalité j de \mathcal{Y} dans la population : \mathbf{D}_Y est la matrice des poids des colonnes.

On note N_J le nuage représentant les modalités de \mathcal{Y} ainsi défini :

$$N_J = \{(y_j, \nu_j)\}_{j=1 \dots L} \quad (\text{A.4.25})$$

N_J est inclus dans un hyperplan affine de \mathbb{R}^P .

Le centre de gravité de N_J est :

$$g_Y = \sum_{i=1}^P \sqrt{\nu_i} \cdot e_i' \quad (\text{A.4.26})$$

$\overrightarrow{OG_Y} = g_Y$ représente la distribution marginale, ou profil moyen de \mathcal{X} . Il est orthogonal en G_Y , défini par $\overrightarrow{OG_Y} = g_Y$, à l'hyperplan affine.

L'inertie de N_J par rapport à son centre de gravité vaut, par raison de symétrie :

$$I(N_J) = \frac{\chi^2}{f} = \varphi^2 = I(N_I) \quad (\text{A.4.27})$$

A.4.3 Analyse du nuage des modalités de \mathcal{X} :

L'analyse du nuage N_I consiste à chercher les axes principaux du nuage préalablement centré, de façon à obtenir des représentations planes de N_I restituant avec le plus de fidélité possible les distances du chi-deux entre les modalités i de \mathcal{X} .

A.4.3.1 Matrices d'inertie

Propriété A.4.3. La matrice d'inertie de N_I par rapport à son centre de gravité vaut :

$$\mathbf{V}^{\mathbf{X}} = \sum_{i=1}^P \nu_i (x_i - g_X)(x_i - g_X)^t \quad (\text{A.4.28})$$

La matrice d'inertie de N_I par rapport à l'origine vaut :

$$\mathbf{V}_0^{\mathbf{X}} = \sum_{i=1}^P \nu_i x_i x_i^t \quad (\text{A.4.29})$$

On montre la relation suivante qui est une généralisation du théorème de Huygens :

$$\mathbf{V}_0^{\mathbf{X}} = \mathbf{V}^{\mathbf{X}} + g_X g_X^t \quad (\text{A.4.30})$$

A.4.3.2 Axes principaux

D'après la propriété A.4.2, l'axe $\Delta(O, g_X)$ est orthogonal en G_X (le centre de gravité de N_I) au support S du nuage : c'est donc un axe principal, associé à la valeur propre 0. On vérifie d'ailleurs facilement que $\mathbf{V}^{\mathbf{X}} g_X = 0$, car g_X est orthogonal à $x_i - g_X \forall i$. On notera cet axe avec l'indice 0 ($\mu_0 = 0$).

Les $L - 1$ autres axes principaux sont engendrés par les autres vecteurs propres normés de $\mathbf{V}^{\mathbf{X}}$:

$$\mathbf{V}^{\mathbf{X}} u_\alpha = \mu_\alpha u_\alpha \quad , \quad \alpha = 1 \dots L - 1 \quad , \quad \|u_\alpha\| = 1 \quad (\text{A.4.31})$$

A.4.3.3 Facteurs principaux

Les facteurs principaux sont les vecteurs de \mathbb{R}^P :

$$c^\alpha = \begin{bmatrix} c_1^\alpha \\ \vdots \\ c_i^\alpha \\ \vdots \\ c_P^\alpha \end{bmatrix} = \mathbf{X}u_\alpha = \sum_{j=1}^L u_\alpha^j x^j \quad , \quad \alpha = 1 \dots L-1 \quad (\text{A.4.32})$$

où $c_i^\alpha = x_i^t u_\alpha$ coordonnée de la modalité i sur l'axe Δu_α , x^j j -ème vecteur-colonne de \mathbf{X} .

Remarque A.4.5. La composante principale c^0 est le vecteur nul.

Le vecteur c^α définit une application de I dans $\mathbb{R} : i \rightarrow c_i^\alpha$. L'application $C^\alpha = c^\alpha \circ \mathcal{X}$ de l'ensemble Ω des individus dans \mathbb{R} est une variable numérique définie sur \mathbb{R} : on dit que l'on a défini un codage de la variable \mathcal{X} .

A.4.3.4 Propriétés des codages

Propriété A.4.4. On considère les codages C^α définis par les composantes principales c^α , $1 \leq \alpha \leq L-1$ (le codage associé à c^0 est constant et égal à 0). On a les propriétés suivantes qui résultent des propriétés des facteurs vues dans la partie A.2 :

1. Les codages sont centrés :

$$\frac{1}{P} \sum_{k=1}^P C^\alpha(k) = \frac{1}{P} \sum_{i=1}^P f_i c_i^\alpha = \sum_{i=1}^P \nu_i c_i^\alpha = 0$$

2. Les variances des codages sont les valeurs propres μ_α

$$\frac{1}{P} \sum_{k=1}^P [C^\alpha(k)]^2 = \frac{1}{P} \sum_{i=1}^P f_i (c_i^\alpha)^2 = \sum_{i=1}^P \nu_i (c_i^\alpha)^2 = \mu_\alpha$$

3. Les covariances entre codages différents sont nulles

$$\frac{1}{P} \sum_{k=1}^P C^\alpha(k) C^\beta(k) = \frac{1}{P} \sum_{i=1}^P f_i c_i^\alpha c_i^\beta = \sum_{i=1}^P \nu_i c_i^\alpha c_i^\beta = 0$$

A.4.3.5 Aspects techniques de l'analyse

Propriété A.4.5. $\mathbf{V}^{\mathbf{X}}$ et $\mathbf{V}_0^{\mathbf{X}}$ ont les mêmes vecteurs propres, relatifs aux mêmes valeurs propres, à l'exception du vecteur g_X : vecteur propre de $\mathbf{V}^{\mathbf{X}}$ pour la valeur propre 0, il est aussi vecteur propre de $\mathbf{V}_0^{\mathbf{X}}$, mais pour la valeur propre 1.

Démonstration A.4.2. Un vecteur propre u_α ($\neq g_X$) de $\mathbf{V}^{\mathbf{X}}$ ($1 \leq \alpha \leq L-1$) relatif à la valeur propre μ_α , est orthogonal à g_X , puisque les vecteurs propres de $\mathbf{V}^{\mathbf{X}}$ sont deux à deux orthogonaux ; donc :

- $\mathbf{V}_0^{\mathbf{X}} u_\alpha = \mathbf{V}^{\mathbf{X}} u_\alpha + g_X^t g_X u_\alpha = \mu_\alpha u_\alpha + 0 = \mu_\alpha u_\alpha$
- $\mathbf{V}_0^{\mathbf{X}} g_X = \mathbf{V}^{\mathbf{X}} g_X + g_X^t g_X g_X = 0 + g_X = g_X$ (car g_X est normé)

Pour déterminer les axes principaux de N_I , on diagonalise la matrice $\mathbf{V}_0^{\mathbf{X}}$ (qui est plus simple que $\mathbf{V}^{\mathbf{X}}$), en ne tenant pas compte du vecteur propre g_X associé à la valeur propre triviale 1 (qui est, comme on le verra plus tard, supérieure aux autres valeurs propres). On notera Δu_α ($1 \leq \alpha \leq L-1$) les axes principaux de N_I .

Ceci revient à faire l'analyse « non centrée » du nuage (i.e. par rapport à l'origine), au lieu de l'analyse « centrée » (i.e. par rapport à G_X). C'est l'orthogonalité de l'axe principal $\Delta(O, g_X)$ avec l'hyperplan affine S support du nuage qui explique, géométriquement, l'équivalence entre les deux analyses : les projections de N_I sur un axe $\Delta(0, u)$ et sur un axes $\Delta(G_X, u)$, où u est un vecteur quelconque orthogonal à g_X , sont les mêmes.

La matrice d'inertie par rapport à l'origine (« non centrée ») $\mathbf{V}_0^{\mathbf{X}}$ s'exprime de manière matricielle avec :

$$\mathbf{V}_0^{\mathbf{X}} = \sum_{i=1}^P \nu_i x_i x_i^t = \mathbf{X}^t \mathbf{D}_X \mathbf{X} = (\mathbf{D}_X^{-1} \Phi \mathbf{D}_Y^{-1/2})^t \mathbf{D}_X (\mathbf{D}_X^{-1} \Phi \mathbf{D}_Y^{-1/2}) = \mathbf{D}_Y^{-1/2} \Phi^t \mathbf{D}_X^{-1} \Phi \mathbf{D}_Y^{-1/2}$$

soit :

$$\mathbf{V}_0^{\mathbf{X}} = \mathbf{H}^t \mathbf{H} \quad \text{avec } \mathbf{H} = \mathbf{D}_X^{-1/2} \Phi \mathbf{D}_Y^{-1/2} \quad (\text{A.4.33})$$

\mathbf{H} est la matrice, de format (P, L) , de terme général :

$$\frac{\nu_{ij}}{\sqrt{\nu_i \nu_j}} \quad (1 \leq i \leq P, 1 \leq j \leq L)$$

A.4.4 Analyse du nuage des modalités de \mathcal{Y} :

L'analyse du nuage N_J est analogue à celle de N_I ; il suffit de permuter les indices i et j , et les notations X et Y .

A.4.4.1 Matrices d'inertie

La matrice d'inertie de N_Y par rapport au centre de gravité G_Y est :

$$\mathbf{V}^{\mathbf{Y}} = \sum_{j=1}^J \nu_j (x y_j - g_Y)(y_j - g_Y)^t = \mathbf{V}_0^{\mathbf{Y}} - g_Y g_Y^t \quad (\text{A.4.34})$$

avec :

$$\mathbf{V}_0^{\mathbf{Y}} = \mathbf{H} \mathbf{H}^t \quad (\text{A.4.35})$$

A.4.4.2 Axes principaux

$\Delta(O, g_Y)$ est un axe principal de N_J , associé à la valeur propre 0 (respectivement 1) dans l'analyse centrée (respectivement non centrée). Les autres axes principaux sont $\Delta(G_Y, v_\alpha)$, notés Δv_α (pour $1 \leq \alpha \leq P-1$), où les v_α sont les vecteurs propres normés de $\mathbf{V}_0^{\mathbf{Y}}$, ou de $\mathbf{V}^{\mathbf{Y}}$, orthogonaux à g_Y , associés aux valeurs propres μ_α .

A.4.4.3 Facteurs principaux

Ce sont les vecteurs de \mathbb{R}^L :

$$d^\alpha = \begin{bmatrix} d_1^\alpha \\ \vdots \\ d_j^\alpha \\ \vdots \\ d_L^\alpha \end{bmatrix} = \mathbf{Y}v_\alpha = \sum_{i=1}^P v_\alpha^i y^i, \quad \alpha = 1 \dots L-1 \quad (\text{A.4.36})$$

où $d_j^\alpha = y_j^t v_\alpha$ coordonnée de la modalité j sur l'axe principal Δv_α , y^i i -ème vecteur-colonne de \mathbf{Y} .

L'application $D^\alpha = d^\alpha \circ \mathcal{Y}$ définit un codage sur J de la variable \mathcal{Y} . Les codages D^α ont des propriétés analogues à celles des codages C^α .

A.4.5 Liens entre les deux analyses :

A.4.5.1 Valeurs propres, axes principaux

Les matrices d'inertie (par rapport à l'origine) des deux nuages ont pour expression :

$$\mathbf{V}_0^{\mathbf{X}} = \mathbf{H}^t \mathbf{H} \quad \mathbf{V}_0^{\mathbf{Y}} = \mathbf{H} \mathbf{H}^t \quad \text{avec } \mathbf{H} = \mathbf{D}_{\mathbf{X}}^{-1/2} \Phi \mathbf{D}_{\mathbf{Y}}^{-1/2} \quad (\text{A.4.37})$$

Propriété A.4.6. On en déduit les propriétés suivantes :

1. Le nombre d'axes principaux associés à des valeurs propres non nulles est le même dans les deux analyses. On le notera q . On a $q \leq \inf(P, L)$, en comptant l'axe principal associé à la valeur propre triviale 1.
2. Les inerties expliquées par les axes principaux sont identiques dans les deux analyses. Les inerties totales des deux nuages étant égales, les parts d'inertie expliquée par les axes principaux sont identiques dans les deux analyses.

Remarque A.4.6. Pour le calcul des parts d'inertie, on ne tient pas compte de la valeur propre triviale égale à 1 :

$$\tau_\alpha = \frac{\mu_\alpha}{\sum_{\alpha=1}^{q-1} \mu_\alpha} \quad (\text{A.4.38})$$

où $\sum_{\alpha=1}^{q-1} \mu_\alpha$ est l'inertie du nuage par rapport à son centre de gravité.

A.4.5.2 Axes principaux et facteurs principaux

Les facteurs principaux d'un nuage se déduisent des axes principaux de l'autre nuage :

$$c^\alpha = \mathbf{X}u_\alpha = \mathbf{D}_{\mathbf{X}}^{-1/2} \mathbf{H}u_\alpha = \frac{1}{\sqrt{\mu_\alpha}} \mathbf{D}_{\mathbf{X}}^{-1/2} \mathbf{H} \mathbf{H}^t v_\alpha = \frac{1}{\sqrt{\mu_\alpha}} \mathbf{D}_{\mathbf{X}}^{-1/2} (\mu_\alpha v_\alpha)$$

soit :

$$c^\alpha = \sqrt{\mu_\alpha} \mathbf{D}_{\mathbf{X}}^{-1/2} v_\alpha \quad (\text{A.4.39})$$

De même :

$$d^\alpha = \sqrt{\mu_\alpha} \mathbf{D}_Y^{-1/2} u_\alpha \quad (\text{A.4.40})$$

On déduit des relations précédentes :

$$\sqrt{\mu_\alpha} d^\alpha = \sqrt{\mu_\alpha} \mathbf{Y} v_\alpha = \mathbf{Y} \mathbf{D}_X^{1/2} c^\alpha$$

soit d'après l'équation A.4.21 :

$$\sqrt{\mu_\alpha} d^\alpha = \mathbf{D}_Y^{-1} \Phi^t c^\alpha \quad (\text{A.4.41})$$

De même :

$$\sqrt{\mu_\alpha} c^\alpha = \mathbf{D}_X^{-1} \Phi^t d^\alpha \quad (\text{A.4.42})$$

Ces relations peuvent s'écrire sous la forme suivante

$$\begin{aligned} \sqrt{\mu_\alpha} c_i^\alpha &= \sum_{j=1}^L \frac{\nu_{ij}}{\nu_i} d_j^\alpha = \sum_{j=1}^L \nu_j^i d_j^\alpha \\ \sqrt{\mu_\alpha} d_j^\alpha &= \sum_{i=1}^P \frac{\nu_{ij}}{\nu_j} c_i^\alpha = \sum_{i=1}^P \nu_i^j c_i^\alpha \end{aligned} \quad (\text{A.4.43})$$

Ces formules permettent donc de « transiter » de \mathbb{R}^P dans \mathbb{R}^L , et réciproquement, pour représenter les facteurs principaux des deux analyses. Elles justifient la représentation graphique simultanée des résultats des deux analyses que l'on utilise habituellement en AFC : on confond formellement l'axe (G_X, u_α) de \mathbb{R}^L et l'axe (G_Y, v_α) de \mathbb{R}^P , ainsi que (G_X, u_β) et (G_Y, v_β) . Sur le graphe déterminé par ces deux axes (fig. A.4), notés Δ_α et Δ_β , on représente la modalité i , notée X_i , de \mathcal{X} par un point de coordonnées (c_i^α, c_i^β) , et la modalité j , notée Y_j , de \mathcal{Y} par un point de coordonnées (d_j^α, d_j^β) . On superpose ainsi les projections de chacun des deux nuages N_I et N_J sur des plans engendrés par des axes de même rang (bien que N_I et N_J soient situés dans des espaces différents).

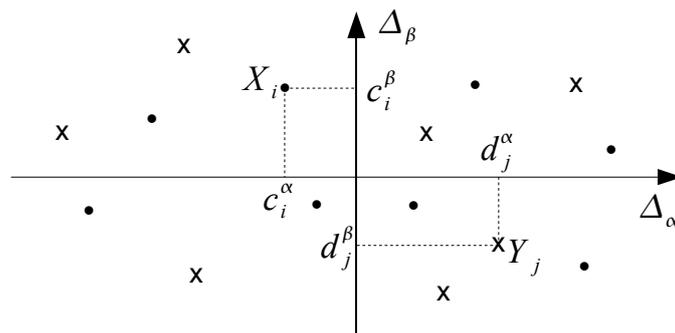


FIG. A.4 – Plan factoriel.

L'origine de ce graphe représente à la fois G_X et G_Y , centres de gravité des nuages N_I et N_J .

Les relations ci-dessus peuvent se traduire de la façon suivante :

- le point de coordonnées $(\sqrt{\mu_\alpha} d_j^\alpha, \sqrt{\mu_\beta} d_j^\beta)$ est le barycentre des points (c_i^α, c_i^β) représentant les modalités i de \mathcal{X} , affectés de poids égaux aux fréquences conditionnelles ν_i^j .

- de même le point $(\sqrt{\mu_\alpha}c_i^\alpha, \sqrt{\mu_\beta}c_i^\beta)$ est le barycentre des points de modalité j affectés des poids f_j^i .

Ces formules de transition sont aussi appelées relations « barycentriques ».

Comme conséquence des relations barycentriques on montre que :

$$\forall \alpha \quad 0 \leq \mu_\alpha \leq 1 \tag{A.4.44}$$

A.4.6 Reconstitution des données :

A.4.6.1 Reconstitution des profils-lignes

On cherche à reconstituer les valeurs $x_i^j = \frac{\nu_{ij}}{\sqrt{\nu_{.j}\nu_{i.}}} = \frac{\nu_j^i}{\sqrt{\nu_{.j}}}$.

On se place dans un plan engendré par deux axes principaux Δu_α et Δu_β ; on note \hat{e}_j la projection orthogonale sur ce plan du vecteur e_j de la base canonique de \mathbb{R}^L .

On obtient une approximation de la valeur $x_i^j = \sqrt{\nu_{.j}}$ (car l'origine est le centre de gravité g_X de N_I) par :

$$(x_i - g_X | \hat{e}_j) = c_i^\alpha w_\alpha^j + c_i^\beta w_\beta^j = \text{proj}(X_i | \hat{e}_j) \| \hat{e}_j \|$$

en notant $\text{proj}(X_i | \hat{e}_j)$ la coordonnée de la projection orthogonale du point x_i représentatif de la modalité i , notée X_i , sur l'axe engendré par \hat{e}_j (fig. A.5).

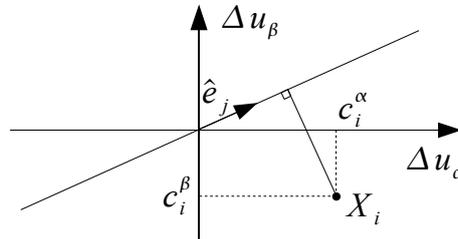


FIG. A.5 – Plan factoriel utilisé pour la reconstitution du profil-ligne.

On obtient donc une reconstitution de la fréquence conditionnelle $\nu_j^i = \nu_{ij}/\nu_{i.}$ du profil-ligne :

$$\begin{aligned} \hat{\nu}_j^i &= \nu_{.j} + \sqrt{\nu_{.j}} \text{proj}(X_i | \hat{e}_j) \| \hat{e}_j \| \\ &= \nu_{.j} + \sqrt{\nu_{.j}} (c_i^\alpha w_\alpha^j + c_i^\beta w_\beta^j) \end{aligned} \tag{A.4.45}$$

A.4.6.2 Reconstitution des profils-colonnes

La méthode est analogue à celle du paragraphe précédent. On a cette fois :

$$\begin{aligned} \hat{\nu}_i^j &= \nu_{i.} + \sqrt{\nu_{i.}} \text{proj}(Y_j | \hat{e}_i) \| \hat{e}_i \| \\ &= \nu_{i.} + \sqrt{\nu_{i.}} (d_j^\alpha v_\alpha^i + d_j^\beta v_\beta^i) \end{aligned} \tag{A.4.46}$$

A.4.6.3 Reconstitution du tableau de fréquences

La reconstitution de $\nu_j^i = \nu_{ij}/\nu_i$, présentée ci-dessus (eq. A.4.45) s'écrit en utilisant les équations A.4.39 et A.4.40 :

$$\widehat{\nu}_j^i = \frac{\widehat{\nu}_{ij}}{\nu_i} = \nu_{.j} + \sqrt{\nu_{.j}} \left(c_i^\alpha \sqrt{\nu_{.j}} \frac{1}{\sqrt{\mu_\alpha}} d_j^\alpha + c_i^\beta \sqrt{\nu_{.j}} \frac{1}{\sqrt{\mu_\alpha}} d_j^\beta \right) = \nu_{.j} \left(1 + \frac{1}{\sqrt{\mu_\alpha}} c_i^\alpha d_j^\alpha + \frac{1}{\mu_\beta} c_i^\beta d_j^\beta \right)$$

Soit :

$$\widehat{\nu}_j^i = \nu_{.j} \left(1 + \frac{1}{\sqrt{\mu_\alpha}} c_i^\alpha d_j^\alpha + \frac{1}{\sqrt{\mu_\beta}} c_i^\beta d_j^\beta \right)$$

La reconstitution de ν_i^j conduit à la même expression pour $\widehat{\nu}_{ij}$, qui est symétrique en i et j .

Plus généralement, la formule de reconstitution à partir des k premiers axes $\Delta_1, \dots, \Delta_\alpha, \dots, \Delta_k$ s'écrit :

$$\widehat{\nu}_{ij}^{(k)} = \nu_i \nu_{.j} \left(1 + \sum_{\alpha=1}^k \mu_\alpha^{-1/2} c_i^\alpha d_j^\alpha \right) \quad (\text{A.4.47})$$

L'AFC apparaît ainsi comme une méthode de reconstitution d'un tableau de fréquences à partir de ses fréquences marginales :

- à l'étape 0, l'approximation du tableau est le tableau « produit des marges » :

$$\widehat{\nu}_{ij}^{(k)} = \nu_i \nu_{.j} \quad (\text{A.4.48})$$

(correspondant à la situation d'indépendance entre \mathcal{X} et \mathcal{Y})

- les approximations successives sont de plus en plus fidèles au tableau de départ, au sens d'une distance entre tableaux
- on obtient une reconstitution exacte du tableau de départ en considérant les q axes factoriels associés aux q valeurs propres non nulles :

$$\nu_{ij} = \nu_i \nu_{.j} \left(1 + \sum_{\alpha=1}^{q-1} \mu_\alpha^{-1/2} c_i^\alpha d_j^\alpha \right) \quad (\text{A.4.49})$$

On peut remarquer que les marges du tableau reconstitué, quel que soit le nombre d'axes retenus, sont toujours égales aux marges du tableau de départ. Par exemple :

$$\widehat{\nu}_i. = \sum_j \widehat{\nu}_{ij} = \sum_j \nu_i \nu_{.j} \left(1 + \sum_{\alpha=1}^k \frac{1}{\sqrt{\mu_\alpha}} c_i^\alpha d_j^\alpha \right) = \nu_i \left[1 + \sum_{\alpha=1}^k \frac{1}{\sqrt{\mu_\alpha}} c_i^\alpha \left(\sum_j \nu_{.j} d_j^\alpha \right) \right] = \nu_i \quad (\text{A.4.50})$$

car $\sum_j \nu_{.j} d_j^\alpha = 0 \forall \alpha$ (l'origine du graphe représente le centre de gravité de N_I).

