



République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la
recherche scientifique

Université Larbi Tébessi – Tébessa

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département : Mathématiques et Informatique



Mémoire de fin d'études
Pour l'obtention du diplôme de MASTER 2

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Systèmes d'information

Thème

**Multi-spectral Satellitaires Image Segmentation by using
FCM method Combined with AT**

Présenté Par :
ATMANIA RIADH

Devant le jury :

<i>Dr BENOIR Akram</i>	<i>MCA Université Larbi Tébessi</i>	<i>Président</i>
<i>Dr ZAGARI Ahmed</i>	<i>MCB Université Larbi Tébessi</i>	<i>Examineur</i>
<i>Mr NOUIOUA Tarek</i>	<i>MAA Université Larbi Tébessi</i>	<i>Encadreur</i>

Date de soutenance : 15/09/2020

Je dédie ce travail

A toutes ma grande famille

Par souci de confiance et de patience, je leur présente cet humble travail qui, j'espère de tout mon cœur, sera le fruit d'un travail et d'une diligence et sera à la hauteur de leurs aspirations.

En témoignage de mes sincères reconnaissances pour les efforts qu'ils ont consentis pour me soutenir au cours de mes études. Que dieu nous garde toujours unis.

Pour le souvenir du décès de mon père et ma mère et ma sœur Fatima Zahra, ceux qui j'aime profondément

À MA PETITE FAMILLE MA FEMME, MON FILS ABDERRAOUF

ET MA FILLE SARA

À Toutes les personnes qui m'ont aidée à réaliser mon projet

Atmania Riadh

Remerciements

C'est avec un grand plaisir que je réserve ces quelques lignes en signe de gratitude et de profonde reconnaissance à tous ceux qui de près ou de loin, ont contribué à la réalisation et l'aboutissement de ce travail.

*Tout d'abord, je tiens à remercier **Allah** tout puissant, de m'a permis de mener à bien ce mémoire, et de m'a orienté au chemin du savoir.*

*Ensuite, je remercie sincèrement **Dr. NOUIOUA Tarek**, MCB au sein du département mathématiques et informatique de la faculté des sciences exactes et sciences de la nature et de la vie à l'université de Tébessa, pour son encadrement, son assistance, son soutien, sa disponibilité et ses précieux conseils.*

*Mes vifs remerciements s'adressent à **Dr. BENOUR Akram**, MCA, et **Dr. ZAGARI Ahmed**, MCA, au sein du département mathématiques et informatique de la faculté des sciences exactes et sciences de la nature et de la vie à l'université de Tébessa, pour l'honneur qu'ils m'ont accordé en acceptant de juger mon travail.*

Enfin, j'adresse mes chaleureux remerciements à mes enseignants pour la qualité de l'enseignement qu'ils ont bien voulu me prodiguer durant mes études afin de me fournir une formation efficiente.

Merci à toutes et tous.

Résumé

Ce mémoire présente beaucoup plus une renaissance de travaux faits en littérature mais son implémentation n'est disponible en aucune forme, ni en forme exécutable ni en forme libre ou à acheter. Il s'agit d'implémenter une méthode qui sera une combinaison entre deux algorithmes, qui sont bien connus dans plusieurs domaines d'imagerie, que ce soit industriel, médical ou encore satellitaire chose que nous devons être vigilant lors de l'implémentation de cette combinaison à cause de la taille des images satellitaires et encore de la consommation en temps de calcul qui augmente avec la taille des images et la méthodologie avec laquelle on applique la méthode. Cette méthode est une segmentation consistant à être utilisée dans un domaine de fouille de donnée avec des images multi-spectrales c'est une méthode combinant la méthode connue floue c-moyennes avec la méthode du seuillage adaptatif pour but de fournir un système d'aide à la décision surtout quand elle soit automatique.

Mots clés : floue c-moyennes, seuillage adaptatif, fouille des données, système d'aide à la décision spatiale, traitement d'image satellitaire

Abstract

This thesis presents much more a revival of work done in literature, but its implementation is not available in any form, neither in executable form nor in free form or for purchase. It is a question of implementing a method which will be a combination between two algorithms, which are well known in several fields of imagery, whether industrial, medical or satellite. We must be careful when implementing this combination because of the size of the satellite images and the consumption of computing time which increases with the size of the images and the methodology with which the method is applied. This method is a segmentation consisting in being used in a data mining domain with multi-spectral images.

It is a method combining the known fuzzy c-means method with the adaptive thresholding method in order to provide a decision support system especially when it is automatic.

Keywords: fuzzy c-means, adaptive thresholding, data mining, spatial decision support system, satellite image processing.

ملخص

تقدم هذه الأطروحة أكثر بكثير من إحياء العمل المنجز في الجانب العلمي النظري و لكن تنفيذها غير متاح بأي شكل ، سواء في شكل قابل للتنفيذ أو في شكل مجاني أو للشراء. إنها مسألة تنفيذ طريقة تكون عبارة عن مزيج بين خوارزميتين معروفتين في العديد من مجالات التصوير ، سواء كانت صناعية أو طبية أو ساتيليتية. يجب توخي الحذر عند تنفيذ هذه المجموعة بسبب حجم صور الأقمار الصناعية واستهلاك وقت الحوسبة الذي يزداد مع حجم الصور والمنهجية التي يتم بها تطبيق الطريقة. هذه الطريقة عبارة عن تجزئة تتكون من استخدامها في مجال التنقيب عن البيانات مع صور متعددة الأطياف. إنها طريقة تجمع بين طريقة المناطق المتوسطة الضبابية المعروفة مع طريقة العتبة التكيفية من أجل توفير نظام دعم القرار خاصة عندما يكون تلقائيًا .

الكلمات المفتاحية : المناطق المتوسطة الضبابية، العتبة التكيفية، البحث عن البيانات، نظام دعم القرار المكاني، معالجة صور الأقمار الصناعية.

Liste des Abréviations

- **SIG** : Système d'Information Géographique
- **GOES** : Geostationary Operational Environmental Satellite
- **SPOT** : Satellite Pour l'Observation de la Terre
- **ECD** : L'Extraction de Connaissances à partir des Données
- **BD** : Base des données
- **FD** : Fouille des données
- **KDD** : Knowledge Discovery in Database
- **IA** : Intelligence Artificielle
- **FDS** : La fouille de données spatiales
- **DMS** : Le Data Mining Spatial
- **FCM** : Fuzzy C-Means
- **AT** : Adaptive Thresholding
- **SA** : Seuillage Adaptatif
- **UML** : Unified Modeling Language
- **ASAL** : L'Agence Spatiale Algérienne

Liste de Figures

Figure 1.1. Exemple d'une image 2 D	4
Figure 1.2. groupe de pixel formant la lettre A	5
Figure 1.3. Representation d'une image 3D	9
Figure 1.4. Exemple d'une Image vectorielle	10
Figure 1.5. Représentation de voisinage	12
Figure 1.6. Fond et Forme de connexité	14
Figure 1.7. Exemples de champ d'application de traitement image.....	19
Figure 1.8. Satellite à défilement	21
Figure 1.9. Satellites géostationnaire	21
Figure 2.1. Méthodes d'apprentissage	26
Figure 2.2. Schéma générale d'un Classifieur	28
Figure 2.3. Graphe de voisinage et matrice de voisinage	30
Figure 3.1. Le modèle décisionnel proposé	40
Figure 3.2. Outils d'Investigations	43
Figure 3.3. Système Décisionnel FCM-AT Proposé	44
Figure 3.4. Diagramme UML de Cas d'utilisation	46

Figure 3.5. Diagramme de Séquences	47
Figure 3.6. Exemple d'ensemble flou avec $x = 3$	48
Figure 3.7. Trame hexagonal	51
Figure 3.8. Matrice M1.....	52
Figure 3.9. Matrice M2.....	52
Figure 4.1. élimination de la zone d'étude	55
Figure 4.2. Les trois canaux TM 1, 3, 4	55
Figure 4.3.a. FCM paramétrique $k=4$ et U_{ik} équitable (Oran)	57
Figure 4.3.b. FCM paramétrique $k=4$ et U_{ik} non équitable (Oran).....	57
Figure 4.4.a. FCM paramétrique $k=6$ et U_{ik} équitable (Oran)	57
Figure 4.4.b. FCM paramétrique $k=6$ et U_{ik} non équitable (Oran)	57
Figure 4.5.a. FCM paramétrique $k=10$ et U_{ik} équitable (Oran)	58
Figure 4.5.b. FCM paramétrique $k=10$ et U_{ik} non équitable (Oran)	58
Figure 4.6. FCM automatique (Oran)	59
Figure 4.7.a. FCM paramétrique $k=3$ et U_{ik} équitable (Sebkha)	60
Figure 4.7.b. FCM paramétrique $k=3$ et U_{ik} non équitable (Sebkha)	60
Figure 4.8.a. FCM paramétrique $k=5$ et U_{ik} équitable (Sebkha)	60
Figure 4.8.b. FCM paramétrique $k=5$ et U_{ik} non équitable (Sebkha)	60
Figure 4.9. FCM automatique (Sebkha)	61
Figure 4.10. Le seuillage adaptatif manuel avec masque central 1 et seuil = 50 et 10 et 3 (de gauche à droite) (Partie nord d'Oran)	61
Figure 4.11. Le seuillage adaptatif manuel avec masque central 0 et seuil = 50 et 10 et 3 (de gauche à droite) (Partie nord d'Oran)	62
Figure 4.12. Seuillage adaptatif automatique (Partie nord d'Oran)	63
Figure 4.13. Le seuillage adaptatif manuel avec masque central 1 et seuil = 40 et 20 et 3 (de gauche à droite) (Partie Sebkha)	63
Figure 4.14. Le seuillage adaptatif manuel avec masque central 0 et seuil = 40 et 20 et 3 (de gauche à droite) (Partie Sebkha)	64
Figure 4.15. Le Seuillage adaptatif automatique (Partie Oran Sebkha)	64
Figure 4.16.a. Fusion de FCM paramétrique et seuillage automatique (Oran)	65
Figure 4.16.b. Fusion de FCM paramétrique et seuillage manuel (Oran)	65
Figure 4.17.a. Fusion de FCM automatique et seuillage automatique (Oran)	66
Figure 4.17.b. Fusion de FCM automatique et seuillage manuel (Oran)	66

Figure 4.18.a. Fusion de FCM paramétrique et seuillage automatique (Sebkha)**66**

Figure 4.18.b. Fusion de FCM paramétrique et seuillage manuel (Sebkha)**66**

Figure 4.19.a. Fusion de FCM automatique et seuillage automatique (Sebkha)**67**

Figure 4.19.b. Fusion de FCM automatique et seuillage manuel (Sebkha)**67**

Table des Matières

Dédicaces	i
Remerciements	ii
Résumé	iii
Liste des Abréviation.....	vi
Liste des figures	vii
Table des matières.....	x
Introduction Générale.....	1
Chapitre 1 : Généralité sur l'imagerie... ..	3
1. Introduction	3
2. Définition de l'image	4
➤ Définition d'une image numérique	4
3. Caractéristiques d'une image numérique	5
3.1. LE PIXEL	5
3.2. LA DIMENSION	5
3.3. LE BRUIT	6
3.4. L'HISTOGRAMME	6
3.5. LES CONTOURS ET TEXTURES	6
3.6. Brillance, Variance et Contraste	6
4. Différents types d'images	7
4.1. Images Noir et blanc	7
4.2. Images couleur	8
5. La Qualité de l'image numérique	9
6. Images bitmap et images vectorielles	10
6.1. Les images bitmap	10
6.2. Les images vectorielles	10
7. Représentation d'une image	11

8. Différente perception d'une image	11
9. Connexité dans les images	11
9.1. Voisinage d'un pixel	12
9.2. Connexité	12
9.3. Chainage de pixel	12
9.3.1. Région connexe composante connexe	13
9.3.2. Fond, Forme de connexité	14
9.4. Etiquetage de composantes connexes	14
10. Distances entre pixels	15
11. Les systèmes de traitement d'images	15
11.1. Acquisition des données images	15
11.2. Dispositifs de numérisations d'images	16
11.3. Prétraitement Et Post-Traitement	16
11.4. Notions du filtrage	16
11.4.1. Définition	16
11.4.2. Types de filtres	17
11.5. Notions de lissage	18
11.6. Traitement Numérique Des Images	18
11.7. Visualisation	18
11.8. Domaines d'application traitement d'images	19
12. Définition d'un satellite	20
13. Différent type de satellite	20
13.1. Satellites à défilement	20
13.2. Satellites géostationnaires	20
14. Définitions de SIG (Systèmes d'information Géographique)	21
15. Segmentation et classification des images Multi spectrales	22
16. Conclusion	23
Chapitre 2 : La Fouille de Données Spatiales	24
1. Introduction	24
2. Méthodes d'apprentissage	25
2.1. Les systèmes d'apprentissage automatique supervisé	26
2.2. Les systèmes d'apprentissage automatique non supervisé	27
3. Fouille de données	27

➤ Vocabulaires	27
4. La fouille de données (data mining) spatiales	28
4.1. Définition	29
4.2. Spécificités de la fouille de données spatiales	29
4.3. Caractéristiques des données spatiales	29
4.4. Matrice et graphe de voisinage	30
4.5. Travaux connexes	30
4.6. Panorama des méthodes de fouille de données spatiales	31
4.6.1. Phase exploratoire	31
4.6.2. Phase décisionnelle	33
5. Conclusion	35
Chapitre 3 : La Classification des images satellitaires et la méthode FCM-AT	36
1. Introduction	36
2. Classification et Analyse d'images satellitaires	37
➤ Types de classification et ensemble flou	37
3. Segmentation d'image	38
3.1. Définition de la segmentation	38
3.2. Le choix d'une technique de segmentation	39
4. Présentation globale du cadre d'application	39
5. Le Système de classification proposé	39
5.1. La phase de structuration du modèle	41
5.2. La phase d'exploitation du modèle	41
5.3. Les Outils d'investigation pour l'élaboration de FCM-AT	42
5.4. Le système FCM-AT	44
5.5. Diagramme de cas d'utilisation du système proposé	45
5.6. Diagramme de Séquences	47
6. K-moyennes Floues (Fuzzy C-Means) FCM	48
7. Seuillage Adaptative (Adaptive Thresholding)	51
8. Conclusion	53
Chapitre 4 : Mise en Œuvre	54
1. Introduction	54
2. Description de la zone d'étude	54
3. Exploitation des données	55
4. Les Résultats	56

4.1 Chargement des images	56
4.2. Colorisation et raffinement	56
4.3. Segmentation par le seuillage adaptatif	61
4.4. Fusion (FCM et Seuillage Adaptatif)	65
5. Conclusion	67
Conclusion Générale	68
Références Bibliographiques	70

Introduction Générale

Grâce à la télédétection, la surface de la terre est maintenant regroupée sous une forme où les distances sont avec une prodigieuse capacité, on peut passer du détail à la globalité d'une manière très simple avec un clic, en manipulant l'échelle temps ou distances. Toutefois, les images satellitaires dans leurs états bruts ne reflètent pas la réalité du terrain, cela est dû à plusieurs facteurs liés à la position des objets, aux effets du reliefs, et aux transitions entre régions.

L'aide à la décision avec l'évolution des nouvelles technologies, permet de surpasser ce problème surtout lorsqu'il s'agit de l'observation de la terre pour différentes études. Par ailleurs, les décideurs sont confrontés à un problème spatial complexe où sa résolution est conflictuels avec plusieurs objectifs. Une variété de techniques d'analyse a été développée pour aider les décideurs à résoudre des problèmes avec des critères multiples.

Les systèmes d'aides à la décision spatiale aide à résoudre ces problèmes spatiaux complexes. Les systèmes classiques ne prennent pas adéquatement en compte la prise de décision, parce qu'ils manquent de capacités de modélisations analytiques. Avec les nouvelles technologies, on peut disposer d'une image satellitaire bien détaillée. Les systèmes d'information géographique (SIG) sont largement utilisés pour l'aide à la décision spatiale.

Dans ce présent mémoire, nous utiliserons une image satellitaire de la région d'Oran, image qui est disponible pour nous, es cela ne signifie pas que l'implémentation se limite à cette région seulement. La procédure entreprise prend en charge toute la chaine de traitement d'image satellitaire, la démarche suivie permet l'extraction de régions homogènes

(les classes : mer , sol, forêts, urbain....) à partir d'une image Landsat 5 TM en exploitant la méthode de C-mayennes Floues (CMF) en combinaison avec le seuillage adaptatif (AT).

L'objectif de ce mémoire est d'implémenter cette fusion des deux méthodes pour concevoir un système de classification des images multi-spectrales satellitaires.

Ce mémoire est organisé comme suit :

Chapitre 1: une théorie consistant à présenter de généralité sur l'imagerie

Chapitre 2: présente La fouille de données spatiales

Chapitre 3: présente La classification des images satellitaires et la méthode FCM-AT

Chapitre 4: une implémentation de la fusion des deux méthodes : La mise en œuvre

Chapitre 1

Généralité sur l'imagerie

1. Introduction

Avant que les satellites soient lancés dans l'espace, l'homme n'a jamais eu l'occasion de pouvoir avoir une vision globale du monde qu'il habite. La mise en orbite du premier satellite lui a ouvert les voies vers une découverte du ciel et par là de voir la planète où il habite de façon plus claire et globale. Dès qu'ils sont mis en service, et que la technologie a évolué les satellites d'observation sont devenus indispensables parce qu'ils offrent une multitude de raisons d'observer la Terre depuis l'espace.

Les satellites de plus en plus performants et équipés de systèmes technologiques de visions de hautes performances sont capables de couvrir de manière systématique et répétitive des portions de territoires immenses en produisant des images très détaillées.

Les satellites sont tout à fait adaptés à l'étude de l'environnement terrestre. Leurs avantages sont appréciables : rapidité de déplacement, discrétion, étendue du champ de vision, survol de régions d'accès difficile au sol (zones polaires, désertiques...).

Des satellites d'observation de la Terre utilisent différents orbites pour pallier au problème de résolution. Ces images que nous envoient ces satellites sont une source capitale d'informations objectives, globales et actualisées. Elles permettent de collecter et mettre à jour les informations sur les ressources de la Terre et de surveiller entre autre, l'impact des activités humaines.

L'observation de la Terre par ces photos constitue donc un moyen privilégié d'obtenir des informations géographiques, indispensables dans de nombreuses applications.

Nombreux systèmes d'informations notamment Géographique (SIG) permettent d'intégrer des données tridimensionnelles pour en réaliser des visualisations impressionnantes de réalisme, permettre de s'y promener ou de survoler les différentes zones modélisées.

2. Définition de l'image

L'image est une représentation d'une scène ou d'un objet par la peinture, la sculpture, le dessin, la photographie, le film, etc. C'est aussi un ensemble structuré d'informations qui, après affichage sur l'écran, a une signification pour l'œil humain [1].

Elle peut être décrite sous la forme d'une fonction $I(x, y)$ de brillance analogique continue, définie dans un domaine borné, tel que x et y sont les coordonnées spatiales d'un point de l'image et I est une fonction d'intensité lumineuse. Sous cet aspect, l'image est inexploitable par la machine, ce qui nécessite sa numérisation.

➤ Définition d'une image numérique

Une image numérique est constituée de Pixels où ses informations sont codées sous forme chiffrée. Un pixel est caractérisé par un rapport des couleurs : Rouge, Vert, Bleu (RVB).

Par exemple, si pour chaque couleur on a 256 niveaux (2 puissance 8 niveaux (bit)), le point constitué de trois couleurs aura $256 \times 256 \times 256 = 16\,777\,216$ couleurs possibles. On dit que la couleur est codée sur 24 bits (3 fois 8 bits) ou à une profondeur de couleur de 24 bits [1].



Fig.1.1: exemple d'une image 2 D

3. Caractéristiques d'une image numérique

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants:

3.1. LE PIXEL

Contraction de l'expression anglaise " Picture éléments ": éléments d'image, le pixel est 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 d'affichage ou d'impression. La lettre A, par exemple, peut être affichée comme un groupe de pixels dans la figure 1.2 :

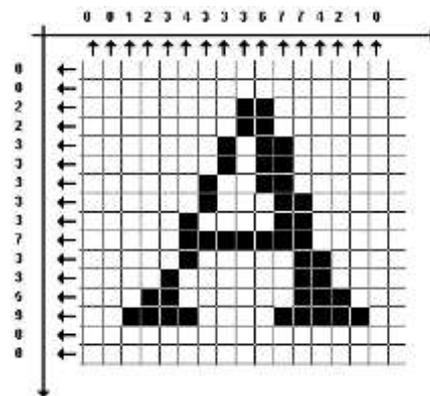


Fig.1.2 : groupe de pixel formant la lettre A

La quantité d'information que véhicule chaque pixel donne des nuances entre images monochromes et images couleurs. Dans le cas d'une image monochrome, chaque pixel est codé sur un octet et, la taille mémoire nécessaire pour afficher une telle image est directement liée à la taille de l'image.

Dans une image couleur (R.V. B.), un pixel peut être représenté sur trois octets : un octet pour chacune des couleurs : rouge (R), vert (V) et bleu (B) [2].

3.2. LA DIMENSION

C'est la taille de l'image. Cette dernière 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 dans une image [3].

3.3. LEBRUIT

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 [4].

3.4. L'HISTOGRAMME

L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image.

Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents ou, encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme correspondant [4-5].

Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans le cas d'une image trop claire ou d'une image trop foncée.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci.

3.5. LES CONTOURS ET TEXTURES

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 [6].

Les textures décrivent la structure de ceux-ci. L'extraction de contour consiste à identifier dans l'image les points qui séparent deux textures différentes [5].

3.6. Brillance, Variance et Contraste

On définit la brillance B, qui est l'intensité moyenne des niveaux de gris appartenant à tous les pixels de la région image. Ce paramètre représente l'emplacement de l'histogramme sur l'échelle des niveaux de gris.

$$B = \text{MOY} = \frac{1}{N} \sum_{ij} g(i, j) \dots \dots \dots (1)$$

On définit la variance VAR, elle mesure la répartition des niveaux de gris autour de la valeur moyenne comme suit :

$$\text{VAR} = \frac{1}{N} \sum_{ij} (g(i, j) - B)^2 \dots \dots \dots (2)$$

On définit le contraste C, comme suit :

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.[7]

$$C = \sqrt{\text{VAR}} = \sqrt{\frac{1}{N} (g(i,j) - B)^2} \dots \dots \dots (3)$$

Une autre formule pour définir le contraste :

$$C = \frac{L_{\text{fond}} - L_{\text{test}}}{L_{\text{fond}}} \dots \dots \dots (4)$$

Cette formule dite de Weber, de plus :

$$L_{\text{moy}} = L_{\text{max}} + L_{\text{min}} \dots \dots \dots (5)$$

Tel que L : la luminance, cette grandeur est mesurée en candéla/m².

