HAL
open science

Nouvel Algorithme pour la Réduction de la Dimensionnalité en Imagerie Hyperspectrale

Jihan Khoder

To cite this version:

Jihan Khoder. Nouvel Algorithme pour la Réduction de la Dimensionnalité en Imagerie Hyperspectrale. Traitement du signal et de l'image [eess.SP]. Université de Versailles-Saint Quentin en Yvelines, 2013. Français. NNT : . tel-00939018

HAL Id: tel-00939018

https://theses.hal.science/tel-00939018v1

Submitted on 29 Jan 2014

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.

THÈSE EN COTUTELLE

Présentée par

Jihan Fawaz Khoder

POUR OBTENIR LE GRADE DE DOCTEUR DELIVREE PAR

L'Université de Versailles St Quentin-en-Yvelines

Et L'Université Libanaise

SPÉCIALITÉ : GENIE INFORMATIQUE, AUTOMATIQUE ET TRAITEMENT DE SIGNAL

Nouvel Algorithme pour la Réduction de la Dimensionnalité en Imagerie Hyperspectrale

---

Soutenue publiquement le : 24 Octobre 2013

Devant la commission d'examen composée de :

Stephane SERFATY	Professeur, Université de Cergy-Pontoise	Président
Fahed ABDALLAH	MC-HDR, Université de Technologie de Compiègne	Rapporteur
Hassan GHAZIRI	Professeur, Beirut Research and Innovation Center	Rapporteur
Paul HONEINE	MC-HDR, Université de Technologie de Troyes	Examinateur
Clovis FRANCIS	Professeur, Université Libanaise, Tripoli	Examinateur
Fethi B.OUEZDOU	Professeur, Université de Versailles, Versailles	Co-Directeur
Rafic YOUNES	Professeur, Université Libanaise, Beyrouth	Co-directeur
Gérad SINPRASEUTH	Président de MyDon SAS, Paris	Invité

DEDICACES

À mes chers et tendres parents, Fawaz et Joumana,
sans vous, rien n'aurait pu être possible,
que Dieu vous garde pour moi
et vous prête une longue vie
pleine de santé et de prospérité…

À mes sœurs et mes frères,
merci de m'avoir soutenu tout au long de mes études
et tout au long de cette aventure,
que Dieu vous préserve…

À la personne que rien ne lui récompense,
à l'acteur principal derrière la réussite de ce travail,
à la personne qui a tant sacrifié et qui était à tout moment à côté de moi,
que Dieu me donne le pouvoir de te récompenser…

REMERCIEMENTS

Je tiens à adresser en premier lieu mes plus chaleureux remerciements à mon directeur de thèse Pr. Rafic Younes. Il n'a pas simplement accepté de diriger ma thèse; il m'a aussi transmis la passion de la recherche et n'a jamais cessé de m'encourager et de me soutenir durant ces trois années. J'ai pu apprécier non seulement sa dimension scientifique, mais aussi sa dimension humaine. J'en profite pour lui exprimer ici ma plus profonde gratitude.

Mes remerciements s'adressent ensuite au Pr. Fethi Ben Ouezdou qui m'a accueilli dans son équipe. Il m'a assuré le soutien scientifique et morale au cours des années de ma recherche.

Pr. Fahd Abdallah et Pr. Hassan Ghaziri m'ont fait l'honneur d'être rapporteurs sur ma thèse. ils ont pris le temps de m'écouter et de discuter avec moi. Leurs remarques m'ont permis d'envisager mon travail sous un autre angle. Je les remercie infiniment.

J'associe à ces remerciements Pr. stéphane Serfaty, Pr. Clovis Francis, Pr. Paul Honeiné et M. Gérad Sinpraseuth pour avoir accepté d'examiner mon travail.

L'association Azm & Saade et L'Ecole Doctorale EDST au Liban meritent de ma part une attention particulière dans ces remerciements. Qu'ils trouvent ici toute ma reconnaissance pour leur soutien financier.

J'adresse toute ma gratitude à tous mes ami(e)s et à toutes les personnes qui m'ont aidé dans la réalisation de ce travail. Je n'oublie surement pas Pr. Luc Chassagne, directeur du LISV, qui m'a epargné de traiter beaucoup des difficultés administratifs. Je dois aussi mentionner le coté amical et professionnel de Patricia Chartier, Dominique Maillet et Véronique Delahaye.

Enfin, je souhaite remercier ma famille pour leur soutien constant. Cette thèse, aboutissement de longues années d'études, je la dois beaucoup à mon frère, mes sœurs et mes parents exceptionnels avec qui j'ai vécu dans un climat toujours serein, à l'abri de tous soucis affectifs. Il m'est impossible de trouver les mots pour leur dire à quel point je suis fière d'eux, et à quel point je les aime.

Une pensée pour terminer ces remerciements pour toi papa qui n'a pas vu l'aboutissement de mon travail mais je sais que tu en aurais été très fière de ta fille !! Je t'aime papa.

RESUME

Mots clés :
Image hyperspectrale ; Méthodes de projection ; Sélection de bandes ; Indice de validité ; Critère de stabilité ; Estimation de la dimension intrinsèque ; Critère de similarité ; Perte d'information ; Réduction de la dimension.

En Imagerie hyperspectrale, les volumes de données acquises atteignent souvent le gigaoctet pour une seule et même scène observée. De ce fait, l'analyse de ces données au contenu physique complexe passe obligatoirement par une étape préliminaire de réduction de la dimensionnalité. Cette réduction a un double objectif, le premier consiste à réduire la redondance et le second permet de faciliter les traitements postérieurs (extraction, classification et reconnaissance de formes) et donc l'interprétation des données. La classification automatique est une étape importante du processus d'extraction de connaissances à partir des données. Elle vise à découvrir la structure intrinsèque d'un ensemble d'objets en formant des regroupements qui partagent des caractéristiques similaires.

Dans cette thèse, nous nous intéressons à la réduction de dimension dans le cadre de la classification non supervisée des bandes spectrales. Différentes approches existent, comme celles basées sur la projection (linéaire ou non-linéaire) des données de grandes dimensions sur des sous-espaces de représentation bien choisis ou sur les techniques de sélection de bandes spectrales exploitant des critères de complémentarité-redondance d'information qui ne permettent pas de préserver toute la richesse de l'information apportée par ce type de données.

1 - Nous avons accompli une étude comparative, sur la stabilité et la similarité des algorithmes des méthodes non paramétriques et non supervisée de la projection et aussi de la sélection des bandes utilisées dans la réduction de la dimensionnalité à différents niveaux de bruit déterminés. Les tests sont effectués sur des images hyperspectrales, en classant ces derniers en trois catégories selon leur degré de performance de préserver la quantité d'informations.

2 - Nous avons introduit une nouvelle approche de critère basée sur la di-similarité des attributs spectraux et utilisée dans un espace local sur des matrices de données ; L'approche a servi pour définir un taux de préservation d'un évènement rare dans une transformation mathématique donnée. Cependant, nous avons limitée son application au contexte de la thèse liée à la réduction de la taille des données dans une image hyperspectrale.

3 – Les études comparatives ont permis une première proposition d'approche hybride pour la reduction de la taille d'une image hyperspectrale permettant une meilleure stabilité : *BandClustering* avec *Multidimensional Scaling* (MDS). Des exemples sont donnés pour démontrer l'originalité et la pertinence de l'hybridation (BandClust / MDS) de l'analyse effectuée.

4 – La tendance de l'hybridation a été généralisée par la suite en présentant un algorithme hybride adaptatif non supervisé basé sur la logique flou (Fuzzy C means), une méthode de projection comme l'analyse en composante principale (ACP) et un indice de validité d'une classification. Les classifications effectuées par Fuzzy C means permettent d'affecter chaque pixel d'une image hyperspectrale à toutes les classes avec des degrés d'appartenance variant entre 0 et 1. Cette propriété rend la méthode FCM intéressante pour la mise en évidence soit des transitions progressives entre les différentes bandes spectrales ou des hétérogénéités spectrales. Grâce à des méthodes conventionnelles appelées indices de validité de classes, nous avons déterminé le nombre optimal de classes de FCM ainsi que le paramètre de flou. Nous montrons que cette hybridation conduit à un taux de réduction pertinent dans l'imagerie hyperspectrale. Par conséquent, Cet algorithme appliqué à différents échantillons de données hyperspectrales, permet une imagerie spectrale beaucoup plus informative, notamment au niveau de l'hétérogénéité spectrale.

ABSTRACT

Keywords: Fuzzy Hybridized algorithm ; Band selection ; Projection method ; Estimator dimension reduction ; Hyperspectral imagery ; Validity index ; Dimensionality reduction.

In hyperspectral imaging, the volumes of data acquired often reach the gigabyte for a single scene observed. Therefore, the analysis of these data complex physical content must go with a preliminary step of dimensionality reduction. Therefore, the analyses of these data of physical content complex go preliminary with a step of dimensionality reduction. This reduction a two objective, the first is to reduce redundancy and the second facilitates post-treatment (extraction, classification and recognition) and therefore the interpretation of the data. Automatic classification is an important step in the process of knowledge extraction from data. It aims to discover the intrinsic structure of a set of objects by forming groups that share similar characteristics.

In this thesis, we focus on dimensionality reduction in the context of unsupervised classification of spectral bands. Different approaches exist, such as those based on projection (linear or nonlinear) of high-dimensional data in a representation subspaces or on the techniques of selection of spectral bands exploiting of the criteria of complementarity-redundant information do not allow to preserve the wealth of information provided by this type of data.

1- We have made a comparative study on the stability and similarity of nonparametric algorithms and unsupervised projection methods and also the selection of spectrales bands used in the dimensionality reduction at different noise levels determined. The tests are conducted on hyperspectral images, classifying them into three categories according to their level of performance to preserve the amount of information.

2- We introduced a new approach of criterion based on the dissimilarity of the attributes spectral and used into a local space on matrices of data; the approach was used for to define a rate of safeguarding of a rare event in a given mathematical transformation. However, we limited his application to the context of the thesis related to the reduction of the size of the data in a hyperspectral image.

3- The comparative studies allowed a first hybrid proposal for an approach for the reduction of the size of an image hyperspectral allowing a better stability: BandClustering with Multidimensional Scaling (MDS). Examples are given to show the originality and the relevance of the hybridization (BanClust\MDS) of the analysis carried out.

4- The trend of hybridization was generalized thereafter by presenting an adaptive hybrid algorithm unsupervised based on Fuzzy logic (Fuzzy C-means), a method of projection as the analysis in principal component (PCA) and a validity index of a classification. The classifications carried out by Fuzzy C-means, make it possible to assign each pixel of an image hyperspectral to all the classes with degrees of membership varying between 0 and 1. This property makes the FCM interesting method for the detection of progressive or between different spectral bands or from spectral heterogeneities. With conventional methods known validity indices of classes, we determined the optimal number of classes of FCM and the fuzzy parameter. We show that this hybridization leads to a relevant reduction rate in hyperspectral imaging. Consequently, this algorithm applied to different samples of hyperspectral data, allows a spectral imagery much more informative, in particular on the level of spectral heterogeneity.

LISTE DES PUBLICATIONS

Articles publiés dans des conférences et journaux

J. Khoder, R. Younès: "*Dimensionality Reduction on Hyperspectral Images: A Comparative Review Based on Artificial Datas*", The 4th International Congress on Image and Signal, 15-17 October 2011, Shanghai, China.

J. Khoder, R. Younès, F.B. Ouezdou: "*Similarity of Dimensionality Reduction Methods Applied on Artificial Hyperspectral Images*", The 2012 World Congress in Computer Science, Compute Engineering, and Applied Computing, 16-19 July, 2012, Las Vegas, Nevada, USA.

J. Khoder, R. Younès, F.B. Ouezdou: "*Stability of Dimensionality Reduction Methods Applied on Artificial Hyperspectral Images*", Lecture Notes in Computer Sciences, Volume 7594, pp 465-474, 2012.

J. Khoder, R. Younès, F.B. Ouezdou: "*Potential of Hybridization Methods to Reducing the Dimensionality for Multispectral Biological Images*", The 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13), July 3-7, 2013, Osaka, Japan.

J. Khoder, R. Younès, F.B. Ouezdou: "*Proposal for Preservation Criteria to Rare Event. Application on Multispectral/Hyperspectral Images*", 25nd International Conference on Microelectronics, IEEE ICM 2013, Beirut, December 2013.

Articles soumis à des conférences et journaux

J. Khoder, R. Younès, F.B. Ouezdou: "*Adaptive Fuzzy Hybrid Method to Reducing the Dimensionality for Multispectral/Hyperspectral Images*", IEEE Transactions on Pattern Analysis and Machine Intelligence.

LISTE DES FIGURES

Figure 1. Du monochrome à l'hyperspectral 1
Figure 2. Même scène dans diverses bandes spectrales à 458 nm, 664 nm, 712 nm et 1211 nm. 1
Figure 3. Données hyperspectrales 2
Figure 4. Termes utilisés pour la position des pixels dans le cube hyperspectral. 2
Figure 5. Comparaison des coefficients de corrélation spatiale et spectrale 3
Figure 6. Corrélation spectrale (inter bandes), sur les images AVIRIS de Railroad Valley (a) et sur Moffett (b). 4
Figure 7. Principe d'acquisition par un capteur pushbroom. 7
Figure 8. Effet de smile : une ligne de l'image a une projection déformée sur le CCD. 11
Figure 9. Effet de keystone : la projection sur le CCD des bandes spectrales d'un même pixel est oblique. 11
Figure 10. Principe de la sélection de variables ou bandes. 18
Figure 11. Illustration de la carte de kohonen 27
Figure 12. Architecture standard d'un réseau multicouche. 28
Figure 13. Bandes 10 D'AVIRIS Indiana Pines 34
Figure 14. Procedure de recherche de FPBS, BPBS et BBBS 36
Figure 15. Mécanisme "clustering" 37
Figure 16. Les 3 bandes sélectionnées 40
Figure 17. Image de verdure avec évènement rare 53
Figure 18. Facteur de pertinence pour deux méthodes de réduction 54
Figure 19. Les Quatre images artificielles. 56
Figure 20. Type de cellules anormales pour le carcinome de côlon 57
Figure 21. Image de l'Indiana Pines *Figure 22. Image du Vallée Salinas* 58
Figure 23. Image : Rangement des objets. 58
Figure 24. 1ere et 2eme catégorie de courbes sur la perte d'information des méthodes de projection 59
Figure 25. 1er et 2 ^{eme} catégorie des méthodes de projections testés avec MAE, MQE et MAD 60
Figure 26. 1er et 2eme catégorie des méthodes de projections testés avec 1er , 2eme derivée spectral 61
Figure 27. 1er et 2 ^{eme} categorie des méthodes de projections testés avec l'entropie. 61
Figure 28. Méthode proposée pour l'étude de la stabilité appliquée sur les images hyperspectrales. 62
Figure 29. Méthodes de projection, 1ere Catégorie des méthodes de réduction 63
Figure 30. Méthodes de projection, 2eme Catégorie des méthodes de réduction 64
Figure 31. Méthodes de projection, 3eme Catégorie des méthodes de réduction 64
Figure 32. Méthodes de projection, 1ere Catégorie des méthodes de réduction 65
Figure 33. Méthodes de projection, 2eme Catégorie des méthodes de réduction 65
Figure 34. Méthodes de projection, 3eme Catégorie des méthodes de réduction 66
Figure 35. La sélection de bandes, Courbes de performances pour la stabilité 67
Figure 36. Pourcentage de similarité obtenus pour les diverses méthodes de réduction 68
Figure 37. 1ere et 2eme catégorie des méthodes de projections testés avec le temps de calcul. 69
Figure 38. L'Algorithme hybride BandClust/MDS 70
Figure 39. NCC et Structural Content for Hybrid BandClust/MDS Algorithm 70
Figure 40. Pourcentage de la similarité BandClust/MDS 70
Figure 41. Principe de notre approche hybride adaptative 72
Figure 42. Modèle hybride flou adaptative pour la réduction de la dimension. 73
Figure 43. Sept bandes réduites de cellules du côlon anormal. 83

LISTE DES TABLEAUX

Tableau 1. Estimation de l'entropie à différents ordres d'une image hyperspectrale. Cette estimation devient rapidement peu fiable à cause du manque de données pour évaluer la loi de probabilité. 7

Tableau 2. Capteurs et Caractéristiques pour images Hyperspectrales. 10

Tableau 3. Classification des méthodes de projection pour la réduction de dimensionnalité selon les critères relatifs aux pertes d'informations. 62

Tableau 4. Classification des méthodes de projection pour la réduction de dimensionnalité selon les critères relatifs à la stabilité. 65

Tableau 5. Les diverses méthodes de la réduction par projection utilisée, de manière à classer, sur des données artificielles avec des variances de bruit différents. 69

Tableau 6. Performance moyenne de l'AHFARD sur les trois applications. 84

Tableau 7. Comparaison entre le PCA seule et notre approche. 85

NOMENCLATURE

m	La dimension de l'espace de départ
r	La dimension de l'espace réduit
$\boldsymbol{n_\lambda}$	Le nombre de bandes
$\boldsymbol{n_x}$	Le nombre de pixels en colonnes
$\boldsymbol{n_\lambda}$	Le nombre de bandes spectrales
S	Covariance
H	L'entropie
Ω	La matrice de transformation
$\boldsymbol{n_y}$	Le nombre de pixels en lignes
V	La variance
$\mathbf{d_{ij}}$	La distance entre deux points dans l'espace d'origine(r dimensions).
$\boldsymbol{\delta_{ij}}$	La distance entre deux points dans l'espace réduite.
$\boldsymbol{H_u}$	Histogramme globale associé à la bande $\boldsymbol{B_u}$
$\boldsymbol{H_v}$	Histogramme globale associé à la bande $\boldsymbol{B_v}$
$\boldsymbol{P_l}$	La distribution de probabilité associée à la bande l.
$\boldsymbol{P_k}$	La distribution de probabilité associée à la bande k.
$\boldsymbol{M_T}$	La fiabilité
$\boldsymbol{M_C}$	La continuité
C(x,y)	Le contraste entre deux faisceaux
I(x,y)	La fonction de comparaison de luminance
$\boldsymbol{\sigma_{xy}}$	La similarité par la perte de corrélation entre les deux vecteurs
K(x,y)	La fonction de noyau
$\mathbf{W_{ij}}$	Le noyau de chaleur
$\boldsymbol{I(x,y,\lambda)}$	La matrice qui représente l'image hyperspectrale
$\boldsymbol{\tilde{I}(x,y,\lambda)}$	La matrice moyenne qui représente l'image hyperspectrale
F	Le coefficient de partition
$\mathbf{V_{fs}}$	Indice de Fukuyama and Sugeno
$\mathbf{V_{XB}}$	Indice de Xie and Beni
$\mathbf{V_K}$	Indice de Kwon
$\mathbf{V_{PBMF}}$	Indice de Pakhira
$\mathbf{V_{SV}}$	Indice de Kim
$\mathbf{V_{cwb}}$	Indice de rezaee
DMLE	Maximum de vraisemblance
$\mathbf{I_A}$	La fonction indicatrice de l'événement A
P	Exposant de proportion
N (r)	Le nombre de hyper sphères de rayon r

SOMMAIRE

Chapitre 1 ..... **1**

Les contraintes de l'hyperspectral **1**
1.1 Pourquoi faire des images hyperspectrales **1**
1.2 Proprietes des images hyperspectrales **3**
1.3 Problème de la normalisation **4**
1.4 Un espace presque vide **5**
1.5 Estimation de l'entropie **5**
1.6 Acquisition des images hyperspectrales **7**
1.6.1 aviris *7*
1.6.2 CASI, CASI-2 ..... *8*
1.6.3 MIVIS *8*
1.6.4 Hymap *8*
1.6.5 Hypérion *8*
1.6.6 CHRIS-PROBA *8*
1.6.7 COIS-NEMO *8*
1.6.8 Spectra *8*
1.6.9 Hydice *8*
1.6.10 MERIS *9*
1.7 Caractéristiques des capteurs hyperspectraux **9**
1.8 Dégradation typiques sur les données **10**
1.9 Les applications **12**
1.9.1 Agriculture de précision *12*
1.9.2 Foresterie *12*
1.9.3 Aménagement du territoire *12*
1.9.4 Qualité de l'eau *12*
1.9.5 Détection des algues *13*
1.9.6 Les géosciences-science de la terre solide *13*
1.9.7 Les sciences médicales *13*
1.9.8 Microscopie *13*
1.9.9 Les applications de la surveillance et militaires *13*
1.9.10 Fabrication de processus *14*
1.9.11 Fabrication pharmaceutique *14*
1.9.12 L'urbain *14*
1.9.13 Remote sensing - Télédétection *15*
1.9.14 Recherches spatiales et capteurs satellitaires *15*
1.9.15 Fabrication pharmaceutique *15*
1.9.16 Antiquités, œuvres d'art et la validation du document *15*
1.9.17 L'atmosphère *16*
1.9.18 Le Traitement du Signal et de l'Image *16*
1.9.19 La défense *16*
1.10 Problématique de la thèse **17**
1.11 Plan du manuscrit **18**

Chapitre 2 ..... **21**

État de l'art sur la réduction de la dimmensionalité **21**
2.1 Introduction **21**
2.2 Approches par projection linéaire **21**
2.2.1 Analyse en Composante Principale(ACP) *22*
2.2.2 Locality Preserving Projections (LPP) : Projections Préservées Localement *23*
2.2.3 Positionnement Multidimensionnel (MDS) *23*
2.3 Approche par projection non linéaire **24**

2.3.1 ISOMAP (Cartographie des Caractéristiques Isométriques) ... 24
2.3.2 Coordination Localement Linéaire (LLC) ... 25
2.3.3 Manifold Charting : Cartographies Multiples ... 25
2.3.4 Diffusion Maps : Cartes de Diffusion ... 26
2.3.5 Hessian-based LLE : Hessien basée sur LLE ... 26
2.3.6 Multilayer-Autoencoders (Auto-encodeurs Multicouches) ... 27
2.3.7 Sammon Mapping : Cartes de Sammon ... 28
2.3.8 Maximum Variance Unfolding (MVU) : Séquences sur le Maximum de Vraisemblance ... 29
2.3.9 Plongement Localement Linéaire (LLE) ... 29
2.3.10 Local Tangent Space Analysis (LTSA) : Analyse Tangent et Local de l'Espace ... 31
2.3.11 Laplacien Eigenmaps ... 31
2.3.12 ACP Kernelisée ... 32
2.4 Approches par selection de bandes spectrales ... 33
2.4.1 La propagation d'affinité (AP) ... 33
2.4.2 Séléction par multicritères ... 34
2.4.3 La Sélection Progressive des Bandes (PBS) ... 35
2.4.4 Méthodes hiérarchiques ... 36
2.4.5 Sélection par Prédiction et Projection ... 37
2.4.6 Méthodes basés sur la Théorie d'Information ... 38
2.4.7 Méthode basée sur l'Histogramme ... 39
2.5 Conclusion ... 40

Chapitre 3 ... 42

Critères d'évaluation des méthodes de réduction de la dimensionnalité ... 42
3.1 Critères locaux ... 42
3.1.1 Local Continuity Meta–Criterion (LCMC): Le Critère de la Continuité Locale ... 42
3.1.2 La Fiabilité ($\boldsymbol{MT}$) ... 43
3.1.3 La Continuité ($\boldsymbol{MC}$) ... 43
3.1.4 Mean Relative Rang Error (MRRE): Le Rang d'Erreur Moyenne Relative ... 43
3.2 Critères relatives à la perte d'information ... 43
3.2.1 L'Entropie ... 44
3.2.2 La Variance ... 44
3.2.3 Dérivée Spectral d'Ordre 1 ... 44
3.2.4 Erreur Moyenne Absolu (EMA) ou Moyenne de $\boldsymbol{l1}$ ... 44
3.2.5 Erreur Quadratique Moyenne (EQM) ou Moyenne de $\boldsymbol{l22}$ ... 44
3.2.6 Erreur Quadratique Moyenne Relative (RMSE) ... 45
3.2.7 Average RMSE (a-rmse) : la Moyenne du RMSE ... 45
3.2.8 Peak RMSE (p-rmse) : Maximum du RMSE ... 45
3.2.9 Maximum Absolute Distorsion : Distorsion Maximale Absolue ... 45
3.2.10 Percentage MAD : Pourcent sur MAD ... 45
3.2.11 Spectral Information Divergence (SID) : la Divergence de l'Information Spectrale ... 45
3.2.12 Information Mutuelle ... 46
3.3 Critères relatives à la stabilité ... 46
3.3.1 Structural Content (SC) : Contenu Structurel ... 46
3.3.2 Normalized Cross-Correlation (NCC): L'Inter-Corrélation Normalisée ... 46
3.4 Critères relatives à la similarité ... 46
3.4.1 La Fidélité (F) ... 47
3.4.2 Similarité Spectrale (SS) ... 47
3.4.4 Indice de Similarité Structurale ... 47
3.5 Critères basés sur les vecteurs spectraux ... 48
3.5.1 Distance Euclidienne Minimale (EMD) ... 48
3.5.2 Average RMSE : La Moyenne du RMSE ... 49
3.5.3 Peak RMSE (P-RMSE) : Le Maximum du rmse ... 49
3.5.4 Maximum Spectral Angle : Angle Spectral Maximal ... 49
3.5.5 Distance Mahalanobis ... 49
3.6 Critères pour les évènements rares ... 50

3.6.1 Etat de l'Art *50*
3.6.2 Proposition d'un nouveau critère : *51*
3.6.3 Application sur une image réelle : *53*
3.7 Conclusion **54**

Chapitre 4 ..... **55**

Etude comparative des méthodes de réduction de la dimensionnalité **55**
4.1 Introduction **55**
4.2 Données expérimentales **55**
4.2.1 Application aux données synthétiques *56*
4.2.3 Application à l'imagerie biomédicale *56*
4.2.2 Application à l'Indiana Pines et Salinas *57*
4.2.4 Image contenant un rangement d'objets *58*
4.3 Etudes comparatives sur la perte d'informations **59**
4.4 Etudes comparatives sur la stabilité **62**
4.4.1 Méthodes de projection : Normalized Cross Correlation(ncc) *63*
4.4.2 Méthodes de projection: Structural Content (sc) *65*
4.4.2 Méthodes de sélection de bandes *66*
4.5 Etudes comparatives sur la similarité **67**
4.6 Etudes Comparatives sur le temps de calcul **69**
4.7 Conclusion et perspectives **69**

Chapitre 5 ..... **72**

l'Approche hybride proposée **72**
5.1 Introduction **72**
5.2 Approche hybride pour la réduction de la dimensionnalite **73**
5.3 Classification fcm **74**
5.4 Estimation de la dimension intrinsèque **75**
5.4.1 Estimateurs locaux *75*
5.4.2 Estimateurs globaux *76*
5.5 Les indices de validité **77**
5.5.1 Coefficient de Partition (F) *78*
5.5.2 Entropie de Partition (H) *78*
5.5.3 Exposant de Proportion (P) *79*
5.5.4 Indice de Fukuyama et Sugeno, V_{fs} *79*
5.5.5 Indice de Xie et Beni, V_{XB} *79*
5.5.6 Indice de Kwon, V_K *80*
5.5.7 Indice de rezaee, V_{cwb} *80*
5.5.8 Indice de Pakhira, V_{PBMF} *81*
5.5.9 Indice de Kim, V_{SV} *81*
5.6 Etude expérimentale **82**
5.6.1 Discussion reliée à l'Indiana Pines *82*
5.6.2 Discussion reliée à l'image biomédicale *83*
5.6.3 Discussion reliée au jeu artificiel de objets *83*
5.7 Effet des critères de qualité **84**
5.8 Conclusion **85**

Chapitre 6 ..... **86**

Conclusions et perspectives **86**

7. Références **89**

CHAPITRE 1
LES CONTRAINTES DE L'HYPERSPECTRAL

1.1 POURQUOI FAIRE DES IMAGES HYPERSPECTRALES

L'observation de la Terre depuis l'espace s'est d'abord faite en monochrome, principalement dans le domaine militaire avec une récupération physique des films. Les capteurs électroniques, se sont par la suite vite développés, présentant des avantages évidents pour la récupération des images. Il est ensuite apparu que l'observation de la même scène à plusieurs longueurs d'onde permettait une meilleure exploitation des données. Un très grand nombre de capteurs multi spectraux se sont ainsi développés, le premier étant Landsat au début des années 70 [4]. Pour permettre la formation d'images aux couleurs naturelles, l'observation est alors souvent faite dans les trois couleurs usuelles (rouge, vert et bleu) dans le proche infrarouge. Cette dernière bande est particulièrement intéressante pour l'observation des végétaux qui ont une réponse spectrale forte à cause de la présence de chlorophylle.

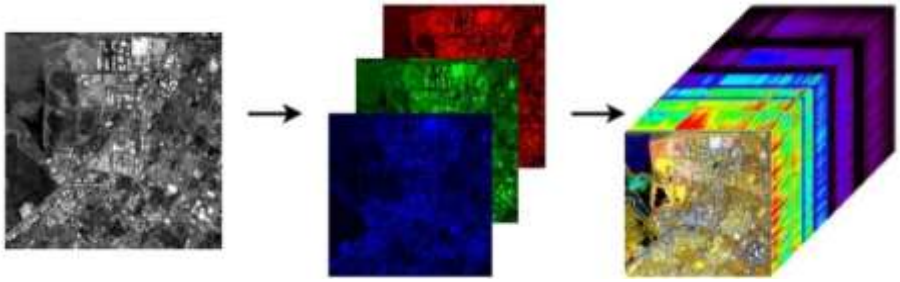


Figure 1. Du monochrome à l'hyperspectral

L'évolution naturelle des capteurs d'images a conduit à l'acquisition non pas d'une, trois ou quatre bandes spectrales mais plutôt de plusieurs centaines (**Figure 1**).
L'ajout de bandes spectrales permet d'augmenter le pouvoir discriminant des données acquises (**Figure 2**). On peut ainsi arriver à différencier deux matériaux possédant une couleur identique à l'œil. Par exemple, une peinture verte et une feuille qui ont la même couleur, i.e. la même réponse spectrale dans le rouge, le vert et le bleu, ne pourront pas être différenciées à l'œil, l'ajout d'autres bandes spectrales permettront de faire la différence. Les images hyperspectrales tirent parti de ces propriétés.

Figure 2. Même scène dans diverses bandes spectrales à 458 nm, 664 nm, 712 nm et 1211 nm.

Une image hyperspectrale est obtenue grâce à un spectro-imageur. L'acquisition d'une même scène est réalisée dans plusieurs bandes spectrales. L'évolution entre les capteurs multispectraux et hyperspectraux suit la même logique qu'entre les capteurs monochromes et multispectraux. La différence par rapport aux images multispectrales tient au nombre important de bandes (100 à 200), à leur largeur fine (10 à 20 nm) et au fait qu'elles soient contigües. Cette dernière propriété permet une reconstruction du spectre de chaque pixel: On réalise en fait un échantillonnage du spectre. Cet échantillonnage doit être assez fin pour permettre une bonne reconstruction. Les techniques dites de Full Spectral Imaging (FSI) devraient conduire de plus en plus à cette vision spectrale des données [1]. Les sondeurs atmosphériques actuels arrivent à plusieurs milliers de bandes (2378 pour l'instrument AIRS de la NASA) et on parle alors d'ultra-spectral, mais ces capteurs ne sont pas encore imageurs. Certains instruments, comme **Hymap**, ne conservent pas les valeurs dans les bandes d'absorption de l'eau qui sont situées dans l'infrarouge (autour de 1.4 et 1.9 µm), mais comme ces bandes contiennent plus de bruit que de signal, on peut en faire abstraction.

Les données hyperspectrales sont donc acquises selon trois dimensions: deux spatiales et une spectrale. Une représentation possible de ces données peut être faite sous la forme d'un cube hyperspectral (**Figure 3**). La face supérieure du cube correspond à la scène spatiale, souvent une composition colorée de trois bandes spectrales. Toutes les scènes pour les différentes longueurs d'onde sont ensuite empilées pour donner le cube. Les autres faces du cube représentent alors respectivement les luminances, selon les longueurs d'onde (λ), des lignes (x) et des colonnes (y) en bordure du cube.

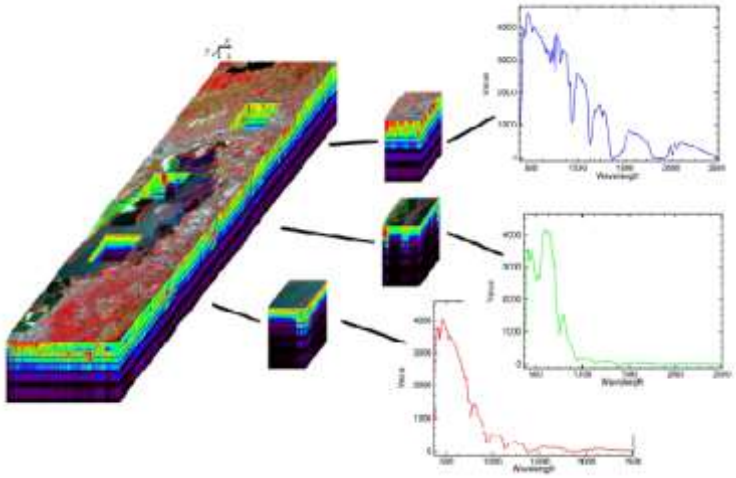


Figure 3. Données hyperspectrales

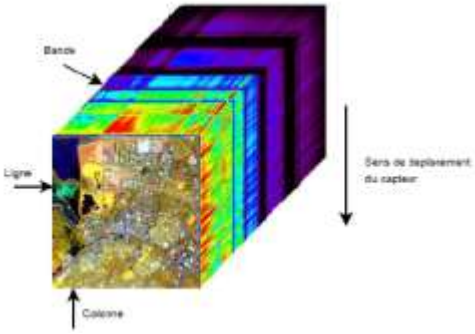


Figure 4. Termes utilisés pour la position des pixels dans le cube hyperspectral.

Sur la droite de la figure 3, sont représentés des exemples de spectres des pixels de cubes extraits de l'image totale. L'abscisse correspond aux longueurs d'onde et l'ordonnée aux luminances.

Sur la **figure 4** (le cube de données ayant tourné pour avoir la scène spatiale sur l'avant du cube), sont détaillés les termes utilisés pour préciser la position des pixels dans une image hyperspectrale. On notera I(x, y, λ) la valeur sur la colonne x (on parle également d'échantillon), la ligne y et dans la bande spectrale λ. Les conditions d'acquisition de ces images ainsi que les contraintes spécifiques du spatial vont être précisées ci-après, mais on peut déjà entrevoir que ces images hyperspectrales constituent un flot de données considérable à transmettre. Ainsi, l'objectif principal de cette thèse est de proposer une méthode de réduction de la dimension sans perte d'informations.

1.2 PROPRIETES DES IMAGES HYPERSPECTRALES

Une première manière de voir les données est de considérer chaque valeur indépendamment des autres, où chaque valeur correspondant à la luminance d'un pixel pour une longueur d'onde donnée. On considère alors les trois directions comme équivalentes. En réalité, les propriétés de ces valeurs sont différentes selon les directions spectrales et spatiales. Dans les directions spatiales, la corrélation est forte à faible distance et décroit rapidement quand le décalage augmente (**Figure 5(a)**). Au contraire, la corrélation spectrale est présentée pour tout le spectre (**Figure 5(b)**). Les propriétés statistiques sont donc différentes selon la direction considérée.

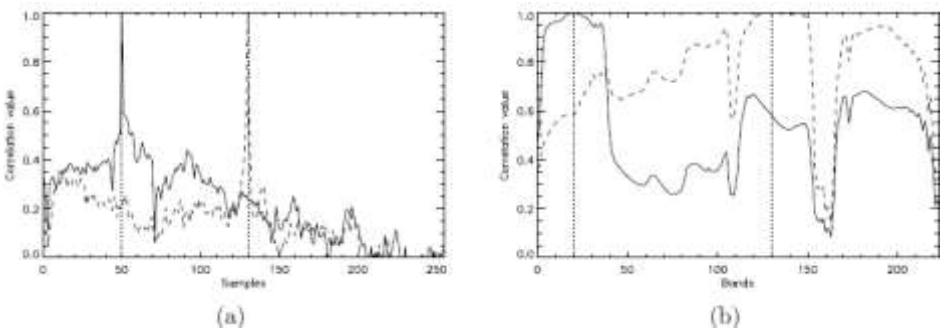


Figure 5. Comparaison des coefficients de corrélation spatiale (a) et spectrale (b) en fonction du décalage entre les lignes ou entre les bandes.
(a) ligne continue : corrélation entre la colonne 50 et les autres ; ligne pointillée : corrélation entre la colonne 130 et les autres.
(b) ligne continue : corrélation entre la bande spectrale 20 et les autres; ligne pointillée: corrélation entre la bande spectrale 130 et les autres. On remarque que la corrélation spectrale est beaucoup plus forte.

La matrice de corrélation spectrale, dont l'utilité dans l'analyse des données hyperspectrales a été montrée par Simmons et Brower [2], représente le facteur de corrélation entre les différentes bandes spectrales. La **figure 6** présente deux exemples de cette matrice sur deux images différentes. Le coefficient (i, j) de cette matrice de corrélation représente la corrélation entre les bandes i et j et est défini par un coefficient de corrélation situé donc dans l'intervalle [−1; 1], une valeur de 1 indiquant une égalité entre les deux bandes.

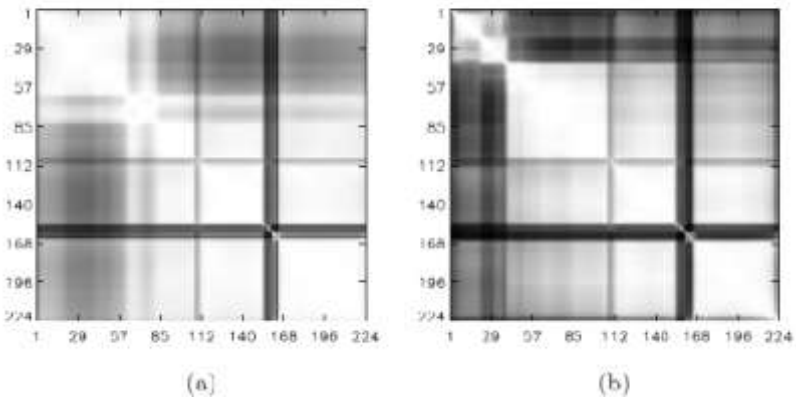


Figure 6. Corrélation spectrale (inter bandes), sur les images AVIRIS de Railroad Valley (a) et sur Moffett (b). Le pixel (i, j) de l'image de corrélation représente la valeur du coefficient de corrélation r (i, j) entre les bandes i et j, blanc correspondant à la valeur 1 et noir à 0. Les bandes noires très marquées correspondent au domaine d'absorption de la vapeur d'eau où le signal est très faible et très bruité. Il apparait que des corrélations existent entre des bandes très éloignées.

Les valeurs étant en fait fortement corrélées, il y a rarement apparition de coefficients négatifs en pratique. Pour expliquer la corrélation spectrale, on peut remarquer que le profil spectral présent des caractéristiques indépendantes du type du terrain observé. La forme générale des spectres présents des traits communs à toutes les zones observées. Cela est dû à l'influence de l'atmosphère (bandes d'absorption) et à l'éclairement solaire, dont l'énergie est maximale dans le domaine du visible et s'atténue dans le domaine de l'infrarouge. Ces caractéristiques modulent fortement le spectre reçu au niveau du capteur (spectre en luminance et non en réflectance).

1.3 PROBLEME DE LA NORMALISATION

La question de la normalisation des données est aussi une question importante pour la généralisation des résultats. Classiquement, pour se ramener à une variable centrée réduite, on normalise la variable aléatoire x de la manière suivante :

$$x_n = \frac{x - E[x]}{\sigma_x} \qquad (1)$$

$E[x]$ Étant l'espérance de x, $\sigma_x$étant sa variance.

Le problème pour l'hyperspectral est de savoir de quelle manière on considère les données. Est-ce une image (i.e. bande spectrale) pour différentes longueurs d'onde, ou un spectre pour différents points de l'image. Pour certains, l'image hyperspectral est vue comme une extension du multi-spectral. Dans ce cas, la normalisation est effectuée pour chaque image : On calcule alors la moyenne et la variance pour les valeurs d'une bande spectrale et on normalise grâce à ces facteurs (1). La normalisation est dépendante du contenu spatial de l'image : le même spectre sur deux images différentes donnera après normalisation deux spectres différents. On parlera de normalisation bande par bande.

Une autre possibilité est de faire la normalisation pour chaque pixel. On calcule alors la moyenne et la variance de chaque spectre, puis on normalise par rapport à ces valeurs. Ceci correspond à une normalisation par spectre. Le spectre normalisé obtenu est alors

indépendant du contenu spatial de l'image : un même spectre donnera le même spectre normalisé quel que soit l'image. Par contre, le fait de considérer ou non les bandes bruitées (de valeur faible) changera les résultats.

Une autre manière de normaliser les données, de façon à accorder la même importance à toutes les bandes par la suite, est de diviser chaque bande par sa valeur moyenne (2). L'intérêt de cette opération est d'équilibrer un peu la situation entre les bandes qui reçoivent peu d'énergie et Celles qui en reçoivent plus selon l'éclairement solaire et les phénomènes atmosphériques.

$$x_n = \frac{x}{E[x]} \qquad (2)$$

1.4 UN ESPACE PRESQUE VIDE

Pour les applications, une image hyperspectrale est souvent vue comme une représentation des différentes réalisations d'une variable aléatoire à n_λ dimensions, n_λ étant ici le nombre de bandes. Les différentes réalisations de cette variable aléatoire sont les différents pixels. Cette vision des données est intéressante pour la classification ou la discrimination. Les pixels proches (notion de distance) dans cet espace vectoriel sont des pixels semblables, la classification regroupera donc ces pixels proches dans une même classe.