Remarque A.4.7. D'après [22], du point de vue de l'interprétation, l'approximation d'ordre k du tableau, que donne la formule :

$$\widehat{f}_{ij}^{(k)} = f \nu_i \nu_{.j} \left(1 + \sum_{\alpha=1}^k \mu_\alpha^{-1/2} c_i^\alpha d_j^\alpha \right) \quad (\text{A.4.51})$$

apparaît comme une schématisation ne tenant compte que des qualités associées aux k premiers facteurs. D'une part, cette schématisation est économique en ce que $\widehat{f}_{ij}^{(k)}$ est complètement engendré par

les facteurs et les marges ; soit au total, pour un grand tableau, beaucoup moins de nombres que n'en contient le tableau initial (e.g. si Card $I = 152 \times 91 = 13\,832$ pixels vecteurs, Card $J = 61$ canaux, $k = 3$, les 3 premiers facteurs sur I représentent $3 \times 13\,832 = 41\,496$ nombres ; ceux de $J : 3 \times 61 = 183$ nombres ; les marges $\nu_{i.}$ et $\nu_{.j} : 13\,832 + 61 = 13\,893$ nombres ; soit, au total, 55 572 nombres au lieu de $13\,832 \times 61 = 843\,752$ nombres qui constituent le tableau \mathbf{F}). D'autre part, si les facteurs de rang supérieur à k n'ont pas reçu d'interprétation, cette schématisation est l'expression numérique de ce que l'on est parvenu à comprendre des données. D'un point de vue statistique, on peut ajouter que ces facteurs de rang élevé peuvent être dépourvus de toute signification, s'ils sont seulement l'expression de fluctuations statistiques.

Ainsi, le tableau approché à l'ordre k ne sera pas seulement plus simple que le tableau initial. Il sera véritablement plus fidèle au réel, plus juste : en éliminant le résidu d'inertie associé aux derniers axes, i.e le bruit, la formule de reconstitution où n'entrent que des nombres directement porteurs de signification aura filtré le tableau brut de ce qu'il contenait d'aléatoire.

A.4.7 Quelques démonstrations concernant les distances

Dans cette partie nous démontrons quelques résultats concernant la distance du chi-deux et la distance euclidienne.

A.4.7.1 Écriture de la distance du chi-deux selon les profils lignes selon les pixels vecteurs

La distance du chi-deux est une distance entre profils lignes ν_j^i et $\nu_j^{i'}$ (eq. A.4.13, p : 207) :

$$d_{\chi^2}(\nu_j^i, \nu_j^{i'}) = \sqrt{\sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{i.}} - \frac{\nu_{i'j}}{\nu_{i'.}} \right)^2} = \sqrt{\sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\nu_j^i - \nu_j^{i'} \right)^2}$$

Par extension elle correspond à une distance entre pixels vecteurs $\mathbf{f}_\lambda(x_i)$ et $\mathbf{f}_\lambda(x_{i'})$:

$$d_{\chi^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = \sqrt{\sum_{j=1}^L \frac{N}{f_{.j}} \left(\frac{f_{\lambda_j}(x_i)}{f_{x_i.}} - \frac{f_{\lambda_j}(x_{i'})}{f_{x_{i'}.}} \right)^2}$$

avec $f_{.j} = \sum_{i=1}^P f_{\lambda_j}(x_i)$, $f_{x_i.} = \sum_{j=1}^J f_{\lambda_j}(x_i)$ et $N = \sum_{j=1}^L \sum_{i=1}^P f_{\lambda_j}(x_i)$.

Montrons l'égalité entre ces deux écritures.

Comme $\nu_{ij} = f_{\lambda_j}(x_i)/N$, $\nu_{.j} = f_{.j}/N$ et $\nu_{i.} = f_{x_i.}/N$ on a l'égalité :

$$d_{\chi^2}(\nu_j^i, \nu_j^{i'}) = d_{\chi^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) , \quad (\text{A.4.52})$$

A.4.7.2 Équivalence entre distance du chi-deux dans l'espace image et distance euclidienne dans l'espace factoriel

La métrique du chi-deux, entre spectres, dans l'espace image est équivalente à la métrique euclidienne dans l'espace factoriel si tous les axes factoriels sont conservés. Si seule une partie des axes factoriels est conservée, alors la métrique euclidienne dans l'espace factoriel est une bonne approximation de la métrique du chi-deux entre spectres dans l'espace image [21]. On se propose de démontrer cette propriété.

Démonstration A.4.3 (Distance du chi-deux et distance euclidienne).

$$\begin{aligned} d_{\chi^2}^2(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) &= d_{\chi^2}(\nu_J^i, \nu_J^{i'}) \quad (\text{cf. équation A.4.52}) \\ &= d_E^2(x_i, x_{i'}) \quad (\text{cf. équation A.4.15, p : 208}) \end{aligned} \quad (\text{A.4.53})$$

Or d'après l'équation 2.2.30 (p : 15) la projection des données s'écrit $\mathbf{C} = \mathbf{X}\mathbf{U}$. La reconstitution s'écrit donc :

$$\begin{array}{ccc} \mathbf{X} & \simeq & \mathbf{C} \times \mathbf{U}^t \\ (P, L) & & (P, K) \quad (K, L) \end{array} \cdot$$

Si $K = (L - 1)$ l'équation de reconstitution est alors exacte $\mathbf{X} = \mathbf{C}\mathbf{U}^t$. On en déduit que le vecteur ligne x_i de \mathbf{X} s'écrit :

$$x_i = \sum_{j=1}^L x_i^j e_j \simeq \sum_{j=1}^L \left(\sum_{\alpha=1}^K u_j^\alpha c_i^\alpha \right) e_j$$

avec $\{e_j\}_{j=1\dots L}$ la base canonique de \mathbb{R}^L .

La distance euclidienne entre vecteurs lignes x_i et $x_{i'}$ vaut alors :

$$\begin{aligned} d_E^2(x_i, x_{i'}) &\simeq \sum_{j=1}^L \left[\left(\sum_{\alpha=1}^K u_j^\alpha c_i^\alpha \right) - \left(\sum_{\alpha=1}^K u_j^\alpha c_{i'}^\alpha \right) \right]^2 \quad (\text{A.4.54}) \\ &\simeq \sum_{j=1}^L \left[\sum_{\alpha=1}^K u_j^\alpha (c_i^\alpha - c_{i'}^\alpha) \right]^2 \\ &\simeq \sum_{j=1}^L \left[\sum_{\alpha=1}^K (u_j^\alpha (c_i^\alpha - c_{i'}^\alpha))^2 \right] \quad \text{car } \sum_{j=1}^L u_j^\alpha u_j^{\alpha'} = 0 \\ &\simeq \sum_{\alpha=1}^K \left[(c_i^\alpha - c_{i'}^\alpha)^2 \sum_{j=1}^L (u_j^\alpha)^2 \right] \\ &\simeq \sum_{\alpha=1}^K [(c_i^\alpha - c_{i'}^\alpha)^2] \\ &\simeq d_E^2(c_i, c_{i'}) \\ &\simeq d_E^2(c_\alpha^{\mathbf{f}}(x_i), c_\alpha^{\mathbf{f}}(x_{i'})) \end{aligned}$$

D'après les équations A.4.53 et A.4.54, on a l'équivalence entre distance du chi-deux entre spectres dans l'espace image et distance euclidienne entre spectres des facteurs pixels dans l'espace factoriel AFC :

$$d_{\chi^2}(\mathbf{f}_\lambda(x_i), \mathbf{f}_\lambda(x_{i'})) = d_E(c_\alpha^{\mathbf{f}}(x_i), c_\alpha^{\mathbf{f}}(x_{i'})) \quad (\text{A.4.55})$$

A.5 L'AFC et le bruit

Cette partie vient compléter la partie 2.4.2.

On observe l'influence du bruit sur l'inertie des axes factoriels, c'est à dire les valeurs propres de la matrice d'inertie. Le phénomène va être étudié sur l'image « woman face » que l'on a bruitée avec un bruit d'écart type σ constant $n_{\lambda_j} = \mathcal{N}(0, \sigma)$ (eq. 2.4.10, p : 32). Les canaux n'ont donc par un RSB constant.

On trace l'inertie des axes factoriels pour différents écart types σ du bruit (fig. A.6). On constate que l'inertie du premier axe diminue avec l'augmentation du bruit alors que l'inertie des autres axes augmente avec le bruit. Or lorsque le bruit augmente, la part d'inertie créée par l'information est plus faible alors que la part d'inertie correspondant au bruit augmente. Donc, le premier axe dont l'inertie diminue avec le bruit, est porteur de l'information ; alors que les autres axes contiennent principalement du bruit, i.e. leur inertie augmente.

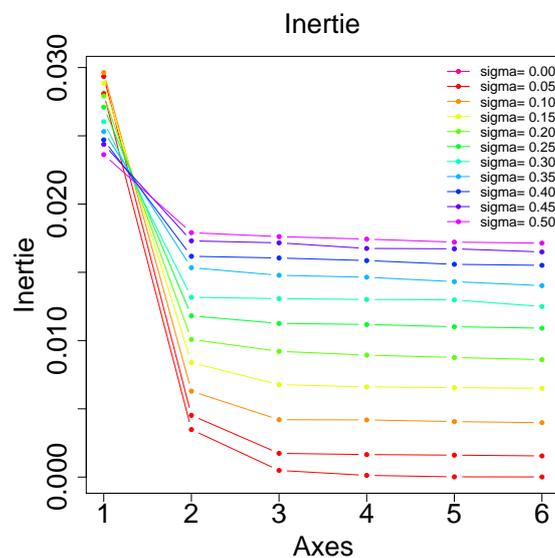


FIG. A.6 – Comparaison des inertie des axes factoriels pour différents écart types σ du bruit.

A.6 Composition d'AFC et de reconstitutions pour le filtrage spectral

Dans cette partie, plusieurs études sur le filtrage spectral des données bruitées par AFC ont été ajoutés afin de compléter l'étude sur la bruit (annexe partie A.6). Ces résultats présentent un intérêt pour la compréhension de la réduction de bruit. Ils sont placés en annexe, et non dans le corps du manuscrit, afin de ne pas entraver la compréhension du lecteur.

A.6.1 Rapport [S]/[B] pour un nombre d'axes variable

(annexes)

Une AFC récursive est effectuée sur l'image bruitée par un bruit blanc gaussien d'écart type $\sigma = 0,25$. On utilise l'image « woman face » que l'on a bruitée avec un bruit d'écart type σ constant $n_{\lambda_j} = \mathcal{N}(0, \sigma)$ (eq. 2.4.10, p : 32). Les canaux n'ont donc par un RSB constant. Le nombre d'axes décroît selon les itérations de la manière suivante :

- Première itération : 6 axes factoriels ;
- Deuxième itération : 3 axes factoriels ;
- Troisième itération : 1 axe factoriel.

Remarque A.6.1. Le nombre d'axes ne croît pas. En effet, lorsque l'on reconstitue, après avoir choisi un nombre donné d'axes, l'espace image est dégénéré et de dimension égale au nombre d'axes. Ainsi lors de l'AFC suivante le nombre d'axes doit être inférieur ou égal au nombre d'axes de l'AFC précédente.

On observe le canal $f_{\lambda_{39}}$ non bruité, après bruitage $f_{\lambda_{39}}^{br}$ et après plusieurs itérations d'AFC $\widehat{f_{\lambda_{39}}^{br}}^{(n)}$ (fig. A.7). Les images des facteurs pixels pour les différentes itérations d'AFC-reconstitutions sont visibles sur la figure A.9.

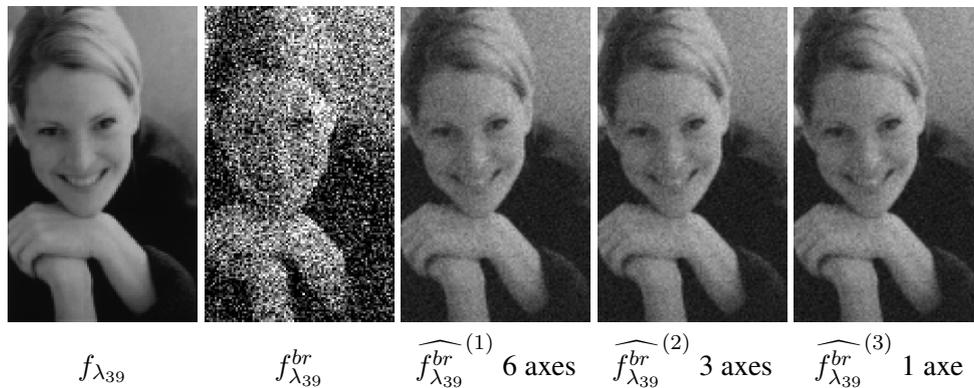


FIG. A.7 – Canal 39 de l'image sans bruit $f_{\lambda_{39}}$, bruitée $f_{\lambda_{39}}^{br}$ et reconstituée après différentes itérations $\widehat{f_{\lambda_{39}}^{br}}^{(n)}$ (bruit blanc gaussien d'écart type $\sigma = 0, 25$).

On mesure le rapport signal sur bruit (RSB) hyperspectral selon les itérations comme dans l'équation 2.4.13 (p : 32).

Sur la figure A.8, on constate que le RSB est bien meilleur à la troisième itération lorsque le nombre d'axes décroît que lorsqu'il reste constant. En effet les facteurs pixels sur les axes 2 et 3 contiennent beaucoup de bruit, à toutes les itérations (fig. A.9). Ces axes diminuent les performances de la composition d'AFC-reconstitution à nombre d'axes constant. En outre, dans cet exemple, seuls les facteurs pixels sur le premier axe contiennent réellement de l'information, d'où l'importance du choix d'un nombre d'axes pertinent dès la première itération.

Une AFC récursive avec nombre d'axes décroissant (bien choisis !) est donc meilleure.

A.6.2 Comparaison entre une composition d'AFC sur quelques canaux et sur tous les canaux

On veut observer l'influence que peut avoir le nombre de canaux sur le rapport signal sur bruit pour la composition d'AFC-reconstitutions sur une image bruitée f_{λ}^{br} . Pour cela on effectue :

- une AFC-reconstitution sur tous les canaux (61 canaux) ;
- une AFC-reconstitution sur une image contenant les canaux $f_{\lambda_1}^{br}, f_{\lambda_{10}}^{br}, f_{\lambda_{20}}^{br}, f_{\lambda_{30}}^{br}, f_{\lambda_{40}}^{br}, f_{\lambda_{50}}^{br}$ et $f_{\lambda_{60}}^{br}$ (7 canaux).

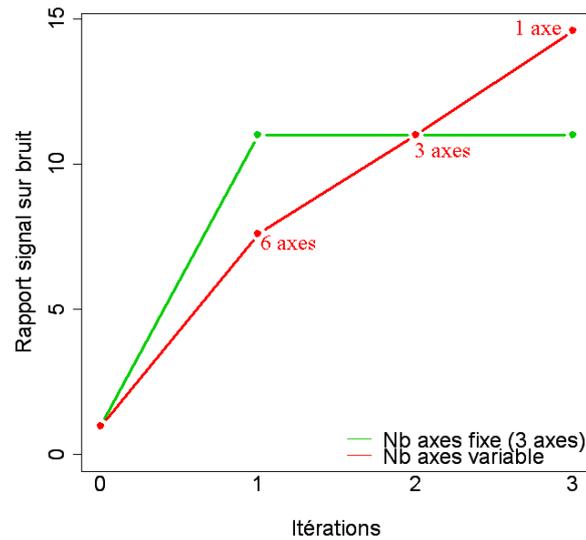


FIG. A.8 – Comparaison des RSB hyperspectraux de l'image bruitée f_{λ}^{br} pour un nombre d'axes fixe ou variable pour différentes itérations d'AFC-reconstitutions (bruit blanc gaussien d'écart type $\sigma = 0,25$).

On considère un image bruitée par un bruit blanc gaussien d'écart type $\sigma = 0,25$. On mesure les RSB hyperspectraux entre les images reconstituées $\widehat{f_{\lambda}^{br}}^{(n)}$ et l'image de départ f_{λ} pour différentes itérations d'AFC-reconstitutions (fig. A.10). Les RSB sont plus élevés pour l'image à 61 canaux que pour l'image à 7 canaux.

C'est donc bien le nombre de canaux, c'est à dire la redondance d'information entre les canaux, qui permet de réduire le bruit.

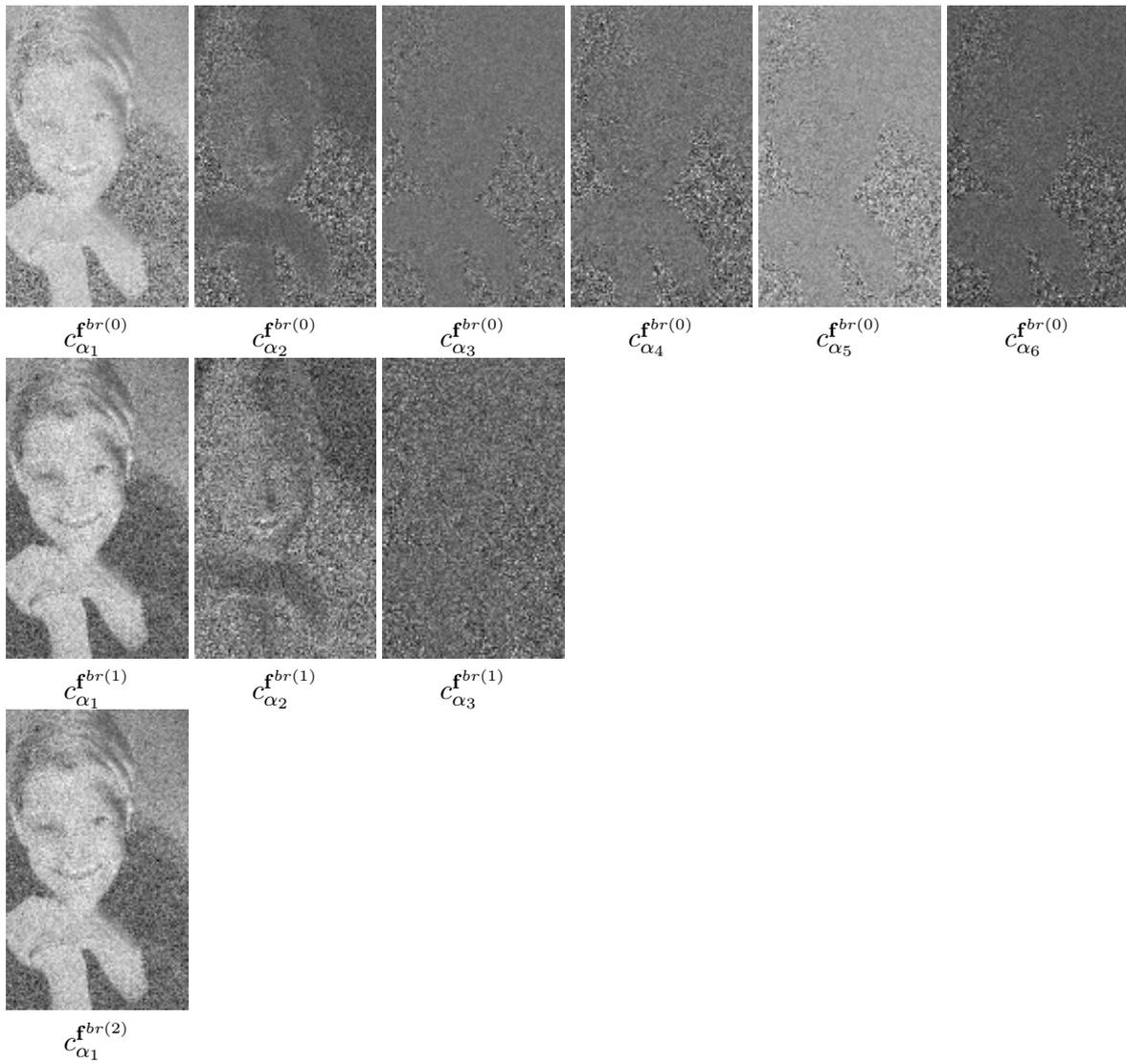


FIG. A.9 – Images des facteurs pixels de l'image bruitée pour différentes itérations d'AFC-reconstitution (bruit blanc gaussien d'écart type $\sigma = 0,25$).

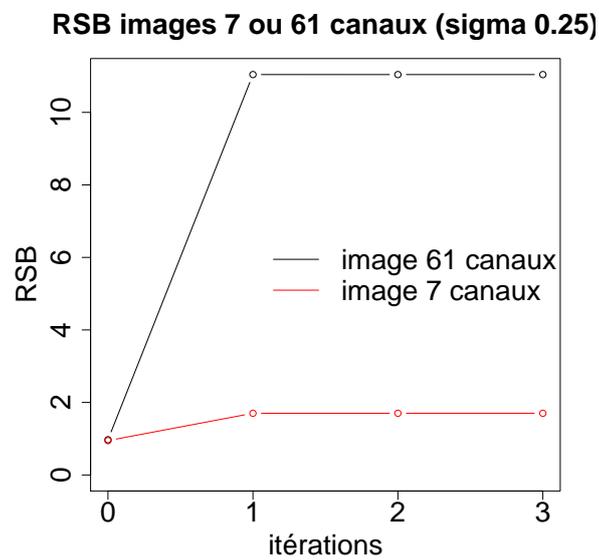


FIG. A.10 – Comparaison des RSB hyperspectraux de l'image bruitée à 7 canaux et de l'image bruitée à 61 canaux f_{λ}^{br} pour différentes itérations (bruit blanc gaussien d'écart type $\sigma = 0,25$).

A.7 Influence de la translation des données sur l'AFC

Dans cette partie, on étudie l'influence de la translation, également appelée décalage, des données sur l'AFC. Pour cette étude, une seule AFC est effectuée, c'est à dire que la composition d'AFC-reconstitutions n'est pas utilisée.

A.7.1 Image sans bruit

Afin d'observer l'influence de la translation sur l'AFC on ajoute à l'image hyperspectrale \mathbf{f}_λ , sans bruit, une constante positive ε en chaque point (x_i, λ_j) . L'image obtenue est appelée image décalée et est notée \mathbf{f}_λ^* :

$$f_{\lambda_j}^*(x_i) = f_{\lambda_j}(x_i) + \varepsilon = t_\varepsilon(f_{\lambda_j})(x_i) \quad (\text{A.7.1})$$

avec la translation t_ε qui s'exprime sous la forme :

$$t_\varepsilon : \begin{cases} \mathcal{T}^L & \rightarrow \mathcal{T}^L \\ \mathbf{f}_\lambda & \rightarrow t_\varepsilon(\mathbf{f}_\lambda) \text{ tel que } \forall i = 1 \dots P, \forall j = 1 \dots L \quad t_\varepsilon(\mathbf{f}_{\lambda_j})(x_i) = \mathbf{f}_{\lambda_j}(x_i) + \varepsilon \quad \varepsilon \in \mathbb{R} \end{cases} \quad (\text{A.7.2})$$

La valeur de ε est faible par rapport à la dynamique des niveaux de gris des canaux. En général, elle est de l'ordre de grandeur du minimum, en valeur absolue, de l'image reconstituée après une AFC sans translation : $\widehat{\mathbf{f}}_\lambda$.