Le contraste est le plus souvent exprimé en pourcentage. 90% étant un contraste très élevé, 10% étant un contraste très faible. Mais, pour la plupart des auteurs étudiant la Fonction de Sensibilité au Contraste, le contraste de luminance est la différence entre le maximum de luminance des bandes claires (L_{max} des bandes claires) et le minimum de luminance des bandes sombres (L_{min} des bandes sombres), rapportée à la somme de ces deux luminances. Le contraste de modulation (M) ou contraste de Michelson est défini par :

$$C = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}} \dots \dots \dots (6)$$

L'inverse du contraste (C) est la sensibilité (S) :

$$S = \frac{1}{C} \dots \dots \dots (7)$$

4. Différents types d'images

4.1. Images Noir et blanc

Ces images sont dites à niveaux de gris, car on ne prend pas en compte ici la couleur mais seulement l'intensité lumineuse (l'exemple classique correspond aux photographies noir et blanc).

Parmi ces images on peut trouver:

- **Images binaires** : Où chaque pixel est représenté par un bit (0 ou 1) avec en général (0 : noir, intensité nulle ; 1 : blanc, intensité maximale).

Notons que la plupart des systèmes de traitement d'images placent chaque pixel dans un octet (code 0 ou 255 (pour coder le 1 de l'image binaire)) pour des facilités d'accès et d'écriture des algorithmes.

- **Images en niveaux de gris** : Dans ce cas on dispose d'une échelle de teintes de gris, et la plupart du temps on dispose de 256 niveaux de gris avec:

0 → noir,127 → gris moyen,, 255 → blanc ceci est commode car l'unité d'information est l'octet (unsigned char en langage C).

Certaines images peuvent être codées sur deux octets ou plus (certaines images médicales, des images astronomiques,...) ce qui peut poser des problèmes dans la mesure où les systèmes de traitement d'images courants supposent utiliser des pixels d'un octet !

Notons au passage que l'humain standard ne reconnaît au plus que 64 niveaux de gris .

4.2. Images couleur

Ces images sont en général codées en utilisant le codage des trois couleurs fondamentales (rouge,vert,bleu), on parle alors d'images RVB.

(Cela correspond au codage des téléviseurs couleur)

Chaque couleur est codée sous forme d'un octet, d'où Composante Rouge, intensité de 0 à 255 image RVB + composante Verte, intensité de 0 à 255 + composante Bleue, intensité de 0 à 255 on code ainsi :

$2^{24} = 16\ 777\ 216$ couleurs différentes.

La couleur peut aussi être codée sur un octet (256 couleurs) et l'affichage étant réalisé après passage dans une table de couleurs (transcodage).

Le codage peut ainsi être réalisé en affectant 3 bits au rouge et au vert et, *2 bits au bleu (pour tenir compte par exemple de la plus faible sensibilité de la vision humaine au bleu)*.

Il existe d'autres techniques de représentation de la couleur pour les images (on passe d'un espace 3D, l'espace RVB, à un autre espace 3D défini par une autre base).

Ainsi le standard T.V. US **NTSC** utilise un autre codage appelé **YIQ**.

- **Images 3D** :

Des images 3D sont des images qui représentent une scène en trois dimensions. Le « pixel » est alors appelé un **Voxel**, et représente un volume élémentaire.

Ces images peuvent être évidemment, d'un des deux types définis précédemment (N/B ou couleur).

Des exemples d'images de ce type se rencontrent dans les images médicales. Les images tomographiques axiales sont ainsi des images construites à partir de plusieurs radiographies faites sous des angles de vue différents

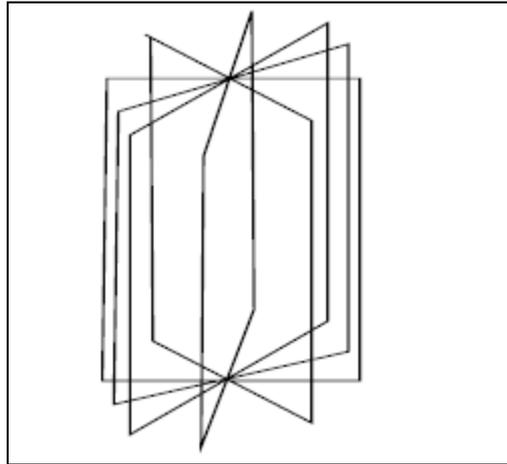


Fig.1.3:Representation d'une image 3D

Autres exemples : les images scannées, les images de résonance magnétique...

Toutes ces images pouvant être éventuellement « fusionnées » pour former des images plus complexes (les rayons X fournissant l'ossature, le scanner les tissus, et la RMN décrivant les fonctions physiologiques par exemple).

Cas particuliers :

Les images obtenues par un laser tournant pour obtenir une partie de l'information de forme des objets d'une scène.

❖ Images stéréographiques :

On dispose alors d'une paire d'images (N/B ou couleur) prises sous des points de vue différents. A partir de telles images, il est possible d'obtenir de l'information sur la scène 3D [8].

5. La Qualité de l'image numérique

Elle dépend, d'une part, de la qualité des images d'origine et, d'autre part, des moyens mis en œuvre pour convertir un signal analogique en signal numérique. Elle dépend aussi de :

- ✓ La qualité des périphériques de numérisation de l'image, du nombre de niveaux de gris ou de couleurs enregistrées, etc.

- ✓ La qualité de l'affichage à l'écran : définition de l'écran, nombre de teintes disponibles simultanément, calibrage de l'écran, etc.

Les critères d'appréciation de la qualité d'une image, tels que cités succinctement ci-dessus, dépendent largement de la structure même de l'image réaliste ou conceptuelle et de son mode de représentation (bitmap ou vectorielle) [9].

6. Images bitmap et images vectorielles

On distingue généralement deux grandes catégories d'images :

6.1. Les images bitmap

(Appelées aussi images raster) : il s'agit d'images pixellisées, c'est-à-dire un ensemble de points (pixels) contenus dans un tableau, chacun de ces points possédant une ou plusieurs valeurs décrivant sa couleur.

6.2. Les images vectorielles

Les images vectorielles sont des représentations d'entités géométriques telles qu'un cercle, un rectangle ou un segment. Celles-ci sont représentées par des formules mathématiques (un rectangle est défini par deux points, un cercle par un centre et un rayon, une courbe par plusieurs points et une équation). C'est le processeur de l'ordinateur qui sera chargé de «traduire» ces formes en informations interprétables par la carte graphique [10].

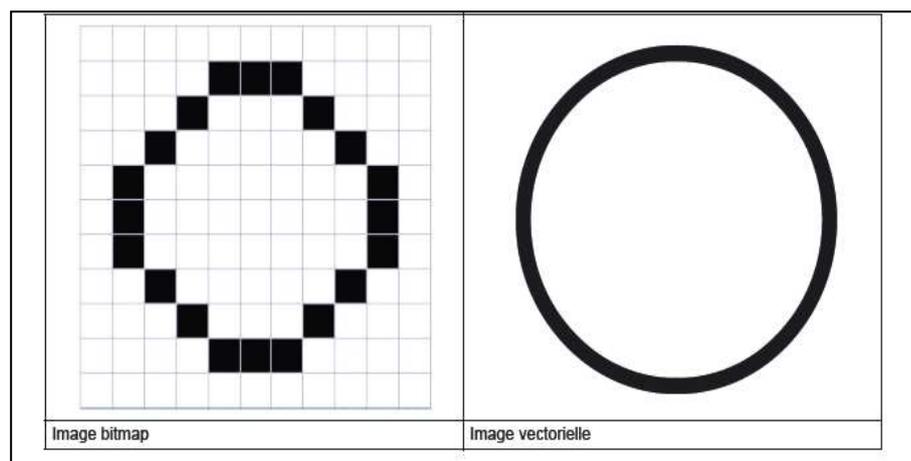


Fig.1.4 : exemple d'une Image vectorielle

7. Représentation d'une image

Le but d'un processus de vision par ordinateur est de trouver une relation entre une image en entrée et un modèle du monde réel.

Au cours du processus, l'information devient plus dense et la sémantique de l'image plus pertinente → plusieurs niveaux de représentation sont nécessaires :

On peut distinguer 4 niveaux :

- image brute : matrice de pixels également le résultat des opérations de prétraitements.
- image segmentée : différentes parties de l'image sont regroupées.
- représentation géométrique : connaissance des formes 2D et 3D.
- modèle relationnel : plus haut niveau d'abstraction permettant des traitements prenant en compte la connaissance du milieu [11].

8. Différente perception d'une image

- Une matrice de points de z^2
- un signal bidimensionnel à support et à valeurs bornées $A [i, j]$: L'image résulte de l'échantillonnage d'un signal continu.
- un processus stochastique $A[s]$
- un vecteur aléatoire $A = (A_1, \dots, A_s, \dots)$ où A_s est une variable aléatoire associée au site s .
- une surface $(i, j, A [i, j])$ de l'espace Z^3 [11].

9. Connexité dans les images

Le problème que nous voulons aborder ici concerne les relations entre pixels dans les images numériques ceci, pour préparer l'étude des divers algorithmes de traitement d'images impliquant des opérations entre pixels voisins.

Notamment, la notion de connexité entre pixels est une notion particulièrement importante pour la détection des frontières d'objets dans une image et des pixels composant un objet :

Deux pixels seront considérés comme connexes (appartenant au même objet donc) s'ils satisfassent deux critères:

- d'une part un critère de similarité (par exemple même niveau de gris)
- s'ils sont adjacents (voisins) [8].

9.1. Voisinage d'un pixel

Un pixel $I(x, y)$ possède quatre voisins horizontaux et verticaux qui forment ce qu'on appelle le 4-voisinage de $I(x, y)$. Si on considère un pixel comme un carré élémentaire, les pixels présentant un côté commun avec $I(x, y)$ sont appelés les 4-voisins de (x, y) .

Le pixel $I(x, y)$ possède aussi quatre voisins diagonaux, ce sont les pixels qui ont un sommet commun avec (x, y) .

L'ensemble des huit voisins du pixel (x, y) représentent ce qu'on appelle les 8-voisins (8-voisinage). Ces huit voisins forment la fenêtre 3x3 du pixel $I(x, y)$.

On dit aussi que ces pixels sont 8-adjacents de (x, y) .

9.2. Connexité

La connexité est une propriété de liaison entre deux pixels qui fait qu'on les considère comme faisant partie de la même région dans une image.

En supposant que deux pixels P et Q vérifient déjà un certain critère de similarité, on peut définir différents types de connexités:

❖ Connexité 2D

- ✓ 4-connexité ou connexité par arête : Les deux pixels sont tels que Q est un des 4-voisins de P
- ✓ 8-connexité ou connexité par face : Les deux pixels sont tels que Q est un des 8-voisins de P
- ✓ Connexité mixte : ou bien P et Q sont 4-voisins ou bien P et Q sont voisins diagonaux et aucun des 4-voisins communs à P et Q ne sont 4-connexes.

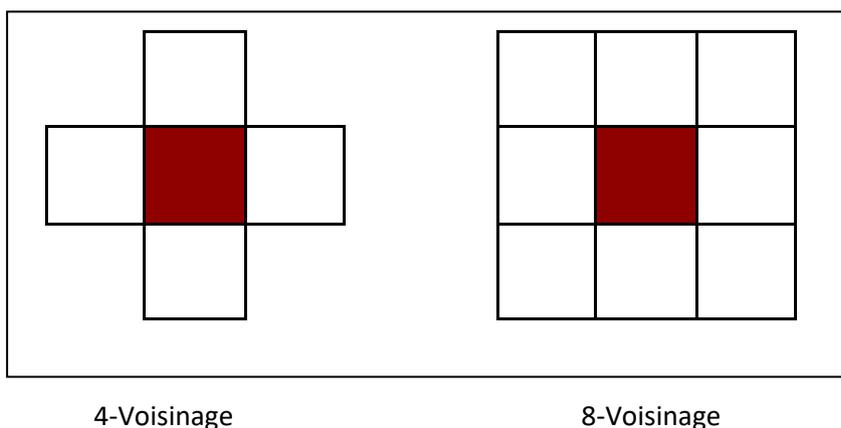


Fig. 1.5 : Représentation de voisinage

On remarque que la « méthode » de connexité mixte consiste à appliquer en priorité la 4-connexité, et la 8-connexité là où la 4-connexité ne “marche” pas.

❖ Connexité 3D

Un élément d'une image est appelé un voxel (volume élément) et, est représenté par un cube unitaire.

- ✓ 6-connexité ou connexité par face
- ✓ 18-connexité ou connexité par arête
- ✓ 26-connexité ou connexité par sommet

Chaque pixel/voxel ne représente pas seulement un point de l'image mais une petite zone de l'image.

Le nombre de pixel pour une image donnée détermine la résolution spatiale, plus elle est basse, moins l'on voit le sujet et plus les transitions semblent importantes.

9.3. Chainage de pixel

La notion de connexité permet de relier des pixels adjacents (vérifiant le même critère de similitude). Il existe un chaînage de longueur n entre deux pixels P et Q si on peut trouver une suite de $(n+1)$ pixels $P_0=P, P_1, P_2, \dots, P_n=Q$ tels que, deux pixels consécutifs sont toujours connexes.

On parlera de 4-chainage ou de 8-chainage suivant le type de connexité choisie.

Les deux pixels P et Q sont alors dits connexes et la relation de connexité qui les lie est une relation d'équivalence qui a la propriété d'être réflexive, symétrique et transitive (notons que la connexité mixte n'est pas transitive).

9.3.1. Région connexe composante connexe

Un sous ensemble R de pixels (satisfaisant le même critère de similarité) d'une image est une région connexe si quelque que soit un couple de pixels P et Q de ce sous ensemble, il existe au moins un chaînage entre ces pixels formé de pixels de R .

Soit S , un sous ensemble de pixels d'une image (ces pixels satisfaisant le même critère de similarité), alors quel que soit un pixel P de S , l'ensemble des pixels de S connectés par chaînage à P forment une composante connexe (région connexe) de S .

On en déduit que deux composantes connexes distinctes sont disjointes.

9.3.2. Fond, Forme de connexité

Lorsque qu'on examine une image considérée comme un ensemble de composantes connexes (la **forme** : rassemblant tous les pixels satisfaisant le même critère de similarité) sur un **fond** (rassemblant les autres pixels); on imagine aisément que les composantes connexes considérées seront différentes suivant que l'on utilise la 4-connexité, la 8-connexité ou la connexité mixte.

Exemple: soit la portion d'image ou le fondes en noir (0) et la forme en blanc (1)

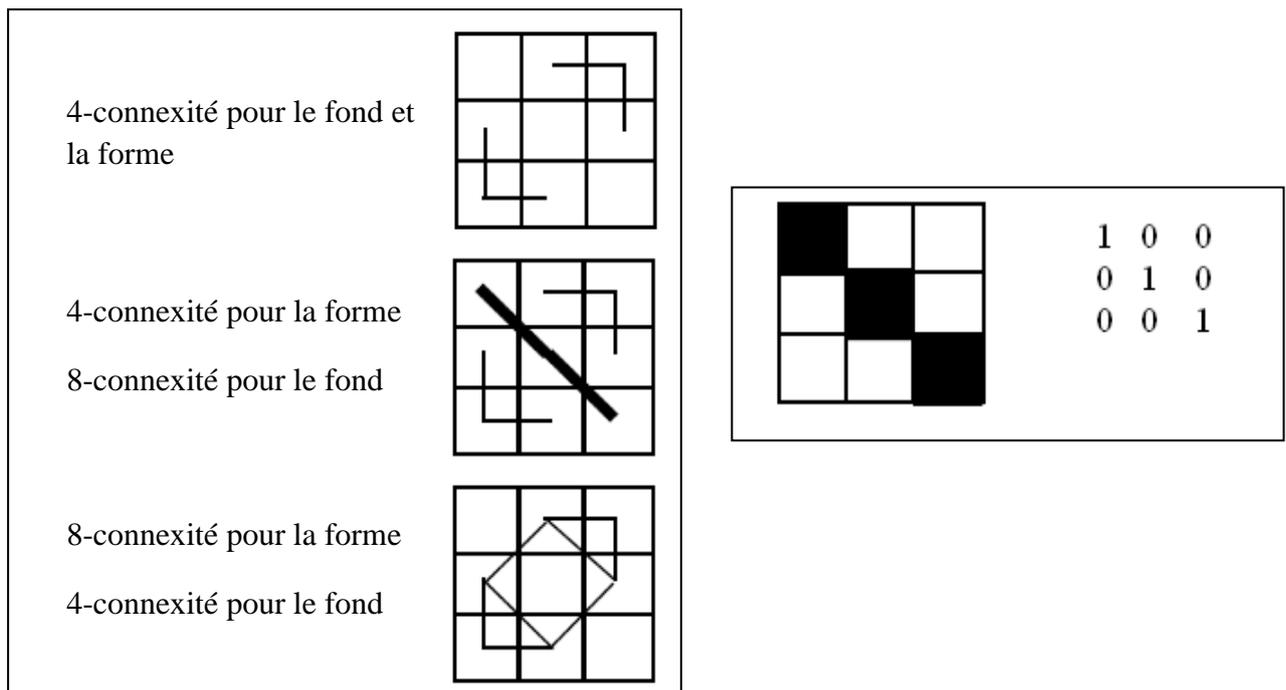


Fig. 1.6 : Fond et Forme de connexité

Ceci explique que la recherche des régions connexes est souvent réalisée en prenant les conventions suivantes:

- 4-connexité pour la forme et 8-connexité pour le fonds
- 8-connexité pour la forme et 4-connexité pour le fonds.

9.4. Etiquetage de composantes connexes

C'est l'opération qui consiste dans une image binaire (forme/fonds) à trouver toutes les classes d'équivalence de la forme, c'est à dire les différentes composantes connexes et, à les étiqueter.

L'étiquetage consiste à affecter une étiquette identique (par exemple un numéro) à tous les pixels d'une même composante connexe.

Exemple d'algorithme général :

1- scanner les pixels de l'image ligne par ligne et ne considérer que ceux de la forme (critère de similarité, pixel à 1,...).

Si le pixel n'a pas de voisin connexe déjà étiqueté, créer une nouvelle étiquette et l'affecter à ce pixel.

Si le pixel a exactement un seul pixel connexe étiqueté, lui affecter cette étiquette.

Si le pixel a plus d'un pixel connexe (avec des étiquettes différentes) lui affecter une de ces étiquettes et mémoriser que toutes ces étiquettes sont équivalentes.

2- Refaire une passe sur l'image en groupant les étiquettes équivalentes sur une seule étiquette.

10. Distances entre pixels

Il existe différentes mesures de distances entre deux pixels (x_1, y_1) et (x_2, y_2)

Distance Euclidienne : $\sqrt{((X_1 - X_2)^2 + (Y_1 - Y_2)^2)}$

➤ Distance « City block »: $|X_1 - X_2| + |Y_1 - Y_2|$

➤ Distance « chessboard »: $\max(|X_1 - X_2|, |Y_1 - Y_2|)$

Les deux premières distances indiquées sont respectivement l'expression des normes vectorielles

L_1, L_2, L_∞ . Soit un vecteur (v_1, v_2, \dots, v_n) alors $L_p(v) = \|V\|_p = (\sum_{i=1}^n |V_i|^p)^{1/p} \dots \dots (8)$

11. Les systèmes de traitement d'images

Un système de traitement numérique d'images est composé de :

11.1. Acquisition des données images

L'acquisition d'images constitue un des maillons essentiels de toute chaîne de conception et de production d'images. Pour pouvoir manipuler une image sur un système informatique, il est avant tout nécessaire de lui faire subir une transformation qui la rendra lisible et manipulable par ce système. Le passage de cet objet externe (l'image d'origine) à sa représentation interne (dans l'unité de traitement) se fait grâce à une procédure de numérisation. Ces systèmes de saisie, dénommés optiques, peuvent être classés en deux catégories principales : les caméras numériques et les scanners [7].

A ce niveau, notons que le principe utilisé par le scanner est de plus en plus adapté aux domaines professionnels utilisant le traitement de l'image comme la télédétection, les arts graphiques, la médecine, etc.

Le développement technologique a permis l'apparition de nouveaux périphériques d'acquisition appelés cartes d'acquisition, qui fonctionnent à l'instar des caméras vidéo, grâce à un capteur C.C.D. (Charge Coupled Device). La carte d'acquisition reçoit les images de la camera, de la T.V. ou du scanner afin de les convertir en informations binaires qui seront stockées dans un fichier.

11.2. Dispositifs de numérisations d'images

Suivant l'objet ou le document à numériser et le domaine d'application dans lequel l'image va être utilisée, il existe divers dispositifs de numérisation d'images allant du simple scanner à main au satellite de télédétection.

11.3. Prétraitement Et Post-Traitement

Le prétraitement concerne les images acquises tandis que le post-traitement concerne les images traitées. Dans les deux cas, les images ne peuvent échapper aux effets de dégradations dus essentiellement aux phénomènes physiques tels que :

- La diffraction (déviation) du système optique.
- Le flou dû au mouvement de l'image durant son acquisition.

Pour pallier à ces dégradations, on utilise en général le filtrage.

11.4. Notions du filtrage

11.4.1. Définition

Le filtrage est une opération qui consiste à réduire le bruit contenu dans une image. Il est considéré comme une transformation de l'image.

Le système linéaire utilisé est appelé filtre numérique.

Parmi ces systèmes, nous distinguons : les filtres passe-bas (lissage), filtres passe-haut (accentuation), filtres passe-bande (différenciation) et filtres directionnels [3].

11.4.2. Types de filtres

➤ **Filtres linéaires**

Les opérateurs sont caractérisés par leur réponse impulsionnelle (RI) $h(x,y)$ et,

$$C(x,y) = A * h(x,y) = \iint A(\alpha, \beta) \cdot h(x - \alpha; y - \beta) \cdot d\alpha d\beta \dots \dots \dots (9)$$

$A(x,y)$ est une image a traiter dont l'information utile est un échelon unitaire $U(x,y)$ perturbé par un bruit blanc stationnaire centré : $B(x,y)$ de densité spectrale N_0^2 .

Le filtre linéaire remplace chaque pixel par une valeur égale à une somme pondérée de ses voisins (convolution). Soit l'image de taille $N \times N$. Le filtre (ou masque) est une matrice carrée ($n \times n$) avec n impaire et $N \gg n$.

❖ Le filtre moyenneur :

- Le moyennage est un filtre passe-bas.
- Rend l'image floue, en particulier les contours.
- Élimine les dégradations locales de faibles dimensions. Valide lorsque les objets présents dans l'image sont de dimensions supérieures aux dégradations.
- La réponse impulsionnelle est donné par :

$$h(x,y) = \frac{1}{t^2} \dots \dots \dots (10)$$

❖ Le filtre gaussien :

- Utilisé par David Marr, l'un des pionniers du traitement de l'image.
- La réponse impulsionnelle est donnée par :

$$h(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \dots \dots \dots (11)$$

➤ **Filtres non linéaires**

❖ Le filtre médian : Les filtres de moyennage ont tendance à rendre l'image floue et donc à perdre de l'information sur les contours caractérisés par des fortes variations d'intensité. Pour diminuer cet effet, on ne moyenne plus sur le voisinage mais on prend la valeur médiane sur ce voisinage. C'est le filtre médian.

- Élimine le bruit impulsionnel.
- Préserve l'information de contour et peut être appliqué itérativement.
- Élimine les contours très ns. Un voisinage adapté permet de limiter cet effet.

- Filtrage d'un bruit impulsif (Salt & Pepper noise).

