

17/621.751

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université 8Mai 1945 – Guelma
Faculté des Sciences et de la Technologie
Département de Génie Electrotechnique et Automatique



Domaine : Sciences et Technologie
Filière : Automatique et Informatique industrielle
Spécialité : Commande et Diagnostic
des Systèmes Industriels

Mémoire de fin d'études
pour l'obtention du diplôme de Master Académique

Classification Des Images Satellitaires Par La Méthode De
L'analyse En Composantes Principales

Présenté par :
Kouahla Hichem

Sous la direction de :
Monsieur : Aidoud Mohamed

JUIN 2012



Remerciements



C'est avec un grand plaisir que nous apportons ce modeste travail à tous ceux qui nous ont gratifiés de leur soutien et de leur confiance.

Louanges à Dieu, qui nous a donné vie et santé pour le parachèvement de ce travail.

Je tenais à exprimer mon profonde gratitude à mon encadreur Monsieur Aidoud.M, pas Je le remercie pour sa disponibilité, ses encouragements, son soutien moral, sa sympathie et son amitié sincère dont il nous fait preuve de le long de ce travail. Nous apprécions la confiance qu'il nous a témoigné et les conseils avisés qu'il nous a prodigués et n'oublier pas aussi monsieur boualeg a.halim .

Je remercie mafamille, qui durant mes études, me ont toujours donné la possibilité de faire ce que je voulais et qui ont toujours cru en moi.

Enfin, je remercie tous ceux qui ont soutenu et aider de près ou de loin pour la réalisation de ce travail.

Kouahla HICHEM

Dédicaces

*Je remercie le Dieu de m'avoir donné le
courage et la force pour accomplir ce modeste travail
que je dédie à :*

*Mon premier enseignant c'est mon père et ma
chère mère qu'est sacrifiés pour moi.*

Mes chers frères : Mouhamed Nadjib, Zinedinne,

Mes chère sœur Soumia, Aicha

Tous mes amis,

Toutes les personnes qui me sont chères,

Ma famille

Kouahla et Douakha

Ainsi , toute la promotion 2012 .

Kouahla HICHEM

Résumé

Les images satellitaires sont d'une importance capitale. Elles représentent un vecteur essentiel d'informations. L'analyse de ces images permet aux acteurs en charge de la gestion de l'espace et de l'aménagement du territoire de localiser les zones d'intérêt et de suivre leur évolution au cours du temps.

Le sujet abordé dans ce mémoire entre dans le cadre du traitement des images satellitaires, dont le but principal est de classer des images satellitaires en zones selon la texture et la couleur donc de synthétiser de nouvelles images à partir des images originales sur laquelle une région de texture et/ou de couleur uniforme sera désignée par une même couleur (attribut utilisé pour discriminer deux régions différentes). Pour cela, nous avons appliqué la méthode de l'analyse en composantes principales utilisant la distance euclidienne pour la décision.

Nous avons effectué des tests sur la base de projection elle-même, mais avec des prises de vues différentes, cela a conduit à des résultats très satisfaisants où le taux d'apprentissage a atteint les 100%. Ensuite, nous avons effectué les tests de reconnaissance sur la base de tests. Les résultats sont illustrés aussi dans le tableau, où nous constatons que le système de reconnaissance que nous venons de mettre en œuvre est acceptable et donne des résultats satisfaisants 95.8333%

Sommaire

Titre	Page
<i>Introduction générale :</i>	03
Chapitre 1: TRAITEMENT D'IMAGES SATELLITAIRE	
1. Introduction :	05
2. L'image :	05
2.1. Types d'images :	10
2.1.1. Image matricielle (ou image bitmap) :	10
2.2 Représentation des couleurs :	11
2.2.1. Images 24 bits (ou couleurs vraies) :	12
2.2.2. Images à palettes« images en 256 couleurs (8 bits) :	12
2.2.3. Images en teintes (ou niveaux) de gris :	13
2.3. Formats d'images :	14
3. Image satellitaire :	14
3.1 Qu'est une image satellitaire ?	14
3.2. Intérêts de l'image satellitaire :	15
3.3. Système de télédétection :	15
3.4. Application de l'image satellitaire :	16
4. Types d'images satellite :	18
4.1. Image panchromatique :	18
4.2. Image multi bande :	19
4.3. Image radar :	19
5. Techniques de traitement de l'image satellitaire :	21
5.1. Compression d'image :	21
5.2. La segmentation :	21
6. Domaines d'application du traitement d'images :	23
7. Conclusion :	23
Chapitre 2: LA CLASSIFICATION	
1. Introduction :	24
1.1 Classification non supervisée (ou non dirigée) :	24
1.2 Classification supervisée ou dirigée :	25
2. Principe de la segmentation introduite dans les étapes de la classification :	25
3. Utilisation de la segmentation dans la classification :	27
3.1 Dans la classification : vers une classification par objet :	27
3.2. Pour l'amélioration de classification a posteriori :	28
3.3. Aide à l'extraction d'objets :	28
3.4. Pour l'analyse du paysage :	28
4. Les algorithmes de classification développés :	28
4.1. Les modes opératoires des classifications :	29
5. Algorithme général de la classification :	30
5.1. Version détaillée :	31
5.1.1. Apprentissage :	31
5.1.2. Classification initiale :	31
5.1.3. Itérations extérieures :	32
5.1.4. Faire décroître la température :	32
5.1.5. Poursuivre les balayages pour un nombre déterminé d'itération ou jusqu'à la réalisation d'un critère d'arrêt :	33
6. Conclusion :	33
Chapitre 3: ANALYSE EN COMPOSANTES PRINCIPALES34	
1. Introduction :	34
2. Analyse Multifactorielles :	34
3. Présentation Générale de l'Analyse de Composantes Principales (ACP) :	35
4. Objectif de l'Analyse en Composantes Principales :	36

Sommaire

5. Fonctionnement de l'ACP :	37
5.1. Principe :	37
5.2. Formulation mathématique de l'ACP :	37
6. Représentation des individus lors d'une ACP :	40
7. Interprétation des résultats d'une ACP :	40
7.1. Base sur les variables :	41
7.2. Base sur les individus :	42
8. Domaines d'applications :	42
9. Avantages et inconvénients de L'ACP :	43
9.1. Avantages :	43
9.2. Inconvénients :	43
10. Conclusion :	43
Chapitre 4: CONCEPTION ET REALISATION	
1 .Introduction :	44
2. Besoin matériel:	44
3. Besoin logiciel :	44
3.1. Matlab :	44
3.2 Les images utilisées & Google Earth :	46
1 .Analyse en composantes principales :	47
4.1. Les étapes Principale de L'ACP pour la classification des données :	47
4.2 : ACP détaillée :	48
4.3 : Organigramme de L'ACP:	49
5 .Mise en œuvre de la base de données :	49
6. Application de L'ACP :	50
6.1. Reconnaissance ;	53
7. Test & évaluation :	54
8. Conclusion :	55
<i>Conclusion générale et perspectives :</i>	56
<i>Bibliographie :</i>	57

Introduction générale

Depuis l'espace, on peut observer la terre et utiliser ces observations pour prévoir le temps, étudier l'évolution des climats, mesurer si le trou dans la couche d'ozone s'agrandit, détecter et suivre des feux de forêt ou des nappes de pétrole, créer des Cartes géographique.

Les images satellitaires sont d'une importance capitale. Elles représentent un vecteur essentiel d'informations. L'analyse de ces images permet aux acteurs en charge de la gestion de l'espace et de l'aménagement du territoire de localiser les zones d'intérêt et de suivre leur évolution au cours du temps.

Les images de la Terre, en prise de vue verticale (image satellitaire ou issue d'une campagne aérienne classique) ou oblique (image prise du sol ou d'un aéronef léger), se caractérisent par la diversité de leurs modes d'enregistrement (verticale, oblique, altitudes diverses, panchromatique, radar, etc.) et par la masse considérable de données qu'elles représentent. Le problème se pose alors de l'automatisation de certaines tâches de repérage et de préparation à l'interprétation intelligente. Si l'opérateur humain reste le plus efficace pour extraire les informations qualitatives et sémantiques, il est parfois mal adapté pour l'extraction rapide d'informations quantitatives (mesures de superficies, dénombrement d'objets).

Notre travail consiste à classifier des images satellitaires en zones selon la texture et la couleur donc de synthétiser de nouvelles images à partir des images originales ou nous avons appliqué la méthode de l'analyse en composantes principales

La classification des images satellitaires par l'analyse en composantes principales (ACP), est le travail global se partage en quatre chapitres, *chapitre 1, traitement d'images satellitaires*, Nous avons consacré la première partie de ce chapitre aux concepts généraux de l'image : Sa numérisation, ces caractéristique allant du pixel à la texture, et ce traitement : la compression, et la segmentation...etc. Et pour finaliser nous avons abordé les différents domaines d'application du traitement d'images.

Dans la deuxième partie, nous l'avons consacré aux images satellitaires, où nous avons commencé par définir l'image satellitaire, puis on a présenté les différentes étapes du système de télédétection, leur domaine d'application ainsi que leur type.

Introduction générale

Une étape d'une grande importance qui est le traitement des images satellitaire a été développé en fin de chapitre ; allant de l'utilité de cette dernière jusqu'à la technologie utilisée pour ce système. Le deuxième *chapitre est consacré* à définir le concept de classification avec ces deux types supervisé et non supervisé ; après nous avons représenté les différentes méthodes de classification et surtout ces intérêts (les méthodes) qui ont apportés sur les images satellitaires. Le chapitre trois est consacré de présenté l'ACP (l'Analyse en Composante Principale) faisant partie des méthodes d'analyse Multifactorielle et considéré comme méthode d'analyse de donnée qualitatives souvent exploité pour la classification. Nous avons aussi présenté les objectifs de la méthode ainsi que les défèrent étapes nécessaires à sa réalisation. et enfin Nous avons vu dans le dernier chapitre la réalisation de l'application de l'algorithme d'Analyse des Composant Principale sur des images satellitaires, on peut constater que l'objectif qui a été mis au début a été réalisé dans la totalité, avec un système de classification sur un type d'image difficile à maîtriser, la similarité entre eux.

Nous terminons notre travail par une conclusion générale ainsi que les perspectives de notre travail.

CHAPITRE 1

TRAITEMENT D'IMAGES SATELLITAIRE

1. Introduction :

Le traitement d'image voit son origine en physique des particules, dans les années 1950, Lorsque les scientifiques cherchent à déterminer les trajectoires de particules en analysant des milliers d'images par expérience dans des chambres à bulles. Jusqu'à la fin des années 1960, les images acquises sont généralement de mauvaise qualité et très volumineuses. Leur Traitement consiste, en premier lieu, à les restaurer pour corriger les défauts liés à l'acquisition ; les améliorer pour rendre l'image plus « belle » à l'affichage ; et les compresser pour réduire leur volume. Très vite, il est apparu nécessaire d'interpréter les images et l'analyse de scènes est devenue un aspect essentiel du traitement.

2. L'image ?

Contrairement aux images dessinées sur du papier, les images manipulées par un ordinateur sont numériques (représentées par une série de bits). L'image numérique est l'image dont la surface est divisée en éléments de tailles fixes appelés cellules ou pixels (Voir la figure X), ayant chacun comme caractéristique un couleur prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description [2].

La description de l'image est la conversion de celle-ci de son état analogique (distribution continue d'intensités lumineuses dans un plan xOy .) en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques $f(x,y)$ où :

- x, y : coordonnées cartésiennes d'un point de l'image.
- $F(x, y)$: niveau de gris en ce point

L'image est un ensemble structuré d'informations caractérisé par plusieurs paramètres : Pixel, Résolution, Contraste, Luminance, Bruit, Histogramme, Dimension, Contours, Texture.

Pixel ou « Picture éléments »: le plus petit point de l'image, c'est une entité calculable qui peut recevoir une structure et une quantification. Si le bit est la plus petite unité d'information que peut traiter un ordinateur, le pixel est le plus petit élément que peuvent manipuler les matériels et logiciels (Voir la figure 1.1). [2]

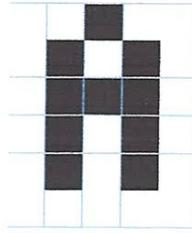


Figure 1.1 : La lettre 'A' en groupe de pixel

Résolution, c'est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. La résolution détermine le nombre de points par unité de surface, exprimé sur les moniteurs d'ordinateurs en points par pouce (PPP), en anglais DPI pour Dots Per Inch). Un pouce représentant 2.54cm. La résolution permet ainsi d'établir le rapport entre le nombre de pixels d'une image et la taille réelle de sa représentation sur un support physique. (Voir la figure 1.2). [2]



Figure 1.2 : Exemple sur La résolution

Contraste, C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images. Si L_1 et L_2 sont les degrés de luminosité respectivement de deux zones voisines A_1 et A_2 d'une image, le contraste C est défini par le rapport :

$$C = \frac{L_1 - L_2}{L_1 + L_2}$$

Luminance, C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, qui correspond à l'éclat d'un objet. Une bonne luminance se caractérise par :

- Des images lumineuses (brillantes),
- Un bon contraste : il faut éviter les images où la gamme de contraste tend vers le blanc ou vers le noir ; ces images entraînent des pertes de détails dans les zones sombres ou lumineuses,
- L'absence de parasites.



Figure 1.3 : Exemple sur La résolution luminance

Bruit, Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.



Figure 1.4 : Exemple de bruit d'image

Histogramme, est un graphique statistique permettant de représenter la distribution des intensités des pixels pour chaque intensité lumineuse. Par convention un histogramme représente le niveau d'intensité en abscisse en allant du plus foncé (à gauche) au plus clair (à droite). Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire l'information utile de celle-ci. [BS06]

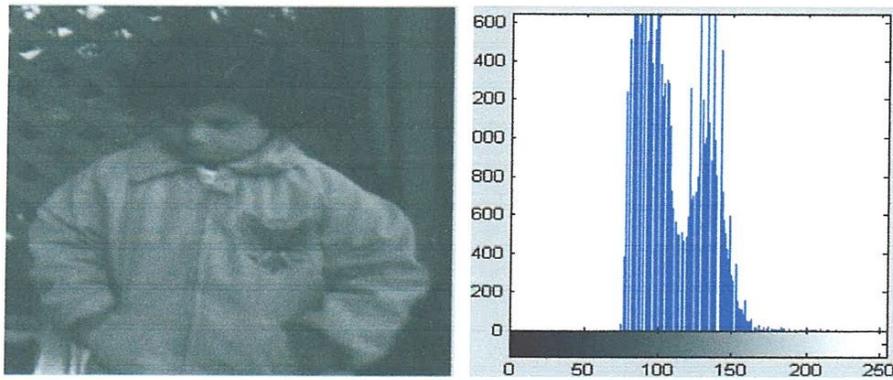


Figure 1.5 : Exemple d'histogramme d'image au niveau gris

Dimension, L'image se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels qui représente ce qu'on appelle par la dimension informatique ou taille de l'image.