Sur l'image dont les valeurs ont été traduites \mathbf{f}_λ^* , on applique une AFC. On ne conserve que les facteurs pixels $\mathbf{c}_{\alpha}^{\mathbf{f}^*} = (c_{\alpha_1}^{\mathbf{f}^*}, c_{\alpha_2}^{\mathbf{f}^*}, c_{\alpha_3}^{\mathbf{f}^*})$ associés aux 3 premiers axes factoriels. A partir de ces axes factoriels on reconstitue l'image $\widehat{\mathbf{f}}_\lambda^*$. On retranche ensuite la constante ε à cette image reconstituée :

$$\widehat{\mathbf{f}}_\lambda^{*\varepsilon} = \widehat{\mathbf{f}}_\lambda^* - \varepsilon = t_{-\varepsilon}(\widehat{\mathbf{f}}_\lambda^*) = t_{-\varepsilon}(\widehat{t_\varepsilon(\mathbf{f}_\lambda)}) \quad (\text{A.7.3})$$

On mesure le rapport signal sur bruit entre l'image \mathbf{f}_λ et l'image reconstituée $\widehat{\mathbf{f}}_\lambda^{*\varepsilon}$, eq. 2.4.13, ainsi qu'entre les axes factoriels de l'image de départ $\mathbf{c}_{\alpha}^{\mathbf{f}}$ et les axes de l'image traduite $\mathbf{c}_{\alpha}^{\mathbf{f}^*}$:

$$RSB_{\alpha_k} = \frac{\text{var}(\text{signal})}{\text{var}(\text{bruit})} = \frac{\text{var}(c_{\alpha_k}^{\mathbf{f}})}{\text{var}(c_{\alpha_k}^{\mathbf{f}^*} - c_{\alpha_k}^{\mathbf{f}})} \quad (\text{A.7.4})$$

La somme S_{χ^2} des carrés des distances du chi deux entre l'image \mathbf{f}_λ et l'image décalée et reconstituée $\widehat{\mathbf{f}}_\lambda^{*\varepsilon} = \widehat{\mathbf{f}}_\lambda^* - \varepsilon$ centrée sur la distribution marginale ν_{λ_j} , i.e. la référence, s'écrit :

$$S_{\chi^2} = \sum_{i=1}^P d_{\chi^2}^2(\mathbf{f}_\lambda(x_i), \widehat{\mathbf{f}}_\lambda^{*\varepsilon}(x_i)) = \sum_{i=1}^P \sum_{j=1}^L \frac{1}{\nu_{.j}} \left(\frac{\nu_{ij}}{\nu_{i.}} - \frac{\widehat{\nu}_{ij}^{*\varepsilon}}{\widehat{\nu}_{i.}^{*\varepsilon}} \right) = \sum_{i=1}^P \sum_{j=1}^L \frac{f}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{\widehat{f}_{ij}^{*\varepsilon}}{\widehat{f}_{i.}^{*\varepsilon}} \right) \quad (\text{A.7.5})$$

avec $f = \sum_{i=1}^P \sum_{j=1}^L f_{ij}$. Cette somme permet de mesurer les écarts, dans l'espace image, entre l'image \mathbf{f}_λ et l'image décalée et reconstituée $\widehat{\mathbf{f}}_\lambda^{*\varepsilon}$.

La somme des carrés des écarts entre les facteurs pixels de l'image de départ $\mathbf{c}_{\alpha}^{\mathbf{f}}$ et ceux de l'image décalée $\mathbf{c}_{\alpha}^{\mathbf{f}^*}$ vaut :

$$SCE = \sum_{i=1}^P \sum_{k=1}^K \left(c_{\alpha_k}^{\mathbf{f}}(x_i) - c_{\alpha_k}^{\mathbf{f}^*}(x_i) \right)^2 \quad (\text{A.7.6})$$

Cette somme permet de mesurer les écarts, dans l'espace factoriel, entre les facteurs pixels de l'image de départ \mathbf{c}_α^f et ceux de l'image décalée \mathbf{c}_α^{f*} .

Les courbes correspondantes sont tracées sur la figure A.11 ainsi que les facteurs canaux dans le plan factoriel (fig. A.12).

Remarque A.7.1. Comme les axes factoriels peuvent changer de sens selon les analyses [21], il est important de les mettre tous dans le même sens. Ainsi pour les facteurs pixels, comme pour les facteurs canaux, on calcule le coefficient de corrélation entre les facteurs de l'image de départ et les facteurs de l'image décalée. Si ce coefficient est négatif, on change le signe des facteurs de l'image décalée pour l'affichage dans le plan factoriel.

Observations :

- La somme des distances du chi-deux S_{χ^2} entre l'image de départ \mathbf{f}_λ et l'image décalée et reconstituée $\widehat{\mathbf{f}}_\lambda^{*- \varepsilon}$ augmente avec la valeur de la translation ε .
- Le RSB de l'image décalée et reconstituée $\widehat{\mathbf{f}}_\lambda^{*- \varepsilon}$ diminue avec l'augmentation de la translation.
- La somme des carrés des écarts SCE entre les facteurs pixels de l'image de départ \mathbf{c}_α^f et les facteurs pixels de l'image décalée \mathbf{c}_α^{f*} augmente avec la valeur de la translation ε .
- Le RSB des facteurs pixels de l'image décalée $\mathbf{c}_{\alpha_k}^{f*}$ diminue avec la valeur de la translation ε .
- L'inertie de chacun des axes de l'image décalée diminue avec la valeur de la translation ε .
- Les facteurs canaux sont « plus resserrés » dans le plan factoriel engendré par les deux premiers axes lorsque la translation augmente.

Interprétations : La translation ε a pour effet d'éloigner de l'image de départ \mathbf{f}_λ sa reconstitution $\widehat{\mathbf{f}}_\lambda^{*- \varepsilon}$. Il en va de même pour les facteurs pixels sur les axes factoriels de l'image reconstituée \mathbf{c}_α^{f*} qui s'éloignent de ceux de l'image de départ \mathbf{c}_α^f . Ceci se traduit par une diminution du RSB en fonction de la valeur de la translation et une augmentation de la somme des carrés des distances du chi-deux S_{χ^2} dans l'espace image. De même, il se produit une augmentation de la somme des carrés des écarts SCE dans l'espace factoriel.

Remarque A.7.2. Il est à noter, qu'à cause de la translation, le centre de gravité du nuage des pixels et du nuage des canaux est changé.

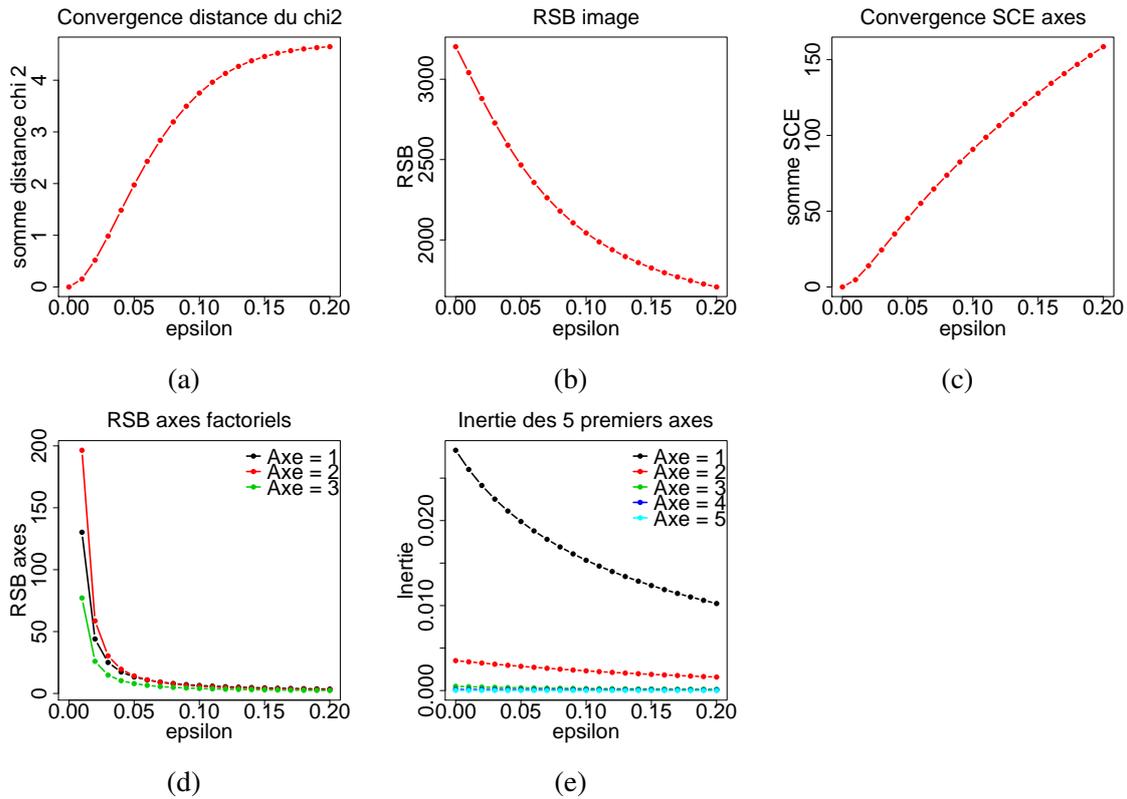


FIG. A.11 – Influence de la translation ϵ sur l'AFC entre l'image de départ f_λ et l'image décalée et reconstituée $\widehat{f_\lambda^{*-\epsilon}}$: (a) Somme des distances du chi deux (référence image f_λ), (b) RSB de l'image décalée et reconstituée, (c) Somme des carrés des écarts entre les facteurs pixels c_α^f et $c_\alpha^{f^*}$, (d) RSB pour les facteurs pixels $c_{\alpha_k}^{f^*}$ par rapport à $c_{\alpha_k}^f$ (e) Inertie des 5 premiers axes factoriels de l'image décalée et reconstituée.

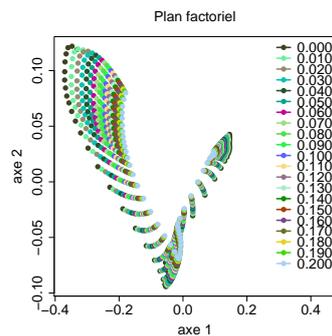


FIG. A.12 – Influence du décalage $\epsilon \in [0; 0, 2]$ sur les facteurs canaux dans le plan factoriel formé par les 2 premiers axes.

A.7.2 Image bruitée

A.7.2.1 Image bruitée avec un bruit blanc gaussien d'écart type constant

On considère une image de référence sans bruit \mathbf{f}_λ . A cette image on ajoute un bruit blanc gaussien centré d'écart type $\sigma = 0,05$, $n = \mathcal{N}(0, \sigma = 0,05)$, et une constante positive ε en chaque point (x_i, λ_j) . L'image obtenue est notée \mathbf{f}_λ^{br*} :

$$f_{\lambda_j}^{br*}(x_i) = f_{\lambda_j}(x_i) + n + \varepsilon = f_{\lambda_j}^{br}(x_i) + \varepsilon = t_\varepsilon(f_{\lambda_j}^{br})(x_i) \quad (\text{A.7.7})$$

Sur l'image bruitée dont les valeurs ont été translatées \mathbf{f}_λ^{br*} , on applique le même procédé que dans la partie A.7.1. On ne conserve que les facteurs pixels $\mathbf{c}_\alpha^{\mathbf{f}^{br*}} = (c_{\alpha_1}^{\mathbf{f}^{br*}}, c_{\alpha_2}^{\mathbf{f}^{br*}}, c_{\alpha_3}^{\mathbf{f}^{br*}})$ associés aux 3 premiers axes factoriels de l'AFC et on reconstitue l'image $\widehat{\mathbf{f}_\lambda^{br*}}$. On retranche ensuite la constante ε à cette image reconstituée :

$$\widehat{\mathbf{f}_\lambda^{br*-\varepsilon}} = \widehat{\mathbf{f}_\lambda^{br*}} - \varepsilon = t_{-\varepsilon}(\widehat{\mathbf{f}_\lambda^{br*}}) = t_{-\varepsilon}(t_\varepsilon(\widehat{\mathbf{f}_\lambda^{br}})) \quad (\text{A.7.8})$$

Les courbes correspondantes sont sur la figure A.13. On trace également l'inertie des 5 premiers axes et les facteurs canaux dans le plan factoriel (fig. A.14).

Observations :

- La somme des distances du chi-deux S_{χ^2} entre l'image de départ \mathbf{f}_λ et l'image bruitée décalée et reconstituée $\widehat{\mathbf{f}_\lambda^{br*-\varepsilon}}$ diminue avec la valeur de la translation ε .
- Le RSB de l'image bruitée décalée et reconstituée $\widehat{\mathbf{f}_\lambda^{br*-\varepsilon}}$ augmente avec l'augmentation de la translation ε .
- La somme des carrés des écarts SCE entre les facteurs pixels de l'image de départ $\mathbf{c}_\alpha^{\mathbf{f}}$ et les facteurs pixels de l'image bruitée-décalée $\mathbf{c}_\alpha^{\mathbf{f}^{br*}}$ diminue avec la valeur de la translation ε .
- Le RSB des facteurs pixels de l'image bruitée-décalée $c_{\alpha_k}^{\mathbf{f}^{br*}}$ présente un maximum en fonction de ε . Ce maximum n'est pas le même selon les axes et augmente avec ε selon l'ordre des axes factoriels.
- Le RSB hyperspectral sur l'ensemble des facteurs pixels de l'image bruitée-décalée $\mathbf{c}_\alpha^{\mathbf{f}^{br*}}$ présente également un maximum en fonction de ε .
- L'inertie de chacun des axes diminue avec l'augmentation du décalage.
- Les facteurs canaux de l'image bruitée et décalée $\mathbf{c}_\alpha^{\mathbf{f}^{br*}}$ sont différents de ceux de l'image sans bruit $\mathbf{c}_\alpha^{\mathbf{f}}$ dans le plan factoriel (fig. A.14 (a)).
- Les facteurs canaux sont « plus resserrés » dans le plan factoriel lorsque le décalage ε augmente (fig. A.14 (b)).

Interprétations : La translation ε a pour effet de rapprocher de l'image de départ \mathbf{f}_λ sa reconstitution $\widehat{\mathbf{f}_\lambda^{br*-\varepsilon}}$. Il est important de noter que le RSB des facteurs pixels de l'image bruitée et reconstituée $\mathbf{c}_\alpha^{\mathbf{f}^{br*}}$ s'améliore jusqu'à un maximum. Il est alors intéressant de se placer à ce maximum pour réduire l'influence du bruit sur les facteurs pixels. En outre la somme des carrés des écarts entre les facteurs pixels de l'image bruitée-décalée $\mathbf{c}_\alpha^{\mathbf{f}^{br*}}$ et ceux de l'image de référence $\mathbf{c}_\alpha^{\mathbf{f}}$ diminue. Ceci montre que l'AFC avec décalage permet de réduire encore plus le bruit.

Conclusions : Ajouter une constante, correspondant au maximum du RSB des facteurs pixels, à l'image permet d'obtenir des axes factoriels et une image reconstituée moins bruités.

Remarque A.7.3. Après cette observation très importante, comment trouver a priori la translation ε correspondant à ce maximum ? Nous proposons de choisir comme valeur de ε le minimum, en valeur absolue, de l'image après reconstitution \widehat{f}_λ .

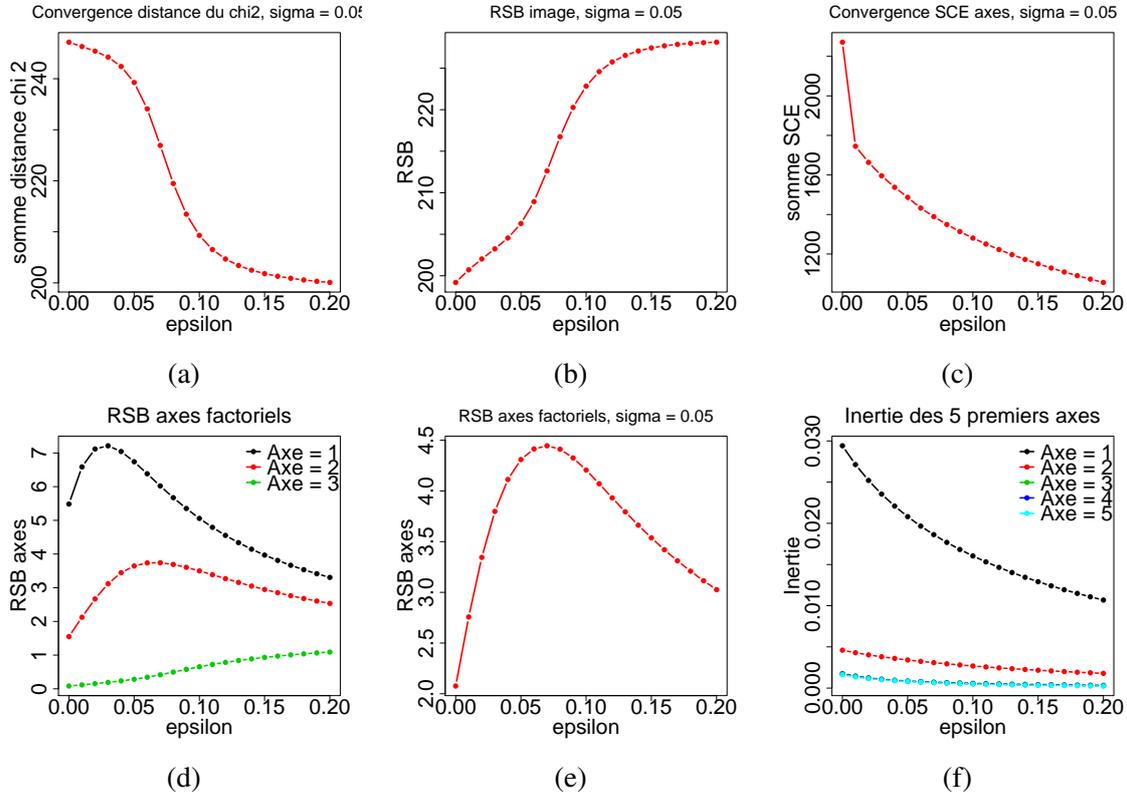


FIG. A.13 – Influence de la translation ε sur l'AFC entre l'image de départ f_λ et l'image bruitée décalée et reconstituée $\widehat{f}_\lambda^{br*-\varepsilon}$: (a) Somme des distances du chi deux (référence image f_λ), (b) RSB de l'image bruitée décalée et reconstituée, (c) Somme des carrés des écarts entre les facteurs pixels c_α^f et c_α^{br*} , (d) RSB pour les facteurs pixels $c_{\alpha_k}^{br*}$ par rapport à $c_{\alpha_k}^f$ (e) Inertie des 5 premiers axes factoriels de l'image décalée et reconstituée.

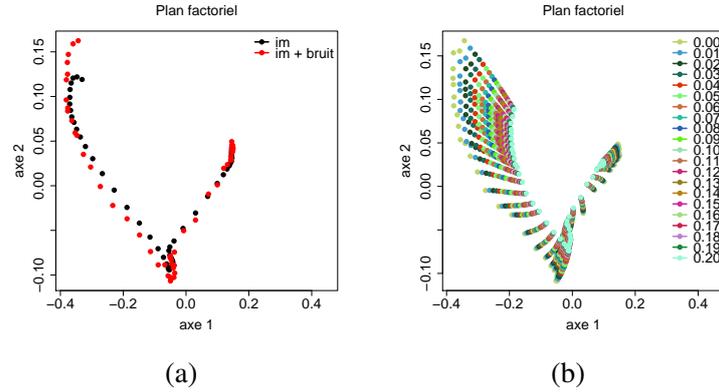


FIG. A.14 – (a) Comparaison entre les facteurs canaux de l'image sans bruit \mathbf{f}_λ et de l'image bruitée \mathbf{f}_λ^{br*} ($\sigma = 0,05$) (b) Influence du décalage $\varepsilon \in [0; 0,2]$ sur les facteurs canaux dans le plan factoriel formé par les 2 premiers axes.

A.8 Covariance et rapports signaux à bruits

Dans la partie 2.4 (p : 22 et suivantes), est expliqué comment mesurer les RSB par ouverture morphologique et comment sont estimées les covariances spatiales \bar{g}_{λ_j} et spectrales $\bar{\Sigma}_{\lambda_j, \lambda_{j'}}$.

Sur la figure A.15, les covariances spatiales et spectrales sont tracées pour les canaux $f_{\lambda_{50}}$ et $f_{\lambda_{13}}$.

Pour le canal $f_{\lambda_{50}}$, on constate que :

- $\bar{g}_{\lambda_{50}}(0) = 2\,207\,890$
- $\bar{\Sigma}(\lambda_{50}, \lambda_{50}) = 2\,208\,140$

donc $\bar{g}_{\lambda_{50}}(0) \simeq \bar{\Sigma}(\lambda_{50}, \lambda_{50})$. L' écart relatif de $\bar{g}_{\lambda_{50}}(0)$ par rapport à $\bar{\Sigma}(\lambda_{50}, \lambda_{50})$ vaut 0,01 %.

Pour le canal $f_{\lambda_{13}}$, on constate que :

- $\bar{g}_{\lambda_{13}}(0) = 4\,080\,225$
- $\bar{\Sigma}(\lambda_{13}, \lambda_{13}) = 4\,080\,688$

donc $\bar{g}_{\lambda_{13}}(0) \simeq \bar{\Sigma}(\lambda_{13}, \lambda_{13})$. L' écart relatif de $\bar{g}_{\lambda_{13}}(0)$ par rapport à $\bar{\Sigma}(\lambda_{13}, \lambda_{13})$ vaut 0,01 %.