11.5. Notions de lissage

Le lissage est utile dans le cas des bruits "poivre et sel", c'est-à-dire dans le cas des bruits très localisés et très perturbants. Ces bruits induisent des trous ou des coupures sur des formes ou des frontières pas très nettes. La solution est alors d'appliquer un lissage local ou semi-local. Soient W une fenêtre de taille $[-m ; m] \times [-n ; n]$, g l'image brute et g_w l'image traitée ou lissée.

$$g_w(x, y) = \sum_{i, j \in W} w(i, j) \cdot g(x - i, y - j) \dots \dots \dots (12)$$

Les $g_w(x, y)$ sont définis par l'opérateur de lissage.

11.6. Traitement Numérique Des Images

On peut citer les traitements suivants :

➤ La Convolution

La convolution est le remplacement de la valeur d'un pixel par une combinaison linéaire ou non linéaire de ses voisins. Elle consiste à faire balayer une fenêtre (masque) sur l'ensemble des points de l'image [12].

➤ La Segmentation

La segmentation d'images est l'une des opérations les plus importantes dans le domaine de l'analyse d'images. L'objectif de cette opération est de diviser l'image en zones homogènes appelées régions. Ainsi, la segmentation d'une image permet de retrouver des formes ou des zones ayant un sens pour l'utilisateur ou pour une autre opération de traitement d'images [13].

11.7. Visualisation

Tout système de traitement d'image est doté d'un dispositif de visualisation qui permet l'affichage des images.

L'utilisation de différents types de reconstituteurs permet de transformer le signal numérique qu'est la matrice image en un signal analogique visible par l'œil de l'observateur. Pour cela, différents types de supports peuvent être employés : moniteur vidéo, clichés photographiques, impression sur papier. Dans tous les cas et pour chaque échantillon de l'image numérique, on recrée un nouvel élément d'image ou un nouveau pixel dont on choisit la forme de façon à reconstituer une image analogique qui soit la plus proche possible de l'image avant

numérisation compte tenu, des erreurs introduites lors de l'acquisition, de la numérisation et de la transmission .

11.8. Domaines d'application traitement d'images

Il existe plusieurs domaines d'applications traitement d'image on a essayé montrée quelques exemple dans La figure suivante :



Fig 1.7 : Exemples de champ d'application de traitement image.

➤ **Dans notre vie :**

- Grand public
- Médical
- Imagerie Aérienne
- Robotique
- Analyse
- Cinéma.

➤ **Dans l'industrie**

- Contrôle de processus, de qualité, Etc.

➤ **Dans le monde de la recherche**

- Le domaine militaire.
- Météorologie.
- La climatologie.
- L'océanographie.

12. Définition d'un satellite

Les satellites sont des objets qui sont en orbite autour d'un autre objet dans notre cas, la terre. Contrairement à la lune qui est un satellite naturel de la terre, les satellites artificiels sont des plates-formes, placés en orbite pour les besoins de la télédétection, des communications etc....

Un satellite est généralement composé de deux ensembles:

- 1- La plate-forme qui correspond à la structure destinée à supporter une ou plusieurs charges utiles et équipées pour leur fournir les ressources nécessaires à leur fonctionnement.
- 2- La charge utile : est l'ensemble d'éléments que peut transporter un satellite et est destinée à remplir une mission bien déterminée.

A bord d'un satellite sont installés des capteurs, pour enregistrer et analyser les rayonnements émis ou réfléchis par l'objet terrestre. [14]

13. Différent type de satellite

Ont une altitude d'environ 36000 kilomètres et se déplacent à une vitesse égale à celle de la terre, donnant l'impression qu'ils sont stationnaires. Cela permet au satellite d'observer la même région spécifique (voir figure.1.8). Les satellites de communication et de météorologie sont des satellites géostationnaires : exemple GOES (Geostationary Operational Environmental Satellite) en Amérique, METEOSAT (Meteorology Satellite) en Europe et en Afrique.

13.1. Satellites à défilement

Ont une orbite allant pratiquement du nord au sud et inversement. Ceci, combiné à la rotation de la terre (ouest-est), fait qu'au cours d'une certaine période, les satellites observent presque la totalité de la surface de la terre. Ils ont une altitude comprise entre 800 et 1000 km

Exemple de satellite à défilement : SPOT (Satellite Pour l'Observation de la Terre), LANDSAT et QUICK BIRD [15].

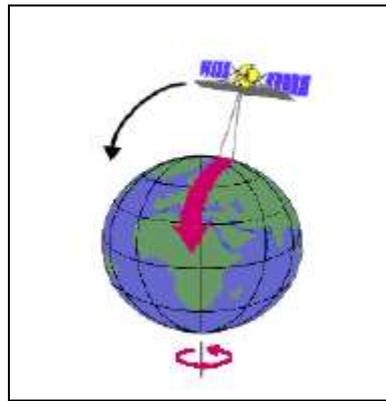


Fig.1.8 : Satellites à défilement

13.2. Satellites géostationnaires

Ont une altitude d'environ 36000 kilomètres et se déplacent à une vitesse égale à celle de la terre, donnant l'impression qu'ils sont stationnaires. Cela permet au satellite d'observer la même région spécifique (voir figure.1.9). Les satellites de communication et de météorologie sont des satellites géostationnaires : exemple GOES (Geostationary Operational Environmental Satellite) en Amérique, METEOSAT (Meteorology Satellite) en Europe et en Afrique.

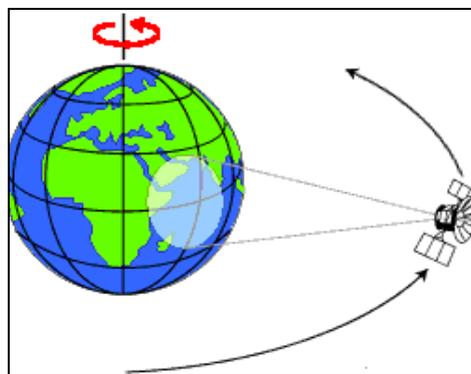


Fig.1.9 : Satellite géostationnaire

14. Définitions de SIG (Systèmes d'information Géographique)

Une constatation fréquente est la difficulté que l'on rencontre à définir un Système d'information géographique. Il n'y a pas véritablement de définition officielle. En fait, le vocable SIG désigne à la fois :

- le concept de chaîne de traitement numérique de l'information géographique.

- les outils informatiques qui permettent de l'instrumenter.
- et les applications qui sont construites avec cet outil.

Les définitions sont donc relativement générales pour ne pas perdre ce caractère polysémique.

Une définition simple présente un SIG comme " un système pour collecter, stocker, vérifier, manipuler, analyser et restituer des données spatialement référencées à la surface de la Terre " (DoE, 1987). Dans le même registre, un SIG est un " Système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer et de combiner, d'élaborer et de représenter des informations localisées géographiquement contribuant notamment à la gestion de l'espace ". (SFPT, Strasbourg 1989).

15. Segmentation et classification des images Multi spectrales

Il existe plusieurs classifications utilisées pour obtenir une carte d'image objective et il existe deux types de méthodes de classification:

Les non supervisées et les supervisées. Les résultats finaux diffèrent peu entre ces deux familles de classifieurs. L'intérêt d'utiliser l'une ou l'autre dépend du contexte.

Une classification va nous permettre :

- D'obtenir des cartes thématiques en classifiant les images Landsat et en labellisant sémantiquement les régions. Ce travail consiste alors à partitionner les images en régions homogènes en regroupant sous un même label les pixels de l'image répondant à un critère déterminé. Rappelons que le but principal est l'automatisation et la standardisation de la stratégie d'analyse d'image pour obtenir les variables environnementales.

Pour toutes les méthodes de classification, la discrimination des pixels est réalisée par leurs propriétés statistiques et/ou mathématiques. Il n'existe pas de méthode unique pour classifier une image. Le choix d'une technique sera généralement lié :

- la nature de l'image
- Les opérations de prétraitement effectuées
- Les primitives à extraire : Contours, formes des régions, textures, ...
- Les contraintes d'exploitation :

Complexité algorithmique, taille de mémoire disponible en machine, traitement en temps réel ou non.

La classification d'une image consiste donc à associer chaque pixel de l'image à une classe, le nombre de classe étant limité. Chaque classe correspond à un élément de structure du paysage. Autrement dit, il s'agit d'établir une cartographie de l'image en s'appuyant soit sur les valeurs radiométriques des pixels soit sur différents paramètres calculés conjointement (texture, indices...).

16. Conclusion

La représentation des images fixes est un des éléments essentiels des applications multimédias, comme dans la plupart des systèmes de communication.

La manipulation des images pose cependant des problèmes beaucoup plus complexes que celle du texte. En effet, l'image est un objet à deux dimensions, censée représenter un espace à trois dimensions, ce qui a deux conséquences majeures :

- Le volume des données à traiter est beaucoup plus important.
- La structure de ces données est nettement plus complexe.

Il en résulte que la manipulation, le stockage et la représentation de ces données se heurtent à certaines limitations.

Grâce au traitement d'image, ces contraintes sont levées ou contournées. En effet, ce domaine cherche à détecter la présence de certaines formes, certains contours ou certaines textures d'un modèle connu, c'est le problème de la détection.

Chapitre 2

La Fouille de Données Spatiales

1. Introduction

L'Extraction de Connaissances à partir de Données (*ECD*) constitue aujourd'hui un enjeu stratégique, économique, commercial et scientifique, en raison des intérêts majeurs qu'elle promet. Elle suscite de grandes espérances chez les entreprises, les administrations, les hôpitaux, les agences touristiques, etc. qui espèrent, grâce à elle, mieux évaluer les risques financiers, mieux cibler les consommateurs, mieux comprendre les mécanismes sociaux, mieux diagnostiquer les patients, etc. Hormis ces domaines, l'*ECD* peut être, globalement, partout où le volume d'information est largement grand, ce qui fait de son domaine d'application un domaine remarquablement vaste.

L'*ECD* est, d'un côté pratique, une démarche qui, en la suivant, permet d'extraire de nouvelles informations à partir des données, donc le terme *ECD* englobe tout un processus de découverte de connaissances dans des bases de données. Dans ce contexte, chaque nouvelle information, utile, intéressante, non triviale et non redondante, extraite des données, entre sous l'appellation de connaissance. Parmi les étapes qui rentrent en jeu dans le processus de l'*ECD* nous pouvons citer[16]:

- La collecte des informations et leurs organisations dans une *BD*.
- Le nettoyage de la base par un prétraitement ou des transformations des données.
- L'élaboration du modèle en appliquant les méthodes de *FD*.
- L'évaluation des résultats obtenus par *ECD*.

Dans ce processus, c'est dans la troisième étape que des méthodes seront appliquées et des modèles seront bâtis, elle est appelée '*étape de FD* ', et est considérée comme la plus importante.

2. Méthodes d'apprentissage

Pour extraire de la connaissance d'un ensemble de données, et la présenter sous forme d'un modèle, on fait appel aux méthodes d'ECD (phase de modélisation). Plusieurs disciplines s'intéressent à la problématique d'ECD comme le *Data Mining* ou *Knowledge Discovery in Database (KDD)*, la Reconnaissance de Formes, etc.

Ces disciplines s'appuient grandement sur les recherches réalisées dans d'autres domaines, notamment en Apprentissage Automatique qui est à l'origine de nombreux algorithmes exploités et utilisés en ECD.

L'Apprentissage Automatique avait été présenté vers la fin des années 70's dans la communauté '*Intelligence Artificielle*' pour dissoudre les problèmes que rencontrait l'ancien type d'apprentissage, l'Apprentissage par cœur [17]. En effet, à la même époque, l'approche Système Expert, qui concrétisait le type d'apprentissage par cœur, était très répandue mais commençait à montrer certaines limites comme la difficulté à acquérir l'expertise de l'expert humain et à la modéliser. Comme solution idéale à ces limites, est apparu l'Apprentissage Automatique avec la caractéristique principale est que l'apprentissage est effectué par la machine et non plus par l'homme [17].

Il existe d'autres types d'apprentissage, ceux basés sur la manière d'apprendre comme l'apprentissage par l'action, par l'explication et l'apprentissage inductif. Ce dernier se divise à son tour, en apprentissage supervisé et non supervisé. Dans le premier cas, on suppose qu'on possède un ensemble d'exemples pour lesquels on connaît à priori leurs classes. Le but est de construire une fonction de classement capable d'attribuer une nouvelle observation à l'une des classes prédéfinies. Dans le deuxième cas, on suppose avoir un ensemble d'observations, le but étant la détection de classes ou de clusters. L'apprentissage inductif supervisé est le plus utilisé. D'autres types d'apprentissage existent dans la littérature et qui sont liés à d'autres disciplines comme la Discrimination ou la Reconnaissances de Formes [18].

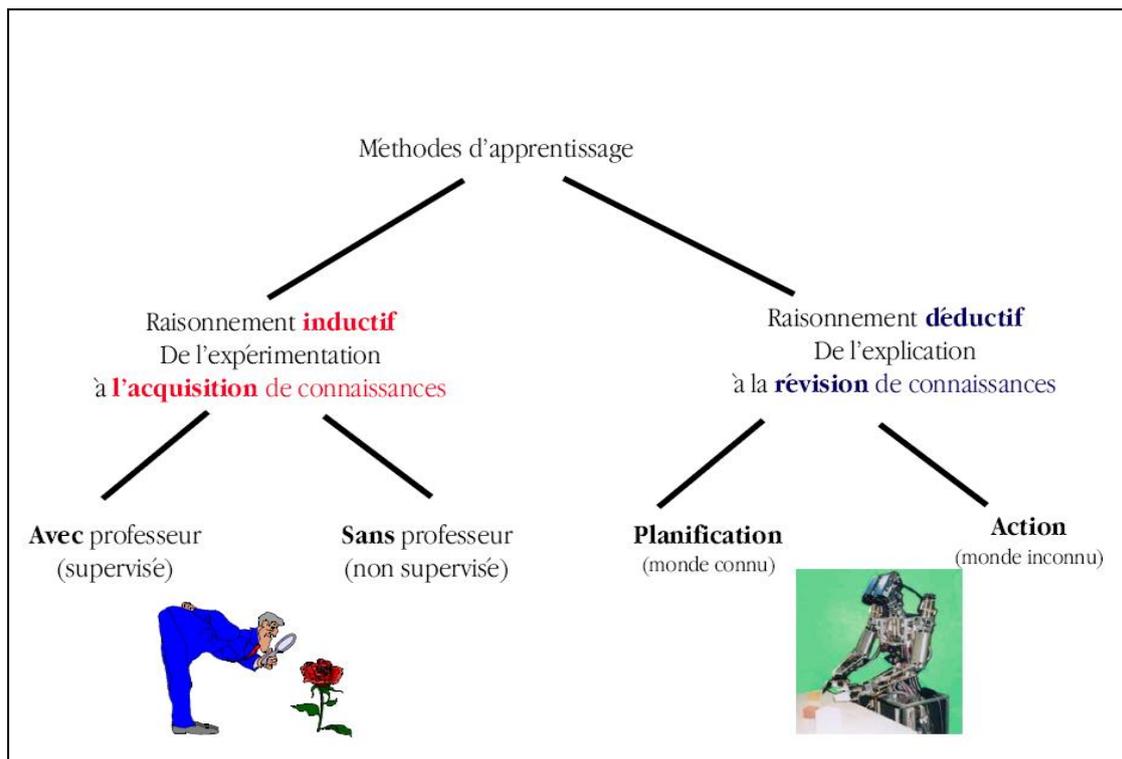


Fig 2.1 : Méthodes d'apprentissage

2.1. Les systèmes d'apprentissage automatique supervisé

Les systèmes d'apprentissage automatique supervisé fournissent les algorithmes d'apprentissage en quantités pour soutenir les décisions futures. Supervisé ou non supervisé, l'apprentissage automatique est utilisé surtout pour les chatbots, les véhicules autonomes, les logiciels de reconnaissance faciale, systèmes experts et robots. Les systèmes d'apprentissage supervisé sont principalement liés à l'IA basés sur la récupération, mais peuvent également être basés sur un modèle d'apprentissage génératif.

Les données utilisées pour l'apprentissage supervisé sont une série d'exemples constitués de paires de sujets d'entrée et de sorties attendues (également appelés signaux de supervision). Prenons l'exemple d'un système d'apprentissage supervisé pour le traitement d'images dans lequel on introduit des photos de véhicules appartenant aux catégories voitures et camions. Après un temps d'observation suffisant, le système doit être capable de faire la distinction entre plusieurs images non étiquetées et de les catégoriser ; une fois cet objectif atteint, l'apprentissage peut être considéré comme terminé.

Les modèles d'apprentissage supervisé présentent certains avantages sur les modèles non supervisés, mais ils ont aussi des limites. Par exemple, ils sont plus susceptibles de prendre des décisions auxquelles les humains peuvent s'identifier car elles sont basées sur des données fournies par les

humains. Mais dans le cas de la récupération, les systèmes d'apprentissage supervisé ont du mal à traiter les nouvelles informations. Si le système qui connaît les catégories des voitures et camions, par exemple, recevait l'image d'un vélo, il devrait le placer dans l'une de ces deux catégories, ce qui sera incorrect. Alors que si le système était générateur, il ne saurait pas forcément reconnaître le vélo, mais il serait capable de l'identifier comme appartenant à une autre catégorie.

2.2. Les systèmes d'apprentissage automatique non supervisé

L'apprentissage non supervisé est en train d'apprendre à un algorithme d'IA des informations qui n'ont pas été catégorisées et permet à cet algorithme d'interagir avec ces informations sans supervision.

Dans ce mode d'apprentissage, le système d'IA peut regrouper des informations non classifiées en fonction de leurs similitudes et de leurs différences, même si aucune classe n'est spécifiée. Les systèmes d'IA capables d'utiliser l'apprentissage non supervisé sont souvent associés à des modèles d'apprentissage génératifs, mais ils peuvent aussi fonctionner avec une approche basée sur la récupération (souvent associée à l'apprentissage supervisé). Les approches d'apprentissage supervisé ou non supervisé sont utilisées notamment dans les chatbots, les véhicules autonomes, les programmes de reconnaissance faciale, les systèmes experts et les robots.

Dans l'apprentissage non supervisé, les données fournies au système d'IA ne sont ni étiquetées, ni classées, et les algorithmes du système traitent les données sans aucun entraînement préalable. La sortie dépend des algorithmes codés. L'introduction dans un système d'une approche d'apprentissage non supervisé est un moyen d'expérimenter l'intelligence artificielle.

Les algorithmes d'apprentissage non supervisé peuvent exécuter des tâches de traitement plus complexes que les systèmes d'apprentissage supervisé, mais ils peuvent aussi être plus imprévisibles. Même si un système d'IA d'apprentissage non supervisé parvient tout seul, par exemple, à faire le tri entre des chats et des chiens, il peut aussi ajouter des catégories inattendues et non désirées pour y classer des races inhabituelles, créant la confusion au lieu de mettre de l'ordre.

3. Fouille de données

Les méthodes d'ECD forment par leur diversité une grande boîte à outils à la disposition des utilisateurs. Dans cette section, nous définirons brièvement l'étape la plus importante dans l'ECD qui est la fouille de données, mais avant cela, un petit paragraphe sera consacré à fixer quelques vocabulaires.

➤ Vocabulaires

Certains de ces vocabulaires ont été utilisés précédemment, certains le seront dans le reste des chapitres, ce qui permettrait d'éclaircir un ensemble d'expressions techniques que l'on rencontrera lors de l'introduction des méthodes.

- L'apprentissage Inductif Supervisé a pour but de construire un modèle de prédiction appelé *Classifieur*.
- Une *population* est un ensemble d'*individus* caractérisés chacun par un ensemble d'*attributs*.
- En apprentissage supervisé, un de ses attributs est à prédire, que l'on nomme *classe*, *variable endogène*, *concept* ou encore *variable à prédire*.
- Les autres attributs sont nommés *variables exogènes*, *variables explicatives* ou *variables prédictives*.
- Un individu dont on dispose de sa classe est un *exemple*.
- Le *Classifieur* est une procédure qui, à partir d'un ensemble d'exemples, peut prédire la classe de toute donnée.

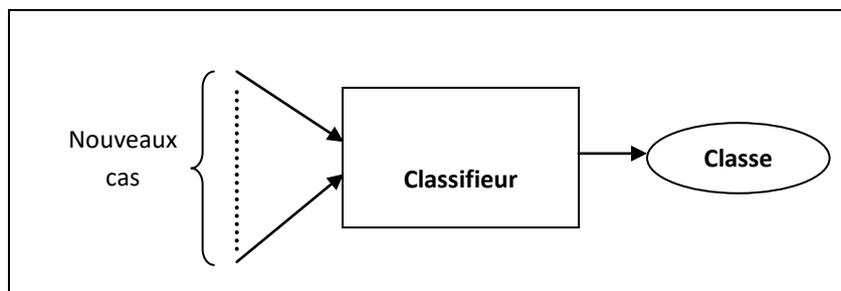


Fig 2.2 : Schéma générale d'un Classifieur

- Une opération d'apprentissage supervisé se déroule en deux étapes, la première dite d'*apprentissage*, où le modèle sera élaboré, et la deuxième dite de *test*, où le modèle sera testé et validé.
- Lors de la phase d'*apprentissage*, ce ne sont pas toutes les données qui sont prises en considération, mais plutôt un sous ensemble de la population nommé *Base d'apprentissage*, le reste sert au test et à la validation du modèle, il est appelé *Base de test*.

4. La fouille de données (data mining) spatiales

La fouille de données spatiales (*FDS*) est un domaine de recherche en pleine expansion. Elle s'attache à effectuer une analyse dirigée par les données sur des bases de données spatiales [55]. Elle offre de nouvelles perspectives pour beaucoup d'applications à caractère décisionnel comme le géo-marketing ou l'analyse de risques d'accidents [56] dont le volume de données ne cesse de croître. Cette tendance à développer de grandes bases de données s'explique par les progrès dans l'acquisition de données spatiales [57].

Le Data Mining Spatial (*DMS*) est né du besoin d'exploiter, dans un but décisionnel, des données à caractère spatial produites, importées ou accumulées, susceptibles de délivrer des

connaissances par le moyen d'analyses exploratoires et de fouille de données. Les données spatiales peuvent être de nature diverse comme des cartes thématiques, des images de satellites ou même des fichiers d'adresses [58].

4.1. Définition

La fouille de données spatiales est définie comme l'extraction de connaissances implicites, de relations spatiales ou d'autres propriétés non explicitement stockées dans la base de données spatiales [59].

Ses avantages sont d'une part son aspect exploratoire dans le sens où les hypothèses sont générées, contrairement à l'analyse classique. D'autre part, les informations sur la localisation spatiale et sur les liens de voisinage sont complètement intégrées.

Les méthodes mises en œuvre pour la fouille de données spatiales utilisent de manière intensive les relations spatiales. C'est ce qui distingue ces méthodes de celles appliquées dans le cas de données de type alphanumérique. Ces relations spatiales jouent donc un rôle primordial dans l'analyse de données spatiales et la découverte de connaissances.

Les travaux sur le data mining ont été développés essentiellement dans le contexte de données tabulaires simples, donnant lieu à des logiciels et des applications opérationnelles. Cependant, l'extension aux données complexes (multimédia, temporelles ou spatiales) constitue un problème de recherche ouvert [56].

Les tâches dans un processus de *FDS* sont définies comme une extension par les critères spatiaux de celles utilisées dans la fouille de données [60]. Ces différentes tâches permettent :

(1) de résumer les données, (2) de trouver des règles de classification, (3) de former des clusters d'objets similaires, (4) de trouver des associations et des dépendances pour caractériser les données, et enfin (5) d'extraire des tendances et de détecter les déviations [61].