Les capteurs hyperspectraux acquérant généralement de l'ordre de 200 bandes, les données se retrouvent alors représentées dans un espace à 200 dimensions. Les espaces de grande dimensionnalité posent des problèmes particuliers qui ont été exposés par Hughes [4]: de tels espaces sont presque vides. En effet, en considérant des données échantillonnées sur 12 bits et acquises pour 200 bandes, il y a alors 2^{12} = 4096 possibilités pour chaque bande et un total de $4096^{200} \approx 10^{720}$ localisations possibles dans l'espace vectoriel de discrimination. Même avec des images de taille importante (de l'ordre de 10^6 pixels), la probabilité d'avoir deux pixels égaux ou voisins dans cet espace vectoriel est très faible. D'autres problèmes apparaissent qui rendent difficile l'extension des idées intuitives en 3 dimensions. Landgrebe [5] a détaillé une grande partie des effets provenant de ces espaces à grande dimension.

Ces considérations ont conduit à des méthodes pour réduire le nombre de dimensions. C'est le cas, par exemple, de l'Analyse en Composantes Principales (ACP) ou l'Analyse en Composantes Indépendantes (ACI). Le nombre de dimensions est alors réduit à environ 10 ou 20. Ces décompositions projettent les données sur les vecteurs propres et sélectionnent les meilleurs vecteurs pour conserver l'énergie maximale. Il y a alors une perte d'information, mais ces pertes ne sont pas prohibitives pour certaines applications.

1.5 ESTIMATION DE L'ENTROPIE

Dans un contexte de compression sans pertes, la notion d'entropie permet d'avoir une évaluation des performances que l'on peut espérer. Dans un message, on comprend intuitivement que tous les symboles n'apportent pas la même information. Un symbole

inattendu (faible probabilité) apportera plus d'information qu'un symbole attendu (probabilité forte). Cette notion intuitive d'information a été formulée par Shannon en 1948 [6]. L'information $I(s_i)$ associée au symbole s_i est fonction de la probabilité d'apparition de s_i : $p(si)$. Cette fonction $F[p(si)]$ est définie par trois conditions :
1 - Si la source ne délivre qu'un seul message (donc une probabilité égale à 1), l'information associée à ce message est nulle;
2 - En considérant si l'union de deux évènements indépendants (s_j et s_k), $s_i = s_j \cup s_k$, avec $p(s_i) = p(s_j)*p(s_k)$, l'information $I(s_i)$ est égale à la somme des informations associées à s_j et s_k : $F[p(s_i)] = F[p(s_j)] + F[p(s_k)]$;
3 - F est continu, monotone et positive.

Une fonction F vérifiant ces trois conditions est $-\lambda$ log(.). L'unité binaire (binary unit ou bit) choisie par Shannon est définie comme la quantité d'information contenue dans une expérience binaire équiprobable : pile ou face. Ainsi, en choisissant d'utiliser les logarithmes en base 2, le facteur λ est égal à l'unité et la quantité d'information associée à un évènement s_i s'écrit :

$$H(s_i) = -log_2(p(s_i))$$

exprimée en bits. L'entropie $H(S)$ d'une source sans mémoire est définie par :

$$H(S) = -\sum_{i=0}^{n-1} p(s_i)\ log_2(p(s_i))$$

où $\{s_0,\ldots, s_{N-1}\}$ représente l'alphabet de la source.

L'entropie est ainsi une mesure de l'information moyenne de chaque symbole de la source. Elle est maximale si tous les symboles de la source sont équiprobables. De plus, l'entropie représente la longueur moyenne minimale d'un codage binaire sans pertes des données de la source. C'est pourquoi il est intéressant d'évaluer l'entropie des images hyperspectrales avant d'étudier des algorithmes de compression.

Dans un premier temps, on peut supposer avoir un modèle de source sans mémoire et estimer l'entropie des images en supposant que tous les pixels sont indépendants. Toutefois, il serait plus réaliste de prendre en compte une dépendance entre les pixels et d'introduire un modèle de Markov d'ordre 1, 2,. . .

Cependant, on se heurte rapidement à un problème d'estimation de probabilités. Dans le cas d'un modèle d'ordre 0, chacun des symboles est une valeur codée sur 16 bits et peut donc prendre 2^{16} = 65536 valeurs. En effet, après calibration, les données acquises sur 12 bits sont généralement codées sur 16. Le nombre de points dans une image hyperspectrale étant typiquement de l'ordre de $15*10^6$ (256 × 256 × 224), on peut considérer que la loi de probabilité est suffisamment bien évaluée.

Dans le cas du modèle d'ordre 1, un symbole est alors un groupe de deux valeurs : on cherche la probabilité qu'un pixel prenne une valeur sachant la valeur du pixel voisin. Ce symbole prend une valeur parmi $(2^{16})^2 \sim 4.3\ 10^9$. L'estimation de la loi de probabilité devient alors délicate. L'entropie a cependant été estimée jusqu' à l'ordre 3 pour donner une idée des ordres de grandeur (**Tableau 1**). Pour des ordres supérieurs, l'estimation de la densité de probabilité n'est plus possible de manière fiable.

Ordre	Entropie par pixels (en bits)
0	11.39
1	9.13
2	6.88
3	5.28

Tableau 1. Estimation de l'entropie à différents ordres d'une image hyperspectrale. Cette estimation devient rapidement peu fiable à cause du manque de données pour évaluer la loi de probabilité.

1.6 ACQUISITION DES IMAGES HYPERSPECTRALES

Il existe principalement deux types de capteurs hyperspectraux: ***whiskbroom*** et ***pushbroom***. Pour échantillonner le signal reçu selon la longueur d'onde, les deux utilisent un élément dispersif : parfois un prisme, mais le plus souvent un réseau. La différence se situe au niveau de l'acquisition d'une ligne (au sens de la **figure 6**). Le ***whiskbroom*** possède une barrette CCD, ou chaque élément acquiert une longueur d'onde différente. L'acquisition d'une ligne complète se fait par un système de miroir mobile pour imager une portion différente de la fauchée (largeur de l'image) a différents instants. Le capteur ***pushbroom*** possède une matrice CCD où une des dimensions correspond aux différentes longueurs d'ondes et l'autre à la fauchée de l'instrument (**Figure 7**). L'acquisition des différentes lignes se fait dans les deux cas par déplacement du capteur.

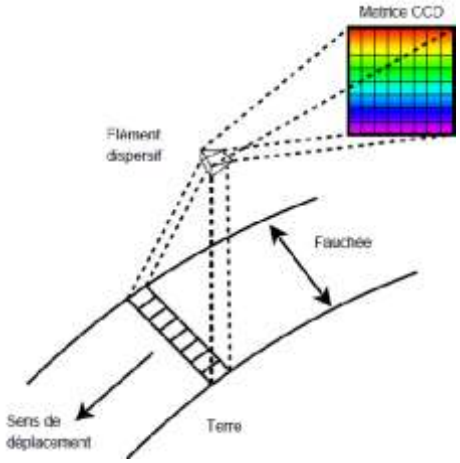


Figure 7. Principe d'acquisition par un capteur pushbroom.

L'inconvénient du ***whiskbroom*** est qu'il y a une partie mobile (le miroir), ce système est moins fiable dans un contexte de système spatial. Le ***pushbroom*** est plus robuste, mais produit en général des images rayées car les différentes colonnes de la matrice CCD n'ont pas toutes la même sensibilité. Ce défaut peut être corrigé en partie par étalonnage. La tendance actuelle pour les capteurs spatiaux est clairement orientée vers les capteurs pushbroom.

1.6.1 AVIRIS

Les spécifications datent de 1987 pour un instrument aéroporté. Aviris a été réalisé pour simuler les performances des futurs instruments hyperspectraux spatiaux [84]. D'autre part, les images sont facilement disponibles et beaucoup d'endroits sont couverts. De plus, la documentation technique disponible est très complète, et toutes les spécifications sont disponibles sur le site.

1.6.2 CASI, CASI-2

CASI est un instrument hyperspectral commercial assez répandu et largement utilisé. C'est l'appareil commercial leader au niveau de l'acquisition d'images hyperspectrales dans le domaine du visible [85].

1.6.3 MIVIS

Instrument principal de la société Sensytech, MIVIS ne possèdant que 20 bandes dans le domaine du visible. Néanmoins, ses quatre spectromètres, couvrant le domaine spectral du visible à l'infrarouge thermique, peuvent fournir des données utiles pour beaucoup d'applications. Cependant, le bruit est très irrégulier en fonction de différentes bandes [86].

1.6.4 HYMAP

Considéré comme un bon modèle pour les futurs capteurs hyperspectraux, cet instrument fournit un spectre complet de 450 à 2480 nm sauf dans les bandes d'absorption de l'eau autour de 1400 et 1900 nm. Il y a environ 5 exemplaires en utilisation actuellement. Certaines différences existent entre les versions (bandes TIR sur certaines versions). L'instrument peut être adapte en fonction des besoins des utilisateurs. HyVista, un opérateur, possède un de ces instruments, et se charge de vendre les images [87, 88].

1.6.5 HYPERION

Premier capteur à transmettre des données hyperspectrales depuis l'espace, Hypérion a été envoyé pour démontrer la faisabilité des applications spatiales hyperspectrales. Le SNR n'est pas très bon comparé avec des instruments aéroportés (AVIRIS par exemple) [89, 90].

1.6.6 CHRIS-PROBA

Ce Capteur satellitaire embarqué à bord de Proba, lance en octobre 2001 et opérationnel depuis avril 2002, il possède différents modes d'acquisition de données, favorisant soit la résolution spatiale, soit la résolution spectrale. Ce compromis est nécessaire à cause de la capacité limitée de transmission. L'objectif de cet instrument est de collecter des BRDF (Bidirectional Reflectance Distribution Fonction) afin de mieux comprendre les réflectances spectrales. Au niveau technologique, CHRIS sert aussi à explorer les capacités des spectro imageurs sur des petits satellites maniables [91].

1.6.7 COIS-NEMO

NEMO a des applications civiles et militaires. Il a été conçu pour l'observation des zones côtières (eaux peu profondes) [92].

1.6.8 SPECTRA

Spectra est actuellement en développement par l'ESA. Toutes les caractéristiques ne sont pas encore fixées, mais il est intéressant de voir les objectifs visés. Les données mentionnées ici ne concernent que la bande 450-2350 nm, la bande 10.3-12.3 µm n'est pas détaillée [93].

1.6.9 HYDICE

L'hydice développé par Hughes Danbury Optical Systems et opérationnel depuis 1994, il a été spécifié pour démontrer l'utilité des techniques hyperspectrales dans les domaines du renseignement, du militaire, mais aussi du civil. Sa configuration est assez similaire à celle de

CASI, mais il dispose d'une plus grande résolution spatiale (1-4 m), mais pour une fauchée plus faible (10°environ). A cause de sa nature militaire, les images ne sont pas disponibles.

1.6.10 MERIS

MERIS (Medium Résolution Imaging Spectrometer) est un capteur spatial européen de l'ESA. Ce n'est pas réellement un capteur hyperspectral, car il ne transmet que 15 bandes sélectionnées, on n'a plus la continuité spectrale. De plus, la résolution spatiale est assez faible (1200 à 300 m). Une étude exhaustive ne sera donc pas réalisée pour ce capteur.

1.7 CARACTERISTIQUES DES CAPTEURS HYPERSPECTRAUX

Le **tableau 2** suivant propose une synthèse concernant les principales caractéristiques des capteurs hyperspectraux et ces diverses applications.

Senseur (agence / compagnie)	**Nombre De bandes**	**Domaine spectral (nm)**	**Largeur de bande à mi-hauteur (nm)**	**IFOV (mrad)**	**FOV (°)**	**Période d'utilisation**	**Applications**
AAHIS / LASH (NRL, USA)	144	432-832	3	1,8	40	1994	Défense: bathymétrie, détection de mines, détection de pollution
AHI	32	7300 –11500	14 cm^{-1}	0,9x2	7	1996	Défense
AHS (INTA)	20 1 42	7 10 430 – 1030 1550 – 1750 994 – 2540 300 – 5400 8200 –12700	30 200 13 300 400	2,5	90		Applications duales, préparation mission satellite
AIRS	2000-12000	1, 2, 4, 8	16 cm^{-1}	1 à 3 (1 à 3) x (8x8pixels)			Défense
AIS (NASA/JPL)	128	800-1600 1200-2400	80 – 200	2,05	7,3	1982	Minéraux, végétation, roches
AISA Eagle Hawk LWIR L120M	340 254 42 24	400 – 970 970 – 2450 8000 – 12000	2,9 8,5 96	0,029 - 0,060° 0,054°- 0,111° 0,063°	29,9 –62,1 17,8 –35,5		Applications civiles, analyse chimique, Détection de gaz
APEX	313 –500 civiles, préparation	400 – 2500	1000 – 277	0,5	28	2008	Applications civiles,préparation mission spatiale
ARES	128 30	400 – 2500 8000– 12000	30 – 125 64 – 100	2	65	2008	Applications civiles, préparation mission spatiale
AVIRIS (JPL)	224	400 – 1060	11,5	0,80	25	1992	Luminances directionnelles des surfaces, validation des algorithmes utilisant des visées obliques
ARES	128 30	380 – 2500	9,7 – 12,0	1	30	1989	Ecologie, neige, océanographie, géologie, hydrologie, atmosphère
CASI (Itres Research)	288	430 – 870	1,8	0,3 – 2,4	40	1990	Applications générales en télédétection

CIS (Chine)	64 24 1 2	400 – 1040 2000 – 2480 3530 – 3940 10500 – 12500	10 20 410 1000	1,2 x 3,6 1,2 x 1,8 1,2 x 1,2 1,2 x 1,2	80	1994	Applications générales en télédétection, préparation d'instruments satellitaires futurs
CHRISS (SAIC)	40	430 – 860	11	0,05	10	1994	Applications générales en télédétection, surveillance de la pollution en mer, environnement, végétation
Scanner GER-63 (GER)	24 4 29 6	400 – 1000 1500 – 2000 2000 – 2500 8000– 12500	25 125 17,2 750	90	90	1986	Applications générales en télédétection, préparation d'instruments satellitaires futurs
DAIS-7915 (GER/DLR/CE)	32 8 32 1 6	400 – 1000 1500 – 1800 2000 – 2500 3000 – 5000 8000–12600	15 – 30 45 20 2µm 0,9µm	3,3	+/- 26	1994	Géologie, écologie terrestre et marine, environnement
HIRIS (LLNL/ LANL)	1 – 14 µm	$1cm^{-1}$			90	1999	Défense (détection de mines et de gaz)
HYDICE (NLR/SITAC)	206	400 – 2500	7,6 – 14,9	0,5	8,94	1994	Agriculture, forêt, environnement, gestion des ressources, cartographie, gestion des désastres
HYMAP	128	400 – 2500	30 – 125	265	90	1994	Applications civiles
HYSPEX	160 160 256	400 – 1000 1000 – 1700 1300 – 2500	4,5 6 6	0,185 0,75 0,75	17 14 14	2006	Applications générales en télédétection.
ISM (DESPA / IAS / OPS)	64	800 – 1600 1600 – 3200	12,5 25	3,3 x 11,7	40 (sélection)	1991	Géologie, nuages, glace, neige, végétation
HYSPEX	160 160 256	400 – 1000 1000 – 1700 1300 – 2500	4,5 6 6	0,185 0,75 0,75	17 14 14	2006	Applications générales en télédétection

Tableau 2. Capteurs et Caractéristiques pour images Hyperspectrales.

1.8 DEGRADATION TYPIQUES SUR LES DONNEES

Bruit radiométrique: Comme tous les capteurs, les instruments hyperspectraux sont affectés par le bruit présent dans le signal. La valeur du rapport signal sur bruit (RSB ou Signal to Noise Ratio, SNR) permet déjà de donner une première caractérisation du bruit. Les rapports, assez faibles pour les premiers instruments hyperspectraux (de l'ordre de 50 : 1 en linéaire, soit 17 dB, pour les premières versions d'AVIRIS), atteignent maintenant des valeurs plus importantes pour les capteurs aéroportés. Les premiers instruments spatiaux présentent des valeurs plus faibles. Cependant, cette valeur seule n'est pas suffisante pour caractériser le bruit. Simmons et Brower [94] ont notamment étudié la corrélation du bruit entre les différentes bandes pour les capteurs HYDICE et AVIRIS.

Déregistration spatiale entre bandes spectrales: Les images hyperspectrales sont à priori fortement comprimables. En effet, les différentes bandes spectrales sont très liées, la

corrélation spectrale est donc importante. Cette corrélation requiert toutefois que les bandes soit registrées, c'est-à-dire que les différentes bandes soit superposables spatialement. Un même pixel sur toutes les bandes correspondra alors à la même zone physique observée.

Dans le cas des images hyperspectrales, les technologies utilisées pour l'acquisition des images sont différentes: acquisition simultanée et par le même capteur des différentes bandes spectrales. Le problème de la déregistration est donc moins important. Dans les spécifications du capteur SPECTRA, on peut voir par exemple que la déregistration pour les différentes bandes d'un même spectromètre est inférieure à 0.2 pixels.

Autres dégradations dues au capteur : D'autres effets, principalement dus à la conception, peuvent affecter les instruments hyperspectraux. Le premier de ces effets est un effet appelé smile. Cet effet est provoqué par une déformation de la projection de la fente image sur la matrice CCD. L'effet au niveau de l'image se traduit par une différence d'étalonnage selon les points de l'image (**Figure 8**).

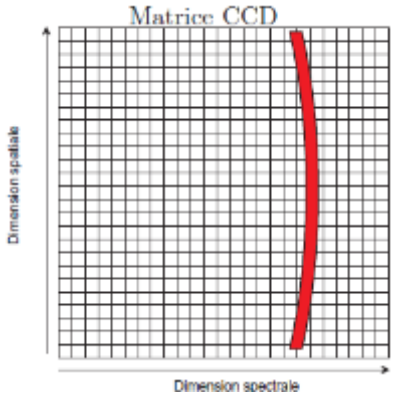


Figure 8. Effet de smile : une ligne de l'image a une projection déformée sur le CCD.

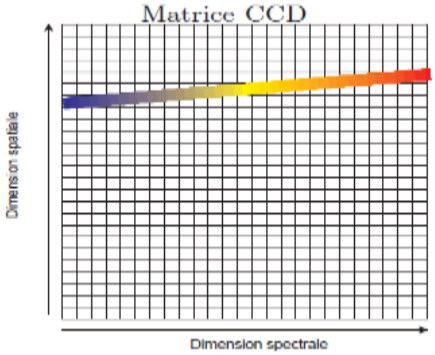


Figure 9. Effet de keystone : la projection sur le CCD des bandes spectrales d'un même pixel est oblique.

Le deuxième effet, appelé keystone, est dû à une inclinaison de l'image d'un point sur la matrice de capteur (**Figure 9**). Cet effet provoque une contamination des valeurs d'un pixel par les pixels voisins en fonction des longueurs d'onde. La dernière version du capteur aéroporté **CASI** est spécifiquement optimisée pour réduire ces distorsions.

D'autre part, les techniques de conception actuelles pour les instruments hyperspectraux éliminent quasiment les problèmes de smile et de keystone [95]. On ne prendra donc pas ces effets en compte.

Dégradations après acquisition: Les post-traitements appliqués aux données causent également des dégradations irréversibles. C'est notamment le cas de la quantification, dont l'utilisation est inhérente aux données numériques.

1.9 LES APPLICATIONS

Aujourd'hui, la richesse et la précision de l'information géo-localisée sont des éléments clés de l'aide à la décision dans de nombreuses applications. Grâce à des méthodes de traitement d'image adaptées, il est possible de mettre en évidence et d'interpréter les phénomènes observés, localement comme à grande échelle. Le résultat produit est alors une carte thématique numérique.

1.9.1 Agriculture de précision

La gestion maîtrisée des exploitations agricoles permet à la fois un rendement croissant et un impact environnemental négatif réduit. La télédétection aérienne hyperspectrale aide à la réalisation de ce compromis. Elle intervient notamment dans:

- L'estimation des récoltes
- La détection précoce des maladies
- La caractérisation des sols
- Le traitement contrôlé des parcelles

1.9.2 Foresterie

La conservation et l'entretien du patrimoine forestier représente un enjeu croissant. Dans ce domaine, l'imagerie hyperspectrale constitue un outil de choix pour les actions suivantes:

- La Détection et caractérisation du bocage et du couvert forestier
- L'entretien et conservation du patrimoine forestier
- L'identification des essences, recensement des espèces dominantes et des espèces rares.

1.9.3 Aménagement du territoire

La télédétection hyperspectrale est un outil de choix pour la mise en valeur et la gestion durable de l'environnement :

- Gestion des zones côtières (érosion, qualité des eaux de baignade, parcs aquacoles)
- Localisation et caractérisation des zones sensibles (zones inondables, zones humides, bassins versants, habitats naturels, écosystèmes)
- Cartographie des changements.

1.9.4 Qualité de l'eau

L'imagerie hyperspectrale permet l'évaluation de différents indicateurs révélateurs de la qualité de l'eau comme les matières organiques, les sédiments en suspension, la flore algale ou encore les nappes d'hydrocarbures.

1.9.5 Détection des algues

Le milieu marin est un indicateur précieux de la santé de notre environnement. L'étude des algues apporte de nombreuses informations utiles pour évaluer et corriger l'impact écologique de l'activité humaine. L'imagerie multispectrale permet d'étudier ce milieu parfois difficile d'accès en y recensant les différentes espèces d'algues. Les campagnes d'étude menées par le CEVA en 2007 et 2008 ont permis d'observer précisément la répartition de quatre espèces dominantes dans la vasière du Jaudy (Côtes d'Armor) en Bretagne.

1.9.6 Les géosciences-science de la terre solide

L'exploitation des données hyperspectrales en Géosciences – Sciences de la Terre Solide est bien maîtrisée et son potentiel a été illustré en géologie, minéralogie, pétrologie, prospection minière et pétrolière, environnements miniers, qualité et dégradation des sols, volcanisme, et en gestion de crises. L'imagerie hyperspectrale est mûre pour investir plusieurs marchés parmi lesquels le suivi des sites pollués et la protection des sols, l'érosion et la dégradation des terres, l'ouverture de nouveaux sites de prospection minière.

1.9.7 Les sciences médicales

Pour les applications médicales diagnostiques, l'imagerie Hyperspectrale fournit des moyens extrêmement résolus pour les tissus d'imagerie à un niveau soit macroscopique ou cellulaire et fournit des informations spectrales très précises concernant le patient, le prélèvement de tissu, ou l'état d'une maladie. L'image Hyperspectrale constitue un outil de choix pour les actions suivantes :

- L'Imagerie pour le diagnostic non-invasif
- La biopsie optique
- La Délimitation des tissus
- L'Analyse thérapeutiques

1.9.8 Microscopie

Pour les applications comme la classification d'absorption de drogue cellulaire ou quantifier la présence de nano-particules marquées au sein dans les échantillons de tissus. L'imagerie hyperspectrale représente une extension valable de techniques traditionnelles dans la recherche qui peuvent utiliser des microscopes optiques actuels disponibles dans le laboratoire. Avec des échantillons de recherche positionnés le long de la platine du microscope, l'imagerie spectrale produit des informations analytiques critiques avec l'ajout d'un capteur hyperspectral attaché avec un adaptateur de C-MONT de l'orifice de sortie du microscope. Avec l'étape de microscope qui déplace le secteur d'échantillon dans un "pushbroom" de manière que l'imagerie hyperspectrale donne simultanément des informations précises pour toutes les longueurs à travers la gamme spectrale complète du sensor.

1.9.9 Les applications de la surveillance et militaires

La surveillance hyperspectrale est l'exécution de la technologie de balayage pour les objectifs de surveillance. Elle est notamment utile dans la surveillance militaire à cause des mesures que les entités militaires prennent maintenant pour éviter la surveillance aéroportée. Cette dernière a été exécutée puisque les soldats ont utilisé des ballons captifs

pour espionner les troupes au cours de l'American Guerre civile, et depuis ce temps ils ont appris non seulement à se cacher de l'œil nu, mais pour masquer la signature thermique pour se fondre dans l'environnement et d'éviter ce balayage infrarouge, aussi bien. L'idée qui conduit la surveillance hyperspectrale est ce balayage qui tire l'information d'une si grande partie du spectre léger qui tout objet donné devrait avoir la signature spectrale unique dans quelques-unes des nombreuses bandes.

1.9.10 Fabrication de processus

La mise en œuvre pour la surveillance des processus critiques de la formulation et de l'inspection. Il a mis en œuvre à représenter un objet de valeur analytique pour capturer des données spectrales importantes à l'entretien. Des étapes principales fonctionnent dans une opération pour la fabrication de ce processus. Dans le champ visuel du capteur, l'imagerie hyperspectrale donne simultanément des informations précises pour tous les longueurs d'onde à travers la gamme spectrale complète disponible. Traditionnellement, les proches gammes infrarouges (NIR) de 900 à 1700 nanomètres et de l'infrarouge (VNIR prolongé) de 600 à 1600 nanomètres sont un intérêt considérable pour les applications de processus. Avec un champ visuel large, et de résolutions spectrales et spatiales optimisés. L'imagerie hyperspectrale constitue un outil de choix pour les actions suivantes:

- LCD Contrôle de la qualité
- Les opérations des semi-conducteurs
- Pharmaceutiques
- Photovoltaïque
- Inspection des Wafers

1.9.11 Fabrication pharmaceutique

L'utilisation de l'imagerie hyperspectrale n'expédie pas seulement sur le processus de la découverte des médicaments, mais présente des avantages clairs et distincts pour les fabrications de produits pharmaceutiques. Un avantage évident de la représentation hyperspectrale est la capacité de numériser des séries multiples de comprimés simultanément pendant que ces derniers se déplacent à travers une chaîne de fabrication. Le nombre de comprimés qui peuvent être balayés simultanément est basé sur le champ visuel exigé (FOV) et résolution spectrale/spatiale exigée (IFOV, champ visuel instantané). Ces paramètres sont spécifiques à l'application et peuvent être modélisée pour configurer les composants de capteurs spécifiques et d'atteindre les performances de débit d'imagerie.

1.9.12 L'urbain

L'imagerie hyperspectrale dispose d'un potentiel intéressant pour étudier les objets qui composent l'espace urbain et les zones vertes. Les méthodologies de traitement des images en milieu urbain ne sont pas harmonisées à ce jour. On note un besoin fort en résolution spatiale (meilleure que 5m) et un intérêt pour l'infrarouge thermique (un canal unique pour la température et un mode hyperspectral pour la qualité de l'air). La complémentarité d'informations multispectrales à haute résolution spatiale et hyperspectrales offre des capacités intéressantes actuellement pour l'étude de ces milieux.

1.9.13 Remote sensing- Télédétection

C'est également connu sous le nom de détection chimique, offre aux chercheurs et aux biologistes de réaliser l'analyse spectrale aéroportée et stationnaire pour des applications de télédétection. La représentation hyperspectrale aéroportée représente une technique de télédétection établie pour capturer des données spectrales importantes critiques aux applications de télédétection. Dans le champ d'exploration du détecteur, elle se produit simultanément des informations précises pour toutes les longueurs d'ondes à travers la gamme spectrale complète disponible. Avec la création du cube hyperspectral, un ensemble de données qui inclut toute les informations spatiales et spectrales, les chercheurs peuvent produire et analyser des données spectrales en profondeur écologiques. Dans ce domaine, l'imagerie hyperspectrale constitue un outil de choix pour les actions suivantes:

- Le Génie civil et environnemental
- La Surveillance de l'environnement
- La Détection de la pollution
- La Gestion forestière
- L'Exploration minière

1.9.14 Recherches spatiales et capteurs satellitaires

Pour les applications spatiales, la famille de capteurs d'imagerie hyperspectrale offrent des avantages significatifs et uniques pour les chercheurs. Conçu pour l'imagerie dans les environnements durs, les capteurs d'imagerie sont personnalisés pour avoir de qualité supérieure par l'utilisation de l'optique du rendement extrêmement élevé en rapportant la résolution spectrale et spatiale exceptionnelle, tout en comportant une image très grande pour un champ d'exploration très large.

- La Surveillance de l'environnement
- La télédétection
- Les systèmes de petits satellites

1.9.15 Fabrication pharmaceutique

L'utilisation de l'imagerie hyperspectrale n'expédie pas seulement sur le processus de la découverte des médicaments, mais présente des avantages clairs et distincts pour les fabrications de produits pharmaceutiques. Un avantage évident de la représentation hyperspectrale est la capacité de numériser des séries multiples de comprimés simultanément pendant que ces derniers se déplacent à travers une chaîne de fabrication. Le nombre de comprimés qui peuvent être balayés simultanément est basé sur le champ visuel exigé (FOV) et résolution spectrale/spatiale exigée (IFOV, champ visuel instantané). Ces paramètres sont spécifiques à l'application et peuvent être modélisée pour configurer les composants de capteurs spécifiques et d'atteindre les performances de débit d'imagerie.

1.9.16 Antiquités, œuvres d'art et la validation du document

La représentation hyperspectrale a été employée pour indiquer les secrets des documents célèbres tels que les manuscrits de la mer morte aussi bien que les objets façonnés archéologiques tels que les tessons de poterie (ostracons), qui sont la plus ancienne représentation connue de l'écriture hébraïque. Cette technique spectrale non destructive est inestimable pour un large éventail de recherche historique et d'analyse de la science

légale. La représentation hyperspectrale rapporte simultanément l'information précise pour toutes les longueurs d'onde à travers le domaine spectral du capteur. Avec la création du cube de données hyperspectrales, un ensemble de données qui comprend toutes les informations spatiale et spectrale dans le champ de vision, les équipes de recherche sont en mesure d'évaluer plus rigoureusement les documents et autres preuves scène de crime qui augmenteront considérablement la connaissance de composition et de l'unicité spectrales de ces échantillons. Elle intervient dans :

- La recherche archéologique
- Examen de documents et de vérifications
- Analyse artefact
- Inspection des peintures et des œuvres d'art

1.9.17 L'atmosphère

Des axes originaux de recherche en imagerie hyperspectrale ont été proposés bien qu'il n'apparaisse pas de besoin fort et clairement identifié d'une mission spatiale pour ce thème, l'analyse de l'atmosphère étant principalement un support pour la correction des images brutes. Ces axes visent l'exploitation de paramètres atmosphériques déduits des images pour la détection de phénomènes de surface (feux, volcans, émissions de méthane) et pour l'estimation de la pollution dans la couche limite (en particulier les espèces NO_2 et NO). On note un intérêt fort de cette communauté pour l'infrarouge thermique.

1.9.18 Le Traitement du Signal et de l'Image

Le traitement du signal et de l'image est suffisamment mûr pour répondre aux besoins des diverses applications thématiques. Il se décompose principalement en une étape de prétraitements, une étape d'extraction des propriétés optiques et une étape de traitement numérique des images (classification, détection, etc.). A cela peut éventuellement s'ajouter une étape de compression du signal à bord. La communauté nationale dispose d'une compétence forte dans ces domaines et se distingue en particulier par une prise en compte approfondie de la physique de l'image.

1.9.19 La défense

La défense a identifié que l'imagerie hyperspectrale pouvait contribuer à répondre aux besoins prioritaires et aux programmes en cours suivant 5 axes d'effort : ciblage, vulnérabilité au leur rage et au camouflage des systèmes d'observation (VULNOBS), détection et caractérisation des effluents, traficabilité, caractérisation des petits fonds complexes. Un groupe de travail (GTHS Défense) a été créé en 2004 par l'EMA afin de coordonner de nombreuses actions (une ETO et un PEA en particulier). Les travaux de la défense ont démontré l'apport de l'imagerie hyperspectrale pour deux missions satellitaires (détection de bâtiments fixes sous couvert végétal partiel et détermination du type de bâtiments pour l'aide au ciblage ; caractérisation des sols pour l'aptitude au mouvement des forces). La détection et la caractérisation à petite échelle spatiale de bâtiments et de voies d'accès sont réalisables à l'aide d'un instrument hyperspectral dont les bandes sont sélectionnées judicieusement.

1.10 PROBLEMATIQUE DE LA THESE

L'exploitation des données de grandes dimensions comme l'imagerie hyperspectrale pour l'aide à la décision va connaître dans la prochaine décennie un développement spectaculaire, grâce à l'évolution rapide des moyens technologiques et aux besoins croissants suscités par de nombreux domaines applicatifs.

L'imagerie hyperspectrale est une technologie permettant la représentation d'une même scène suivant un grand nombre de bandes spectrales étroites paramétrables dans de gammes de longueur d'ondes spécifiques. La génération actuelle de capteurs d'imagerie hyperspectrale présente un intérêt scientifique, technique et économique indéniable, car elle fournit de grandes quantités d'information précise sur la nature et l'évolution spatio-temporelle des zones à analyser.

Cette technologie mature et accessible est devenue essentielle pour apporter des éléments objectifs d'analyse et d'expertise, et donc des solutions rapides et appropriées dans les domaines de l'environnement, de l'agronomie, de l'océanographie, de la sécurité industrielle et sanitaire, des transports, de l'urbanisme, etc. En télédétection, la technologie sophistiquée des imageurs, embarqués sur un vecteur aérien, permet de produire instantanément et à distance une grande quantité d'information permettant d'accéder à une mesure de signature spectrale des matériaux imagés pour une meilleure discrimination. Par exemple, embarquée sur un avion, une caméra hyperspectrale fonctionnant dans le domaine visible-proche infrarouge (400-1000 nm) peut enregistrer simultanément plusieurs dizaines, voire centaines de bandes spectrales de faible largeur, et accéder ainsi à une mesure échantillonnée de la luminance spectrale des matériaux pour leur discrimination.

Les volumes de données (cubes de données) acquises atteignent souvent le gigaoctet pour une seule et même scène observée. De ce fait, l'analyse de ces données au contenu physique complexe passe obligatoirement par une étape préliminaire de réduction de la dimensionnalité. Cette réduction a un double objectif, le premier consiste à réduire la redondance et le second permet de faciliter les traitements postérieurs (extraction, classification et reconnaissance de formes) et donc l'interprétation des données.

Plusieurs approches existent, comme celles basées sur la projection (linéaire ou non-linéaire) des données de grandes dimensions sur des sous-espaces de représentation bien choisis (Analyse en Composantes Principales, Minimum Noise Fraction, Analyse en Composantes Curvilignes, Analyse Canonique, etc.), sur la séparation de sources (Analyse en Composantes Indépendantes) ou sur les techniques de sélection de bandes spectrales exploitant des critères de complémentarité-redondance d'information (entropie, contraste, corrélation, information mutuelle) qui ne permettent pas de préserver toute la richesse de l'information apportée par ce type de données.

La préservation de la richesse de l'information apportée par l'imagerie multi-composante, pose donc des difficultés. La résolution optimale d'un tel problème passe par la préservation des détails même rares apportés par certaines bandes spectrales et par la réduction maximale de toute redondance.

Le problème posé consiste mathématiquement à déterminer un sous espace réduit à partir d'un ensemble de données X de la forme $\overrightarrow{x_i}$ ={$x_{1,i}$, $x_{2,i}$, …, $x_{m,i}$}^{T} i=1,…n. composé de n individus de dimensions m. Chaque individu est représenté par un vecteur $\overrightarrow{x_i}$ dans l'espace de m dimensions. Chaque vecteur est la réalisation de m variables aléatoires (X_1, X_2,…,X_n). L'ensemble des données est représenté sous la forme d'une matrice X : m x n contenant en ligne les dimensions et en colonne les individus : X = {$x_{j,i}$;j=1,..,m i=1,..,n} où n est le nombre d'individus et m est le nombre de dimensions de chaque élément.

Il s'agit de trouver une autre représentation de dimension réduite $\overrightarrow{y_i}$= {$y_{1,i}$, $y_{2,i}$, …, $y_{m,i}$}^{T} avec r $\leq$ m, qui exprime la même information que les données originales (**Figure 10**).

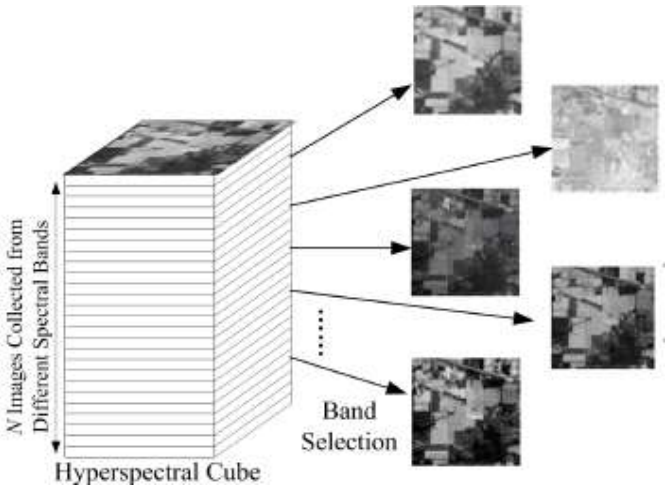


Figure 10. Principe de la sélection de variables ou bandes.

$$\begin{bmatrix} P_{11} & P_{12} \dots . & P_{1m} \\ \vdots & \vdots & \vdots \\ P_{n1} & P_{n2} \dots . & P_{nm} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \begin{bmatrix} PS_{11} & PS_{12} \dots . & PS_{1r} \\ \vdots & \vdots & \vdots \\ PS_{n1} & PS_{n2} \dots . & PS_{nr} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

n pixels x m bandes n pixels x r bandes (r <m bandes)

Pour rendre facilement exploitables ces données et répondre rapidement aux besoins suscités par les différentes applications potentielles, pour réaliser le meilleur compromis entre la réduction et la préservation de la quantité d'information acquise deux approches générales des méthodes non paramétriques, non supervisées de réduction de la dimensionnalité des images hyperspectrales sont proposées.

- La sélection de bandes spectrales: Consiste à choisir des bandes spectrales les plus représentatives dans l'espace de mesure.
- Méthodes de projection : Vise à sélectionner des caractéristiques d'un espace transformé à un espace de projection.

1.11 PLAN DU MANUSCRIT

Ce mémoire comprend deux parties. La première partie (chapitres 1 à 3) présente un état de l'art des méthodes de réduction de la dimension par projection et par sélection des bandes spectrales ainsi que les critères répondant aux points de cahier de charges. La

deuxième partie (chapitres 4 et 5), décrit les contributions apportées par cette thèse et expose divers cadres expérimentaux proposés afin de valider ces idées.

Le premier chapitre présente les prérequis pour comprendre le contexte de la réduction de la dimensionnalité des images hyperspectrales.

Le second chapitre, présente les deux catégories des méthodes de la réduction (Méthodes de projection et méthodes de sélection de bandes spectrales), leurs fondements théoriques, leurs caractéristiques principales et leurs principales limitations.

Dans le troisième chapitre, on recense des critères de qualités adaptées aux images hyperspectrales. Une démarche est définie pour déterminer les propriétés de ces critères en termes de sensibilité spécifique à un type de dégradations. Enfin, trois critères sont finalement conservés pour leur bonne complémentarité sur la stabilité, la similarité de l'algorithme, la préservation de la quantité d'informations et la minimisation des redondances.

Concernant le quatrième chapitre, on décrit les études comparatives faites sur la stabilité, la similarité et sur la perte d'informations des algorithmes de réduction de la dimension, testées sur des images hyperspectrales artificielles et naturelles.

L'avant dernier chapitre, présente une approche hybride pour la réduction des données hyperspectrales. Il se termine par une optimisation de cet algorithme.

Le chapitre 6 présente les conclusions de ce travail et quelques perspectives.

CHAPITRE 2

ÉTAT DE L'ART SUR LA RÉDUCTION DE LA DIMMENSIONALITÉ

2.1 INTRODUCTION

Généralement, La taille des données peut être mesurée selon deux dimensions : le nombre de variables et le nombre d'exemples. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors de l'exploration et l'analyse de ces données. Pour cela, il est fondamental de mettre en place des outils de traitement de données permettant une meilleure compréhension de la valeur des connaissances disponibles dans ces données. La réduction des dimensions est l'une des plus vieilles approches permettant d'apporter des éléments de réponse à ce problème. Son objectif est de sélectionner ou d'extraire un sous-ensemble optimal de caractéristiques pertinentes selon un critère fixé auparavant. Cette sélection/extraction permet donc de réduire la dimension de l'espace des exemples et de rendre l'ensemble des données plus représentatif du problème. Cette réduction a un double objectif, le premier consiste à réduire la redondance et le second permet de faciliter les traitements postérieurs (extraction, classification et reconnaissance de formes, réduire l'espace de stockage nécessaire, réduire le temps d'apprentissage et d'utilisation, identifier les facteurs pertinents,…) et donc l'interprétation des données.

Deux approches sont utilisées pour classer les techniques mathématiques non- supervisée et non-paramétriques pour la réduction de la dimensionnalité : celles basées sur la projection (linéaire ou non-linéaire) des données de grandes dimensions sur des sous-espaces de représentation bien choisis, celles basées sur les techniques de sélection de bandes spectrales exploitant des critères de complémentarité-redondance d'information (Entropie, contraste, corrélation, information mutuelle) qui consiste à choisir des bandes spectrales les plus représentatives dans l'espace de mesure.

Les méthodes utilisées, dans chacune de ces deux classes, pour l'extraction de caractéristiques sont très variées. Nous rappellerons brièvement les principes des méthodes de projection linéaires/non linéaires, puis nous décrirons les méthodes de sélection de bandes spectrales qui ont fait l'objet de nombreuses études depuis cinq ans.