On va comparer le RSB de chacun des canaux par ouverture dans le domaine spatial et dans le domaine spectral. Pour le domaine spatial on procède par ouverture bidimensionnelle avec un élément structurant 3×3 pixels (eq. 2.4.11 p : 32). Pour le domaine spectral, on effectue une ouverture unidimensionnelle (1D) avec un élément structurant (e.s.) centré de longueur 3 pixels :

$$RSB_{\lambda_j}^{spectral} = \frac{Var(signal)}{Var(bruit)} = \frac{\gamma \bar{\Sigma}(\lambda_j, \lambda_j)}{\bar{\Sigma}(\lambda_j, \lambda_j) - \gamma \bar{\Sigma}(\lambda_j, \lambda_j)} \quad (A.8.1)$$

On observe que sur la figure A.16, pour le canal $f_{\lambda_{50}}$, la covariance spectrale présente un minimum local au point λ_{50} de valeur $\bar{\Sigma}(\lambda_{50}, \lambda_{50})$. Ainsi l'ouverture de la covariance en ce point $\gamma \bar{\Sigma}(\lambda_{50}, \lambda_{50})$ est égale à la valeur avant ouverture $\bar{\Sigma}(\lambda_{50}, \lambda_{50})$. Comme la variance du bruit est estimée par la différence $\bar{\Sigma}(\lambda_{50}, \lambda_{50}) - \gamma \bar{\Sigma}(\lambda_{50}, \lambda_{50})$ et que cette différence est nulle, on trouve une valeur nulle pour la variance du bruit et donc un RSB infini !

La covariance spectrale $\bar{\Sigma}(\lambda_{50}, \lambda_j)$ ne devrait pas présenter de minimum local en λ_{50} . En fait l'échantillonnage dans la dimension temporelle ne respecte pas le théorème de Shannon. En effet,

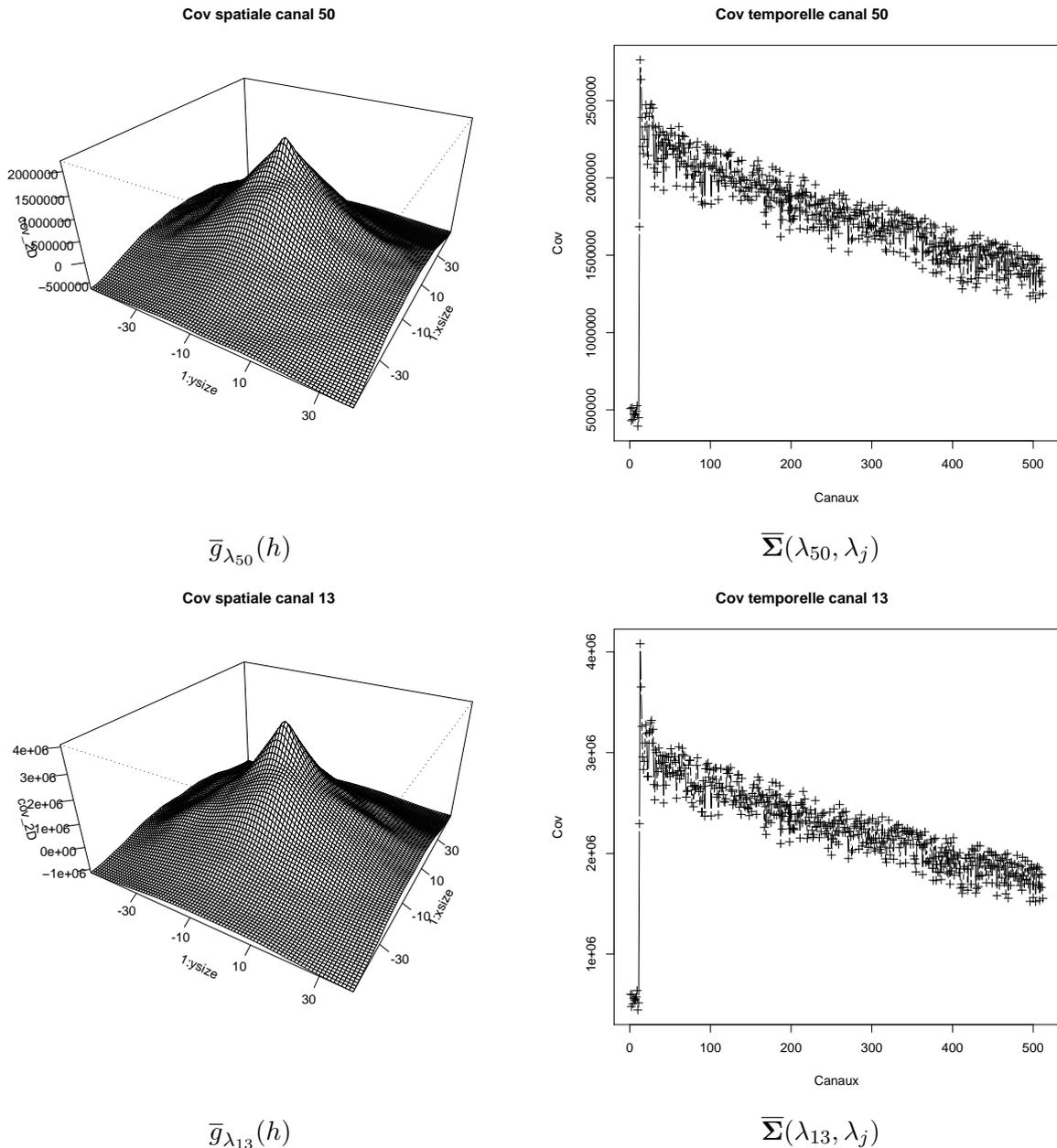


FIG. A.15 – Comparaison entre la covariance spectrale $\bar{\Sigma}(\lambda_{50}, \lambda_j)$ et la covariance spatiale $\bar{g}_{\lambda_{50}}(h)$ ainsi que $\bar{\Sigma}(\lambda_{13}, \lambda_j)$ et $\bar{g}_{\lambda_{13}}(h)$.

d'après les médecins, entre deux canaux il y a eu 4 ou 5 battements cardiaques et la fréquence d'acquisition n'est a priori pas en phase avec la fréquence cardiaque. L'échantillonnage spectral ne permet donc pas d'estimer le RSB spectral par ouverture sur le canal $f_{\lambda_{50}}$.

Par contre, pour le canal $f_{\lambda_{13}}$ on a :

- $\bar{\Sigma}(\lambda_{13}, \lambda_{13}) = 4\,080\,688$.
- $\gamma \bar{\Sigma}(\lambda_{13}, \lambda_{13}) = 3\,260\,230$.

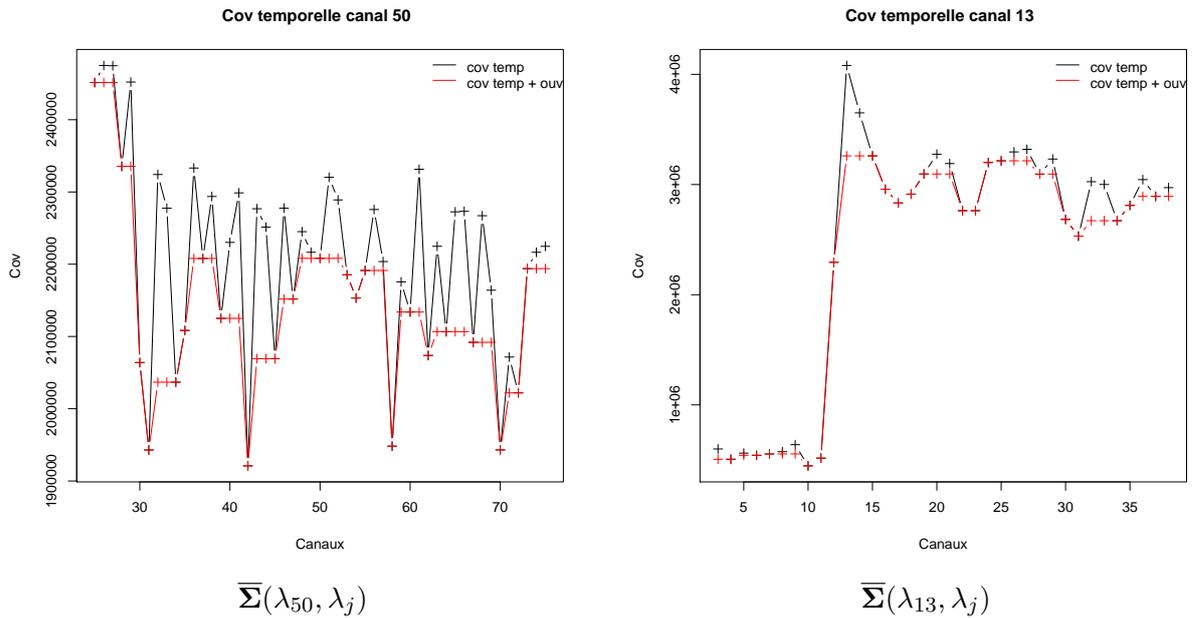


FIG. A.16 – Covariances spectrales $\overline{\Sigma}(\lambda_j, \lambda_{j'})$ pour les canaux $f_{\lambda_{50}}$ et $f_{\lambda_{13}}$ avant et après ouverture γ unidimensionnelle par un e.s. centré de taille 3 pixels.

- le RSB spectral vaut donc $RSB_{\lambda_{13}}^{spectral} = 3,97$.

Cette valeur correspond plutôt à un ordre de grandeur qu'à une valeur précise car l'échantillonnage spectral ne respecte pas le théorème de Shannon.

En utilisant la covariance spatiale $\overline{g}_{\lambda_j}(h)$ (fig. A.15) on trouve pour le canal $f_{\lambda_{13}}$:

- $\overline{g}_{\lambda_{13}}(0) = 4\,080\,225$.
- $\gamma \overline{g}_{\lambda_{13}}(0) = 3\,827\,412$.
- le RSB spatial vaut donc $RSB_{\lambda_{13}} = 15,14$.

De même pour le canal $f_{\lambda_{50}}$:

- $\overline{g}_{\lambda_{50}}(0) = 2\,207\,890$.
- $\gamma \overline{g}_{\lambda_{50}}(0) = 2\,049\,623$.
- le RSB spatial vaut donc $RSB_{\lambda_{50}} = 12,95$.

Au vu de ces résultats, on constate que le RSB spectral ne peut pas être calculé pour tous les canaux, et que même pour les canaux pour lesquels il peut être calculé on n'obtient qu'un ordre de grandeur. Par conséquent, on utilise uniquement le RSB spatial.

Annexe B

Segmentation morphologique

La morphologie mathématique est une méthode non-linéaire de traitement d'image initiée dans les années 60 par les travaux de *G. Matheron* [113, 115] et *J. Serra* [159, 160] à l'Ecole des Mines de Paris. Aujourd'hui, la morphologie mathématique est considérée comme une branche très importante de l'analyse d'image avec des opérateurs capables de réaliser des tâches très sophistiquées sur des images binaires, à niveaux de gris, couleurs, hyperspectrales, des images 3D et des séquences vidéo ...

Cette partie n'étant pas introduction à la morphologie, a pour but de rappeler quelques notions de segmentation morphologique utilisées dans ce mémoire. Le lecteur intéressé pourra se reporter aux ouvrages de référence de : *G. Matheron* [113, 115], *J. Serra* [159, 160] ainsi qu'aux livres de *Coster et Chermant* [44], *Heijmans* [77] et *Soille* [172]. Pour la morphologie couleur la thèse de *J. Angulo* [2] et [3] constituent une bonne référence.

Pour présenter cette partie nous nous sommes inspirés de [2] qui présente de manière concise la morphologie mathématique. Ce chapitre vient compléter celui sur la segmentation (chapitre 3).

B.1 Notions élémentaires

Introduisons tout d'abord quelques notions sur les treillis et quelques propriétés de base des transformations morphologiques.

B.1.1 Treillis et morphologie mathématique

La structure fondamentale en morphologie mathématique est le treillis complet et les lois premières sont le supremum et l'infimum.

Définition 24 (Treillis complet). Un treillis complet \mathcal{L} est un ensemble muni d'une relation d'ordre partiel \leq ($x \leq x$, $x \leq y$, $y \leq x \Rightarrow x = y$, $x \leq y$, $y \leq z \Rightarrow x \leq z$, $x, y, z \in \mathcal{L}$) tel que pour tout sous-ensemble \mathcal{K} fini ou non il, existe un infimum (plus grand minorant) $\wedge \mathcal{K}$ et un supremum (plus petit majorant) $\vee \mathcal{K}$; où $\mathcal{K} = \{X_i, i \in I\}$. L'infimum et le supremum sont donnés par $\wedge X_i$ et $\vee X_i$ respectivement.

Les opérateurs morphologiques agissent sur des ensembles en morphologie binaire et sur des fonctions en morphologie numérique. Nous allons donc présenter les treillis correspondant à ces deux types d'images,

Définition 25 (Treillis des parties d'un ensemble). Soit E un ensemble arbitraire non vide. L'ensemble $\mathcal{P}(E)$ de tous les sous-ensembles de E , dont l'ordre est défini par la relation d'inclusion \subseteq , est un treillis complet pour les opérations \cup (union) et \cap (intersection). De plus pour chaque élément $X \in \mathcal{P}(E)$ il existe un unique $X^C \in \mathcal{P}(E)$, appelé le complément de X , tel que $X \cap X^C = \emptyset$ et $X \cup X^C = E$.

L'ensemble des réels $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ est un treillis complet. De même pour l'ensemble des entiers relatifs $\overline{\mathbb{Z}} = \mathbb{Z} \cup \{-\infty, +\infty\}$.

Définition 26 (Treillis des fonctions numériques). Étant donné E un espace arbitraire et \mathcal{T} désignant $\overline{\mathbb{R}}$, $\overline{\mathbb{Z}}$ ou une de leurs parties fermées, les fonctions $\mathcal{F} = \{f : E \rightarrow \mathcal{T}\}$ forment à leur tour un treillis, noté \mathcal{F}^E ou $\mathcal{F}(E, \mathcal{F})$, pour l'ordre suivant $\forall f, g \in \mathcal{F}, f \leq g \Leftrightarrow \forall x \in E, f(x) \leq g(x)$; et où le supremum et l'infimum dérivent des sup et inf numériques, i.e. $f = \vee f_i \Leftrightarrow f(x) = \sup f_i(x)$ et $f' = \wedge f_i \Leftrightarrow f(x) = \inf f_i(x), \forall x \in E$.

Dans [2], cette démarche s'étend directement aux produits de treillis de type \mathcal{T} , c'est à dire aux fonctions multivariées (images couleur).

B.1.2 Propriétés de base des transformations morphologiques

En général, les transformations morphologiques sont dotées de propriétés importantes relatives aux opérations sur les treillis. Nous allons présenter ces propriétés pour les treillis de l'ensemble des parties de E , $\mathcal{P}(E)$, et des fonctions numériques (i.e. scalaires ou à niveaux de gris), $\mathcal{F}(E, \mathcal{T})$.

Définition 27 (Croissance). Une transformation ψ est dite croissante si et seulement si elle préserve l'ordre :

$$\forall X, Y \in \mathcal{P}(E), X \subseteq Y \Rightarrow \psi(X) \subseteq \psi(Y), \text{ ou } \forall f, g \in \mathcal{F}, f \leq g \Rightarrow \psi(f) \leq \psi(g).$$

Dans le cas contraire ($X \subseteq Y \Rightarrow \psi(X) \supseteq \psi(Y)$, ou $f \leq g \Rightarrow \psi(f) \geq \psi(g)$), la transformation ψ est dite décroissante. La préservation de l'ordre dans les opérateurs (i.e. opérateurs croissants) est une caractéristique fondamentale des transformations morphologiques premières. Elle assure que l'ordre, structure de base du treillis, est conservé lors d'un filtrage morphologique.

Définition 28 (Extensivité). Une transformation ψ est dite extensive si et seulement si son résultat est toujours plus grand que l'original :

$$\forall X \in \mathcal{P}(E), X \subseteq \psi(X), \text{ ou } \forall f \in \mathcal{F}, f \leq \psi(f).$$

Dans le cas contraire ($X \supseteq \psi(X)$, ou $f \geq \psi(f)$), la transformation ψ est dite anti-extensive.

Définition 29 (Idempotence). Une transformation ψ est dite idempotente lorsqu'elle est invariante par itération :

$$\psi(\psi) = \psi \circ \psi = \psi.$$

Définition 30 (Dualité). Deux transformations ψ_1 et ψ_2 sont duales si et seulement si appliquer l'une revient à appliquer l'autre sur le complémentaire (symétrie ou négation pour les fonctions) de l'initial puis à complémenter le résultat final :

$$\forall X \in \mathcal{P}(E), \psi_1(X) = [\psi_2(X^C)]^C, \text{ ou } \forall f \in \mathcal{F}, \psi_1(f) = -(\psi_2(-f)).$$

En fait, la dualité fondamentale entre le supremum et l'infimum se transmet à tous les outils de la morphologie mathématique. En général, les transformations vont par paires qui se correspondent par dualité (par exemple l'érosion et la dilatation). On trouve aussi des opérateurs autoduaux, i.e. $\psi(X) = [\psi(X)]^C$.

B.2 Théorie générale de la segmentation

Dans cette section nous résumons les résultats principaux de la définition formelle de la segmentation d'image dans le cadre de la théorie des treillis introduite récemment par Serra [166].

B.2.1 Critère et segmentation

Habituellement, on appelle critère une clef qui permet de distinguer entre des assertions vraies et fausses. Dans le domaine de la segmentation, la notion est formalisée de la manière suivante [166] :

Définition 31 (Critère). Soient E et T deux ensembles arbitraires et soit \mathcal{F} une famille de fonctions de E dans T . Un critère σ sur la classe \mathcal{F} est une fonction binaire de $\mathcal{F} \otimes \mathcal{P}(E)$ dans $\{0, 1\}$ telle que, pour chaque fonction $f \in \mathcal{F}$, et pour chaque ensemble $A \in \mathcal{P}(E)$, l'on ait :

ou bien $\sigma[f, A] = 1$ (le critère est satisfait sur A)

ou bien $\sigma[f, A] = 0$ (le critère est refusé sur A).

De plus, de manière conventionnelle, on décide que, pour toutes les fonctions, tous les critères sont satisfaits sur l'ensemble vide, i.e.

$$\sigma[f, \emptyset] = 1 \quad \forall f \in \mathcal{F}$$

Par exemple, le critère qui est satisfait si

$$x \in A, \quad t_0 \leq f(x) \leq t_1$$

avec t_0 et t_1 fixés, et $t_0 \leq t_1$, définit le critère de seuillage appliqué à la fonction f .

Définition 32 (Partition). Soit E un ensemble arbitraire. On appelle partition de l'espace E toute application $D : E \rightarrow \mathcal{P}(E)$ telle que :

- (i) $\forall x \in E : x \in D(x)$ et
- (ii) $\forall x, y \in E : D(x) = D(y)$ ou $D(x) \cap D(y) = \emptyset$.

$D(x)$ se nomme la classe de la partition qui contient x .

L'ensemble des partitions d'un ensemble arbitraire E est ordonné comme suit.

Définition 33 (Ordre de partitions). Une partition \mathcal{A} est dite plus fine (respectivement plus grossière) qu'une partition \mathcal{B} , $\mathcal{A} \leq \mathcal{B}$ (respectivement $\mathcal{A} \geq \mathcal{B}$), quand chaque classe de \mathcal{A} est incluse dans une classe de \mathcal{B} .

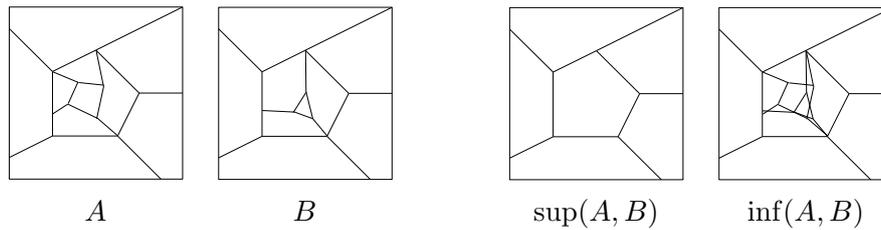


FIG. B.1 – Exemple de supremum et d'infimum de deux partitions A et B .

Ceci conduit à la notion de hiérarchie ordonnée de partitions $\Pi_{i=1}^N D_i$, telle que $D_i \leq D_{i+1}$.

Définition 34 (Hiérarchie de partitions). Soit Π un ensemble de partitions $\{D_i\}$, $i \in I$ associées à l'espace E . Nous dirons que Π forme une hiérarchie de partitions s'il est possible d'établir un ordre d'inclusion parmi toute paire d'éléments de l'ensemble $\Pi = \{D_i\}$, i.e. $\forall i, j D_i \cap D_j \in \{D_i, D_j\}$.

Le fait d'avoir un ordre entre les partitions implique que la hiérarchie de partitions forme un treillis complet [166].

Définition 35 (Treillis de partitions). Pour l'ensemble arbitraire E , les partitions $D : E \rightarrow \mathcal{P}(E)$ forment un treillis complet \mathcal{D} , où le plus grand élément de \mathcal{D} est E lui-même, et le plus petit celui qui pulvérise E en la totalité de ses points. Étant donnée la famille de partitions indexées $\{D_i, i \in I\}$, l'application $D : E \rightarrow \mathcal{P}(E)$ définie à partir de ses classes $D(x) = \bigwedge \{D_i, i \in I\}$ génère une partition où $\forall x \in E$, $D(x)$ est le plus grand élément de $\mathcal{P}(E)$ qui est contenu dans chaque $D_i(x)$. Par conséquent, \mathcal{D} est l'inf des D_i . L'expression du sup, $D = \bigvee D_i$, signifie que $\forall x \in E$, $\forall i \in I$, la classe $D_i(x)$ est le plus petit ensemble qui est l'union des classes $D_i(y)$, $y \in E$, voir figure B.1.

Pour une partition donnée, on peut toujours construire une partition plus fine, à savoir celle qui réduit l'espace $\mathcal{P}(E)$ à tous ses singletons. Cette remarque nous conduit à remplacer la notion de partition par la notion de segmentation.

Quand $\sigma[f, A] = 1$, on dit que f est homogène dans A pour le critère σ .

Définition 36 (Segmentation). Étant donné une fonction $f \in \mathcal{F}$, un ensemble arbitraire $A \subseteq E$ et un critère σ , soit $\{D_i(A), i \in I\}$ la famille de toutes les partitions de l'ensemble A en zones homogènes de f selon σ . Le critère σ segmente la fonction f lorsque,

- (i) $\sigma[f, \{x\}] = 1, \forall \{x\} \in \mathcal{P}(E)$;
- (ii) la famille $\{D_i(A)\}$ admet un supremum $\bigvee D_i(A)$.

Alors la partition supremum $\bigvee D_i(A)$ définit la segmentation de f sur l'ensemble A selon le critère σ .

Cette définition est cohérente : selon le premier axiome, la partition de A en tous ses singletons appartient à la famille $\{D_i(A)\}$, qui, donc, n'est jamais vide. En particulier, quand A coïncide avec l'espace E de définition de f , alors $\bigvee D_i(E)$ est la partition maximale de tout l'espace. L'axiome $\sigma[f, \{x\}] = 1$ nous garantit que quel que soit le critère considéré, il y a toujours au moins une manière de partitionner E en zones (les singletons) qui le satisfasse. C'est une condition nécessaire pour pouvoir alors parler de plus grande partition.

La partition de segmentation $D = \bigvee \{D_i(A)\} = \bigvee \{D_i(f)\}$ décompose E en zones qui sont :

- disjointes et qui couvrent la totalité de l'espace E ;

- où la fonction f est homogène selon le critère σ ;
- et où la classe de la partition en chaque point est la plus grande possible à vérifier le critère σ .