Une image possédant 640 pixels en largeur et 480 en hauteur aura une définition de 640 pixels par 480, notée 640*480.

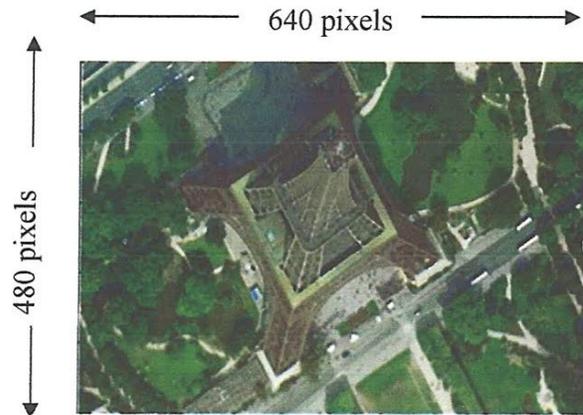


Figure 1.6 : Exemple de dimension

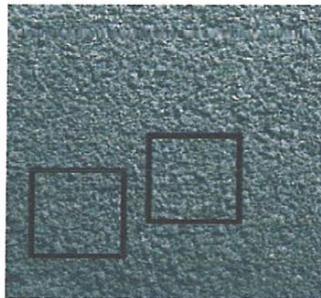
Contours, Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative.

La texture, Une **texture** est une région dans une image numérique qui a des caractéristiques homogènes. Ces caractéristiques sont par exemple un motif basique qui se répète, ou des caractéristiques fréquentielles. Une texture est composée de texels, l'équivalent des pixels.

La texture est un concept qui traduit un aspect homogène local de la surface de l'objet. C'est un concept très important est largement utilisé en traitement d'image. Malgré cette importance, la notion de texture reste un concept sans définition universelle satisfaisante. De nombreuses définitions ont été proposées mais aucune ne convient parfaitement aux différents types de texture rencontrés.

La définition littéraire de la texture est la suivante: "répétition spatiale d'un même motif dans différentes directions de l'espace". Cette définition est limitative car elle caractérise l'objet indépendamment d'un observateur humain.

Donc, il n'existe pas de définition précise de la texture. Une définition générale peut caractériser une texture comme un ensemble de primitives arrangées selon des règles particulières de placement. Une primitive est un ensemble connexe plus ou moins important de pixels de niveaux de gris à peu près semblables : il s'agit en fait d'un motif de base.



Figur1.7 : Exemple de texture

En pratique, on distingue deux grandes classes de textures, qui correspondent à deux niveaux de perception : [Ga83]

- Le micro texture : qui présentent un aspect régulier, sous formes de motifs répétitifs spatialement placés selon une règle précise (ex : peau de lézard, mur de brique).

- Le macro texture : Présentent des primitives " microscopiques" distribuées de manière aléatoire (ex : sable, laine tissée, herbe).

2.1 .Types d'images :

Parmi les façons de stockage des images qui existent :

- Images matricielles.
- Images vectorielles.

2.1 .1. Image matricielle (ou image bitmap) :

Elle est composée comme son nom l'indique d'une matrice (tableau) de points à plusieurs dimensions, chaque dimension représentant une dimension spatiale (hauteur, largeur, profondeur), temporelle (durée) ou autre (par exemple, un niveau de résolution). Il y a trois types : image 2D, image 3D, image stéréoscopique.

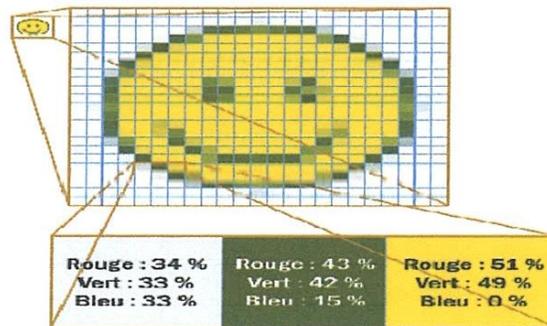


Figure 1.8 : Exemple d'image matricielle

Dans le cas des images à deux dimensions (le plus courant), les points sont appelés pixels. D'un point de vue mathématique, on considère l'image comme une fonction de $\mathbb{R} \times \mathbb{R}$ dans \mathbb{R} où le couplet d'entrée est considéré comme une position spatiale, le singleton de sortie comme un codage. Ce type d'image s'adapte bien à l'affichage sur écran informatique (lui aussi orienté pixel) ; il est en revanche peu adapté pour l'impression, car la résolution des écrans informatiques, généralement de 72 à 96 ppp (point par pouce).



Figure 1.9 : Exemple d'image 2D

Ces cas sont une généralisation du cas 2D, la dimension supplémentaire représentant respectivement le temps, une dimension spatiale ou une échelle de résolution.

D'un point de vue mathématique, il s'agit d'une fonction de $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ dans \mathbb{R}

- Lorsqu'une image possède une composante temporelle, on parle d'animation.
- Dans le cas des images à trois dimensions les points sont appelés des voxels. Ils représentent un volume.



Figure 1.10 : Exemple d'image 3D

2.2 Représentation des couleurs :

S'il existe plusieurs modes de représentation de la couleur, le plus utilisé pour le maniement des images numériques est l'espace couleur Rouge, Vert, Bleu (R, V, B). Cet espace couleur est basé sur la synthèse additive des couleurs, c'est à dire que le mélange des trois composantes (R, V, B) donne une couleur. Il existe différents types d'images couleurs en fonction du nombre de bits utilisés pour le stockage de l'information couleur:

- Images en vraies couleurs (24 bits).
- Images à 256 couleurs (8bits).
- Images en teintes de gris.

2.2.1. Images 24 bits (ou couleurs vraies) :

Il s'agit d'une appellation trompeuse car le monde numérique (fini, limité) ne peut pas rendre compte intégralement de la réalité (infinie). Le codage de la couleur est réalisé sur trois octets, chaque octet représentant la valeur d'une composante couleur par un entier de 0 à 255. Ces trois valeurs codent généralement la couleur dans l'espace RVB. Le nombre de couleurs différentes pouvant être ainsi représenté est de 256 x 256 x 256 possibilités, soit près de 16 millions de couleurs. Comme la différence de nuance entre deux couleurs très proches mais différentes dans ce mode de représentation est quasiment imperceptible pour l'œil humain, on considère commodément que ce système permet une restitution exacte des couleurs, c'est pourquoi on parle de « couleurs vraies ».

Les images bitmap basées sur cette représentation peuvent rapidement occuper un espace de stockage considérable, chaque pixel nécessitant trois octets pour coder sa couleur.

R	V	B	Couleur
0	0	0	noir
0	0	1	nuance de noir
255	0	0	rouge
0	255	0	vert
0	0	255	bleu
128	128	128	gris
255	255	255	blanc

Tableau 1.1 : Valeur des couleurs

2.2.2. Images à palettes« images en 256 couleurs (8 bits) :

Ici l'information couleur est codée sur 1 octet au lieu de trois. En utilisant une palette de couleur "attachée" à l'image. Chaque pixel va non plus véhiculer le code couleur RVB qui lui est affecté, mais simplement un chiffre compris entre 0 et 255. A chacun de ces chiffres va

correspondre une couleur, définie par son code RVB et stockée dans une palette avec les données de l'image.



Figure 1.11: Image numérisée avec 256 couleurs

2.2.3. Images en teintes (ou niveaux) de gris :

En général, les images en niveaux de gris renferment 256 teintes de gris. Image à 256 couleurs, simplement chacune de ces 256 couleurs est définie dans la gamme des gris.

Par convention la valeur zéro représente le noir (intensité lumineuse nulle) et la valeur 255 le blanc (intensité lumineuse maximale).

000	008	016	024	032	040	048	056	064	072	080	088	096	104	112	120	128
255	248	240	232	224	216	208	200	192	184	176	168	160	152	144	136	

Tableau 1.2 : Valeurs des niveaux de gris et teintes de gris correspondantes



Figure 1.12: Image en niveaux de gris

2.3. Formats d'images :

Un format d'image est une représentation informatique de l'image, associée à des informations sur la façon dont l'image est codée et fournissant éventuellement des indications sur la manière de la décoder et de la manipuler.

La plupart des formats sont composés d'un en-tête contenant des attributs (dimensions de l'image, type de codage, etc.), suivi des données (l'image proprement dite). La structuration des attributs et des données diffère pour chaque format d'image. De plus, les formats actuels intègrent souvent une zone de métadonnées (metadata en anglais) servant à préciser les informations concernant l'image comme :

- la date, l'heure et le lieu de la prise de vue,
- les caractéristiques physiques de la photographie (sensibilité ISO, vitesse d'obturation, usage du flash...)

Ces métadonnées sont par exemple largement utilisées dans le format EXIF (extension du format JPEG), qui est le format le plus utilisé dans les appareils photo numériques.

Quelques précautions à prendre concernant les formats d'images :

- Il faut prêter attention aux différentes versions que peut recouvrir un format particulier. Notamment pour le format TIFF qui varie selon les versions ; certaines d'entre elles ne sont pas reconnues par certains logiciels. [BG05]

3. IMAGE SATELLITAIRE:

3.1 Qu'est une image satellitaire ?

Depuis l'espace, on peut observer la terre et utiliser ces observations pour prévoir le temps, étudier l'évolution des climats, mesurer si le trou dans la couche d'ozone s'agrandit, détecter et suivre des feux de forêt ou des nappes de pétrole, créer des Cartes géographiques...etc.

A l'heure actuelle, on voit des images satellitaires partout: dans les informations météo à la TV, dans les journaux et même parfois dans les publicités.

Les images satellitaires, ou images de **télé-détection** sont des images prises à partir d'une 'lune artificielle' (un satellite) qui orbite bien au-dessus de la terre.

On appelle également les images satellitaires des images **'Remote Sensing'**:

Remote est le terme anglais pour éloigné ou à grande distance et *Sensing* signifie sentir ou balayer.



Figure 1.13 : Exemple d'image satellitaire

3.2. Intérêts de l'image satellitaire :

L'image satellitaire présente plusieurs intérêts :

- Elle est obtenue de manière numérique et ne nécessite donc pas de conversion de données.
- Elle est universelle : elle ne connaît pas de frontière géographique ou politique.
- Elle est aussi synthétique, flexible, précise et actuelle (renouvellement de l'information, automatique ou à la demande).

3.3. Système de télédétection :

Le rayonnement émis par une source d'énergie ou d'illumination (A) parcourt une certaine distance et interagit avec l'atmosphère (B) avant d'atteindre la cible (C).

L'énergie interagit avec la surface de la cible, en fonction des caractéristiques du rayonnement et des propriétés de la surface. Le rayonnement est réfléchi ou diffusé vers le capteur (D), qui l'enregistre et peut ensuite transmettre l'énergie par des moyens électroniques à une station de réception (E) où l'information est transformée en images (numériques ou photographiques).

Une interprétation visuelle et/ou numérique de l'image (F) est ensuite nécessaire pour extraire

l'information que l'on désire obtenir sur la cible. La dernière étape du processus consiste à utiliser l'information extraite de l'image pour mieux comprendre la cible, pour nous en faire

découvrir de nouveaux aspects ou pour aider à résoudre un problème particulier (G).

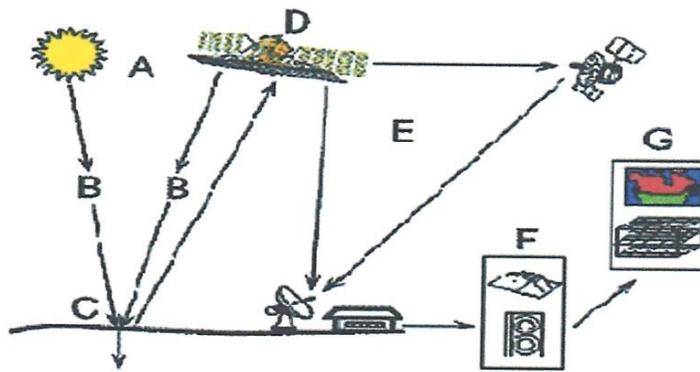


Figure 1.14: Etapes de la télédétection

3.4. Application de l'image satellitaire :

a. Les spatiocartes, Une spatiocarte numérique est un produit image obtenu à partir d'une ou plusieurs scènes originales (assemblées par mosaïquées numérique). Elle couvre un site précis quelles que soient les contraintes d'acquisition des données. Ces spatiocartes sont destinées aux personnes ayant besoin d'informations géographiques et sont donc présentées dans un découpage cartographique normalisé afin d'être accessibles à tous les utilisateurs. [1]

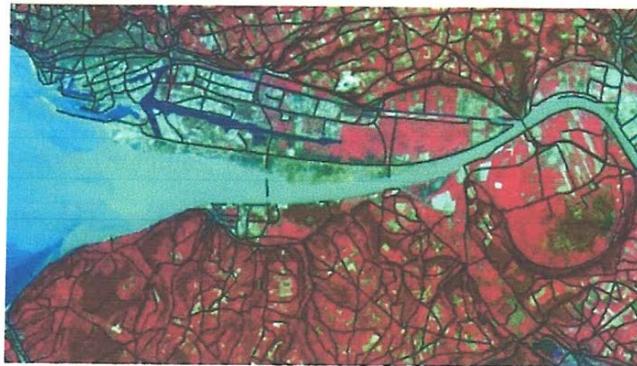


Figure 1.15 : Exemple de modèle spatiocartes

b. Les modèles numériques de terrain (MNT) Le modèle numérique de terrain est le fichier maillé des altitudes d'une région. IL est présenté sous la forme d'une grille régulière, Une altitude tous les 20m par exemple.

Les MNT, éventuellement croisés avec d'autres informations, permettent la réalisation cartes de pentes, d'ensoleillement ou de ruissellement, utilisées dans le cadre de l'aménagement de sites (parcs de loisirs, prévention des risques naturels,... etc.). Ils peuvent aussi être utilisés pour prévoir les tracés des constructions linéaires (canaux, lignes de TGV,

d. Suivi des subsidences La subsidence ou son corollaire l'inflation de terrain ont des causes diverses: exploitations minières souterraines, stockage de gaz souterrain, installations d'équipement souterrain, pompages de nappes... etc.

L'impact sur l'environnement et la sécurité des ouvrages souterrains nécessitent une activité de surveillance. L'objectif de cette surveillance est de préciser le comportement à moyen et long terme des cavités, d'évaluer leurs répercussions en surface, de mettre au point des méthodes de surveillance des zones à risques et enfin de concevoir des procédés de réduction des risques.

Parmi les techniques utilisées, certaines sont fondées sur des méthodes de mesures directes (nivellement ou positionnement par satellite GPS, extensomètres...), d'autres sur des données obtenues à partir d'images satellitaires par interférométrie optique ou radar. Les performances de ces méthodes sont liées à la précision des mesures de déformations, à la densité des réseaux de capteurs et à la pertinence géostatistique de ces réseaux.