2.2 APPROCHES PAR PROJECTION LINEAIRE

Les techniques de projection linéaire des données en sous espace de dimension inférieure est une étape préliminaire dans la réduction de dimension. Elle permet de réduire la dimension de l'espace originale et rendre l'ensemble des données plus représentatif du problème. La nouvelle représentation des données est une combinaison linéaire des données originales, c'est-à-dire ;

$$y_{ki} = w_{k1} \,.\, x_{1i} + w_{k2} \,.\, x_{2i} + \cdots + w_{km} \,.\, x_{mi} \quad \forall\, K \in \{1,2,..,r\}\,,\ \forall\, i \in \{1,2,..,n\} \qquad (3)$$

Ou en notation matricielle : Y = Ω X avec

$$X = \{ x_{ij}, 1 \leq j \leq m \; 1 \leq i \leq n \}, Y = \{y_{ki}, 1 \leq k \leq r, 1 \leq i \leq n \} \quad (4)$$

et $\Omega = \{w_{kj}, 1 \leq k \leq r, 1 \leq j \leq m\}$ la matrice de transformation linéaire.

Plusieurs méthodes de projection linéaire existent dans la littérature, et sont introduits dans les paragraphes ci-dessous.

2.2.1 Analyse en Composante Principale(ACP)

L'analyse en composantes principales [1, 27] est une méthode projective non supervisée, dont le critère à maximiser est la variance dans les données projetées. On fait l'hypothèse que les données de départ se trouvent dans un hyperplan et que l'on peut les exprimer au moyen des vecteurs qui définissent cet hyperplan. Si l'hypothèse se révèle vraie, on trouve la dimensionnalité intrinsèque des données. Les vecteurs centrés de départ $X \subset R^m$ sont projetés vers le sous espace dimensionnel $Y \subset R^r$, $r < m$ tel que $Y = XW$, avec $W \subset R^{m \times r}$.

$$E[XW] = E[X]\,W = \mu W \quad (5)$$

$$Var(XW) = E[(XW - \mu W)^2] = E[W'(X-W)'(X-\mu W)W] = W'\Sigma W \quad (6)$$

La composante principale de la projection est W_1. Comme la direction de ce vecteur est l'information la plus pertinente, on considère que $\|W_1\| = 1$

$$Y_1 = X W_1 \text{ , selon (2) } Var(Y_1) = W'\Sigma W_1 \quad (7)$$

On cherche W_1 tel que $Var(Y_1)$ soit maximisée en respectant la contrainte $W_1' W_1 = 1$. Le lagrangien est: $L(\alpha) = W_1'\Sigma W_1 - \alpha (W_1' W_1 - 1)$. Ainsi, en dérivant par rapport à W_1 on trouve $\Sigma W_1 - \alpha W_1 = 0$.

La relation $\Sigma W_1 - \alpha W_1 = 0$ est valable si W_1 est un vecteur propre de Σ et α est la valeur propre correspondante. On a alors : $\alpha \; W_1' W_1 = \alpha$

On choisit le vecteur propre avec la valeur propre la plus grande pour que la variance soit maximale. La deuxième composante principale doit aussi maximiser la variance, posséder une norme unitaire et, en plus, être orthonormé à W_1 .

$$\text{Alors, Max } L(\alpha) = W_2'\Sigma W_2 - \alpha (W_2' W_2 - 1) - \beta (W_1' W_1 - 1) \quad (8)$$

En dérivant par rapport à W_1 et égalisant à 0:

$$2 \Sigma W_2 - \alpha W_2 - \beta W_1 = 0 \quad (9)$$

Si l'on multiplie l'équation (9) par W_1' on s'aperçoit que $\beta = 0$. On a : $\Sigma W_2 = \alpha W_2$. Alors, W_2 est aussi un vecteur de Σ dont la valeur propre est la 2me plus grande. La $r^{ième}$ composante principale peut être dérivée de la même manière. Σ est symétrique, singulière, son rang (la dimensionnalité intrinsèque des données) est r avec $r < m$. On centre les données de départ avant la projection en faible dimension, alors : $Y = (X - \mu) / W$.

Après cette transformation linéaire, nous avons un espace r-dimensionnel dont les dimensions sont les vecteurs de Σ et les variances sur ces dimensions sont leurs valeurs propres associées. Les vecteurs W_i, i=1,…r définissent une base orthonormée pour l'espace de départ. Si l'on regarde le spectre des valeurs propres, on obtient une indication de la dimensionnalité intrinsèque des données. Les directions qui correspondent aux valeurs propres petites sont normalement interprétées comme du bruit, bien que pour certaines tâches de classification, elles puissent devenir utiles. L'analyse en composante principale

(**ACP**) permettant de supprimer les redondantes mais ne préserve pas toutes les informations complémentaires pendant la projection en particulier l'évènement rare.

2.2.2 Locality Preserving Projections (LPP) : Projections Préservées Localement

Un autre algorithme linéaire dans la réduction de dimension est le LPP [2,45]. Il construit un graphe de voisinage de l'ensemble de données en utilisant la notion du graphe de Laplacian, ensuite une matrice de transformation qui fait correspondre les points de données à un sous espace. Il est relativement insensible aux valeurs aberrantes et aux bruits.

L'algorithme de LPP est formellement indiqué ci-dessous

a. Construire le graphe adjacent G avec k nœuds des arrêts i et j si x_i et x_j sont voisins. Il existe deux variantes:

b. Choisir le poids: nous avons deux variantes pour la pondération

- Noyau de chaleur avec le paramètre t ϵ R. Si les sommets i et j sont reliés,

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (10)$$

- Noyau de chaleur avec le paramètre t = 1. Si les sommets i et j sont reliés, W_{ij}= 1.

c. Calculer les vecteurs et valeurs propres pour le problème généralisé

$$XL\, X^T a = \lambda\, XD\, X^T\, a \quad (11)$$

Où D est une matrice diagonale des poids dont les éléments sont la somme de chaque ligne de W, $D_{ii}=\sum_j W_{ij}$, L est la matrice du Laplacian L = D - W. L est une matrice symétrique semi définie positive, avec $a_0, a_1, \ldots, a_{k-1}$ sont les solutions de l'équation (11) qui correspond à des valeurs $\lambda_0 < \lambda_1 < \ldots < \lambda_{k-1}$, $x_i \rightarrow y_i = A^T x_i$, A= $(a_0, a_1, \ldots, a_{m-1})$, y_i est un vecteur à m dimensions dans l'espace réduit.

2.2.3 Positionnement Multidimensionnel (MDS)

La méthode de positionnement multidimensionnel –" Multidimensionnel Scaling" (MDS) – [3] permet de construire une autre réduction linéaire. L'algorithme MDS est basé sur une recherche de valeurs propres. MDS permet de construire une configuration de n points dans R^m à partir des distances entre n objets [25]. L'exemple classique est d'obtenir la carte d'un pays en partant de la connaissance des distances entre chaque paire de villes. MDS supprime les redondantes. Par contre, cette méthode est très sensible à la variation de paramètres de données et à la dégradation des bruits [4] selon des tests sur deux bases de données naturelles et artificielles. Et ne préserve pas l'évènement rare [3].

Algorithme

On observe n(n - 1)/2 distances. Il est toujours possible de générer un positionnement de n points en m dimensions qui respecte exactement les distances fournies. MDS calcule une approximation en dimension r < m. L'algorithme est le suivant:

i. Moyennes des distances carrées par rangées : $\mu = \frac{1}{N}\sum_j d_{ij}$

ii. Double centrage (distance carrée vers produit scalaire) :

$$P_{ij} = -\frac{1}{2}\left(d_{ij} - \mu_i - \mu_j + \sum_i \mu_i\right) \quad (12)$$

iii. Calcul des vecteurs propres v_j et valeurs propres λ_j principales de la matrice P (avec les λ_j les plus grands),

iv. La $i^{ème}$ coordonnée réduite de l'exemple j est: $\sqrt{\lambda_j}$

2.3 APPROCHE PAR PROJECTION NON LINÉAIRE

Dans de nombreuses applications, la distance euclidienne n'a pas de grand sens ; elle suppose en particulier que toutes les variables sont comparables entre elles (elles doivent donc avoir été convenablement normalisées). La théorie des espaces de Hilbert permet de définir d'autres produits scalaires, basés sur des fonctions noyaux K(x, y). K est alors une mesure de similarité entre les points de l'ensemble à traiter. Le noyau K défini implicitement une application de l'espace d'origine vers un "espace de caractéristiques" H. La dimension de l'espace H est éventuellement infinie. De nombreuses méthodes statistiques peuvent s'exprimer en ne recourant qu'à des produits scalaires entre les points à traiter et les exemples d'apprentissage. Si l'on remplace le produit scalaire habituel par un noyau K, on rend la méthode non-linéaire ; c'est le truc du noyau –" kernel trick "-, qui a fait l'objet de nombreuses recherches depuis son introduction par "Vapnik "dans le cadre des machines à vecteurs de support (SVM). Parfois les données de départ appartiennent à des sous-espaces non linéaires. Les méthodes non-linéaires d'apprentissage de variétés récemment développées sont alors un outil intéressant pour analyser les données de haute dimensionnalité.

Le principe général de ces méthodes est de considérer l'ensemble des données de départ comme des échantillons d'une variété de Riemann. A partir de là, est construite une solution approchée de la métrique de la variété qui est utilisée pour éliminer les dimensions peu significatives de variabilité de l'ensemble. Le but reste d'obtenir les structures de faible dimension existant dans les espaces de haute dimensionnalité.

2.3.1 ISOMAP (Cartographie des Caractéristiques Isométriques)

La méthode de réduction de dimensionnalité Isomap (Isometric Feature mapping) a été introduite par Tenenbaum [5] et s'inspire de la méthode "Multidimensional Scaling" (MDS) pour la connaissance d'une matrice de dissimilarité entre les paires d'individus mais en lui donnant comme métrique la distance géodésique (ou curviligne).

Le but est cette fois de trouver une variété (non linéaire) contenant les données. La méthode Isomap estime la distance géodésique [6, 7] de la façon suivante : Dans un premier temps, le voisinage de chacun des points est calculé. Une fois les voisinages connus, un graphe est construit en reliant tous les points voisins. Chaque arrêt du graphe est ensuite pondéré par la distance euclidienne [8] entre les deux extrémités de l'arrête. Enfin, la distance géodésique entre deux points est estimée par la somme des longueurs des arêtes le long du plus court chemin entre ces points. En pratique, le plus court chemin entre deux sommets du graphe est calculé en utilisant l'algorithme de Dijkstra.

Algorithme

D'un point de vue algorithmique, l'Isomap projette une surface d'un espace à m dimensions, représentée par n points, de la façon suivante :

i. Recherche des vecteurs de l'ensemble X = { x_i,..., x_n} $\subset R^m$ voisins. D'après les distances euclidiennes d_{ij} on détermine un voisinage pour chaque point x_i, soit

avec le critère des k-plus proches voisins soit en considérant tous les points à l'intérieur d'une sphère de rayon α centré sur x_i. On considère que les distances euclidiennes approchent les géodésiques quand les vecteurs se trouvent à petite distance.

ii. On fait une estimation des distances géodésiques d(G_{ij}) entre tous les points x_i . Isomap construit un graphe dont les sommets sont les points et les arêtes sont les distances entre eux.

iii. On applique l'algorithme de Dijkstra afin d'obtenir la matrice des distances D; elle contient les distances géodésiques entre chaque point et ses voisins. Un sommet est adjacent à un autre seulement s'ils ont été définis comme voisins (Étape 1). La distance géodésique est estimée entre chaque paire de données par la distance la plus courte parcourue sur le graphe (algorithme de Floyd ou Djikstra).

iv. Finalement, la méthode MDS est appliquée à la matrice de distances (dG_{ij}), i, j =1,…, n pour obtenir un nouveau système de coordonnées euclidiennes Y ⊂ R^r, (r < m) qui préserve la géométrie intrinsèque de la variété. Parfois, des valeurs propres de petite magnitude sont obtenues (et ignorées) à la sortie du MDS : elles sont dues à des inconsistances mineures du calcul des distances géodésiques et/ou au bruit dans X.

2.3.2 Coordination Localement Linéaire (LLC)

La coordination localement linéaire LLC calcule un certain nombre de modèles linéaires et effectue par la suite un alignement global [9]. Ce processus se compose de deux étapes:

i. Mélange des modèles linéaires locales sur les données au moyen de l'algorithme espérance-maximisation (expectation maximisation EM en anglais).

ii. En alignant les modèles linéaires locales afin d'obtenir une représentation des données dans l'espace réduit en utilisant une variante de LLE.

LLC construit m mélange d'analyse factorielle en utilisant l'algorithme EM [37, 38, 39]. Le modèle linéaire locale et leurs correspondantes {r_{ij}, j=1,.., m} pour chaque point de données x_i, r_{ij} décrit dans quelle mesure les points des données x_i correspond au modèle j en satisfaisant $\sum_{j=1}^{m} r_{ij}$=1. Ensuite la responsabilité pondérée des représentations de données est calculée par u_{ij}= r_{ij}. z_{ij}, stockées dans D bloc d'une matrice U de dimension nxm. Cet alignement des modèles locaux est basé sur deux matrices U et M donné par M = $(I-W)^T(I-W)$, ici la matrice W contient les poids reconstruites par LLE, I la matrice d'identité n x n.
LLC aligne les modèles locaux en résolvant ; A v = λB v. La représentation des données dans la faible dimension est ainsi obtenu par : Y = U L, L est la matrice formée par les d vecteurs v_i, i=1,…, n.

2.3.3 Manifold Charting : Cartographies Multiples

Cette méthode est similaire à LLC. Il s'agit de construire une représentation des données à faible dimension en alignant les 2 modèles MoFA (Mixture Of Factor Analysis) ou MoPPCA (Mixture Of Probabilistic PCA) [10]. Cette méthode minimise une fonction de coût convexe, qui mesure la quantité de désaccord entre les modèles linéaires sur l'ensemble des données de coordonnées globale :

$$\varphi(Y)= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{m} r_{ij} r_{ik} \left\| y_{ij} - y_{ik} \right\|^2$$

La minimisation de cette fonction de coût s'effectue par la résolution d'un problème généralisé. La méthode «Manifold Charting» applique tout d'abord l'algorithme EM pour obtenie les facteurs d'analyses, afin de conclure les représentations des données z_{ij} et ces correspondants r_{ij} où j=1,…,m et i=1,…n. Enfin, on trouve une cartographie linéaire de z_{ij} de coordonnées globale y_i qui minimise la fonction de coût.

2.3.4 Diffusion Maps : Cartes de Diffusion

Les cartes de diffusion (DM) est une méthode qui s'appuie sur l'hybridation entre un modèle de diffusion de la chaleur et un processus stochastique (chaine de Markov) [11, 40]. Soient

$$X = \{x^{(1)}, x^{(2)},\dots, x^{(k)}\} \in R^m,\ x^{(i)}=\{ x_1^{(i)}, x_2^{(i)},\dots, x_m^{(i)}\}\ i=1,\dots,k$$

$$Y= \{y^{(1)}, y^{(2)},\dots, y^{(k)}\} \in R^r\ ,\ y^{(i)}=\{ y_1^{(i)}, y_2^{(i)},\dots, y_r^{(i)}\},\ j =1,\dots,k$$

Avec $r \ll m$, $\{x^{(1)}, x^{(2)},\dots, x^{(k)}\} \in M$, ou M une variété dans R^m, un graphe pondéré a été reconstruit de ces ensembles de donnée pondéré, les nœuds sont les points de données et les arrêts pondérés reliant les nœuds. Une représentation matricielle d'un graphe comme la matrice de laplacien. La procédure est décrite ci-dessous en quatre étapes.

Algorithme

i. Construire la matrice d'adjacence W du graphe dont les poids W_{ij} reliant les nœuds correspondants i et j par le noyau : W_{ij}= exp (- $\frac{\|x_i-x_j\|^2}{\varepsilon}$), $\varepsilon = 2\sigma^2$
   Construire la matrice normalisée diagonale k x k et la matrice Laplacien respectivement par $D_{ii} = \sum_{j=1}^{k} W_{ij}$ et L= D-W. Notez également que L est symétrique et semi définie positive.
ii. Calculer les valeurs propres et vecteurs propres du problème généralisé :

$$L\,y = \lambda\, D\, y, \quad (13)$$

Les vecteurs propres $y_0, y_1,\dots, y_{k-1}$ correspondants à des valeurs propres $\lambda_0, \lambda_1,\dots,\lambda_{k-1}$ sont solutions de l'équation (13).

2.3.5 Hessian-based LLE : Hessien basée sur LLE

Grimes et Donoho [13] proposent une amélioration de la méthode LLE se basant sur le fait de considérer la surface Σ, vue comme une surface de Riemann dans l'espace euclidien R^m, comme localement isométrique à un ouvert θ de l'espace euclidien R^r. Le cadre théorique de cette méthode repose sur la forme quadratique:

$$H(f)= \int \left\|H_f(\sigma)\right\|_f^2 d\sigma \quad (14)$$

Où f : $\Sigma \to R$, H_f est le hessien de f et H(f) est la moyenne de la norme de Frobenius du hessien sur Σ. Le hessien est défini en utilisant les coordonnées orthogonales dans le plan tangent à σ. L'observation principale est que si σ est réellement localement isométrique à θ dans R^r alors H(f) a (r+1) dimensions nulles. H(f) est constituée de fonctions constantes et d'un sous espaces r dimensions, engendrée par les coordonnées isométriques originales. La théorie détaillée relative à la HLLE est développée dans [13] ; on y trouvera également les preuves des théorèmes de convergence. Cette technique peut donc être vue comme une amélioration de "Locally Linear Embedding" où le Laplacian Eigenmaps est modifié : la forme quadratique basée sur le Laplacian est remplacée par une forme quadratique basée sur le hessien.

2.3.6 Multilayer-Autoencoders (Auto-encodeurs Multicouches)

Le concept de réseau non supervisé est issu des observations des neurobiologistes sur la cartographie cérébrale qui s'auto-structure dans les premières années de l'être humain [14]. Le principe est globalement basé sur : lorsque l'on présente la base de données au réseau lors de l'apprentissage, un seul neurone de la couche de sortie est active par une ou plusieurs informations d'entrée (pour autant que celles-ci soient suffisamment voisines) [41]. Les neurones de sortie se spécialisent pendant l'apprentissage en reconnaissant certaines caractéristiques qui leur deviennent spécifiques et des neurones de sortie voisins reconnaissent des objets d'entrée dont les caractéristiques sont voisines. La couche de sortie est stabilisée après de multiples présentations des données d'entrée unie ou bidimensionnelles de la base étudiée [24].

On expose ci–après le principe de l'algorithme SOFM "Self Organizing Feature Map" ou carte de Kohonen. La figure 10 illustre la projection vectorielle effectuée par l'algorithme. La base de données d'entrée est représentée à gauche de la figure par n vecteurs x_k. La carte de kohonen est située à droite. En choisissant une distance euclidienne pour chacun de la base des données, on définit $d_{ij}= \sum(x_k - w_{ijk})^{1/2}$ comme étant la distance entre le site ij (représenté par ses k composantes initiales w_{ij} et le vecteur $\mathrm{x_k}$). Si le nombre d'objet $\mathrm{x_k}$ présentés à chaque site ij est n, il existe pour chacun des n objets un site ij pour lequel la distance est minimale. Chaque objet de la population se trouve ainsi affecté au site ij en question. On notera que suivant les dimensions de n et d'_{ij}, plusieurs objets peuvent se trouver dans le même site ij de la carte et que, certains sites peuvent ne recevoir aucun objet. Certaines zones de la carte de Kohonen pouvant par ailleurs rester aveugles.

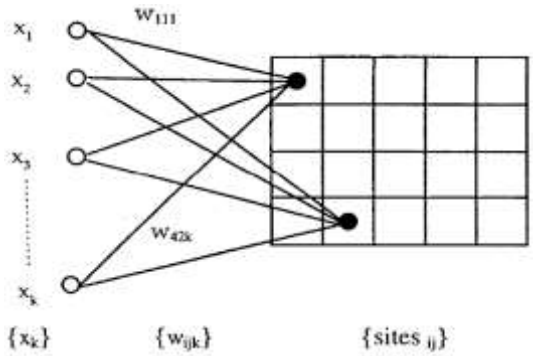


Figure 11. Illustration de la carte de kohonen

Cette première projection de la base d'entrée sur un espace à deux dimensions étant effectuée, on définit pour chaque site qui a capté un ou plusieurs objets un voisinage topologique de n sites voisins. Le voisinage pourra correspondre à un espace de un, deux ou plusieurs sites. Pour tous les sites de l'espace de sortie qui captent des vecteurs de la base des données, on modifie les vecteurs poids W_{ij} de ces sites d'une façon telle que ;

$$W_{ij(t+1)} = W_{ij} + 1_{(t)}[\mathrm{x_k} - W_{ij}] \text{ si ij appartient au voisinage}$$
$$= W_{ij(t)} \qquad \text{autrement}$$

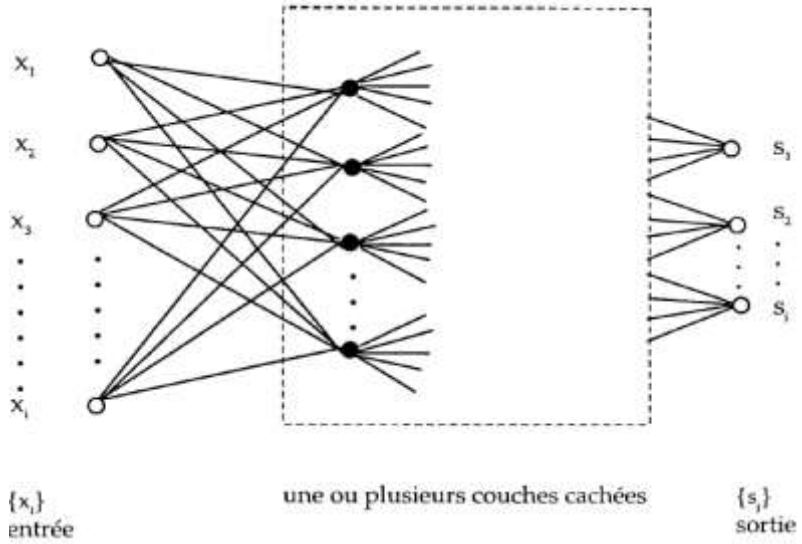


Figure 12. Architecture standard d'un réseau multicouche.

Où $1_{(t)}$ est un paramètre d'apprentissage qui varie avec le temps d'apprentissage (nombre d'itération t). Une fois le poids $w_{(ij)}$ est modifié, on présente à nouveau l'ensemble des objets de la base décrits par leurs x_k et on recommence le processus tout en réduisant progressivement la dimension du voisinage et le paramètre d'apprentissage. La projection est terminée dès que la grille est stabilisée [42].

Auto Associative Neural Network est une approche basée sur les réseaux de neurones [15, 44] pour l'extraction des caractéristiques à partir des données hyperspectrales. Sa topologie ressemble à celui d'un perceptron multicouche comprenant trois couches cachées. L'objectif est de se rapprocher autant que possible des données originales. Cette approche a été utilisée dans la compression, télédétection et dans la classification sur des données hyperspectrales. Les composantes non linéaires sont considérées comme des entrées pour le perceptron multicouche pour effectuer la classification des pixels. Les résultats montrent que la méthodologie mise en œuvre est positive à discriminer parmi 9 classes de couverture terrestre produisant une précision globale de 97% testées sur 50 000 pixels.

2.3.7 Sammon Mapping : Cartes de Sammon

Sammon mapping est un algorithme non linéaire [16], qui est lié à la mise d'échelle multidimensionnelle (MDS). Le but de l'algorithme de projeter des données issues d'un espace à r dimension en m dimensions tel que r<<m. Durant la projection, le Sammon-mapping tend à préserver les distances entre tous les points du jeu de données pris deux à deux afin de préserver la structure des données, de telle sorte que les distances correspondantes soient proches des distances au sein de l'espace d'origine. Considérons un ensemble de n objets. Chaque objet est représenté par un point dans un espace à r dimensions. d_{ij}, ∀ i, j =1,…n est la distance entre deux points dans l'espace d'origine à r dimensions. δ_{ij}, ∀ i, j =1,…n est la distance entre deux points dans l'espace d'arrivé à m dimensions.

Afin d'identifier les n points au sein de l'espace d'arrivé dont les distances deux à deux approchent au mieux celles dans l'espace d'origine, on utilise un critère de cout. Ce dernier

est généralement nommé "fonction de stress" noté S. Il mesure l'écart entre les configurations dans les deux espaces.

$$S = \frac{1}{\sum_{i=1}^{n}\sum_{i=j+1}^{n}\delta_{ij}}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{(\delta_{ij}-d_{ij})^2}{\delta_{ij}} \quad (15)$$

La principale faiblesse de cette technique est qu'elle se concentre principalement sur les grandes distances et non pas sur des petites distances deux à deux qui sont beaucoup plus importantes pour la géométrie des données. Par contre, elle a le pouvoir de minimiser la fonction d'erreur, qui est souvent appelé le stress Sammon et de conserver la structure locale des données (le poids, la distance).

2.3.8 Maximum Variance Unfolding (MVU) : Séquences sur le Maximum de Vraisemblance

MVU est un algorithme non linéaire, similaire à l'Isomap dans le sens qu'il définit un graphe de voisinage sur les données et conserve les distances [17]. Le principe de MVU est de trouver une représentation qui maximise la variance tout en préservant les distances entre points voisins, sous la contrainte que les distances restent inchangées dans le graphe de voisinage. Le problème d'optimisation se résout en utilisant le programme semi définie positive SDP ou OSDP [43].

Algorithme

i. La construction du graphe G, dans lequel chaque point de données, x_i i=1,.., n est relié à son k –plus proche voisins x_{ij}, j=1,…, k.
ii. Maximiser la somme des carrés des distances euclidiennes entre tous les points de données, sous la contrainte que les distances à l'intérieur de voisinage du graphe G sont préservés. En d'autres termes MVU effectue les opérations suivantes du problème d'optimisation

$$\sum_{ij}\|y_i - y_j\|^2 \text{ s.t } \|y_i - y_j\|^2 = \|x_i - x_j\|^2 \ \forall\ (i, j) \in G \quad (16)$$

2.3.9 Plongement Localement Linéaire (LLE)

Roweis et Saul [18] ont proposé une méthode de réduction de dimensionnalité appelée "Locally Linear Embedding" qui repose cette fois, non plus sur des méthodes neuronales, mais algébriques. Cette méthode permet de projeter une variété de vecteurs appartenant à un espace de dimension m dans un espace de dimension inférieure r en considérant que les données sont localement linéaires.

Local Linear Embedded (LLE) a été présenté en même temps qu'Isomap [19] et aborde le même problème par une voie différente. Chaque point est ici caractérisé par sa reconstruction à partir de ses plus proches voisins. LLE construit une projection vers un espace linéaire de faible dimension préservant le voisinage.

La méthode LLE considère que les données, globalement non linéaires, sont linéaires dans de petits voisinages. L'idée est de déterminer pour chaque vecteur $x_i \in E$ ses k voisins et de considérer que dans ce voisinage les données sont linéaires. Une fois les voisins de x_i

déterminés, on calcule les w_{ij} poids de reconstruction associés à chaque couple (x_i, x_j), où x_j appartient au voisinage V(x_i). Ces poids de reconstruction, qui représentent le voisinage de chaque vecteur x_i, permettront de reconstruire la topologie du voisinage dans l'espace de projection. La méthode LLE enchaîne trois étapes principales ayant les fonctions suivantes :

i. Recherche des k voisins,
ii. Calcul des poids de reconstruction w_{ij}
iii. Projection des vecteurs en minimisant la fonction de coût E(W)

Recherche des voisins : Cette étape détermine le voisinage de chaque vecteur dans lequel on va faire l'hypothèse que les données sont linéaires. On recherche simplement les vecteurs les plus proches du vecteur x_i dont on veut déterminer le voisinage, et pour ce faire on utilise simplement la distance euclidienne.

Calcul des poids de reconstruction Les voisinages de chacun des vecteurs de l'ensemble des données étant déterminés, on doit maintenant définir la fonction de coût qu'il faudra minimiser pour déterminer la matrice w poids de reconstruction. La fonction de coût E(W) est définie par :

$$E\ (w)= \sum_{i=1}^{n}\left|x_i - \sum_{j\in V(x_i)} w_{ij}x_i\right|^2 \qquad (17)$$

La minimisation de E(w) se fait sous deux contraintes :

i. chaque vecteur x_i est reconstruit uniquement grâce à ses voisins, i.e. w_{ij}= 0 si x_j n'est pas voisin de x_i
ii. $\forall$ i $\sum_{j\in V(x_i)} x_i w_{ij}$=1

Avec les contraintes ci-dessus, les poids de reconstruction peuvent être calculés de façon exacte. Soit x_i un vecteur particulier de la surface à déplier dans l'espace E; soit (ξ_1 ,..., ξ_k) les plus proches voisins de x_i et w_{ij} les poids de reconstruction associés. Le principe est donc de reconstruire linéairement le voisinage de x_i grâce à ses voisins dans l'espace initial. Pour chaque vecteur x_i

$$E_i(w) = \left|\sum_{j=1}^{k} w_{ij}(\ x_i - x_j)\right|^2 = \sum_{j=1}^{k}\sum_{n=1}^{k} w_{ij}w_{in}\emptyset_{jn}^{i} \qquad (18)$$

On a la fonction de coût suivante : où ϕ^iest une matrice K x K définie par ϕ_{jn}^{i}= $(x_i - x_j)^t$ (x_i-x_n) on pose R^i = $(\phi^i)^{-1}$. Résoudre le problème $\sum_{j=1}^{k} w_{ij}$=1 qui est un problème des moindres carrés, avec la contrainte, conduit à :

$$w_{ij}=\frac{\sum_{n=1}^{k} R_{jn}^{i}}{\sum_{p=1}^{k}\sum_{q=1}^{k} R_{pq}^{i}} \qquad (19)$$

Projection des vecteurs

Cette dernière étape a pour objectif de reconstruire la topologie locale, représentée par la matrice des poids W, dans l'espace euclidien F. Par conséquent, cela permet bien de déplier la surface initiale représentée par les vecteurs (x_i) dans un espace de dimension inférieure. Ce dépliage est obtenu en minimisant la fonction de coût suivante :

$$\phi(y)=\sum_i\left|y_i - \sum_i w_{ij}y_j\right|^2 \qquad (20)$$

Où y est l'ensemble des vecteurs projetés sur F. La fonction ϕ peut-être réécrite sous la forme quadratique suivante :

$$\phi(y)=\sum_i\left|y_i - \sum_i w_{ij}y_j\right|^2 = \sum_i\left[y_iy_j - 2\sum_j w_{ij}y_iy_j + \sum_j\sum_k w_{ij}w_{ik}(y_iy_k)\right]$$
$$=\sum_i[\sum_j \delta_{ij}(y_iy_i) - 2\sum_j w_{ij}y_iy_j + \sum_j\sum_k w_{ij}w_{ik}(y_iy_k)] \qquad (21)$$

Ce qui s'écrit sous forme matricielle : ϕ (y) = Y^TM Y,

Cette minimisation doit être conduite sous les deux contraintes suivantes :

- Toute translation des vecteurs y_j ne doit pas modifier ϕ . Pour supprimer ce degré de liberté, on va imposer aux vecteurs d'être centrés en l'origine $\sum_i y_j$ = 0
- Pour ne pas avoir de solutions dégénérées, les vecteurs y_j sont contraints d'avoir une même covariance $\frac{1}{N}\sum_i y_i \otimes y_i$= I

La dernière étape consiste à aligner les coordonnées globales. On calcule le (d+1) plus petits vecteurs propres de B correspond à la (d+1) valeurs propres.

2.3.10 Local Tangent Space Analysis (LTSA) : Analyse Tangent et Local de l'Espace

Local Tangent Space alignement [12] similaire à HLLE, estime le plan tangent à la variété autour de chaque point (à partir de ses voisins), et trouve des coordonnées globales (i.e. calculées indépendamment dans chaque voisinage par projection sur le plan tangent) en minimisant l'erreur de l'alignement :

$$\arg\min_{y} \sum_{i=1}^{N} \|W_i Y_i\|^2 \qquad (22)$$

Sous les deux contraintes $\sum_{i=1}^{N} y_j$=0, Y $Y^T = I_{dxd}$

Avec $Y_i = [y_{i_1}, y_{i_2}, .., y_{i_k}]$, W_i = I- G_i G^T

Algorithme

i. Extraire l'information locale : pour chaque i =1,…n. on devrait déterminer le k plus proche voisins x_{ij} de x_i et de calculer les d vecteurs propres $g_1, g_2, \ldots, g_d$ de la matrice de corrélation $(X_i - \overline{x_i}\, e^T)^T$ (X_i-$\overline{x_i}$ e^T) et G_i = [e/$\sqrt{k}$, $g_1, g_2, \ldots, g_d$]. X_i est un matrice D x (k+1) formé par les w x_i et ses k plus proches voisins $\{g_1, g_2, \ldots g_k\}$, $\overline{x_i}$représente le moyenne de x_{ij} et e un vecteur colonne de k-dimension.
ii. Construction de la matrice d'alignement : Former une matrice B par les sommes locales, i.e. B (I_i, I_i) I - G_i G^T → B (I_i, I_i), I_i= $\{i_1, i_2, \ldots i_k\}$, est l'ensemble des indices de voisinages des points x_i.
iii. Aligner les coordonnées globales : On calcule le (d+1) plus petits vecteurs propres de B correspond à la (d+1) valeurs propres.

2.3.11 Laplacien Eigenmaps

L'algorithme Laplacian Eigenmaps calcule une représentation de faible dimension Y de l'ensemble X, en préservant les relations de proximité des données dans l'espace de départ [20]. L'algorithme ressemble un peu à la méthode LLE, cependant il possède un fondement théorique différent : l'information de voisinage est récupérée à l'aide d'un graphe mais les coordonnées de faible dimension sont obtenues à partir de la notion du Laplacien du graphe [20].

Algorithme

Détermination des voisins pour chaque point (avec les techniques définies pour Isomap et LLE). Dans la construction du graphe, les sommets sont les points et les arêtes sont non nulles seulement si x_i et x_j sont proches. Dans la Pondération des arêtes, il y a deux possibilités :

- Noyau avec un paramètre t ∈ R. Si les sommets i et j sont reliés,

$$w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (23)$$

- Noyau avec un paramètre t = 1. Si les sommets i et j sont reliés, w_{ij} = 1

D'après le graphe G(s; a) défini auparavant, on calcule les vecteurs et valeurs propres du système :

$$L f = \lambda \Gamma D f \quad (24)$$

Où D est une matrice diagonale des poids dont les éléments sont la somme de chaque ligne de W, $D_{ii} = \sum_j w_{ij}$. L est la matrice du Laplacien L = D - W. L est une matrice symétrique semi définie positive qui peut être considérée comme l'opérateur Laplacien des fonctions définies dans le graphe G.

2.3.12 ACP Kernelisée

Schölkopf a introduit en 1998 le concept du noyau de kernel pour rendre non linéaire l'ACP traditionnelle [46]. En effet, le calcul de l'ACP ne fait intervenir que des produits scalaires entre les points (pour le calcul de la matrice de covariance) et ne considère jamais les coordonnées d'un point isolé. Si l'on remplace le produit scalaire par un noyau, on intègre les points ne faisant pas partie de l'ensemble d'apprentissage, ce qui n'est pas le cas de toutes les méthodes de réduction de dimension non linéaires. On calcule donc les composantes principales dans l'espace de caractéristiques H, et on peut ainsi accéder à des corrélations d'ordre supérieur entre les variables observées. Remarquons que l'on peut calculer la projection d'un point ne faisant pas partie de l'ensemble d'apprentissage, ce qui n'est pas le cas de toutes les méthodes de réduction de dimension non linéaires [47].

Soit l'ensemble d'apprentissage {x_i, i=1,…n}, soit ϕ: $R^m \rightarrow$ F notre fonction de projection, un noyau de Mercer permet d'exprimer directement le produit scalaire de la fonction de projection :

$$\phi(x) * \phi(y) \equiv k(x,y) \quad (25)$$

Parmi ces noyaux de Mercer, on peut citer les noyaux gaussiens k(x; y) = exp (-$\|x-y\|^2$/c) (F de dimension infinie avec ϕ explicite). La matrice de noyaux K de taille n x n est définie par $K_{ij} = \phi(x_i) * \phi(x_j)$ et les noyaux polynomiaux k(x; y) = $(1 + x.y)^d$

Le problème revient alors à déterminer les vecteurs propres de la K lème composante non-linéaire de l'image ϕ d'un individu x est le paramètre de variabilité β_k :

$$\beta_k = (V^k * \phi(x)) = \sum_{i=1}^{n} \alpha_i^k k(x, x_i) \quad (27)$$

Si les observations ne sont pas centrées dans F, on utilise la matrice de Gramm à la place de la matrice à noyau : $\widetilde{K_{ij}} = K_{ij} - \frac{1}{n}\sum_{p=1}^{N} K_{ip} - \frac{1}{n}\sum_{q=1}^{N} K_q + \frac{1}{n^2}\sum_{p.q=1}^{N} K_{pq}$ (28)

La reconstruction sur les premières composantes pose le problème de l'antécédent f d'un point de F, parfois inexistant. Le problème est souvent abordé [48] en minimisant la distance entre ϕ (f) et la projection de notre forme réduite à ses premières composantes :

$$\min_{f \in R^m} \|\phi(f) - P_n \varphi(Z)\|^2 = \min (K(f,f)) - 2 \sum_{k=1}^{N} \beta k \left(\sum_{i=1}^{n} \alpha_i^k k(f, x_i) + w\right) \quad (29)$$

Cette minimisation dépend de la forme du noyau choisi. On peut remarquer que le dernier terme de la quantité à minimiser est constant et que les paramètres de variabilité sont connus à priori.

2.4 APPROCHES PAR SELECTION DE BANDES SPECTRALES

La sélection de bandes consiste à choisir un sous-ensemble optimal de bandes pertinentes à partir d'un ensemble de bandes. L'intérêt de la sélection de bandes est de réduire la dimensionnalité des données hyperspectral, les coûts d'informations et d'accélérer la découverte de connaissance, et d'autres tâches dans des domaines comme la reconnaissance des formes, la modélisation de séries chronologiques et l'identification de processus. Nous présentons ci-dessous les méthodes non supervisée, non paramétrique basées sur la sélection de bandes spectrales.

2.4.1 La propagation d'affinité (AP)

La propagation d'affinité est un algorithme récent de partitionnement de données "clustering", qui permet de déterminer les éléments d'un ensemble qui sont les plus représentatifs [23, 50]. Elle a été appliquée à la vision par ordinateur et à la bio-informatique. Les auteurs ont prouvé que cet algorithme non seulement donne de bons résultats de classification mais également plus rapide que les autres méthodes de classification (KNN, SVM et le classificateur GP). Les exemplaires produits par l'AP ont un degré d'importance très élevés et aussi moins de corrélation/similarité entre les bandes lors de la combinaison des informations. Cette performance de la sélection de la bande est évaluée pour chaque pixel de l'ensemble de données.

Algorithme

C'est un algorithme itératif basé sur le partage des affinités. Chaque élément $x_i \in R^D$ i=1,…, N de l'ensemble de données repère dans son voisinage un élément qui lui ressemble suffisamment, et augmente son affinité pour cet élément.

AP commence par construire la matrice de similarité S $\in R^{L\times L}$ dont ces éléments S(i, k) mesure le degré de similarité entre la bande i et la bande k pour i≠k.

La distance euclidienne est choisie pour mesurer la similarité

$$S(i,k) = \sum_{m,n=1}^{M,N}(R_{mni} - R_{mnk})^2 \quad (30)$$

Avec S(k,k) reflète la pertinence à priori de la bande k pour servir comme exemple sous le nom "préférence" au lieu de fixer le nombre d'exemplaire à l'avance, les préférences sont utilisés pour contrôler le nombre de bandes sélectionnés comme exemplaires. Initialement les disponibilités sont mises à zéro. Ensuite les disponibilités et les responsabilités sont calculées selon les règles suivantes [22]:

$$r(i,k) \leftarrow s(i,k) - max_{k'st.k'\not\equiv k}\{a(i,k') + s(i,k')\} \quad (31)$$

$$a(i,k) \leftarrow s(i,k) - \min\{0, r(k,k) + \sum_{i'sti'} max_{k'st.k'\not\equiv\{i,k\}}\{0, r(i',k)\}, \quad (32)$$

En outre, l'auto disponibilité est mise à jour différemment:

$$a(k,k) \leftarrow \sum_{i'sti'} max_{k'st.k'}\{0, r(i,k')\} \quad (33)$$

Les règles de mise à jour conduisent souvent à des oscillations provoquées par le dépassement de la solution, de sorte que la responsabilité et la disponibilité sont évaluées comme suit.

$$M_{new} = \lambda M_{old} + (1-\lambda) M_{new} \quad (34)$$

Ou M_{new} et M_{old} sont les valeurs des messages de l'itération précédente et actuelle, λ est le facteur d'amortissement compris entre 0 et 1. Dans toutes les expériences illustrées dans ce qui précède, le facteur d'amortissement a été pris par défaut de valeur 0.5. Certains critères d'arrêt de la procédure de transmission de messages sont définis par le nombre de classes ou de la variation relative de l'erreur quadratique moyenne. Les auteurs ont comparé expérimentalement cet algorithme, et ont montré sa performance par rapport aux deux techniques: Méthode à divergences (Climbing Discrepancy Search) [83] et l'Analyse en Composantes Principales (PCA) [58], en utilisant deux types de données hyperspectrales. La première donnée de test est une section d'image prise à Washington (500x307 pixels, 210 bandes, 7 classes), la seconde donnée hyperspectral est une section d'une scène prise au nord-ouest de l'Indiana Pines (**Figure 13**, 145x145pixels, 220 bandes). L'évaluation a été réalisée à travers des méthodes de classification : KNN, SVM et le classificateur GP sont adoptés pour mesurer la performance de chaque méthode de la sélection de bande. Elle apporte de bons résultats dans la technique de réduction par sélection, évalué par le taux de classification (KNN, SVM, GP), par contre souffre du problème dans la complexité de calcul.