Pour utiliser cette notion en pratique, à moins que nous nous mettions à vérifier toutes les partitions possibles pour chaque fonction f , nous avons besoin d'un théorème qui relie le concept de segmentation avec des outils manipulables. La notion qui convient ici est le concept de critère connectif.

B.2.2 Critères connectifs et théorème de la segmentation

Il faut ensuite caractériser les conditions que doit vérifier le critère σ pour devenir un outil de segmentation. L'apparition de partitions maximales, comme aussi bien l'exemple des zones plates nous orientent vers l'axiomatique des connexions par l'intermédiaire du concept de critère connectif.

Définition 37 (Critère connectif). Un critère $\sigma : \mathcal{F} \otimes \mathcal{P}(E) \rightarrow [0, 1]$ est connectif :

- quand il est vérifié sur la classe \mathcal{S} des singletons

$$\forall f \in \mathcal{F}, \{x\} \in \mathcal{S} \Rightarrow \sigma[f, \{x\}] = 1$$

- et quand pour toute fonction $f \in \mathcal{F}$ et pour toute famille $\{A_i\}$ dans $\mathcal{P}(E)$ on a

$$\bigcap A_i \neq \emptyset \text{ et } \bigwedge \sigma[f, A_i] = 1 \Rightarrow \sigma[f, \bigcup A_i] = 1$$

Autrement dit, si le critère est vérifié par la fonction f sur un ensemble de régions A_i de l'espace, et que ces régions ont un point en commun, alors il est aussi vérifié sur réunion $\bigcup A_i$. Les critères connectifs caractérisent les processus de segmentation comme le montre le théorème suivant :

Théorème B.2.1 (Segmentation - J. Serra). *Considérons deux ensembles arbitraires E et \mathcal{T} , et une famille \mathcal{F} de fonctions $f : E \rightarrow \mathcal{T}$. Soit d'autre part σ un critère sur la classe \mathcal{F} . Alors les trois propositions suivantes sont équivalentes :*

1. le critère σ est connectif;
2. pour chaque fonction $f \in \mathcal{F}$, les ensembles sur lesquels le critère σ est satisfait génèrent une connexion \mathcal{C} :

$$\mathcal{C} = \{A \mid A \in \mathcal{P}(E) \text{ et } \sigma[f, A] = 1\};$$

3. le critère σ segmente toutes les fonctions de la famille \mathcal{F} .

La segmentation de la fonction f par rapport au critère connectif σ est la partition $D_\sigma(f(x))$ ($\sigma[f] \rightarrow D_\sigma(f(x))$) dont la classe pour chaque point x est donnée par :

$$D_\sigma(f(x)) = \bigcap \{A : A \in \mathcal{P}(E), x \in A, \sigma[f, A] = 1\}, x \in E.$$

Le théorème n'aurait pas pu être énoncé dans le cadre de la seule connectivité par arcs : beaucoup de critères de segmentation ne mettent pas en jeu de chemins, comme par exemple la connexion par sauts [166]. En fait la notion de connexion possède exactement le niveau de généralité pour que son identification à la segmentation soit possible.

Très souvent en pratique, lorsque la fonction f est fixée, la fonctionnelle marginale $\sigma[f, A] : \mathcal{P}(E) \rightarrow \{0, 1\}$ est décroissante, i.e.

$$\sigma[f, A] = 1 \text{ et } B \subseteq A \Rightarrow \sigma[f, B] = 1.$$

Si cette implication est vraie pour toutes les fonctions $f \in \mathcal{F}$, on dit que le critère est décroissant. Le théorème de la segmentation est donc accompagné de la proposition suivante :

Proposition B.2.2. *Si un critère σ sur $\mathcal{F} \otimes \mathcal{P}(E)$ est décroissant, les deux énoncés suivants sont équivalents :*

1. *le critère σ est connectif;*
2. *le critère σ segmente toutes les fonctions de la famille \mathcal{F} .*

Le théorème et la proposition associée appartiennent à cette catégorie de résultats qui reformulent une définition impossible à mettre en œuvre (comment savoir si une partition du plan est la plus grande possible ?) en critères plus faciles à manipuler. Cependant, en lui-même, le théorème ne fournit aucune « recette » pour découvrir tel ou tel « bon » critère connectif.

B.2.3 Demi-treillis des critères connectifs

Nous considérons maintenant que la fonction $f \in \mathcal{F}$ est fixée, et nous étudions la structure logique de plusieurs critères connectifs qu'on peut appliquer à f . Voici les deux propositions de *Serra* qui aident à combiner les segmentations entre elles.

Proposition B.2.3. *Lorsque la fonction f est fixée, la famille Σ de tous les critères connectifs forme un demi-treillis où l'infimum correspond à l'intersection logique des critères, et où le plus petit élément est la partition de l'espace en ses singletons.*

Par exemple, le critère « la fonction f est constante sur A » conduit à partitionner E selon toutes les valeurs de seuil de f . Si l'on impose en plus que A doive être connexe, on obtient alors la segmentation de f en zones plates et connexes.

La bonne propriété de « ET » logique des critères connectifs n'a malheureusement pas de contrepartie pour le « OU », car le supremum de partitions ne met pas en jeu des classes sur lesquelles un seul critère au moins serait vérifié. La proposition suivante montre cependant qu'en posant le problème différemment, on peut préciser sous quelles conditions le « OU » logique de critères connectifs signifie quelque chose.

Proposition B.2.4. *Soient $f \in \mathcal{F}$ une fonction donnée, et $\{D_i, i \in I\}$ une famille de segmentations de f de critères $\{\sigma_i, i \in I\}$. Le supremum $D' = \vee D_i$ est aussi la plus grande partition de E en classes où un critère σ_{i_0} au moins est vérifié si et seulement si chaque classe $D'(x)$ de D' est simultanément une classe d'une partition D_i au moins.*

En pratique cette proposition sert surtout lorsque la classe non ponctuelle $D_{i_0}(x)$ recouvre des classes non ponctuelles pour tous les autres critères $\sigma_i, i \neq i_0$.

B.3 Connexion

La notion de connexion a fait l'objet d'importants travaux théoriques récents, notamment les contributions de *Serra* [161, 162, 163], *Ronse* [151, 152] et *Braga-Neto et Goutsias* [32, 33]. Nous allons suivre le raisonnement de *Serra* dans [165] pour arriver au concept de connexion.

En mathématiques, la connexité est une notion ensembliste de nature topologique, qui s'exprime de deux façons. D'une manière générale, on dit qu'un ensemble est connexe quand on ne peut pas le partitionner en deux régions non vides. Toutefois, par sa généralité même, elle ne peut tirer avantage de la régularité de certains contextes, comme l'espace euclidien. C'est ici qu'apparaît la connexité par arcs, selon laquelle un ensemble A est connexe, lorsque pour toute paire de points $a, b \in A$, on peut trouver une application continue ψ de $[0, 1]$ dans A telle que $\psi(0) = a$ et $\psi(1) = b$. Cette connexité est plus restrictive que la première ; cependant dans \mathbb{R}^n , les deux notions coïncident pour les ensembles ouverts. Dans le cas discret, les connexités numériques transposent la démarche euclidienne par arcs, en introduisant des arcs élémentaires entre points voisins (ce qui correspond au concept de voisinage à 4 ou 8 connexité de la trame carrée).

L'une et l'autre définition mènent à la propriété suivante, qui est plus riche de sens. Si l'on prend une famille d'ensemble connexes qui possèdent au moins un point en commun, alors leur réunion est elle aussi connexe :

$$\{A_i \text{ connexes}\} \text{ et } \{\cap A_i \neq \emptyset\} \Rightarrow \{\cup A_i \text{ connexe}\}. \quad (\text{B.3.1})$$

Cependant, la définition de connexité topologique n'est guère pratique qu'en traitement des images binaires. En effet, les objets d'étude ne sont pas toujours des ensembles, mais surtout des fonctions numériques ou vectorielles. De plus, une bonne définition de la connexité devrait être plus opérante et généralisable. Cette critique et d'autres sont à l'origine de la notion de connexion proposée par Serra [160] :

Définition 38 (Connexion). Soit E un ensemble arbitraire non vide. On appelle classe connexe ou connexion \mathcal{C} toute famille de $\mathcal{P}(E)$ telle que :

- (0) $\emptyset \in \mathcal{C}$
- (i) quel que soit $x \in E$, $\{x\} \in \mathcal{C}$,
- (ii) pour toute famille $\{C_i, i \in I\}$ de \mathcal{C} , $\cap_i C_i \neq \emptyset$ entraîne $\cup_i C_i \in \mathcal{C}$.

où $\{x\}$ désigne un singleton, i.e. un élément ponctuel de $\mathcal{P}(E)$.

A partir de cette définition, tout ensemble C d'une classe connexe \mathcal{C} est dit connexe. En outre, un espace topologique donné induit une connexité unique, alors qu'on trouvera en général une infinité de connexions dans $\mathcal{P}(E)$. Il est clair qu'une telle définition ne nécessite aucune notion topologique. Les notions classiques (e.g. la connexité numérique ou par arcs euclidiens) sont donc des cas particuliers.

Étant donné un ensemble $A \in \mathcal{P}(E)$ et un point $x \in A$, l'axiome (ii) suggère de prêter une attention particulière à la réunion, elle-même connexe, de toutes les composantes connexes contenant x et incluses dans A :

$$\gamma_x(A) = \cup \{C; C \in \mathcal{C}, x \in C \subseteq A\}. \quad (\text{B.3.2})$$

Voici maintenant deux propriétés de base des connexions, qui sont très intéressantes pour la théorie de la segmentation.

Proposition B.3.1 (Treillis des connexions). *L'ensemble de toutes les connexions sur $\mathcal{P}(E)$ est stable pour l'intersection ; c'est même un treillis complet dans lequel le supremum de la famille $\{C_i, i \in I\}$ est la plus petite connexion contenant $\cup C_i$ [161].*

Proposition B.3.2 (Croissance). *Si \mathcal{C} et \mathcal{C}' sont deux connexions sur $\mathcal{P}(E)$, avec $\mathcal{C} \subseteq \mathcal{C}'$, alors toute \mathcal{C}' -composante de A est le supremum des \mathcal{C} -composantes de A qu'elle majore [160].*

B.4 Nivellements

Les nivellements ont été introduits par *Meyer* [119, 120, 127, 128], comme opérateurs connexes agissant sur les fonctions numériques et définies sur \mathbb{R}^2 muni d'une connexité par arcs. La théorie a aussi été développée par *Matheron* [116], qui a proposé une généralisation à un espace arbitraire (sans une connexion a priori). *Serra* [162] a également ré-interprétés les nivellements comme des opérateurs binaires dans le cadre d'un espace E quelconque muni de la connexion \mathcal{C} pour ensuite les étendre aux fonctions numériques.

Nous présentons dans cette section quelques résultats sur cette classe de filtres morphologiques qui devient de plus en plus importante dans les applications de segmentation [63, 190] ; car les nivellements simplifient les images sans déformer ni déplacer les objets.

Nous présentons les nivellements en terme de voisinage. Pour ce faire nous avons suivi le raisonnement de *Zanoguera* [190]. Il existe une autre formulation en terme de ensembliste définie par *Serra* [162]. Pour en avoir un aperçu, nous invitons le lecteur à se reporter à [2].

Étant donnée une fonction numérique $f \in \mathcal{F}(E, \mathcal{T})$, nous noterons ici par f_s la valeur de la fonction f au point $s \in E$. Quelques définitions préalables sont nécessaires :

Définition 39 (Zones plates). Deux points $x, y \in E$ appartiennent à la même zone plate d'une fonction f si et seulement s'il existe un chemin de points (p_0, p_1, \dots, p_n) tel que $p_0 = x$ et $p_n = y$ et, $\forall i$, $(p_i, p_{i+1}) \in E^2$ sont voisins et $f_{p_i} = f_{p_{i+1}}$.

Notons qu'il n'y a pas de restriction sur la taille minimale ni maximale des zones plates, qui peuvent d'ailleurs être constituées d'un seul point. Remarquons aussi que l'ensemble des zones plates d'une image constitue une partition du support de cette image puisque tout point de l'image appartient à une et une seule zone plate. Par opposition avec les zones plates, nous définissons les zones de transition :

Définition 40 (Zones de transition). Deux points voisins (p, q) forment une transition s'ils n'appartiennent pas à la même zone plate, i.e. $f_p > f_q$ ou $f_p < f_q$.

On note le niveau de gris de la fonction f au point p : $f(p) = f_p$.

Cette définition nous permet de définir un opérateur connexe ψ associé à une certaine connexion :

Définition 41 (Opérateur connexe). Un opérateur connexe ψ transforme une fonction f en une fonction $g = \psi(f)$ de façon que $\forall (p, q)$ voisins on ait $f_p = f_q \Rightarrow g_p = g_q$.

C'est à dire qu'à toute zone plate dans f correspond forcément une zone plate dans g et par conséquent g a, au plus, le même nombre de zones plates que f ; et à toute transition dans g doit correspondre une transition dans f . Un opérateur connexe agit donc par fusion des zones plates de la fonction de départ.

A partir de ces différentes notions, nous pouvons, maintenant, introduire les nivellements.

Définition 42 (Sous-nivellement). Une fonction g est un sous-nivellement d'une fonction f si et seulement si pour toute paire de points voisins p et q :

$$g_p > g_q \Rightarrow g_q \geq f_q$$

Définition 43 (Sur-nivellement). Une fonction g est un sur-nivellement d'une fonction f si et seulement si pour toute paire de points voisins p et q :

$$g_p > g_q \Rightarrow g_p \leq f_p$$

Définition 44 (Nivellement). Une fonction g est un nivellement d'une fonction f si et seulement si elle est un sous- et un sur-nivellement de f :

$$g_p > g_q \Rightarrow f_p \geq g_p \geq g_q \geq f_q$$

Les nivellements agissent en imposant deux conditions supplémentaires à la condition de connexité :

1. la réduction de la dynamique de transitions à la sortie : c'est à dire que pour chacune des transitions de la fonction g , la transition correspondante sur f a une amplitude égale ou plus grande que celle de g ;
2. le couplage du sens de la variation des transitions : ainsi, une transition de haut en bas sur la fonction g correspondra forcément à une transition de haut en bas sur f , et vice-versa.

Étant données une fonction f , dite fonction *référence*, et une fonction quelconque g qu'on appelle *marqueur*, il existe des algorithmes permettant de transformer la fonction marqueur g en un nivellement $g' = \lambda(f, g)$ de f . On obtient le sous- ou sur- nivellement de f à partir d'un marqueur g par itération jusqu'à idempotence des algorithmes suivants, où δ et ε sont respectivement la dilatation et l'érosion morphologique unitaire :

1. sous-nivellement, niv^δ : pour tout pixel p , si $g_p < f_p$, alors $g'_p = f_p \wedge (\delta g)_p$;
2. sur-nivellement, niv^ε : pour tout pixel p , si $g_p > f_p$, alors $g'_p = f_p \vee (\varepsilon g)_p$.

En appliquant le sous-nivellement lorsque $g < f$ et le sur-nivellement quand $g > f$ on obtient un algorithme parallèle autodual des nivellements qui consiste à itérer jusqu'à idempotence la transformation élémentaire $g' = (f \wedge \delta g) \vee \varepsilon g$. Les nivellements peuvent être programmés de façon plus efficace, voir [63].

Le choix de la fonction marqueur g détermine le degré de simplification obtenu. Ainsi, des simplifications plus ou moins fortes peuvent être obtenues en choisissant des marqueurs plus ou moins éloignés de la fonction de référence f .

B.5 La segmentation morphologique : du SKIZ à la LPE

Dans cette sous-section et la suivante, nous reprenons les explications du livre de *Soille* [172].

La segmentation d'une image f est une partition de son domaine de définition $\mathcal{D}_f = E$ en n ensembles disjoints non vides X_1, X_2, \dots, X_n appelés les segments telle que l'union des segments soit égale à \mathcal{D}_f . Pour une théorie de la segmentation se reporter à la sous-section B.2.

Introduisons tout d'abord les techniques qu'utilise la Ligne de Partage des Eaux ou LPE : la croissance de région et la détection de contours.

Dans le cadre de la croissance de région, les régions homogènes sont d'abord détectées. La croissance de ces régions se fait à partir de mesure de ressemblance qui combinent des attributs spatiaux et spectraux. Cette croissance est appliquée jusqu'à ce que tous les pixels soient assignés à une région. Les frontières de régions sont créés lorsque deux régions croissantes se rencontrent.

Les techniques par détection de contours, procèdent de la manière opposée. Comme les objets de l'image sont supposés avoir de faibles variations de niveaux de gris, leurs contours sont caractérisés par de fortes variations de niveaux de gris dans leur voisinage. La tâche de la détection de contour, consiste à améliorer et à détecter ces variations. Les variations de niveaux de gris locales sont améliorées par un opérateur gradient. L'image du gradient (i.e. sa norme) est alors utilisée pour déterminer une carte de contours. Une approche simpliste consiste à seuiller l'image gradient pour toutes les valeurs du gradient supérieures à un seuil. Malheureusement, les contours résultants sont rarement connexes. Un traitement supplémentaire est nécessaire pour obtenir des contours fermés correspondant aux frontières des objets.

L'approche morphologique de la segmentation d'image combine les techniques de croissance de région et de détection de contours : elle regroupe les pixels autour des minima régionaux de l'image et les frontières des groupes adjacents qui sont précisément situées le long des lignes de crêtes de l'image gradient. Ceci est réalisé par une transformation appelée la Ligne de Partage des Eaux (LPE).

B.5.1 Le squelette par zones d'influence : SKIZ

Dans cette section, nous introduisons le squelette par zones d'influence : SKIZ.

Définition 45 (Squelette par Zones d'Influence : SKIZ - *C. Lantuejoul*). Considérons un compact X de \mathbb{R}^2 . La zone d'influence d'une composante X_i de X est l'ensemble des points du plan qui sont plus près de X_i que de toute autre composante. Le SKIZ est alors la frontière de ces zones d'influence.

Un exemple de SKIZ est montré sur la figure B.2.

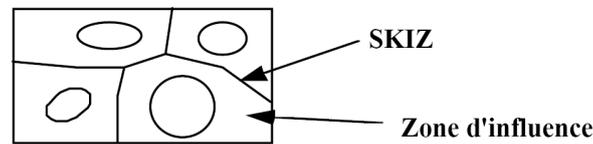


FIG. B.2 – Squelette par zones d'influence (SKIZ). (Source : Cours *J. Serra* [164]).

La définition du SKIZ nécessite l'utilisation d'une distance pour définir les zones d'influence. Dans la segmentation par LPE on sera amené à utiliser les zones d'influences géodésiques et le SKIZ géodésique.

Définition 46 (Zone d'influence géodésique). Soit, dans \mathbb{R}^2 , $Y = \cup_{i \in I} Y_i$ formée de I composantes connexes compactes, toutes incluses dans le compact X . La zone d'influence géodésique de Y_i dans X se définit comme le lieu des points a de X qui sont géodésiquement plus proches de Y_i que de toute autre composante connexe de Y :

$$ZI_X(Y_i) = \{a \in X, \forall k \neq i, d_X(a, Y_j) \leq d_X(a, Y_k)\} \quad (\text{B.5.1})$$

La distance géodésique $d_X(a, Y)$ est la plus petite des distances géodésiques du point a à tous ceux de l'ensemble Y .

Définition 47 (SKIZ géodésique). Le SKIZ géodésique est alors la frontière entre les différentes zones d'influence géodésiques.

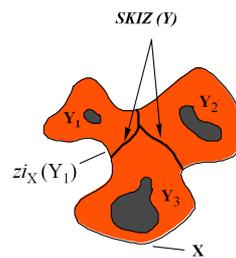


FIG. B.3 – Squelette par zones d’influence géodésique (SKIZ géodésique). (Source : Cours *J. Serra* [164]).

B.5.2 La Ligne de Partage des Eaux (LPE)

Considérons la représentation topographique d’une image à niveaux de gris, i.e. on associe une cote z au niveau de gris du pixel à la position (x, y) . Maintenant, considérons une goutte d’une chute d’eau tombant sur une telle surface. Selon les lois de la gravité, elle va s’écouler le long de la pente la plus raide jusqu’à atteindre un minimum. L’ensemble des points de la surface dont les chemins les plus raides rejoignent un minimum donné constituent un bassin versant associé à ce minimum. Les lignes de partage des eaux sont les zones qui divisent les bassins versants adjacents (fig. B.4).

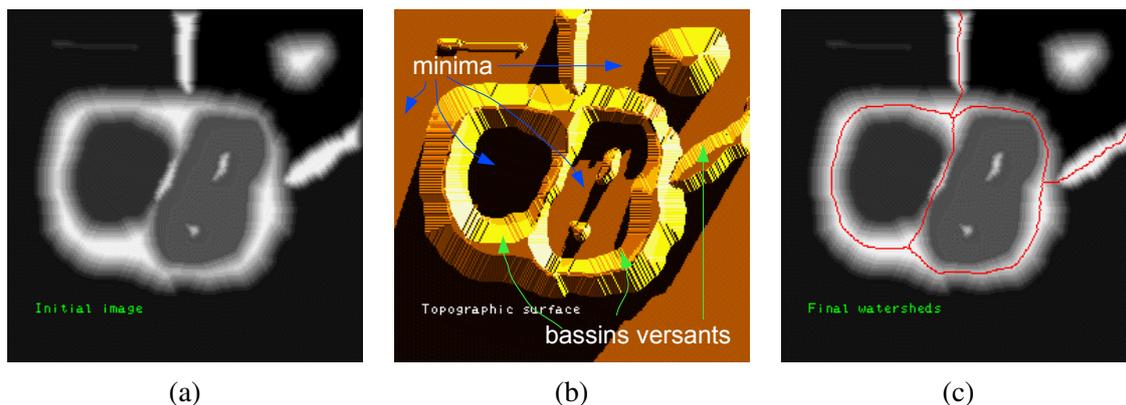


FIG. B.4 – (a) Image de départ, (b) surface topographique avec les bassins versants et les minima, (c) ligne de partage des eaux. (Source : *S. Beucher* - CMM - ENSMP, <http://cmm.ensmp.fr/~beucher/wtshed.html>)

La LPE est un outil très puissant de segmentation. En effet, à condition que l’image d’entrée ait été transformée pour que ses minima marquent les objets pertinents de l’image et que ses lignes de crêtes correspondent aux frontières des objets de l’image, la LPE aura partitionné l’image en régions significatives. Dans la présente partie, nous allons introduire deux définitions équivalentes de la LPE.

B.5.2.1 Définition de la LPE en terme de simulation d’une inondation

Le concept de LPE exprimé en terme d’écoulement d’eau n’est pas approprié pour une définition simple et formelle parce qu’il existe de nombreuses situations pour lesquelles la direction de l’écoulement en un point donné n’est pas déterminé (e.g. les plateaux de pixels ou bien les pixels ayant plus

d'un voisin avec le niveau de gris le plus faible). Cependant, une définition en terme de simulation d'inondations résout ce problème.