4.2. Spécificités de la fouille de données spatiales

La fouille de données spatiales est définie comme l'extraction de connaissances implicites, de relations spatiales ou d'autres propriétés non explicitement stockées dans la base de données spatiales [59]. Ces relations spatiales jouent donc un rôle primordial dans l'analyse de données spatiales et la découverte de connaissances.

Ses avantages sont d'une part son aspect exploratoire dans le sens où les hypothèses sont générées, contrairement à l'analyse classique. D'autre part, les informations sur la localisation spatiale et sur les liens de voisinage sont complètement intégrées.

4.3. Caractéristiques des données spatiales

Les relations spatiales traduisent une caractéristique essentielle du monde réel. Elles mettent en évidence les relations et l'influence du voisinage entre les entités spatiales. La notion de dépendance

spatiale est endémique aux données géographiques et est connue comme la première loi en géographie [59].

Il est donc essentiel d'en tenir compte dans l'analyse et l'interprétation des données localisées.

Les indices d'auto-corrélation spatiale définis en analyse spatiale considèrent uniquement les interactions sur une couche thématique. Or, en règle générale, les couches thématiques sont fortement corrélées. Une carte des précipitations et de la densité de population sont, par exemple, corrélées. En effet, la densité de population dépend de la production agricole qui, elle-même, est liée aux précipitations. Par conséquent, nous distinguons deux types de relations spatiales :

- celles qui lient les objets d'une même classe (intra-thème).
- celles qui sont associées à plusieurs classes (inter-thèmes).

4.4. Matrice et graphe de voisinage

Les relations spatiales sont communément formalisées par la notion de graphe de voisinage, et peuvent être représentées sous forme d'une matrice de voisinage. Celle-ci est une matrice binaire M où $M[i,j]=1$ si l'objet i est voisin de l'objet j et $M[i,j]=0$ dans le cas inverse. Ceci est illustré par la figure (2.3).

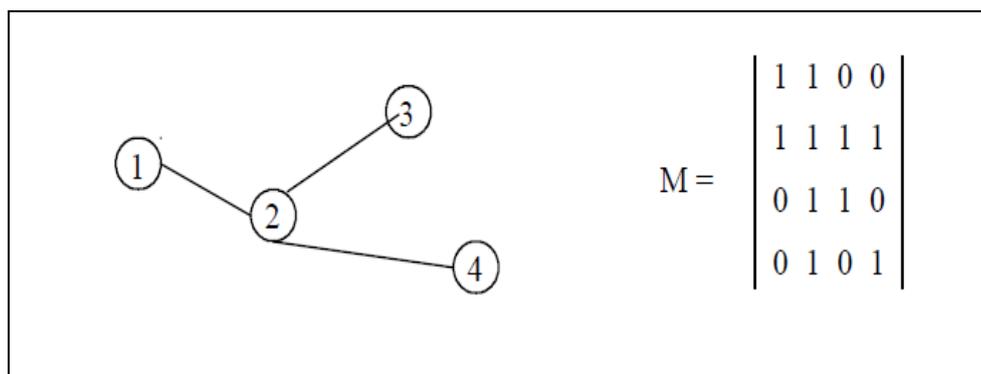


Fig 2.3 : Graphe de voisinage et matrice de voisinage

4.5. Travaux connexes

Les travaux de recherche dans le domaine de la *FDS* visent donc à proposer et à optimiser des méthodes d'analyse en tenant compte des relations spatiales. Ce domaine est à la croisée de plusieurs disciplines : les statistiques et les bases de données [62].

Une littérature abondante existe dans le domaine de l'analyse spatiale [63] et [64], domaine largement antérieur à l'apparition des premiers SIG. Beaucoup de travaux portent sur la mesure de l'auto-corrélation locale et globale [64]. La géostatistique [65] se préoccupe d'analyser des tendances à l'aide de techniques de kriging.

Des recherches récentes en analyse de données exploratoire et interactive ont été développées [66],[67], [68] et [69]. Openshaw [70] propose l'utilisation du calcul parallèle pour identifier des

clusters. D'autres travaux visent également à étendre l'analyse de données multidimensionnelle pour prendre en compte les contraintes de contiguïté [71].

Dans le domaine des bases de données, des algorithmes de fouille de données spatiales ont été proposés et des prototypes ont été développés : GeoMiner [72] et [59] est une extension de DBMiner [73] et est couplé au SIG MapInfo. GeoMiner comme DBMiner ont l'avantage d'intégrer les techniques OLAP (On-Line Analytical Process) et fouille de données. Un moteur SQL étendu, appelé GMQL (*GeoMiner Query Language*), permet d'exécuter les opérations de fouille de données définies. Certains algorithmes sont basés sur une connaissance à priori exprimée par des hiérarchies de concepts.

Par ailleurs, les travaux d'Ester se sont essentiellement articulés autour d'une structure de graphe de voisinage [74]. C'est le cas des méthodes comme la classification, la caractérisation et la recherche de tendances. Dans cette approche, les auteurs ont intégré au sein d'un SGBD géographique les opérateurs de graphe de voisinage. Citons d'autres travaux comme STING [75] centré sur le clustering et l'expression de requêtes dans un langage SQL étendu. Pour une bibliographie plus approfondie, le lecteur pourra se référer à [76].

4.6. Panorama des méthodes de fouille de données spatiales

Les méthodes types sont un prolongement des tâches de fouille de données intégrant les données et les critères spatiaux. Ainsi, une première *phase exploratoire* permet une description synthétique (indice d'auto corrélation globale, généralisation, densité, lissage), de découvrir les écarts donnant les spécificités locales (auto-corrélation locale ou analyse factorielle locale) ou de chercher des regroupements de données (clusters).

Cette première phase permet de guider la *phase décisionnelle*, où l'on procède à une analyse plus fine afin d'expliquer les écarts ou de caractériser les groupes (caractérisation, règles de classement ou d'associations). Nous allons maintenant décrire ces différentes méthodes et les illustrer à l'aide d'exemples issus des travaux de recherche actuels.

4.6.1. Phase exploratoire

Bien avant l'ère des SIG, des mesures du degré de dépendance aux voisins, dites d'auto corrélation spatiale globale ont été étudiées. Elles exploitent, hormis les attributs de l'objet, la matrice de voisinage définie au paragraphe 2.3. Ces mesures sont calculées à l'aide de deux méthodes complémentaires : l'indice de Moran (en 1948) et l'indice de Geary (1954) [76]. Dans le cas où les données seraient corrélées, il faut les simplifier afin de faire apparaître une tendance générale. Pour réaliser cela, il existe différentes approches dont celles basées sur la densité, l'analyse multidimensionnelle lissée ou la généralisation.

- **Généralisation spatiale**

Cette méthode est une extension aux données spatiales de la généralisation basée sur l'induction orientée attribut proposée dans [72]. Elle consiste à substituer les valeurs estimées trop détaillées par des valeurs moins détaillées jusqu'au niveau de détail souhaité, puis à agréger et compter les n-uplets identiques ainsi obtenus. Cette méthode permet de résumer les données et constitue une première étape pour induire des règles d'associations. Elle nécessite au préalable de disposer d'une connaissance a priori que l'on trouve dans des « hiérarchies de concepts » définies par des experts. Une hiérarchie de concepts est définie pour un attribut. Elle décrit le passage des concepts les plus spécifiques - correspondant aux valeurs de l'attribut dans la base de données- aux concepts plus généraux de niveau supérieur. Pour un attribut de type spatial, cette hiérarchie est appelée *hiérarchie spatiale* et correspond à une relation spatiale d'inclusion entre objets. Le découpage administratif en pays, régions, département, communes, etc. en est un exemple. Pour un attribut non spatial, on parle de *hiérarchie thématique*.

La généralisation à dominante spatiale et la généralisation à dominante non spatiale ont été définies dans [77]. Cette méthode a été implémentée dans le prototype *GeoMiner*. Elle permet de générer des règles d'association sur les attributs généralisés qui ne sont pas déductibles au niveau de détail.

Comme elle réduit le nombre de modalités des variables généralisées, elle peut également constituer une étape préalable à d'autres analyses telles que la recherche de règles caractéristiques [78] ou des analyses de correspondances. En outre, ce mécanisme est à la base de l'extension au spatial des techniques *OLAP* dans *GeoMiner*[79] permettant d'explorer les données à différentes échelles sémantiques et spatiales.

- **Analyse globale par lissage**

L'analyse multidimensionnelle lissée est obtenue en remplaçant chaque valeur de la matrice de voisinage par le barycentre de ses voisins. L'application des méthodes d'analyse factorielle se fait ensuite sur le tableau ainsi modifié. L'extension des méthodes factorielles pour la prise en compte de la contiguïté est décrite dans [80].

- **Analyse locale**

A l'inverse de l'analyse globale qui cherche à gommer les particularités, l'analyse locale vise à les faire ressortir pour mettre en évidence les données atypiques. Par exemple, si nous considérons les variables taux de scolarisation et revenu moyen, il n'existe a priori pas de corrélation au niveau global sur l'ensemble du territoire. Par contre, si l'analyse s'effectue au niveau local sur une région bien précise, une corrélation pourra être trouvée. L'analyse locale peut éventuellement contredire les résultats de l'analyse globale. L'auto corrélation locale consiste à calculer un indice local d'association spatiale [67], dérivé de la formule de l'indice global où l'on remplace la matrice de voisinage par le vecteur ou la ligne de la matrice correspondant à l'objet. L'analyse multidimensionnelle locale est

analogue à celle d'analyse lissée présentée ci-dessus. De la même manière, elle procède tout d'abord par transformation du tableau initial en tableau contrasté. Ce dernier correspond à la différence du tableau initial et du tableau lissé présenté dans l'analyse locale. La suite est une analyse factorielle classique sur le tableau ainsi modifié.

- **Clustering**

Le clustering est une méthode de classification automatique non supervisée qui regroupe des objets dans des classes. Son but est de maximiser la similarité intra-classes et de minimiser la similarité inter-classes. Elle est couramment utilisée en fouille de données et elle est bien connue dans le domaine des statistiques [71].

La transposition au domaine spatial des méthodes de clustering s'appuie sur une mesure de similarité d'objets localisés suivant leur distance métrique. Néanmoins, la finalité du clustering en spatial n'est pas tant de former des classes que de détecter des concentrations anormales (par exemple, détecter un point chaud dans l'étude de criminalité, ou des zones à risque en accidentologie). Les travaux sur le clustering spatial sont surtout axés sur l'optimisation des algorithmes. Ainsi [81] propose une extension de l'algorithme *DBSCAN* en utilisant l'index spatial *R**tree, ainsi que des extensions incrémentales des algorithmes *DBCLASS*, *DBLEARN*. Openshaw propose l'utilisation de calculs intensifs et l'indexation spatiale dans sa machine *GAM* [70]. Alors que le clustering à l'origine traite des tableaux à plusieurs variables, l'application aux données spatiales est basée uniquement sur la localisation (souvent sur un ensemble de points). L'extension aux attributs non spatiaux et aux objets de forme autre que ponctuelle a été proposée dans *GDBSCAN* en redéfinissant la fonction de similarité [70].

Cette étape est souvent utilisée en amont d'autres tâches de type décisionnelles comme la recherche d'associations entre groupes et d'autres entités géographiques ou la caractérisation au sein d'un groupe. Un exemple d'application est de former des clusters d'habitations puis de rechercher des caractéristiques communes par cluster.

4.6.2. Phase décisionnelle

Le terme explicatif ici est lié à une intervention de l'analyste qui, à la suite d'une découverte de clusters ou de valeurs atypiques par rapport à une tendance, focalise son analyse sur un sous-ensemble d'objets, sur une partie des variables ou encore sur une zone géographique. Cette partie des données est ensuite analysée dans le but d'expliquer sa particularité par des liens avec certaines valeurs ou par des règles caractéristiques. Ces méthodes, à l'inverse des méthodes précédentes, opèrent sur plusieurs couches thématiques pour permettre d'expliquer un phénomène suivant les propriétés de son entourage. Nous décrivons les méthodes de caractérisation, de règles d'association et de classification.

- **Caractérisation**

Dans [82], l'auteur définit la caractérisation comme l'induction des propriétés caractéristiques d'un sous-ensemble de données. Une règle caractéristique est une assertion qui décrit un concept satisfait par tous ou une grande partie des objets sélectionnés. Appliquée à des bases de données spatiales, la caractérisation découvre en plus le niveau d'extension de ces propriétés aux "voisins". Une propriété caractéristique d'un sous-ensemble S est un prédicat $pi=(attribut= valeur)$ tel que :

- (a) sa fréquence relative dans S et dans son voisinage jusqu'à un ordre n est significativement différente par rapport à sa fréquence relative dans la base (rapport de fréquences supérieur à un seuil donné).
- (b) sa fréquence relative est significativement différente dans le voisinage d'une proportion minimum d'objets du sous-ensemble S (proportion supérieure à un seuil de confiance).

Le module *Geo-Characterizer* de *GeoMiner* permet la découverte de règles caractéristiques après application de la méthode de généralisation à dominante spatiale. Ces règles associent les données non spatiales avec leur localisation et ce pour chaque localisation.

- **Règles d'association**

L'extension de la découverte de règles d'association [83] aux données spatiales permet de générer des règles de type :

$X \rightarrow Y (s, c)$ avec s comme support et c la confiance telles que X et Y sont des ensembles de prédicats spatiaux et non spatiaux. En d'autres termes, ceci revient à trouver des associations entre des propriétés des objets et celles de leur voisinage.

Le module *Geo-associator* de *GeoMiner* permet la découverte d'associations depuis un ensemble de couches thématiques donné, en utilisant certains attributs et des seuils de support et de confiance. Ainsi, la recherche d'associations impliquant les terrains de golf et les autres entités géographiques (bâtis, infrastructure, etc.) génère un ensemble de règles comme par exemple :

$is_a(x, "golf") \rightarrow close_to(x, "zone\ pavillonnaire") (61\%, 70\%)$

L'algorithme proposé dans [84] propose deux phases dans l'évaluation du prédicat spatial. La première fait un test approximatif et génère des candidats pour un test précis du prédicat en seconde phase. Pour cela, des prédicats spatiaux généralisés ont été introduits.

- **Classification**

La recherche de règles de classement vise à structurer un ensemble d'objets en classes d'objets ayant des propriétés communes. Cette tâche est réalisée par apprentissage supervisé qui, à partir de classes fournies partiellement en extension (un échantillon de la base de données), induit une description en intention permettant de classer les prochaines données. On parle de segmentation ou de scoring en statistique. La classification s'exprime généralement sous la forme d'un arbre de décision pour lequel l'algorithme de référence est *ID3*[85].

L'extension au domaine spatial a été définie par l'extension aux propriétés de leurs voisins jusqu'à un ordre N de voisinage [74]. Ainsi, il est possible de trouver une règle de type :

Si population élevée et type de voisin = route et voisin de voisin = aéroport Alors puissance économique élevée (à 95%).

Une approche similaire a été proposée dans [86], à la différence de la précédente, l'utilisateur précise les thèmes à explorer et l'ordre de voisinage est limité au premier (voisins directs).

5. Conclusion

La fouille de données spatiales dérive de la fouille des données traditionnelles. Néanmoins, elle présente une spécificité importante par la prise en compte des relations spatiales. Les matrices et les graphes de voisinage constituent un pivot dans la conception d'opérateurs dédiés à la fouille de données spatiales.

Ce domaine de recherche soulève de réels problèmes, tant scientifiques que technologiques. La résolution de ces problèmes a non seulement des répercussions sur ces deux aspects (scientifique, technologique), mais ont également un très net impact socio-économique à cause de leur très large champ d'applications. Du point de vue scientifique, le Data Mining Spatial étend la problématique du Data Mining aux données dépourvues de toute structure, aux données semi-structurées, ainsi qu'aux données évolutives. A un niveau plus général, il pose la problématique de la complexité opératoire (computationnelle) et aussi systémique des outils à développer dans ce contexte. Cette complexité est très fortement liée à la masse importante des données complexes (images satellitaires par exemple), leur diversité, leur distribution, leur aspect dynamique et au caractère ouvert du spatiale.

Il faut noter que cela nécessite la mobilisation de ressources assez rarement toutes présentes dans la plupart des services :

- De bon spécialiste et staticien.
- De bon spécialiste des bases de données ("data warehouse")
- De bon informaticien programmeur (Et de connaissance en algorithmique avancée et des agents de "mining")

Nous avons détaillé, dans ce chapitre, l'importance du prétraitement dans le processus d'ECD, surtout dans le cas de l'Extraction de Connaissances à partir des Images. Dans le chapitre qui suit nous ferons une présentation générale des concepts structurant la méthodologie de traitement des images satellitaires.

Chapitre 3

La Classification des images satellitaires et la méthode FCM-AT

1. Introduction

La classification de l'image est une partie importante de la télédétection, l'analyse d'image et la forme de reconnaissance. Dans certains cas, la classification elle-même peut faire l'objet de l'analyse. Pour par exemple, la classification de l'utilisation des terres à partir de données produit une carte comme l'image comme produit final de l'analyse (Campbell, 2002).

La classification d'image forme donc un outil important pour l'examen des images numériques. Le terme classificateur se réfère vaguement à un programme informatique qui met en œuvre une procédure spécifique de classification des images (Campbell, 2002). L'analyste doit choisir une classification méthode qui accomplira le mieux une tâche spécifique. À l'heure actuelle, il n'est pas possible d'indiquer quel classificateur est le meilleur pour toutes les situations la caractéristique de chaque image et les circonstances de chaque l'étude varient si grandement. Par conséquent, il est essentiel que chaque analyste comprenne les stratégies alternatives pour la classification des images afin qu'il ou elle peut être prêt à choisir le plus approprié classificateur pour la tâche en main. À l'heure actuelle, il existe différentes procédures de classification des images utilisé à des fins différentes par divers chercheurs (Butera 1983, Ernst and Hoffer 1979, Lo et Watson, 1998, Ozesmi-Bauer 2002, Dean-Smith 2003, Pal-Mather 2003, Liu et coll. 2002) .

Ces techniques se distinguent en deux classifications supervisées et non supervisées. En outre, la classification supervisée comporte différentes sous-classifications méthodes qui sont

nommées parallélépipèdes, probabilité maximale, distances minimales et classificateur Fisher Méthodes. Il existe différents traitement d'image et SIG logiciels utilisant partout dans le monde. Beaucoup d'entre eux ont les propriétés et fonctionnalités similaires pour l'utilisation de la télédétection Fins .

2. Classification et Analyse d'images satellitaires

La classification automatique de documents est un problème connu en informatique, il s'agit d'assigner un document à une ou plusieurs catégories ou classes. Le problème est différent selon la nature des documents en question, en effet la classification de textes diffère de la classification de documents images, vidéo ou encore son. On peut aussi imaginer des classifications selon des paramètres associés aux documents tels que par exemple l'auteur, la date de parution ...

Dans le cadre de ce projet et dans la suite de rapport nous nous basons sur la classification de documents de type image.

➤ Types de classification et ensemble flou

Cette brève présentation de la logique floue montre que celle-ci intervient tout à fait naturellement en classification, puisqu'une classe doit en effet regrouper des individus semblables, concept éminemment flou

Nous examinons dans ce travail l'impact du flou dans les types de classification suivants :

❖ **Classification supervisée** : L'objectif principale de définir des règles permettant de classer des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets. Les méthodes s'étendent souvent à des variables Y quantitatives (régression). On dispose au départ d'un échantillon dit d'apprentissage dont le classement est connu. Cet échantillon est utilisé pour l'apprentissage des règles de classement. Il est nécessaire d'étudier la fiabilité de ces règles pour les comparerait les appliquer, évaluer les cas de sous apprentissage ou de sur apprentissage (complexité du modèle). On utilise souvent un deuxième échantillon indépendant, dit de validation ou de test.

❖ **Classification non supervisée**: Les méthodes de classification non supervisée regroupent les objets en un nombre restreint de classes homogènes et séparées. Homogènes signifie que les éléments d'une classe sont les plus proches possible les uns des autres. Séparées veut dire qu'il y a un maximum d'écart entre les classes.

La proximité et l'écart ne sont pas nécessairement au sens de distance. L'homogénéité et la séparation entrent dans le cadre des principes de cohésion et d'isolation de Cormack.

Cet objectif est à distinguer des procédures de discrimination, ou encore déclassement pour lesquelles une typologie est a priori connue, au moins pour un échantillon d'apprentissage. Nous sommes dans une situation d'apprentissage non-supervisé.

❖ **Classification semi supervisée:** L'apprentissage semi-supervisé trouve ses racines dans les problèmes d'apprentissage en présence de données manquantes. Ainsi de nombreux travaux ont été effectués ce sujet à la fin des années 1970. Cependant l'utilisation des données non étiquetées pour améliorer la précision de la règle de classement apprise connaît un regain d'intérêt depuis la fin des années 1990 ou la communauté du Machine Learning a commencé à s'intéresser à ce sujet. Ceci suite à la disponibilité d'un grand nombre de données acquises de manière automatique grâce aux nouvelles technologies. Ainsi des travaux en classification de texte ont contribué à relancer l'intérêt de l'utilisation des données non classées en vue d'améliorer la précision de la règle de classement apprise [87].

3. Segmentation d'image

3.1. Définition de la segmentation

"La segmentation est un traitement de bas niveau qui consiste à créer une partition de l'image A en sous-ensembles R_i , appelés régions tels qu'aucune région ne soit vide, l'intersection entre deux régions soit vide et l'ensemble des régions recouvre toute l'image.

Une région est un ensemble de pixels connexes ayant des propriétés communes qui les différencient des pixels des régions voisines". [88]

La segmentation est la partition d'une image en un ensemble de régions qui ne se chevauchent pas et dont l'union est l'image entière. Quelques règles à suivre pour obtenir une segmentation sont [89]:

- Les régions doivent être uniformes et homogènes par rapport à certaines caractéristiques (niveau de gris, écart type, gradient).
- Leurs intérieurs doivent être simple et sans beaucoup de petits trous (des parties de région non segmentés).
- Les régions adjacentes doivent avoir des valeurs très différentes par rapport à la caractéristique prise en compte dans la segmentation.
- Les limites de chaque région doivent être simples et spatialement précises.

3.2. Le choix d'une technique de segmentation:

Le choix est lié à :

- La nature de l'image (éclairage, contours, texture, etc).
- Aux opérations en aval de la segmentation (compression, reconnaissance des formes, Mesures, etc).

- Aux primitives à extraire (droites, régions, textures, etc).
- Aux contraintes d'exploitation (temps réel, espace mémoire, etc).

3.3. Objectifs de la segmentation :

- Fournir des régions homogènes (selon un critère donné).
- Localiser de manière précise les contours des régions.
- L'étude et l'interprétation des structures anatomiques.
- Réduction de bruit. [90]

4. Présentation globale du cadre d'application

Dans le cadre de cette présente étude, nous nous intéressons au traitement des images satellitaires. En effet, notre objectif est d'extraire d'une façon automatique, des réseaux à partir d'une image *LANDSAT 5 TM* pour la région d'Oran (Algérie) en procédant par fusion deux méthodes : le *K-Means flou (Fuzzy K-Means)* ou *FCM*, que nous appliquerons sur des images multi spectrales, et la méthode de Seuillage Adaptative ou Adaptive Thresholding (*AT*).