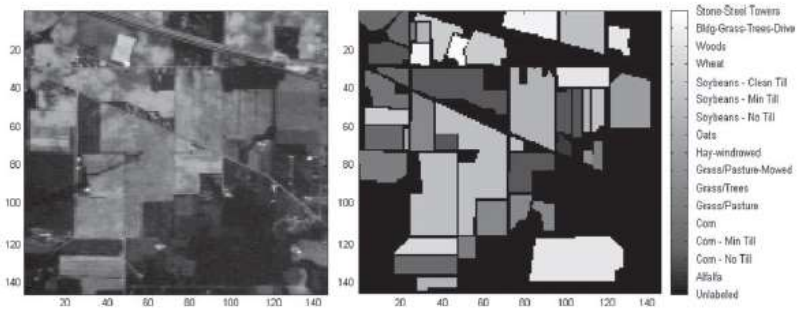


Figure 13. Bandes 10 D'AVIRIS Indiana Pines

2.4.2 Séléction par multicritères

Cette méthode de la sélection de bande est une extension de l'algorithme SPICE "Sparcity Promoting Iterated Constrained Endmember" [26]. Cet algorithme tente de manière autonome d'effectuer la sélection de bandes, tout en effectuant la détermination des composantes spectrales et le demixage, pour maintenir la signification physique des bandes spectrales [31]. La pondération de la bande et la promotion parcimonie d'une bande (BST) sont les termes ajoutés à la fonction objective SPICE qui est l'extension de l'algorithme ICE [29] défini par :

$$J = \eta \frac{RSS_B}{N} + \beta\, SSD_B + SPT + BST \quad (35)$$

Avec $RSS_B = \sum_{i=1}^{N} (wx_i - \sum_{k=1}^{M} P_{ik} wE_k)^T (wE_k - wE_l)$ (36)

$$SSD_B = \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} (wE_k - wE_l)^T (wE_k - wE_l)$$

$$SPT = \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} P_{ik}$$

$$W = \mathrm{diag}(w_1, .., w_d),\ \gamma_k = {}^{T}/_{\sum_{i=1}^{N} P_{ik}} \quad (37)$$

Où w_i est le poids pour la bande i, d est le nombre de bandes, η et β sont des paramètres à coefficient constantes pour les termes RSS_B et SSD_B, T est une constante qui permet de contrôler le degré parmi les composantes spectrales.

Le terme BST est défini comme la somme pondérée des poids de bande

$$BST = \sum_{j=1}^{d} \lambda_j |w_j| = \sum_{j=1}^{d} \lambda_j w_j \quad (38)$$

Où $\lambda_j = \frac{\Lambda\left(\left(\sum_{k=1}^{M}\frac{1}{N}\sum_{i=1}^{N}p_{ik}\,(x_{ij}-e_{jk})^2\right)+1\right)}{\sum_{k=1}^{M}(e_{jk}-\mu_{oj})^2+1}$ (39)

Λ est un paramètre ajustable contrôlant le degré de parcimonie chez les bandes pondérées. μ_0 est la moyenne globale des données, x_{ij} représente la j^ieme^ bande du pixel i, e_{kj} représente la j^ième^ bande de la composante spectrale k. Les bandes pondérées sont soumis à la contrainte suivante: $w_j \geq 0$, j=1,…d $\sum_{j=1}^{d} w_j$=d où d est le nombre de bandes.

2.4.3 La Sélection Progressive des Bandes (PBS)

Une nouvelle approche [28], appelée sélection de bandes progressive (PBS) a été utilisée pour la compression et la transmission des données progressives. Tout d'abord, PBS priorise chaque bande spectrale en fonction de son contenu d'information, mesurée par un certain critère de sélection de bandes, basée sur la statistique d'ordre élevé d'ordre 2, 3 et 4 (Variance, Skewness, Kurtosis), l'entropie, la divergence de Kullback-Leibler (divergence K-L). Ensuite, les bandes sont sélectionnées progressivement selon ses priorités attribuées à chaque bande spectrale. PBS peut être réalisé en trois manières opposées.

- Stratégie ascendante : Forward Selection (**FPBS**),
- Stratégie descendante : Backward Selection (**BPBS**),
- Stratégie bidirectionnelle : Bidirectional Selection (**BBBS**).

Les statistiques de divers ordres peuvent être utilisées comme attributs pour hiérarchiser :

a. Statistiques d'ordre fini

Variance (moment central d'ordre2)

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (40)$$

Où x_i est la valeur de pixel i, $\bar{x}$ est la moyenne de tous les pixels de la bande et n le nombre de pixels.

Skewness (Moment Central d'Ordre 3)

$$\text{Skeweness} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^3} \quad (41)$$

Kurtosis (Moment Central d'Ordre 4)

$$\text{Kurtosis} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^3} \quad (42)$$

b. Statistique d'ordre infini

Deux critères peuvent tirés de statistiques d'ordre infini: l'entropie et la divergence d'information.

- L'entropie :

L'entropie est une mesure qui correspond à la quantité d'information contenue ou délivrée par une source d'information. Défini par :

$$H(X) = -\sum_{i=1}^{n} P(x_i)\, log_2 P(x_i) \quad (43)$$

Avec $p(x)=(p(x_1),\dots,p(x_n))^T$est la fonction de probabilité calculée à partir de l'histogramme des pixels de l'image X.

- Divergence d'information

La divergence d'information [49], est une mesure alternative du contenu de l'information. Il mesure la divergence entre deux distributions de probabilité [69].

$$KL(i,k) = \sum_{i=1}^{n} P(x_i) \log \frac{P(x_i)}{g(x_i)} + \sum_{i=1}^{n} g(x_i) \log \frac{g(x_i)}{q(x_i)} \quad (44)$$

Le mécanisme de PBS commence par **FPBS** qui procède par agrégations successives par ajouts successifs de bandes. Au départ l'ensemble des bandes sélectionnées est initialisé à l'ensemble vide dans le sens où on commence par un faible nombre des bandes et on améliore progressivement la qualité des données en y incluant plusieurs bandes jusqu'à ce que la qualité des données soit satisfaisante.

La méthode **BPBS** est une procédure inverse de la précédente par retraits successifs de bandes, qui commence avec un nombre élevé de bandes spectrales et progressivement les supprime, conformément à leurs priorités de mesure, jusqu'à rendre la qualité des données inférieure à une donnée à tolérer.

La procédure **BBBS** effectue sa recherche dans les deux directions - *Forward* et *Backward* - d'une manière concurrentielle. Afin de déterminer les limites inferieures et supérieures sur le nombre de bandes. La procédure s'arrête dans deux cas : quand une des deux directions a trouvé le meilleur sous-ensemble de bandes avant d'atteindre le milieu de l'espace de recherche, quand les deux directions arrivent au milieu.

Il est clair que les ensembles de bandes sélectionnées par **FPBS** et **BPBS** ne sont pas égaux à cause de leurs différents principes de sélection. Néanmoins, cette méthode réduit le temps de recherche puisque la recherche s'effectue dans les deux directions et s'arrête dès qu'il y a une solution quelle que soit la direction. La **figure 14** illustre la stratégie de recherche.

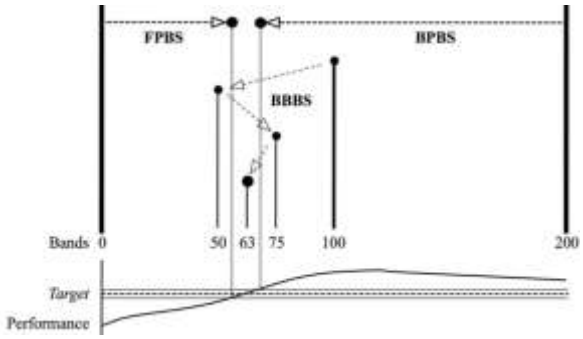


Figure 14. Procedure de recherche de FPBS, BPBS et BBBS

2.4.4 Méthodes hiérarchiques

La classification non supervisée ou classification automatique "**clustering**" [30] est une étape importante dans l'analyse de données. Le concept est étroitement lié à la notion de partition d'un ensemble fini, elle consiste à identifier des groupes d'objets ou d'individus similaires à partir d'un ensemble de données sans en connaître au préalable la structure. Elle ne doit pas être confondue avec la classification supervisée qui consiste à déterminer les règles qui ont permis de séparer un ensemble d'individus en classes connues à priori.

Une classification hiérarchique est une suite de partitions emboîtées $C^{(0)} < \ldots < C^{(N)} = \{\omega\}$ dont le premier terme $C^{(0)}$est la partition la plus fine qui ne contient que des singletons et dont le dernier terme est la partition la plus grossière qui ne comporte qu'une seule partie. La **figure 15** illustre ce concept dans le cas d'un ensemble de quatre objets.

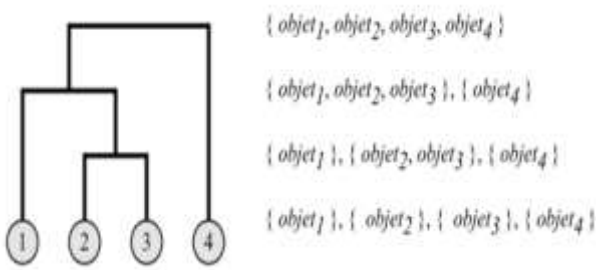


Figure 15. Mécanisme "clustering"

On distingue deux types d'approches de classification hiérarchique: les méthodes descendantes et les méthodes ascendantes.

2.4.1.1 MÉTHODES DESCENDANTES

Elles considèrent l'ensemble des observations ω et procèdent par division successive jusqu'à obtenir une partition formée de singletons.

2.4.1.2 MÉTHODES ASCENDANTES

Elles commencent avec la partition la plus fine de l'ensemble et procèdent ensuite par fusion progressive des parties jusqu'à obtention de la partition la plus grossière. On obtient ainsi un arbre binaire dont la racine correspond à la partition ne comportant qu'une seule partie et dont les feuilles s'identifient aux différents singletons. Les différents nœuds intermédiaires correspondent à la fusion de deux parties. La Classification Ascendante Hiérarchique (CAH) est sans nul doute la méthode la plus largement utilisée de cette catégorie. Différents indices d'agrégation de groupes ont été proposés :

- L'indice du saut minimum est défini comme la distance minimale qui sépare deux éléments issus de groupes différents.
- L'indice du saut maximum correspond à la distance maximale qui sépare deux éléments issus de groupes différents.
- L'indice du saut moyen est l'espérance de la distance qui sépare deux éléments issus de groupes différents.
- La distance entre les centroïdes des groupes qui se calcule au plus en temps linéaire contrairement aux indices précédents dont la complexité est quadratique.
- L'indice de Ward est défini comme l'augmentation de la variance intra-classe résultant de la fusion des deux groupes considérés.

2.4.5 Sélection par Prédiction et Projection

Deux algorithmes de sélection de bandes non supervisées basées sur la similarité métrique de bandes: la prédiction linéaire (LP) et la projection orthogonale (OSP) [52]. Les étapes de base sont décrites comme suit:

1. Initialiser l'algorithme en choisissant une paire de bandes B_1 et B_2 ensuite la sous bande $\emptyset = \{B_1, B_2\}$ est sélectionnée.
2. Trouver une troisième bande B_3 qui est la plus dissemblable à toutes les bandes en utilisant un certain critère. Ensuite, le sous-ensemble choisie est $\emptyset = \emptyset \cup \{B_3\}$. Continuer sur l'étape 2) jusqu'à ce que le nombre de bandes de $\emptyset$ est assez grand.

Algorithme LP

Supposons qu'il existe dans Ø deux bandes B_1et B_2, avec N pixels chacun. Pour trouver un groupe B qui est le plus dissimilaire à B_1, B_2, on utilise l'estimation minimisant l'erreur de prédiction linéaire : e= $\|B - B'\|$ avec B' = $\alpha_1 B_1$+$\alpha_2 B_2 + \alpha_0$ le vecteur a= $(a_0 a_1 a_2)^T$ Il peut être déterminé en utilisant a = $(X^T X)^{-1} X^T$y .
Ou X est une matrice N x 3 dont la première colonne égale à 1, la deuxième colonne comprends tous les pixels de B_1 et la troisième comprends tous les pixels de B_2. y est un vecteur n x 1 avec tous les pixels de B. Le groupe qui donne une erreur maximale est considéré comme bande inégales à B_1et B_2 et sera sélectionné comme B_3 pour Ø

Algorithme OSP

Supposons qu'il existe deux bandes B_1et B_2 dans Ø. Pour trouver un groupe qui est le plus dissemblable à ces deux bandes, un sous espace orthogonale de B_1et B_2 est construit comme suit :

$$P = I - z\,(Z^T Z)^{-1}\, Z^T \qquad (45)$$

I est la matrice d'identité N x N, z est une matrice N x 2 dont la première colonne comprend tous les pixels de B_1et la deuxième colonne tous les pixels dans B_2 ensuite une projection Y_0 = P^Ty est calculée. Le groupe qui donne le maximum des composantes orthogonales sera alors sélectionné comme B_3pour Ø. Cette procédure est réalisée lorsque le nombre de bandes dans Ø > 2.

Une étude expérimentale a été faite sur une image AVIRIS (224 bandes, 350x350 pixels), a montré la performance de LP en termes de conservation de l'information et de séparabilité des classes avec d'autres méthodes de sélection de bandes non supervisée : l'entropie, NFINDR, premier dérivé et l'espace de spectre uniforme (USS) spectrale (FSD).

2.4.6 Méthodes basés sur la Théorie d'Information

L'utilisation de mesures d'informations, comme l'information mutuelle fournit une méthodologie pour trouver des corrélations généralisée entre les bandes d'image. Ainsi, BandClust [35], WaLuMi [51] exploitent ce concept pour la sélection de bande. D'autre part, WaLuDi utilise une mesure de divergence qui a été fréquemment utilisée afin de comparer les distributions de probabilités différentes.

2.4.6.1 BANDCLUST

Une nouvelle approche itérative, non supervisée dans la réduction de bande pour les images hyperspectrales est nommée BandClust [35]. Elle consiste à diviser la première gamme de bandes spectrales en des sous bandes disjoints, basées sur un critère de la minimisation de l'information mutuelle entre la moyenne de sous bandes. L'algorithme fait la subdivision de la plage d'index bande initiale [1, B] en sous bandes fines. Cette procédure est réalisée successivement sur chaque sous bande crée précédemment, en trouvant à chaque itération l'indice de la bande b_{opt} $\in[b_{min}, b_{max}[$qui minimise le critère d'information mutuelle. Une fois la sous bande $[b_k, b_{k+1}]$a été créé, alors les sous bandes voisins sont traitées (k ←k+1) jusqu'à $b_{k+1} =$ B. l'algorithme s'arrête lorsque le sous bandes précédemment crée ne peuvent pas être divisé, c'à d pas de minimum dans le critère d'information mutuelle. Le nombre final de sous bandes est alors considéré comme la solution optimale.

Les auteurs ont été montrés que BandClust donne un meilleur classement en comparaison avec les autres méthodes de réduction non supervisée PCA, LP\OSP, WaLuMi et WaLuDi, testées sur des données hyperspectrales (AVIRIS, ROSIS, CASI) [35].

BandClust a plusieurs avantages :

- Apporte de bons résultats en comparaison avec PCA, LP\OSP, WaLuMi\WaLuDi.
- Elle ne nécessite aucune connaissance de la vérité terrain,
- Elle fournit automatiquement une estimation du nombre optimal de bandes,
- Elle préserve la signification physique des données hyperspectrales.

2.4.6.2 WARD'S LINKAGESTRATEGY USING MUTUAL INFORMATION (WALUMI)

La méthodologie de l'algorithme WaLuMi [32, 51] se résume comme suit: Un espace de similarité est défini entre les bandes, ou une mesure de dissimilarité basée sur l'information mutuelle entre paire de bandes. Le processus commence avec une classification hiérarchique défini dans l'espace de dissemblance. Afin de construire progressivement une famille hiérarchique qui utilise une stratégie de liaison de distance entre les bandes comme fonction objectif à optimiser. Enfin, une bande représentante est choisie pour chaque classe. Les k bandes pertinentes sélectionnés auront ainsi un degré d'indépendance.

2.4.6.3 WARD'S LINKAGE STRATEGY USING DIVERGENCE (WALUDI)

Dans la théorie d'information, la divergence de Kullback-Leibler peut être considéré comme une sorte de distance entre deux densité de probabilités ce n'est pas une mesure de distance réelle car elle n'est pas symétrique [33, 53]. Ainsi une version modifiée de Kullback-Leibler qui est symétrique, utiliser pour mesurer l'écart entre deux distributions de probabilité. En ce sens, il peut être également utilisé pour mesurer la similitude entre les deux bandes, de représenter chaque bande par une distribution de probabilité. Cette mesure de divergence définie comme suit :

$$D\,(p_l, p_k) = \sum_I p_{li} \log\left(\frac{p_{li}}{p_{ki}}\right) + \sum_I p_{ki} \log\left(\frac{p_{li}}{p_{ki}}\right)) \qquad (46)$$

Où p_l, p_k Sont les distributions de probabilité associée à ces deux bandes k et l.

Une étude expérimentale a été faite sur trois types d'images multi spectrales (Aviris, Hymap et RGB) [34]. KNN est utilisée pour évaluer la robustesse de ces deux méthodes WaLiMi, WaLuDi en comparaison avec MVPCA et ID en terme où il existe une forte corrélation entre les bandes sélectionnées.

2.4.7 Méthode basée sur l'Histogramme

Une approche [36] basée sur l'histogramme a été étudiée dans la sélection de bandes, par une analyse des histogrammes globaux construites à partir des pics significatifs des histogrammes locaux pour chaque bande. Le but est de sélectionner les bandes pertinentes en éliminant les redondances. Tout d'abord, une agrégation des bandes qui permet de former des classes de bandes similaires en minimisant la dissimilarité entre deux histogrammes globaux construites associées aux 2 classes adjacentes.

$$D\,(H_u, H_v) = 1 - \frac{\sum_i u_i v_i}{\sum_i u_i^2 + \sum_i v_i^2 - \sum_i u_i v_i} \qquad (47)$$

Avec $H_u = (u_1, u_2, .. u_n)$ représente l'histogramme globale associé à la bande B_u.

$H_v = (v_1, v_2, .. v_n)$ représente l'histogramme globale associé à la bande B_v.

Une étude expérimentale a été menée sur une image multispectrale compose de dix bandes spectrales (1024x512). Trois classes de bandes similaires ont été obtenues par le processus de l'agrégation. On peut remarquer dans la **figure 16** que la différence est claire de ces 3 classes de bandes similaires sélectionnés, et que dans le sens de niveau de gris, les différentes régions qui composent les bandes originales sont bien détectées et les détails ne sont pas perdus.

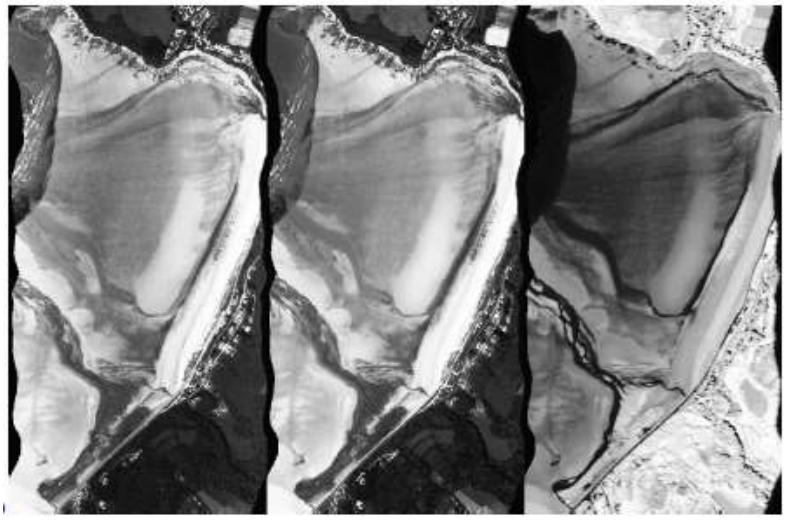


Figure 16. Les 3 bandes sélectionnées

En conclusion, cet algorithme basé sur l'histogramme ne préserve pas les évènements rares, en minimisant la dissimilarité entre deux histogrammes globaux associés à chaque bande, en maximisant le critère d'entropie. Cette méthode peut être améliorée, de maximiser la dissimilarité en minimisant le critère d'entropie en gardant l'évènement rare.

2.5 CONCLUSION

Au cours de ce chapitre, nous avons présenté les deux approches générales non paramétriques et non supervisées pour la réduction de la dimensionnalité des images à grande dimension, et nous avons rappelé les principes de ces méthodes. Avant de poursuivre, rappelons que cette thèse s’inscrit dans le cadre de l’apprentissage non supervisé, non paramétrique en préservant l'évènement rare et que dans ce contexte, nous nous intéressons aux méthodes automatiques de réduction de la dimensionnalité pour les images de grande dimension.

Les techniques non supervisées par projection sont, soit limitées par leur caractère linéaire (ACP, MDS), soit difficilement utilisables à cause de leur complexité algorithmique (Isomap, LLE,..) lorsque l’on travaille sur de grandes bases données et la majorité d'entre elles sont sensibles aux différentes variations qu'elles subissent (paramètres, dégradation de bruits) et à la perte d'informations. C'est la raison qu'elles ne répondent pas totalement au cahier des charges de la thèse. Tenant compte des critères d'évaluation locaux /globaux présentées dans le chapitre suivant, les auteurs ont montré, la pertinence de certains algorithmes de réduction de la dimension par projection (LTSA, H.LLE, Isomap, MVU) en comparaison avec les autres techniques (PCA, K.PCA, DM, LLE, LEM, Sammon, Autoencoder, LLC, MC, LP, LE) pour préserver la structure géométrique locale/globale de données.

La stabilité et la préservation de l'évènement dans quelques méthodes (LTSA, LLE, LE, MVU, HLLE, Isomap, Autoencoder's, RBM, KPCA, MDS), a fait l'objet de quelques études comparatives dans la littérature scientifique [54, 55, 56]. Les auteurs ont montré que les algorithmes MVU, LTSA et HLLE sont assez sensibles à la variation des paramètres et à la dégradation de bruits en comparaison avec les autres techniques de réduction. Concernant la préservation de l'évènement rare, Oleg Kubeyda a cherché [57] à déterminer dans le cas d'un algorithme SVD semi supervisé a priori le rang du sous espace à rechercher. Ceci ne répond pas directement aux points dans le cahier de charge.

La réduction de la dimensionnalité basée sur la sélection de bandes spectrales non supervisée est considéré comme un processus de regroupement des données qui partitionne les données hyperspectrales en groupes d'objets similaires (classes), exploitant des critères de complémentarité d'information ou bien sur la pondération des bandes spectrales. Toutefois, nombreuses sont celles qui s'appuient sur une vérité terrain qui est en générale imparfaite ou qui ne permettent pas de préserver toute la richesse de l'information statistiquement rares apportée par certains bandes spectrales. C'est la raison qui nous ramènera par la suite à sélectionner dans notre étude celles qui fonctionnent automatiquement (sans connaissance a priori). Selon les études comparatives sur la réduction de la dimension par sélection de bande, les algorithmes BANDCLUST et celui basé sur l'histogramme semblent les plus pertinents en comparaison avec les autres méthodes non supersivée WaluMi, WaLuDi, LP/OP et PCA, mais le problème de la préservation de l'évènement rare reste non abordé dans la littérature scientifique comparative pour cette classe des méthodes de réduction.

Il nous semble malgré tout naturel de se focaliser sur les techniques de sélection qui, relativement aux méthodes par projection, permettent de rester dans l'espace des observations et de ne pas imposer d'interprétation de nouvelles variables à l'utilisateur. La sélection de bandes en apprentissage non supervisé est un domaine encore peu exploré et les techniques existantes reposent pour beaucoup sur des mesures de similarité entre attributs (bandes spectrales) ou sur des mesures de variances. Il s'agit d'un problème qui est plus difficile que dans le cas supervisé, car aucune information n'est disponible pour guider la procédure. La détermination automatique du nombre de bandes est un problème associé très important et ces deux problèmes interfèrent l'un avec l'autre. La méthode basée sur l'histogramme doit être amélioré de le sens de maximiser la dissimilarité entre deux histogrammes globaux et de minimiser le critère d'entropie.

CHAPITRE 3
CRITÈRES D'ÉVALUATION DES MÉTHODES DE RÉDUCTION DE LA DIMENSIONNALITÉ

Après avoir défini les particularités du domaine de l'hyperspectral et les approches qui ont été développées pour la réduction de la dimensionnalité, il est nécessaire de recenser un ou plusieurs critères d'évaluation de la qualité d'image. Ces critères permettront de comparer les différentes méthodes de réduction et, en particulier, de mesurer les différents types de dégradation causés par les diverses méthodes de réduction.

Ces critères permettent d'analyser l'efficacité des méthodes de réduction de la dimension, à réduire l'information, ainsi que leur capacité à préserver ou non la totalité de la richesse de l'information (spectrale et spatiale) liée à la géométrie (globale et locale) apportée par l'imagerie hyperspectrale. L'évaluation peut être réalisée soit directement par des critères locaux et globaux, soit indirectement par des méthodes de classification. Dans cette section, nous présentons les quatre catégories principales de critères relatives à la perte d'information, à la préservation de l'évènement rare, à la stabilité et à la similarité. Ensuite, nous proposons une approche de mesurer localement l'évènement rare et de l'appliquer sur l'image hyperspectrale.

3.1 CRITÈRES LOCAUX

Ces critères mesurent la préservation de la structure géométrique locale des données, et d'analyser la situation de chacun de l'ensemble de données par rapport à son voisinage. Dans cette partie, les critères LCMC [62], la fiabilité, la continuité [59] et l'erreur moyenne du rang relatif (MRRE) [60, 61, 62] ont été étudiés.

3.1.1 Local Continuity Meta–Criterion (LCMC): Le Critère de la Continuité Locale

Les auteurs introduisent un critère qui permet de mesurer la fidélité locale de l'ensemble des données dans l'espace réduit. LCMC calcule la moyenne entre deux ensembles de voisinages dans l'espace original avec son représentant dans l'espace réduit. LCMC est défini par :

$$Q_{LC}=1-\sum_{i=1}^{n}(|n_K^X(i)\cap n_K^Y(j)|)-\frac{K^2}{n-1}=\sum_{(k,l)\in\{1,\dots,K\}\times\{1,..,K\}} q_{kl} \qquad (46)$$

Avec $n_K^X(i)$ = {j:1 ≤ ρ_{ij} ≤ K}, $n_K^Y(j)$ = {j: 1≤ r_{ij} ≤ K}. Le symbole δ_{ij} indique la distance de x_i à x_j dans l'espace original. De même d_{ij}est la distance de y_i à y_j dans l'espace réduit. On peut calculer les rangs à partir des distances. Le rang de x_i par rapport à x_j dans l'espace original est écrit par : ρ_{ij} = |{k : $\delta_{ik} < \delta_{ij}$ ou ($\delta_{ik} = \delta_{ij}$et k<j)}|.

De même le rang de y_j par rapport à $y_i$dans l'espace réduit est : r_{ij} =|{k : $d_{ik} < d_{ij}$ ou ($d_{ik} = d_{ij}$et k<j)}|. Ainsi les rangs réflexives sont égales à zéro ($\delta_{ii} = \delta_{jj}$), q_{kl} =|{(i,j): ou ($d_{ik} = d_{ij}$et k<j)}|. q_{kl} =|{(i,j): $\rho_{ik} = k$et r_{ij}=1)|, K représente le nombre de plus proche voisins.

3.1.2 La Fiabilité (M_T)

La mesure de la fiabilité "Trusworthiness" consiste à sélectionner les k-plus proches voisins de points dans l'espace d'origine, ensuite on les identifie avec leurs correspondants dans l'espace réduit. On considère $U_k(x_i)$ comme l'ensemble des points qui ne sont pas dans le voisinage de x_i, tandis que les points correspondants sont dans l'espace réduit et sont dans le voisinage de x_i . M_T a une valeur compris entre 0 et 1 et définie par :

$$M_T = 1- A(k)\sum_{i=1}^{n}\sum_{x_j \in U_k(x_i)}(r(x_i, x_j) - k)) \quad (47)$$

3.1.3 La Continuité (M_C)

La continuité M_C mesure le degré de la continuité des points qui sont dans le voisinage, en vérifiant s'ils ont gardé la même position dans l'espace réduit. On considère $V_k(y_i)$ l'ensemble des points projetés qui ne sont pas dans le voisinage de y_i tandis que ces points correspondants dans l'espace d'origine sont dans le voisinage de x_i. M_C a une valeur compris entre 0 et 1 et définie par:

$$\mathbf{M_C} = 1- A(k) \sum_{i=1}^{n}\sum_{y \in V_k(y_i)}(\hat{r}(y_i, y_j) - k) \quad (48)$$

$$A(k)= {}^{2}/_{nK} \ (2n-3K-1) \quad ; \quad V_k(y_i) = \{y_j / \ y_j \notin \widehat{C_k}(y_i) \cap \in C_k(y_i)$$

3.1.4 Mean Relative Rang Error (MRRE): Le Rang d'Erreur Moyenne Relative

MRRE reflète localement la cohérence dans le voisinage des points sélectionnés à l'origine et leurs correspondants, définie par:

$$Q_T = \beta\, W_T + (1- \beta)\, W_C,\ \beta \in [0,1] \quad (49)$$

Avec $W_T = 1- \frac{1}{H_K}\sum_{i=1}^{n}\sum_{j \in U_K^i}\frac{|r(i,j)- \widehat{r(i,j)}|}{r(i,j)}$ et $W_C = 1- \frac{1}{H_K}\sum_{i=1}^{n}\sum_{j \in V_K^i}\frac{|r(i,j)- \widehat{r(i,j)}|}{r(i,j)}$ (50)

Où $H_K = \sum_{i=1}^{k}\frac{(1-2i+1)}{i}$ est un facteur de normalisation $U_k(x_i) = \{x_j / \ x_j \in C_k(x_i) \cap \notin \widehat{C_k}(x_i)\}$

3.2 CRITÈRES RELATIVES À LA PERTE D'INFORMATION

Dans cette partie, les critères sont directement tirés des mesures statistiques classiques, ils sont basées sur la distance entre l'image originale notée I et l'image après passage dans le processus de réduction notée $\tilde{I}$. Ces critères se reposent sur la sensibilité à l'erreur entre I et $\tilde{I}$. Les images hyperspectrales sont représentées sous la forme d'une matrice tridimensionnelle I(x, y, $\boldsymbol{\lambda}$), x est la position du pixel dans la ligne, y est le numéro de la ligne et $\boldsymbol{\lambda}$ la bande spectrale considérée. $\boldsymbol{n_x, n_y}$ ***et*** $\boldsymbol{n_\lambda}$ sont respectivement le nombre de pixels par ligne, le nombre de lignes et le nombre de bandes spectrales. Ces mesures sont ensuite normalisées de différentes manières pour les rendre plus ou moins indépendantes de la taille et de la nature des données originales.

Les mesures les plus utilisées sont issues des normes $\boldsymbol{l_p}$ et basées sur les distances associées. La norme $\boldsymbol{l_p}$ est définie par:

$$\boldsymbol{l_p}(I) = \|\boldsymbol{I}\|_p = \left(\sum_{x,y,\lambda}|\boldsymbol{I}(x, y, \lambda)|^p\right)^{1/p}$$

La distance correspondante, également appelée distance de Minkowski, est définie par :

$$d_p(\boldsymbol{I}, \tilde{I}) = \boldsymbol{l_p}(\boldsymbol{I} - \tilde{I}) = \left(\sum_{x,y,\lambda}\left|\boldsymbol{I}(x, y, \lambda) - \tilde{I}(x, y, \lambda)\right|^p\right)^{1/p}.$$

3.2.1 L'Entropie

L'entropie de Shannon, due à Claude Shannon [63], est une fonction mathématique qui, intuitivement, correspond à la quantité d'information contenue ou délivrée par une source d'information. Pour une source qui est une variable aléatoire discrète x, comportant n symboles $\boldsymbol{x_i}$, ayant une probabilité $\boldsymbol{p_i}$ d'apparaître, l'entropie H de la source X est définie comme :

$$H = \sum_{i=0}^{N-1} \boldsymbol{H_i} = \sum_{i=0}^{N-1} \boldsymbol{P_i}\, \boldsymbol{Q_i}, \text{ avec } \boldsymbol{Q_i} = -\mathbf{log_2}\, \boldsymbol{P_i}. \quad (51)$$

Où $\boldsymbol{P_i} = \frac{\boldsymbol{n_i}}{N}$ est la probabilité associée à l'apparition de l'évènement i. Ainsi, $\boldsymbol{P_i}$ varie entre 0 et 1. S'il est très petit, la quantité H est très grande, alors que si $\boldsymbol{P_i}$ est très grande, elle est presque nulle.

3.2.2 La Variance

Parmi les techniques statistiques, l'analyse de la variance reste encore un peu à l'ombre, parce que, quoiqu'elle ait été définie dans son principe depuis assez longtemps, elle n'a été mise à la portée des statisticiens non spécialistes qu'avec la diffusion générale des moyens de calcul modernes. Alors qu'elle est en effet d'une logique relativement simple dans son principe, sa pratique exige des calculs nombreux qui étaient extrêmement fastidieux aussi longtemps que l'ordinateur n'a pas pu être chargé de leur réalisation. Le but qu'on se propose dans ce type de calcul est en fait une sorte de généralisation de la comparaison des moyennes ou de la comparaison des pourcentages lorsqu'il y a plus de deux valeurs à comparer.

$$\widehat{\boldsymbol{\Sigma}} = \frac{\mathbf{1}}{\mathbf{I_1 I_2}} \sum_{\mathbf{i_1=1}}^{\mathbf{I_1}} \sum_{\mathbf{i_2=1}}^{\mathbf{I_2}} (\mathbf{r_{i_1 i_2}} - \widehat{\boldsymbol{\mu}})^{\mathbf{2}} \quad (52)$$

$$\text{Avec } \widehat{\boldsymbol{\mu}} = \frac{\mathbf{1}}{\mathbf{I_1 I_2}} \sum_{\mathbf{i_1=1}}^{\mathbf{I_1}} \sum_{\mathbf{i_2=1}}^{\mathbf{I_2}} \mathbf{r_{i_1 i_2}}$$

3.2.3 Dérivée Spectral d'Ordre 1

La plage de la longueur d'onde de chaque bande est une variable dans un capteur hyperspectral [73]. Cette méthode explore la largeur de bande variable en fonction de l'information ajoutée. Il est évident que si deux bandes adjacentes ne diffèrent pas beaucoup, alors la propriété de la sous-jacente géo-spatiale peut être caractérisée avec une seule bande. La description mathématique est représenté ci-dessous, où I représente la valeur hyperspectral, x est une position spatiale et λ est la longueur d'onde centrale. Ainsi, si $\mathbf{D_1}$ est égal à zéro, alors l'un des groupes est redondant. En général, les bandes adjacentes qui diffèrent de manière significative doivent être conservés, tandis que des bandes adjacentes similaires peuvent être réduites.

$$\mathbf{D_1}(\boldsymbol{\lambda_i}) = \sum_{\mathbf{x}} \|\mathbf{I}(\mathbf{x}, \boldsymbol{\lambda_i}) - \mathbf{I}(\mathbf{x}, \boldsymbol{\lambda_{i+1}})\| \quad (53)$$

3.2.4 Erreur Moyenne Absolu (EMA) ou Moyenne de l_1

On moyenne ici par le nombre de pixels pour rendre la mesure indépendante de la taille de l'image.

$$\text{MAE}(\mathbf{I}, \tilde{I}) = \frac{\mathbf{1}}{\mathbf{n_x n_y n_\lambda}} \sum_{\mathbf{x,y,\lambda}} |\mathbf{I}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) - \bar{\mathbf{I}}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda})| \quad (54)$$

3.2.5 Erreur Quadratique Moyenne (EQM) ou Moyenne de l_2^2

Ici, on moyenne également sur le nombre de pixels.

$$\text{MSE}(\mathbf{I}, \tilde{I}) = \frac{\mathbf{1}}{\mathbf{n_x n_y n_\lambda}} \sum_{\mathbf{x,y,\lambda}} \left(\mathbf{I}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) - \tilde{I}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda})\right)^{\mathbf{2}} \quad (55)$$

Cette mesure peut être reliée à des notions de moyenne et variance. En notant μ_U la moyenne de l'ensemble U et σ_U^2 sa variance.

$$\text{MSE}= \sigma^2_{\mathbf{I}-\tilde{I}} + \mu^2_{\mathbf{I}-\tilde{I}}$$

3.2.6 Erreur Quadratique Moyenne Relative (RMSE)

On normalise en plus ici avec le niveau du signal, ce qui sous-entend qu'une erreur de même amplitude sera moins gênante sur un signal important que sur un signal faible.

$$\text{RMSE} = \sqrt{\frac{\mathbf{1}}{\mathbf{n_x n_y n_\lambda}} \sum_{\mathbf{x,y,\lambda}} \left(\frac{\mathbf{I(x,y,\lambda)}-\tilde{I}\mathbf{(x,y,\lambda)}}{I(x,y,\lambda)}\right)^{\mathbf{2}}} \qquad (56)$$

3.2.7 Average RMSE (A-RMSE) : La Moyenne du RMSE

Ce critère réalise une moyenne du RMSE calculé pour chaque vecteur [75, 79].

$$\text{ARMSE} = \frac{\mathbf{1}}{\mathbf{n_x n_y}} \sum_{\mathbf{x,y}} \sqrt{\frac{1}{n_\lambda} \sum_\lambda \left(\mathbf{I(x,y,\lambda)} - \tilde{I}\mathbf{(x,y,\lambda)}\right)^{\mathbf{2}}} \qquad (57)$$

3.2.8 PEAK RMSE (P-RMSE) : Le Maximum du RMSE

Ce critère représente le RMSE maximum calculé selon les vecteurs

$$\text{PRMSE}= max_{x,y} \sqrt{\frac{1}{n_\lambda} \sum_\lambda \left(\mathbf{I(x,y,\lambda)} - \tilde{I}\mathbf{(x,y,\lambda)}\right)^{\mathbf{2}}} = \frac{1}{n_\lambda} max_{x,y}\{l_2(v - v')\} \qquad (58)$$

3.2.9 Maximum Absolute Distorsion : Distorsion Maximale Absolue

Ce critère l_∞ utilisé par Motta [64], borne l'erreur sur l'ensemble de l'image

$$\text{MAD} = l_\infty(\mathrm{I} - \tilde{\mathrm{I}}) = \max_{x,y,\lambda}\{|\mathrm{I}(x,y,\lambda) - \tilde{\mathrm{I}}(x,y,\lambda)|\} \qquad (59)$$

3.2.10 Percentage MAD : Pourcent sur MAD

C'est une mesure très utilisée, notamment par Pickering et Ryan [65, 66, 67]. La définition se trouve dans [68] :

$$\text{PMAD} = \max_{x,y,\lambda} \left\{\frac{|\mathrm{I}(x,y,\lambda)-\tilde{\mathrm{I}}(x,y,\lambda)|}{\mathrm{I}(x,y,\lambda)}\right\} \text{ x } 100 \qquad (60)$$

3.2.11 Spectral Information Divergence (SID) : La Divergence de l'Information Spectrale

Ce critère se dérive de la distance de Kullback-Leibler [75]. En notant v = (v_0, v_1, . , v_i,.) on a :

$$\text{SID } (v,v') = D(v||v')+D(v'||v)$$

Avec D(v||v') la distance de Kullback-Leibler, ou divergence entropique, ou discrimination définie entre deux vecteurs par :

$$D(v||\ v')= \sum_i P_i \log(\frac{P_i}{Q_i})$$

Où $P_i = \frac{v_i}{\|v\|_1}$ et $\widetilde{p_i}= \frac{\widetilde{v_i}}{\|\tilde{v}\|_1}$. Ainsi :

$$\text{SID } (v,v')= \sum_i (p_i - \widetilde{p_i}) \log(\frac{P_i}{\widetilde{p_i}}) \qquad (61)$$

De même, ce critère mesure les distorsions spectrales tandis que les distorsions radiométriques n'impliquent pas nécessairement de distorsions spectrales.

3.2.12 Information Mutuelle

L'entropie d'un ensemble de valeurs discrètes X a valeurs dans [1, 2q] est définie par Shannon [81] comme

$$H(X) = -\sum_{i=1}^{2q} P(i) \ln P(i) \quad (62)$$

Où P(i) désigne la proportion de X prenant la valeur i. L'entropie conjointe entre X et Y, qui représente l'information apportée par X et par Y, est alors définie par :

$$H(X, Y) = -\sum_{i=1}^{2q} \sum_{i=1}^{2q} P(i, j) \ln P(i, j). \quad (63)$$

L'information mutuelle entre X et Y correspond à la quantité d'information commune à X et Y : I(X, Y) = H(X) + H(Y) − H(X, Y),

Le critère considéré est alors le rapport : $\frac{I(x,y)}{H(\frac{X}{Y})}$

Ce critère n'est pas largement employé et n'est utilisé que dans peu d'articles [78]. Une étude plus approfondie serait nécessaire pour mesurer son apport sur les images hyperspectrales.

3.3 CRITÈRES RELATIVES À LA STABILITÉ

Dans ce paragraphe, nous proposons de critères relatifs à la stabilité particuliers aux méthodes de réduction des images hyperspectrales.