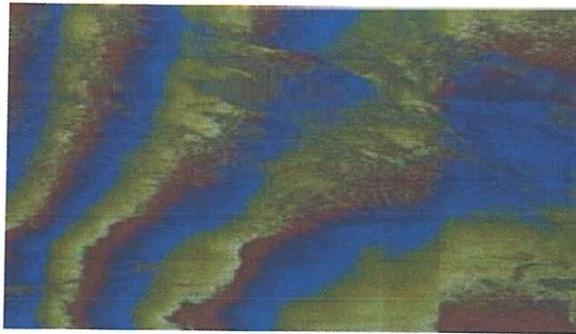


Figure 1.18: Exemple de Suivi des subsidences

4. Types d'images satellite :

Selon l'application que l'on souhaite obtenir, on dispose de plusieurs types d'image :

4.1. Image panchromatique :

Elle est acquise par un capteur numérique qui mesure la réflectance dans une large bande électromagnétique. Les données panchromatiques sont représentées sous forme d'images en noir et blanc. Elle sert principalement à obtenir des informations de type " géométrique " (formes, dimensions, surface). [1]

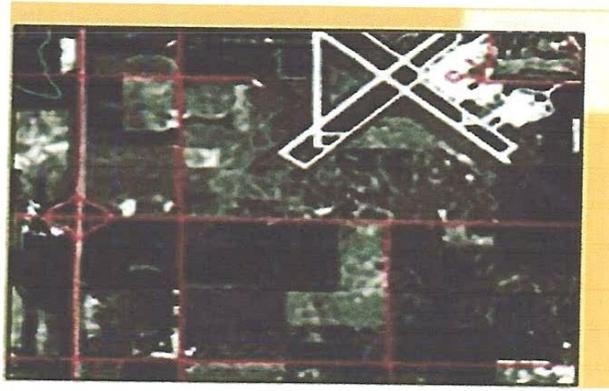


Figure 1.19: Image panchromatique

4.2. Image multi bande :

Elle est acquise par un capteur numérique qui mesure la réflectance dans de nombreuses bandes spectrales (typiquement, une dizaine). Ces multiples valeurs de réflectance se combinent pour créer des images couleur. Elle sert principalement à obtenir des informations à caractère " qualitatif " (cx. classification des types de végétation sur des parcelles).

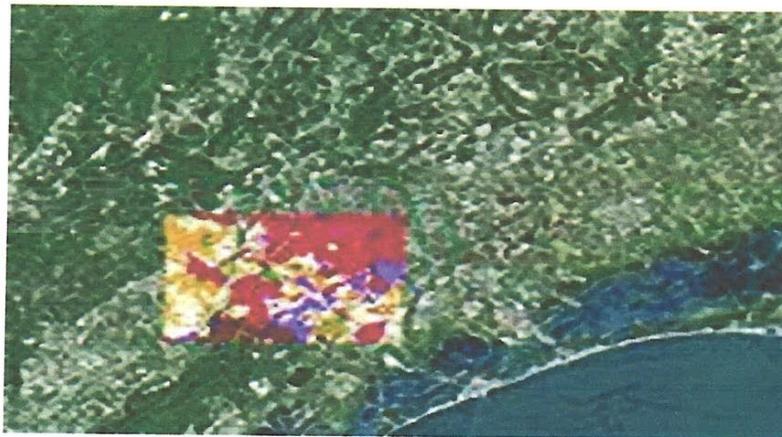


Figure 1.20 : Image multi bande

4.3. Image radar :

Le radar imageur, système actif, qui permet l'élaboration d'images (typiquement 100 km x 100 km) de jour comme de nuit même à travers les nuages. Le radar est sensible aux propriétés diélectriques des milieux (teneur en eau), et à leur rugosité. Il est également sensible aux structures géométriques à l'échelle de la longueur d'onde. [1]

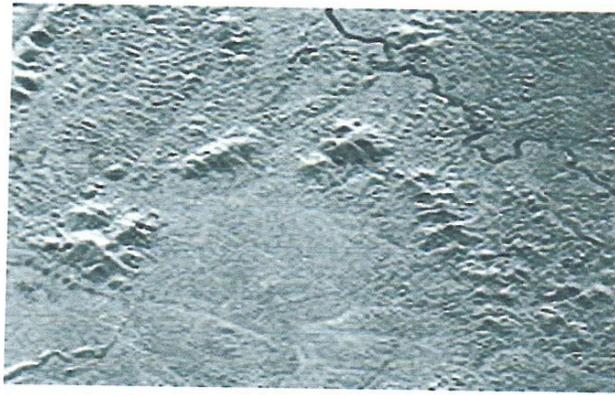


Figure 1.21 : Image radar

Utilité

Le traitement des images de la terre permet de tirer profit de l'information apportée par les caractéristiques spectrales et spatiales de l'image. Les signatures spectrales peuvent révéler la lithologie des roches de surface, alors que la texture. L'utilisation des images radar à des fins cartographiques a beaucoup progressé depuis quelques années permettant la réalisation automatisée de cartes topographiques ainsi que des cartes thématiques portant sur la végétation, la géologie, l'océanographie, ... etc. De nombreuses applications se trouvent aussi dans le domaine militaire, dans la prospection minière, la météorologie, l'hydrologie ainsi que plusieurs autres. [Ro98]

Evolution

L'utilité des images satellitaires s'est accrue par l'introduction de nouvelles technologies pour capter les images et par l'utilisation d'algorithmes poussés de traitement d'images et de techniques de l'intelligence artificielle telle les réseaux Neuroniques.

Technologie

L'évolution technologique a permis d'utiliser les domaines du visible, de l'infrarouge et des micro-ondes et de les combiner afin d'améliorer la résolution et d'augmenter la quantité d'information contenue dans l'image. Des instruments beaucoup plus sophistiqués sont utilisés comme le radar à synthèse d'ouverture qui, grâce à sa sensibilité à la rugosité et à la morphologie de surface, nous fournit un nouvel outil pour entreprendre l'étude de cet environnement. Dans notre projet, les images traitées sont celles acquises par google earth qui utilise le satellite Spot. [Ro98]

5. Techniques de traitement de l'image satellitaire :

Le traitement des images satellitaires consiste à détecter des caractéristiques tels la réflectance spectrale, la texture, dans le cas où les structures géologiques auraient une influence marquée sur le paysage, des phénomènes physiques plus profonds. Par exemple, les différentes textures rencontrées dans une image servent à la classer en zones de végétation différentes, en zones urbaines, etc.

Auparavant, les géologues se servaient de beaucoup moins d'outils dans leur traitement. De nos jours, l'évolution dans le traitement d'image et dans l'intelligence artificielle conduit à de nombreuses applications dans le domaine, par exemple, la correction géométrique, le filtrage de bruit, la détection du gradient avec des filtres directionnels. [Ro98]

Les techniques de traitement des images satellitaire se basent sur les techniques de traitement de base, dans la suite on va passer en revue les principales techniques existe dans la littérature

5.1. Compression d'image :

Compression d'image est le procédé inhérent à la compression d'image consiste à réduire la dimension du fichier de données en ne retenant que les informations essentielles. Le taux de compression est spécifié en utilisant un coefficient de compression CR (compression rate), défini comme le rapport des dimensions du fichier non comprimé et du fichier compressé. Les meilleurs procédés de compression sont ceux qui offrent le meilleur rapport entre un CR élevé et les informations essentielles retenues. Évidemment, cela est fonction de l'application visée. Par exemple, dans une image binaire comportant seulement des informations textuelles, l'information pertinente n'implique que le texte, comparativement à une image médicale dont l'information pertinente à extraire peut consister en des détails très fins.

5.2. La segmentation :

La segmentation d'image est une opération de traitement d'images qui a pour but de séparer des objets du fond d'une image numérique. L'homme sait naturellement séparer des objets dans une image. Pour séparer un objet d'un fond, l'humain se base sur des connaissances de haut niveau qui lui permettent de détecter ce qui l'intéresse dans l'image. À ce jour, il existe de nombreuses méthodes de segmentation, que l'on peut regrouper en quatre principales classes d'algorithmes: [3]

1. Segmentation basée sur les régions (en anglais : *region-based segmentation*). On y trouve par exemple : la croissance de région (en anglais : *region-growing*), décomposition/fusion (en anglais : *split and émerge*)
2. Segmentation basée sur les contours (en anglais : *edge-based segmentation*).
3. Segmentation basée sur une approche globale de l'image (par exemple : seuillage(en anglais : *thresholding*), histogramme, approches basées sur le nuage couleur...etc.)
4. Segmentation basée sur la coopération entre les trois premières segmentations

a. Segmentation par approche « région »

De cette technique est d'assigner un numéro (aussi appelé *label* ou *étiquette*) à chaque région; chacun de ces identifiants est associé à une couleur ou à un niveau de gris, ce qui permet de construire une *carte des régions* représentant visuellement le découpage en régions de l'image. Un objet peut ensuite être défini comme un ensemble de régions. [GM93]

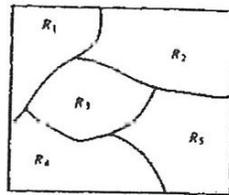


Figure 1.22 : Exemple de segmentation par approche région

b. Segmentation par approche « frontière »

Cette approche est basée sur la détection la différences dans la texture et dans les régions voisines. Ainsi des frontières sont détectées où il y a des différences dans la texture. Cette méthode n'a pas besoin de savoir le nombre de régions texturées dans l'image à l'avance. [CB06]

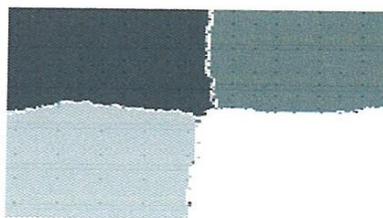


Figure 1.23 : La segmentation sépare les régions de textures différentes

c. Segmentation basée sur une approche globale de l'image (classification)

On part ici d'un rapport qu'entretient chaque pixel individuellement avec des informations calculées sur toute l'image, comme par exemple la moyenne des niveaux de gris de l'ensemble des pixels, ou la médiane, permettant de segmenter l'image en 2 régions comportant le même nombre de pixels. Ces informations permettent de construire des classes de pixels, les pixels appartenant à une même classe et étant connexes formant des régions.

6. Domaines d'application du traitement d'images :

Le traitement d'images possède l'aspect multidisciplinaire. On trouve ses applications dans des domaines très variés tels que les télécommunications (T.V., vidéo, publicité,...), la médecine (radiographie, ultrasons,...), biologie, astronomie, géologie, l'industrie (robotique, sécurité), la météorologie, l'architecture, l'imprimerie, l'armement (application militaire), la segmentation des images satellitaire. De nouvelles applications pratiques sont possibles aujourd'hui et touchent tous les domaines d'activités, tels que : métiers du spectacle, de la radio, créations artistiques,...etc.

7. Conclusion :

Nous avons consacré ce chapitre aux concepts généraux de l'image : Sa numérisation, ces caractéristique allant du pixel à la texture, et ce traitement : la compression, et la segmentation... etc. Et pour finaliser nous avons abordé les différents domaines d'application du traitement d'images.

Dans la deuxième partie nous l'avons consacré aux images satellitaires, où nous avons commencé par définir l'image satellitaire, puis on a présenté les différentes étapes du système de télédétection, leur domaine d'application ainsi que leur type. Une étape d'une grande importance qui est le traitement des images satellitaire a été développé en fin de chapitre ; allant de l'utilité de cette dernière jusqu'à la technologie utilisée pour ce système.

CHAPITRE 2

LA CLASSIFICATION

1. Introduction :

Une classification ou système de classification est un système organisé et hiérarchisé de catégorisation d'« objets ».

La classification est un processus de reconnaissance des formes. En télédétection, elle consiste à effectuer la correspondance entre les éléments d'une scène de l'image matérialisés généralement par leurs valeurs radiométriques, et des classes connues a priori ou non par un utilisateur. La correspondance est réalisée par des fonctions discriminantes sous forme de règle de décision telle que le "maximum de vraisemblance" des probabilités, ou des distances géométriques.

La classification peut être interprétée comme un problème de données incomplètes. En chaque site (Dans notre cas le pixel) s'expriment des données de deux types différents : la donnée observée et l'autre *cachée, la classe que l'on cherche à trouver*. Ainsi, la classification vise à attribuer aux pixels des étiquettes dont l'origine est thématique.

Très généralement, soit un ensemble de points auxquels sont attribués M scalaires représentant des "caractéristiques", pour une image multidimensionnelle elles correspondent aux différents canaux.

Deux types d'approches peuvent être distinguées : les classifications non supervisées sans aucune connaissance *a priori* sur les données, et les classifications supervisées. Ces méthodes qu'elles soient supervisées ou non, partagent des principes communs.

1.1 Classification non supervisée (ou non dirigée) :

La classification non supervisée ou non dirigée détermine automatiquement les classes. Les algorithmes des centres mobiles, (nuées dynamiques appliquées au centre de gravité) et des distances adaptatives sont des méthodes itératives qui, à chaque itération calculent respectivement les centres de gravité ou les distances associées aux classes. Les fonctions discriminantes dépendent, soit des paramètres des densités de probabilité, soit d'un jeu de coefficients qui peuvent être évalués théoriquement par des études physiques, la classification est alors totalement non supervisée. Elle peut être semi-supervisée quand certains paramètres sont fixés par l'utilisateur en fonction du type d'image et de scène, ou que certains objets sont reconnus par l'utilisateur qui peut ainsi initialiser le processus de recherche des classes pour certaines d'entre elles.

L'estimation non dirigée de θ à partir de Y est approchée par des algorithmes basés sur le maximum de vraisemblance, la solution des équations de vraisemblance est approchée par des méthodes itératives.

1.2 Classification supervisée ou dirigée :

Les classes sont définies à partir des besoins de l'utilisateur et correspondent à des unités sémantiques de l'image, ce qui nécessite une étape d'apprentissage préalable à la classification. La classification supervisée consiste à déterminer les classes par un apprentissage qui peut être effectué soit sur chaque scène traitée, soit pour chaque type d'application (agriculture, forêt, littoral, géologie, ...) ou encore à chaque type de capteur utilisé (optique : Landsat, Spot, Ikonos, Noaa, ou radar : ERS, JERS1, Radarsat...). Cette étape influence fortement les résultats de la classification.

Pour l'estimation dirigée (avec apprentissage), on dispose pour chaque classe de l'image des vecteurs caractéristiques de plusieurs pixels de l'image (échantillons des classes) dont on connaît *a priori* la classe d'appartenance. En supposant le processus ergodique, les paramètres sont estimés à l'aide des méthodes classiques d'estimation telles que la méthode du maximum de vraisemblance.