Considérons de nouveau l'image à niveaux de gris comme une surface topographique et supposons que des trous aient été percés en chaque minimum régional de la surface. La surface est alors lentement immergée dans l'eau. Partant des minima de plus faible altitude, l'eau va progressivement inonder les bassins versants de l'image. En outre, des barrages sont dressés aux lieux pour lesquels les eaux provenant de différents minima se rencontrent (fig. B.5). A la fin de ce processus d'inondation, chaque minimum est entouré de barrages délimitant son bassin versant. L'ensemble des barrages correspondant à la ligne de partage des eaux, forment une partition de l'image d'entrée en ses différents bassins versants.

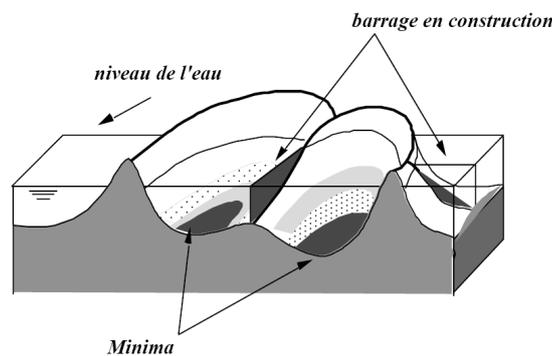


FIG. B.5 – Construction de barrages sur les lieux où les eaux provenant de deux minima différents se rencontrent. (Source : Cours J. Serra [164]).

Formalisons maintenant ce processus d'inondation. La plus petite valeur prise par l'image à niveaux de gris f sur son domaine \mathcal{D}_f est notée h_{min} et la plus grande valeur h_{max} . Le bassin versant associé à un minimum M est noté $BV(M)$. Les points de ce bassin versant qui ont une altitude inférieure ou égale à h sont notés $BV_h(M)$:

$$BV_h(M) = \{p \in BV(M) \mid f(p) \leq h\} = BV(M) \cap T_{t \leq h}(f)$$

avec $T_{t \leq h}(f)$ le seuillage de la fonction f : $[T_{t \leq h}(f)](x) = \begin{cases} 1, & \text{si } (f(x) = t) \leq h ; \\ 0, & \text{sinon .} \end{cases}$

On note X_h le sous-ensemble de tous les bassins versants qui ont une valeur de niveau de gris inférieure ou égale à h :

$$X_h = \cup_i BV_h(M_i)$$

Finalement, l'ensemble des points appartenant aux minima régionaux d'altitude h sont notés $RMIN_h(f)$.

Les bassins versants sont maintenant progressivement construits en simulant le processus d'inondation. Les premiers points de l'image qui sont atteints par l'eau sont les points de plus faible niveau de gris. Ces points appartiennent aux minima régionaux de l'image au niveau h_{min} . Ils sont aussi équivalents à $X_{h_{min}}$:

$$X_{h_{min}} = T_{h_{min}}(f) = RMIN_{h_{min}}(f) .$$

La définition de $X_{h_{min}+1}$ est fondée sur l'analyse du processus d'inondation jusqu'à l'altitude $h_{min} + 1$. Soit l'eau agrandit les régions des bassins versants déjà atteints par l'eau soit l'eau commence à inonder les bassins versants dont les minima ont une altitude égale à $h_{min} + 1$ (fig. B.6). Plus précisément, il y a trois possibilités d'inclusion entre une composante connexe Y de $T_{t \leq h_{min}+1}(f)$ et l'intersection entre Y et $X_{h_{min}}$:

(a) $Y \cap X_{h_{min}} = \emptyset$ (fig. B.6 a). Il s'ensuit que Y est un nouveau minimum régional de f au niveau $h_{min} + 1$ puisque

$$i) \forall p \in Y \quad \begin{cases} p \notin X_{h_{min}} & \Rightarrow f(p) \geq h_{min} + 1; \\ p \in Y & \Rightarrow f(p) \leq h_{min} + 1. \end{cases}$$

ii) $\forall p \in \delta^{(1)}(Y) \setminus Y, f(p) > h_{min} + 1$.

avec $\delta^{(1)}(Y)$ la dilatation géodésique de Y (eq. 48). L'ensemble de tous les minima de niveau $h_{min} + 1$, i.e. $RMIN_{h_{min}+1}$, sera considéré pour définir $X_{h_{min}+1}$.

(b) $Y \cap X_{h_{min}} \neq \emptyset$ et sont connexes (fig. B.6 b). Dans ce cas, Y correspond exactement aux pixels appartenant au bassin versant associé au minimum $Y \cap X_{h_{min}}$ et ayant un niveau de gris inférieur ou égal à $h_{min} + 1$:

$$Y = BV_{h_{min}+1}(Y \cap X_{h_{min}}) = ZI_Y(Y \cap X_{h_{min}}).$$

avec $ZI_Y(Y \cap X_{h_{min}})$ la zone d'influence de $Y \cap X_{h_{min}}$ dans Y (eq. B.5.1, p : 242).

(c) $Y \cap X_{h_{min}} \neq \emptyset$ et sont connexes (fig. B.6 c). Par conséquent, Y contient plus d'un minimum de f de niveau h_{min} . Notons Z_1, Z_2, \dots, Z_k ces minima, et soit Z_i un de ceux-ci. A cette étape, la meilleure approximation possible de $BV_{h_{min}+1}(Z_i)$ correspond à la zone d'influence géodésique de Z_i à l'intérieur de Y :

$$BV_{h_{min}+1}(Z_i) = ZI_Y(Z_i).$$

La dilatation géodésique binaire se définit de la manière qui suit.

Définition 48 (Dilatation géodésique binaire). La dilatation géodésique binaire de taille 1 de l'ensemble $X \subset \mathcal{D}_f$ par rapport à f est notée $\delta^{(n)}(f)$ et est définie par l'ensemble des points de \mathcal{D}_f dont la distance géodésique d_{geo} à X est inférieure ou égale à n :

$$\delta_f^{(n)}(X) = \{p \in \mathcal{D}_f \mid d_{geo}(p, X) \leq n\}$$

Les deux dernières relations d'inclusion correspondent à l'expansion des régions des bassins versants déjà atteints par l'eau. Ces régions en expansion peuvent être définies en terme d'une unique zone d'influence géodésique, i.e. la zone d'influence de $X_{h_{min}}$ dans $T_{t \leq h_{min}+1}$. Il s'ensuit que $X_{h_{min}+1}$ est défini comme l'union de ces zones d'influence géodésiques avec les nouveaux minima régionaux découverts :

$$X_{h_{min}+1} = RMIN_{h_{min}+1}(f) \cup ZI_{T_{t \leq h_{min}+1}(f)}(X_{h_{min}}).$$

Cette relation récursive est valable pour tous les niveaux h . Elle est illustrée sur la figure B.7.

L'ensemble des bassins versants d'une image à niveaux de gris f est égal à l'ensemble $X_{h_{max}}$, i.e. lorsque tous les niveaux ont été inondés :

(i) $X_{h_{min}} = T_{h_{min}}(f)$;

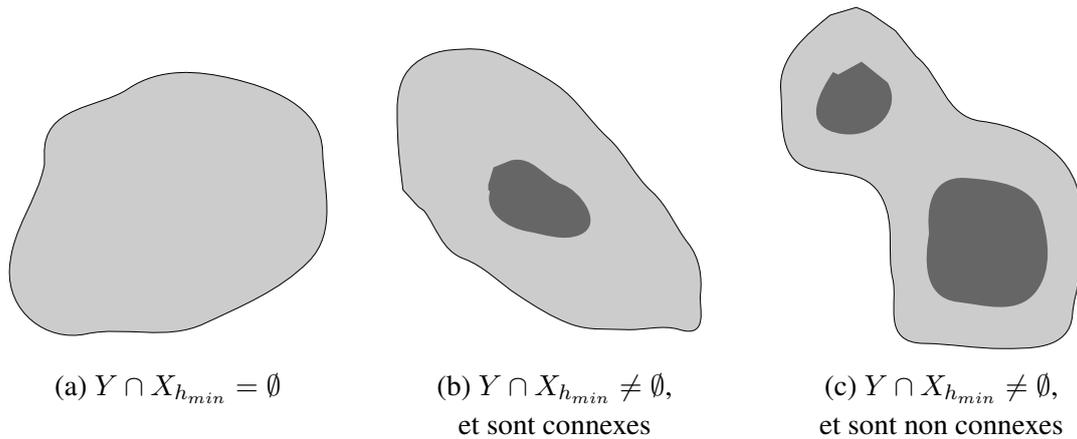


FIG. B.6 – Relations d’inclusions apparaissant durant l’inondation d’une image à niveaux de gris. Les régions sombres indiquent les pixels qui ont déjà été inondés au niveau h_{min} , i.e. les pixels de $X_{h_{min}}$. Les régions en gris clair sont inondées quand l’eau atteint le niveau $h_{min} + 1$. (a) Un nouveau minimum a été découvert au niveau $h_{min} + 1$. (b) La région inondée d’un bassin versant dont le minimum est au niveau $h_{min} + 1$ est étendue. (c) Les régions inondées des bassins versants de deux minima distincts au niveau h_{min} s’étendent et fusionnent.

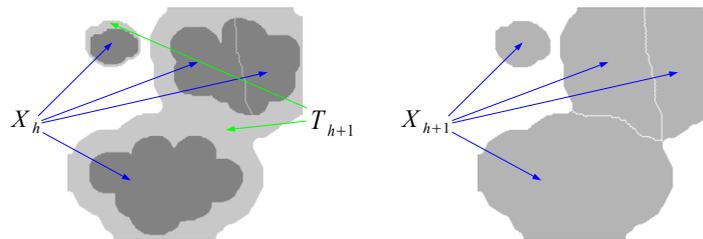


FIG. B.7 – Relation récursive entre X_h et X_{h+1} . (Source : Cours *J. Serra* [164]).

$$(ii) \forall h \in [h_{min}, h_{max} - 1], X_{h+1} = RMIN_{h+1}(f) \cup ZI_{T_{t \leq h_{min}+1}(f)}(X_h).$$

L’image des bassins versants BV d’une image à niveaux de gris est représentée comme une image d’étiquette, chaque région étiquetée correspond au bassin versant d’un minimum régional de l’image d’entrée. La Ligne de Partage des Eaux LPE de f correspond aux frontières des bassins versants de f . La figure B.8 illustre la construction itérative des bassins versants sur une image de synthèse.

B.5.2.2 Définition de la LPE en termes géodésiques

Les bassins versants d’une image à niveaux de gris, peuvent aussi être considérés comme les zones d’influence des minima régionaux de l’image. Dans ce sens ci, la LPE n’est autre que le squelette par zones d’influence géodésique (SKIZ géodésique) de l’image des minima. Cette idée peut être formalisée avec des notions géodésiques.

Les minima régionaux $RMIN$ de l’image d’entrée f sont tout d’abord mis à la valeur minimale

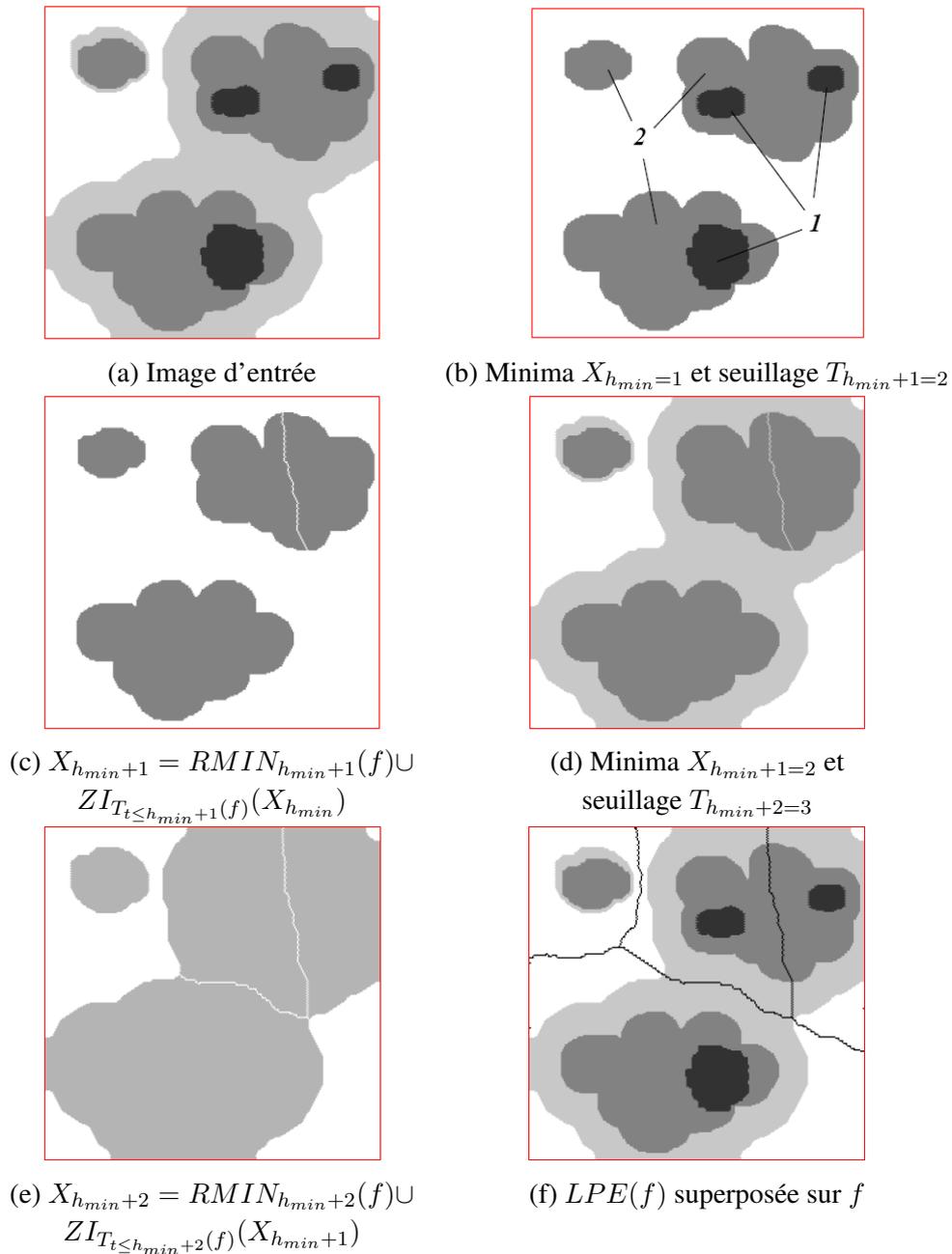


FIG. B.8 – LPE sur une image de synthèse composée de 4 niveaux de gris : illustration de la définition à partir de seuillages successifs et de zones d'influences géodésiques. (Source : Cours *J. Serra* [164]).

h_{min} de l'image. L'image résultante est notée f' :

$$f'(x) = \begin{cases} h_{min}, & \text{si } p \in RMIN(f); \\ f(x), & \text{sinon.} \end{cases} \quad (\text{B.5.2})$$

Les figures B.9 (a) et (b) illustrent cette transformation sur un signal 1D.

Considérons un chemin $\mathcal{P} = (x_1, \dots, x_P)$ défini sur le domaine de définition d'une image à

niveau de gris ou d'une fonction intégrable f . Le temps nécessaire pour parcourir \mathcal{P} est défini comme l'intégrale des valeurs de f le long de \mathcal{P} et noté $\tau_f(\mathcal{P})$:

$$\tau_f(\mathcal{P}) = \int_{\mathcal{P}} f(s) ds.$$

Par conséquent, la fonction temps géodésique correspond à une intégration des niveaux de gris du masque géodésique (qui définit le chemin géodésique), à partir d'un ensemble de référence donné. En appliquant ce procédé sur le gradient morphologique interne $\varrho^-(f')$ de f' et en utilisant les minima régionaux comme ensemble de référence, l'image f' peut être reconstruite (fig. B.9 (d)) :

$$f' = \tau_{\varrho^-(f')} [RMIN(f)],$$

où le gradient interne est défini comme suit.

Définition 49 (Gradient morphologique interne). Le gradient morphologique interne ϱ^- est défini comme la différence entre l'image originale f et son érodée $\varepsilon_B(f)$ sur l'élément structurant B :

$$\varrho_B^-(f) = f - \varepsilon_B(f). \quad (\text{B.5.3})$$

La figure B.9 (c) montre un exemple de LPE en termes géodésiques sur un signal 1D.

Les points pour lesquels les fronts d'ondes, provenant de deux minima distincts, se rencontrent définissent la LPE de l'image originale à niveaux de gris :

$$LPE(f) = SKIZ_{\varrho^-(f')} [RMIN(f')]. \quad (\text{B.5.4})$$

Les bassins versants sont définis en utilisant les zones d'influences géodésiques correspondantes :

$$BV(f) = ZI_{\varrho^-(f')} [RMIN(f')]. \quad (\text{B.5.5})$$

La première étape qui consiste à mettre tous les minima à la valeur minimale de l'image garantit que le processus d'intégration sera retardé proportionnellement à la hauteur de ces minima. Ceci imite le processus d'inondation qui commence par les minima les plus faibles et atteint progressivement tous les pixels de l'image.

Il existe également un lien entre la LPE d'une image et son squelette à niveaux de gris. En effet, la LPE est une sous-ensemble des pixels du squelette à niveaux de gris qui n'appartiennent pas à un minimum régional :

$$LPE(f) \subseteq [RMIN[SK(f)]]^c.$$

B.5.3 La LPE contrôlée par marqueurs

Dans cette partie, on présente tout d'abord la sélection des minima par filtrage, puis l'imposition des minima. Le principe de la segmentation contrôlée par marqueurs. Quelques remarques sont également faites sur l'implémentation de la LPE.

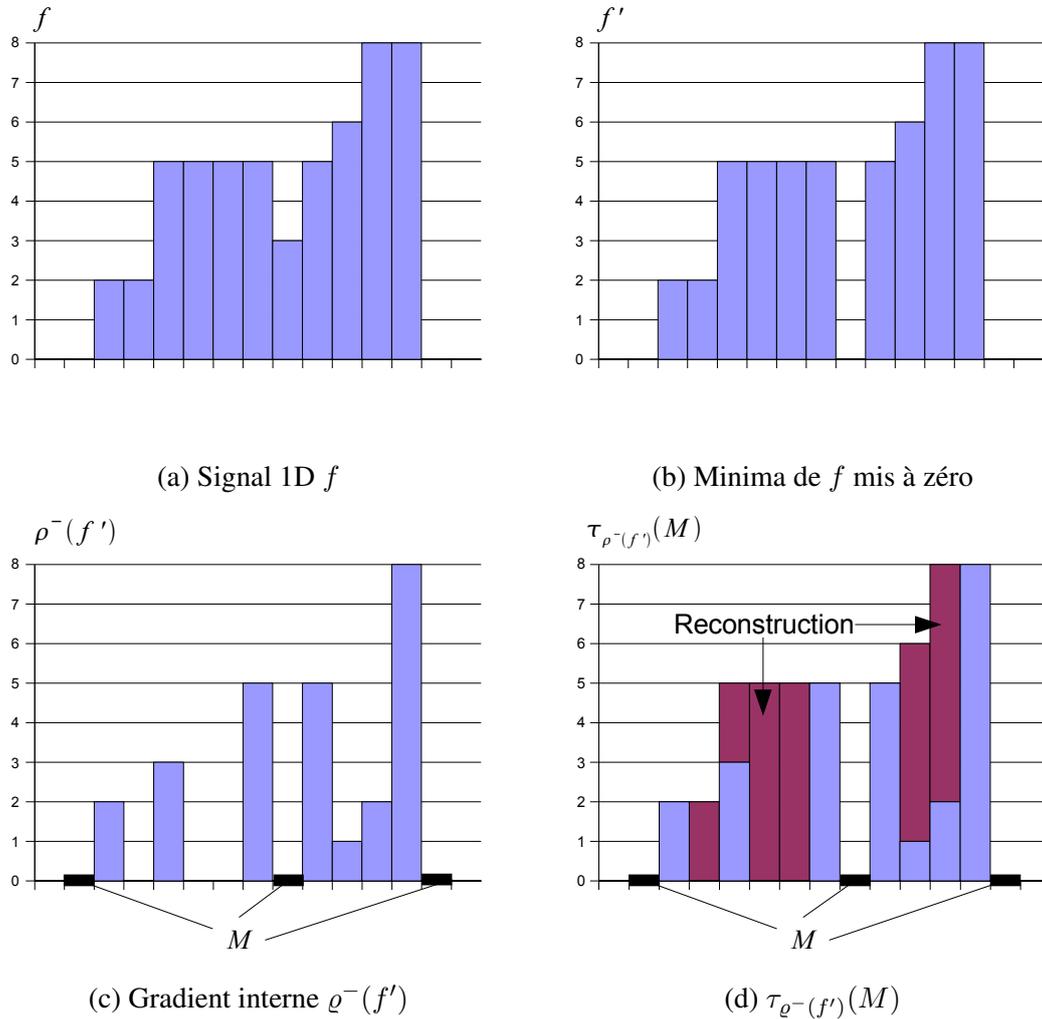


FIG. B.9 – Les bassins versants sont des zones d’influence généralisées. Le gradient interne du signal d’entrée est tout d’abord calculé et utilisé comme masque géodésique à niveau de gris pour calculer les zones d’influence géodésiques des minima M de l’image d’entrée.

B.5.3.1 Sélection des minima par filtrage

En règle générale, les images ont trop de minima, et le calcul sans précaution de la LPE conduit à une sur-segmentation désastreuse. Pour avoir des minima significatifs, on peut procéder par filtrage (fig. B.10) :

- soit « horizontal » par filtres alternés plans avec reconstruction ou non ;
- soit « vertical » par les fermetures de dynamique $h \varphi^{rec}(f, f + h)$. En particulier, pour $h = 1$ on extrait tous les minima.

Rappelons maintenant comment calculer la fermeture par dynamique ou contraste à partir de la reconstruction géodésique.