3.3.1 Structural Content (SC) : Contenu Structurel

Le contenu structurel ou Strucutral Content (SC) est le rapport des densités spectrales de puissance (DSP) des deux images réduites (l'image perturbée réduite $\tilde{\boldsymbol{I}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ sur l'image de référence $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$) présenté dans [80] et est défini comme:

$$SC = \frac{\sigma_{\tilde{I}}^2 + \mu_{\tilde{I}}^2}{\sigma_I^2 + \mu_I^2} \quad (64)$$

Et pour les images hyperspectrales : $SC = \frac{\sum_{x,y,\lambda}[\tilde{\boldsymbol{I}}(x,y,\lambda)]^2}{\sum_{x,y,\lambda}[\boldsymbol{I}(x,y,\lambda)]^2}$

3.3.2 Normalized Cross-Correlation (NCC): L'Inter-Corrélation Normalisée

L'inter-corrélation normalisée ou Normalised Cross-Corrélation (NCC) est mentionnée dans les mêmes articles que ceux proposant d'utiliser la fidélité.

$$NCC = \frac{\sum_{x,y,\lambda} \boldsymbol{I}(x,y,\lambda)\tilde{\boldsymbol{I}}(x,y,\lambda)}{\sum_{x,y,\lambda}[\boldsymbol{I}(x,y,\lambda)]^2} \quad (65)$$

L'interprétation physique de ces trois derniers critères (fidélité, SC, NCC) reste plus difficile. Dans [74], Eskicioglu mesure les performances de ces trois critères. L'intercorrélation normalisée ainsi que le contenu structurel ne donnent pas de très bons résultats. La fidélité semble plus performante.

3.4 CRITÈRES RELATIVES À LA SIMILARITÉ

Les critères présentés ici sont spécialement adaptés au domaine de l'hyperspectral. Certains, comme la similarité spectrale est largement utilisée. On définit par v et $\mathbf{v}'$

respectivement un vecteur de l'image originale et un vecteur de l'image reconstruite contenant les n composantes spectrales du pixel (x, y).

3.4.1 La Fidélité (F)

Ce critère a été développé par Eskicioglu [69] et utilisé dans [70]. On définit la fidélité par un rapport des densités spectrales de puissance (DSP) de l'image d'erreur et de référence. U et V deux ensembles de valeurs (initialement les pixels de deux images) dont on veut évaluer la différence. La fidélité est définit par:

$$F(U, V) = 1- \frac{L_2^2(U-V)}{\sigma_u^2+\mu_u^2} \quad (66)$$

Soit pour les images hyperspectrales, une fidélité spatial sera définit comme :

$$F(\boldsymbol{I}, \widetilde{\boldsymbol{I}}) = 1- \frac{\sum_{x,y,}[\boldsymbol{I}(x,y,\lambda)-\tilde{\boldsymbol{I}}(x,y,\lambda)]^2}{\sum_{x,y,}[\boldsymbol{I}(x,y,\lambda)]^2}$$

La fidélité est égale à un, lorsque l'image de sortie est égale à l'image d'entrée. Du fait de la mesure quadratique, les faibles erreurs seront minimisées.

3.4.2 Similarité Spectrale (SS)

Ce critère apparaît dans [71, 72]. Il cherche à mesurer la ressemblance entre deux spectres vus comme des vecteurs a $\mathbf{n_\lambda}$ dimensions, **v** et $\mathbf{v}'$ défini par :

$$SS(\boldsymbol{v}', \boldsymbol{v}) = \sqrt{\boldsymbol{RMSE}(\boldsymbol{v}, \boldsymbol{v}')^2 + (\mathbf{1} - \boldsymbol{corr}(\boldsymbol{v}, \boldsymbol{v}')^2)^2} \quad (67)$$

Avec $RMSE = \sqrt{\frac{\sum_{\lambda=1}^{n_\lambda}(v(\lambda)-v'(\lambda))^2}{n_\lambda}}$ et $Corr(\boldsymbol{v}, \boldsymbol{v}') = \frac{\frac{1}{n_\lambda-1}\sum_{\lambda=1}^{n_\lambda}(v(\lambda)-\mu_v)(v'(\lambda)-\mu_{v'})}{\sigma_v\sigma_{v'}}$

3.4.4 Indice de Similarité Structurale

La mesure de la similarité structurelle est généralisée à partir du même indice de similarité entre deux signaux ou encore entre deux patches spatiaux x et y extraits respectivement de l'image originale et l'image dégradée. La similarité compare la luminance, le contraste et la corrélation entre chaque couple de vecteurs.

La luminance est estimée par la mesure de l'intensité moyenne de chaque faisceau spatial:

$$\boldsymbol{\mu_x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (68)$$

Le contraste de chaque faisceau est mesuré par:

$$\boldsymbol{\sigma_x} = \left(\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{\mu_x})^2\right)^{\frac{1}{2}} \quad (69)$$

La perte de corrélation entre deux vecteurs est calculée suivant:

$$\boldsymbol{\sigma_{xy}} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{\mu_x})(\boldsymbol{y_i} - \boldsymbol{\mu_y}) \quad (70)$$

La fonction de comparaison de luminance notée par *L*(*x*, *y*) est une fonction de $\boldsymbol{\mu_x}$ et $\boldsymbol{\mu_y}$. Le contraste entre deux faisceaux spatiaux noté par *C*(*x*, *y*), comparent les variances de $\boldsymbol{\sigma_x}$ et $\boldsymbol{\sigma_y}$. La troisième fonction compare les structures des deux vecteurs spatiaux donnés par la fonction *S*(*x*, *y*) en fonction de deux vecteurs normalisés $(\boldsymbol{x} - \boldsymbol{\mu_x})$ et $(\boldsymbol{y} - \boldsymbol{\mu_y})$. Finalement la fonction mesurant la similarité, est une fonction de l(x, y), du contraste c(x, y), et de structure S(x, y). elle prend la forme :

$$SSIM(x, y) = f[L(x, y), C(x, y), S(x,y)] \quad (71)$$

Une propriété intéressante est que les trois composants de la fonction *f*(·) sont relativement indépendants. Aussi, la fonction *S*(*x*, *y*) doit satisfaire les conditions suivantes :

1) SSIM(x, y)= SSIM (y, x)
2) SSIM(x, y)≤ 1
3) SSIM(x, y)=1 si et seulement si x=y ($x_i = y_i$ ∀ i= 1,…n)

Wang et al ont proposé une fonction de comparaison de luminance de la forme :

$$L(x, y) = \frac{2\mu_x+\mu_y+C_1}{\mu_x{}^2+\mu_y{}^2+C_2} \quad (72)$$

La fonction de comparaison de contraste prend une forme similaire :

$$C(x, y) = \frac{2\sigma_x+\sigma_y+C_1}{\sigma_x{}^2+\sigma_y{}^2+C_2} \quad (73)$$

Notons que la corrélation entre les deux vecteurs est une mesure simple et effective de la similarité structurelle. Alors la fonction de comparaison structurelle est donnée par :

$$S(x, y) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad (74)$$

Avec $\boldsymbol{C_1 = (k_1 \cdot L)^2}$, $\boldsymbol{C_2 = (k_2 \cdot L)^2}$, et $\boldsymbol{C_3 = (C_2/2)}$ sont des constantes; $\boldsymbol{k_1}$ est une constante de très faible valeur. La constante$\boldsymbol{(k_1 L)^2}$ évite l'instabilité de la fonction de comparaison lorsque $\boldsymbol{\mu_x{}^2+\mu_y{}^2}$ sont très proches de zéro. Dans le cas des images à niveau de gris, la constante est *L* = 255.

Finalement, la mesure de similarité résultante des trois comparaisons est donnée par une expression simplifiée de l'index de similarité structurelle entre x et y :

$$\text{SSIM}(x, y) = [\boldsymbol{L(x,y)}]^{\alpha}\,[\boldsymbol{C(x,y)}]^{\beta}\,[\boldsymbol{S(x,y)}]^{\gamma}$$

avec $\boldsymbol{\alpha}$ >0, $\boldsymbol{\beta}$ >0 et $\boldsymbol{\gamma}$ >0 sont des paramètres utilisés pour ajuster l'importance relative de chaque composant. Pour simplifier, il est commode de prendre $\boldsymbol{\alpha} = \boldsymbol{\beta} = \boldsymbol{\gamma} = 1$. Une deuxième forme plus simplifiée de SSIM est aussi utilisée :

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x{}^2+\mu_y{}^2+C_2)(\sigma_x{}^2+\sigma_y{}^2+c_2)} \quad (75)$$

Dans [69] les propriétés statistiques sont calculées sur des fenêtres locales de 11×11. Ces fenêtres sont souvent choisies des gaussiennes symétriquement circulaires et normalisées avec un écart. Les évaluations des propriétés statistiques sont alors modifiées comme suit:

$$\boldsymbol{\sigma_{xy}} = \frac{1}{n-1}\sum_{i=1}^{n} \boldsymbol{w_i(x_i-\mu_x)(y_i-\mu_y)} \quad (76)$$

$$\text{Et } \boldsymbol{\sigma_x} = \left(\frac{1}{n-1}\sum_{i=1}^{n} \boldsymbol{w_i(x_i-\mu_x)^2}\right)^{\frac{1}{2}} \quad (77)$$

Alors la similarité est déterminée à travers les M fenêtres à travers lequel l'index SSIM est appliqué :

$$\text{MSSIM}(x, y) = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{SSIM(x_j, y_j)} \quad (78)$$

3.5 CRITÈRES BASÉS SUR LES VECTEURS SPECTRAUX

Un ensemble de critères classiques peut être proposé:

3.5.1 Distance Euclidienne Minimale (EMD)

$$d_2(v, v') = l_2(v - v') = \sqrt{\sum_{\lambda=1}^{n_\lambda}(v(\lambda)-v'(\lambda))^2}$$
$$= \sqrt{\sum_{\lambda=1}^{n_\lambda}(I(x,y,\lambda)-I'(x,y,\lambda))^2} \quad (79)$$

Avec : EMD= $min_v d_2$ (v ,v')

3.5.2 Average RMSE : La Moyenne du RMSE

Average RMSE (A-RMSE) : Ce critère réalise une moyenne du RMSE calculé pour chaque vecteur [75, 79].

$$\text{ARMSE} = \frac{1}{n_x n_y} \sum_{x,y} \sqrt{\frac{1}{n_\lambda} (I(x,y,\lambda) - \bar{I}(x,y,\lambda))u^{\,2}} \quad (80)$$

$$= \frac{1}{n_x n_y \sqrt{n_\lambda}} \sum_{x,y} l_2\,(\text{v}\,, v')$$

3.5.3 Peak RMSE (P-RMSE) : Le Maximum du RMSE

Ce critère représente le RMSE maximum calculé selon les vecteurs.

$$\text{PRMSE} = max_{x,y} \sum_{x,y} \sqrt{\frac{1}{n_\lambda} (I(x,y,\lambda) - \bar{I}(x,y,\lambda))^2} \quad (81)$$

$$= \frac{1}{\sqrt{n_\lambda}} max_{x,y}\{l_2\,(\text{v} - v')\}$$

3.5.4 Maximum Spectral Angle : Angle Spectral Maximal

Le Maximum Spectral Angle (MSA) est un critère utilisé dans de nombreux articles dont [75, 76, 77]. Le MSA est une mesure d'écart entre deux vecteurs. Pour chaque pixel, on mesure l'angle entre le vecteur original et le vecteur reconstruit. Il est plus rigoureux de parler de Spectral Angle pour définir la mesure entre deux vecteurs, le Spectral Angle Mapper désignant l'application de classification correspondant à cette mesure.

$$\text{SA (v, } v') = cos^{-1}\left(\frac{<\text{v},v'>}{\|v\|_2.\|v'\|_2}\right) \quad (82)$$

Où encore directement en utilisant les valeurs du cube hyperspectral

$$\text{SA (v,}v') = cos^{-1}\left(\frac{\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda).\hat{I}(x,y,\lambda)}{\sqrt{\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda)^2 \sum_{\lambda=1}^{n_\lambda} \hat{I}(x,y,\lambda)^2}}\right) \quad (83)$$

Ce critère peut être considéré en moyenne, ou bien en prenant le maximum de manière à borner l'erreur. Il faut toutefois noter que ce critère mesure les distorsions spectrales et que les distorsions radiométriques n'impliquent pas nécessairement de distorsions spectrales.
On prendra ici le maximum, avec :

$$\text{MSA} = max_{x,y}\{SA_{x,y}\} \quad (84)$$

3.5.5 Distance Mahalanobis

Cette distance est un paramètre statistique entre deux vecteurs v et v'

$$d_S(\text{v}\,, v') = \sqrt{(\text{v} - v')^T S^{-1} (v - v')} \quad (85)$$

S étant la matrice de covariance. Cette distance est couramment utilisée en classification de manière générale. Egalement, dans le domaine particulier de la classification en hyperspectral.

3.6 CRITÈRES POUR LES EVENEMENTS RARES

Un événement rare est par définition, un événement ayant une faible probabilité d'apparition. Les événements rares peuvent potentiellement se produire dans de nombreuses applications, notamment dans l'analyse d'images hyperspectrales ou dans la détection de ces vecteurs particuliers dans diverses applications [114, 115, 116]. Ces événements deviennent parfois d'une importance primordiale. Mais en raison de leur nature sporadique, les vecteurs spectraux portant des informations associées à des événements rares se trouvent souvent dans un grand ensemble de signaux non pertinents et ne sont pas facilement accessibles. Le local des anomalies préservé est très utile dans certaines situations critiques, comme dans les applications liées à la défense nationale et l'application de la loi... etc. Lors la réduction de la dimensionnalité, il est donc essentiel de mesurer le taux de la préservation de la quantité d'information de ces vecteurs rares afin de permettre la détection de cibles et la discrimination dans un contexte non supervisé, lorsque les capacités de stockage et de transmission de données sont limitées (par exemple de petites plates-formes aéroportées).

Mathématiquement, Soit la matrice Y: [PxN] contenant des vecteurs rares. Il existe une décomposition Y = $\prod[Y_1, Y_2]$, avec Y_1 et Y_2 sont deux sous matrices de Y, de dimensions p x N_1et p x N_2, telle que N_1+ N_2= N, $N_1 \gg N_2$, et rang(Y_1) ⊂ rang(Y). Aux termes de cette propriété la matrice Y contient des vecteurs rares si le nombre de colonnes dans la matrice Y qui sont linéairement indépendantes avec toutes les autres colonnes est relativement faible.

En termes de probabilité, lorsque l'on dispose d'un jeu de données, les événements rares apparaissent donc peu souvent, pour cela nous supposons que les $\{v_i\}_{i=1}^{k-1}$ sont des réalisations indépendantes d'une même variable aléatoire, de loi de probabilité P définie sur X. Nous pouvons alors dire que l'événement $v_k \in X$ est un événement rare par rapport à la loi P si :
v_k Est rare ⟺ P (v_k) $< \gamma_{rare}$ Avec γ_{rare} ($0 < \gamma_{rare} << 1$) est un seuil de décision

Ce paragraphe fournit un cadre statistique d'un critère local pour mesurer le taux de la préservation de ces événements rares durant la réduction de la dimensionnalité. L'objectif principal de ce cadre est de proposer une procédure adaptative, se concentrant dans les ressources d'informations sur les segments de l'ensemble de données qui portent l'information pertinente des événements rares. Un accent particulier est mis sur les vecteurs locaux dans chaque bande spectrale dans le but de mesurer le taux de la préservation des vecteurs rares.

3.6.1 Etat de l'Art

Un certain nombre d'approches ont été proposées dans la littérature [118, 119] pour l'estimation d'un sous-espace signal et de son rang, sous l'hypothèse que les données soient indépendantes, fixes, de moyennes nulles et gaussiennes. Il est montré dans [120] que l'analyse en composantes principales (ACP) est optimale dans le sens de maximum-vraisemblance pour l'estimation d'un sous-espace de signal. L'auteur détermine l'espace signal en minimisant la norme l_2 de résidus appartenant au sous-espace complémentaire, qui peut être obtenu par décomposition en valeurs singulières (SVD). Nascimento et Dias [121], proposent une nouvelle approche nommé HySime, conçu pour déterminer à la fois le

sous-espace signal et son rang dans l'imagerie hyperspectrales. La méthode passe par trois étapes: 1) Définition du signal et calcul des matrices de covariance du bruit, 2) Estimation du rang du signal pour trouver le sous-ensemble des valeurs propres qui représente le mieux le sous-espace, dans le-sens de la valeur moyenne des données, 3) Le sous-espace signal est obtenu en appliquant SVD à bruit réduit sur la matrice de covariance des données.

Oleg Kuybeda [57] a abordé le problème de minimisation de la redondance de données à grande dimension en préservant les vecteurs rares. Il a proposé un algorithme, nommé Maximum Orthogonal-Compléments (MOCA) pour déterminer le sous espace d'un signal et son rang, en gardant à la fois les vecteurs abondants et rares par l'utilisation d'une norme $L_{2,\infty}$. Cette norme intègre deux objectifs : L_2 pour minimiser les vecteurs abondants et L_∞ pour minimiser les vecteurs rares. Deux aspects ont été utilisés :

- Estimer le sous espace signal $\widehat{S_k}$= argmin $\|P_{L\perp X}\|_{2,\infty}^2$en déterminant a priori son rang, en minimisant le maximum des données de la forme résiduels suivante: $\|X\|_{2,\infty}^2 = max_{i=1,..,n}\ \|x_i\|_2$
- Utiliser l'algorithme Maximum Orthogonal-Compléments MOCA pour grouper les vecteurs d'anomalies dans Ω_h afin de les représenter dans un sous espace vectoriel des résidus.

Dans [117], les auteurs ont montré que la norme $l_{2,\infty}$ peut être utilisée efficacement pour la détection d'anomalies. Cependant, l'algorithme développé dans [117], nommé Min-Max-SVD (MX-SVD), pour l'estimation d'un sous-espace signal minimise seulement la norme $l_{2,\infty}$ pour les résidus de fausse déclaration. Une nouvelle approche optimale pour la minimisation de la norme $l_{2,\infty}$ nommé Maximum Orthogonal Optimale Subspace Estimation (MOOSE) est développée dans [7]. L'optimisation est réalisée via une méthode d'apprentissage du gradient effectuée sur l'ensemble des sous-espaces de dimension n dans R^m, m>n. Les résultats de l'application MOOSE, MX-SVD, et l'approche basée sur la norme $l_{2,\infty}$ sont mises en évidence sur des données hyperspectrales réelles.

Autres approches liées à la littérature de la détection ont été utilisés par les auteurs [122, 123]. Le but de la recherche est d'identifier un ou plusieurs événements rares parmi tous les n événements donné par l'ensemble de données. Les auteurs ont élaboré un processus de critère d'information optimal qui maximise la décision de la fiabilité soumis aux prescriptions locales en préservant l'information rare. Tout d'abord, les approches existantes en matière de détection séquentielle appliquée au problème de la recherche rapide et optimale est souvent un équilibre entre la fiabilité de décision et une contrainte sur le nombre d'observations. Ceci est en contraste avec la cadre de la recherche qui applique une contrainte dure sur la détection de ressources. Aussi, la détection séquentielle vise souvent à identifier tous les événements rares ou seulement un événement rare avec le test du rapport de probabilité (ESCR) [124] et la somme cumulée (CUMSUM) [125].

3.6.2 Proposition d'un nouveau critère :

Dans cette partie, nous proposons une nouvelle approche statistique pour estimer la pertinence d'une variable, définit par un rapport de similarité locale entre une variable et son voisinage. Soit un jeu de données $\{v_i\}_{i=1}^{k-1}$ qui sont des réalisations indépendantes d'une même image hyperspectrale. L'élément x_i (i=1,.., n<<k) est un évènement rare dans cette réalisation. On définit V(x_i) comme étant le voisinage d'ordre m<n de cet élément,

$\bar{\mu} = \mu - \{x_i\}$ la moyenne représentative de V(x_i) non inclus x_i et σ son ecart-type. Cette mesure est définie par :

$$\text{Pertinence } (x_i) = \frac{\|x_i - \bar{\mu}\|^2}{\mu^2 + \sigma^2} \quad \forall \text{ i=1,...,n}$$

$$\text{Pertinence } (x_i) = \begin{cases} 0 < P(x_i) < \varepsilon = 0.5\text{: Evènement } x_i \text{ non rare} \\ P(x_i) \geq \varepsilon = 0.5\text{: Evènement } x_i \text{ rare} \end{cases}$$

Deux exemples numériques sont présentés ci-dessous pour illustrer la représentation des vecteurs rares dans une matrice de taille 5x5 et 3x3. L'évènement rare dans notre contexte est un nombre différent de son voisinage qui l'entoure. Une faible valeur pour cette indice, indique une absence locale de l'évènement rare.

<u>Exemple 1:</u>

…	…	…	…	…	…	…
…	1	1.1	1.4	1.6	1.3	…
…	1.2	1.5	1.1	1.4	1.2	…
…	1.8	10	1.8	1.3	1.6	…
…	1.1	1.9	1.7	1.2	1.5	…
…	1	1.8	1.4	1.9	1.7	…
…	…	…	…	…	…	…

P(10) = $(10- 1.4375)^2$/ 6.062=12.094

…	…	…	…	…	…	…
…	1	1.1	1.4	1.6	1.3	…
…	1.2	1.5	1.1	1.4	1.2	…
…	1.8	2.5	1.8	1.3	1.6	…
…	1.1	1.9	1.7	1.2	1.5	…
…	1	1.8	1.4	1.9	1.7	…
…	…	…	…	…	…	…

P(2.5) = $(2.5- 1.433)^2$/ 2.3036= 0.49

<u>Exemple 2 :</u>

…	…	…	…	…
…	1	1,1	1,4	…
…	1,2	2	1,5	…
…	1	1	1	…
…	…	…	…	…

P(2)= $(2-1,5)^2$/ 1,6628= 1,85

…	…	…	…	…
…	1	1,1	1,4	…
…	1,2	1.3	1,5	…
…	1	1	1	…
…	…	…	…	…

P (1.3)= (1.3- 1.15)^2/ 1,398= 0.016

3.6.3 Application sur une image réelle :

En conséquence, une mesure quantitative basée sur la pertinence de x_i sera utilisée pour évaluer le taux de préservation de cet évènement rare x_i. Cette mesure, notée T (x_i), est calculée par le rapport de la pertinence avant la réduction $\mathrm{P}(x_i)_{Avant}$ sur la pertinence après la réduction $\mathrm{P}(x_i)_{Après}$. Ce taux de préservation est définie ainsi par :

$$\mathrm{T}\,(x_i) = \frac{\mathrm{P}(x_i)_{Avant}}{\mathrm{P}(x_i)_{Après}} \quad , \ \forall\ i=1,\ldots, n$$

$$\mathrm{T}\,(x_i) = \begin{cases} 0<\mathrm{T}(x_i)< 0.5 & \text{l'évènement rare } x_i \text{ est non preservé} \\ 0.5<\mathrm{T}(x_i)\leq 1 & \text{l'évènement rare } x_i \text{ est preservé} \end{cases}$$

Nous proposons dans ce paragraphe de mener un essai pour ce critère sur une image réelle. Cette image est récupérée sur le site de M. David Foster :
http://personalpages.manchester.ac.uk/staff/david.foster/
Elle représente une scène de verdure dans lequel est intégré une fleur considérée comme rare dans le contexte générale (**Figure 17**). Cette scène de 820×820×31 est composée de 31 bandes chacune de taille 820×820 pixels.

Figure 17. Image de verdure avec évènement rare

Nous procédons par l'application de deux méthodes de projection. L'une est assez simple qui est la méthode PCA et l'autre est une méthode de réduction connue par son potentiel non linéaire et qui est la méthode Sammon. Nous avons scanné tous les réductions possibles pour cette image en partant de 30 bandes jusqu'à atteindre une seule bande. Les courbes suivantes (**Figure 18**) illustrent bien le pouvoir de chacune de ces deux méthodes à la préservation de l'évènement rare comme conçu par notre définition.

Il est clair que la méthode PCA perd tout son potentiel sur une image «difficile» contrairement à la méthode Sammon. La difficulté provient du fait qu'une verdure débordante et assez monotone dans l'image devient très difficile à détecter au cours de la projection. Raison pour laquelle une méthode comme la Sammon devient indispensable pour ce style d'exercice. Cependant, l'image a été prise pour ce style d'essai et son caractère reste proche du contexte mathématique dans lequel nous cherchons à tester une certaine définition.

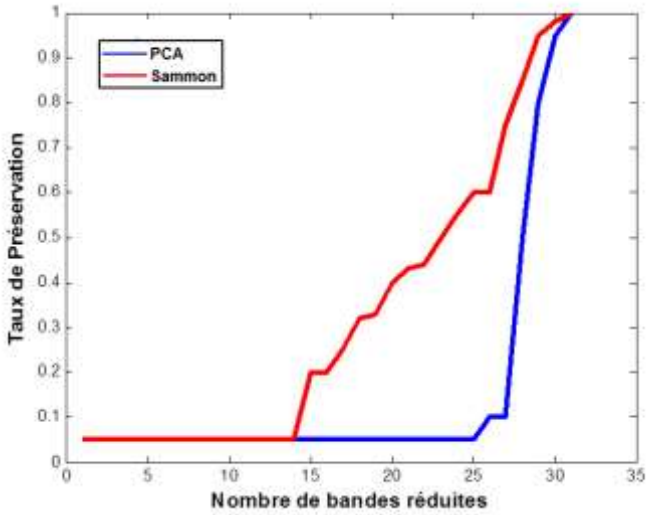


Figure 18. Facteur de pertinence pour deux méthodes de réduction

3.7 CONCLUSION

Ce chapitre introduit bien les critères relatifs à l'évaluation d'une méthode de réduction selon deux catégories : Locale ou Globale. Cette dernière est aussi divisée en quatre classes : Pertes d'information, Stabilité, Similarité et Evènement rare. Cet ensemble de critères doit nous permettre par la suite d'évaluer les distorsions, la sensibilité à l'erreur, la dégradation de donnée initiales, le taux de préservation d'un évènement quelconque,… lors de l'application dans un algorithme quelconque pour la réduction de données.

Chacun de ces critères a ses particularités d'avantages et d'inconvénients. Les critères relatifs à la perte d'information sont directement tirés des mesures de la structure des données, et permettent d'analyser la situation de l'ensemble de données par rapport à son voisinage avant et après réduction de la dimensionnalité. Les critères relatives à la stabilité et à la similarité sont fortement liés à la structure géométrique de l'image, et permettent de décrire la nature et l'importance des dégradations sur l'image reconstruite, mais qui ne reflètent pas un indicateur précis sur l'état local de données surtout par rapport à la préservation de l'événement rare. La cohérence fonctionnelle global/local s'adapte mieux à l'étude structurale de données mais leur utilisation reste rare pour l'évaluation des techniques de réduction de la dimension.

Ce chapitre était également l'occasion d'adapter une mesure de similarité mais dans un contexte locale permettant ainsi d'évaluer le taux de préservation d'un évènement rare dans un processus de réduction de dimension.

CHAPITRE 4
ETUDE COMPARATIVE DES METHODES DE REDUCTION DE LA DIMENSIONNALITE

4.1 INTRODUCTION

L'intérêt pour les données hyperspectrales est croissant au cours des dernières années. Traiter ces quantités de données importantes présente un défi et la plupart des applications sont encore en développement. Pour définir un système de réduction de la dimension de données efficace sans perdre le potentiel important de ces images, la définition de critères de qualité est une étape indispensable.

Ce chapitre présente trois études comparatives sur la stabilité, la similarité et sur la perte d'informations des méthodes non paramétriques et non supervisée de la projection et de la sélection des bandes utilisées dans la réduction de la dimensionnalité. Un nombre important de critères qualité basés sur les normes et les corrélations est défini, puis évalué. Finalement, cinq critères sont retenus pour donner une bonne estimation de l'impact des dégradations sur des applications classiques portant sur la réduction de la dimension. La valeur ajoutée de ces critères peut être illustrée à l'évaluation de la performance de la réduction, si l'on considère la stabilité des deux catégories de méthodes de sélection des bandes et des méthodes de projection. Les performances d'une méthode sont vérifiées sur des ensembles de données synthétiques et naturelles.

Une sélection des courbes obtenues par des tests sur les critères de «Normalized Cross Correlation» (NCC) et de «Structural Content» (SC) est définie ci-dessous. Dans ce qui suit, sont aussi expliquées l'influence du taux de la perturbation introduit à l'observation d'un pixel des données lors de phénomène de la réduction et les conclusions obtenues à partir des simulations effectuées.

En conséquence, une méthode de projection Multidimensional Scaling (MDS) intégrée à un algorithme de classification BandClustering (BandClust) sera présentée dans ce chapitre pour la réduction de la dimensionnalité. Des exemples sont donnés pour démontrer l'originalité et la pertinence de l'hybridation (BandClust/MDS) de l'analyse effectuée dans les parties suivantes pour une meilleure stabilité et similarité.

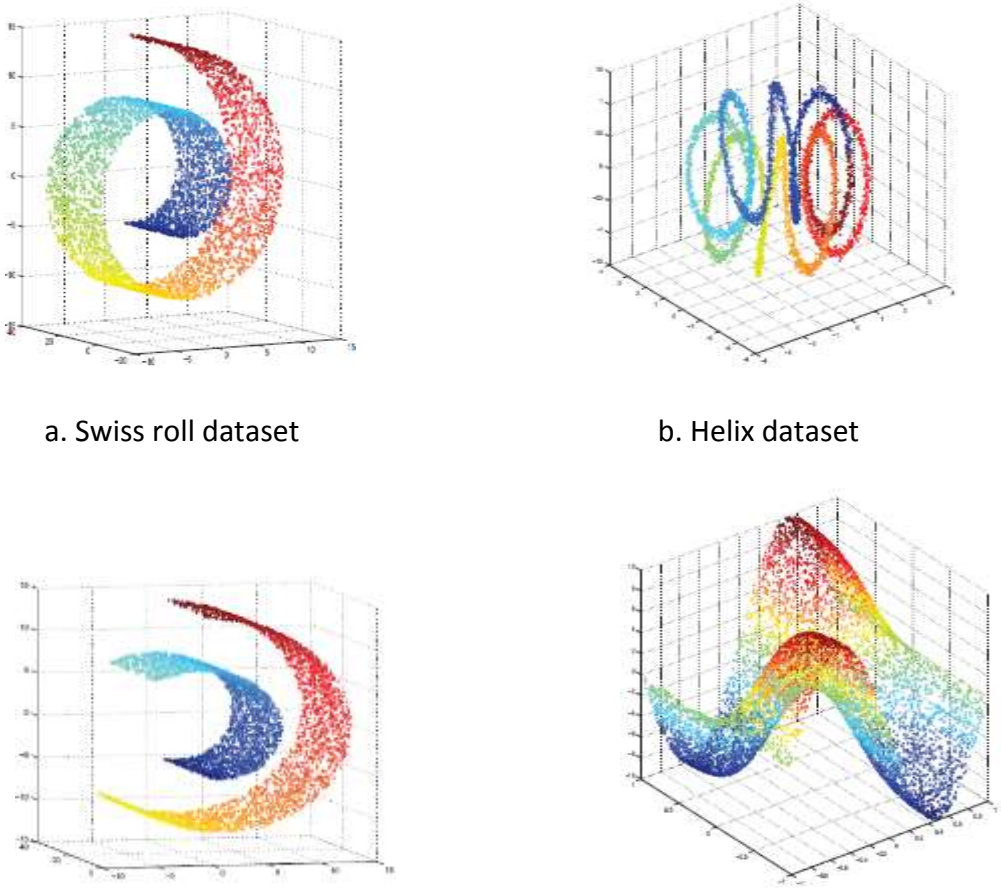
4.2 DONNEES EXPERIMENTALES

Pour valider notre approche, différentes images naturelles et synthétiques seront employées pour évaluer la performance de méthodes de réduction de la dimensionnalité par des critères sur la stabilité, la similarité et la préservation de la quantité d'information. Cependant, quelques jeux de données semblent assez populaires sont présentés brièvement dans ce qui suit.

Plusieurs images hyperspectrales sont sélectionnées, sur lesquelles les algorithmes seront évalués. Ces images sont riches en information et représentent des zones variées, qui peuvent être uniformes ou hétérogènes (désert, ville, plans d'eau . . .).

4.2.1 Application aux données synthétiques

Dans cette catégorie, nous présentons les quatre ensembles de données artificielles sur lesquels nous aurons effectué l'étude expérimentale **(Figure 19)** : l'ensemble des données Swiss roll (a), l'ensemble des données hélice (b), l'ensemble des données TwinPeaks (c) et l'ensemble des données broken swiss roll (d).
Référence: http://ict.ewi.tudelft.nl/ lvandermaaten/



a. Swiss roll dataset

b. Helix dataset

c. Twinpeaks dataset

d. Broken Swiss roll dataset

Figure 19. Les Quatre images artificielles.

4.2.3 Application à l'imagerie biomédicale

Le second jeu de données concerne l'image biomédicale multispectrale. Il est obtenu à partir d'un débit élevé à cristaux liquides à filtre accordable (LCTF) qui a permis la capture de 33 bandes spectrales entre 500 nm et 650 nm avec un pas 4,5454 nm entre chaque bande successive (**Figure 20**). Par conséquent, il y a 33 bande-images séquentielles pour un développement d'un tissu anormal correspondant à un cancer malin. Ce type d'image médicale est de textures multi-spectrales [82] visible grossissement avec une microscopie optique par l'objectif moyen X25.

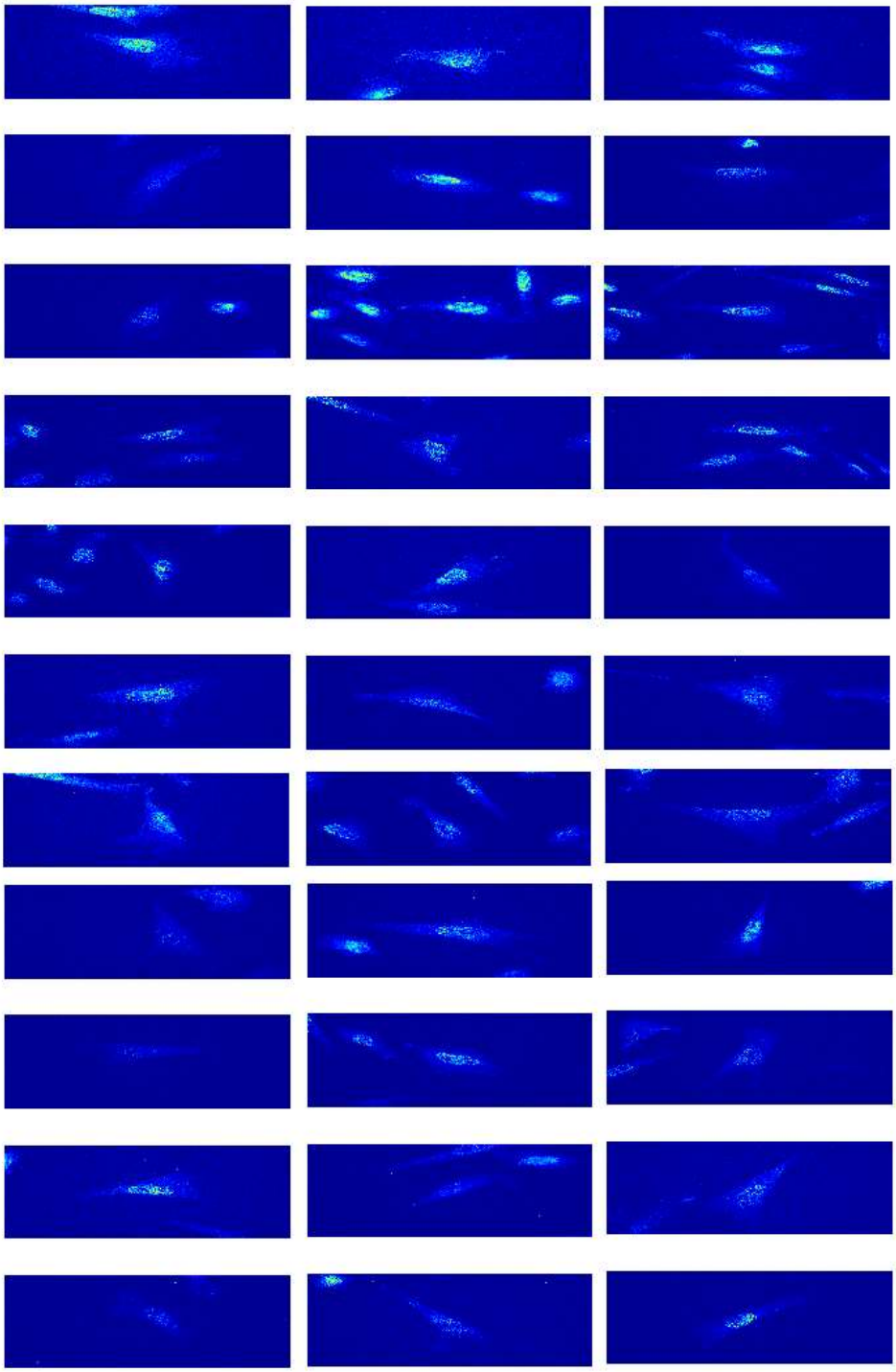


Figure 20. Type de cellules anormales pour le carcinome de côlon

4.2.2 Application à l'Indiana Pines et Salinas

Le premier type de données hyperspectrales utilisé dans cette expérience est l'image **Indian Pines** réalisé en Juin 2012 par un spectromètre AVIRIS (**Figure 21**). L'image est composée de 220 bandes spectrales dans la gamme 0.4-2.5 µm, chaque bande est de dimension spatiale 145 x 145 pixels et de résolution spatiale 20 m par pixel. Cet ensemble de données a été largement utilisé dans la littérature pour la réduction de dimension hyperspectrale [61]. La scène Pines Indian contient l'agriculture, des forêts ou d'autres plantes vivaces naturelles, deux grandes routes à double voie, une voie ferrée, ainsi que

certains logements de faible densité, d'autres structures construites, et des petites routes. Indiana Pines est disponible via le site de Pursue's univeristy.

Une scène similaire a été recueillie par le capteur AVIRIS 224-bande sur la vallée de Salinas, en Californie, et se caractérise par une haute résolution spatiale (3,7 mètres de pixels). La zone couverte comprend 512 lignes par 217 échantillons. Cette image comprend des légumes, des sols nus et un champ vignoble (**Figure 22**).

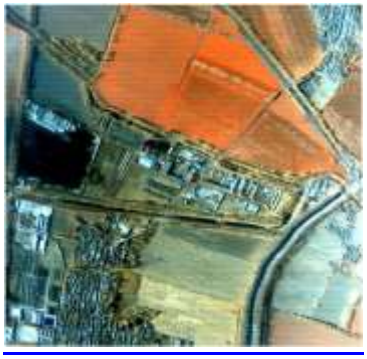


Figure 21. Image de l'Indiana Pines

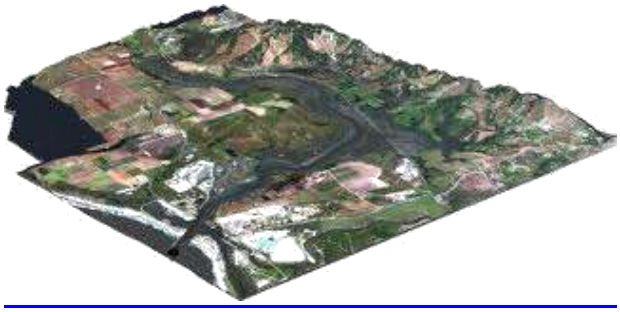


Figure 22. Image du Vallée Salinas

4.2.4 Image contenant un rangement d'objets

Comme ensemble de données, nous avons aussi utilisé l'image hyperspectrale d'une scène naturelle contenant un rangement d'objets divers (**Figure 23**). Elle est acquise par un appareil photo numérique monochrome (Pulnix TM-1010, Pulnix America Inc, Sunnyvale, CA, USA), avec un capteur CCD de 1024 par 1024 pixels et 10-bit de sortie (pour un transfert rapide des données à l'ordinateur, seule la partie centrale de chaque image, ne dépassant pas 820 par 820 pixels, a été capturé habituellement).

Figure 23. Image : Rangement des objets.

L'appareil a un obturateur électronique avec une minuterie commandée par un signal externe. Cette donnée se compose de 33 bandes spectrales dans la gamme [400-720] nm, après retrait de bandes bruitées. La dimension spatiale de cette image est de 820 × 810 pixels. La gamme de longueur d'onde de 400-720 nm a été échantillonnée à des intervalles de 10 nm.

4.3 ETUDE COMPARATIVE SUR LA PERTE D'INFORMATIONS

Cette étude a permis de classer toutes les techniques mathématiques de réduction des dimensions en deux catégories. Les courbes sont tracées pour les 6 critères suivants : l'entropie, la variance, 1ere et 2eme dérivée spectrale, l'erreur moyenne absolue et quadratique. Chaque courbe est tracée en représentant la valeur du critère qualité en ordonnée et en abscisse la valeur du taux de réduction (dimension spectral). Les performances minima et maxima sont sélectionnées pour chaque dégradation (la sélection est faite par rapport aux résultats sur une image artificielle).

Une sélection des courbes obtenues pour tous les critères testés est présentée ci-dessus. Dans ce qui suit, sont expliquées les interprétations que l'on peut en tirer avec les conclusions obtenues à partir des simulations effectuées.