Le système permet d'une part de segmenter l'image originale par les *K-Means flous* et, d'autre part d'appliquer la méthode du seuillage adaptative sur une image mono spectrale. A cet effet, nous aurons besoin des trois canaux de *LANDSAT* dans la mesure où les *K-Means flous* ne sont applicables qu'à des images multi spectrales, et nous transformerons ensuite l'image obtenue en niveau de gris afin qu'elle soit segmentée par le seuillage adaptatif (voir figure 3.1 ci-dessous).

5. Le Système de classification proposé

Le modèle décisionnel du système proposé est largement inspiré du modèle d'aide à la décision adapté au domaine de l'environnement et à la décision spatiale proposé par Pictet [92] et repris par Hamdadou [93]. Le modèle opère, principalement, en trois phases à savoir : La structuration du modèle, son exploitation et la concrétisation des résultats.

La figure 3.1 illustre, en détails, le modèle décisionnel adapté par le système FCM-AT.

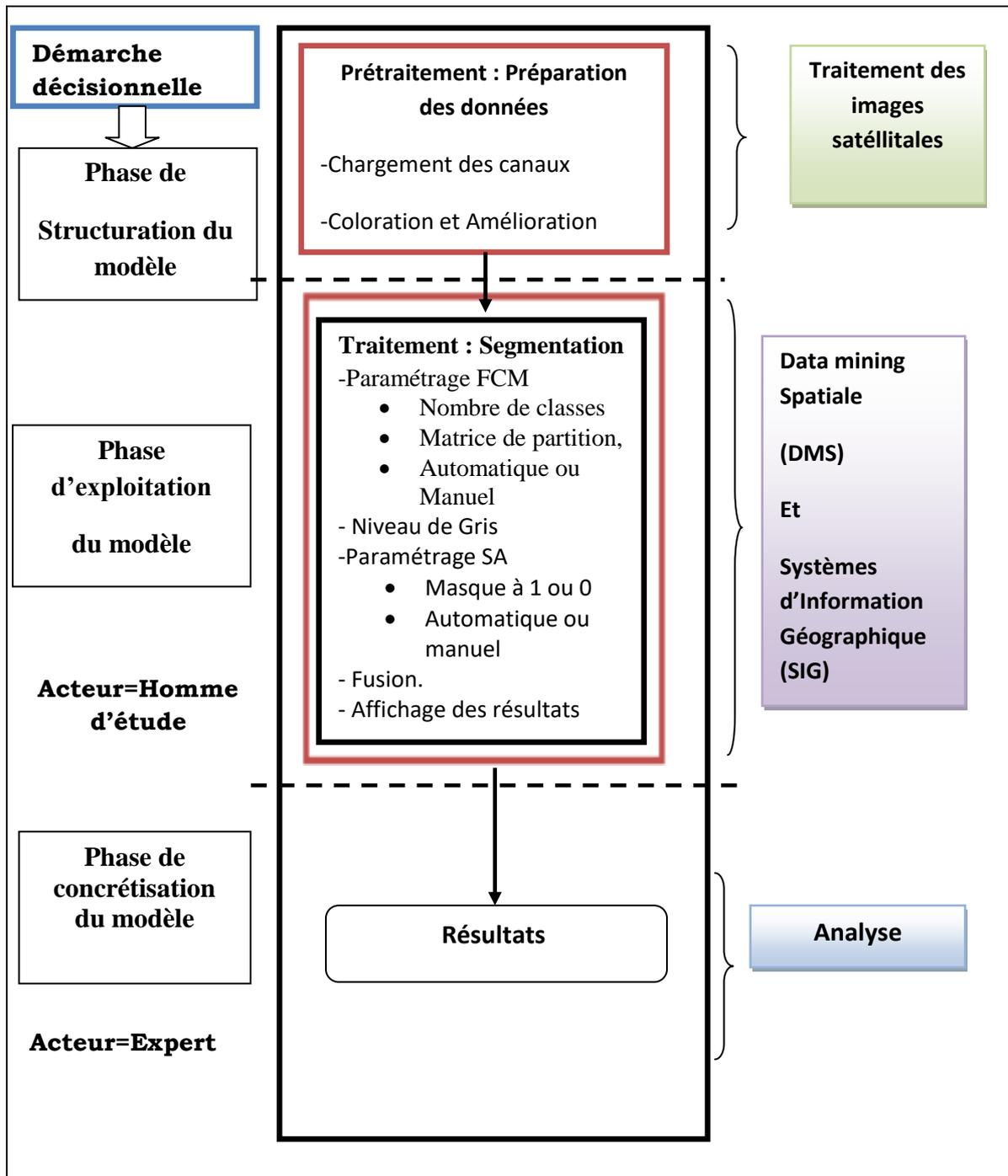


Fig 3.1 : Le modèle décisionnel

5.1. La phase de structuration du modèle

Elle a pour objectif l'identification du problème et les choix fondamentaux sur la manière de l'aborder. Le prétraitement est une étape très sensible dans toute démarche d'aide à la décision, en particulier lorsque la méthode d'analyse procède par fusion de méthodes, il est très important d'identifier l'ensemble des paramètres qu'on va utiliser, et fixer quelques objectifs. Il faut tenir compte qu'on est dans le non supervisé, donc on sait pas réellement ce qu'on recherche, et notre system entre aussi dans le cadre de la fouille de donnée complexe, car on traite sur des données complexes (images Multi-Spectral), la structuration consiste à préparer les images, On identifie aussi les acteurs, on peut constater qu'il ya deux acteurs principaux, le premier est l'homme d'étude, il prend en charge la structuration et 'exploitation du système, le deuxième c'est le décideur, il fait la concrétisation du système.

Cette étape se constitue de plusieurs tâches :

- Formulation du problème : fixer des buts initiaux.
- Choix de régions : on utilise une image multi spectral Landsat Tm d'Oran Algérie, et découpage de l'image *LANDSAT 5 TM*, pour la région d'Oran (Algérie), elle contient plusieurs régions, et pour la découper on a besoin d'un expert.
- Chargement des canaux 1,3, 5 : chaque canal contient des valeurs radiométriques.
- Coloration et Amélioration : affectation des couleurs (R (rouge), V (vert), B (bleu)) aux canaux (1,3,5) (une couleur pour chaque canal), et amélioration de l'image.

5.2. La phase d'exploitation du modèle

C'est la partie la plus analytique du processus d'étude. C'est là ou toutes les méthodes de fouille de données spatiales, et qui font notre approche seront utilisés, Il s'agit essentiellement du FCM et le Seuillage Adaptative.

Notre system est dédié à des expert dans les systèmes d'informations géographique et du spatial, l'objectif visé par l'exploitation du système est de mettre à leur disposition un système simple, souple, robuste, et efficace qui a pour objectif l'exploitation des méthodes FCM, et SA, puis la fusion de ses méthodes.

Et doit inclure les étapes suivantes :

- Affichage de l'image coloré et prête au traitement
- Paramétrage FCM
 - Nombre de classes
 - Matrice de partition,

- Automatique ou Manuel
- Niveau de Gris
- Paramétrage (AT)
 - Masque à 1 ou 0
 - Automatique ou manuel
- Fusion : on peut constater quatre types de fusion de bases
 - FCM Automatique - Seuillage Adaptative Automatique
 - FCM Automatique - Seuillage Adaptative Manuel
 - FCM Manuel - Seuillage Adaptative Automatique
 - FCM Manuel - Seuillage Adaptative Manuel
- Affichage des résultats d'une façon qu'on peut comparer entre les différentes possibilités de fusion.
 - Sélection des zones qu'on doit optimiser
 - Insertion manuel ou automatique
 - Lancement du processus d'optimisation
 - Affichage des solutions voisines

➤ La concrétisation des résultats

Elle vise essentiellement l'acceptation sociale du résultat. Cependant, elle comprend aussi la mise en œuvre de la décision ainsi que son contrôle.

➤ Acteurs

Le concept d'acteur se réfère à une entité concrète, localisée (dans un contexte) ; c'est une unité d'action et de décision, individuelle ou collective, à laquelle on peut attribuer des ressources, une finalité et une stratégie. Dans FCM-AT on peut identifier deux types d'acteurs: Le premier c'est **l'homme d'étude**, il a pour rôle de faire la structuration et l'exploitation du modèle en général c'est l'informaticien, et le deuxième acteur c'est **l'Expert**, il fait la concrétisation du modèle, la mise en œuvre et le contrôle de l'application, et de tester et valider ou non les résultats obtenus.

5.3. Les Outils d'investigation pour l'élaboration de FCM-AT

La puissance du système d'aide à la décision FCM-AT réside dans sa pluridisciplinarité. En effet, il exploite (voir figure 3.2.):

- Le Data Mining Spatial : pour extraire les connaissances à partir des données, ici on utilise le Datamining Spatial car nos données sont des images multi spectral, on

peut dire aussi, que c'est de la fouille de données complexe, car on traite sur des données complexes (images).

- Traitement d'images satellitaires : avec toutes une panoplie de techniques, de chargement d'images, colorations, amélioration, détection de régions, et de contours, jusqu'à la segmentation, pour en final afficher les résultats.
- Le SIG : s'inscrit comme outil d'analyse et de décision, donner des résultats (cartes colorés), les expliquer avec des thèmes.

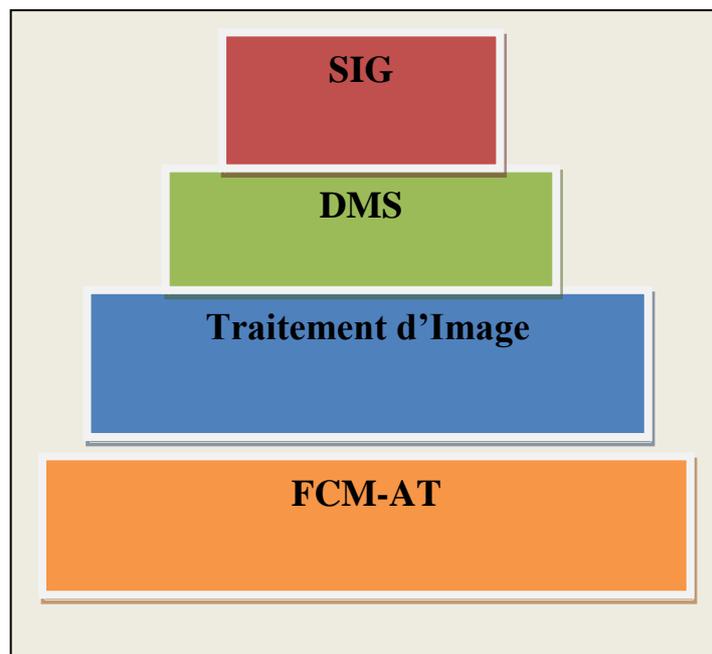


Fig 3.2 : Outils d'Investigations

5.4. Le système FCM-AT

La figure 3.3 est un organigramme général du système décisionnel FCM-AT

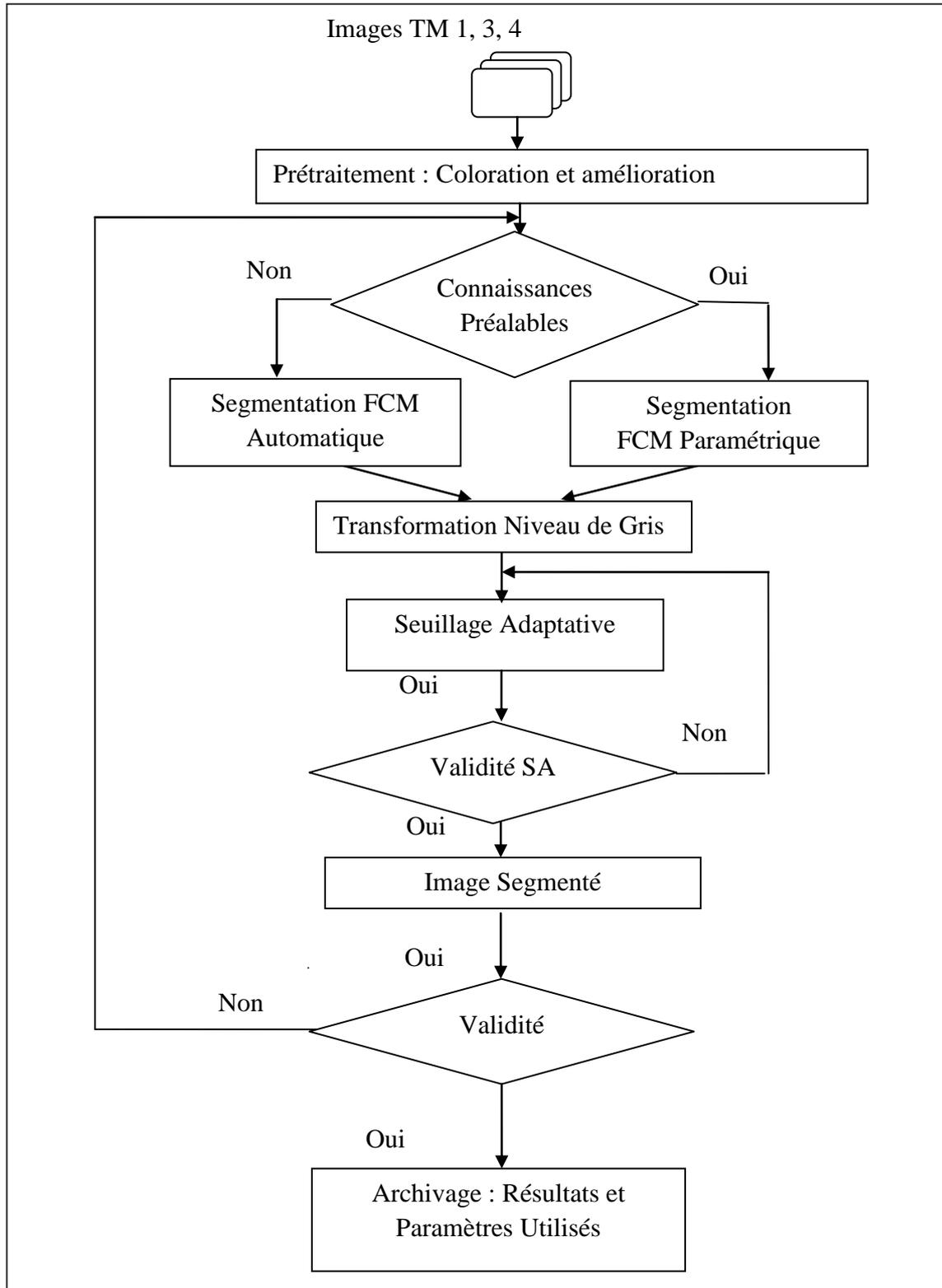


Fig 3.3 : Système Décisionnel FCM-AT

5.5. Diagramme de cas d'utilisation du système proposé :

Le diagramme de cas d'utilisation est une réflexion sur les fonctionnalités attendues du système avant la conception, il permet de fixer une idée sur les principaux modules du système et voir les fonctionnalités que doit fournir chaque composant, Ces fonctionnalités vont aider les utilisateurs à effectuer leurs missions.

Dans le contexte de notre application, nous identifions 7 phases d'utilisations illustrées par la Figure 3.4 :

- **Phase 1** : Chargement des valeurs radiométriques des 3 canaux du satellite (TM1, TM2, TM3).
- **Phase 2** : Coloration et affichage de l'image.
- **Phase 3** : Traitement Histogrammes (affichage Histogramme, Egalisations, étalement, Expansion Dynamique ...).
- **Phase 4** : Débruitage et amélioration par l'application des Filtres (Linéaires gaussien, Moyenner, Non Linéaires médian ...).
- **Phase 5** : Détection de Contours (par Sobel, Laplacien, Robert, Newton, Delewit, ...).
- **Phase 6** : Segmentation de l'image par la méthode FCM (Choisir nombre de classe, degrés d'appartenance).
- **Phase 7** : Optimisation.

Et enfin L'affichage de résultat.

Systèmes de Classification des Images Satellitaires

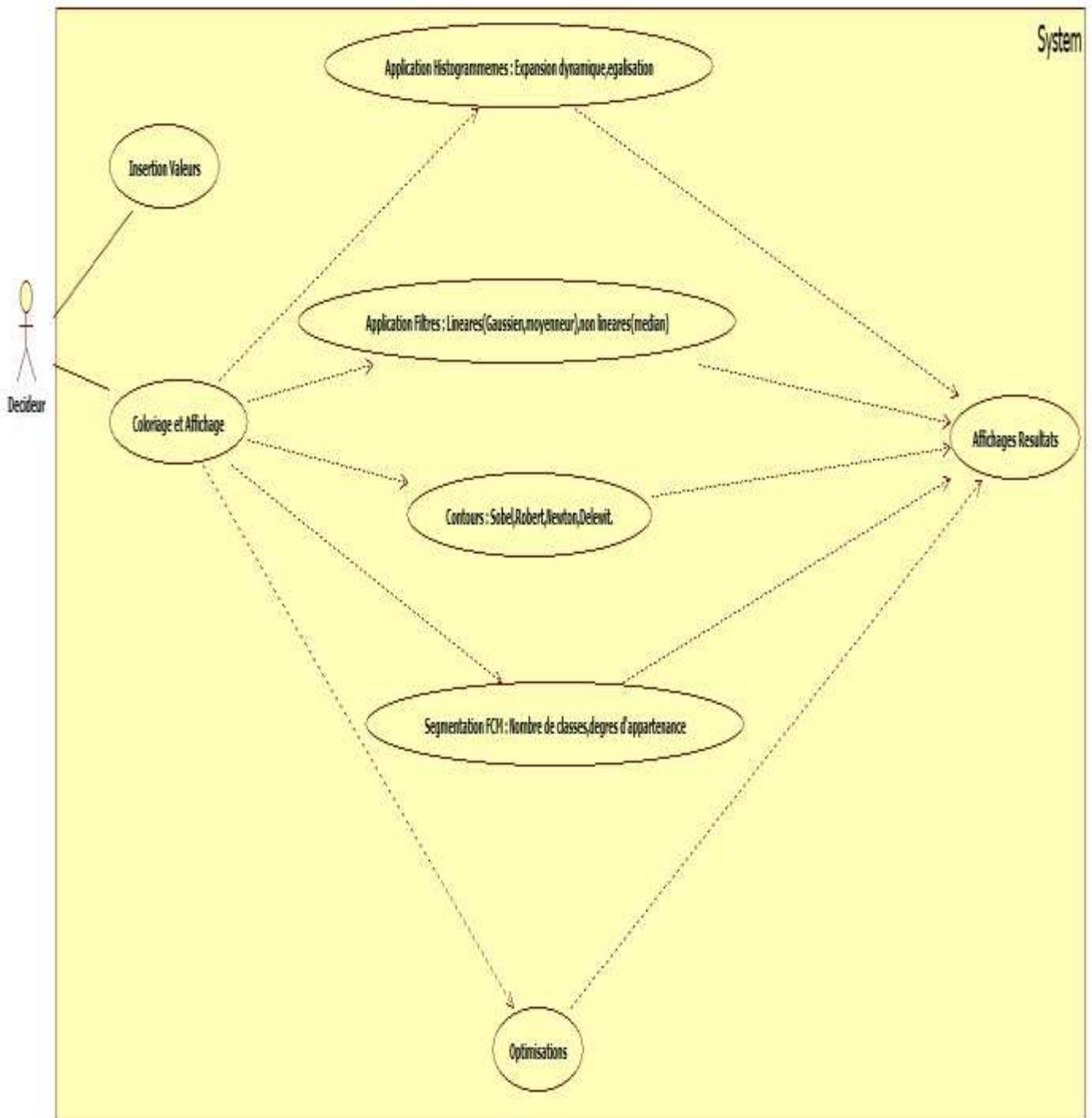


Fig 3.4 :. Diagramme UML de Cas d'utilisation

5.6. Diagramme de Séquences

Le diagrammes de Séquences permettant de représenter des collaborations entre les objets selon un point de vue temporel, on y met l'accent sur la chronologie des requêtes envoyées. Le diagramme de Séquences associé à la démarche décisionnelle adoptée par notre application est illustré par la (Figure 3.5)

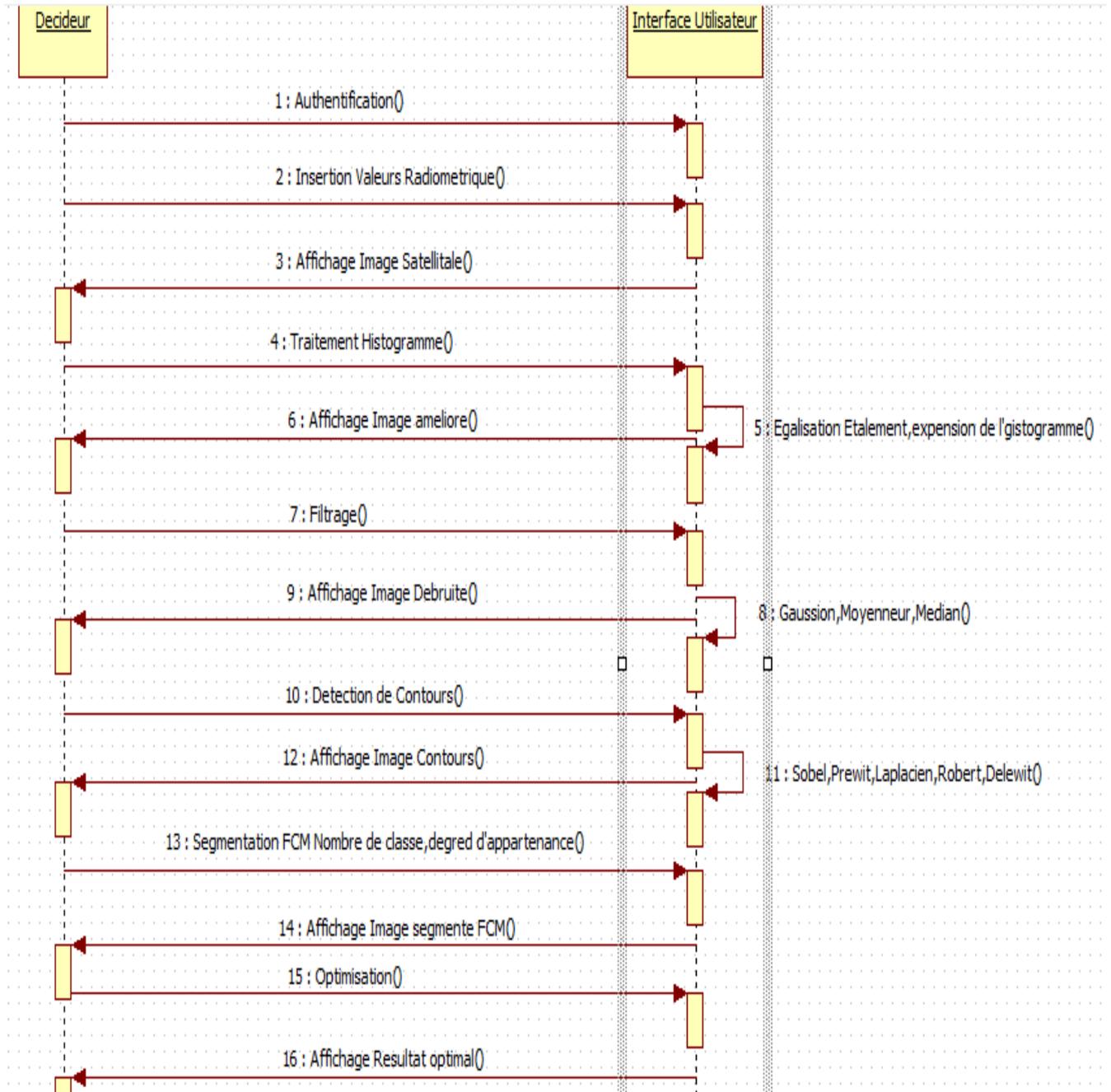


Fig 3.5 : Diagramme de Séquences

6. K-moyennes Floues (Fuzzy C-Means) FCM

➤ Définitions

Un ensemble flou est un ensemble dont les bords sont mal définis. Cela se traduit par une fonction d'appartenance comprise dans l'intervalle $[0, 1]$, alors que dans le cas classique, la fonction d'appartenance prend deux valeurs 0 ou 1 [94]. Un exemple d'ensemble flou est celui contenant des nombres réels "proches" de 3. Ainsi, 3 appartient complètement à cet ensemble et il a, par conséquent, un degré 1 et plus on s'éloigne de 3, plus le degré d'appartenance diminue (figure 3.6).