Définition 50 (Érosion géodésique numérique). Soit $f \in \mathcal{F}(E, \mathcal{T})$ une fonction numérique sur le

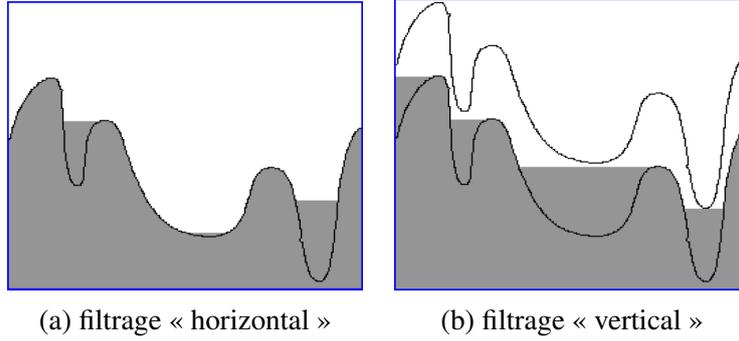


FIG. B.10 – Filtrage des marqueurs. (Source : Cours *J. Serra* [164]).

masque géodésique $g \in \mathcal{F}(E, T)$ ($f \geq g$), l'érosion géodésique de taille n est définie par :

$$\varepsilon_g^{(n)}(f) = \varepsilon_g^{(1)}[\varepsilon_g^{(n-1)}(f)]$$

où l'érosion géodésique unitaire est donnée par :

$$\varepsilon_g^{(1)}(f) = \varepsilon_B(f) \vee g$$

Définition 51 (Fermeture par reconstruction numérique). La fermeture par reconstruction numérique, ou reconstruction par érosion numérique, d'une image f selon une image g ($\mathcal{D}_f = \mathcal{D}_g$ et $f \geq g$) est définie comme l'érosion géodésique de f selon g itérée jusqu'à atteindre l'idempotence. Elle est notée $\varphi^{rec}(f, g)$:

$$\varphi^{rec}(f, g) = \varepsilon_g^{(n)}(f) \quad \text{avec } \varepsilon_g^{(n)}(f) = \varepsilon_g^{(n+1)}(f).$$

B.5.3.2 Imposition des marqueurs comme des minima

Les marqueurs n'ont souvent aucune raison de coïncider avec des minima de l'image f . La notion d'imposition des minima ou « swamping » permet de dissocier complètement les deux aspects.

Étant donnés une fonction $f \geq 0$, et un jeu M de marqueurs, le swamping consiste à trouver l'inf des fonctions :

- qui admettent pour minima les seules composantes connexes de M ;
- qui valent zéro sur M et sont supérieures ou égales à f sur le complémentaire de M : M^c .

L'image marqueur f_m est définie comme ce qui suit :

$$f_m(p) = \begin{cases} 0, & \text{si } p \in M ; \\ \max_x(f(x)), & \text{si } p \notin M . \end{cases}$$

L'imposition des minima sur l'image d'entrée f se fait en deux étapes. Tout d'abord, l'image minimum, pixel à pixel, entre l'image d'entrée et l'image marqueur est calculée : $f \wedge f_m$. Par ce moyen, les minima sont créés aux lieux correspondants aux marqueurs (s'ils n'existent pas déjà) et on est assuré que l'image résultante est inférieure ou égale à l'image marqueur. En outre deux minima distincts à imposer peuvent tomber dans un minimum de f de niveau 0. Il est donc nécessaire de considérer $(f + 1) \wedge f_m$ plutôt que $f \wedge f_m$. La deuxième étape consiste en une fermeture par reconstruction de $(f + 1) \wedge f_m$ depuis l'image marqueur f_m (cf. définition 51, p : 250) :

$$\varphi^{rec}(f_m, (f + 1) \wedge f_m).$$

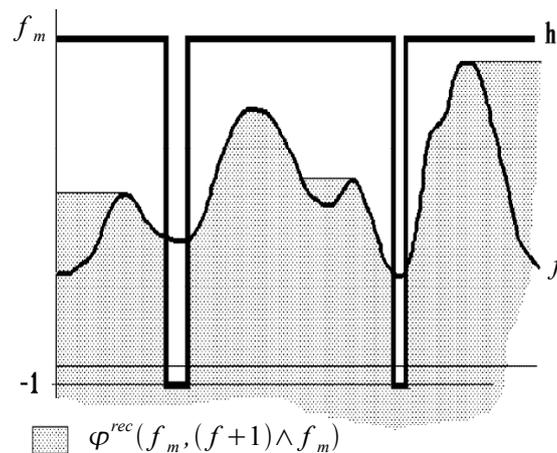


FIG. B.11 – Imposition des minima ou swamping. (Source : Cours J. Serra [164]).

B.5.3.3 Paradigme de la segmentation par LPE

L'idée fondamentale derrière la segmentation contrôlée par marqueurs est de transformer l'image d'entrée de telle manière que la LPE de l'image transformée corresponde aux frontières des objets significatifs. L'image transformée est appelée « fonction de segmentation » ou « fonction à segmenter ». En pratique, un calcul direct de la LPE de la fonction de segmentation fournit une sur-segmentation qui est due à la présence d'une pléthore de minima. Par conséquent, la fonction de segmentation doit être filtrée avant de calculer sa LPE afin de supprimer tous les minima non pertinents. Toute technique de filtrage peut être considérée. Cependant, la technique d'imposition des minima associés à des marqueurs, que nous venons de décrire, est le filtre le plus approprié pour de nombreuses applications. Elle nécessite la détermination d'une fonction marqueur qui marque les objets pertinents de l'image et leur fond. Les marqueurs correspondants sont ensuite utilisés comme un ensemble de minima à imposer à la fonction de segmentation.

Le schéma de cette méthode est présenté dans la figure B.12. Les marqueurs des objets sont extraits de l'image en utilisant des détecteurs de caractéristiques. Le choix des détecteurs de caractéristiques repose sur une connaissance a priori ou des hypothèses sur les propriétés d'un objet de l'image. Les caractéristiques les plus courantes concernent les extrema de l'image, les zones plates (i.e. les composantes connexes de pixels de même niveau de gris), les zones de texture homogène, ... Pour certaines applications, les marqueurs peuvent être définis manuellement. Un marqueur par région est nécessaire puisqu'il existe une correspondance entre le marqueur et le segment associé de la partition finale. Cependant, si la classe de l'objet repéré par chaque marqueur est connue, plusieurs marqueurs de la même classe peuvent être considérés pour chaque objet de l'image. La taille des marqueurs peut aller d'un pixel unique à de grandes composantes connexes de pixels. Lorsque l'on travaille sur des images bruitées, les grands marqueurs fonctionnent mieux que les petits.

La détermination de la fonction de segmentation se fait à partir d'un modèle pour la définition de la frontière de l'objet. Par exemple, si les objets de l'image sont définis comme des régions aux niveaux de gris quasiment constants, un opérateur de gradient morphologique améliorera les frontières des objets. Si les objets de l'image sont des régions à la texture homogène, les opérateurs rehaussant

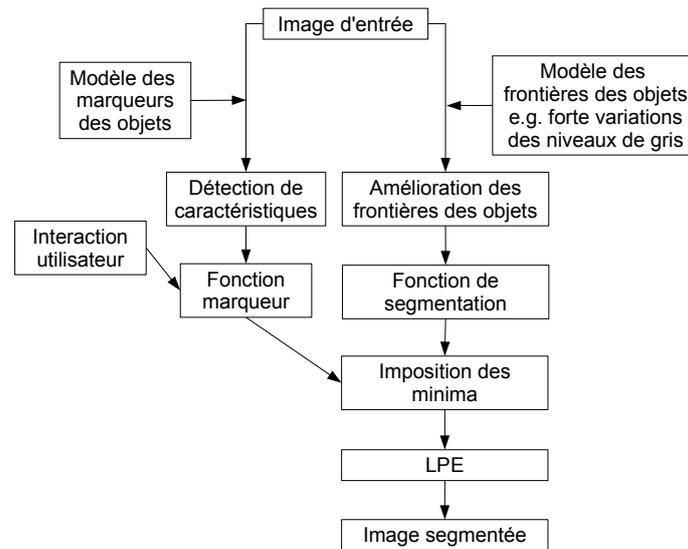


FIG. B.12 – Paradigme morphologique pour la segmentation d’image. La compréhension d’image est faite aux premières étapes du processus. La clef du succès consiste à générer des marqueurs et des fonctions de segmentation pertinents. Alors, le reste de la procédure est non-paramétrique.

les transitions entre deux textures peuvent être utilisés.

Les marqueurs des objets sont ensuite utilisés comme ensemble de marqueurs à imposer à la fonction de segmentation. Finalement, les frontières des objets sont obtenues en calculant les LPE des fonctions de segmentation filtrées.

B.5.3.4 Calcul direct de la LPE avec des marqueurs

Meyer [117] a proposé une extension de l’algorithme de LPE qui combine l’imposition des minima et la calcul de la LPE dans un seul algorithme. Ceci est effectué grâce à des files d’attente hiérarchiques. Pour plus de détail se reporter à [29].

B.5.4 LPE hiérarchique : algorithme des cascades

Meyer [118] a proposé des méthodes utilisant des graphes qui ont abouti à des algorithmes de segmentation hiérarchiques multi-échelles très rapides et efficaces.

Nous présentons ici un autre algorithme de LPE hiérarchique non-paramétrique : l’algorithme des cascades de LPE, proposé par *Beucher* [26]. Si la LPE est construite sur l’image gradient, cette méthode repose sur l’hypothèse que les contours les plus forts correspondent aux objets à segmenter. Concrètement à chaque étape de cette segmentation hiérarchique, on cherche à supprimer les contours les plus faibles entourés de contours de valeurs plus élevées, en fusionnant les bassins versants des contours supprimés. Un tel processus peut être itéré de façon à supprimer, lors d’une nouvelle passe, l’ensemble des contours devenus à leur tour les plus faibles. *Demarty* [49] a proposé un algorithme pour cette procédure de segmentation hiérarchique fondé sur un graphe de voisinage entre régions (sans avoir besoin de l’image mosaïque).

L'implémentation la plus simple de l'algorithme de cascades se fonde sur la reconstruction de la fonction gradient de l'image mosaïque avec sa LPE. Soyons plus précis. L'image à niveaux de gris à segmenter et son gradient sont respectivement f et g , avec $f, g : E \rightarrow T = \{t_{min}, t_{min} + 1, \dots, t_{max}\}$. Soit $LPE(g)$ la LPE du gradient g . L'image mosaïque m est obtenue comme la valeur moyenne de la fonction f dans chaque bassin versant de $LPE(g)$. Puis, on calcule le gradient de la mosaïque, qu'on notera g_1 . L'algorithme de cascades est comme suit :

Algorithme 15 Algorithme des cascades

- 1: Calculer la LPE de la fonction $g_1 : LPE(g_1)$.
- 2: Définir une nouvelle fonction numérique, h :

$$h(x) = \begin{cases} g_1(x) & \text{si } x \in LPE(g_1) \\ t_{max} & \text{si } x \in LPE^c(g_1) \end{cases}$$

On a $f \geq g_1$.

- 3: Reconstruire g_1 à partir de h par érosion géodésique : $g_1^* = \varepsilon_{g_1}^{rec}(h)$. En fait la reconstruction va remplir chaque bassin versant avec un plateau de hauteur égal à la profondeur minimale des lignes de la LPE qui entourent le bassin versant. Par conséquent, s'il y a un bassin versant adjacent avec une hauteur inférieure à celle définie, le plateau généré ne sera pas un minimum.
 - 4: Calculer la LPE de la fonction g_1^* : les bassins versants sont associés aux minima significatifs.
 - 5: Construire une nouvelle mosaïque à partir de cette dernière LPE, m_1 , sur laquelle on calcule le gradient g_2 et ensuite on peut itérer l'algorithme pour obtenir la hiérarchie.
-

Bibliographie

- [1] *Morphological partitioning of multispectral images*, Journal of Electronic Imaging **5** (1996), no. 3, 252–263.
- [2] J. Angulo, *Morphologie mathématique et indexation d'images couleur: application à la microscopie en biomédecine*, Thèse de doctorat en morphologie mathématique, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 2003.
- [3] ———, *Morphological colour operators in totally ordered lattices based on distances : Application to image filtering, enhancement and analysis*, Computer Vision and Image Understanding **107** (2007), 56–73.
- [4] J. Angulo and D. Jeulin, *Stochastic watershed segmentation*, Mathematical Morphology and its Applications to Signal and Image Processing, Proc. of the 8th Int. Symposium on Mathematical Morphology (G. Banon et al., ed.), vol. 1, Instituto Nacional de Pesquisas Espaciais (INPE), 2007, pp. 265–276.
- [5] J. Angulo and J. Serra, *Modelling and segmentation of colour images in polar representations*, Image Vision and Computing **25** (2007), no. 4, 475–495.
- [6] E. Aptoula and S. Lefevre, *A comparative study on multivariate mathematical morphology*, Pattern Recognition (2007), 2914–2929.
- [7] P. Arbelaez, *Une approche métrique pour la segmentation d'images*, Thèse de doctorat, CEREMADE, Université Paris-Dauphine, Paris, France, 2005.
- [8] P. Arbelaez and L. Cohen, *Generalized voronoi tessellations for vector-valued image segmentation*, Proceedings 2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision (VLISM'03) (Nice, France), September 2003.
- [9] R. Archibald and G. Fann, *Feature selection and classification of hyperspectral images with support vector machines*, IEEE Geoscience and remote sensing letters **4** (2007), no. 4, 674–677.
- [10] E. Arzuaga-Cruz, L.O. Jiménez-Rodríguez, and M. Vélez-Reyes, *Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis*, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX (S.S. Shen and P.E. Lewis, eds.), Proceedings of the SPIE, vol. 5093, 2003, pp. 462–473.
- [11] J. Astola, P. Haavisto, and Y. Nuevo, *Vector median filters*, Proceedings of the IEEE, vol. 78, 1990, pp. 678–689.
- [12] H. Attias, *Independent factor analysis*, Neural Computation **11** (1999), no. 4, 803–851.
- [13] D. Balvay, F. Frouin, G. Calmon, B. Bessoud, E. Kahn, N. Siauve, O. Clément, and C.A. Cuenod, *New criteria for assessing fit quality in dynamic contrast-enhanced T_1 -weighted MRI for perfusion and permeability imaging*, Magnetic Resonance in Medicine **54** (2005), 868–877.
- [14] E. Bayro Corrochano, *Handbook of geometric computing*, Springer, 2005.
- [15] C.F. Beckmann and Smith S.M., *Tensorial extensions of independent component analysis for multisubject fmri analysis*, Neuroimage **25** (2005), 294–311.

- [16] R. Beichel, H. Bischof, F. Leberl, and M. Sonka, *Information processing in medical imaging*, Lecture Notes in Computer Science, ch. Robust Active Appearance Model Matching, pp. 114–125, Springer, Berlin, 2005.
- [17] M. Belkin and P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*, Advances in Neural Information Processing Systems, vol. 14, The MIT Press, Cambridge, MA, USA., 2002, p. 585–591.
- [18] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson, *Classification of hyperspectral data from urban areas based on extended morphological profiles*, IEEE Transactions on Geoscience and Remote Sensing **42** (2005), 480–491.
- [19] J.A. Benediktsson, M. Pesaresi, and K. Amason, *Classification and feature extraction for remote sensing images from urban areas based on morphological transformations*, IEEE Transactions on Geoscience and Remote Sensing **41** (2003), no. 9, 1940–1949.
- [20] J.P. Benzécri, *Sur l'analyse factorielle des proximités*, Publications Institut de Statistique, vol. 13, Université de Paris, 1964, pp. 235–282.
- [21] ———, *L'analyse des données, l'analyse des correspondances*, vol. 2, Dunod, Paris, 1973.
- [22] ———, *Pratique de l'analyse des données. analyse des correspondances & classification. exposé élémentaire.*, 2 ed., vol. 1, Dunod, Paris, 1980.
- [23] M. Berman, *The statistical properties of three noise removal procedures for multichannel remotely sensed data*, Consulting Rep. NSW/85/31/MB9, CSIRO Div. of Mathematics and Statistics, P.O. Box 218, Lindfield, NSW 2070, Australia, 1985.
- [24] L. Bertelli, B. Sumengen, and B.S. Manjunath, *Redundancy in all pairs fast marching method*, IEEE International Conference on Image Processing ICIP'06, 2006, pp. 3033–3036.
- [25] S. Beucher, *Segmentation d'images et morphologie mathématique*, Thèse de doctorat, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 1990.
- [26] ———, *Mathematical morphology and its applications to image and signal processing*, ch. Watershed, hierarchical segmentation and waterfall algorithm, pp. 69–76, Kluwer, 1994.
- [27] ———, *Algorithmes sans biais de ligne de partage des eaux*, Note interne, 34 p., Ecole des Mines de Paris, Centre de Morphologie Mathématique, Février 2002.
- [28] S. Beucher and C. Lantuéjoul, *Use of watersheds in contour detection*, Proc. Int. Workshop on image processing, real-time edge and motion detection-estimation, 1979, pp. 17–21.
- [29] S. Beucher and F. Meyer, *Mathematical morphology in image processing*, ch. The Morphological Approach to Segmentation : The Watershed Transformation, pp. 433–481, Marcel-Dekker, New York, 1992.
- [30] G. Borgefors, *Distance transformations in digital images*, Comput. Vision Graph. Image Process. **34** (1986), no. 3, 344–371.
- [31] Y. Boykov and M. P. Jolly, *Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images*, Proceedings of International Conference on Computer Vision (Vancouver, Canada), vol. I, 2001, pp. 105–112.
- [32] U.M. Braga Neto and J. Goutsias, *Connectivity on complete lattices : New results*, Computer Vision and Image Understanding **85** (2002), 22–53.
- [33] U.M. Braga Neto and J. A Goutsias, *Theoretical tour of connectivity in image processing and analysis*, Journal of Mathematical Imaging and Vision **19** (2003), 5–31.
- [34] M. W. Browne, *A comparison of factor analytic techniques*, Psychometrika **33** (1968), 267–334.

- [35] D. Brunner and P. Soille, *Iterative area seeded region growing for multichannel image simplification*, Proceedings ISMM'05, Mathematical Morphology and its Applications to Image and Signal Processing (C. Ronse, L. Najman, and E. Decencière, eds.), 2005, pp. 397–406.
- [36] C. Burges, *A tutorial on support vector machines for pattern recognition*, pp. 1–43, Kluwer Academic, 1998.
- [37] J.-F. Cardoso, *Infomax and maximum likelihood for blind source separation*, Signal Processing Letters, IEEE **4** (1997), no. 4, 112–114.
- [38] R. B. Cattell and J. A. Jaspers, *A general plasmode for factor analytic exercises and research*, Multivariate Behavioral Research Monographs (1967).
- [39] J. Chanussot, *Approches vectorielles ou marginales pour le traitement d'images multi-composantes*, Thèse de doctorat, Université de Savoie, 1998.
- [40] K. Cho and P. Meer, *Image segmentation from consensus information*, Computer Vision and Image Understanding **68** (1996), no. 1, 72–89.
- [41] E. Christophe, *Compression des images hyperspectrales et son impact sur la qualité des données*, Thèse de doctorat, Ecole nationale supérieure de l'aéronautique et de l'espace, 2006.
- [42] L.D. Cohen and R. Kimmel, *Fast marching the global minimum of active contours*, IEEE Proceedings International Conference on Image Processing, vol. 1, 1996, pp. 473–476.
- [43] D. Comaniciu and P. Meer, *Mean shift : a robust approach toward feature space analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 5, 603–619.
- [44] M. Coster and J.L. Chermant, *Précis d'analyse d'images*, Les Presses du CNRS, Paris, 1985.
- [45] T. Cox and M. Cox, *Multidimensional scaling*, Chapman & Hall, London, UK, 1994.
- [46] J. Crespo, *Morphological connected filters and intra-region smoothing for image segmentation*, Phd in electrical engineering, Georgia Institute of Technology, 1993, 4511.
- [47] J. Crespo, R.W. Schafer, J. Serra, C. Gratin, and F. Meyer, *The flat zone approach : A general low-level region merging segmentation method*, Signal Processing **62** (1997), no. 1, 37–60.
- [48] O. Debeir, I. Adanja, N. Warzée, P. Van Ham, and C. Decaestecker, *Phase contrast image segmentation by weak watershed transform assembly*, Proceedings IEEE International Symposium on Biomedical Imaging ISBI 2008, 2008, pp. 724–727.
- [49] C. H. Demarty, *Segmentation et structuration d'un document vidéo pour la caractérisation et l'indexation de son contenu sémantique : Application aux journaux télévisés*, Thèse de doctorat, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 2000.
- [50] D. DeMers and G. Cottrell, *Advances in neural information processing systems*, vol. 5, ch. Non-linear dimensionality reduction, p. 580–587, Morgan Kaufmann, San Mateo, CA, USA, 1993.
- [51] E. Diday, *Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques*, Revue de statistique appliquée **19** (1979), no. 2, 19–33.
- [52] D.L. Donoho and C. Grimes, *Hessian eigenmaps : New locally linear embedding techniques for high-dimensional data*, Proceedings of the National Academy of Sciences, vol. 102, 2005, p. 7426–7431.
- [53] Y.Z. Du, C.I. Chang, H.A. Ren, F. D'Amico, and J.O. Jensen, *A new hyperspectral discrimination measure for spectral similarity*, Algorithms and Technologies for Multispectral, hyperpsctral and ultraspectral imagery IX (Shen S.S. and Lewis P.E., eds.), Proceedings of SPIE, vol. 5093, 2003, pp. 430–439.
- [54] R.O. Duda and P.E. Hart, *Pattern classification and scene analysis*, Wiley, New York, 1973.
- [55] M. Fauvel, J. Chanussot, J.A. Benediktsson, and J.R. Sveinsson, *Spectral and spatial classification of hyperspectral data using svms and morphological profiles*, IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007, 2007, pp. 4834–4837.

- [56] F. Flitti, *Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de markov*, Thèse de doctorat, Université Louis Pasteur, Strasbourg I, 2005.
- [57] G. Flouzat, O. Amram, and S. Cherchali, *Spatial and spectral segmentation of satellite remote sensing imagery using processing graphs by mathematical morphology*, Proceedings IEEE Geoscience and Remote Sensing Symposium, IGARSS '98, vol. 4, 1998, pp. 1769–1771.
- [58] J.H. Friedman and J.W. Tukey, *A projection pursuit algorithm for exploratory data analysis*, IEEE Transactions on Computers **23** (1974), 881–890.
- [59] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, 1972.
- [60] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (1984), 721–741.
- [61] F. Genova, M. Petremand, F. Bonnarel, M. Louys, and C. Collet, *Reduction and segmentation of hyperspectral data cubes*, Astronomical Data Analysis Software & Systems XVI (Tucson, Arizona, USA), october 2006.
- [62] D.D. Gomez, L.H. Clemmensen, B.K. Ersbøll, and J.M. Carstensen, *Precise acquisition and unsupervised segmentation of multi-spectral images*, Computer Vision and Image Understanding **106** (2007), 183–193.
- [63] C. Gomila, *Mise en correspondance de partitions en vue du suivi d'objets*, Thèse de doctorat, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 2001.
- [64] J. Goutsias, H.J.A.M. Heijmans, and K. Sivakumar, *Morphological operators for image sequences*, Computer Vision and Image Understanding **62** (1995), no. 3, 326–346.
- [65] H. Grahn and P. Geladi (Editors), *Techniques and applications of hyperspectral image analysis*, Wiley, 2007.
- [66] V. Grau, A.U.J. Mewes, M. Alcañiz, R. Kikinis, and S.K. Warfield, *Improved watershed transform for medical image segmentation using prior information*, IEEE Transactions on Medical Imaging **23** (2004), 447–458.
- [67] A. Green, M. Berman, P. Switzer, and M.D. Craig, *A transformation for ordering multispectral data in terms of image quality with implications for noise removal*, IEEE Transactions on Geoscience and Remote Sensing **26** (1988), no. 1, 65–74.
- [68] M. J. Greenacre, *Theory and applications of correspondence analysis*, New York : Academic Press, 1984.
- [69] M. Grimaud, *La géodésie numérique en morphologie mathématique. application à la detection automatique de microcalcifications en mammographie numérique*, Thèse de doctorat, Ecole des Mines, Centre de Morphologie Mathématique, 1991.
- [70] L. Görlitz, B.H. Menze, M.A. Weber, B.M. Kelm, and F.A. Hamprecht, *Pattern recognition*, Lecture Notes in Computer Science, vol. 4713/2007, ch. Semi-supervised Tumor Detection in Magnetic Resonance Spectroscopic Images Using Discriminative Random Fields, pp. 224–233, Springer Berlin, 2007.
- [71] J. Gualtieri and R. Crompt, *Support vector machines for hyperspectral remote sensing classification*, Proceedings of the SPIE, vol. 3584, 1999, pp. 221–232.
- [72] Kaplan H., *Practical applications of infrared thermal sensing and imaging equipment*, 2d ed., vol. TT34, ch. Tutorial texts in optical engineering, SPIE Press Book, 1999.
- [73] A. R. Hakstian, W. D. Rogers, and R. B. Cattell, *The behavior of numbers of factors rules with simulated data*, Multivariate Behavioral Research **17** (1982), 193–219.
- [74] A. Hanbury and J. Serra, *Morphological operators on the unit circle*, IEEE Transactions on Image Processing **10** (2001), no. 12, 1842–1850, EX N-40/01/MM.