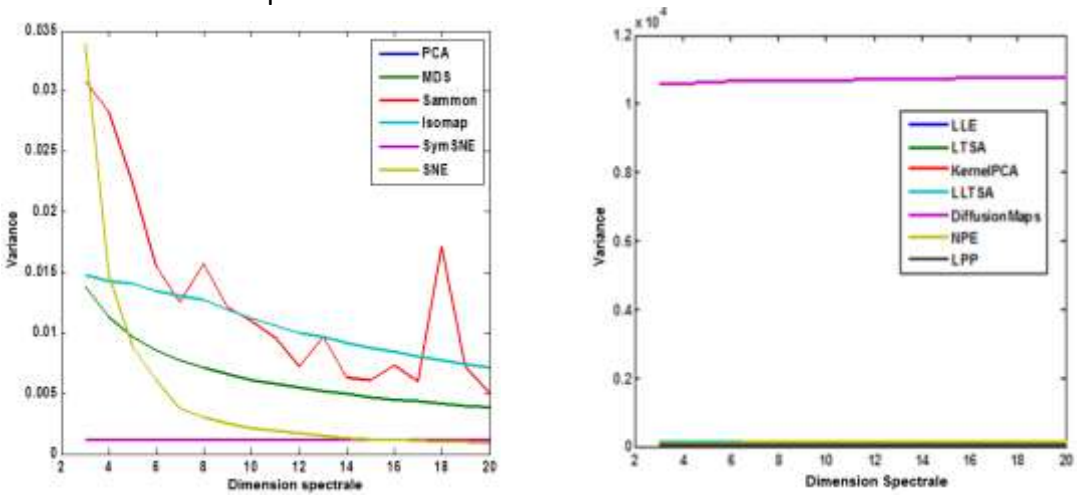


Figure 24. 1ere et 2eme catégorie de courbes sur la perte d'information des méthodes de projection

Les tests sur le critère de la variance montrent globalement un comportement proportionnel à la dimension spectrale pour les 13 méthodes (**Figure 24**). On peut aussi noter une certaine dégradation aléatoire qui semble être caractéristique pour certaines méthodes lors de ce phénomène de réduction comme Sammon après avoir gardé une bonne quantité de données et qui est assez proche de SNE. Pour l'ACP, Isomap, MDS et SymSNE, la courbe de variance testée garde une quantité de l'information assez moyenne presque constante relativement au taux de réduction. Nous pouvons ainsi noter une croissance rapide reflétant la stabilité de ce critère sur les quantités de données gardées par cet ensemble de méthodes. Enfin, l'étude sur NPE, DiffusionMaps, LLTSA, LTSA, LLE, KACP et LPP montrent des mauvaises performances par rapport aux autres méthodes de réduction interprétées ci-dessus malgré la stabilité des valeurs de variance assez petite dans tout le seuil de réduction [2; 20].

Dans la **Figure 25**, on remarque bien des performances similaires pour les méthodes LTSA, LLTSA, NPE, Kernel-ACP, LLE, LPP et DM, d'une part et pour les six autres méthodes SNE, SymSNE, ACP, MDS, Isomap et Sammon d'autres part. Ainsi, les trois critères (MAE, MQE et MAD) reflètent sur la méthode Sammon une bonne performance par rapport aux autres méthodes de la même catégorie. Egalement, ces deux critères sont moins efficaces sur SymSNE, MDS, DM et SNE en montrant une perte des informations lors d'une chute assez faible. Globalement, la 2eme catégorie contenant les six méthodes reflète assez moins la sensibilité de ces critères sur les quantités d'informations. Par contre, les valeurs moyennes, dans la 1ere catégorie, sont trop faibles sous un intervalle de taux de réduction [2-20]

marquent la capacité de ces sept méthodes citées si dessus d'avoir préservé une quantité d'informations favorables. En conclusion, le critère d'erreur moyen absolu et d'erreur quadratique moyenne montrent leurs efficacités sur Sammon et partiellement sur le SNE lors de la réduction mieux que les autres méthodes.

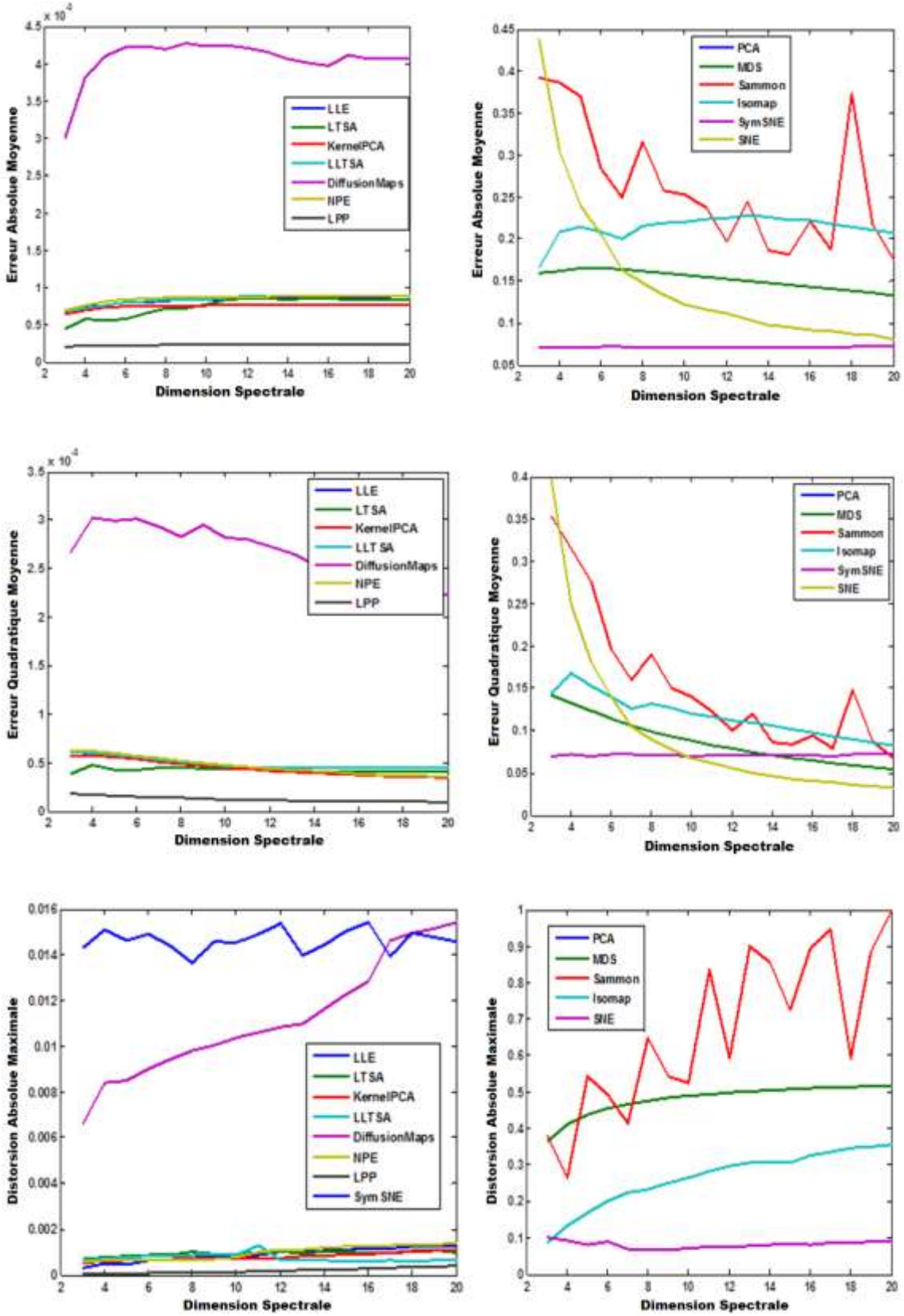


Figure 25. 1er et 2 eme catégorie des méthodes de projections testés avec MAE, MQE et MAD

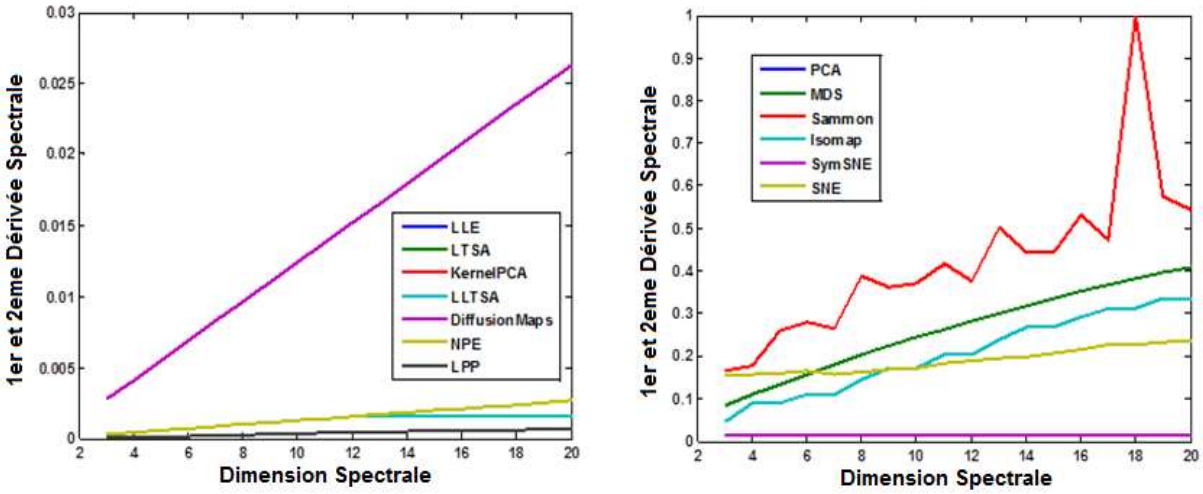


Figure 26. 1er et 2eme catégorie des méthodes de projections testés avec 1er , 2eme derivée spectral

La **figure 26** concerne la 1ere et 2eme dérivée spectrale. Ces deux critères peuvent bien être expliqués comme un taux de variation dans l'information stockée entre deux bandes. Une augmentation de cette valeur (après réduction) peut dévoiler le besoin d'ajouter le nombre de bandes pour ne pas perdre toutes les informations d'un événement donné. Ainsi pour une réduction de la taille d'une image hyper spectrale, on peut bien prévoir l'augmentation de ce critère plus le nombre des bandes est diminuée. Cependant, ce qu'on doit rechercher dans la comparaison entre méthodes de réduction est un maximum pour ce critère. Ceci doit s'interpréter comme un maximum d'information par bande sur un minimum de nombre de bandes. Tenant à cette explication pour ce critère, le Sammon sera classé comme le plus avantageux, par rapport aux méthodes similaires MDS, S-SNE, SNE, Isomap, ACP qui viennent par la suite pas très loin de ce qui précède, et finalement les méthodes LLTSA, LTSA, NPE, LPP, LLE, KACP et Diffusion Maps sont les moins importantes.

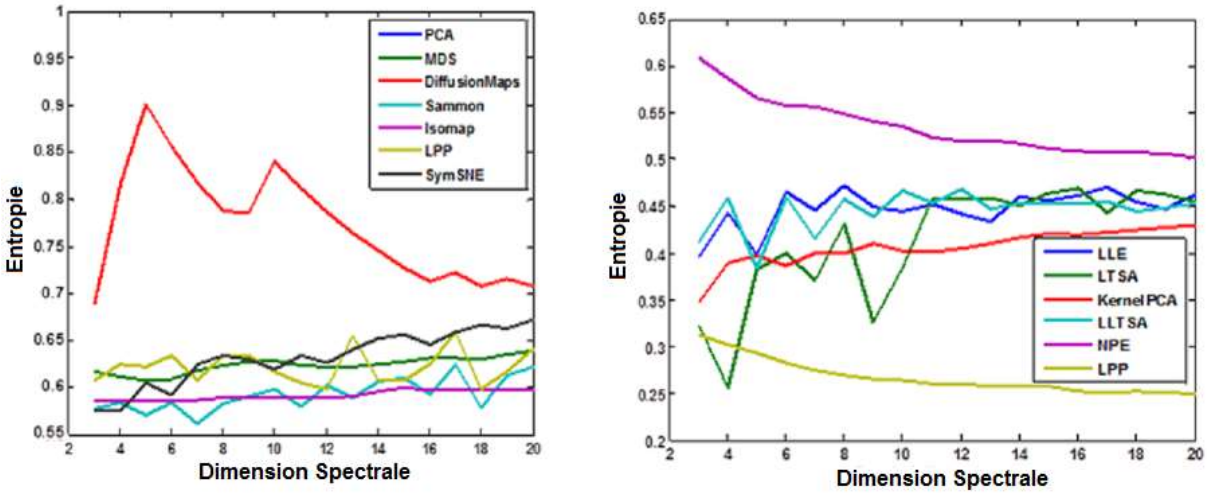


Figure 27. 1er et 2 eme categorie des méthodes de projections testés avec l'entropie.

Les courbes d'entropie (**Figure 27**) expérimentée sur 13 méthodes de réduction montrent les faits suivants: Lors du phénomène de la réduction de dimension, les méthodes LLTSA, LTSA, KACP, LPP, LLE, NPE montrent un comportement similaire : L'entropie est proportionnelle au taux de réduction de l'image. La quantité d'information maximale gardée dans Diffusion Maps est 0.96 de la valeur maximale pour un taux de réduction variant dans

[2, 20] et révèle ainsi une bonne performance par rapport aux autres techniques de réduction. Pour le PCA, MDS, Sammon, Isomap, LPP et SymSNE on remarque aussi des dégradations aléatoires lors de la réduction pouvant refléter une performance partielle de ce critère sur les quantités des données lors de sa décroissance rapide sur l'intervalle [2;20] choisi pour cette étude. La quantité d'information gardée dans LLTSA, LTSA, LLE, LPP, KACP est presque équivalente mais non constante en fonction de la dimension spectrale. On remarque des perturbations qui rendent difficile le choix d'une dimension de réduction optimale, ce qui révèle que ces méthodes sont moins importantes pour garder les quantités d'informations par rapport aux autres méthodes. En conclusion, ce critère présente sa bonne performance illustrée sur DM et un peu moins sur PCA, MDS et Sammon malgré la perte d'informations liées par une chute assez faible par rapport aux autres méthodes de réduction.

Dans le tableau 3, nous classifions les méthodes de réduction de la dimensionnalité par projection en deux catégories, la première, ce sont les bons algorithmes qui ont une valeur proche de l'unité. Cette classe contient aussi les méthodes jugées moyenne dont la valeur des critères MAE, MQE, Entropie, 1ere dérivée spectrale et variance assez loin de zéro en proportion de la dégradation de bruit et la dernière catégorie est la valeur des indices presque nulle, présentant les performances de la demande par rapport à la perte d'information observé par les méthodes de réduction de la dimension.

Tableau 3. Classification des méthodes de projection pour la réduction de dimensionnalité selon les critères relatifs aux pertes d'informations.

Classification des méthodes de réduction de la dimension par projection	
Bonne	Mauvaise
DM - PCA - MDS - Laplacian - Sammon	LPP- LTSA- LLE- KPCA-LLTSA

4.4 ETUDES COMPARATIVES SUR LA STABILITE

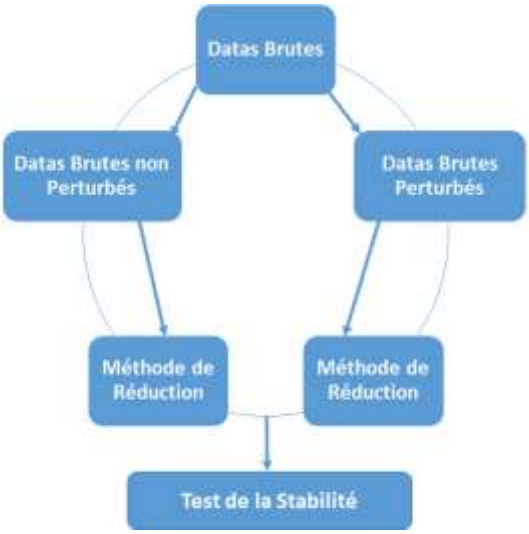


Figure 28. Méthode proposée pour l'étude de la stabilité appliquée sur les images hyperspectrales.

Le principe de cette étude est le suivant (**Figure 28**):

Une dégradation est appliquée aux images afin de simuler au mieux les pertes dues à la réduction ou à des défauts instrumentaux. Ce type de dégradation appliqué est un bruit

blanc gaussien additif dont on modifie la variance. Ce bruit modélise les défauts d'acquisition de l'appareil au niveau électronique, ainsi que le bruit de quantification. On ajoute à l'image un tirage aléatoire de ce bruit blanc. Certains critères ou certaines applications posséderont une sensibilité différente à la réalisation du bruit.

Un examen comparatif sur la stabilité inclut les techniques les plus importantes pour la réduction de la dimensionnalité linéaire/non linéaire, afin de révéler un algorithme efficace pour détecter et supprimer les bruits importants. Le problème de la réduction de la dimensionnalité peut se définir comme suit. Supposons que nous avons un ensemble de données X représentées dans la matrice n × m constitué par les vecteurs de données r xi pour i = {1, 2 ... n} dont la dimension intrinsèque r où r< m.

En pratique, le signal de perturbation ξ_i-sous-espace de vecteurs observés, doit satisfaire le modèle général suivant:

$$\xi_i = x_i + Z_i \quad i = 1 \dots n \tag{86}$$

Où $x_i \in R^m$ est le vecteur aléatoire observée, $Z_i \in$ est le bruit d'acquisition de données et / ou le modèle. Enfin, le test de la stabilité sera utilisé sur un ensemble de données et les données perturbées.

4.4.1 Méthodes de projection : Normalized Cross Correlation(NCC)

Une sélection des courbes obtenues pour les critères de l'inter-corrélation normalisée et le Structural content testés sont présentées ci-dessus. Dans ce qui suit, sont expliqués l'influence de taux de la perturbation d'un pixel introduit à l'observation des données artificielles lors de phénomène de la réduction et les conclusions obtenues à partir des simulations effectuées.

L'inter-corrélation normalisée valorise numériquement la stabilité des méthodes, plus cette valeur est proche de l'unité, plus la méthode est stable. Nous perturbons numériquement l'image de base, puis nous calculons la valeur de ce critère proportionnellement à la dégradation de bruits sur une échelle [0%; 10%]. Ces courbes NCC expérimentée sur 15 méthodes de réduction montrent les faits suivants:

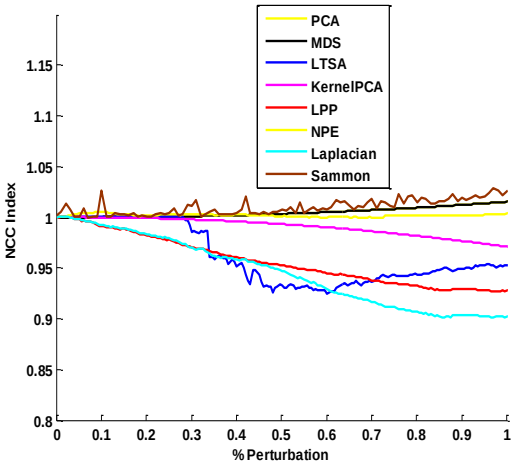


Figure 29. Méthodes de projection, 1ere Catégorie des méthodes de réduction

On remarque bien des performances similaires pour les méthodes PCA, MDS, NPE LTSA, Laplacian, KernelPCA, LPP et Sammon d'une part et pour les autres catégories des techniques non linéaires DiffusionMaps, LLTSA, LLE, KernelLDA d'autres part (Figure 27 et

28). Ainsi, Ce critère NCC testés sur NPE, MDS et PCA reflètent une bonne performance par rapport aux autres méthodes de la même catégorie Sammon, LTSA, Laplacian, LPP et KernelPCA, de valeur NCC variante autour de l'unité [1; 1.025] qui offrent une stabilité assez accrue sous une échelle de bruits [0%; 10%] suivi par d'assez faible de déformation de l'image de base. La quantité d'information gardée dans diffusionMaps, LLE, LLTSA et DM est presque équivalente mais non constante en fonction de la dégradation de bruits liés par une déformation géométrique de valeur NCC se croit tout autour de l'unité proportionnellement à la variation de bruits, Ce qui révèle la performance partielle sur les quantités de données gardées par cet ensemble de méthodes.

Enfin, l'étude sur Isomap et Landmark Isomap (**Figure 30**) montrent des mauvaises performances par rapport aux autres méthodes de réduction interprétées ci-dessus, dont certaines présentent quelques perturbations pouvant aussi diminuer la valeur globale du potentiel de ces méthodes, illustrée par la déformation d'informations liées par une chute de dégradations assez forte par rapport aux autres méthodes de réduction sur un échelle de bruits [2%;10%] de valeur NCC [0.2;1.8] assez proche de l'unité, au contraire que celles obtenus par HessianLLE (**Figure 31**), ce qui révèle l'influence totale de la perturbation de bruits sur cette méthode.

En conclusion, ce critère montre une meilleure performance partielle sur le PCA, MDS suivi par la méthode NPE. Le taux de variation de bruits égale à 40 semble être un choix optimal.

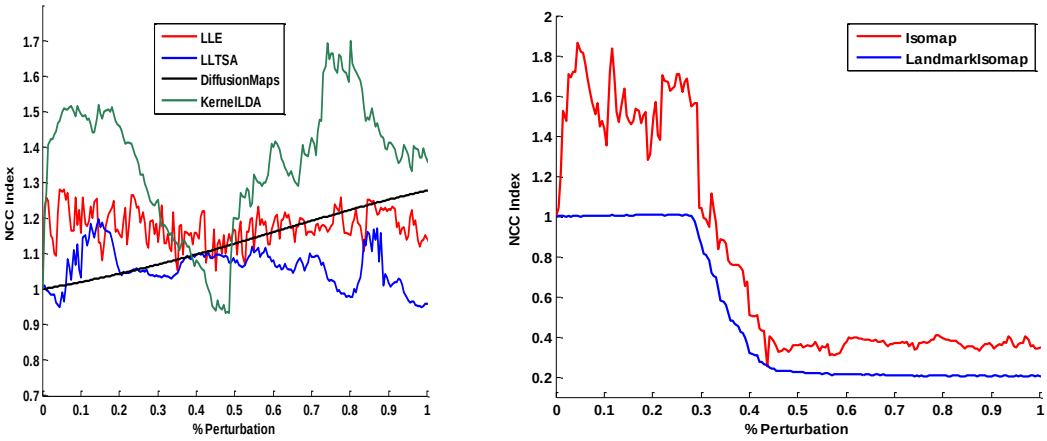


Figure 30. Méthodes de projection, 2eme Catégorie des méthodes de réduction

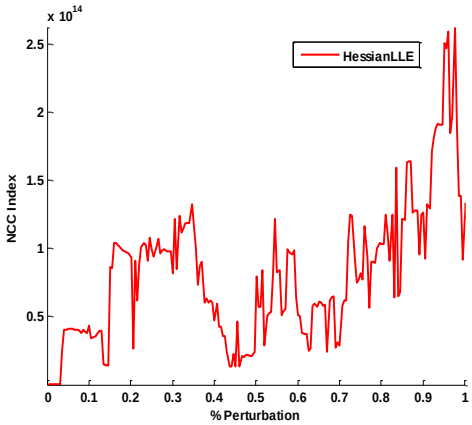


Figure 31. Méthodes de projection, 3eme Catégorie des méthodes de réduction

Dans le tableau 4, nous classifions les méthodes de réduction de la dimensionnalité par projection en trois catégories, la première, ce sont les bons algorithmes qui ont une valeur NCC et SC qui varie dans un intervalle [0,95; 1,5[, la deuxième sont les méthodes jugées moyenne dont la valeur du NCC et SC varie dans [1,5; 2[, en s'éloignant de l'unité en proportion de la dégradation de bruit, et la dernière catégorie est la valeur des indices > 2, présentant les performances de la demande par rapport à la stabilité observé par NCC et SC. La valeur de la NCC a une faible sensibilité aux dégradations de bruits faites que la valeur SC.

Tableau 4. Classification des méthodes de projection pour la réduction de dimensionnalité selon les critères relatifs à la stabilité.

Classification des méthodes de réduction de la dimension par projection		
Bonne	Moyenne	Mauvaise
PCA – MDS – NPE - LTSA- LPP KernelPCA - Laplacian-Sammon	Isomap -LLE- Kernel LDA - Landmark Isomap	HessianLLE

4.4.2 Méthodes de projection: Structural Content (SC)

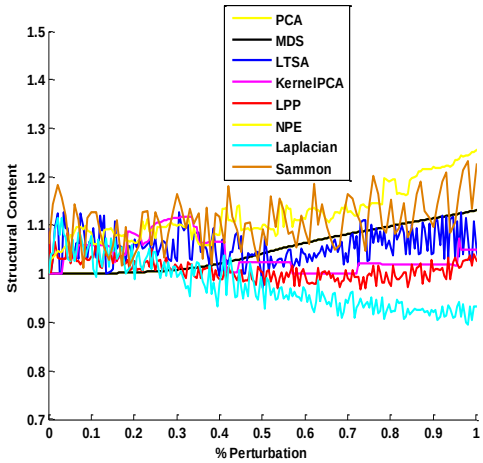
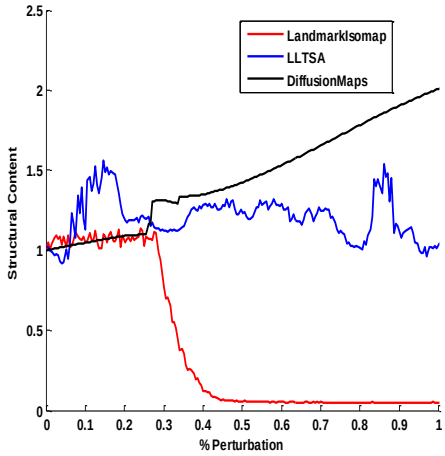


Figure 32. Méthodes de projection, 1ere Catégorie des méthodes de réduction



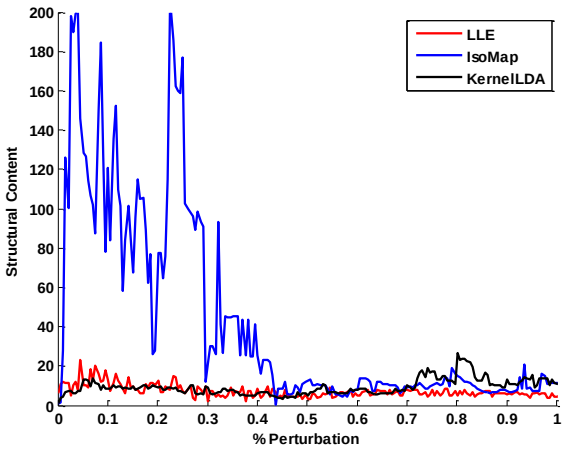


Figure 33. Méthodes de projection, 2eme Catégorie des méthodes de réduction

Ce critère semble réagir d'une manière similaire au critère de l'inter-corrélation normalisée. Ce critère reflète le potentiel d'une méthode de réduction de conserver la structure géométrique globale et locale d'informations, dont le rapport de la valeur

d'information est le plus proche de l'unité. On remarque bien des performances similaires pour les méthodes NPE, LTSA, LPP, PCA, MDS, Sammon, KPCA et Laplacian d'une part (**Figure 32**) et pour les autres catégories de méthodes avec Landmark Isomap, LLTSA, Diffusion Maps, LLE, Isomap, KernelLDA (**Figure 33**). Ainsi, les courbes SC illustrées avec les méthodes de la première catégorie à valeurs SC d'oscillantes autour de l'intervalle [1.1;1.5], présentent une performance en conservant la structure géométrique de la quantité d'information de base de donnée dans tout l'intervalle de taux de variation de bruits [10;100], malgré les faibles perturbations qui révèlent une certaine stabilité partielle pour ce critère. En comparant avec celle de la deuxième catégorie des techniques de réduction comme Landmark Isomap, DM et LLTSA dont l'ordre à valeur Structural content varie [0.2;1.8], liées par des fortes déformations géométriques sur la base de données de valeur SC appartenant à une boule d'unité.

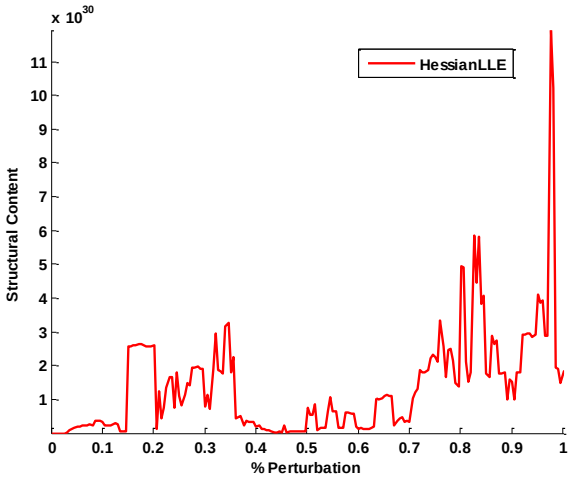


Figure 34. Méthodes de projection, 3eme Catégorie des méthodes de réduction

Globalement, la 2eme catégorie contenant les trois méthodes LLE, Isomap et KernelLDA, de la troisième catégorie contenant Hessian LLE (**Figure 34**) dont l'ordre SC est très grande, reflètent la sensibilité de ce critère sur les quantités d'informations préservées par ces méthodes. Certaines présentent quelques perturbations, pouvant aussi diminuer la valeur globale du potentiel de ces méthodes d'avoir préservé une quantité d'informations favorables après réduction par rapport à la présence de bruits.

La courbe SC associée à ces méthodes montre de mauvaise performance par rapport à l'ensemble des méthodes de réduction citées ci-dessus de pouvoir garder la forme géométrique locale et globale de valeur SC très loin de l'unité. En conclusion, ce critère montre ses meilleures performances sur le MDS.

4.4.2 Méthodes de sélection de bandes

Cette catégorie des méthodes semble donner différents avantages par rapport à la première catégorie de projection linéaire ou non-linéaire. En effet, les deux indices permettent la conclusion d'un comportement stable pour les trois méthodes comparées (Figure 35). La méthode BandClust utilisant l'information mutuelle entre deux bandes comme critère de sélection, fournit une stabilité plutôt élevé.

Plusieurs méthodes de classification sont des candidats potentiels pour le regroupement des bandes similaires [50]: K-means, RBG, K-médoıdes, Clara, LDA, SVM, K plus proches voisin. D'autre part, l'influence de ce paramètre de bruit sur cette catégorie des méthodes

favorise probablement l'utilisation des images de grande dimension. A noter que le pourcentage de bruit de 20% n'était pas suffisant pour perturber les performances de ces méthodes.

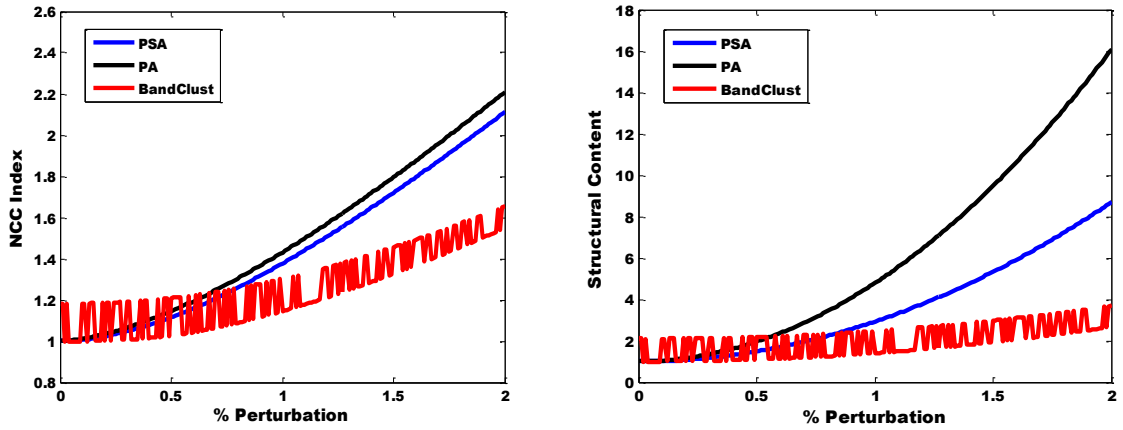
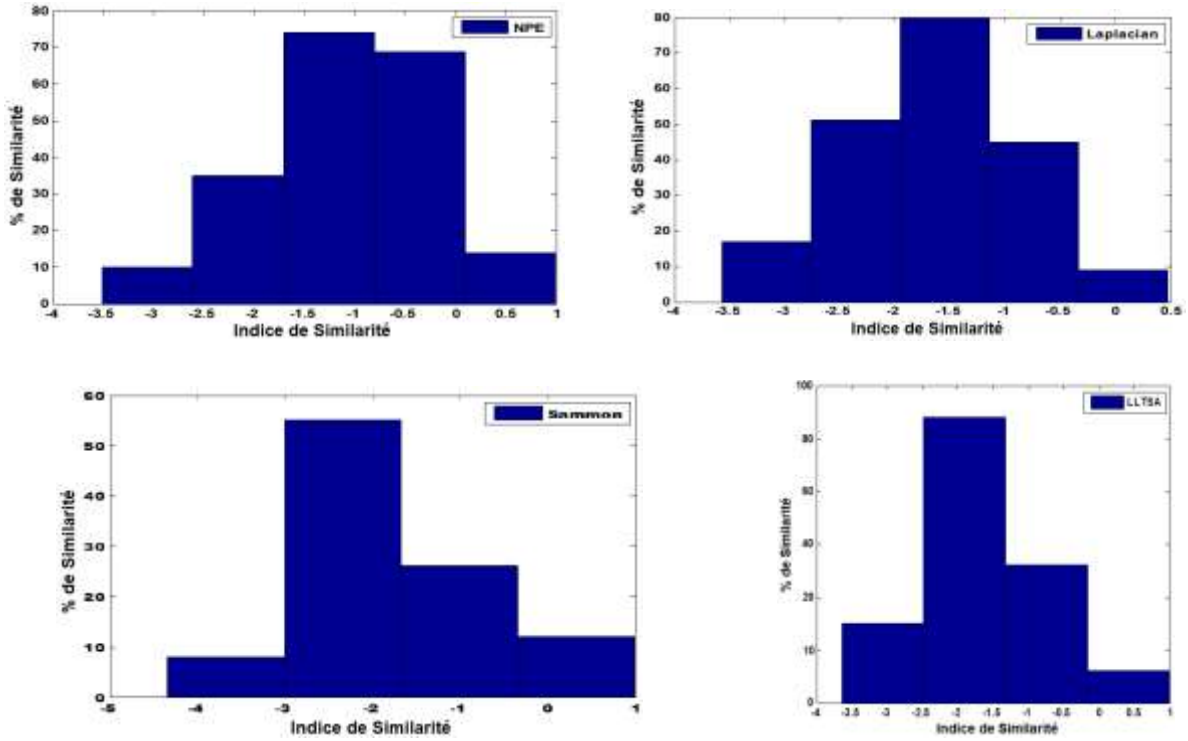


Figure 35. La sélection de bandes, Courbes de performances pour la stabilité

4.5 ETUDES COMPARATIVES SUR LA SIMILARITE

Nous avons exploité ce critère de similarité index (F) pour étudier l'efficacité de ce critère relativement à la dégradation de bruits sur la base des données, et l'influence d'évènements rares sur la réduction des données. Il semble important de pousser la discussion plus loin pour favoriser une nouvelle sélection relativement aux indices similarité déjà définis plus haut. Les courbes ci-dessous (**Figure 36**) montrent aussi trois classes pour ce groupe de méthodes jugées suffisamment différentes.



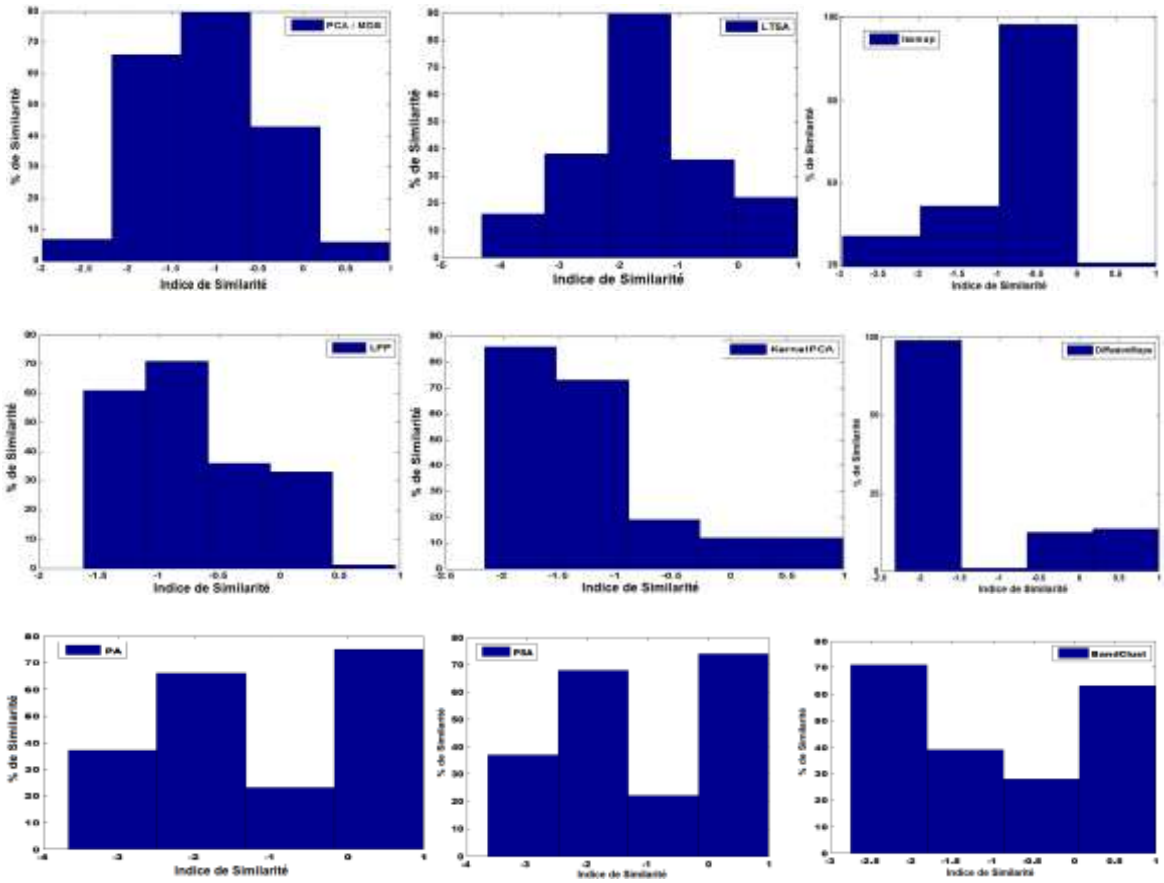


Figure 36. Pourcentage de similarité obtenus pour les diverses méthodes de réduction

On se base sur la notion de différence pour mesurer la ressemblance entre deux vecteurs à D-dimensions. Les deux critères permettent de bien choisir un groupe favori des méthodes de réduction de la dimension : PCA, MDS, LTSA, KernelPCA, LPP, NPE, Laplacian et Sammon. Par contre, les valeurs de similarité, dans la 1ere catégorie des méthodes de projection, sont presque équivalente mais non constante en fonction de la dégradation de bruits, liés par une déformation géométrique sensible à la variation de bruits et bien détectée par le critère de similarité. Ceci révèle la performance partielle sur les quantités de données gardées par cet ensemble de méthodes sous un intervalle de taux de réduction [0%;10%], ce qui marque la faiblesse de ces méthodes citées si dessus d'avoir préservé une quantité d'information favorable, en comparaison avec celles des résultats de similarité obtenus par la 2eme catégorie de sélection de bandes (BandClust/MDS, PA, PSA), qui garde une quantité maximale meilleure que les autres méthodes de réduction.

Pour les méthodes de projection et malgré les limitations que peut engendrer les choix des datas artificiels, il semble qu'un nouveau classement par pourcent de similarité décroissante peut être conclue : LTSA, NPE, KernelPCA, Laplacian, PCA/MDS et finalement LPP. Par contre la sélection de bande (PSA, PA et BandClust) offre une performance de similarité accrue tout en gardant un comportement presque identique entre les trois méthodes choisies ici.

Dans le tableau 5, une classification des méthodes de projection et de sélection de bandes est classée en trois catégories (bonne, moyenne ou mauvaise), en présentant les

performances de l'application en comparaison avec le critère de similarité observée ci-dessus.

Tableau 5. Les diverses méthodes de la réduction par projection utilisée, de manière à classer, sur des données artificielles avec des variances de bruit différents.

Classification des méthodes de réduction		
Bonne	**Moyenne**	**Mauvaise**
PCA - MDS - NPE - LTSA - KernelPCA - LPP - Laplacian - Sammon – BandClust	Isomap - Landmark isomap - LLE - Kernel LDA - PSA - PA	HessianLLE

4.6 ETUDES COMPARATIVES SUR LE TEMPS DE CALCUL

Un autre critère classique consiste à classifier les méthodes de réduction de dimension selon le temps d'exécution. La nature du processeur influe certainement sur cette valeur mais on peut toujours réaliser une étude comparative des valeurs normalisées pour ce critère. Le premier groupe des méthodes (PCA, MDS, LLE, LPP, LTSA, KPCA, DiffusionMaps, LLTSA, Isomap, NPE) sont particulièrement intéressantes et équivalentes en temps de calcul. Si nous prenons la moyenne du temps de calcul de ces méthodes comme valeur de référence, alors les méthodes SNE, Sammon et SymSNE présentent respectivement 15, 30 et 100 fois plus de temps pour réaliser leur performances de réduction. Ces valeurs assez élevées peuvent s'interpréter par l'utilisation des techniques d'optimisation itératives par ce dernier groupe pour effectuer la réduction. Cependant, cet inconvénient peut toujours être admis en observant leurs bonnes performances présentées dans certains critères.

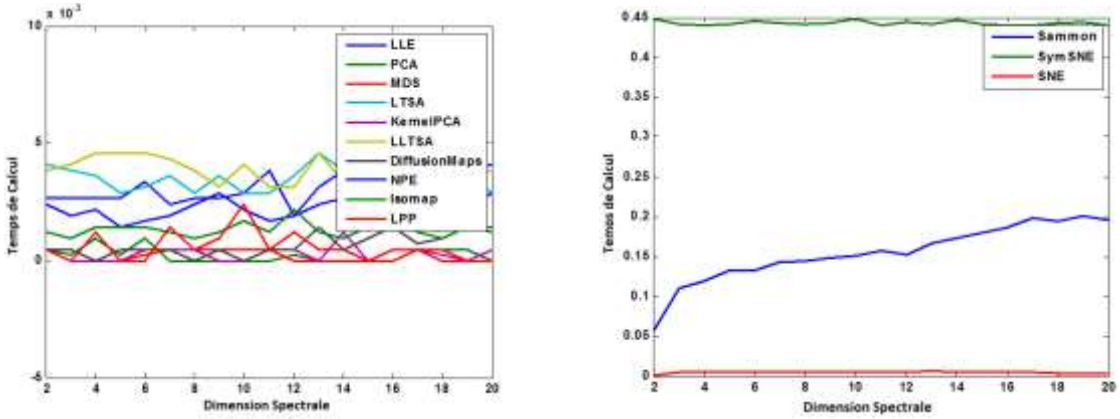


Figure 37. 1ere et 2eme catégorie des méthodes de projections testées avec le temps de calcul.