Les échantillons sont sélectionnés sur des observations faites sur le terrain (vérités terrain) en repérant au sol des parcelles représentatives des différents éléments devant être présent dans la classification de la zone étudiée, si possible à une date idéalement identique à celle des prises de vues. A ces vérités terrain peuvent s'ajouter des connaissances thématiques obtenues auprès d'un expert. Les échantillons de classes stables temporellement (routes, urbain, etc...), peuvent être sélectionnés sur la base de cartes précises de la région.

Les classifications supervisées sont nécessaires lorsque les besoins thématiques sont pointus. En effet les techniques non supervisées se limitent souvent à la reconnaissance de grandes classes thématiques (forêt, urbain, surface nue, végétation haute et basse et surface d'eau).

2. Principe de la segmentation introduite dans les étapes de la classification :

La méthode de segmentation a été développée à l'ENSEEIH (IRIT/LIMA) [Fjortoft, Marthon *et al* 95] pour son application aux images optiques et modifiée au CESBIO pour les images radar.

L'approche locale de détection de contours consiste à balayer l'image avec une fenêtre définissant la zone d'intérêt. Un opérateur est appliqué sur les pixels de la fenêtre afin d'estimer s'il y a une transition significative au niveau de l'attribut choisi. L'opérateur choisi

est celui de Shen et Castan [Shen & Castan 1992] car, il est robuste au bruit, il permet une bonne localisation, pour les images optiques et le ROEWA (Ratio Of Exponentially Weighted Average) qui tient compte du speckle, pour les images radars [Fjortoft 99]. L'opérateur de Shen et Castan est optimal pour un bruit additif blanc gaussien et une distribution de Poisson de contours. Il obtient de très bons résultats selon des critères de bonne détection de contours : robustesse au bruit, bonne localisation, et faible multiplicité des maxima dus au bruit. Cet opérateur consiste à calculer la différence entre les moyennes calculées de chaque côté du pixel central, en introduisant une pondération exponentielle négative au niveau des moyennes de support infinie. L'opérateur ROEWA adapte l'opérateur de Shen et Castan pour les données RSO à partir du rapport des moyennes.

Une fois l'image de puissance de contour créée, il faut en extraire des contours fermés et amincis pour établir une segmentation. L'algorithme de ligne de partage des eaux, est une méthode issue de la morphologie mathématique, qui extrait directement des contours fermés et squelettisés à partir d'une image de puissance de contour (image de module de gradient). L'image de puissance est interprétée comme une surface, dont les lignes de crête, sont supposées correspondre aux contours de l'image origine. Pour détecter les lignes de crête, on simule une inondation de la surface. Comme la méthode crée une région pour chaque minimum local dans l'image de puissance de contour, elle a tendance à être sur-segmentée, d'autant plus que l'image est bruitée. Une méthode plus robuste au bruit est appliquée : le seuillage des dynamiques de bassin à partir d'une carte de puissance de contour, réalisant ainsi une représentation hiérarchique des différents niveaux de segmentation de l'image [Fjortoft *et al.* 1998,1999, Lemaréchal *et al.* 1998]. Cette méthode est particulièrement intéressante pour réduire le nombre de faux contours. La dynamique du bassin associée à un minimum local est définie comme la différence d'altitude entre ce minimum, et le point de crête le plus bas qui doit être franchi pour arriver dans un bassin ayant un minimum plus bas que le premier. On définit la dynamique d'un arc de contour comme étant la valeur maximale du seuil pour lequel cet arc de contour subsiste, lorsqu'on supprime tous les bassins ayant une dynamique inférieure au seuil

L'image de dynamique de contour n'est pas binaire. Chaque arc de contour a la valeur de sa dynamique de contour. Le concept de dynamique de contour permet de créer une représentation hiérarchique et compacte des segmentations obtenues en appliquant différents seuils aux dynamiques de bassin.

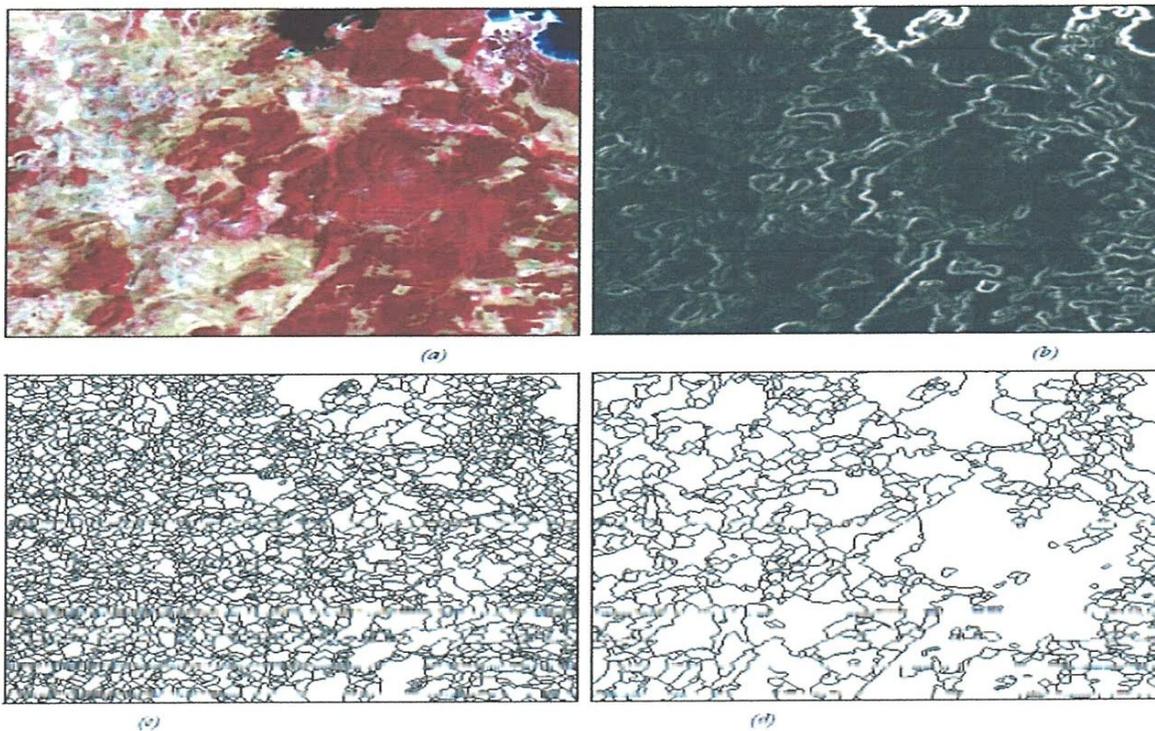


Figure 2.1 (a) extrait d'une image TM d'une zone forestière et agricole (b) image de puissance, (c) niveau de segmentation 10, (d) niveau de segmentation 30

3. Utilisation de la segmentation dans la classification :

Le système de classification et d'analyse du paysage qu'utilise la segmentation à plusieurs niveaux : dans la classification, en post classification, dans l'analyse du paysage, dans l'apprentissage automatique. Le choix du seuil de segmentation est délicat, car il dépend des objectifs de la segmentation. Une part de décision est laissée à l'utilisateur, afin de le moduler en fonction des résultats souhaités.

3.1 Dans la classification : vers une classification par objet :

La classification peut être effectuée sur les régions déterminées par la segmentation, on ne s'intéresse plus au pixel mais à un objet correspondant à un écosystème ou à une partie de cet écosystème. Il est préférable d'utiliser une segmentation fine (sur-segmentée), car il vaut mieux de petits segments, la classification permettant de corriger d'éventuelles erreurs de segmentation. Ainsi plusieurs petits segments seront affectés à la même classe, ces segments

appartenant au même objet. Par contre il ne faut pas classer deux objets de nature différente dans une même classe. Ceci peut se produire en augmentant le seuil de segmentation : des parcelles proches au sens de l'homogénéité, mais différentes sémantiquement, sont rassemblées, ce qui ne sera pas le cas en sur-segmentant.

Pour les images RSO, les parcellaires agricoles, constituées de champs homogènes, sont généralement bien adaptées à une classification basée sur la segmentation. En général les structures apparaissent bien dans ce type d'image.

3.2. Pour l'amélioration de classification *a posteriori* :

Nous avons développé diverses méthodes qui utilisent le contexte, dans le cas le plus simple, le principe correspond à éliminer des points isolés et effectuer un lissage. L'apport d'une image segmentée permet de conserver des contours nets et les structures donc il faut une segmentation plutôt moyenne s'approchant de la parcelle.

3.3. Aide à l'extraction d'objets :

La segmentation est une aide intéressante pour l'extraction d'objets tels que les nuages, les villes, les rivières ... Elle constitue une information supplémentaire (mais non suffisante) aux algorithmes d'extraction.

3.4. Pour l'analyse du paysage : calcul de l'hétérogénéité,

L'indice d'hétérogénéité se calcule à partir de la segmentation. Des combinaisons d'images sous segmentées et sur segmentées, permettent de détecter des sous classes, dans des parcelles hétérogènes.

4. Les modes opératoires des classifications :

Trois modes opératoires de classification ont été implantés : le mode *ponctuel*, *contextuel* ou local et *global*. Rappelons que la plupart des méthodes présentées ont été réalisées de façon supervisée ou non.

Les méthodes non supervisées sont itératives. Nous avons mis au point leur version supervisée qui est initialisée à partir d'échantillons et exécutée avec peu d'itération. Ce procédé d'abord utilisé pour les images radar car chaque itération réduit le speckle, s'est également révélé intéressant pour les images optiques car il diminue des confusions de classes, et fait disparaître les points isolés, et comme une image segmentée est introduite, les structures sont mieux dessinées. Ces méthodes supervisées ou non, partagent des principes communs.

4.1 Classification ponctuelle :

Les méthodes de classification bayésiennes ponctuelle (pixel à pixel), utilisent un jeu de probabilités *a priori*, constant sur toute l'image, qui ne dépend pas du contexte du pixel. En pratique, les caractéristiques d'un pixel dépendent, sinon de l'ensemble des pixels de l'image, tout au moins d'un voisinage de celui-ci. Par la suite nous verrons que la modélisation du mécanisme de formation des régions de l'image par un processus markovien permet de ramener la considération des pixels de l'image entière à celle de ses plus proches voisins.

La première étape consiste à déterminer les termes d'observation (probabilité conditionnelle) et d'*a priori*. Pour le terme d'observation, on approche la distribution $P(y_s|x_s=k)$ par la distribution empirique de l'image sur les échantillons correspondants à la classe k (pour le supervisé). Une autre solution consiste à choisir un système de distributions paramétriques et à ajuster les paramètres aux échantillons prélevés de manière à obtenir les mêmes moments. Dans le cas d'images RSO, on approche la distribution $P(y_s|x_s=k)$ de chaque classe par des lois Gamma ayant la même moyenne et le même écart-type que les échantillons. Si les régions sont homogènes, la loi de Gauss peut être utilisée, sinon les lois K ou le système de Pearson seront préférés. Pour les images optiques, la loi de Gauss est appropriée.

Une méthode a permis d'introduire la notion de texture dans le cas ponctuel comme la méthode de la moyenne des classes sous hypothèse gaussienne et la texture sous hypothèse d'une distribution gamma dans le modèle multiplicatif du speckle (gamma-texture) (Nezry et al (96)).

Pour la deuxième étape, il suffit d'affecter à chaque pixel la classe qui maximise sa probabilité d'observation. Le problème d'optimisation se limite ici à la comparaison directe de N_c probabilités. La qualité de la classification est liée au taux de recouvrement des classes. Dans le cas d'images RSO brutes c'est à dire non filtrées, la dynamique conduit à un fort recouvrement des classes. Les données radar, offrent des aspects très différents. Par exemple, une route peut apparaître très réfléchissante ou non selon son orientation (pente). En y ajoutant le chatoiement, le recouvrement des classes est alors trop important pour garantir des résultats fiables. Les résultats sont insuffisants, et donnent à la classification ponctuelle un aspect *pointilliste*. Le taux de classification pour un canal de type ERS/JERS/RADARSAT est de 15% à 25% de pixels bien classés en moyenne suivant ces capteurs. Pour avoir des résultats corrects avec ces méthodes, il vaut mieux, au préalable, filtrer les images. Il peut être préférable de transformer les images par le logarithme des amplitudes dans ce cas l'approximation gaussienne est plus acceptable Pour le cas multi temporel (à partir d'au

moins 3 dates) et si les images sont filtrées, les taux de bien classés sont de l'ordre de 60 à 70%, ce taux augmentant avec le nombre de dates. Pour les images optiques la loi Normale multi variée est une approximation correcte de la distribution :

$$p(\mathbf{y} | \omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |C_i|^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(\mathbf{y}-\mu_i)^T C_i^{-1}(\mathbf{y}-\mu_i) \right]}$$

où

d : la dimension de la variable aléatoire correspondant au nombre de canaux traités simultanément C_i : la matrice de covariance et μ_i le vecteur moyen de la classe ω_i . L'avantage de cette méthode est la simplicité d'application, la prise en compte des corrélations inter-dimensions, notamment temporelles, la possibilité d'ajouter aux canaux de base, des canaux synthétiques tels que des textures locales ou autres s'ils ont une distribution approchant une gaussienne.

-La loi K

Pour les images radar, la loi K est plus précise que la distribution gaussienne, quand la réflectivité est distribuée selon une loi Gamma. Bien que la loi K soit une meilleure approximation, l'estimation des paramètres nécessaires à ces fonctions discriminantes peut générer des erreurs annulant le gain de cet ajustement plus précis surtout quand le nombre de pixels des échantillons est faible.

5. Algorithme général de la classification :

- Choisir une configuration initiale, aussi proche que possible d'une configuration optimale : classification initiale par maximum de vraisemblance.
- Balayer l'ensemble des sites s de Ω , et en chaque site, calculer l'ensemble des probabilités conditionnelles déduites de $P_T(\mathbf{x})$, et sélectionner l'état qui correspond à la probabilité maximale. On donne à \mathbf{x} l'étiquette qui minimise E .
- Poursuivre les balayages jusqu'à convergence ou pour un nombre déterminé d'itérations ou jusqu'à la réalisation d'un critère d'arrêt.

5.1. Version détaillée :

5.1.1. Apprentissage :

Echantillonnage

- ◆ Dans le cas de la classification *supervisée*, cette étape consiste seulement à récupérer les vérités terrains fournies par l'utilisateur pour obtenir un jeu d'échantillons pour l'estimation des statistiques initiales et un autre pour la vérification.
- ◆ Pour la classification *non supervisée*, les échantillons sont créés automatiquement de différentes façons suivant l'utilisation :
 - espacés régulièrement de taille variable en fonction des dimensions de l'image, l'écart entre les échantillons de forme carrée est variable afin d'éviter le chevauchement.
 - tirés aléatoirement.
- ◆ Dans le cas *semi supervisé* il y a introduction de vérités-terrain, ou d'échantillons interprétés visuellement sur l'image auquel s'ajoute des échantillons créés automatiquement, espacés régulièrement ou aléatoirement.