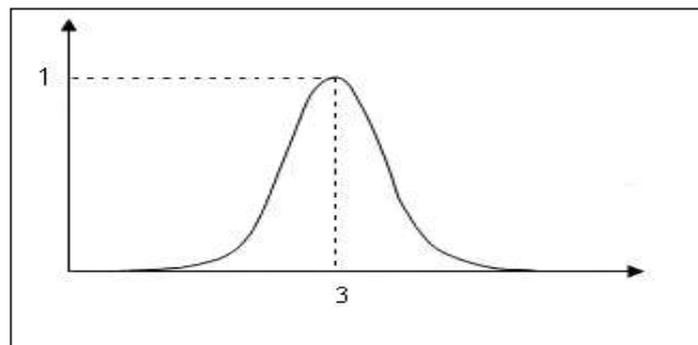


Fig 3.6 : Exemple d'ensemble flou avec $x = 3$

Cette propriété est exploitée en traitement d'images, plus précisément en classification où les classes, appelées aussi régions, sont représentées par des ensembles flous. Cela est fort utile lorsque les régions ne peuvent pas être définies de manière nette et précise. Leur manipulation, en gardant le caractère flou, permet de traiter des données imprécises, incertaines et/ou redondantes d'une manière plus flexible [95]. Dans ce qui suit nous présentons un algorithme de classification basé sur la logique floue, connu et utilisé pour son efficacité et sa robustesse.

L'algorithme des C- moyennes floues (*fuzzy C-Means*) est un algorithme de classification non supervisée flou issu de l'algorithme C-moyennes (*C-means*). L'algorithme introduit la notion d'ensemble flou dans la définition des classes. Chacun des points de l'ensemble des données appartient à une classe avec un certain degré d'appartenance, et toutes les classes sont caractérisées par leurs centres de gravité. L'algorithme est basée sur l'optimisation d'un critère quadratique, du type des moindres carrés, exige au préalable la connaissance du nombre de classes et génère les classes par un processus itératif en minimisant une fonction objective [96].

L'algorithme utilise l'ensemble des pixels $A = \{x_1, x_2, x_3, \dots, x_n\}$ où x_i est un vecteur de trois composantes rouge (R), vert (V), bleu (B), et le nombre de région K . Les valeurs des degrés d'appartenance sont regroupées dans une matrice $U = [u_{ik}]$ où chaque élément désigne le degré d'appartenance du pixel i à la classe k [97].

Pour avoir une bonne partition, on impose aux éléments de U les contraintes suivantes :

- ✓ $u_{ik} \in [0,1]$;
- ✓ $\sum_k u_{ik} = 1$; Ceci $\forall i$;

L'algorithme du FCM fait évoluer la partition (Matrice U) en minimisant la fonction objectif $J_m(U, C)$ suivante appelée aussi fonction de coût [98]:

$$J_m(U, C) = \sum_i \sum_k (u_{ik})^m \cdot \|x_i - c_k\|^2 \dots\dots\dots(13)$$

Avec :

- ✓ $m > 1$: est un paramètre contrôlant le degré de flou appelé le facteur de flou.
- ✓ C_k : Le centre de la classe k .
- ✓ $\| \|$: Une distance quelconque entre pixel i et le centre C_k .

a. Etapes de l'algorithme FCM

- 1- Choisir le nombre de classes K .
- 2- Initialiser la matrice de partition U_i d'une manière aléatoire et les centres C_k .
- 3- Faire évoluer la matrice de partition et les centres.

On montre que la minimisation de la fonction de coût s'effectue lorsque les conditions d'optimalité de Kuhn-Tucker suivantes sont vérifiées [99]:

$$U_{ik} = 1 / \left(\sum_{j=1,c} (d_{ik} / d_{ij})^{(2/(m-1))} \right) \dots\dots\dots(14)$$

Cette condition fait la mise à jour des degrés d'appartenance où : $d_{ij} = \|x_i - c_j\|$

$$C_k = \left(\sum_i (u_{ik})^m \cdot x_i \right) / \left(\sum_i (u_{ik})^m \right) \dots\dots\dots(15)$$

$$4\text{-Test d'arrêt : } |j^{(t+1)} - j^{(t)}| < \textit{seuil} \dots\dots\dots(16)$$

La partition optimale est obtenue par des itérations successives jusqu'à la convergence des résultats. Autrement dit après une initialisation des différents paramètres de l'algorithme, notamment, le nombre de classe K , la partition initiale, le facteur de flou " m " et le seuil de convergence désiré [98].

On détermine les centres des nouvelles classes à partir l'équation (15). La position initiale des centres de gravité conditionne le résultat final de telle sorte que les centres de gravité doivent être initialement placés le plus loin possible les uns des autres. Ceci signifie que les objets à l'intérieur de chaque classe soient aussi proches que possible les uns des autres et aussi loin que possible des objets d'autres classes [98]. Ensuite on détermine les degrés d'appartenance des formes à ces classes par l'équation (14).

On répète ces deux phases jusqu'à ce que l'écart entre deux vecteurs centres successifs satisfasse le test d'arrêt.

Remarque. Le choix de m pour le Fuzzy C-means (FCM) joue un rôle très important. L'équation (14) montre que si m se rapproche de 1, le FCM tend à produire des degrés d'appartenance "durs" c'est-à-dire la partition tend à être classique. Par conséquent, la mise à jour basée sur l'équation (14) fait qu'une certaine valeur U_{ik} est trop proche de 1. Cela entraîne que la classe correspondante à cette valeur est mise à jour de façon beaucoup plus importante que les mises à jour faites pour les $(k-1)$ classes restantes. Ce phénomène peut créer une propriété indésirable vu que certains prototypes convergent rapidement vers les clusters alors que d'autres convergent beaucoup moins vite et inversement, lorsque m augmente, la partition devient floue [97].

L'algorithme proposé donne un résultat appréciable mais présente deux inconvénients majeurs : D'une part, il nécessite le choix préalable du nombre K de classes, ce qui rend l'automatisation de la méthode impossible et d'autre part il requière un temps de calcul souvent élevé du fait de sa nature itérative [98].

7. Seuillage Adaptative (Adaptive Thresholding)

Dans cette section, nous étudierons la deuxième approche qui est basée sur le seuillage adaptatif et que nous appliquerons sur les images satellitaires afin de réaliser une segmentation par classification. Cette approche utilise la valeur d'un seuil qui est changé dynamiquement afin de rendre compte de l'intensité locale des différentes zones de l'image d'entrée (en niveau de gris). Elle présente l'avantage de rehausser le contraste et les contours, par contre elle a le désavantage de représenter les grandes surfaces de pixels à un niveau d'intensité constant. La méthode du seuillage adaptatif présente quelques caractéristiques [99] :

- a. L'image de sortie présente une résolution spatiale ainsi qu'un contraste supérieur à l'image en niveau de gris (originale). Le contraste de l'image originale doit être rehaussé bien qu'il y ait une perte d'information au niveau des intensités du fond de l'image.
- b. Bien que l'entropie de l'image de sortie soit relativement faible, l'image binarisée peut être compressée en utilisant les mêmes techniques permettant le stockage de données comme que le texte ou les graphiques alors que les images en niveau de gris ne peuvent pas être compressées par ces techniques.
- c. L'algorithme du seuillage dynamique peut être implémenté en hardware à moindre coût.
- d. Ce même algorithme peut être aussi appliqué à du texte ou à des informations graphiques avec une qualité garantie, pour les images résultantes.

Tout le problème réside dans le choix de la décision, lors de la binarisation, d'affecter la valeur 1 ou 0 aux différents pixels de l'image. Le principe de cette méthode est de déplacer sur toute l'image un masque et de calculer en chaque pixel la valeur $I(i, j)$ comme indiqué sur la figure 3.7. Il y aura donc encore une fois des problèmes pour le traitement des bords.

Trame hexagonale : est un élément structurant que nous déplaçons sur la totalité de l'image.



Fig 3.7 : Trame hexagonal

Etape1 : On applique le masque à 1 sur toute l'image (voir figure 3.8).

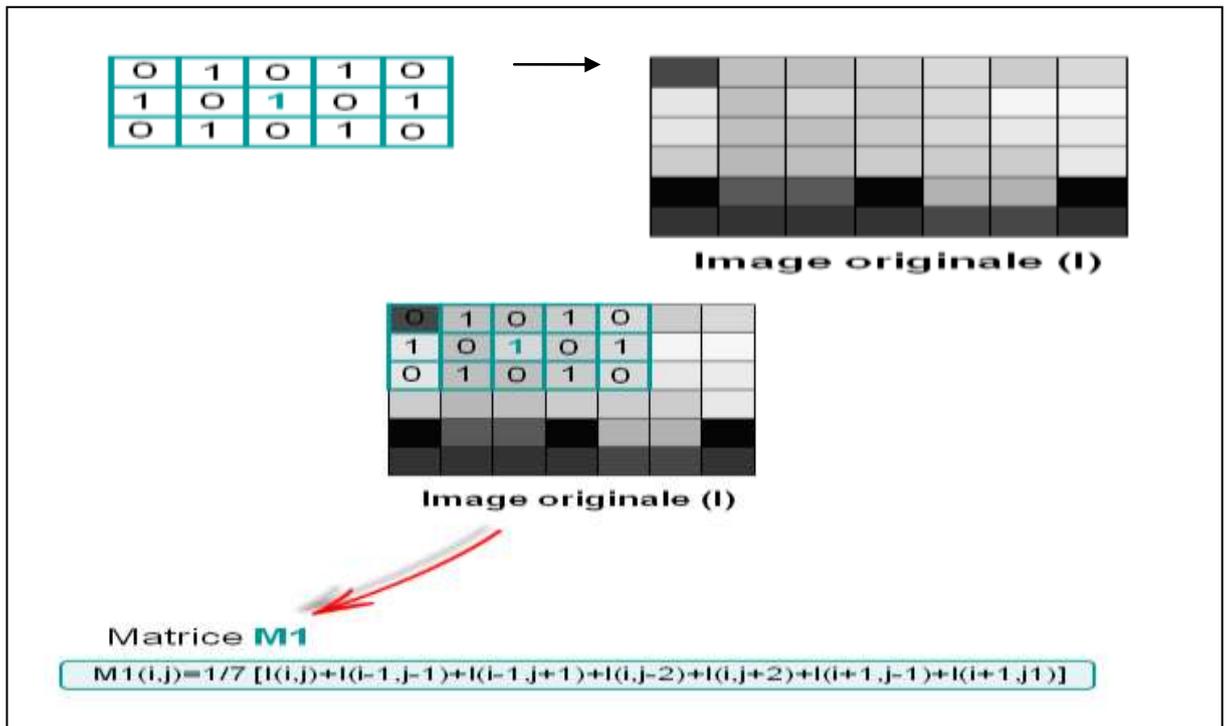


Fig 3.8 : Matrice M1

Etape 2. Une deuxième matrice sera appliquée : (figure 3.9).

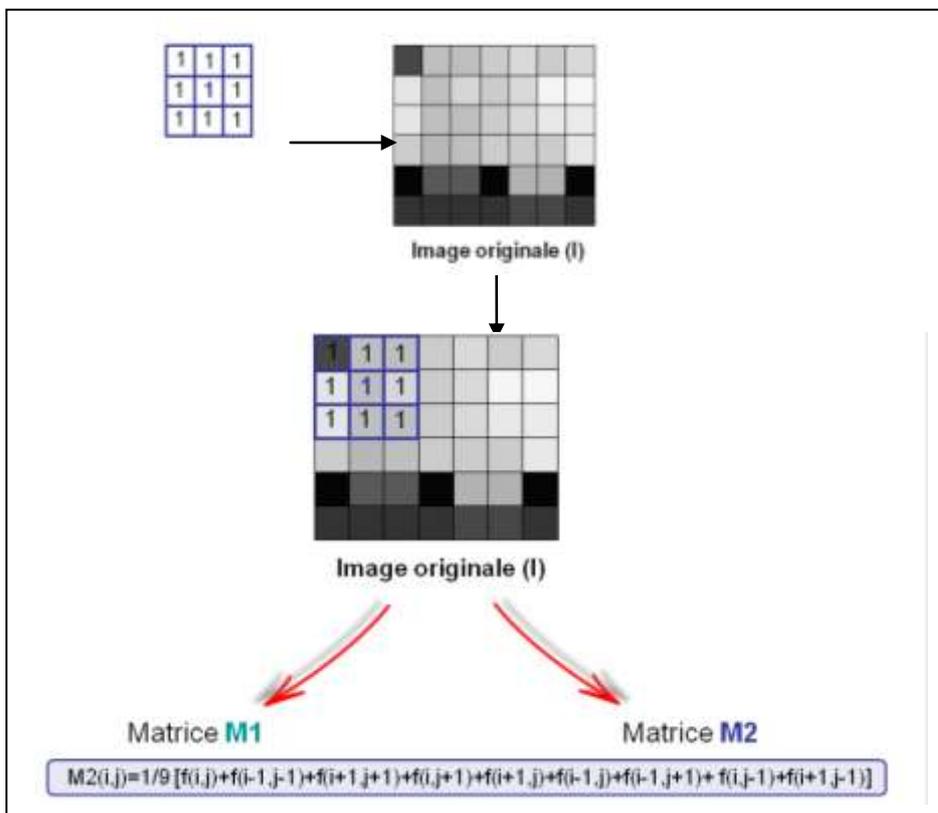


Fig 3.9 : Matrice M2

On obtient :

$$M2(i,j)=1/9[f(i,j)+f(i-1,j-1)+f(i+1,j+1)+f(i,j+1)+f(i+1,j)+f(i-1,j)+f(i-1,j+1)+f(i,j-1)+f(i+1,j-1)]$$

Pour segmenter l'image originale par le seuillage adaptative on effectue l'opération suivante :

Si $I(i,j)-M1(i,j) < I(i,j)-M2(i,j)$ **Alors** $IS(i,j) = 255$ *noir*

Sinon 0 *blanc*.

8. Conclusion

Le SIG s'inscrit de plus en plus comme un outil d'analyse et d'aide à la décision dans les domaines économiques et sociaux.

Le SIG est accessible à tous et offre des services, hier jugés futiles, aujourd'hui indispensables dans la vie quotidienne au travers des médias courants : Internet, téléphonie, navigation embarquée...

Dans ce chapitre nous avons proposé un Système Décisionnel FCM-AT fusionnant deux méthodes de traitement d'images satellitaires. La première est basée utilise le Fuzzy C-Means et est dédiée aux zones uniformes tandis que la deuxième basée sur le seuillage adaptative est applicable zones texturées (pour délimiter frontières de chaque région). Notre approche sera appliquée à une étude de cas avec une étude de performances de l'algorithme proposé, tout cela dans le chapitre 5 consacré à la mise en œuvre de notre approche.

L'exploitation et la lisibilité du système est intimement liée à la capacité de l'homme à l'interpréter. Simplifier la représentation de l'information dans une approche métier, en améliorer la précision et la rendre accessible à tous est un des enjeux de demain au-delà de la fonctionnalité même proposée par FCM-AT.

Chapitre 4

Mise en Œuvre

1. Introduction

Le système de classification d'images satellitaires vise à faciliter la prise de décision du fonctionnaire au niveau de l'entreprise ou ailleurs. Un système de classification d'images satellite bien conçu est un programme interactif qui aide les décideurs à extraire des informations utiles à partir de données spatiales, de connaissances personnelles et de modèles commerciaux, afin d'identifier, de résoudre et de prendre des mesures. Les décisions.

Ce chapitre est consacré à notre contribution au développement d'un modèle de prise de décision basé sur la fusion de FCM-AT (FCM plus AT), qui peut aider les décideurs dans plusieurs domaines à réaliser des projets de développement.

Ce chapitre présente un résumé de tous les résultats obtenus par l'application du système proposé sur des images satellitaires réelles de la ville d'Oran captées par la satellite Landsat.

2. Description de la zone d'étude

Les procédures utilisées dans l'approche de segmentation proposée ont été testées sur des données numérisées des images satellitaires de type LANDSAT – TM (Figure 4.1).

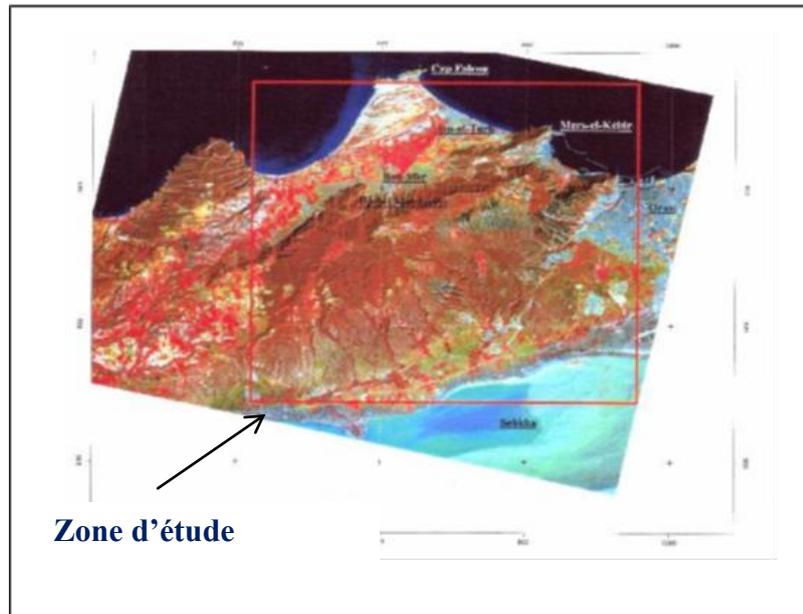


Fig 4.1 : Délimitation de la zone d'étude

3. Exploitation des données

Tout d'abord, nous commençons à segmenter l'image d'origine avec les FCM. Comme les FCM sont appliqués uniquement aux images multispectrales, nous utiliserons les trois canaux TM 1,3,4 de LANDSAT illustrés sur la figure (4.2), qui sont des fichiers contenant des valeurs de mesure radio pour chaque pixel de l'image.

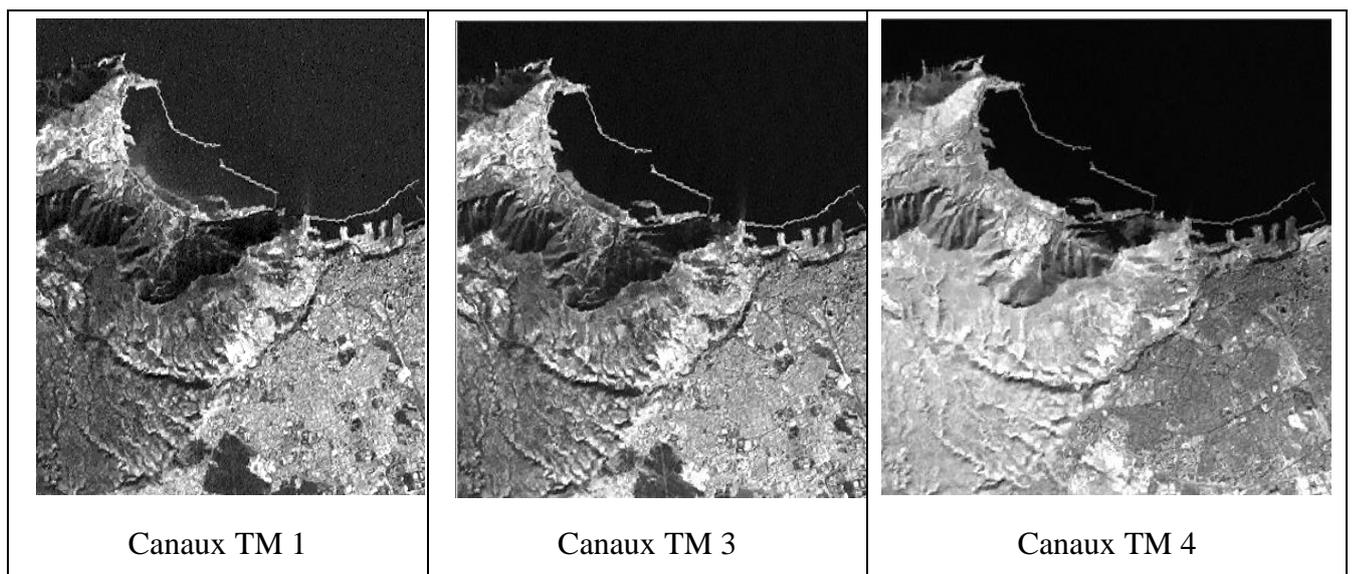


Fig.4.2 : Les trois canaux TM 1, 3, 4

4. Les Résultats

Il y a quatre étapes de base que nous devons suivre dans l'application:

4.1 Chargement des images

Nous chargeons les trois canaux qui composent l'image nécessaire à notre application.

4.2. Colorisation et raffinement

Les composantes RVB (coloration) sont obtenues en obtenant la scène à travers trois filtres: rouge, vert et bleu. Le filtre bleu correspond à une couleur bleue plus foncée. Les filtres vert et rouge permettent des effets de contraste en imagerie industrielle et fine. Le filtre rouge est également utilisé pour corriger la brume atmosphérique en photographie aérienne et pour éliminer le bleu en photographie infrarouge. Le résultat final est un uniforme trichromatique appelé composition de couleurs où les traits sont mis en évidence par différentes nuances dans ces trois couleurs primaires.

Nous mettons en œuvre des différentiels dynamiques pour obtenir un contraste maximal sur toute la plage radio métrique qui nous intéresse. Par conséquent, nous parlons d'amélioration du contraste.

Les k-moyennes floues donnent un résultat très efficace pour notre application. Dans notre travail nous avons appliqué les *FCM* sur des images multi spectrales.