- [75] J. A. Hartigan and M. A. Wong, *A k-means clustering algorithm*, Applied Statistics **28** (1979), 100–108.
- [76] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning : Data mining, inference, and prediction*, Springer Series in Statistics, Springer, 2003.
- [77] H.J.A.M. Heijmans, *Morphological image operators*, Academic Press, Boston, 1994.
- [78] G.E. Hinton and R.R. Salakhutdinov, *Reducing the dimensionality of data with neural networks*, Science **313** (2006), no. 5786, 504–507.
- [79] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology **24** (1933), no. 6, 417–441.
- [80] ———, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology **24** (1933), no. 7, 498–520.
- [81] G. Hughes, *On the mean accuracy of statistical pattern recognizers*, Information Theory, IEEE Transactions on **14** (1968), 55–63.
- [82] A. Hyvärinen and E. Oja, *Independent component analysis : Algorithms and applications*, Neural Networks **13** (2000), 411–430.
- [83] A. Ifarraguerri and Chein-I Chang, *Unsupervised hyperspectral image analysis with projection pursuit*, IEEE Transactions on Geoscience and Remote Sensing **38** (2000), no. 6, 2529 – 2538.
- [84] D. Jeulin, *Modèles morphologiques de structures aléatoires et de changement d'échelle*, Thèse de doctorat ès sciences physiques, Université de Caen, 1991, 2236.
- [85] L.O. Jimenez and D.A. Landgrebe, *Hyperspectral data analysis and supervised feature reduction via projection pursuit*, IEEE Transactions on Geoscience and Remote Sensing **37** (1999), no. 6, 2653–2667.
- [86] L. Journaux, *Analyse multispectrale d'images satellitaires et analyse multi-tableaux : application à la répartition des populations d'oiseaux et à la structure du paysage*, Thèse de doctorat, Université de Bourgogne, 2006.
- [87] H. F. Kaiser, *The application of electronic computers to factor analysis*, Educational and Psychological Measurement **20** (1960), 141–151.
- [88] ———, *The scree test for the number of factors*, Multivariate Behavioral Research **1** (1966), 245–276.
- [89] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes : Active contour models*, International Journal of Computer Vision **1** (1988), no. 321–331.
- [90] L. Kaufman and P.J. Rousseeuw, *Finding groups in data. an introduction to cluster analysis*, John Wiley and Sons, 1990.
- [91] B. M. Kelm, *Evaluation of vector-valued clinical image data using probabilistic graphical models : Quantification and pattern recognition*, Ph.D. thesis, University of Heidelberg, 2007.
- [92] N. Keshava, *Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries*, IEEE Transactions on Geoscience and Remote Sensing **42** (2004), no. 7, 1552 – 1565.
- [93] M. Köppen and K. Franke, *A color morphology based on pareto-dominance relation and hypervolume measure*, Proceedings IS&T's Fourth European Conference on Color in Graphics Imaging and Vision CGIV 2008 (Terrassa - Barcelona, Spain), June 9-13 2008, pp. 423–426.
- [94] J.B. Kruskal, *Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis*, Psychometrika **29** (1964), 1–27.
- [95] S. Lafon and A.B. Lee, *Diffusion maps and coarse-graining : A unified framework for dimensionality reduction, graph partitioning, and data set parameterization*, IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006), no. 9, 1393–1403.

- [96] D. Landgrebe, *Hyperspectral image data analysis*, IEEE Signal Processing Magazine **19** (2002), no. 1, 17–28.
- [97] ———, *Signal theory methods in multispectral remote sensing*, John Wiley and Sons, 2003.
- [98] C. Lantuéjoul and S. Beucher, *Geodesic distance and image analysis*, 5th ICS (Salzburg, Autriche, Septembre 1979), *Mikroskopie*, vol. 37, 1980, pp. 138–142.
- [99] ———, *On the use of the geodesic metric in image analysis*, Journal of microscopy **121** (1981), no. 1, 39–49.
- [100] C. Lantuéjoul and H. Digabel, *Iterative algorithms*, Proc. of 2nd European Symposium on Quantitative Analysis of Microstructures in Material Sciences, Biology and Medecine, 1977, pp. 39–49.
- [101] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.J. Lefèvre-Fonollosa, *Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated pleiades-hr images*, Information Fusion **6** (2005), 193–212.
- [102] J.B. Lee, A.S. Woodyatt, and M. Berman, *Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform*, IEEE Transactions on Geoscience and Remote Sensing **28** (1990), no. 3, 295–304.
- [103] A.C. Legrand, F. Meriaudeau, and P. Gorria, *Active infrared non-destructive testing for glue occlusion detection within plastic lids*, NDT & E International **35** (2002), no. 3, 177–187.
- [104] M. Lennon, *Méthodes d'analyse d'images hyperspectrales. exploitation du capteur aéroporté casi pour des applications de cartographie agro-environnementale en bretagne*, Thèse de doctorat, Université de Rennes - Rennes I, 2002.
- [105] M. Lennon, G. Mercier, and L. Hubert-Moy, *Classification of hyperspectral images with nonlinear filtering and support vector machines*, IEEE International Geoscience and Remote Sensing Symposium. IGARSS '02, vol. 3, 2002, pp. 1670–1672.
- [106] M. Lennon, G. Mercier, M.C. Mouchot, and L. Hubert-Moy, *Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images*, Geoscience and Remote Sensing Symposium IGARSS '01, vol. 6, IEEE 2001 International, 2001, p. 2893–2895.
- [107] P. Lewicki and T. Hill, *Statitics methods and applications*, Statsoft Inc., 2008.
- [108] J. Li and Y. Li, *Multivariate mathematical morphology based on principal component analysis : initial results in building extraction*, Proceedings of the 20th ISPRS (Istanbul, Turkey), vol. 35, 2004, p. 1168–1173.
- [109] P. Li and X. Xiao, *Evaluation of multiscale morphological segmentation of multispectral imagery for land cover classification*, Proceedings IEEE Geoscience and Remote Sensing Symposium, IGARSS '04, vol. 4, 2004, pp. 2676–2679.
- [110] R. L. Linn, *A monte carlo approach to the number of factors problem*, Psychometrika **33** (1968), 37–71.
- [111] A.R.S. Marcal and L. Castro, *Hierarchical clustering of multispectral images using combined spectral and spatial criteria*, IEEE Geoscience and Remote Sensing Letters **2** (2005), no. 1, 59–63.
- [112] B. Marcotegui, *Segmentation de séquences d'images en vue du codage*, Thèse de doctorat, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 1996.
- [113] G. Matheron, *Elements pour une théorie des milieux poreux*, Masson, Paris, 1967.
- [114] ———, *La théorie des variables régionalisées et ses applications*, Les cahiers du Centre de Morphologie Mathématique de Fontainebleau 5, Centre de Morphologie Mathématique de Fontainebleau, Ecole des Mines de Paris, France, 1970.
- [115] ———, *Random sets and integral integral geometry*, Wiley, New York., 1975.

- [116] Georges Matheron, *Les nivellements*, note manuscrite, CG/CMM Fontainebleau, ENSMP, Février 1997, 6004.
- [117] F. Meyer, *Un algorithme optimal de ligne de partage des eaux*, Proceedings 8ème Congrès AFCET (Lyon-Villeurbanne, France), 1991, pp. 447–457.
- [118] ———, *Mathematical morphology and its applications to image and signal processing*, ch. Minimum spanning forests for morphological segmentation, pp. 77–84, Kluwer, 1994.
- [119] ———, *From connected operators to levelings*, Mathematical Morphology and its Applications to Image and Signal Processing, Proc. ISMM'98 (Amsterdam, June 1998) (Henk J.A.M. Heijmans and Jos B.T.M. Roerdink, eds.), Kluwer, 1998, pp. 191–198.
- [120] ———, *The levelings*, Mathematical Morphology and its Applications to Image and Signal Processing, Proc. ISMM'98 (Amsterdam, June 1998) (Henk J.A.M. Heijmans and Jos B.T.M. Roerdink, eds.), Kluwer, 1998, pp. 199–206.
- [121] ———, *Morphological multiscale and interactive segmentation*, Proceedings of IEEE-EURASIP Workshop on Non-Linear Signal and Image Processing (NSIP'99) (Antalya, Turquie), 1999.
- [122] ———, *Mathematical morphology and its applications to image processing, proc. ismm'00*, ch. Flooding and segmentation, pp. 189–198, Kluwer, 2000.
- [123] F. Meyer, *An overview of morphological segmentation*, International Journal of Pattern Recognition and Artificial Intelligence **15** (2001), no. 7, 1089–1118.
- [124] F. Meyer, *Levelings, image simplification filters for segmentation*, Journal of Mathematical Imaging and Vision **20** (2004), 59–72.
- [125] ———, *Grey-weighted, ultrametric and lexicographic distances*, Proceedings ISMM'05, Mathematical Morphology and its Applications to Image and Signal Processing (C. Ronse, L. Najman, and E. Decencière, eds.), 2005, pp. 289–298.
- [126] F. Meyer and S. Beucher, *Morphological segmentation*, Journal of Visual Communication and Image Representation **1** (1990), no. 1, 21–45.
- [127] F. Meyer and P. Maragos, *Nonlinear scale-space representation with morphological levelings*, Journal of Visual Communication and Image Representation **11** (2000), no. 2, 245–265.
- [128] Fernand Meyer, *Alpha-beta flat zones, levelings and flattenings*, Mathematical Morphology VI, Proc. ISMM 2002 (Sydney, Australia, April 3-5, 2002) (Hugues Talbot and Richard Beare, eds.), CSIRO, 2002, pp. 47–68.
- [129] S.C. Mitchell, B.P.F. Lelieveldt, R.J. van der Geest, H.G. Bosch, J.H.C. Reiver, and M. Sonka, *Multiscale hybrid active appearance model matching : segmentation of left and right ventricles in cardiac mr images*, IEEE Transactions on Medical Imaging **20** (2001), no. 5, 415–423.
- [130] A. Mohan, G. Sapiro, and E. Bosch, *Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images*, IEEE Geoscience and Remote Sensing Letters **4** (2007), no. 2, 206–210.
- [131] J. Moreau, P.A. Doudin, and P. Cazes, *L'analyse des correspondances et les techniques connexes. approches nouvelles pour l'analyse statistique des données*, Springer, 2000.
- [132] P. Mouttou and H. Mariotte, *Analyse des données (cours)*, S1625, Ecole des Mines de Paris, 60 Boulevard Saint-Michel, 75 272 Paris cedex 06, 2001.
- [133] B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis, *Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems*, Applied and Computational Harmonic Analysis **21** (2006), 113–127.
- [134] G. Noyel, J. Angulo, and D. Jeulin, *Morphological segmentation of hyperspectral images*, Image Analysis and Stereology **26** (2007), 101–109.

- [135] ———, *Morphological segmentation of hyperspectral images*, Proceedings of 12th International Congress for Stereology, ICS XII (Saint Etienne, France), August 30 to September 7 2007.
- [136] ———, *On distances, paths and connections for hyperspectral image segmentation*, Mathematical Morphology and its Applications to Signal and Image Processing. Proc. of the 8th Int. Symposium on Mathematical Morphology (G. Banon et al., ed.), vol. 1, Instituto Nacional de Pesquisas Espaciais (INPE), 2007, pp. 399–410.
- [137] ———, *Random germs and stochastic watershed for unsupervised multispectral image segmentation*, KES 2007/ WIRN 2007 (B. Apolloni et al., ed.), LNAI 4694, vol. III, Knowledge-Based Intelligent Information and Engineering Systems, Springer-Verlag, 2007, pp. 17–24.
- [138] ———, *Classification-driven stochastic watershed. application to multispectral segmentation*, Proceedings IS&T's Fourth European Conference on Color in Graphics Imaging and Vision CGIV 2008 (Terrassa - Barcelona, Spain), June 9-13 2008, pp. 471–476.
- [139] ———, *Filtering, segmentation and region classification by hyperspectral mathematical morphology of dce-mri series for angiogenesis imaging*, Proceedings IEEE International Symposium on Biomedical Imaging ISBI 2008, 2008, pp. 1517–1520.
- [140] J.P. Orfeuil, *Une méthode de filtrage des données*, Rapport N-314, CG, Ecole des Mines de Paris, 1973.
- [141] R.L. Ornberg, B.M. Woerner, and D.A. Edwards, *Analysis of stained objects in histological sections by spectral imaging and differential absorption*, Journal of Histochemistry and Cytochemistry **47** (1999), 1307–1314.
- [142] Scheunders P., *Multivalued image segmentation based on first fundamental form*, Proceedings IEEE 11th International Conference on Image Analysis and Processing, 2001, pp. 185–190.
- [143] P. Paclík, R.P.W. Duin, G.M.P. Van Kempen, and Kohlus R., *Segmentation of multi-spectral images using the combined classifier approach*, Image and Vision Computing **21** (2003), 473–482.
- [144] Matthieu Petremand, *Détection des galaxies à faible brillance de surface, segmentation hyperspectrale dans le cadre de l'observatoire virtuel*, Thèse de doctorat, Université Louis Pasteur - Strasbourg I, 2006.
- [145] I. Pitas and P. Tsakalides, *Multivariate ordering in color image filtering*, IEEE Transactions on Circuits and Systems for Video Technology **1** (1991), no. 3, 247–259.
- [146] A. Plaza, J.A. Benediktsson, J.W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P.E. Gamba, A. Gaultieri, and J.C. Tilton, *Advanced processing of hyperspectral images*, IEEE International Conference on Geoscience and Remote Sensing Symposium IGARSS 2006 (2006), 1974–1978.
- [147] A. Plaza, P. Martinez, R.M. Perez, and J. Plaza, *Spatial/spectral endmember extraction by multidimensional morphological operations*, IEEE Transactions on Geoscience and Remote Sensing **40** (2002), no. 9, 2025–2041.
- [148] G. Polder, *Spectral imaging for measuring biochemicals in plant material*, Ph.D. thesis, Technische Universiteit Delft, 2004.
- [149] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, *La poursuite de projection pour la classification d'image hyperspectrale texturée*, Rapport de Recherche 4152, INRIA, 2001.
- [150] N. Renard, S. Bourennane, and J. Blanc-Talon, *Denosing and dimensionality reduction using multilinear tools for hyperspectral images*, IEEE Geoscience and Remote Sensing Letters **5** (2008), no. 2, 138 – 142.
- [151] C. Ronse, *Set theoretical algebraic approaches to connectivity in continuous or digital spaces*, Journal of Mathematical Imaging and Vision **8** (1998), 41–58.
- [152] C. Ronse and J. Serra, *Geodesy and connectivity in lattices*, Fundamenta Informaticae **46** (2001), 349–395.

- [153] S.T. Roweis and L.K. Saul, *Nonlinear dimensionality reduction by locally linear embedding*, *Science* **290** (2000), no. 5500, 2323–2326.
- [154] P. Salembier, L. Garrido, and D. Garcia, *Auto-dual connected operators based on iterative merging algorithms*, Proceedings ISMM'98, International Symposium on Mathematical Morphology (H. Heijmans and J. Roerdink, eds.), Kluwer, 1998, pp. 183–190.
- [155] P. Salembier and J. Serra, *Morphological multiscale image segmentation*, Proc. SPIE, Visual Communications and Image Processing'92, Boston, November 1992, vol. 1818, 1992, 2745, pp. 620–631.
- [156] ———, *Flat zones filtering, connected operators, and filters by reconstruction*, *IEEE Transactions on Image Processing* **4** (1995), no. 8, 1153–1160.
- [157] B. Schölkopf, A.J. Smola, and K.-R. Müller, *Nonlinear component analysis as a kernel eigenvalue problem*, *Neural Computation* **10** (1998), no. 5, 1299–1319.
- [158] F. Schmidt, *Classification de la surface de mars par imagerie hyperpsectrale omega. suivi spatio-temporel et étude des dépôts saisonniers de CO_2 et H_2O .*, Thèse de doctorat, Université Joseph Fourier - Grenoble I, France, Laboratoire de Planétologie de Grenoble, Octobre 2007.
- [159] J. Serra, *Image analysis and mathematical morphology*, vol. I, Academic Press, London, 1982.
- [160] ———, *Image analysis and mathematical morphology : Theoretical advances*, vol. II, Academic Press, London, 1988.
- [161] ———, *Connectivity on complete lattices*, *Journal of Mathematical Imaging and Vision* **9** (1998), 81–96.
- [162] ———, *Lecture notes in computer science*, vol. 1658, ch. Set connections and discrete filtering, pp. 191–206, Springer, 1999.
- [163] ———, *Connectivity for sets and functions*, *Fundamenta Informaticae* **41** (2000), 147–186.
- [164] ———, *Cours de morphologie mathématique*, Tech. report, Centre de Morphologie Mathématique - ENSMP, <http://cmm.ensmp.fr/~serra/cours/>, 2000.
- [165] ———, *Connexions et segmentation d'image*, *Traitement du signal* **20** (2003), no. 3, 243–254.
- [166] ———, *A lattice approach to image segmentation*, *Journal of Mathematical Imaging and Vision* **24** (2006), 83–130.
- [167] ———, *Segmentation connective*, Exposé au CMM 25 Juillet 2007 (http://cmm.ensmp.fr/Seminaires/20070725_segmentation_serra/SegmentationConnectiveCMM25juillet2007.pdf), 2007.
- [168] J. Serra and M. Mlynarczuk, *Morphological merging of multidimensional data*, Proceedings 6th International Conference Stereology and Image Analysis in Materials Science STERMAT'00, 2000, pp. 385–390.
- [169] J.A. Sethian, *A marching level set method for monotonically advancing fronts*, *Proceedings Nat. Acad. Sci. USA*, vol. 93, 1996.
- [170] ———, *Level set methods and fast marching methods : Evolving interfaces in computational geometry, fluid mechanics, computer vision and materials science*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 1999.
- [171] P. Soille, *Morphologie mathématique : du relief à la dimensionalité - algorithmes et méthodes*, Thèse de doctorat en sciences agronomiques, Université Catholique de Louvain, 1992, 2505.
- [172] ———, *Morphological image analysis*, Springer-Verlag, Berlin Heidelberg, 1999.
- [173] Pierre Soille, *Generalized geodesy via geodesic time*, *Pattern Recognition Letters* **15** (1994), no. 12, 1235–1240.

- [174] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis & and machine vision*, Thomson Engineering, 2008.
- [175] B. Sumengen, L. Bertelli, and B.S. Manjunath, *Fast and adaptive pairwise similarities for graph cuts-based image segmentation (cvprw'06)*, Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, IEEE, 2006, p. 179.
- [176] Y.W. Teh and S.T. Roweis, *Automatic alignment of hidden representations*, Advances in Neural Information Processing Systems, vol. 15, The MIT Press. Cambridge, MA, USA., 2002, p. 841–848.
- [177] J.B. Tenenbaum, *Mapping a manifold of perceptual observations*, Advances in Neural Information Processing Systems, vol. 10, The MIT Press. Cambridge, MA, USA., 1998, p. 682–688.
- [178] J.B. Tenenbaum, V. de Silva, and J.C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), no. 5500, 2319–2323.
- [179] C. Thomas, *Fusion d'images de résolutions spatiales différentes*, Thèse de doctorat, Ecole des Mines de Paris, Centre Energétiques et Procédés, 2006.
- [180] C.E. Thomaz, J.P. Boardman, S. Counsell, D.L.G. Hill, J.V. Hajnal, A.D. Edwards, M.A. Rutherford, D.F. Gillies, and D. Rueckert, *A multivariate statistical analysis of the developing human brain in preterm infants*, Image and Vision Computing **25** (2007), no. 6, 981–994.
- [181] Z. Tu and S.C. Zhu, *Image segmentation by data-driven markov chain monte carlo*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), 657–673.
- [182] L. R. Tucker, R. F. Koopman, and R. L. Linn, *Evaluation of factor analytic research procedures by means of simulated correlation matrices*, Psychometrika **34** (1969), 421–459.
- [183] C. Vachier, *Extraction de caractéristiques, segmentation d'image et morphologie mathématique*, Thèse de doctorat, Ecole des Mines de Paris, Centre de Morphologie Mathématique, 1995.
- [184] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik, *Dimensionality reduction : A comparative review*, Submitted to Neurocognition (2008).
- [185] F. van der Meer, *The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery*, International Journal of Applied Earth Observation and Geoinformation **8** (2006), no. 1, 3–17.
- [186] M.A.O. Vasilescu and D. Terzopoulos, *Multilinear analysis of image ensembles*, Proc. of the European Conf. on Computer Vision (ECCV'02), 2002, pp. 447–460.
- [187] ———, *Multilinear analysis of image ensembles*, IEEE Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, 2005, pp. 547 – 553.
- [188] L. Vincent and P. Soille, *Watersheds in digital spaces : An efficient algorithm based on immersion simulations*, Pattern Analysis and Machine Intelligence **13** (1991), no. 6, 583–598.
- [189] K.L. Vincken, A.S.E. Koster, and M.A. Viergever, *Probabilistic multiscale image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997), 109–120.
- [190] F. Zanoguera, *Segmentation interactive d'images fixes et de séquences vidéo basée sur des hiérarchies de partitions*, Thèse de doctorat, Ecole des Mine de Paris, Centre de Morphologie Mathématique, 2001.
- [191] F. Zanoguera and F. Meyer, *On the implementation of non-separable vector levelings*, Proceedings International Symposium on Mathematical Morphology ISMM'02 (Sydney, Autralie) (H. Talbot and R. Beare, eds.), CSIRO, 2002, pp. 369–377.
- [192] Z. Zhang and H. Zha, *Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment*, SIAM Journal of Scientific Computing **26** (2004), no. 1, 313–338.