4.7 CONCLUSION ET PERSPECTIVES

Vu l'évolution de cette discussion, il est important de remarquer que l'hybridation entre les meilleurs techniques de projections et les méthodes de sélection de bande pourra aboutir vers des réductions de meilleure qualité en terme de stabilité, similarité et préservation d'informations.

Après plusieurs essais d'hybridations, nous avons ainsi conclu d'utiliser BandClust pour la classification des bandes similaires en terme d'information mutuelle puis d'utiliser MDS pour projeter chaque classe vers une seule bande (**Figure 38**).

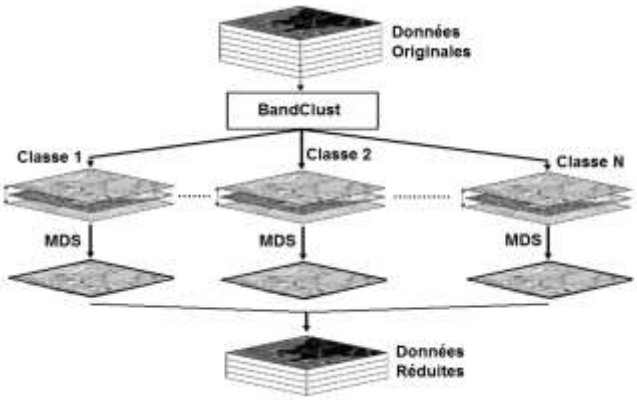


Figure 38. L'Algorithme hybride BandClust/MDS

La **figure 39** concernant les valeurs de l'inter-corrélation normalisée et Structural Content nous informent sur les bonnes performances obtenues par cette hybridation. Aussi la **figure 40** résume le potentiel de cette hybridation pour garder la similarité entre l'image réduite et l'image de base.

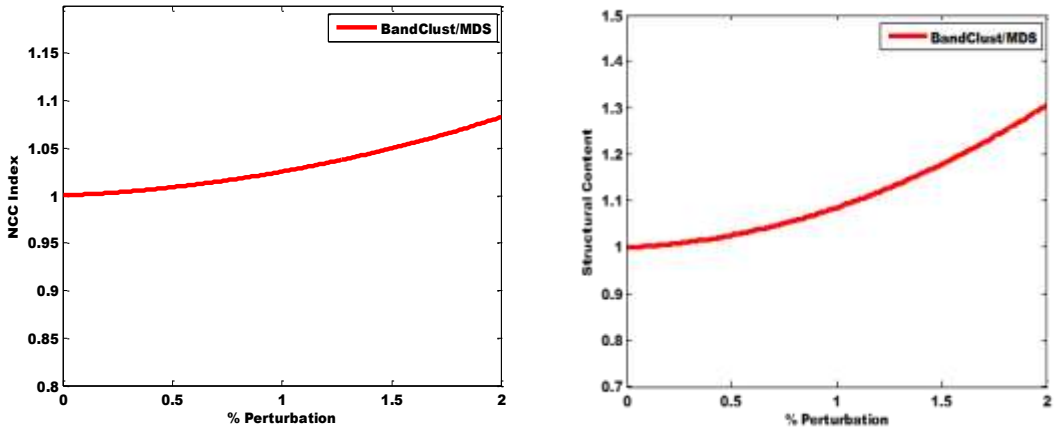


Figure 39. NCC et Structural Content for Hybrid BandClust/MDS Algorithm

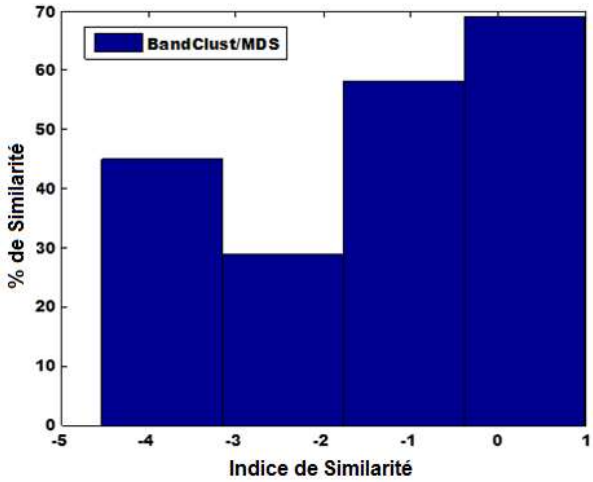


Figure 40. Pourcentage de la similarité BandClust/MDS

Nous avons accompli à travers cette étude comparative une hybridation avec les meilleures techniques de réduction de la dimension moyennant un jugement basé sur des critères de stabilité et similarité. Ces critères (l'inter-corrélation normalisée NCC, Structural Content SC) reflètent le potentiel de l'hybridation BandClust/MDS dans la réduction pour

préserver l'information, dont la distance est éloignée de la valeur moyenne de l'information, proportionnellement à la dégradation de bruits sur une échelle [0%; 10%]. Pour ces deux critères NCC et SC plus cette valeur est proche de l'unité, plus la méthode est stable.

BandClust/MDS offre un résultat optimal correspondant à des valeurs de l'inter-corrélation normalisée et Structural Content arrondi très proche de l'unité et plus précisément dans un intervalle [1 ; 1.3]. Ceci marque bien sa performance de conserver la structure géométrique et la quantité d'information dans la base de données pour un taux de variation de bruits dans [0%; 2%]. En outre, l'efficacité de cette hybridation en termes de similarité (F) est presque constante en fonction de la dégradation de bruits, ce qui révèle la performance de cette hybridation pour garder la structure géométrique dans la matrice des informations. Les tests de simulation ont porté sur un taux de réduction de 30% associé à une perturbation de 2%. BandClust/MDS offre des résultats de similarité meilleure que les techniques de la réduction par projection linéaire\non linéaire et par sélection de bandes.

Ainsi, ces trois critères (NCC, SC et F) testé sur BandClust/MDS reflète une bonne performance de préserver la structure locale/globale géométrique et structurelle des volumes de données réduits. La tendance de l'hybridation a été généralisée par la suite en présentant un algorithme hybride adaptatif non supervisé basé sur la logique flou.

Cette étude est à caractère comparative sur la stabilité des techniques non paramétriques, non supervisées pour la réduction de la dimensionnalité. Plusieurs critères ont été appliquées sur des datas artificielles avec distinction entre les critères d'évaluation de la similarité et ceux de la stabilité. Nous révélons de ce travail élaboré les points suivants :

Nous classifions en trois catégories les techniques de la réduction par projection et également ceux de la réduction par sélection de bandes spectrales. Cette classification établit une priorité sur les techniques de sélection de bandes spectrales BandClust, PA et PSA d'une part et celles de la projection PCA, MDS et NPE d'autre part. Leurs valeurs de stabilité semblent plus pertinentes sur les données artificielles utilisées, en comparaison avec les autres techniques de projection (LLTSA- LTSA- Diffusion Maps- HLLE-LPP-HLLE-KPCA-Isomap-Sammon) pour préserver une structure géométrique locale/globale de données réduites.

En deuxième point, l'étude propose une hybridation BandClust/MDS qui a permis des résultats très encourageants sur la similarité et la stabilité des données artificielles réduites. Cette hybridation présente l'avantage de combiner deux méthodes assez simple de point de vue algorithmique.

En perspectives, nous prévoyons un changement vers le développement de technique pour la réduction de dimensionnalité avec la fonction Hybride BandClust/MDS qui a montré son efficacité à travers les simulations effectués sur les critères de la similarité et la stabilité lors de la réduction. En effet BandClust/MDS pourrait être améliorée en maximisant la dissimilarité entre chaque deux bandes spectrales, et l'utilisation d'un critère d'information impliquant plusieurs objectifs créatifs, pouvant conserver les informations statistiquement rare et minimiser la redondance spectrale.

CHAPITRE 5
L'APPROCHE HYBRIDE PROPOSEE

5.1 INTRODUCTION

Pour améliorer l'efficacité de la réduction de dimensionnalité, nous proposons dans ce chapitre une hybridation basée sur la classification floue Fuzzy-C-Means avec une méthode de projection assez simple qui est l'Analyse en Composante Principale (ACP). A l'origine, la classification Fuzzy-C-means (FCM) ne s'utilise pas dans un objectif visant de réduire la dimension de données. Nous visons à travers cette utilisation d'obtenir un ensemble de données réduit contenant éventuellement un maximum des variables non corrélées. Le principe schématique de ce modèle est illustré sur la **figure 41**. Au début de notre démarche, il était normal de procéder par le plus simple. Nous avons ainsi initié une hybridation entre l'algorithme Fuzzy-C-Means et l'ACP. Cependant, après plusieurs tests avec d'autres méthodes de projection et sur plusieurs types d'images, aucune amélioration significative n'a était constatée sur la réduction des données hyperspectrales.

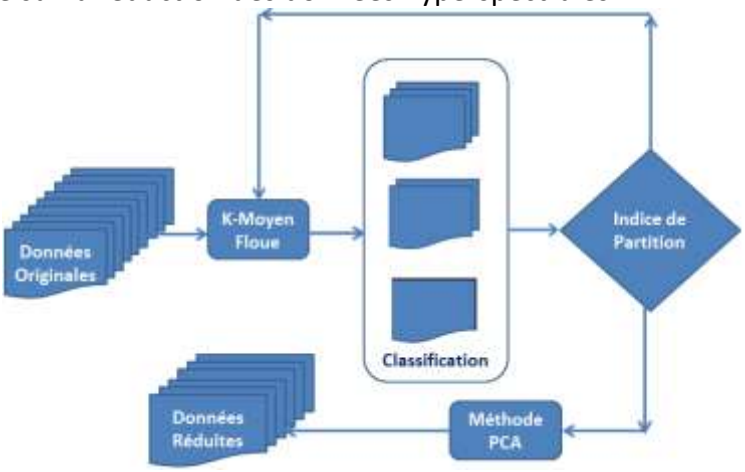


Figure 41. Principe de notre approche hybride adaptative

La classification floue permet d'affecter à chaque pixel d'une image hyperspectrale une appartenance à une classe avec des degrés d'appartenance variant entre 0 et 1 et cela toujours d'une manière non supervisée [96]. Ce type d'affectation est basé sur la distance entre un élément et tous les centres de classes, distance également appelée fonction d'appartenance. Cette fonction d'appartenance a été introduite par Zadeh [98] dans une des premières études sur les ensembles flous. Cette propriété rend la méthode FCM intéressante pour la mise en évidence soit des transitions progressives entre les différentes bandes spectrales ou des hétérogénéités spectrales.

Cependant, pour la classification FCM, comme pour les classifications « dures », le nombre de classes doit être défini *a priori* par l'utilisateur. De plus, elle nécessite de fixer *a priori* un paramètre supplémentaire appelé paramètre de flou. Dans un premier temps, grâce à des méthodes conventionnelles portant sur les indices de validité d'une classification, nous avons déterminé le nombre optimal de classes de FCM ainsi que le paramètre de flou. Nous montrerons ainsi que cette hybridation FCM/ACP devient aussi adaptative et conduit à un taux de réduction pertinent dans l'imagerie hyperspectrale. Par conséquent, Cet algorithme

appliqué à différents échantillons de données hyperspectrales, permet une imagerie spectrale réduite mais suffisamment informative, notamment au niveau de l'hétérogénéité spectrale.

5.2 APPROCHE HYBRIDE POUR LA REDUCTION DE LA DIMENSIONNALITE

Apres avoir défini le principe de notre approche, nous détaillons dans ce paragraphe tous ses composants. La **figure 42** met en relief ce besoin sur laquelle nous constatons la présence des éléments suivants :

- L'algorithme Fuzzy-C-Means,
- L'Indice de Validité,
- Le Test de Dis-similarité,
- L'Estimateur de Dimension Intrinsèque.

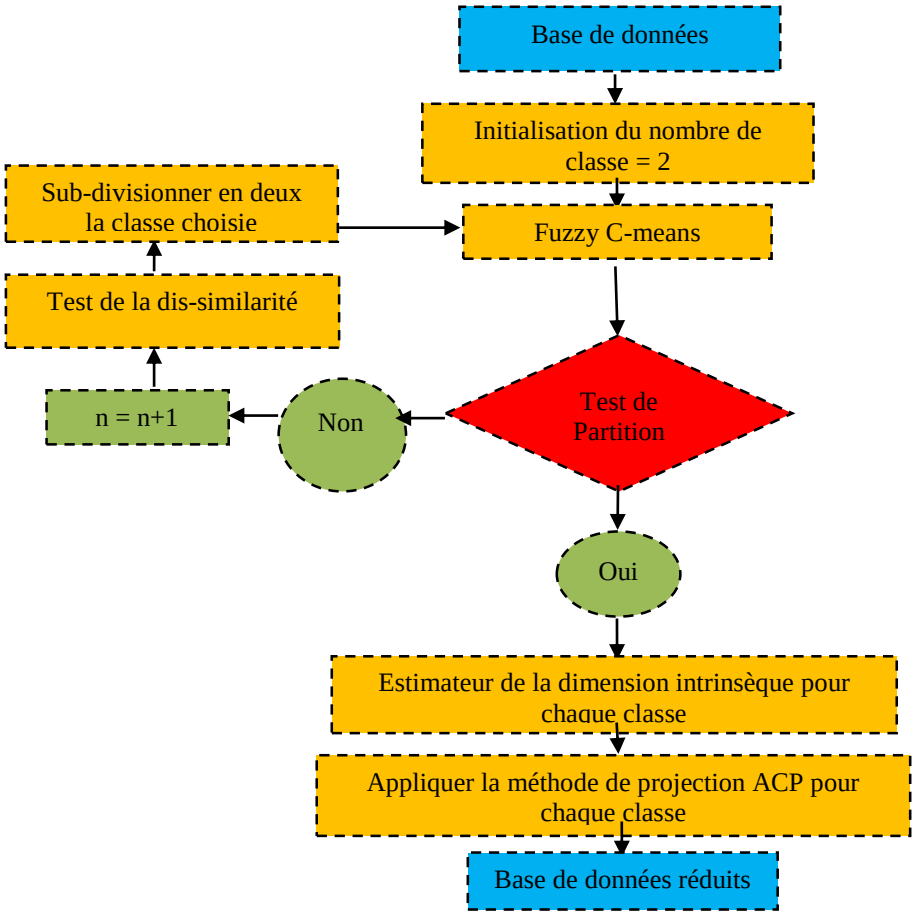


Figure 42. Modèle hybride flou adaptative pour la réduction de la dimension.

La classification FCM vise à diviser un ensemble de données en différents «paquets» homogènes, en ce sens que les données de chaque sous-ensemble partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité (similarité informatique) que l'on définit en introduisant des mesures et classes de distance entre objets. Pour valider une certaine classification FCM, on utilise un indice de validité, qui est une méthode statistique d'analyse des données, pour mesurer la pertinence de la partition de données. S'il est inférieur d'un seuil à déterminer, alors on réapplique

l'application FCM pour augmenter d'une seule classe. Par contre, si l'indice de partition de l'ensemble étudié est satisfait, on applique sur chaque classe un estimateur de la dimension intrinsèque de la réduction. Cette valeur sera considérée comme dimension de réduction optimale utilisée par la suite par la méthode de projection ACP.

Nous détaillons par la suite une étude bibliographique sur les éléments principaux qui composent notre algorithme que nous convenons de l'appeler : Approche Hybride Floue et Adaptative pour la Réduction de la Dimension (AHFARD).

5.3 CLASSIFICATION FCM

Le premier algorithme de FCM a été développé par Dunn [97] et a été amélioré par la suite par Bezdek [96]. A la différence de l'algorithme de Dunn, l'algorithme FCM de Bezdek introduit un paramètre de flou m : $1 \leq m < \infty$. Le but d'une classification floue est de minimiser la fonction objective définie par l'équation suivante:

$$J_m(X, U, V) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ij}^{m} \left\| x_i^{(j)} - c_j \right\|^2 \quad (87)$$

Où $\{x_1, x_2, \ldots, x_n\}$ représente l'ensemble des données, $V = \{v_1, v_2, \ldots, v_k\}$ l'ensemble des centres des classes correspondants, $\|.\|$ la norme L2 (distance euclidienne) entre l'élément x_i et le centre de classe V_K et U la matrice contenant tous les degrés d'appartenance u_{ik}.

Le degré U_{ik} d'appartenance doit satisfaire les conditions suivantes :

$$0 \leq U_{ik} \leq 1 \quad i=1,\ldots,n,\ k=1,\ldots,K \quad (88)$$

$$\sum_{k=1}^{K} u_{ik} = 1 \quad (89)$$

Autrement dit, le degré d'appartenance d'un élément à toutes les classes varie entre 0 (pas d'appartenance à une classe) et 1 (degré d'appartenance maximal à une classe). De plus, la somme des appartenances d'un élément à tous les classes doit être égale à 1, introduisant ainsi une notion probabiliste, ou de nuance, dans l'appartenance d'un élément aux classes. Le degré d'appartenance d'un élément est basé sur la distance entre cet élément et son centre de classe.

Pour minimiser la fonction objective (87), l'algorithme proposé par Bezdek est un algorithme itératif, convergent, décrit par les étapes suivantes [96] :

- Fixer le nombre de classes K=2. où $2 \leq K < n$
- Initialiser toutes les valeurs d'appartenance, u_{ik} , i =1,...,n , k =1,...,K avec des valeurs aléatoires telles qu'elles satisfassent les conditions (88, 89)
- Calculer les centres de classes v_k en utilisant la relation :

$$v_k = \text{Center}_K = \frac{\sum_{i=1}^{n} (u_{ij})^m x_i}{\sum_{i=1}^{n} (u_{ij})^m} \quad (90)$$

- Mettre à jour les valeurs d'appartenance avec la relation :

$$u_{ij}(x) = \frac{1}{\sum_{p=1}^{K} \left(\frac{d_{ik}}{d_{ip}}\right)^{\frac{2}{m-1}}} \quad \forall\ i=1,\ldots,n,\ k=1,\ldots,K \quad (91)$$

Où $d_{ik} = \|x_i - v_k\|$ et $d_{ip} = \|x_i - v_p\|$

5.4 ESTIMATION DE LA DIMENSION INTRINSEQUE

La dimension intrinsèque d'un vecteur aléatoire X est généralement définie comme le nombre des paramètres «Indépendants» nécessaires pour représenter X. Dans cette section, nous présentons les techniques d'estimation de la dimensionnalité intrinsèque qui peuvent être utilisées afin de contourner le problème de la sélection d'une dimension cible appropriée pour réduire le jeu de données X = (X_1, X_2, .., X_i, .., X_n) tel que X_i = (x_{i1}, x_{i2}, .., x_{im}) avec i représente l'indice de bandes spectrales.

Dans cette partie, nous discutons six techniques pour l'estimation de la dimensionnalité intrinsèque. Pour plus de commodité, on désigne la dimension intrinsèque de l'ensemble de données X par d et son estimation par $\hat{d}$.

Officieusement, l'approche de base à l'estimation de la dimension intrinsèque est basée sur les propriétés géométriques des données plutôt que des techniques de projection. Cette famille des méthodes en général ne nécessitent aucune hypothèse explicite sur le modèle de données sous-jacente, ni à définir les paramètres d'entrée. La plupart des méthodes géométriques utilise l'estimateur de la dimension par la corrélation, en raison de la simplicité de calcul de son estimation. Formellement, La définition est basée sur l'observation que, dans un ensemble de dimension m, le nombre de paires de points proches les uns des autres que r est proportionnelle à r^m.

Les techniques d'estimation de la dimensionnalité intrinsèque peuvent être subdivisés en deux groupes principaux: (1) des estimateurs basés sur l'analyse des propriétés locales des données et (2) des estimateurs basés sur l'analyse des propriétés globales des données. Pour formaliser cette procédure intuitive, nous présentons les définitions suivante.

5.4.1 Estimateurs locaux

Nous présentons brièvement dans ce qui suit, les trois estimateurs locaux de la dimension intrinsèque des données.

- Estimation de corrélation

Etant donné un ensemble de donnés Sn = {x_1, x_2,.., x_n} dans un espace métrique χ.

Où I_A est la fonction indicatrice de l'événement A.

$$C_n(r) = 2/n(n-1) \sum_{i=1}^{n} \sum_{j=i+1}^{n} I_A,\ A = \{\|xi - xj\| < r\},$$

Pour un ensemble dénombrable S ={x_1, x_2,....}⊂ χ. L'intégrale de corrélation est défini C(r) = $\lim_{n\to\infty} C_n(r)$. Si la limite existe, la corrélation dimension de S est définie comme

$$\widehat{D_{corr}}(r_1, r_2) = (\log(cr_1) - \log(cr_2)) / (\log(r_2) - \log(r_1)), \quad (92)$$

- Estimation du plus proche voisin ou *k-nearest neighbor*

En intelligence artificielle, la méthode des *k* plus proches voisins est une méthode d'apprentissage supervisé. Dans ce cadre, on dispose d'une base de données d'apprentissage constituée de *N* couples « entrée-sortie ». Pour estimer la sortie associée à une nouvelle entrée *x*, la méthode des *k* plus proches voisins consiste à

prendre en compte (de façon identique) les *k* échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée *x*, selon une distance à définir.

$$\widehat{d}\,(k_1, k_2) = \log\,(C\,(k_1) - (C(k_2))/\ \log\,((k_2))\text{-}(k_1)),\ \text{avec}\ \ C(k)=\frac{1}{n}\sum_i T_{k(}x_{i)}, \qquad (93)$$

Où $T_k(x_i)$ représente le rayon de la plus petite sphère de centre x_i qui couvre k points de datas voisins.

- Estimation du maximum de vraisemblance ou Maximum likelihood

L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Cette méthode a été développée par le statisticien Ronald Aylmer Fisher en 1922. Défini par

$$\text{DMLE} = 1\ /\ (k_2\text{-}k_1\text{+}1)\ \sum_{k=k1}^{k2} d_k \qquad (94)$$

Avec $d_k = 1/m \sum_{i=1}^{m} d_k(xi)$ et $d_k(xi) = [1/k - 1 \sum_{j=1}^{k-1} \log d(xi, xik)\ /d(xi, xij)]^{-1}$

5.4.2 Estimateurs globaux

Les estimateurs globaux considèrent les données comme un ensemble lors de l'estimation de la dimension intrinsèque. Ce paragraphe décrit les trois estimateurs globaux de dimensionnalité intrinsèques: (1) l'estimateur basé sur la valeur propre ou Eigenvalue estimator, (2) l'estimateur base sur l'arbre couvrant du poids minimal et (3) packing number estimator. Les trois estimateurs globaux de la dimensionnalité intrinsèques sont décrits séparément.

- Eigenvalue-based estimator : Estimateur par la valeur propre

L'estimateur de la dimensionnalité intrinsèque par la valeur propre performe PCA sur l'ensemble de données X de grande dimension et évalue les valeurs propres correspondant à la composante principale [99]. Son objectif est de maximiser la fonction $\frac{X^T A\, X}{X^T X}$.
Les valeurs de ces valeurs propres reflètent un aperçu de la quantité de variance qui est décrit par leurs vecteurs correspondant. Après normalisation, les valeurs propres peuvent être tracées afin de fournir un éclairage sur la dimension intrinsèque des données.

- Packing numbers estimator : Estimateur par le r-recouvrement du nombre

L'estimateur Packing numbers [100] est basé sur l'intuition que le nombre r de recouvrement N(r) (**r-covering number**) est proportionnel à r^{-d}. N(r) est le nombre de hypersphères de rayon r qui est nécessaire pour couvrir tous les points de données x_i dans l'ensemble de données X. Parce que N(r) est proportionnel à r^{-d}, la dimension intrinsèque d'un ensemble de données X est donnée par :

$$D = -\lim_{r\to 0} \frac{logN(r)}{logr}$$

En général, la recherche de la r-recouvrement du nombre N(r) d'un ensemble de données X est un problème qui est infaisable. L'estimateur Packing numbers peut contourner ce problème en utilisant le nombre r-packing M(r) au lieu de l'option N(r). Le nombre r-packing M(r) est défini comme étant la taille maximale d'un r sous-ensemble séparé de l'ensemble X. Autrement dit, le nombre r-packing M (r) est le nombre maximal de points de données en X qui peut être couvert par une seule hypersphère de rayon r. La dimension intrinsèque d'un ensemble de données X peut être trouvé en évaluant la limite :

$$d= -\lim_{r\to 0} \frac{logM(r)}{logr}$$

Parce que cette limite ne peut être évaluée de façon explicite, la dimension intrinsèque des données est estimée (similaire à l'estimateur de la dimension de corrélation) par :

$$\widehat{d_{pack}}(r_2, r_1) = -\log(M(r_2) - M(r_1))/(\log(r_2) - (r_1)). \quad (95)$$

- GMST estimator : Estimateur par l'arbre géodésique couvrant de poids minimal

Le minimum géodésique Spanning Tree (GMST) est un estimateur basé sur l'observation que la longueur de la fonction d'un arbre géodésique minimal est fortement dépendante de la dimension intrinsèque. Le GMST est le minimum sur un arbre du graphe de voisinage défini sur le jeu de données X. La fonction de la longueur GMST est la somme des distances euclidiennes correspondant à toutes les arêtes de l'arbre avec une distance géodésique minimum.

Similaire à l'isomap, l'estimateur GMST construit un graphe G sur le jeu de données X, dans laquelle chaque point de données x_i est relié à son k plus proche voisin x_{ij}. La géodésique minimum spanning tree T est défini comme le minimum du graphe sur X, qui a une longueur $L(x)=\min_{t\in T} g_e$

Où T est l'ensemble de tous les sous-arbres de graphe G, et g_e est la distance euclidienne correspondant à l'arête e.

Dans l'estimateur GMST, un certain nombre de sous-ensembles $A \subset X$ de l'ensemble de données X sont construits avec différentes tailles m, et les longueurs L(A) des sous-ensembles de GMST sont calculées. Théoriquement, le rapport $\frac{\log L(A)}{logm}$ est linéaire et peut être estimée par une fonction de la forme y = ax + b. Les variables a et b sont estimés au moyen de moindres carrés.

5.5 LES INDICES DE VALIDITE

Pour déterminer le nombre de classes optimale K_{opt} et apprécier la qualité des partitions, il existe des méthodes conventionnelles appelées indices de validité de classes. Dans ce travail nous avons utilisé l'indice de validité couramment utilisé qui est le coefficient de partition. Les différents indices de validité sont présentés ci-dessous.

Trois types des mesures de validité de classes peuvent être distingués d'une manière générale : ceux basés sur des critères externes, internes ou relatifs [103, 104].

- Les critères externes évaluent les résultats d'un algorithme de classification par rapport aux structures de classes prédéfinies. Par exemple, un critère externe mesure le degré de correspondance entre les classes obtenues et les étiquettes des classes à partir d'une structure assignée a priori.
- les critères internes évaluent la justesse entre les structures de classes et les données elles-mêmes. Essentiellement, il s'agit d'une mesure qui ne peut résulter que de la matrice de similarité. Cette matrice exprime en quelque sorte la qualité de la classification obtenue.
- les critères relatifs évaluent les résultats de classification afin de déterminer la classification qui représente le mieux les données à analyser.

Dans cette thèse, nous nous sommes concentrés sur des données réelles issues du domaine médical et satellitaire. Les bandes spectrales analysées sont complexes, riche en informations et le nombre de bandes différentes ne peut être qu'approximativement

estimé. Les connaissances *a priori* sur les composantes spectrales ne seront donc pas suffisantes pour fixer manuellement le nombre de classes nécessaire à l'étape de la projection. L'utilisation des indices de validité devrait donc permettre de choisir le meilleur schéma de classification des bandes spectrales, le nombre de classes pouvant ainsi être identifié automatiquement.

Les différents indices testés s'appliquent sur plusieurs partitions calculées par FCM pour différentes valeurs de *K* et *m*. Pour chaque couple (*K*,*m*) , l'indice de validité étudié est calculé à partir des degrés d'appartenance et des centres de classes estimés par FCM, et parfois en utilisant également les données elles-mêmes.

Nous présentons brièvement dans les sous-sections suivantes les indices de validité de classes testés dans ce travail. Rappelons que u_{ik} représente le degré d'appartenance d'un élément i au centre de classe k, x_i le $i^{ème}$ spectre, v_k le centre de la kème classe, N le nombre de spectres disponibles d'une acquisition.

5.5.1 Coefficient de Partition (F)

En 1981, Bezdek a introduit le coefficient de partition, indice de validité de classes dépendant uniquement des degrés d'appartenance U_{ik} [96]. Le coefficient de partition *F* est défini comme :

$$F(K)=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(u_{ik})^2 \qquad (96)$$

Ce coefficient est borné, F(K) ∈ $[\frac{1}{K}; 1]$. Cette gamme a été obtenue à partir de deux cas extrêmes. Le premier se produit lorsque les degrés d'appartenance de chaque élément sont soit égaux à 0 soit à 1, un élément appartenant donc à une seule classe. Dans tel cas, la valeur *F* est égale à 1, indiquant que toutes les classes sont bien délimitées. Dans le second cas, chaque élément appartient à toutes les classes avec le même degré d'appartenance et F(K) indiquant que la classification est la plus floue possible. Par conséquent, avec l'augmentation de la qualité de la classification, la valeur de *F* augmente également.
La gamme de valeurs possibles dépendant du nombre de classes et le coefficient ayant une évolution monotone en fonction de *K*, une normalisation de F a été proposée par Dave [102].

$$F^*(K)= 1-\frac{K}{1-K}(1-F(K))$$

Pour trouver la valeur optimale du nombre de classes avec ce critère, il faut optimiser la relation suivante :

$$K_{opt}^{F}(K)=\text{argmax}\,\{F(K)\}_{\,K\in\{2,..,N-1\}}$$

5.5.2 Entropie de Partition (H)

L'entropie de partition, *H* ne dépendant également que des degrés d'appartenance, a été définie toujours par Bezdek [96] comme :

$$H(K) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}u_{ik}\log_a(u_{ik}) \qquad (97)$$

Où *a* est la base du logarithme. Cette mesure est également bornée, *H(K)* ∈ $[0, log_a K]$. D'une façon similaire, il y a deux cas extrêmes qui forment la gamme de valeurs possibles pour *H*. Lorsque les classes sont bien séparées, la valeur de *H* est proche du zéro, en

revanche, lorsque la classification est plus floue, la valeur de *H* s'approche du $log_a K$. Le nombre de classes optimal est donné par la relation :

$$\widehat{K}_{opt}^{H} = \text{argmin}\ \{\text{H(K)}\}_{\ \text{K}\in\{2,..,\text{N}-1\}}$$

5.5.3 Exposant de Proportion (P)

Windham a proposé une autre mesure appelée exposant de proportion (*proportion exponent*), *P*, [101] :

$$\text{P (K)} = -\log_e[\prod_{i=1}^{N}[\sum_{j=1}^{u_i^{-1}}(-1)^{j+1}\binom{K}{j}(1-ju_i)^{k-1}]] \qquad (98)$$

où u_i = max $\{u_{ik}\}$, k ∈ [1,K]. Cette mesure utilise la valeur maximale du degré d'appartenance d'un élément à toutes les classes. Une grande valeur de *P* indique l'indentification la plus distincte d'une structure sous-jacente dans l'ensemble de données.

Le nombre de classes optimal est donné par :

$$\widehat{K}_{opt}^{P} = \text{argmax}\ \{\text{P(K)}\}_{\ \text{K}\in\{2,..,\text{N}-1\}}$$

Le principal avantage de ces trois indices de validité de classes est qu'ils sont faciles à calculer. Cependant, ils sont réellement utiles que s'il y a un nombre faible de classes bien séparées [105]. De plus, ces indices ont quelques limitations comme :
En première point, en calculant *F* et *H* sur de données hyperspectrales, les valeurs de *F* suivent une décroissance monotone et les valeurs de *H* une croissance monotone quand le nombre de classes *K* augmente.
En deuxième point, en utilisant uniquement les valeurs des degrés d'appartenance des données (c'est-à-dire sans utilisation des données elles-mêmes), ces indices n'ont pas non plus un rapport direct avec les propriétés géométriques inhérentes aux données.

Afin de surmonter ces limitations, et notamment prendre en compte les propriétés géométriques des données, beaucoup de fonctions, également appelées indices de validité, ont été proposées dans la littérature de spécialité. Il faut noter également que d'autres indices de validité utilisant uniquement les degrés d'appartenance aux classes existent, comme par exemple ceux proposés par Kim ou par Chen, mais ils n'ont pas été étudiés dans ce travail [106, 107].

5.5.4 Indice de FUKUYAMA et SUGENO, V_{FS}

Fukuyama et Sugeno ont proposé un premier indice de validité utilisant à la fois les degrés d'appartenance et les données définies comme [108]

$$\text{V}_{\text{FS}} = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}{}^{m}\ \|x_i - v_k\|^2 - \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}{}^{m}\ \|x_i - \bar{v}\|^2 \qquad (99)$$

Où $\bar{v}$ est la moyenne globale des données calculée comme la moyenne de centres de classes. Le premier terme combine les degrés d'appartenances aux classes avec la compacité géométrique de la représentation des données x_i par les centre de classe V_k . Le second terme combine les degrés d'appartenances avec la distance entre le prototype *k* et la moyenne globale des données. Le nombre de classes optimal est donné par la relation :

$$\widehat{K}_{opt}^{FS} = \text{argmin}\ \{\text{V}_{\text{FS}}\text{(K)}\}_{\ \text{K}\in\{2,..,\text{N}-1\}}$$

5.5.5 Indice de XIE et BENI, V_{XB}

Un autre indice de validité proposé par Xie et Beni [109] est défini comme :

$$V_{XB}= \frac{\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}{}^{m}\|x_i-v_k\|^2}{Nmin\,\|v_k-v_q\|^2}, k\neq q \qquad (100)$$

Cet indice comporte deux mesures : la compacité de la partition floue, le numérateur, qui n'est rien d'autre que la fonction objective à minimiser par l'algorithme FCM, et la séparation entre les classes, le dénominateur, qui indique la force de la séparation entre les classes. Selon les auteurs, une bonne classification produit une faible valeur pour la compacité et des centres de classes bien séparées produisent une valeur élevée pour la séparation.

Le nombre de classes optimal est donné par la relation :

$$\widehat{K}_{opt}^{XB} = \text{argmin}\,\{V_{XB}(K)\}_{\,K\in\{2,..,N-1\}}$$

5.5.6 Indice de KWON, V_K

En 1998, Kwon a étendu l'indice de Xie et Beni afin d'éliminer sa décroissance monotone lorsque le nombre de classes s'approche du nombre total des éléments de l'ensemble [110] de validité initial **:**

$$V_k(K)=\frac{\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}{}^{2}\|x_i-v_k\|^2+\frac{1}{K}\sum_{k=1}^{K}\|v_k-\bar{v}\|^2}{min\,\|v_k-v_q\|^2}, k\neq q \qquad (101)$$

Où $\bar{v}$ est la moyenne globale des données, Le premier terme du numérateur mesure la similarité intra-classe, c'est-à-dire la compacité des classes. Plus les classes sont similaires (compactes), plus le premier terme est faible. Cela est indépendant du nombre de classes *K*. Le second terme du numérateur représente une fonction de correction, utilisée pour éliminer la tendance de décroissance lorsque le nombre de classes *K* devient très important.

Une valeur plus élevée de celui-ci signifie que toutes les classes sont bien séparées. Il faut noter dans cette définition que la valeur du paramètre flou *m* a été tout simplement considérée comme étant égale à 2. Le nombre de classes optimal en utilisant cet indice est donné comme précédemment en minimisant la relation :

$$\widehat{K}_{opt}^{K} = \text{argmin}\,\{V_k(K)\}_{\,K\in\{2,..,N-1\}},$$

5.5.7 Indice de REZAEE, V_{CWB}

En 1998, Ramze Rezaee *et al.* proposent un indice de validité défini par [111] :

$$V_{CWB}= \alpha\,\frac{\frac{1}{K}\sum_{k=1}^{K}\|\sigma(v_k)\|}{\|\sigma(X)\|}+\frac{D_{max}}{D_{min}}\sum_{k=1}^{K}\left(\sum_{q=1}^{k}\left\|v_k-v_q\right\|\right)^{-1} \qquad (102)$$

Où D_{max} = maximum $\{\|v_a-v_b\|\}$, D_{min} = minimum $\{\|v_a-v_b\|\}\forall$ a, b $\in\{2,\ldots,K\}$, α est un facteur de ponderation et $\|.\|$la norme de L_2.

Le premier terme de *V*(*K*) *CWB* représente la moyenne de la variation de la dispersion au sein des classes pour un nombre de classes *K*. Une faible valeur pour ce terme indique une partition compacte. Quand la dispersion au sein des classes augmente, celles-ci deviennent moins compactes et ce terme est donc une bonne indication de la compacité moyenne des classes. Le second terme de l'indice de validité représente la séparation totale de la dispersion entre les classes. Généralement, ce terme va augmenter avec le nombre de classes et est influencé par la géométrie des centres de classes. Puisque les valeurs de ces deux termes sont d'une gamme différente, un facteur de pondération est nécessaire afin de les contrebalancer d'une manière appropriée.

Le nombre de classes optimal en utilisant cet indice est donné par :

$$\widehat{K}_{opt}^{CWB} = \text{argmin}\,\{V_{CWB}(K)\}_{K\in\{2,..,N-1\}},$$

5.5.8 Indice de PAKHIRA, V_{PBMF}

En 2004, Pakhira *et al.* proposent un indice de validité appelé l'indice PBMF, développé à la fois pour les classifications « dures » et floues. Dans le cas des classifications floues, cet indice est défini comme [112] :

$$V_{PBMF}=(\frac{1}{K}\times\frac{E_1}{J_m}\times D_k)^2,\ E_1=\sum_{i=1}^{N} u_{ik}\|x_i - v\| \qquad (103)$$

L'indice PBMF est une composition des trois facteurs, à savoir : 1/ K , E_1 / J_m et D_K. Le premier, indiquant la divisibilité du système de classification, devient faible avec l'augmentation de K. Le deuxième facteur inclut la somme des distances intraclasses pondérées pour l'ensemble intégral de données considéré comme une seule classe et pour le système à K classes. Ici, le dénominateur diminue avec l'augmentation de K, le numérateur ne variant pas. Le numérateur est fixé afin d'éliminer les possibilités que le deuxième facteur devient très faible. Ce facteur est une mesure de la compacité du système de classification. Le troisième facteur est une mesure de la séparation maximale inter-classes.

Le nombre de classes optimal en utilisant cet indice est donné par :

$$\widehat{K}_{opt}^{PBMF} = \text{argmax}\,\{V_{PBMF}(K)\}_{K\in\{2,..,N-1\}},$$

5.5.9 Indice de KIM, V_{SV}

En 2001, Kim *et al.* proposent un nouvel indice de validité qui se formule comme suit [113] :

$$V_{SV}(K)=\frac{v_{u(K)}-Vu_{min}}{vu_{max}-vu_{min}}+\frac{v_{O(K)}-VO_{min}}{vO_{max}-vO_{min}} \qquad (104)$$

Où $Vu_{max}= max_k V_u(K)$, $Vu_{min}= min_k V_u(K)$, $VO_{max}= max_k V_o(K)$, $VO_{max}= min_k V_o(K)$, pour K= 2,3, ..,K_{max}

Cet indice est composé de deux fonctions normalisées des mesures de partition, les auteurs distinguant ici deux états de partition de classes : sous-partition, quand $K_{opt} < K$ et sur-partition, quand $K_{opt} > K$, K_{opt} étant le nombre optimal de classes.
Le premier terme de l'indice représente une fonction de mesure de sous-partition :

$$V_U(K)=\frac{1}{K}\sum_{k=1}^{K} MD_k,$$

Où $MD_k= \sum_{x\in X_k} \|v_k - x\|/\eta_k$ est La distance de la k^{ieme} classe ayant η_k éléments. Ce premier terme, montrant la moyenne des distances intra-classes moyennes, mesure la compacité structurelle des classes. Quand la partition des données est optimale ou bien quand nous avons une sur partition, toutes les classes deviennent compactes et cela rend la valeur de V_u faible. De plus, quand le nombre de classes K devient très élevé et s'approche du nombre total d'éléments, N, cette distance moyenne tend vers 0. Cependant, dans le cas d'une sous-partition, V_u devient relativement élevé car certaines classes compactes peuvent se grouper dans une classe unique.
Le second terme de l'indice V_{SV} représente une fonction de mesure de sur-partition des classes :

$$V_O(K) = \frac{K}{d_{min}},$$

Où d_{min} représente la distance minimale entre les centres des classes, mesurant ainsi la séparation inter-classe. Quand la partition des données est optimale ou bien quand nous avons une sous-partition, d_{min} devient important ce qui rend la valeur V_o faible. Cependant, dans le cas d'une sur-partition, d_{min} devient très faible car certaines classes compactes peuvent se diviser en plusieurs classes.

Le nombre de classes optimal est donné par la relation :

$$\widehat{K}_{opt}^{SV} = \text{argmin}\ \{V_{SV}(K)\}_{\ K\in\{2,..,N-1\}},$$

5.6 ETUDE EXPERIMENTALE

L'Approche Hybride Floue et Adaptative pour la Réduction de la Dimension (AHFARD) est appliquée sur trois types d'imagerie hyperspectrales Indiana pines, une image cancéreuse et sur un jeu de objets réelles. Nous montrons le potentiel de la réduction de dimension avec l'algorithme proposée, en comparaison avec d'autres méthodes de projection.

Les critères « Normalized Cross Correlation », « Fidélité », « Mean Absolute Error », « Maximum Absolute Error » et « Mean Quadratic Error » sont utilisés pour refléter le potentiel de l'hybridation dans la réduction de préserver l'information. Les deux premiers mesurent la stabilité et la similarité et les trois derniers mesurent le niveau de préservation globale de l'information.