Le nombre de classes est adapté à la distribution dans ces deux cas.

Estimation initiale des statistiques

Cette partie correspond à la détermination des vecteurs statistiques des différents attributs (canaux,...)

Poser $T=T_0$, la température initiale.

5.1.2. Classification initiale :

Chaque pixel est classé dans chaque lot selon la règle du maximum de vraisemblance et une loi de distribution au choix (gaussienne, Gauss-Wishart, Pearson ou gaussienne complexe) ou barycentrique. Puis il est affecté à la classe dont la moyenne des probabilités sur tous les lots est la plus élevée. Cette classification initiale est effectuée de manière ponctuelle pour les images optiques, ou de manière contextuelle pour les images radar. Si un lot rassemble plusieurs jeux de données, la fusion de ces données est réalisée lors de la classification globale.

5.1.3. Itérations extérieures : cette étape est composée de quatre étapes :

- *Calcul de la matrice transition inter dates* : calcul des contraintes temporelles (éventuellement)
- Balayer l'ensemble des sites s de Ω , et en chaque site :
 - Calculer l'ensemble des probabilités conditionnelles déduites de $P_T(\mathbf{x})$.
 - Le type de règle de classification peut être différent selon les données : modèle barycentrique, gaussien pour les images optiques (Spot, Landsat ...). Système Pearson pour les images radars (ERS, JERS ...), distribution gaussienne complexe (SLC).
 - Le classement est effectué dans une certaine proportion fixée par un poids qui permet d'accorder plus ou moins d'importance à une donnée.

Les images par dates (ou lots) peuvent être classées en parallèle, ou séparément puis une image multi-lots (ou multi-dates) est faite.

✚ *Itérations intérieures*

Ces itérations consistent à classer les lots de façon séquentielle (séparément), selon la règle du maximum de vraisemblance. Chacune de ces classifications tient compte du contexte (voisinage proche), des contours (contrainte spatiale à partir de la segmentation), des paramètres de transition, contraintes temporelles, forts diffuseurs et de la température.

✚ *Classification par lot*

Il s'agit des mêmes calculs que ceux effectués dans les itérations intérieures, mais les lots sont traités de façon parallèle (chaque pixel est classé dans chacun des lots).

✚ *Fusion des lots : classification globale*

Cette étape consiste à combiner pour chaque pixel les classes trouvées dans les lots (fusion), et d'en produire une image classée finale.

✚ *Ré-estimation des statistiques*

Il s'agit des mêmes calculs que ceux effectués lors de l'estimation initiale des statistiques.

5.1.4. Faire décroître la température : $T_{k+1} = T_k$ où k est le numéro de l'itération.

5.1.5. Poursuivre les balayages : pour un nombre déterminé d'itération ou jusqu'à la réalisation d'un critère d'arrêt.

Le nombre de classe peut être imposé par l'utilisateur ou être donné automatiquement, c'est-à-dire déterminé par une formule dépendant du nombre de la taille de l'image.

Des itérations extérieures, sans contrainte sont d'abord exécutées seules. Les itérations intérieures avec contraintes sont ensuite effectuées à température constante, puis la température est diminuée avec la nouvelle itération extérieure ainsi de suite jusqu'à la convergence vers la solution. Ce procédé raccourci les traitements et permet d'introduire les contraintes plus proches de la solution.

6. Conclusion :

Nous avons consacré ce chapitre à définir le concept de classification avec ces deux types supervisé et non supervisé ; après nous avons représenté les différentes méthodes de classification et surtout ces intérêts (les méthodes) qui ont apportés sur les images satellitaires.

CHAPITRE 3

ANALYSE EN COMPOSANTES PRINCIPALES

1. Introduction :

L'ACP est une technique de statistique descriptive dont le principe est simple mais qui met en œuvre des calculs numériques importants, pour cette raison elle n'a pu se développer qu'avec l'apparition des ordinateurs.

Il est à présent possible de traiter des données de taille conséquente sur micro-ordinateur.

L'ACP est à conseiller pour un premier examen, une mise en forme ou une présentation synthétique de données abondantes croisant des individus avec des variables quantitatives. On n'omettra cependant pas d'examiner préalablement les données par les méthodes statistiques usuelles (moyenne, écart-type, graphiques, corrélation, etc.).

Un reproche fréquemment adressé à l'ACP et aux techniques connexes est qu'elles ne révéleraient que des évidences. Le propos est injuste, mais il est rassurant que souvent les premiers axes retrouvent et confirment ce qui était déjà connu.

Comme avec les autres méthodes descriptives, il faut être très prudent pour inférer des modèles explicatifs ou causaux à partir des configurations obtenues

2. Analyse Multifactorielles :

Les méthodes multifactorielles permettent d'obtenir des représentations graphiques qui constituent le meilleur résumé possible de l'information contenue dans un grand tableau de données. Pour cela, il faut consentir à une perte d'information afin de gagner en lisibilité. En fonction des phénomènes que l'on veut étudier et de la nature du tableau de données dont on dispose, on appliquera telle ou telle méthode multifactorielle. En effet, il n'existe pas une méthode factorielle d'analyse des données, mais un ensemble de méthodes, reposant toutes sur les mêmes théories mathématiques. Ainsi, on trouvera les principales méthodes suivantes :

- **ACP** : Analyse en Composantes Principales, pour les tableaux de variables quantitatives.
- **AFTD** : Analyse Factorielle d'un Tableau de Distances, pour les tableaux de distances.
- **AFC** : Analyse Factorielle des Correspondances, pour les tableaux de contingence.
- **ACM** : Analyse des Correspondances Multiples, pour les tableaux de variables qualitatives.
- **STATIS** : Structuration des Tableaux A Trois Indices de la statistique.
- **AFM** : Analyse Factorielle Multiple.

- **DACP** : Double Analyse en Composante Principale.

Ces méthodes reposent toutes sur les mêmes notions théoriques, mais chacune produit un genre de résumé spécifique et s'applique sur un type de donnée précis. C'est pourquoi on devra choisir la méthode la plus adaptée au type d'information que l'on possède et aux phénomènes qu'on veut étudier.

3. Présentation Générale de l'Analyse de Composantes Principales (ACP) :

La plus simple et la plus connue des techniques d'Analyse de Données, qui a pour objectif de remplacer les variables initiales (numériques) par de nouvelles variables (également numériques) dites "**Composantes Principales**", qui ont les deux propriétés suivantes :

1) Elles peuvent être classées par ordre d'importance décroissante (on sait donner un sens précis à ce terme). Ainsi, les premières Composantes Principales (donc les plus importantes) suffisent pour rendre compte des données avec une perte minimale d'information. En d'autres termes, les données initiales peuvent être remplacées par de nouvelles données dans lesquelles figurent les mêmes individus, mais décrits par des variables en plus petit nombre.

2) Ces nouvelles variables sont deux-à-deux dé corrélées.

L'ACP peut donc être vu comme une technique de réduction de dimensionnalité.

En fait, l'ACP n'a pas été inventé dans ce but, mais dans celui de donner des individus et des variables des représentations planes pour examen visuel. Ces représentations (ou "projections") sont obtenues en ne retenant que deux Composantes Principales parmi les plus significatives. L'utilisation traditionnelle de l'ACP met l'accent sur l'interprétation des diagrammes visuels ainsi obtenus, cette interprétation exigeant de la part des utilisateurs une bonne expérience.

Quand on ne prend pas compte que les deux premières composantes principales, l'ACP peut être vue comme une projection du nuage de points des individus sur le plan passant de façon "optimale" à travers ce nuage.

On attend d'une ACP :

* Une interprétation des premières Composantes Principales en termes "métier" plus pertinents que ceux attachés aux variables originales.

* En conséquence, une interprétation des différentes zones des quelques plans de

projection les plus importants.

* Une aide à la clustering (parfois) visuelle des données.

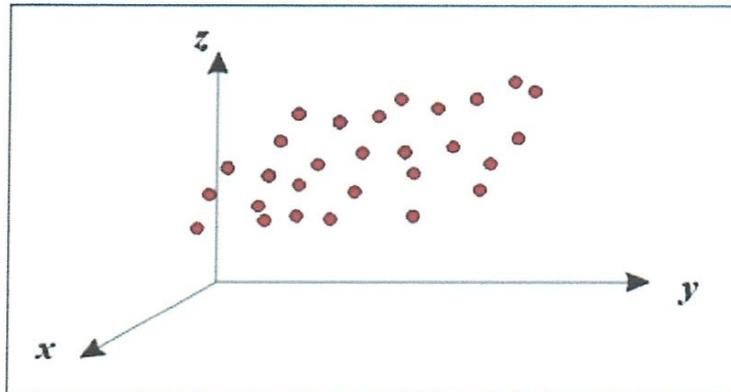


Figure : 3.1 : nuage de points représenté par 3 composantes (ou sur 3D) .

4. Objectif de l'Analyse en Composantes Principales :

L'ACP est une analyse multi variée qui cherche à identifier les axes principaux qui expliquent le mieux des corrélations entre variables descriptives. Cette méthode (ACP) peut être présentée comme une méthode numérique qui a pour but de :

- Décrire et représenter les ressemblances entre les individus par rapport à l'ensemble des variables.
- Décrire et représenter les corrélations linéaires entre les variables.
- Réduire efficacement le nombre de dimensions étudiées (et ainsi simplifier l'analyse), en cherchant à exprimer le plus fidèlement possible l'original de données grâce aux relations détectées entre les variables.

L'ACP est donc une bonne méthode pour déceler des liens entre variables dans un ensemble de données multidimensionnelles afin de comprendre leurs interactions et faciliter leur étude.

5. Fonctionnement de l'ACP :

5.1. Principe :

On a vu que l'objectif principal de l'ACP était de réussir à exprimer un système complexe de données de dimension quelconque, représenté par un tableau dit Individus/Variables, dans un petit nombre de dimensions, tout en minimisant la perte d'informations engendrée.

Ce processus consiste à construire de nouveaux caractères synthétiques à partir de facteurs (d'où le terme d'analyse multifactorielle) obtenus par combinaisons linéaires des variables initiales. Ce sont ces nouveaux caractères synthétiques qui sont appelés composantes principales.

La problématique consiste alors à trouver les n composantes principales qui contiennent la plus grande quantité d'information originale (n étant le nombre de dimensions auxquelles on réduit l'étude). Dans de nombreux cas, on se limite à 2 l'étude des données pourra alors être réduite à ces composantes principales, si elles sont suffisamment représentatives. Il faut donc pouvoir quantifier l'efficacité des composantes principales, pour ne retenir que les plus significatives.

Dans le cadre de la représentation graphique, les composantes principales sont assimilées aux axes. Ainsi, dans le cadre d'une représentation de données dans un espace à deux dimensions, l'axe des abscisses représentera la première composante principale, celle qui explique la plus grande partie de l'information d'origine, et l'axe des ordonnées sera la deuxième composante principale, celle qui explique la plus grande partie de l'information restante. [Bo04]

5.2. Formulation mathématique de l'ACP :

La mise en œuvre mathématique de l'ACP peut être divisée en 6 étapes principales: [Bo04]

Etape 1 : Préparation des données

On part d'un tableau de données rectangulaire, représentant toutes les données, en plaçant en ligne les individus et en colonnes les variables. Soit X ce tableau de données.

Nous avons alors : $X = \text{individus} \begin{bmatrix} x_1^1 & \dots & x_p^1 \\ \dots & \dots & \dots \\ x_1^n & \dots & x_p^n \end{bmatrix}$

- P variable, représentées en colonnes,
- n individus, représentés en lignes,
- les valeurs prises par chaque variable, pour chaque individu, notée :

$$X = (x_{ij}^i) (1 \leq i \leq n, 1 \leq j \leq p) \text{ et } \forall (i,j), x_{ij}^i \in \mathbb{R}$$

Le problème est que si nous analysons directement la matrice X, les résultats seraient faussés par les valeurs relatives des variables (Par exemple si les valeurs ont été mesurées dans des unités différentes). Préparer les données pour le traitement consiste donc à transformer le tableau de données de façon à réduire ces effets.

Nous pouvons alors travailler sur une version centrée_réduite X_{cr} de la matrice X.

Pour centrer la matrice X, on soustrait à chacune des valeurs la moyenne de sa variable.

Pour réduire la matrice centrée, on divise chacune des valeurs par l'écart-type de sa variable.

Etape 2 : Calcul de la matrice des coefficients corrélation

Dans cette étape, nous calculons la matrice de corrélations des données contenues dans le tableau X_{cr} , notée « Corr ». [Bo04]

Matrice de variance_covariance

$$X_{cr} = 1/n \sum_{k=1}^n X_{cr_{ki}} \cdot X_{cr_{kj}}$$

Matrice des coefficients de corrélation

$$\text{Corr} = \frac{X_{cvij}}{X_{cvii} \cdot X_{cvjj}}$$

Etape 3 : Calcul des valeurs et vecteurs propres de la matrice de corrélation

Les valeurs λ_i et vecteurs propres u_i de la matrice « Corr » sont les facteurs utilisés pour construire les composantes principales.

Où λ_i sont les solutions de l'équation détection $(R - \lambda_i * I) = 0$:

$$R * u_i = \lambda_i * u_i$$

Il s'agit d'utiliser la matrice des composantes principales précédemment calculée pour représenter les individus, sous forme de nuages de points, dans des plans ou espaces factoriels de deux ou trois composantes principales.

6. Représentation des individus lors d'une ACP :

Le problème initial était d'obtenir une représentation du nuage N dans des espaces de dimension réduite. On connaît maintenant les axes définissant ces espaces. Pour pouvoir obtenir les différentes représentations, il suffit de déterminer les coordonnées de la projection de tous les points du nuage sur chaque axe factoriel. Soit C_1^i, \dots, C_n^i ces n coordonnées pour l'axe i .

Le vecteur $C^i = \begin{bmatrix} c_1^i \\ \dots \\ c_n^i \end{bmatrix}$ est appelé $i^{\text{ème}}$ composante principale.

On peut alors obtenir "l'image" du nuage N dans un plan factoriel quelconque (u_i, u_j) grâce aux composantes principales c^i et c^j . La représentation dans le premier plan factoriel est obtenue grâce à c^1 et c^2 . En utilisant conjointement la représentation du plan (u_1, u_3) , on peut "voir" le nuage dans le sous-espace E_3 .

Le calcul des composantes principales se fait par changement de base. Il suffit de faire une projection orthogonale sur les nouveaux vecteurs de base. Ainsi, pour la $i^{\text{ème}}$ composante principale, on a : avec métrique de choisie noté : M

$$c^i = (c_j^i)_{1 \leq j \leq n} \text{ avec } c_j^i = M(u_i, x^j)$$

D'où l'expression de la composante principale : $c^i = XM u_i$

7. Interprétation des résultats d'une ACP :

Le but initial est atteint : on dispose d'un nombre réduite des variable « les composantes principales », qui sont non corrélés, mais que signifient ces nouvelles variables et comment les interpréter ? deux approches existent l'une base sur les individus et l'autre sur les variables.