Les images utilisées dans nos expériences montrent la région d'Oran (ici par manque de données d'imagerie satellitaire nous nous sommes limités à la région d'Oran), et nous voyons:

- **Test1** : les paramètres utilisés :

a/ $m=2$, on a fixé le nombre de classes $K=4$, et le degré d'appartenance équitable $U_{ik}=0.2$ pour toutes les classes. Les résultats obtenus après segmentation sont illustrés sur la figure (4.3.a.).

b/ On garde les mêmes paramètres que dans **(a)** sauf que le degré d'appartenance est non équitable, la première classe a un *degré* $=0.5$ et les classes restantes $U_{ik}=0.1$. Les résultats obtenus après segmentation sont illustrés sur la (figure 4.3.b).

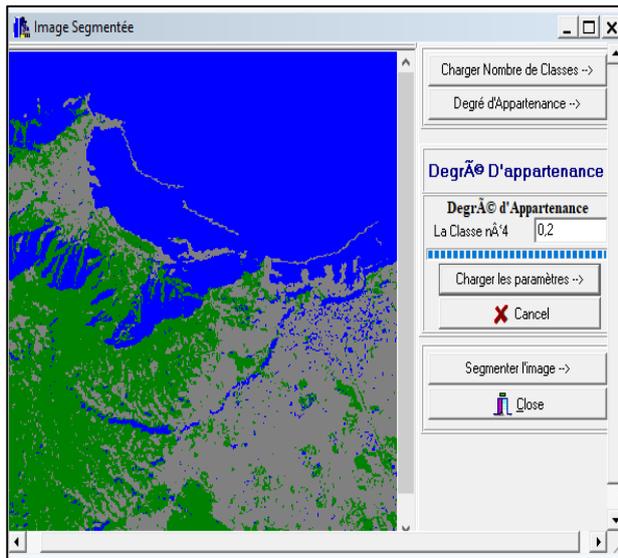


Fig 4.3.a : FCM paramétrique $k=4$ et U_{ik} équitable (Oran)

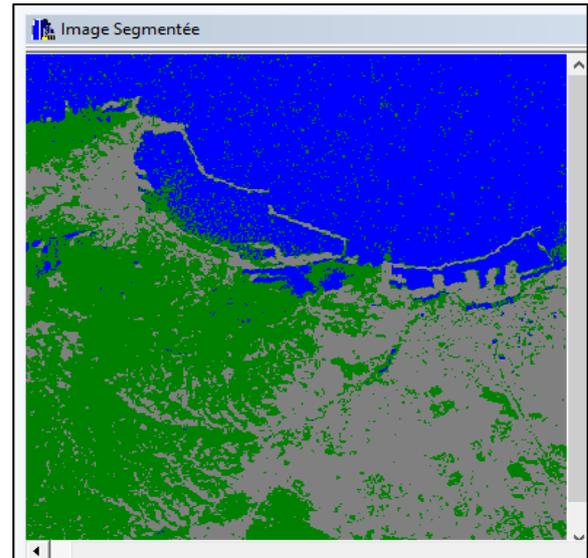


Fig 4.3.b : FCM paramétrique $k=4$ et U_{ik} non équitable (Oran)

- **Test2** : utilisés les paramètres suivants :

a/ $m=2$, on fixe le nombre de classe à $K=6$ et on a pris degré d'appartenance équitable $U_{ik}=0.15$ pour les classes .les résultats obtenus après segmentation sont présentés par la (figure 4.4.a).

b/ On garde les mêmes paramètre que (a) sauf que le degré d'appartenance est non équitable, la première classe a un degré = 0.4 et les classes restantes $U_{ik}=0.1$. Les résultats obtenus après segmentation sont présentés par la (figure 4.4.b).

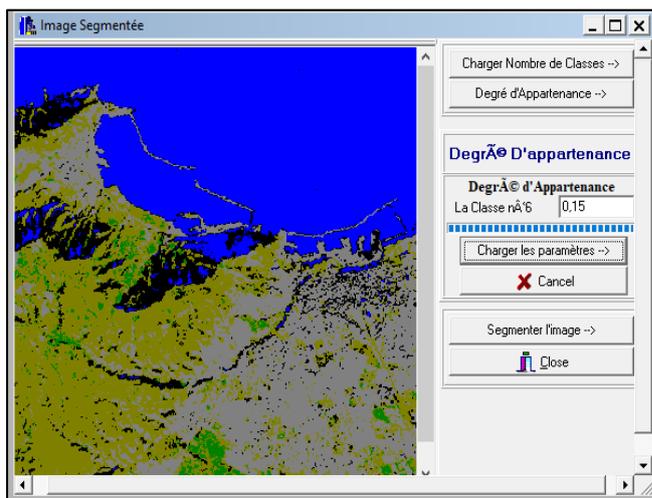


Fig 4.4.a:FCM paramétrique $k=6$ et U_{ik} équitable (Oran)

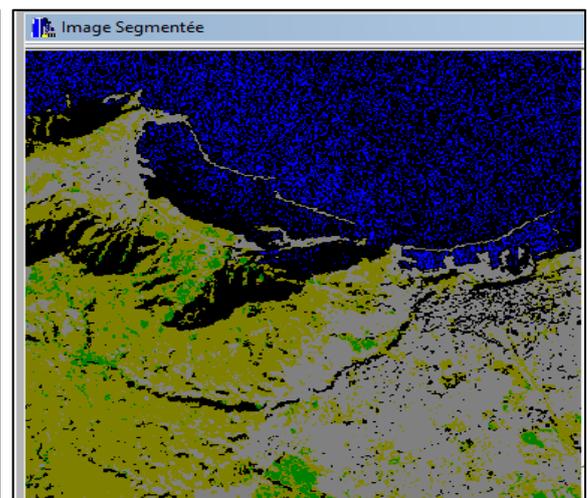


Fig 4.4.b : FCM paramétrique $k=6$ et U_{ik} non équitable (Oran)

- **Test3** : utilisés les paramètres suivants :

a/ $m=2$, on fixe le nombre de classes à $K=10$ et on a pris degré d'appartenance équitable $U_{ik}=0.1$ pour les classes. Les résultats obtenus après segmentation sont présentés par la figure (4.5.a.).

b/ On garde les mêmes paramètre que (a) sauf que le degré d'appartenance est non équitable, la première classe a un *degré* =0.2 et les classes restantes $U_{ik}=0.08$. Les résultats obtenus après segmentation sont présentés par la figure (4.5.b.)

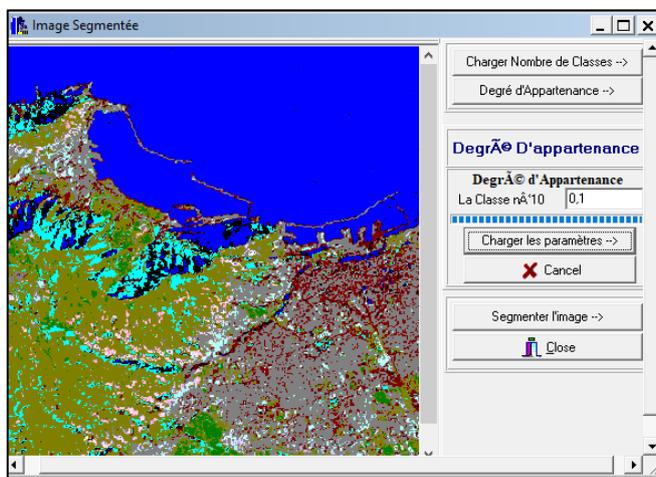


Fig 4.5.a : FCM paramétrique $k=10$ et U_{ik} équitable (Oran)

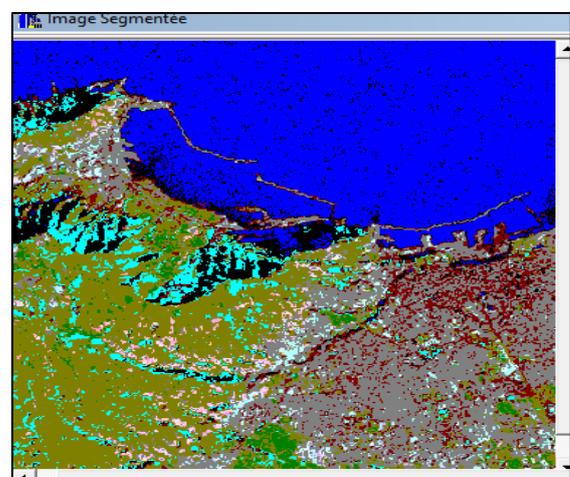


Fig 4.5.b : FCM paramétrique $k=10$ et U_{ik} non équitable (Oran)

Nous confirmons que l'efficacité du FCM est d'autant plus grande que le contraste de luminance est grand, et nous remarquons que plus le nombre des classes sont élevé, plus le FCM est efficace, mais ce processus nécessite un temps de calcul très long.

Sur la base des résultats que nous avons obtenus, nous pouvons confirmer que le choix du degré d'appartenance joue un grand rôle dans le partitionnement des régions. Ce choix pose le problème de la préférence de régions lors de l'attribution des degrés sans connaissance préalable du terrain.

Pour résoudre ce problème, nous avons retravaillé nos relations d'algorithmes en ajoutant les connaissances de l'expert du terrain (spécialiste), afin d'automatiser le traitement. Nous avons constaté qu'en automatisant l'algorithme, nous pouvons nous débarrasser de tous les inconvénients du FCM (étude faite en 2016) [100].

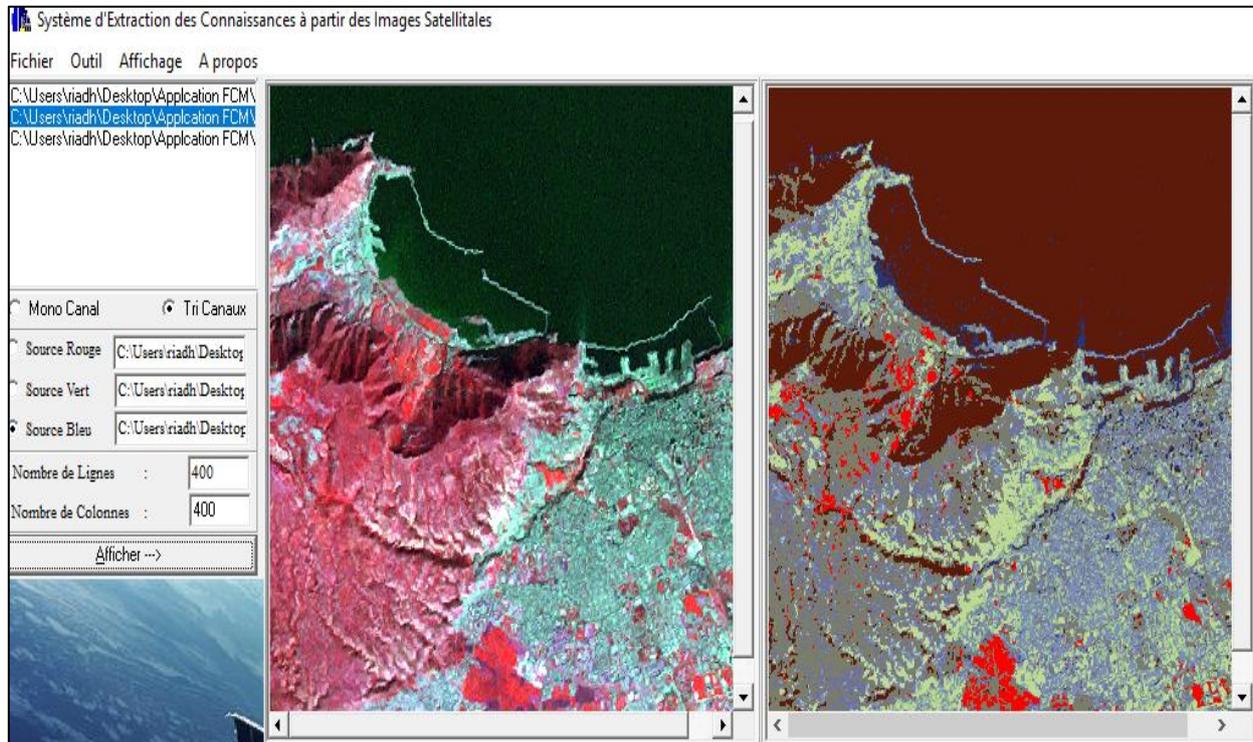


Fig 4.6: FCM automatique (Oran)

- **Test 4 :** les paramètres utilisés:

a/m=2, on fixe le nombre de classe à $K=3$ et on prend le degré d'appartenance équitable $U_{ik}=0.3$ pour les classes. Les résultats obtenus après segmentation sont présentés par la figure (4.7.a).

b/ On garde les mêmes paramètres que dans (a) sauf que le degré d'appartenance est non équitable, la première classe a un *degré* =0.5 et le degré d'appartenance pour les classes restantes est $U_{ik}=0.2$. Les résultats obtenus après segmentation sont présentés par la figure (4.7.b).

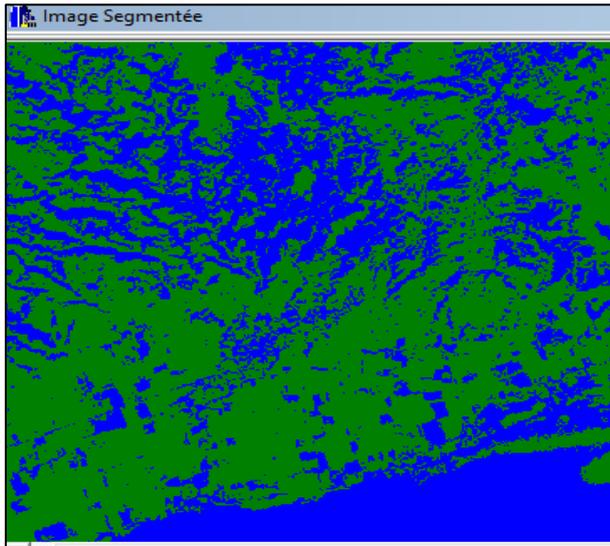


Fig 4.7.a : FCM paramétrique $k=3$
et U_{ik} équitable (Sebkha)

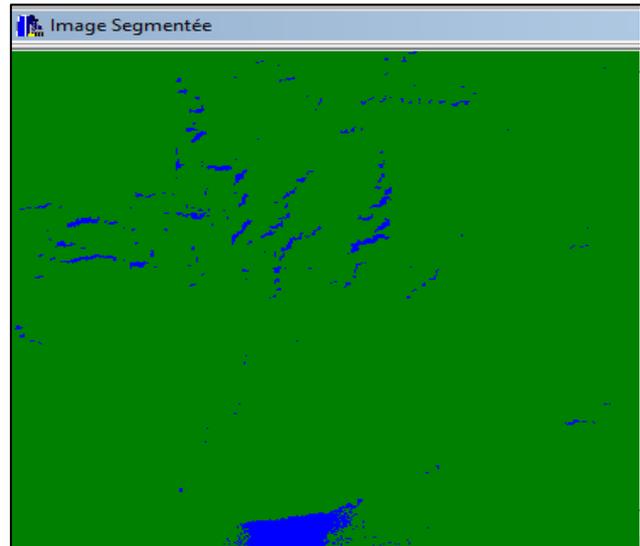


Fig4.7.b : FCM paramétrique $k=3$
et U_{ik} non équitable (Sebkha)

a/ $m=2$, on fixe le nombre de classe à $K=5$ et on a pris degré d'appartenance équitable $U_{ik}=0.15$ pour les classes .les résultats obtenus après segmentation sont présentés dans la figure (4.8.a.).

b/ On garde les mêmes paramètres que dans (a) sauf que le degré d'appartenance est non équitable, la première classe a un *degré* $=0.5$ et pour les classes restantes un $U_{ik}=0.12$. Les résultats obtenus après segmentation sont présentés sur la figure (4.8.b).

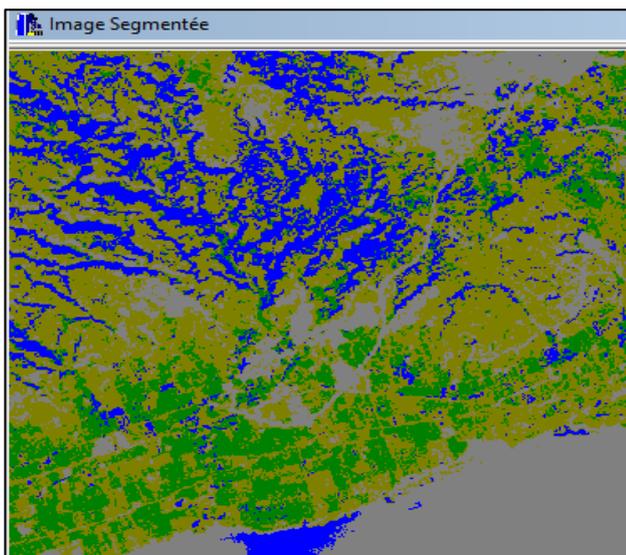


Fig 4.8.a : FCM paramétrique $k=5$
et U_{ik} équitable (Sebkha)

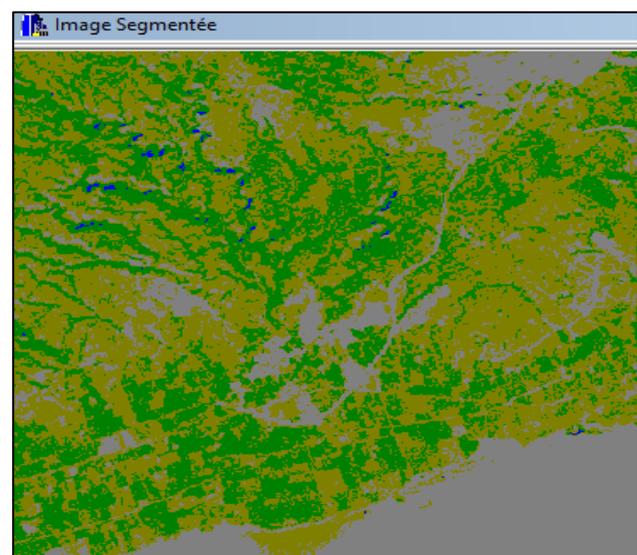


Fig 4.8.b : FCM paramétrique $k=5$
et U_{ik} non équitable (Sebkha)

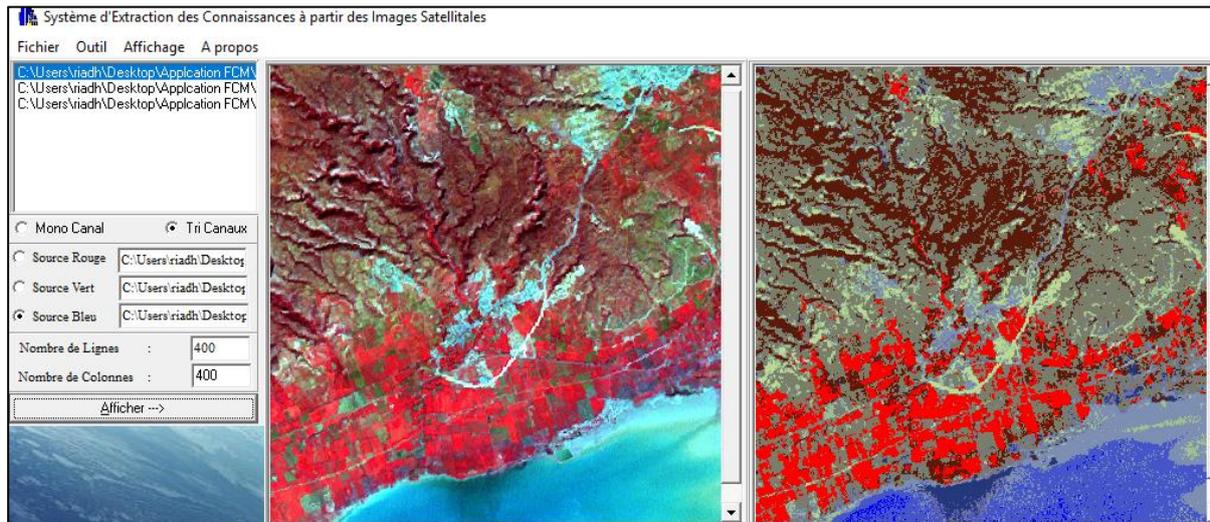


Fig 4.9 : FCM automatique (Sebkha)

4.3. Segmentation par le seuillage adaptatif

Nous effectuons le transfert dans le niveau de gris parce que le seuillage adaptatif est applicable aux images à spectrales unique.

- a. **L'image originale de partie nord d'Oran** : Les deux figures suivantes illustrent les images obtenues en appliquant un seuillage manuel avec un masque central 1 et un masque central 0 et en utilisant des valeurs de seuil différentes, Figure. (4.10) et Figure (4.11).

✓ *En utilisons le seuillage adaptatif manuel avec masque central 1*

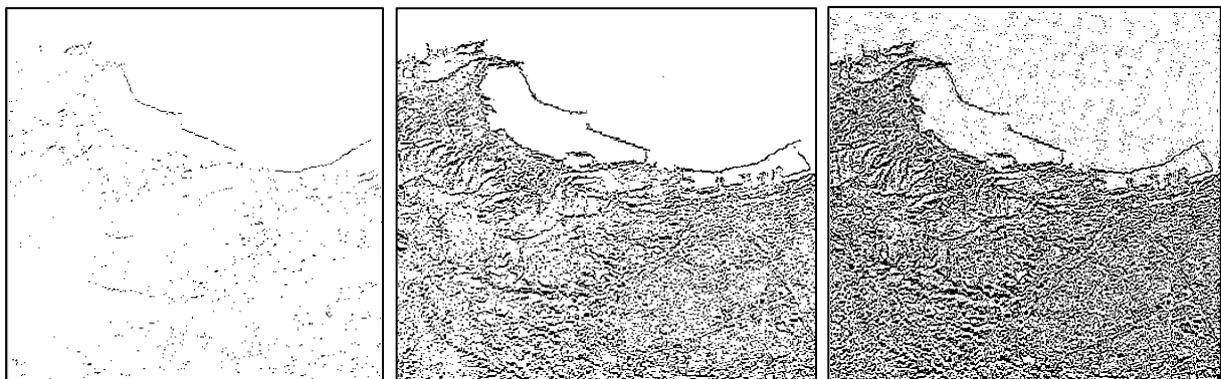


Fig 4.10 : Le seuillage adaptatif manuel avec masque central 1 et seuil= 50 et 10 et 3 (de gauche à droite) (Partie nord d'Oran)

✓ *En utilisons le seuillage adaptatif manuel avec masque central 0*

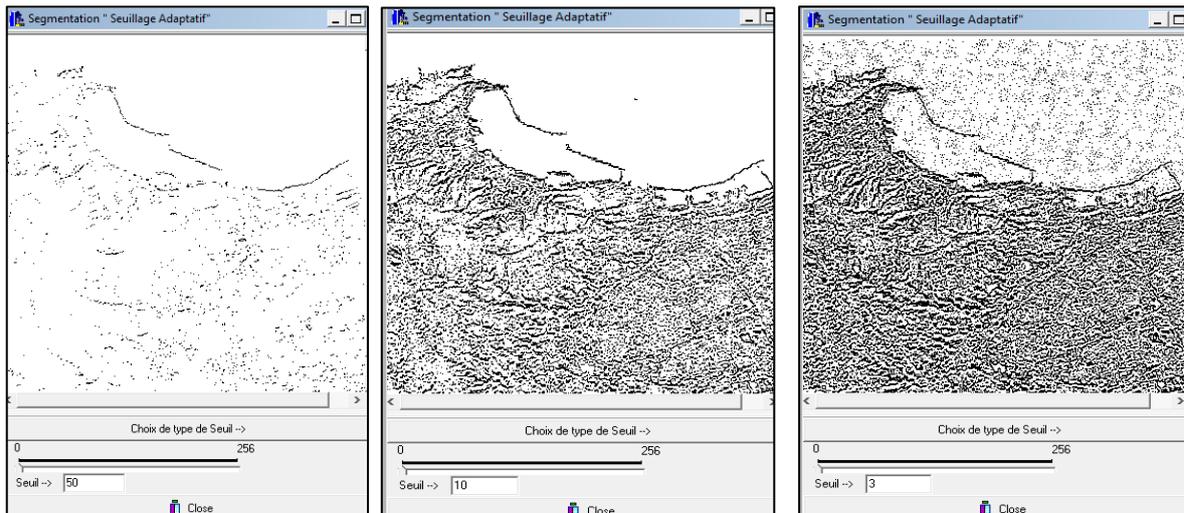


Fig 4.11: Le seuillage adaptatif manuel avec masque central 0 et seuil= 50 et 10 et 3 (de gauche à droite) (Partie nord d'Oran)

Nous constatons que:

- ✓ Plus le seuil du comparateur est bas, plus le bruit est important.
- ✓ Plus le seuil du comparateur est élevé, plus l'annulation du bruit est importante.
- ✓ La trame centrale à 1 est moins sensible au bruit que la trame à 0. Cela est dû au fait que les performances moyennes sont bien meilleures et plus efficaces.