5.6.1 Discussion reliée à l'Indiana Pines

Une première application de la méthode proposée a été effectuée sur l'image de la **figure 19**. Les bandes optimales données par AHFARD fournissent une valeur de taux de réduction de dimension égale à 23 bandes spectrales sur un total de 224 bandes originales.

La quantification de cette approche par les critères choisis offre les valeurs suivantes :

(71,45%) pour l'erreur absolue moyenne, (73,22%) pour la valeur maximale d'erreur quadratique moyenne et le plus élevé fournit par Max absolu (84,73%). Ainsi, ces trois critères (MAE, MSE et MQE) testés sur AHFARD reflète une bonne capacité de préserver la structure locale/globale physique sur les volumes de données réduits.

L'algorithme proposé valorise ses performances, relativement au nombre de bandes retenue, par (90,77%) correspondant à la plus haute valeur de l'indice de similarité de la Fidélité et (93,05%) comme valeur moyenne pour la Normalized Cross correlation. Ces résultats sont obtenus en utilisant toutes les bandes des ensembles de données d'origine. Ce fait révèle la performance totale sur les quantités de données stockées par cet ensemble avec la méthode hybride. En effet, les critères NCC et SC reflète le potentiel de la réduction pour préserver la structure géométrique globale de l'information. Lorsque le rapport est proche de l'unité, la préservation des informations atteint son plus haut niveau.

5.6.2 Discussion reliée à l'image biomédicale

Afin de vérifier la validité de la méthode proposée, une deuxième application a été menée sur l'image biomédicale de la **figure 20**. Les bandes optimales données par AHFARD fournissent une valeur de taux de réduction de dimension égale à 7 bandes spectrales sur un total de 33 bandes originales. Le vecteur suivant représente le regroupement des bandes en 7 classes spectrales (**Figure 43**).
Classification: [2 1 3 5 5 6 6 4 6 6 4 6 6 4 6 4 6 6 4 7 7 7 7 7 7 7 7 5 7 5 7 7 7] pour l'image biomédicale.

La quantification de cette approche par les critères offre les valeurs suivantes :

(72,84%) pour l'erreur absolue moyenne, (78,32%) pour la valeur maximale d'erreur quadratique moyenne et le plus élevé fournit par Max absolu (83,63%). Ainsi, ces trois critères (MAE, MSE et MQE) testé sur AHFARD reflète également une bonne performance de préserver la structure locale/globale physique sur les volumes de données réduits.

L'algorithme proposé valorise également ses performances, relativement au nombre de bandes retenue, par (93,57%) correspondant à la plus haute valeur de l'indice de similarité de la Fidélité et (98,77%) comme valeur moyenne pour la Normalized Cross correlation. Ces résultats sont obtenus en utilisant toutes les bandes des ensembles de données d'origine.

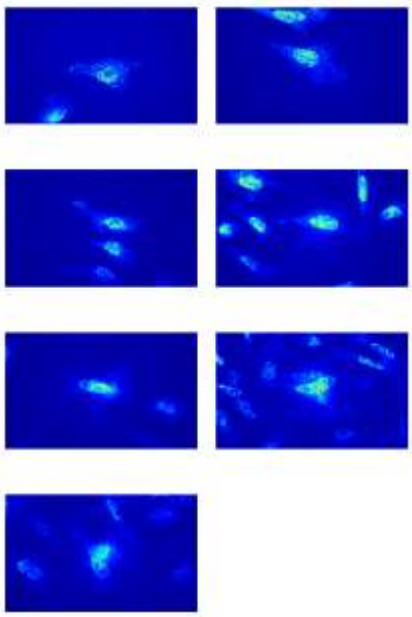


Figure 43. Sept bandes réduites de cellules du côlon anormal.

5.6.3 Discussion reliée au jeu artificiel de objets

Finalement, une troisième application a été menée sur l'image reliée au jeu artificielle des objets de la **figure 23**. Les bandes optimales données par AHFARD fournissent une valeur de taux de réduction de dimension égale à 11 bandes spectrales sur un total de 33 bandes originales. Le vecteur suivant représente le regroupement des bandes en 11 classes spectrales.
Classification: [7 7 7 7 7 1 2 3 5 6 6 6 4 4 4 4 6 6 6 5 5 5 8 8 8 11 11 10 11 10 9 9 9].

La quantification de cette approche par les critères offre les valeurs suivantes :

(74,47%) pour l'erreur absolue moyenne, (75,55%) pour la valeur maximale d'erreur quadratique moyenne et le plus élevé fournit par Max absolu (84,30%). Ainsi, ces trois critères (MAE, MSE et MQE) testé sur AHFARD reflète aussi une bonne performance de préserver la structure locale/globale physique sur les volumes de données réduits.

L'algorithme proposé valorise aussi ses performances, relativement au nombre de bandes retenue, par (92,24%) correspondant à la plus haute valeur de l'indice de similarité de (Fidélité) et (95,06%) comme valeur moyenne pour la Normalized Cross corrélation. Ces résultats sont obtenus en utilisant toutes les bandes des ensembles de données d'origine. Ce fait révèle la performance totale sur les quantités de données stockées par cet ensemble avec la méthode hybride. D'autre part, les critères NCC et SC reflète le potentiel de la réduction pour préserver la structure géométrique globale de l'information. Lorsque le rapport est proche de l'unité, la préservation des informations atteint son plus haut niveau.

5.7 EFFET DES CRITERES DE QUALITE

Les résultats numériques dans le **tableau 6** révèlent la forte performance de l'approche hybride dans la réduction de la dimensionnalité, utilisée sur chacun des trois datas naturelles.

Tableau 6. Performance moyenne de l'AHFARD sur les trois applications.

Données	Performances
Taux de réduction : bandes / Tous les bandes	7/33 23/224 13/33
MaxAbs	75.70%
MeanQuad	84.22%
MeanAbs	72.92**%**
Fidélité	92.20**%**
NormCC	95.62%

Plusieurs faits sont caractéristiques de cette approche :

1- Elle offre, en premier lieu, l'avantage d'optimiser le taux de réduction. Cette caractéristique est absente de toutes les autres méthodes et relève d'une question permanente dans le domaine de la réduction des cubes de données. La combinaison de plusieurs indices (validité d'une classification, test de dis-similarité, estimateur de réduction) a bien permis ce résultat.

2- L'hybridation testée avec la méthode PCA relève assez suffisante. En effet, La classification floue offre l'avantage de regrouper les bandes dans des classes avec forte similarité entre les bandes d'une même classe. Ainsi il devient inutile d'utiliser des méthodes assez fortes ou difficiles qui ont été développées pour traiter des images hyperspectrales avec forte non linéarité entre les bandes.

3- Une comparaison entre la méthode PCA seule et notre approche (hybridation entre F C-means et PCA) montre l'importance de notre approche (**Tableau 7**). FC-means prépare bien les groupes qui seront par la suite réduit avec le PCA. Cette même comparaison peut bien sûr être généralisée sur toutes les autres méthodes, mais elle semble inutile vu la 2eme remarque de ce paragraphe.

Tableau 7. Comparaison entre le PCA seule et notre approche.

Jeu des Objets	**Performances AHFARD**	**Performances PCA**
Taux de réduction	7/33	7/33
MaxAbs	75.70%	70.52%
MeanQuad	84.22%	72.27%
MeanAbs	72.92**%**	70.44**%**
Fidélité	92.20**%**	84.33**%**
NormCC	95.62%	86.84%

4- Il est aussi important de signaler le caractère non supervisée de la méthode. Très peu de paramètres sont à initialiser dans la méthode. Il s'agit de : pourcent de validité pour le test de validité et le taux de dis-similarité à admettre pour recréer une nouvelle division au sein d'une classe. Tous les deux sont indépendant du contexte dans lequel doit fonctionner l'approche de réduction.

5- Le temps de calcul peut être noté comme inconvénient majeur de cette approche. La méthode de réduction PCA est appliquée plusieurs fois sur toutes les classes. La méthode F C-means prend aussi son temps pour chaque itération jusqu'à atteindre une classification valide.

5.8 CONCLUSION

Ce chapitre était l'occasion de proposer une nouvelle approche pour la réduction des cubes de données. Cette approche bénéficie de plusieurs avantages : non supervisée, adaptative et optimale. Des validations sur plusieurs types d'images ont été menées et des faits caractéristiques ont été signalés. Des travaux futurs peuvent être prévus surtout pour Elargir le champ de validation.

CHAPITRE 6
CONCLUSIONS ET PERSPECTIVES

Cette thèse aborde le problème de réduction de la dimensionnalité des images hyperspectrales :

Nous avons commencé par présenter les deux approches générales (les techniques par projection linéaire/non-linéaire et les techniques par sélection des bandes spectrales) non paramétriques et non supervisées de réduction de la dimensionnalité des images à grande dimension. Nous avons rappelé les principes de ces méthodes. Rappelons que cette thèse s'inscrit dans le cadre de l'apprentissage non supervisé, non paramétrique en préservant l'évènement rare et que dans ce contexte, nous nous intéressons aux méthodes automatiques de réduction de la dimensionnalité pour les images de grande dimension.

Après avoir défini les particularités du domaine de l'hyperspectral et les approches qui ont été développées pour la réduction de la dimensionnalité, nous avions bien répertorié et classifié les critères d'évaluation de qualité image en trois catégories :
1- Les critères relatifs à la perte d'information (local/global), qui fonctionnent d'une certaine manière sur des propriétés spectrales par rapport au spatial.
2- Les critères relatifs à la stabilité et à la similarité qui sont fortement liés à la structure géométrique de l'image permettant d'évaluer les distorsions et de mesurer la sensibilité entre deux images (originales et réduites).

Nous avons profité de deux dernières classes de critères pour mesurer le niveau de dégradation de données initiales après réduction. Egalement, l'influence d'une perturbation uniforme introduite sur les données initiales a été analysée pour tous les algorithmes de réduction de la dimension. Ces critères ont permis d'analyser l'efficacité des méthodes de réduction de la dimension à réduire l'information, ainsi que leur capacité à préserver ou non la totalité de la richesse liée à la géométrie (globale et locale) de l'information (spectrale et spatiale) apportée par l'imagerie hyperspectrale.

Une nouvelle approche dans un cadre statistique intégrée dans les critères locaux, basée sur la di-similarité des attributs spectraux, pour mesurer le taux de la préservation des événements rares durant la réduction de la dimensionnalité a ensuite été présentée ; son principe est basé sur un rapport de similarité locale entre une variable et son voisinage. Deux exemples numériques sont présentés dans le chapitre III pour illustrer le taux de la préservation des vecteurs rares dans une matrice de taille 5x5 et 3x3. L'évènement rare dans notre contexte est un vecteur différent de son voisinage qui l'entoure. Une faible valeur pour cet indice, indique une préservation locale de l'évènement rare. En perspective, cette nouvelle approche doit d'une part être validée sur des grandes datas afin de valider son principe et d'autre part utiliser cet outil comme un critère comparatif sur toutes les méthodes de réduction de la dimensionnalité.

Nous avons accompli dans ce travail une étude comparative sur la stabilité et la similarité des algorithmes non paramétriques et non supervisée de la projection et aussi de la sélection des bandes utilisées dans la réduction de la dimensionnalité à différents niveaux de

bruit déterminés. Les tests sont effectués sur des images hyperspectrales, en classant ces derniers en trois catégories selon leur degré de performance de préserver la quantité d'informations. Pour valider notre approche, différentes images naturelles et synthétiques ont été employées pour évaluer la performance des critères de la stabilité, la similarité et la préservation de la quantité d'information sur les méthodes retenus de réduction de la dimensionnalité. Vu l'évolution des résultats, nous avons remarqué que l'hybridation entre les meilleurs des techniques entre les méthodes de projections et les méthodes de sélection de bandes pourra aboutir vers des réductions de meilleure qualité en terme de stabilité, similarité et préservation d'informations.

Après plusieurs essais d'hybridations valorisés par les critères de Normalized Cross Corrélation et Structural Content, nous avons constaté les bonnes performances obtenues par une hybridation BandClust/MDS. MDS pour projeter chaque classe vers une seule bande spectrale, et BandClust pour le regroupement des bandes similaires.

Dans le cadre de la thèse, vu la tendance de l'hybridation et les résultats obtenus dans le chapitre IV, nous étions amené à généraliser notre vision dans l'hybridation par une approche hybride adaptative non supervisée basée sur la logique flou (Fuzzy C means). La méthode est hybride dans le sens présenté ci-dessus, adaptative par son pouvoir de déterminer le meilleur taux de réduction, non supervisée vu qu'elle ne réclame pas de la réadaptation de ses paramètres en fonction des nouvelles applications. Son principe s'appuie sur une classification simultanée entre les différentes bandes spectrales en diminuant l'influence des dimensions redondantes.

Bien que l'ensemble des résultats soit satisfaisant, nous envisageons en perspectives différentes améliorations dans le contexte lié à la réduction de la dimension:

1- La mesure de similarité/dissimilarité entre variables est un point essentiel de l'approche développée et mérite toute notre attention: nous envisageons de remplacer la distance euclidienne qui s'est révélée inadaptée par la distance entre les partitions; nous envisageons plus précisément d'utiliser des méthodes conventionnelles appelées indices de validité de classes proposée par Bezdek's [96, 100].

2- La croissance des données de haute dimension crée un besoin de techniques de réduction de dimensionnalité pour transformer les données dans un espace réduit, facile à gérer et rapidement visualisées. Notre approche est non supervisée, c'est à dire ne nécessite aucune connaissance de la réalité de terrain ou le nombre final de groupes à répartir. L'hybridation floue adaptative BandClust/ACP semble être très facilement réalisable et a pu donner des résultats encourageants sur de véritables données hyperspectale. Cependant, il serait intéressant de l'appliquer sur plusieurs types de données provenant des diverses champs qui font appel à l'imagerie hyperspectrale.

3- L'approche offre un avantage particulier non existant dans les autres méthodes. C'est de fournir automatiquement une estimation du nombre optimal de groupes aux fins de classement réduit. Un travail plus exhaustif est à faire sur les degrés d'appartenance et les indices de validité étudiés par notre approche afin de rendre optimale la réduction.

4- L’évènement rare retient une attention particulière chez les utilisateurs des méthodes de réduction des données. Cependant, la littérature scientifique reste avare concernant ce problème. Notre initiation dans ce domaine a pu proposer une mesure quantitative de ce phénomène. Cette conclusion est notre occasion pour mentionner l’importance de classifier toutes les méthodes de réduction de la dimensionnalité selon leur potentiel de préserver l’ensemble des évènements rares dans les données hyperspectrales.

Il reste bien entendu beaucoup de possibilités d’améliorations, d’idées nouvelles à essayer, de points à préciser. Cette thèse était l’occasion d’apporter notre petite graine dans ce domaine de recherche. Elle était aussi notre opportunité vers un domaine vaste et infini pour lequel une vie entière n’est pas suffisante pour répondre à ces questions mais largement satisfaisant pour passer dedans une vie entière.

7. REFERENCES

1. L. Van Der Maaten, E. Postma, J. Van Den Herik. **Dimensionality Reduction: A Comparative Review**. TiCC, Tilburg University.

2. X. He, P. Niyogi, **Locality preserving projections, in: Advances in Neural Information Processing Systems**, MIT Press, Cambridge, MA, 2004, pp. 37–44.

3. D. Demers and G. Cottrell. **Non-linear dimensionality reduction**. Advances in Neural Information Processing Systems, Vol. 5, pp 580–587, San Mateo, CA, USA, 1993.

4. F.S. Tsai. **Comparative Study of Dimensionality Reduction Techniques**. Journal of artificial intelligence Vol. 3, pp 119-134, 2010.

5. J.B. Tenenbaum, V. de Silva, and J.C. Langford. **A global geometric framework for nonlinear dimensionality reduction**. Science, 290(5500): pp 2319–2323, 2000.

6. K.Q. Weinberger, L.K. Saul, **Unsupervised learning of image manifolds by semi definite programming**. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition. pp 988–995, 2004,

7. H.F. Chen, G.F. Jiang, and K. Yoshihira. **Robust nonlinear dimensionality reduction for manifold learning**. Proceeding of 18th International Conference on Pattern Recognition, pp 447–450, 2006.

8. K.Q. Weinberger, L.K. Saul. **Unsupervised learning of image manifolds by semi definite programming**. International Journal of Computer Vision 70 (1), pp 77–90, (2006).

9. Y.W. Teh and S.T. Roweis. **Automatic alignment of hidden representations**. Advances in Neural Information Processing Systems, 2002.

10. M.I. Jordan, M.J. Kearns, A.A. Solla. **Advances in Neural Information Processing Systems**, Volume 15, pages 841–848, Cambridge, MA, USA, 2002.

11. B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis**. Diffusion maps, spectral clustering and the reaction coordinate of dynamical systems.** Applied and Computational Harmonic Analysis, Vol. 21, pp 113-127, 2006.

12. Z. Zhang and H. Zha. **Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment**. SIAM Journal of Scientific Computing, 26(1):313–338, 2004.

13. D.L. Donoho and C. Grimes. **Hessian eigenmaps: locally linear embedding techniques for high-dimensional data.** Proceedings of the National Academy of Sciences USA 100, 5591–5596, 2003.

14. Trevor F. Cox, Michael AA Cox. **Multidimensional scaling**. Vol. 88. Chapman & Hall/CRC, 2001.

15. H.J. Yin, W.L. Huang. **Adaptive nonlinear manifolds and their applications to pattern recognition**, Information Sciences 180 (2010) 2649–2662.

16. Y. Weiss. **Segmentation using eigenvectors: a unifying view**. Proceedings of the IEEE International Conference on Computer Vision, Vol. 2, pp 975–982, Los Alamitos, CA, USA, 1999. IEEE Computer Society Press.

17. K.Q. Weinberger, F. Sha, and L.K. Saul. **Learning a kernel matrix for nonlinear dimensionality reduction**. Proceedings of the 21st International Conference on Machine Learning, 2004.

18. S.T. Roweis and L.K. Saul. **Nonlinear dimensionality reduction by Locally Linear Embedding**. Science, 290(5500):2323–2326, 2000.

19. O.C. Jenkins and M.J. Mataric. **Deriving action and behavior primitives from human motion data.** In International Conference on Intelligent Robots and Systems, volume 3, pages 2551–2556, 2002.

20. M. Belkin and P. Niyogi. **Laplacian Eigenmaps and spectral techniques for embedding and clustering**.In Advances in Neural Information Processing Systems, volume 14, pages 585–591, Cambridge, MA, USA, 2002. The MIT Press.

21. C. Spearman. **The proof and measurement of association between two things**, American Journal of Psychology 15 (1904) 72–101.

22. L. Chen. **Local multidimensional scaling for nonlinear dimension reduction, graph layout and proximity analysis**, Ph.D. Thesis, University of Pennsylvania, 2006.

23. P. Hansen, E. Ngai, K. Bernard Cheung, and N. Mladenovic. **Analysis of global k-means, an incremental heuristic for minimum sum-of-squares clustering**. Journal of Classification, 22(2), September 2005.

24. H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin. **Exploring strategies for training deep neural networks**. Journal of Machine Learning Research, 10(Jan):1–40, 2009.

25. T. Cox and M. Cox. **Multidimensional scaling**. Chapman & Hall, London, UK, 1994.

26. A. Zare, P. Gader. **Hyperspectral Band Selection and Endmember. Detection using Sparsity Promoting Priors**. IEEE Geoscience and Remote Sensing Letters, Vol. 5, Issue 2, pp 256- 260, April 2008.

27. K. Pearson**. On lines and planes of closest fit to systems of points in space**. Philiosophical Magazine, 2:559–572, 1901.

28. K. Fisher and Chein-I Chang. **Progressive band selection for satellite hyperspectral data compression and transmission.** Journal of Applied Remote Sensing, Vol. 4, Issue 1, Special Section on Satellite Data Compression.

29. J. C. Price, **Band selection procedure for multispectral scanners**. Applied Optics 33 (15), pp 3281 -3288. 1994.

30. S. Lee and Melba M. Crawford. **Hierarchical Clustering Approach for Unsupervised Image Classification of Hyperspectral Data**. IEEE International Symposium of Geoscience and Remote Sensing, IGARSS '04, vol.2, pp 941- 944, 2004.

31. D. C. Heinz and C.-I Chang. **Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery**. IEEE Trans. Geosci. Remote Sens., vol. 39, no. 3, pp. 529–545, Mar. 2001.

32. A. Martinez-Uso, F. Pla, J.M. Sotoca and P. Garcia-Sevilla. **Clustering-based multispectral band selection using mutual information**. ICPR, pp. 760–763 (2006).

33. T.M. Cover, J.A. Thomas. **Elements of Information Theory.** Wiley, Chichester (1991).

34. A. Webb. **Statistical Pattern Recognition**, 2nd edn. Wiley, Chichester (2002).

35. C. Cariou, K. Chehdi, and S. Le Moan. **BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing**. Geoscience and Remote Sensing Letters, IEEE. Vol. 8, Issue 3, pp 565 – 569, May 2011.

36. C. Kermad, K. Chehdi. **Multi-Bands Image Segmentation: A Scalar Approach. International Conference on Image Processing**, Vol.3, pp 468 -471, 2000.

37. M.E. Tipping and C.M. Bishop. **Mixtures of probabilistic principal component analysers**. Neural Computation, 11(2):443–482, 1999.

38. Z. Ghahramani and G.E. Hinton. **The EM algorithm for mixtures of factor analyzers.** Technical Report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996.

39. A. Dempster, N. Laird, and D. Rubin. **Maximum likelihood from incomplete data via the EM algorithm.** Journal of the Royal Statistical Society, Series B, 39(1):1–38, 1977.

40. 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(9):1393–1403, 2006.

41. G.E. Hinton and R.R. Salakhutdinov. **Reducing the dimensionality of data with neural networks**. Science, 313 (5786): 504–507, 2006.

42. B.L. Betechuoh, T. Marwala, and T. Tettey. **Autoencoder networks for HIV classification**. Current Science, 91(11):1467–1473, 2006.

43. L. Vandenberghe and S. Boyd. **Semidefinite programming**. SIAM Review, 38(1):49–95, 1996.

44. G.A Licciardi, Fabio Del Frate. **A Comparison of Feature Extraction Methodologies Applied on Hyperspectral Data**. Proc. Hyperspectral 2010 Workshop, Frascati, Italy. 2010.

45. S.Chen, H. Zhao, M. Kong, & B. Luo. **2D-LPP: a two-dimensional extension of locality preserving projections**. Neurocomputing, 70(4), 912-921. 2007.

46. B. Scholkopf, A.J. Smola, and K.-R.Muller. **Nonlinear component analysis as a kernel eigenvalue problem**. Neural Computation, 10(5):1299–1319, 1998.

47.J. Shawe -Taylor and N. Christianini. **Kernel Methods for Pattern Analysis.** Cambridge University Press, Cambridge, UK, 2004.

48.C. Chatfield and A.J. Collins. **Introduction to Multivariate Analysis**. Chapman and Hall, 1980.

49.T. Cover and J. Thomas. **Elements of Information Theory,** Wiley, New York (1989).

50.Y. Qian, F. Yao, & S. Jia. **Band selection for hyperspectral imagery using affinity propagation.** Computer Vision, IET, 3(4), 213-222. (2009).

51.A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla. **Clusteringbased hyperspectral band selection using information measures.** IEEE Trans. Geosci. Remote Sens., vol. 45, no. 12, pp. 4158–4171, Dec. 2007.

52.Q. Du and H. Yang. **Similarity-based unsupervised band selection for hyperspectral image analysis.** IEEE Geosci. Remote Sens. Lett, vol. 5, no. 4, pp. 564–568, Oct. 2008.

53.C.-I. Chang. **An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis.** IEEE Trans. Inf. Theory, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

54. Ch. Hou, Ch. Zhang, Y. Wu, Y. Jiao. **Stable local dimensionality reduction approaches**. Pattern Recognition, Volume 42, Issue 9, Pages 2054–2066, September 2009.

55.N. Feiping, Sh. Xiang, J. Yangqing, Z. Changchui, Y. Shuicheng. **Trace ratio criterion for feature selection.** Proceedings of the 23rd national conference on Artificial intelligence. p. 671-676. 2008.

56.E.S. Tsai. **Techniques for Data Visualization**. Journal of artificial intelligence 3(3); 119-134. 2010.

57.O. Kuybeda, D. Malah, M. Barzohar. **Rank Estimation and Redundancy Reduction of High-Dimensional Noisy Signals With Preservation of Rare Vectors**. IEEE Transactions on Signal Processing, Vol. 55, No. 12, December 2007.

58. Chein-I Chang. and Su WANG. **Constrained band selection for hyperspectral imagery**. Geoscience and Remote Sensing, IEEE Transactions on, 2006, vol. 44, no 6, p. 1575-1585.

59.J. Venna, S. Kaski. **Local multidimensional scaling, Neural Networks**, 2006, vol.19, no.6, p. 889–899.

60.M. Deyu, Y. Leung, and X. Zongben. **A new quality assessment criterion for nonlinear dimensionality reduction.** Neurocomputing 74.6 (2011): 941-948.

61.J. Lee, M. Verleysen, **Quality assessment of dimensionality reduction: rank- based criteria**, Neurocomputing 72 (7–9) (2009) 1431–1443.

62.F. Kruse, **Introduction to hyperspectral data analysis**. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, HI, USA, 2000.

63. X. He, D. Cai, S. Yan, H. Zhang, **Neighborhood preserving embedding**. IEEE International Conference on Computer Vision, Vancouver, Canada, 2005, pp. 1208–1213.

64. G. Motta, F. Rizzo et J. A. Storer. **Compression of hyperspectral imagery**. Data Compression Conference, DCC, vol. 8, p. 333– 342. IEEE mars 2003.

65. P. Baizert, M. Pickering and M. Ryan**. Compression of hyperspectral data by spatial/spectral discrete cosine transform.** IEEE International Geoscience and Remote Sensing Symposium, IGARSS'01, vol. 4, p. 1859–1861. juil. 2001.

66. M. Ryan and M. Pickering. **An improved M-NVQ algorithm for the compression of hyperspectral data.** Dans IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2000, vol. 2, p. 600–602. juil. 2000.

67. M. Pickering and M. Ryan. **Compression of hyperspectral data using vector quantization and the discrete cosine transform**. IEEE International Conference on Image Processing, ICIP 2000, vol. 2, p. 195–198.sept. 2000.

68. B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, L. Santurri et M. Selva. **Spectral distortion evaluation in lossy compression of hyperspectral imagery**. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'03, vol. 3, p. 1817–1819. juil. 2003.

69. George J. KLIR, and Mark J. Wierman. **Uncertainty-based information: elements of generalized information theory**. Physica-Verlag Heidelberg, 1999.

70. Y.W. Teh, S. Roweis. **Automatic alignment of local representations, in: Advances in Neural Information Processing Systems**. MIT Press, Cambridge, MA, pp. 841–848, 2003.

71. W.E. Arnoldi. **The principle of minimized iteration in the solution of the matrix eigenvalue problem**. Quarterly of Applied Mathematics, 9:17–25, 1951.

72. Chang, Chein-I. **An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis**. Information Theory, IEEE Transactions on, 2000, vol. 46, no 5, p. 1927-1932.

73. T. Tsai, F.,and Philpot, W. (1998). **Derivative analysis of hyperspectral data**. Remote Sensing of Environment, 66(1), 41-51.

74. Z. Wang, H.R. Sheikh, A. Bovik. **Objective video quality assessment**. In The Handbook of Video Databases: Design and Applications (B. Furht and O. Marques,eds.), CRC Press, 2003.

75. B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, L. Santurri et M. Selva**. Spectral distortion evaluation in lossy compression of hyperspectral imagery.** IEEE International Geoscience and Remote Sensing Symposium, IGARSS'03, vol. 3, p. 1817–1819. juil. 2003.

76. S. Qian, M. Bergeron, C. Serele, I. Cunningham and A. Hollinger**. Evaluation and comparison of JPEG2000 and vector quantization based onboard data compression algorithm for hyperspectral imagery**. IEEE International Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. vol. 3, p. 1820–1822. juil. 2003.

77.N. Keshava and J. F. Mustard. **Spectral unmixing**. IEEE Signal Processing Magazine, vol. 19, no 1, p. 44–57, jan. 2002.

78.J. Ma, Q. Tong, L. Zheng et B. Zhang. **An approach to assessing hyperspectral data quality.** Proc. SPIE, Multispectral and hyperspectral image acquisition and processing, vol. 4548, pp. 244–247, 2001.

79.B. Aiazzi, L. Alparone, S. Baronti, F. Lotti, L. Santurri et M. Selva**. Quality issues for hyperspectral data dissemination: radiometric and spectral distortions in lossy compression.** Dans Sensors, Systems, and Next-Generation Satellites VI, vol. 4881, pp. 701–708. SPIE, 2003.

80.A. M. Eskicioglu and P. S. Fisher. **A survey of quality measures for gray scale image compression**. In 9th Computing in Aerospace Conference, 93-4514. AIAA, Oct. 1993.

81.C. E. Shannon and W. Weaver. **The Mathematical Theory of Communication**. Illini Books, 1949.

82.A. Chaddad, C.Tanougast, A. Dandache, A. Al Houseini and A. Bouridane. **Improving of Colon Cancer Cells Detection Based on Haralick‟s Features on Segmented Histopathological Images.** IEEE Conference on Computer Applications and Industrial Electronics, December 2011, pp.87 - 90***.***

83.C.I. Chang, Q. Du, T. L. Sun, and M. L. G. Althouse. **A joint band prioritization and band-deccorelation approach to band selection for hyperspectral image classification.** IEEE Trans. Geosci. Remote Sens.,vol. 37, no. 6, pp. 2631C2641, 1999.

84.W. M. Porter and H. T. Enmark. **A system overview of the airborne visible/ infrared imaging spectrometer (AVIRIS).** maging Spectroscopy II Conference. Society of Photo-Optical Instrumentation Engineers, aout 1987.

85.ITRES Research Limited. **CASI-2 specifications**.http://www.itres.com/docs/casi2.html, oct. 2002.

86.R. Bianchi, R. M. Cavalli, C. M. Marino and S. Pignatti**. MIVIS airborne hyperspectral remote sensing in Europe.** ISSSR-95 proceedings. CNR, LARA Project, nov. 1995.

87.T. Cocks, R. Jenssen, A. Stewart, I. Wilson and T. Shields. **The HYMAP airborne hyperspectral sensor : the system, calibration and performance**. 1st EARSEL workshop on imaging spectroscopy, Zurich. oct. 1998.

88.F. Kruse, J. Boardman, A. Lefkoff and J. Young. **The AIG/HyVista1999 USA HyMap group shoot: overview and analysis examples**. IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2000, vol. 4, p. 1373 –1375. juil. 2000.

89.D. Beiso. **Overview of Hyperion on-orbit instrument performance, stability,and artifacts**. In Proceedings 31st Applied Imagery Pattern Recognition Workshop, 2002, pp. 95–101. oct. 2002.

90.J. Pearlman, C. Segal, L. Liao, S. Carman, M. Folkman, B. Browne,L. Ong and S. Ungar. **Development and operations of the EO-1 Hyperionimaging spectrometer**. Rap. tech., TRW Space and Electronics Group,aout 2000.

91. ESA Scientific Campaign Unit, ESTEC. **Exploitation of CHRIS datafrom the PROBA mission for science and applications experimenters handbook**, oct. 1999.

92. H.J. Kramer. **Observation of the Earth and its Environment : Survey of missions and sensors**. Springer, 4e ed., 2002.

93. European Space Agency. **SPECTRA - Surface Processes and Ecosystem Changes Through Response Analysis**. The Five Candidate Earth Explorer Core Missions. ESA, sept. 2001.

94. R.E. Simmons et B. V. Brower. **Data characterization for hyperspectral image compression.** Multispectral imaging for terrestrial application II, vol. 3119, p. 172–183. SPIE, 1997.

95. J.F. Bolton. **Full spectral imaging : a revisited approach to remote sensing. In Sensors, Systems, and Next-Generation Satellites** VII, vol. 5234, p. 243–251. SPIE, sept. 2003.

96. J.C. Bezdek. **Pattern recognition with fuzzy objective function algorithms**. Plenum Press: New York, 1981.

97. J.C. Dunn. **A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters**. Journal of Cybernetics. (3): 32-57. 1973.

98. L.A. Zadeh. **Fuzzy sets**. Information and Control; (8): 338-53. 1965.

99. K. Fukunaga and D.R. Olsen**. An algorithm for finding intrinsic dimensionality of data**. IEEE Transactions on Computers, C-20:176–183, 1971.

100. N.R. Pal, J.C. Bezdek. **On cluster validity for the fuzzy c-means model**. IEEE Transactions on Fuzzy Systems 3 (3) (1995) 370–379.

101. M.P. Windham. **Cluster validity for fuzzy clustering algorithms**. Fuzzy Sets and Systems,5 (2),177-185,1981.

102. R.N. Dave. **Validating fuzzy partitions obtained through c-shells clustering**. Pattern Recognition Letters, vol. 17,p. 613-623, 1996

103. M. Halkidi, Y. Batistakis, and M. Vazirgiannis**. On clustering validation techniques**. Journal of Intelligent Information Systems, vol.17,p. 107-45, 2001

104. A.K. Jain, and R.C. Dubes. **Algorithms for clustering data**. Prentice-Hall: New Jersy, 1988.

105. H. Sun, S. Wang, and Q. Jiang. **FCM-based model selection algorithms for determining the number of clusters.** Pattern Recognition (37): 2027-37, 2004.

106. D.W. Kim, K.H. Lee, and D. Lee. **On cluster validity index for estimation of the optimal number of fuzzy clusters.** Pattern Recognition 2004; (37): 2009-25.

107. M.Y. Chen, and D.A. Linkens. **Rule-base self-generation and simplification for data-driven fuzzy models.** Fuzzy Sets and Systems 2004; (142): 243-65.

108. Fukuyama Y, and Sugeno M. **A new method of choosing the number of clusters for the fuzzy c-mean method. Proceedings of the 5th Fuzzy Systems Symposium**, pp 247-50, 1989.

109. X.L. Xie, and G. Beni. **A validity measure for fuzzy clustering**. IEEE Transactions on Pattern Analysis and Machine Intelligence 1991; (13): 841-7.

110. S.H. Kwon. **Cluster validity index for fuzzy clustering**. Electronics Letters 1998; (34): 2176-7.

111. R.M. Ramze, B.P.F. Lelieveldt, and J.H.C. Reiber. **A new cluster validity index for the fuzzy c-mean.** Pattern Recognition Letters 1998; (19): 237-46.

112. M.K. Pakhira, S. Bandyopadhyay, and U. Maulik. **Validity index for crisp and fuzzy clusters**. Pattern Recognition 2004; (37): 487-501**.**

113. D.J. Kim, Y.W. Park, and K.J. Park. **A novel validity index for determination of the optimal number of clusters**. IEICE Transactions on Information and Systems 2001; (E84-D): 281-5.

114. C. Han, P. Willett, and D. Abraham. **Some methods to evaluate the performance of Page's test as used to detect transient signals**. IEEE Trans. Signal Process., vol. 47, no. 8, pp. 2112–2127, 1999.

115. V. F. Pisarenko, A. F. Kushnir, and I. Savin. **Statistical adaptive algorithms for estimation of onset moments of seismic plates,** Physics of the Earth and Planetary Interiors, vol. 47, pp. 4–10, 1987.

116. B. Penna, T. Tillo, E. Magli, and G. Olmo. **Hyperspectral image compression employing a model of anomalous pixels**. IEEE Geosci. Remote Sens. Lett., vol. 4, pp. 664–668, 2007.

117. O. Kuybeda, D. Malah, and M. Barzohar. **Global unsupervised anomaly extraction and discrimination in hyperspectral images via maximum-orthogonal complement analysis**. presented at the Eur. Signal Process. Conf. (EUSIPCO), Lausanne, Switzerland, Aug. 2008.

118. P. V. Overshee and B. D. Moor. **Subspace algorithms for the stochastic identification problem**. Automatica, vol. 29, pp. 649–660,1993.

119. M. Viberg. **Subspace-based methods for the identification of linear time-invariant systems.** Automatica, vol. 31, no. 12, pp. 1835–1853,1995.

120. M. Tipping and C. Bishop. **Probabilistic principal component analysis**. J. Roy. Statist. Soc. B, Statist. Methodol., vol. 61, no. 3, pp. 611–622, 1999.

121. J. M. P. Nascimento and J. M. B. Dias. **Signal subspace identification in hyperspectral linear mixtures.** Lecture Notes Comput. Sci., vol. 3523, pp. 207–214, Jan. 2005.

122. M. Marcus and P. Swerling. **Sequential detection in radar with multiple resolution elements**. IEEE Trans. Inf. Theory, vol. 8, no. 3, pp. 237–245,1962.

123. M. Thottan and C. Ji. **Anomaly detection in ip networks**. IEEE Trans. Signal Process., vol. 51, no. 8, pp. 2191–2204, 2003.

124. A. Wald and J. Wolfowitz. **Character of the sequential probability ratio test.** Annals of Mathematical Statistics, vol. 19, no. 3, pp. 326–339,Sep. 1948.

125. L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis. **Quickest search over multiple sequences.** IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5375– 5386, Aug. 2011.

126. O. Kuybeda, D. Malah, and M. Barzohar. **Anomaly Preserving $l_{2,\infty}$-Optimal Dimensionality Reduction over a Grassmann Manifold**. IEEE Transactions on signal Processing. Vol 58, no.2, February 2010.

ANNEXE 1 : EXPLOITATION DU MATLAB

MatLab ne propose aucun toolbox pour le traitement des images hyperspectrales. Certains auteurs proposent sur le net des toolbox pour cet objectif. Cependant, ces toolbox souffrent de certains inconvénients rendant leurs utilisations inefficaces :

- Manque de guide d'utilisation complet,
- Paramétrage difficile des méthodes,
- Deux propositions sur la même méthode ne fournissent pas le même résultat,
- Proposition continue des versions évolutives de ces toolbox dans le temps sans clarification sur le besoin de changement ou les nouveaux apports,
- La reconstruction de l'image après réduction pose un vrai problème non abordé par ces toolbox,
- Toolbox non référencié dans une source scientifique valide.

Voici ci-joint quelques sites potentiels dans ce sens :
http://sourceforge.net/projects/matlabhyperspec/files/
http://www.censsis.neu.edu/software/hyperspectral/hyperspectral.html
http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB
http://matlab-hyperspectral-toolbox.soft112.com/

Ainsi, la 1ere étape de cette thèse était consacrée pour le développement de 20 techniques de réduction sous forme des fonctions matlab. La liste suivante a été réalisée :

- Principal Component Analysis ('pca.m')
- Multidimensional scaling ('mds.m')
- Sammon mapping ('sammon.m')
- Isomap ('isomap.m')
- Landmark Isomap ('landisomap.m')
- Locally Linear Embedding ('lle.m')
- Laplacian Eigenmaps ('le.m')
- Hessian LLE ('hessianlle.m')
- Local Tangent Space Alignment ('ltsa.m')
- Diffusion maps ('diffmaps.m')
- Kernel PCA ('kpca.m')
- Stochastic Neighbor Embedding ('sne.m')
- Symmetric Stochastic Neighbor Embedding ('symsne.m')
- Neighborhood Preserving Embedding ('npe.m')
- Linearity Preserving Projection ('lpp.m')
- Linear Local Tangent Space Alignment ('lltsa')
- Maximum Variance Unfolding ('mvu.m')
- Bands Clustering ('bandclust.m')
- Affinity Propagation ('pa.m')
- Bands Progressif selection ('psa.m')

L'étude comparative a bien pu profiter des fonctions matlab. Cependant, quelques améliorations étaient indispensables pour adapter ces fonctions au contexte hyperspectrales. Par exemple, la fonction de matlab : 'entropy.m' est implantée pour des images classiques dont le pixel pouvant être soit de dimension égale à l'unité, soit un

vecteur de dimension égale à trois (la base RGB de l'image classique). Ainsi, la liste suivante a bien été écrite aussi :

- Variance ('variance.m')
- Mean Absolute Error ('mae.m')
- Mean Quadratic Error ('mqe.m')
- Maximum Absolute Distosion ('mad.m')
- Spectral Derivative ('sd1.m') et ('sd2.m')
- Entropie ('entropie.m')
- Fidelity ('fidelity.m')
- Spectral Similarity Index ('ss.m')
- Structural Content ('sc.m')
- Normalized Cross Corelation ('ncc.m')
- Rare Event Criteria ('rarevent.m')

L'évaluation du temps de calcul d'une méthode de réduction a du profiter des commandes tic et toc déjà prête sous matlab.

A ce stade, il est impératif de clarifier que plusieurs fonctions ont aussi été développées au service d'un programme principale ('mainprog.m') qui a permis de gérer l'ensemble des techniques et des critères sans besoin d'intervention de notre part. Ces fonctions intermédiaires sont :

- Conversion des datas 3D en un tableau 2D ('tri2two.m')
- Conversion d'un tableau 2D en datas 3D ('two2tri.m')
- Perturbation des tableaux 2D ('disturbdatas.m')
- Sauvegarde de tous les résultats pour chaque itération de calcul pour éviter les pertes d'informations lors des fonctionnements imprévus du PC ('saveitern.m')
- Impressions et Affections automatiques des résultats dans différentes figures avec génération des sous-titres et des légendes convenables ('plotfigures.m')

Le dernier chapitre était également l'occasion de mettre en œuvre plusieurs nouvelles fonctions matlab, ainsi que le programme principale ('hybrid.m') qui a permis de valider notre nouvelle approche de réduction des datas. Les sous-éléments additionnels de ce programme sont :

- Fuzzy K-Mean Clustering ('myfcm.m')
- Partition Index ('clustcheck.m')
- Dissimilarity test ('dsstest.m')
- Intrinsinc dimentionality for reduction ('idcorr.m')
- Reconstruction de l'image ('invpca.m')