7.1. Base sur les variables :

L'ACP construit de nouvelles variables artificielles et des représentations graphiques permettant de visualiser des relations entre variables

- Les corrélations, elles sont synthétisées dans un cercle imaginaire à deux axes ou plus. Chaque axe représente une composante principale. On regarde la position des variables par rapport à cet axe.
- Si les variables sont proches de la direction positive de l'axe alors elles sont corrélées.
- Si les variables sont proches de la direction négative de l'axe alors elles sont anti-corrélées.
- Si les variables sont proches d'un autre axe, alors elles sont non corrélées.
- Pour comparer des points entre eux on ne le fera que si ils sont proches de la circonférence du cercle.
- Par contre, on ne peut rien dire quand les variables sont agglomérées au centre du cercle, ou de la sphère unité.
- Les corrélations avec la métrique M choisie, s'observent dans un espace sphérique.

Toutes les corrélations calculées, on les représente dans un plan formé par un couple de composantes principales c^1 et c^2 , où elles figureront dans un cercle appelé cercle de corrélation. Dans ce plan, une variable x_j est repérée par un point de coordonnée $r(c^1, x_j)$ et d'ordonnée $r(c^2, x_j)$.

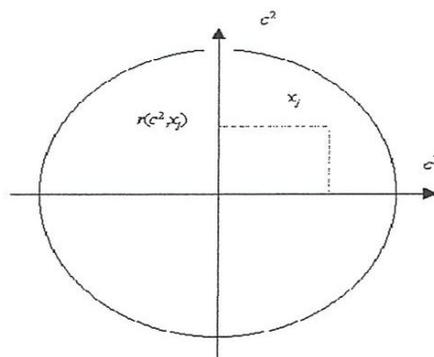


Figure : 3.2 : cercle de corrélation.

7.2. Base sur les individus :

On entant d'interpréter les composants principales à l'aide de la position des individus part à pour à ce composants la. la position est de termine par $c=x.u$ avec $u= \langle u_1, u_2, \dots, u_q \rangle$

- La qualité globale de représentation : est le pourcentage de l'inertie prés en conte par

les sous espace de deux dimension
$$\left[\frac{1}{\sum_{i=1}^p \lambda_i} (\lambda_1 + \lambda_2 + \dots + \lambda_q) \right] * 100$$

Pour interprète les résultat nous besoin de connaître d'autre informations tel que :

- Les contributions : qui permette d'identifier les individus très influents peuvent détermine a seul l'orientation de certain axes

$$CTR(i, \alpha) = \frac{1}{n} * \frac{(c_i^\alpha)^2}{\lambda_\alpha} * 100$$

i : indice de l'individu.

α : indice de la composantes principales.

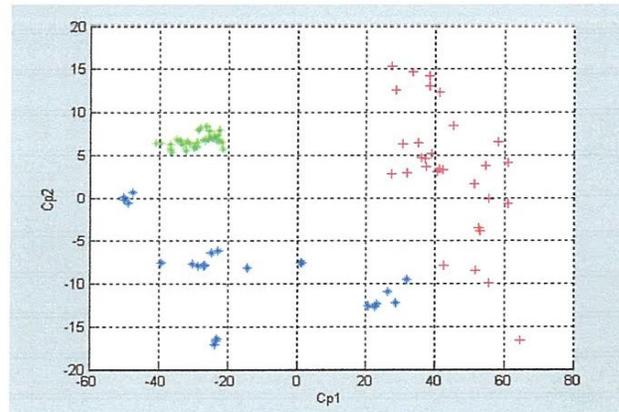


Figure : 3.3 : Diagramme de dispersion des individus selon les composante Cp1 et Cp2.

8. Domaines d'applications :

L'ACP trouve son application dans nombreux domaines :

- Etude des caractéristiques d'une population (sondage, étude de santé...).
- L'assurance qualité et fiabilité des produits.
- Etude des resulta sportifs.
- Exploitation d'expériences et des relevés scientifiques (en physique, chimie, géologie).

- L'indication économique (études de marchés, rentabilité d'investissements,... etc.)
- L'assurance qualité et les fiabilités de produit.
- L'étude de résultats sportifs.
- Le traitement de l'information (reconstitution de photos satellites, reconnaissance vocale, reconnaissance de visage... etc.)

9. Avantages et inconvénients de L'ACP :

9.1. Avantages :

Parmi les avantages proposés par L'ACP on cite : [DZ06]

- *Aune grande simplicité mathématique* : les seuls véritables outils mathématiques utilisés dans L'ACP sont le calcul des valeurs propres d'une matrice.
- *Simplicité des résultats* : Grace aux graphiques qu'elle fournit, L'ACP permet d'appréhender une grande partie de ses résultats simple coup d'œil.
- *Puissance et fiabilité* : C'est une méthode très souple, puisqu'elle s'applique sur un ensemble de données contenu et de taille quelconque, elle offre un résumé et une vue complète des relations existant entre les variables quantitatives d'une population d'étude.

9.2. Inconvénients :

En tant que d'analyse de données, l'ACP n'a pas réellement d'inconvénients en soi, elle s'applique simplement sur des cas précis et pour générer un type de résultat particulier. Ça n'aurait donc aucun sens de dire que c'est un inconvénient de l'ACP qu'elle ne s'applique pas en de hors de ce contexte. De même étant donné qu'il s'agit avant tout d'une technique de résumé de données. La perte d'information forcément en gendre n'est pas un inconvénient, mais plutôt une condition d'obtention du résultat. [Bo04]

10. Conclusion :

Dans ce chapitre, nous avons présenté l'ACP (l'Analyse en Composante Principale) faisant partie des méthodes d'analyse Multifactorielle et considéré comme méthode d'analyse de donnée qualitatives souvent exploité pour la classification.

Nous avons aussi présenté les objectifs de la méthode ainsi que les défèrent étapes nécessaires à sa réalisation.

CHAPITRE 4

CONCEPTION ET REALISATION

1. Introduction :

Dans ce chapitre nous présentons notre application qui consiste à implémenter un système de segmentation des images satellitaires en utilisant la méthode de l'analyse en composantes principales.

Données : *tableau $n \times p$ individus * variables (Les images) ;*

Objectif : recherche d'une typologie ou segmentation, c'est à dire d'une partition ou répartition des n individus dans des classes, sur la base de l'observation de p descripteurs (classification)

Moyen : chaque classe doit être la plus homogène possible et, entre elles, les plus distinctes possibles, au sens d'un critère à définir (ACP).

Question : *Pourquoi ACP ?*

Réponse : ACP = outil puissant pour analyser les corrélations entre plusieurs variables. Et il permet d'obtenir de nouvelles variables non corrélées à forte variance (c'est-à-dire fort pouvoir informatif).

2. Besoin matériel :

Pour le développement de notre logiciel nous avons utilisé une machine qui possède les caractéristiques matérielles suivantes :

- Processeur : *Pentium « IV ».*
- Vitesse d'horloge : *2.80 GHZ.*
- Capacité de disque dur : *74.4 Go.*
- Capacité mémoire : *256 Mo.*

3. Besoin logiciel :

Comme toute implémentation d'un système que ce soit son objectif, on a besoin de certains d'outil de développement.

Dans notre travail et a plusieurs raison nous avons choisis le Matlab comme notre outil de développement ; nous citons par exemple la performance qui le possède dans le domaine de traitement d'image et le calcul mathématique et certes ca facilité d'utilisation surtout par des gens moins expert dans le domaine de programmation.

3.1. Matlab :

Le MATLAB est un acronyme de «**MAT**rix **LAB**oratory», ce produit a été développé par la société MATH WORKS.

MATLAB est un logiciel de calcul matriciel à syntaxe simple. Avec ses fonctions spécialisées ;

MATLAB peut être aussi considéré comme un langage de programmation adapté pour les problèmes scientifiques.

Les instructions de MATLAB sont interprétées et exécutées ligne par ligne, et fonctionnent dans plusieurs environnement tels que Windows, Macintosh et Linux.

Il existe deux modes de fonctionnement :

- Mode interactif: MATLAB exécute les instructions au fur et à mesure qu'elles sont données par l'utilisateur.
- Mode exécutif : MATLAB exécute ligne par ligne un "fichier .M" (programme en langage MATLAB). Voir la figure 5.1

La structure de données la plus manipulable par le MATLAB sont les tableaux qui manipulent essentiellement des images.

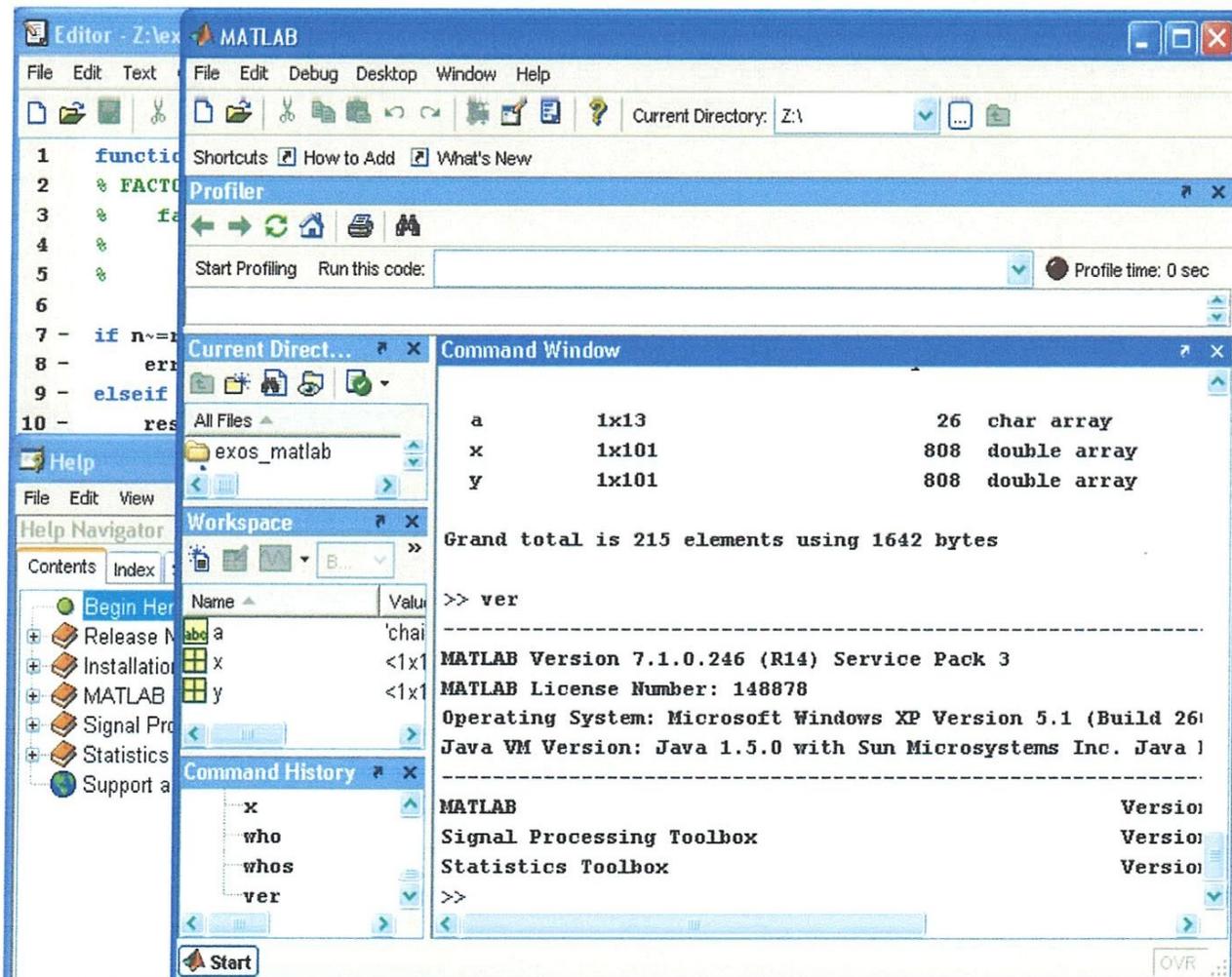


Figure 4.1 : fenêtre principale du MATLAB.

3.2. Les images utilisées & Google Earth :

Depuis sa sortie en août 2005, Google Earth est devenu un véritable phénomène, le but de ce logiciel est de fournir un guide urbain et géographique global, basé sur une représentation 3D détaillée et réaliste de la terre, en théorie, Google Earth permet de rechercher n'importe quoi sur notre belle planète ; une pizzeria, un itinéraire de son hôtel au cinéma le plus proche, le logiciel affiche un niveau de détail tel que l'on peut même avoir accès à une modélisation 3D (simple) des immeubles par-dessus les images satellites, elles-mêmes détaillées et tenant compte du relief. Google Earth est une mine d'informations sur notre planète et met des informations géographiques sur le monde entier à votre portée, en associant la puissance de la recherche Google et des images satellite, des plans, des cartes, des images en relief et des représentations 3D des bâtiments. Google Earth utilise le satellite spot, de résolution entre 2,5 à 5 mètre.

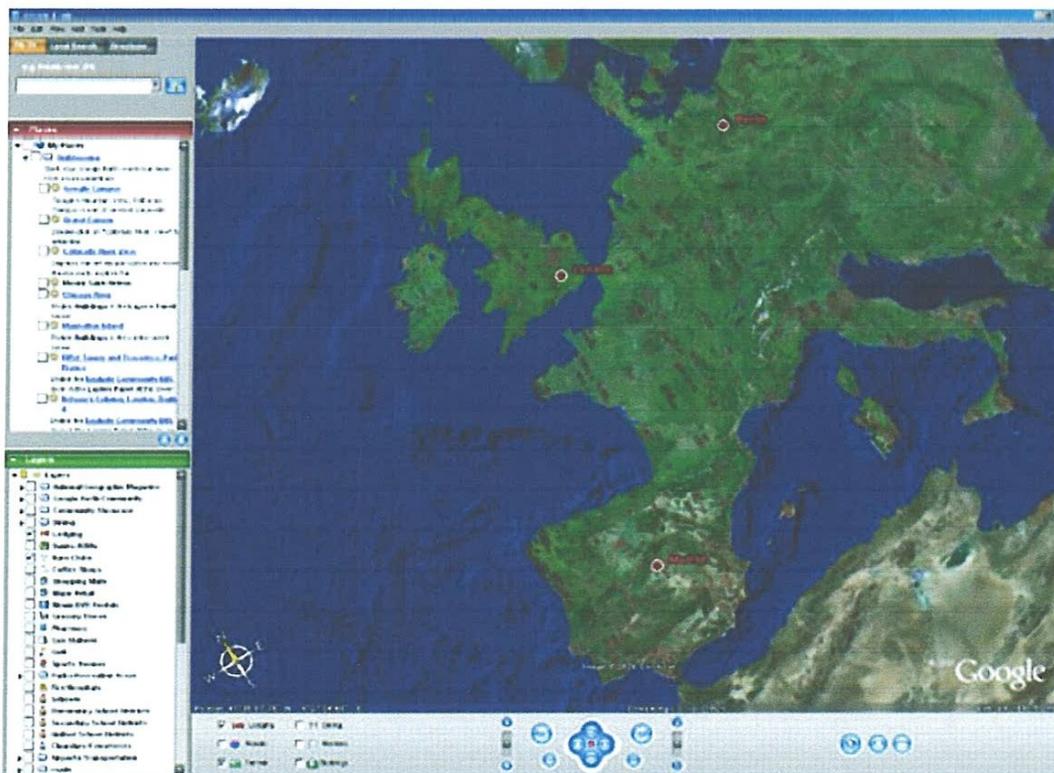


Figure 4.2 : Google Earth.



Figure 4.3 : Image satellitaire.

1. Analyse en composantes principales :

Nous avons utilisé le principe de la méthode d'analyse en composantes principales qui nous permet de passer de l'espace de représentation des échantillons qui est de très grande dimension à un espace de dimension réduit.

4.1. Les étapes Principale de L'ACP pour la classification des données :

Soit X la matrice (base de donnée) de dimension $n * m$, ou n : représente le nombre des de échantillons, et m : le nombre des variables

Les opérations nécessaires à ce traitement sont décrites ci-dessous :

Etape 1 : Calcul du tableau de données centrées et réduites par rapport aux variables \tilde{X}

$$X' = \frac{(X_i - X_m)}{S_i} \quad 4.1$$

Ou X_m : moyenne des X_i

Si : l'écart type par :

$$S_i = \sum_{i=1}^n ((X_i - X_m)^2 / n)^{1/2} \quad 4.2$$

Etape 2 : Calcul des matrices de variances V_{ii} , de covariances V_{ij} et de corrélation R_{ij}

Matrices de variances V_{ii} :

$$V_{ii} = \sum_{i=1}^n (X_i - X_m)^2 / n \quad 4.3$$

Matrices de covariances $V_{ij} : V_{ij} = \sum_{i=1}^n \sum_{j=1}^n (X_i - X_{mi})(X_j - X_{mj}) / n$ 4.4

Matrices de corrélation $R_{ij} : R_{ij} = V_{ij} / (v_{ii} v_{jj})^{1/2}$ 4.5

Etape 3 : Calcul des valeurs propres λ_i (voir Figure 4.2) et vecteur propres V_i de la matrices « Corrélations» sont les facteurs utilisés pour construire les composantes principales.

Où V_i sont les solutions de l'équation déterminant $(R - \lambda_i * 1) = 0$

$$R * V_i = \lambda_i * V_i \tag{4.6}$$

Etape 4 : Classement des vecteurs propres dans l'ordre décroissant des valeurs propres associées.

Etape 5 : Calcul des composantes principales.

La matrice (C) appelée la matrice des composantes principales CP est celle qui contient les coordonnées des échantillons dans l'espace formé par les composantes principales.

CP = matrice centrée réduite * vecteur propre $\Leftrightarrow CP = X_{cr} . U$ 4.7

Composantes principales (axes principaux) au lieu de base IJIS dimensions précédentes.

Etape 6 : Représentations graphiques.

4.2. ACP détaillée :

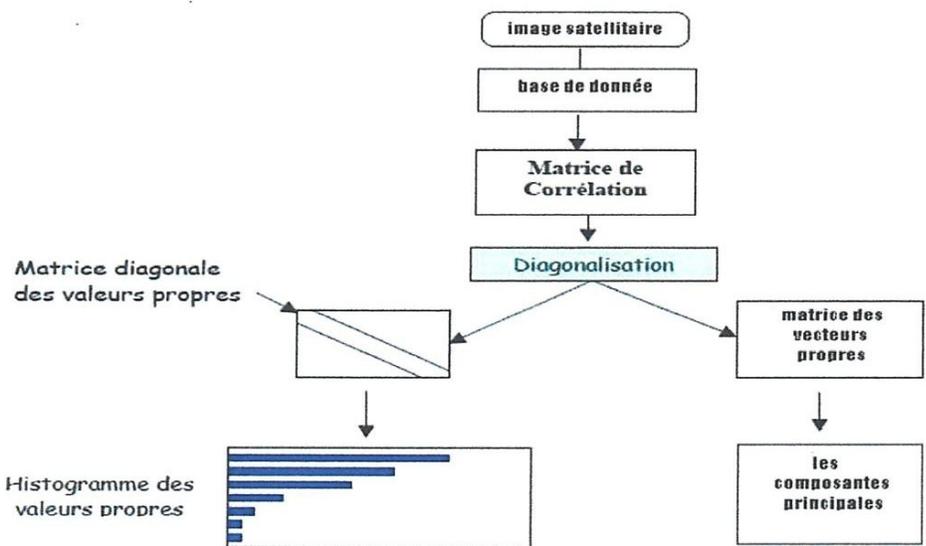


Figure 4.4 : détaillé de l'algorithme de classification ACP

4.3. Organigramme de L'ACP :

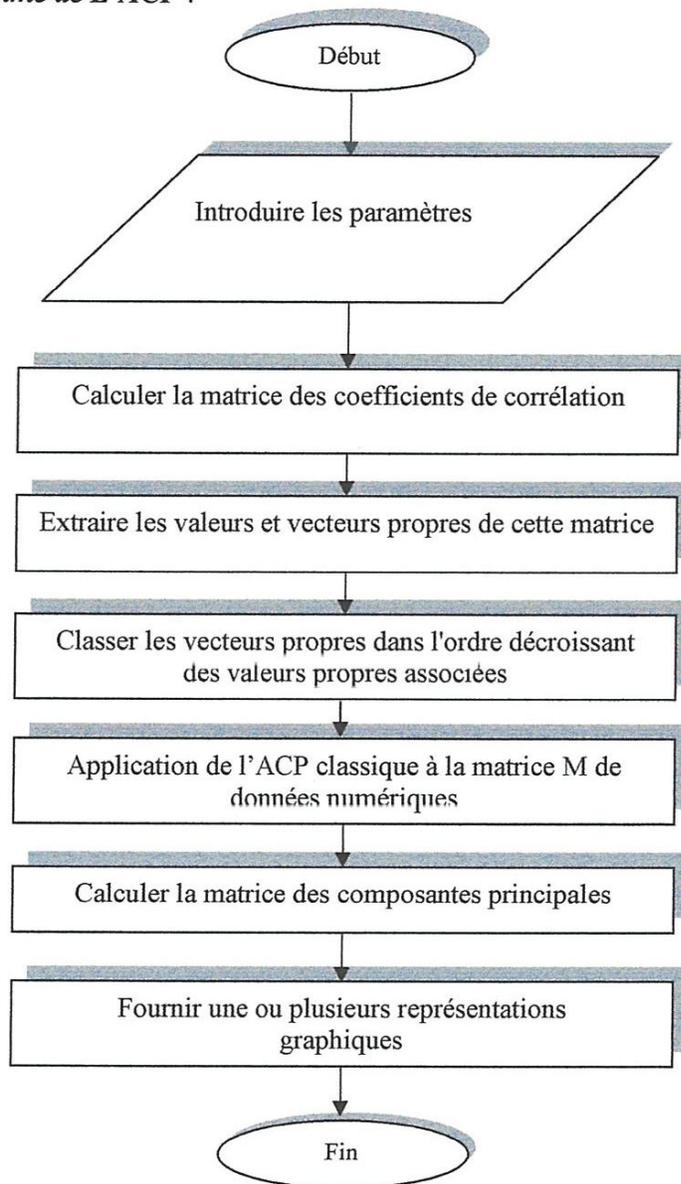


Figure 4.5. : Algorithme de la méthode de l'ACP

5. Mise en œuvre de la base de données :

Les images satellitaires utilisées pour la conception de la base de donnée sont issu de google earth , dont la taille est de $(693*764 \text{ pixels})$, ils sont prise à des distances de la terre (altitude) variant dans l'intervalle $[935 \text{ } 1000] \text{ Km}$.

On décompose chaque image en 12 imageries (partie d'image), de taille $10*10 \text{ pixels}$.

Sachant que l'ensemble des petites images contient des zones d'eau, terre, urbanisme et les arbres, soit 4 classes en totalités.

Sachant que les images utilisé sont des images couleurs (RVB) donc chaque pixel possède 3 attributs (rouge, vert, bleu) qui peuvent prendre de 0 à 255 niveau de couleur (dégradation de couleur).

La base d'image répartie sur 2 catégories : la base d'apprentissage et la base de test.

Base d'apprentissage: contient 24 images 06 échantillons pour chaque classe.

Base de test : contient 24 images 06 échantillons pour chaque classe.

Pour la mise en œuvre de la base d'image nous avons décomposé les images satellitaires en images de taille 10*10 contenant des zones d'eau, terre, urbanisme et les arbres, soit 4 classes en totalités, dont nous choisissons les plus discriminantes.

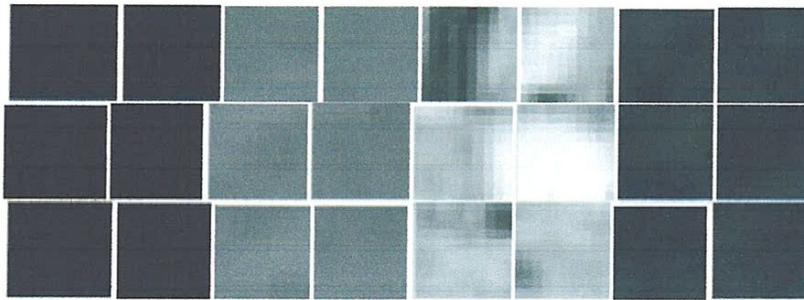


Figure 4.6 : Exemple des images présentes de la base d'apprentissage.

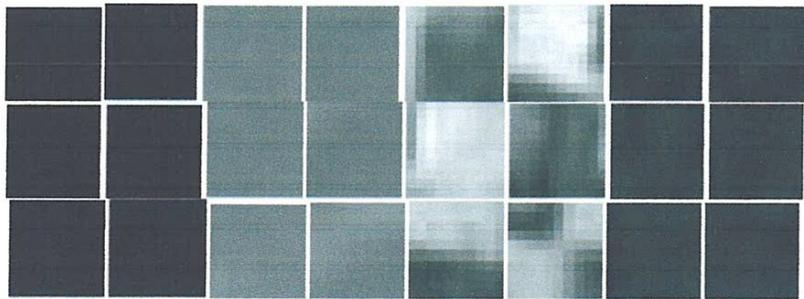


Figure 4.7 : Exemple des images présentes de la base de test.

6. Application de L'ACP :

Pour notre application, nous avons suivi les étapes illustrées dans l'organigramme de la Figure 4.5.

L'ACP comme signalée dans le chapitre 3, cherche à diminuer l'espace de représentation à deux ou trois dimensions sans perdre le maximum d'information. En triant les valeurs propres de la plus grande à la plus petite et même chose pour les vecteurs propres associés (étape 4 de

L'ACP), nous avons obtenu le graphe de distribution des valeurs propres suivant (voir Figure 4.8)

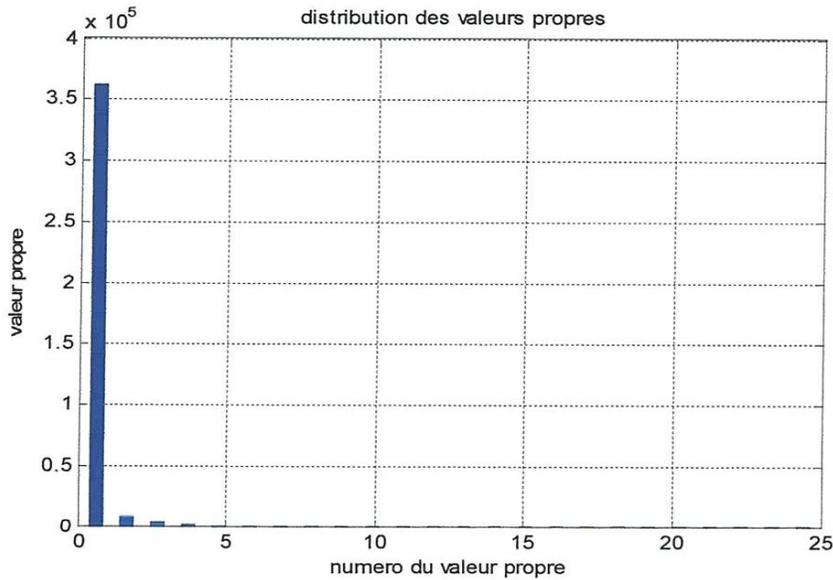


Figure 4.8 : Le graphe de distribution des valeurs propres.

Le calcul des composantes principales permet de donner de nouvelles coordonnées pour chaque classe. Les Figures 4.9 et 4.10 donnent la représentation des quatre classes dans le plan de composantes principales cp1, cp2 et cp1, cp3 pour notre base de données ou nous avons présenté chaque classe par une couleur différent pour :

L'eau par la couleur vert.

- ✓ L'urbanisme par la couleur bleu ciel.
- ✓ La terre par la couleur bleu.
- ✓ Les arbres par la couleur rouge.

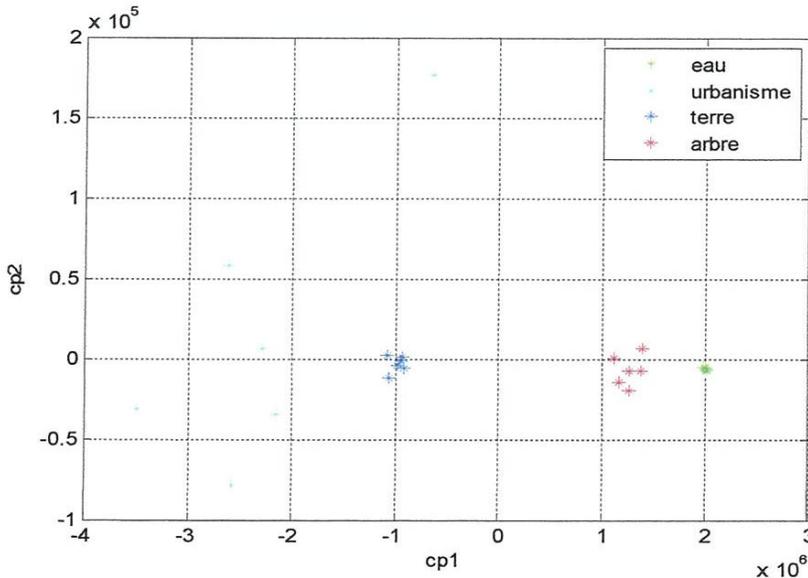


Figure 4.9 : projection de la base suivant cp1 et cp2.

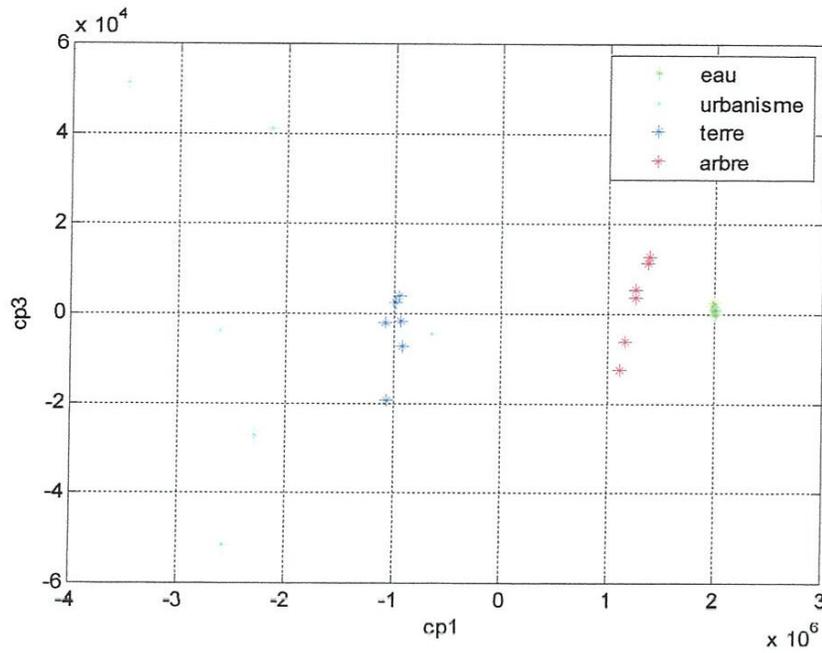


Figure 4.10 : projection de la base suivant $cp1$ et $cp3$.

La figure 4.11 donne la représentation des quatre classes dans le plan de composantes principales $cp1$, $cp2$ et $cp3$ pour notre base de données.

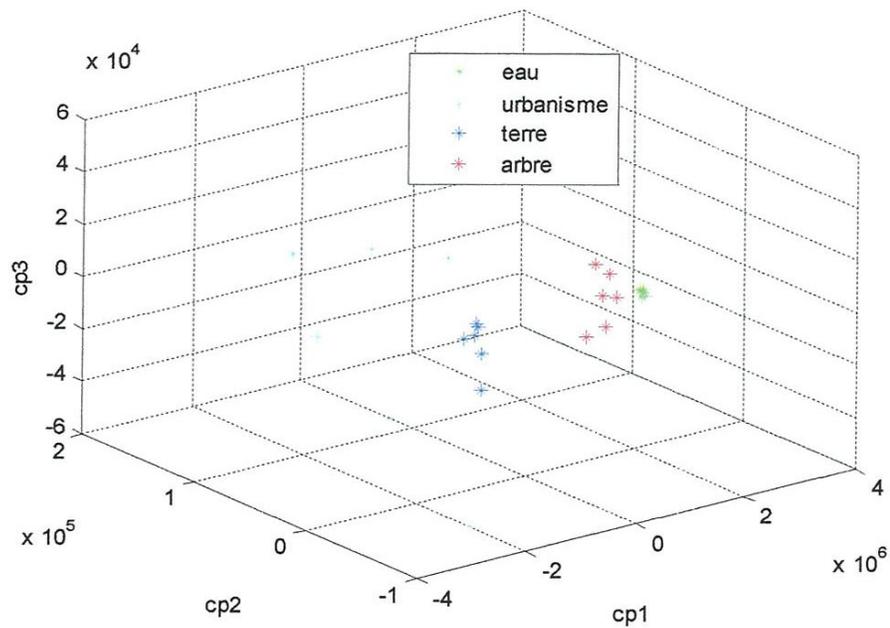


Figure 4.11 : projection de la base suivant $cp1$, $cp2$ et $cp3$.

Suivant cette représentation, nous voyons que chaque classe représentée et séparée de l'autre, il y a deux classes qui son proche (eau et arbre) et aussi (urbain et terre) à cause peut

être de la ressemblance de ces classes entre eux (qui ont des paramètres proches les uns des autres).

6.1. Reconnaissance :

Pour la classification d'une classe à reconnaître, nous projetons ces composantes sur cette représentation ou il peut appartenir à une classe.

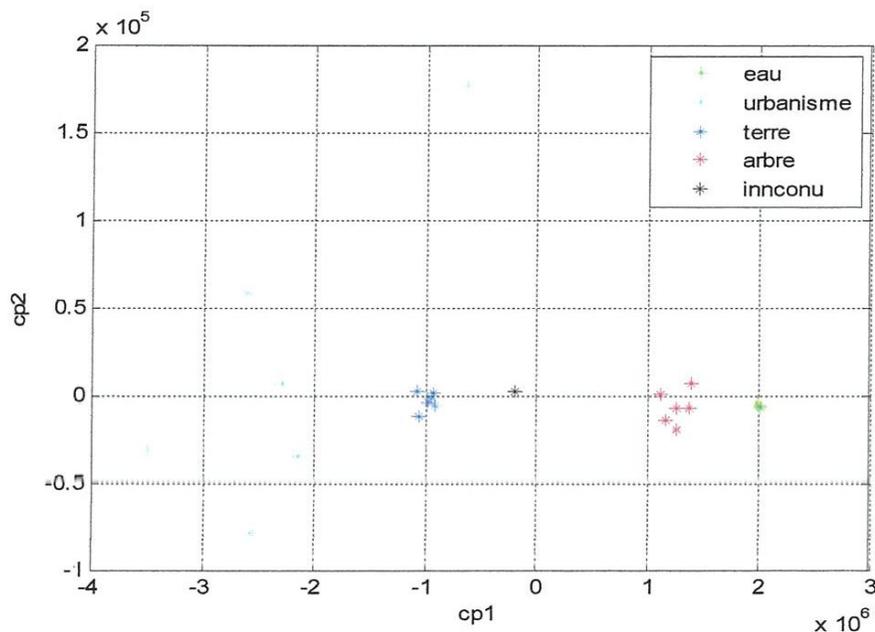


Figure 4.12 : représentation de test de reconnaître sur cp1 et cp2 .

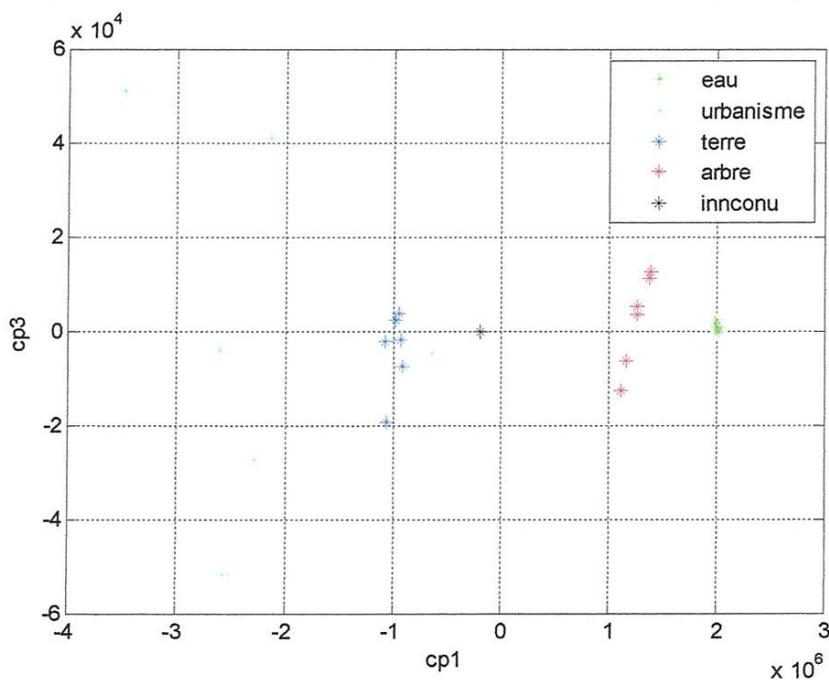


Figure 4.13: représentation de test de reconnaître sur cp1 et cp3.

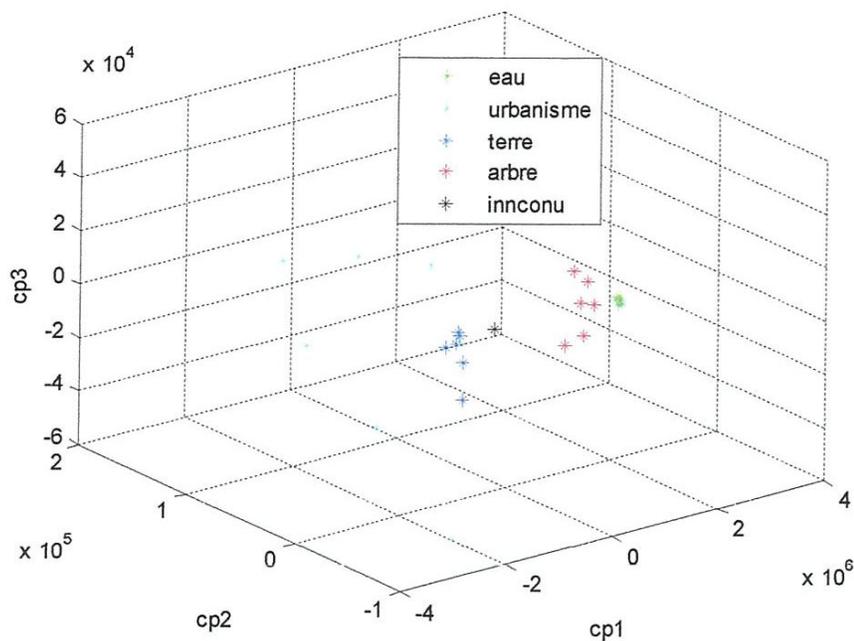


Figure 4.14 : représentation de test de reconnaissance sur cp1, cp2 et cp3.

En fin, pour décider à quelle classe appartient, nous calculons la distance euclidienne du vecteur projeté avec ceux des différentes classes. La valeur la plus petite permet de définir la classe d'appartenance et affiche la classe reconnue.

En revenant à l'exemple représenté sur la figure 4.12, 4.13 et 4.14, la classe a été reconnue parfaitement.

Cette méthode a conduit au très bon résultat résumé dans le tableau 4.1.

7. Test & évaluation :

Pour tester la fiabilité de notre système, nous avons effectué des tests sur la base de projection elle-même, mais avec des prises de vues différentes, cela a conduit à des résultats très satisfaisants où le taux d'apprentissage a atteint les 100%. Ensuite, nous avons effectué les tests de reconnaissance sur la base de tests. Les résultats sont illustrés aussi dans le tableau, où nous constatons que le système de reconnaissance que nous venons de mettre en œuvre est acceptable et donne des résultats satisfaisants 95.8333%

	Taux de Rec. (en%)	Taux d'erreur (en%)
Base d'apprentissage	100%	0%
Base de test	95.8333%	4.1667%

Tableau 4.1 : résultats obtenue par l'ACP

D'après le tableau précédant on trouve que le taux de reconnaissance est très élevés parce que notre base de données n'est pas vaste et l'efficacité de notre classifieur (ACP+distance euclidienne).

8. Conclusion :

L'objectif de notre travail est la conception d'un système capable de reconnaître des images satellitaire.

Nous avons implémentés pour cela l'approche ACP en utilisant une règle de décision. Les performances des algorithmes sont testées sur une base de test composée de 4 classes. Pour Chacune nous avons pris 6 imagerie. Pour la décision qui permet de déterminer aisément le degré d'appartenance d'une forme à une classe en utilisant le minimum de la distance euclidienne.

Conclusion générale

Dans ce mémoire nous présentons une application de l'analyse en composantes principales. Nous avons appliqué cette méthode sur une base de données, de quatre classes pour chacune nous avons pris 12 imagerie (6 pour l'apprentissage et 6 pour le test).

Dans la phase d'apprentissage nous avons utilisé une base de données contient 6 imagerie de chaque classe, on mémorise les informations caractérisant chaque image.

Pour la phase de reconnaissance nous avons utilisé une base de test contient 6 imagerie de chaque classe.

Pour l'identification des images nous avons utilisé la distance euclidienne, ou on cherche la distance minimale d'image projeté avec toute la base d'apprentissage.

Finalement nous estimons que l'étude que nous avons présentée à une grande importance dans le domaine de la classification des images satellitaires.

Ce travail reste ouvert et peut avoir plusieurs perspectives :

- Améliorer la base d'apprentissage (taille et variabilité), facteur essentiel de l'apprentissage.
- Augmenter le nombre de classes pour pouvoir localiser tous types de zones sur les images satellitaires analysées.
- Généraliser l'application sur des images de grandes tailles et en temps réel.
- ...etc

Bibliographie :

[Ga83] A.GAGALOWICZ

« Vers un modèle de texture » Thèse de doctorat d'état et sciences
Mathématiques, Université de Paris VI. 1983.

[GM93] Goesta Granlund Michel Kocher

« Traitement numérique des images » Presses polytechnique et universitaire
romandes et GNET-ENST.1993.

[Ro98] RRouhana

« Détection des linéaments dans les images de radarSAT par les réseaux de
neurone cellulaire » école de technologie supérieur université du QUEBAC 1998.

[CB06] S. Chelhi & E. Bouguern

« Discrimination de texture par perceptron multicoches »
Mémoire d'ingénieur en informatique, Université de Guelma, 2006.

[Mo00] M.Mokhtari

« Matlab 5.2 & 5.3 et simulink 2 & 3 » Valeo Thermique Habitable 8,2000.

[DZ06] S. Brissi, O. Zedadra

« Application d'une méthode Hybride NEUAC Pour la reconnaissance des
vissages ». mémoire d'ingénieur université de Guelma 2006.

[Bo04] A/D. Boualleg

« Reconnaissance automatique des vissages ». mémoire de majister université
de Guelma 2004.

[BS06] N.Béchairai & H. Seïmioul

« Classification supervisée par les réseaux de neurone sartificiels ». mémoire
d'ingénieur en électronique université de Guelma 2006.

[BG05] W, Belguidoum & B.Guerzize

« Système de détection de visages par méthode hybride, détection de
peau et Template Matching » Mémoire d'ingénieur en informatique,
Université de Guelma, 2005.

[1] <http://www.l-space> - imagerie satellitaire et observation du sol.htm

[2] <http://www.kaddour.com/chapl/chapl.htm> #défmition%20images

[3] http://fr.wikipedia.Org/wiki/analyse_spectrale#signaux