➤ ***Le Seuillage adaptatif automatique***

Il existe une autre méthode plus efficace et précise: un seuillage adaptatif automatique, les résultats sont donnés par la figure (4.12)

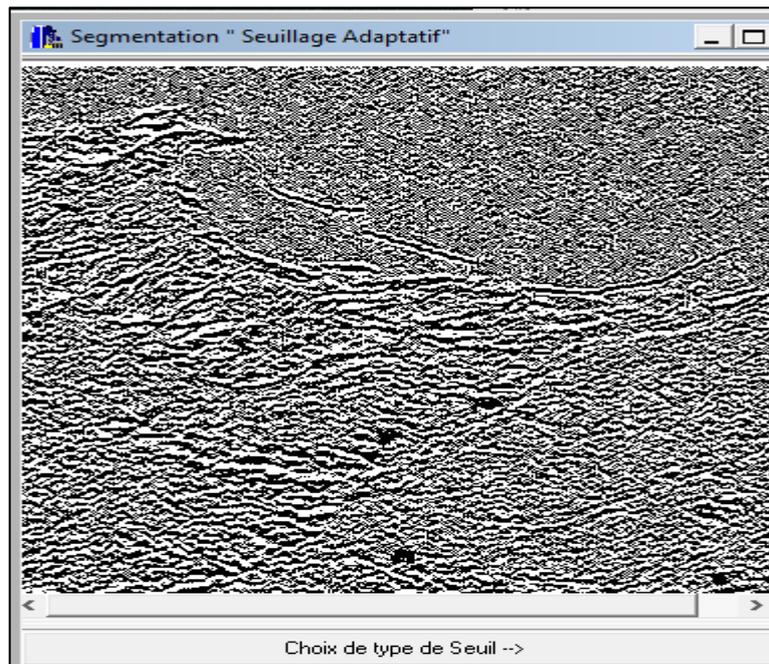


Fig 4.12 : Seuillage adaptatif automatique (Partie nord d’Oran)

b. **L’image originale de partie Oran Sebkh**a : Les deux figures suivantes illustrent les images obtenues en appliquant un seuillage manuel avec un masque central 1 (Figure (4.13)) et un masque central 0, Figure (4.14)) et en utilisant des valeurs de seuillage différentes pour chaque type de masque central.

➤ *En utilisons le seuillage adaptatif manuel avec masque central 1*

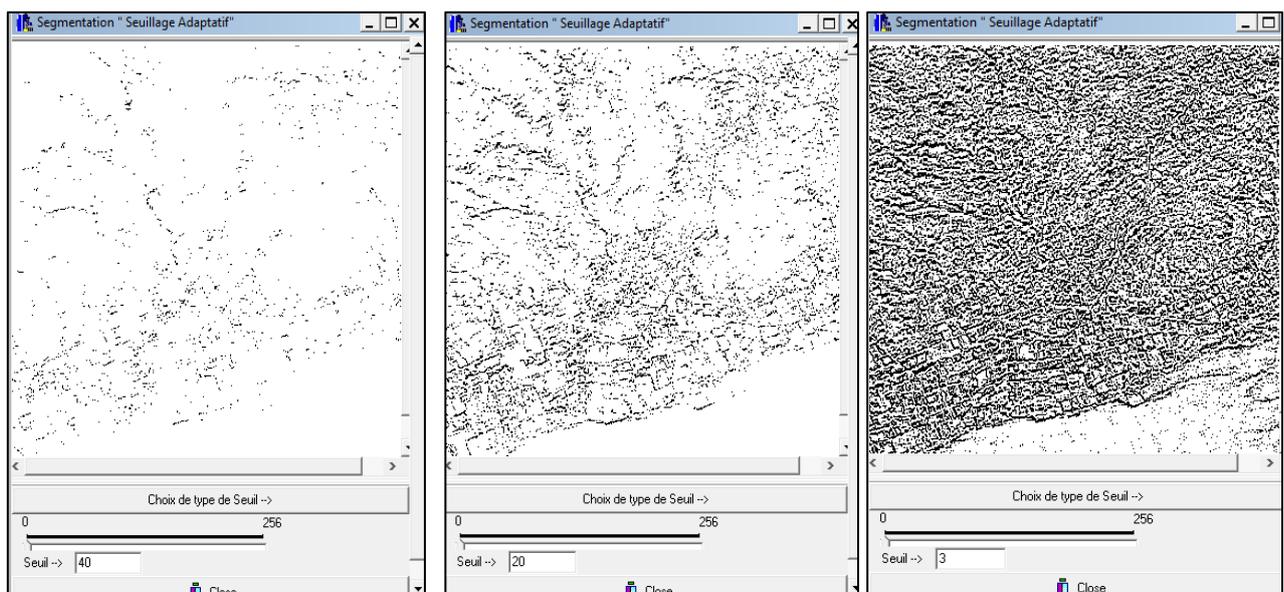


Fig 4.13: Le seuillage adaptatif manuel avec masque central 1 et seuil= 40 et 20 et 3 (de gauche à droite) (Partie Sebkh)a

➤ *En utilisons le seuillage adaptatif manuel avec masque central 0*

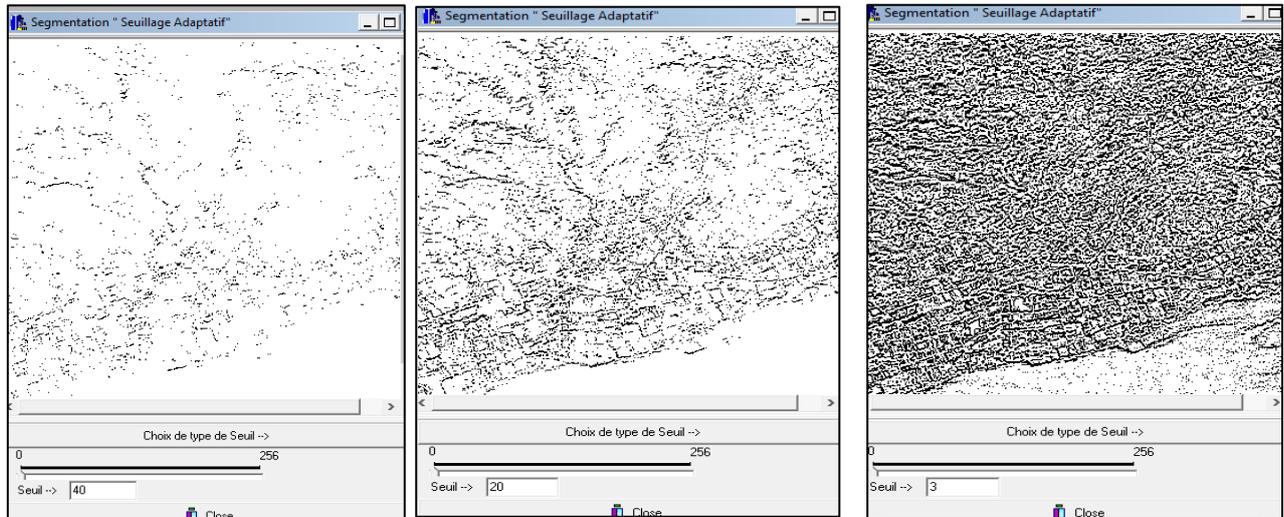


Fig 4.14 : Le seuillage adaptatif manuel avec masque central 0 et seuil = 40 et 20 et 3 (de gauche à droite) (Partie Sebka)

➤ *Le Seuillage adaptatif automatique*

Il existe une autre méthode plus efficace et précise un seuillage adaptatif automatique, les résultats sont donnés par la figure (4.15)

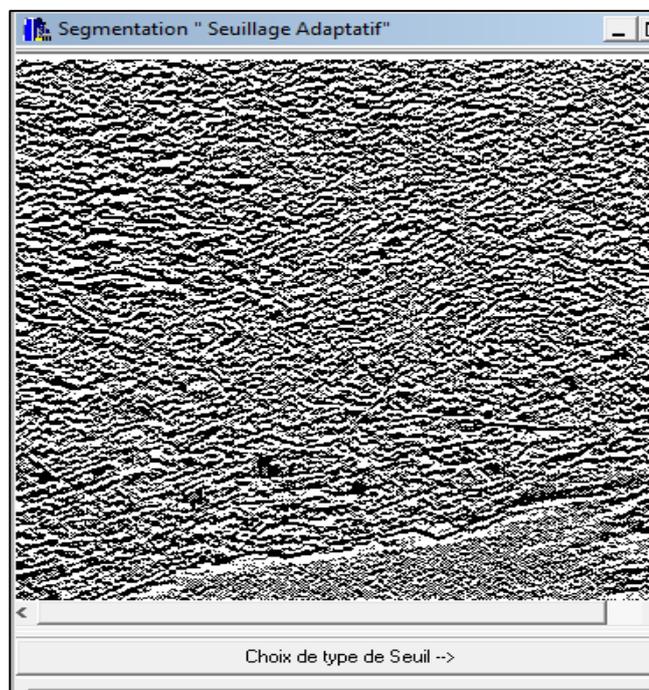


Fig 4.15 : Le Seuillage adaptatif automatique (Partie Oran Sebka)

les résultats obtenus via le seuillage automatique sont plus efficaces que ceux obtenus via le seuillage manuel, quel que soit l'image (Oran nord ou Sebka).

4.4. Fusion (FCM et Seuillage Adaptatif)

La fusion se fait, d'abord par la segmentation de l'image originale par K-moyennes floues, puis l'image résultante est transformée en niveau de gris et enfin segmentée par un seuillage adaptatif.

a. Test 1 : Image originale : Partie Oran nord

Nous avons utilisé la méthode K-moyennes floues paramétriques et les K-moyennes floues automatiques, en utilisant les deux types de seuillage (manuel et automatique).

Les résultats suivants présentés dans les figures (4.16 (a, b)) et (4.17 (a, b)) ont été obtenus en prenant les valeurs suivantes comme paramètres: Le numéro de la classe est égal à 10 et des degrés d'appartenance équitables.

➤ Avec un K-moyennes floue paramétrique

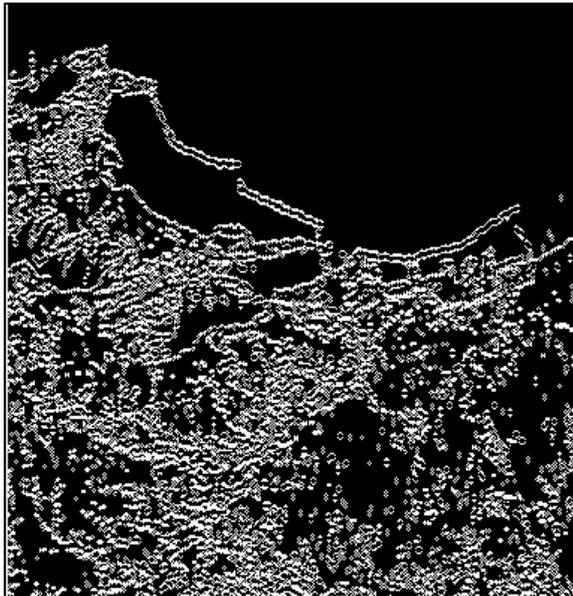


Fig 4.16.a : Fusion de FCM paramétrique et seuillage automatique (Oran) $K=10$ et $U_{ik}=0.08$ $K=10$

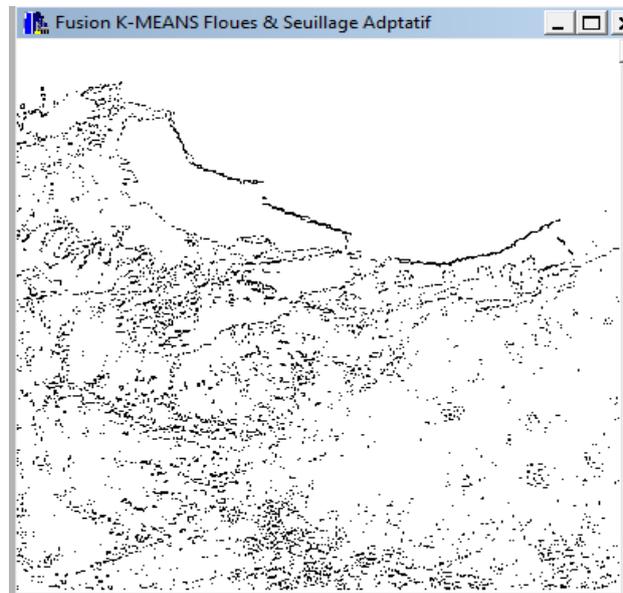


Fig 4.16.b : Fusion de FCM paramétrique et seuillage manuel (Oran) et $U_{ik}=0.08$ avec masque central 0

➤ Avec un K moyennes automatique

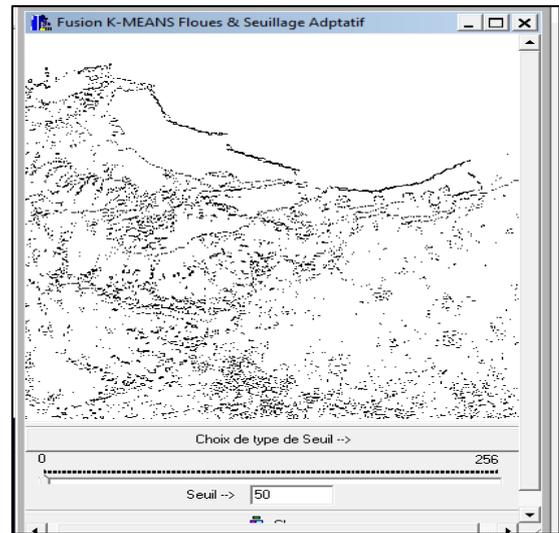
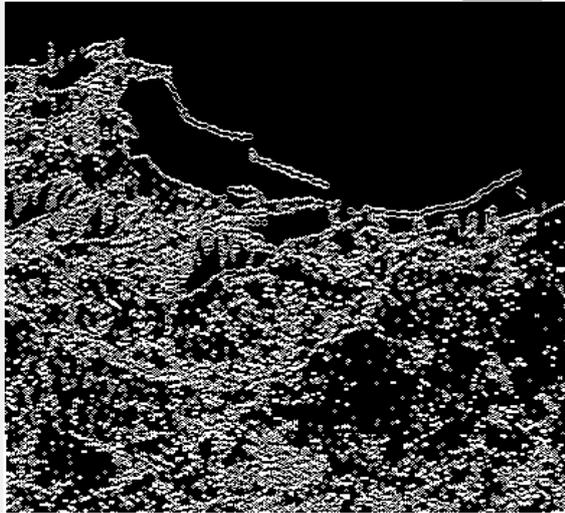


Fig 4.17.a : Fusion de FCM automatique et seuillage automatique (Oran) *Fig 4.17.b* : Fusion de FCM automatique et seuillage manuel (Oran)

a. Test 2 : Image originale : Oran Sebkh

Nous avons refait les mêmes expérimentations que précédemment mais cette fois en utilisant l'image de partie Sebkh et en gardant les mêmes paramètres. Les résultats sont illustrés par les figures (4.18 (a, b)) et les figures (4.19. (a, b)).

➤ Avec un K -means paramétrique

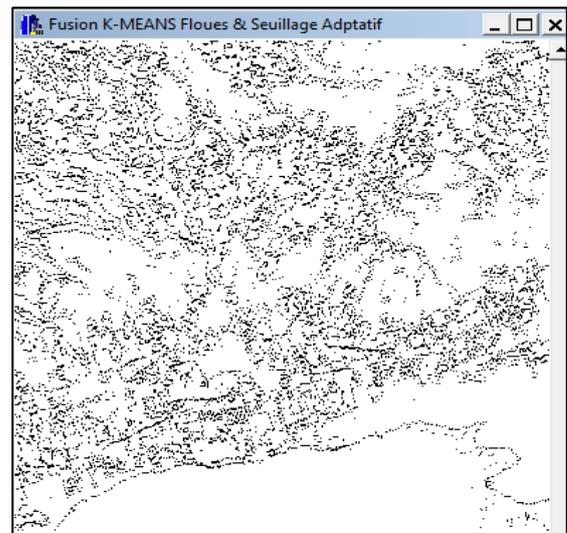
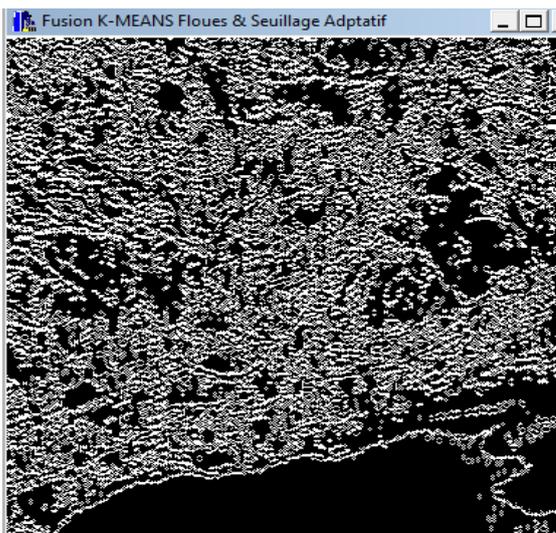


Fig 4.18.a : Fusion de FCM paramétrique et seuillage automatique (Sebkh)
K= 10 et U_{ik} = 0.08

Fig 4.18.b : Fusion de FCM paramétrique et seuillage manuel (Sebkh)
K=10 et U_{ik} = 0.08 avec masque central 0

➤ Avec un *K_means Automatique*

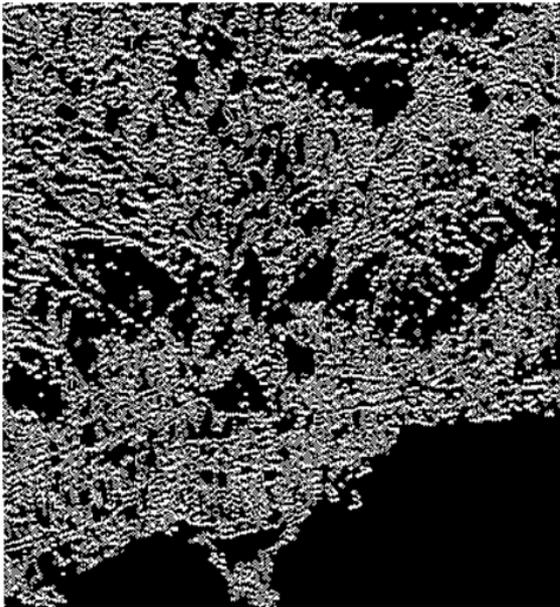


Fig 4.19.a : Fusion de FCM automatique et seuillage automatique (Sebkha)

Fig 4.19.b : Fusion de FCM automatique et seuillage manuel (Sebkha)

5. Conclusion

Dans ce chapitre, nous avons activé la mise en œuvre de notre outil réalisé, appelé *FCM-AT*. Nous l'avons appliqué à une étude de cas consistant en des images satellitaires réelles d'Oran Nord et d'Oran coté Sebkha captées par le satellite *LANDSAT 5*.

D'abord on segmente l'image originale par les K-moyennes floue ensuite on transforme le résultat obtenu en niveau de gris pour être enfin segmenter par le seuillage adaptatif.

Le Système *FCM-AT* est destiné à des utilisateurs connaisseurs et experts (les géographes et spécialistes) ayant de fortes compétences techniques dans les domaines satellitaires et l'exploitation de données graphiques. Cette outil est basé sur une combinaison *FCM-AT* susceptible d'apporter une aide aux décideurs du spatiale dans la réalisation des différents projets d'aménagement.

Les résultats obtenus sont très encourageants et très intéressants et ont montré en conséquence l'efficacité et la robustesse de l'approche proposée.

Conclusion Générale

Dans notre travail, nous avons abordé le problème complexe de l'analyse des scènes satellitaire.

Tout au long de notre étude, nous avons constaté que l'analyse d'image est un domaine très vaste, on peut le partager en deux étapes ; la première, correspond aux différents traitements effectués sur les images pour les améliorer et les préparer à des traitements automatiques ; la deuxième étape est l'ensemble des traitements permettant l'interprétation des images afin de pouvoir identifier les objets espérés parmi les autres objets existants.

L'introduction de la phase de segmentation dans un système de fusion entre deux méthodes FCM-AT complet permet de faciliter l'extraction des objets significatifs dans des images satellitaires. Dans notre document, on utilise les algorithmes de FCM avec AT.

Les résultats obtenus par cette étude sont :

- Présentation de différentes méthodes.
- Implémentation des deux algorithmes très efficaces, reliant une segmentation en région et un étiquetage des composantes connexes.

Les connaissances sont des informations qui concernent les pixels et sont utilisées pour pouvoir l'identifier, dans notre travail les seules informations utilisées pour le regroupement des pixels sont des informations radiométriques :

- Les valeurs de niveau de gris.
- La valeur absolue de la différence de niveau de gris maximum et minimum.
- La variance des niveaux de gris des pixels de la même région.

La qualité de l'interprétation d'une image dépend fortement de celle de la segmentation, malgré la grande diversité de méthodes. Les résultats restent moyens et varient beaucoup en fonction de la technique choisie. Une méthode générale et automatique est difficile à concevoir étant donnés, les différents types de méthodes.

Les résultats obtenus ne sont pas une fin en soit, au contraire ils nous mènent à poser les questions sur d'autres perspectives intéressantes comme, une coopération régions-contours qui met en correspondance nos algorithmes et d'autres.

Nous remarquons que ce qui s'ouvre en perspective est plus large de ce qui a été accompli puisque, l'analyse d'image a dépassé le but qu'elle s'était fixée au début ; ce qui n'était d'autre que de remplacer l'observateur humain par la machine à travers un système d'aide à la décision.

Références Bibliographiques

- [1] : M.ANDRE -Introduction aux techniques de traitements d'images, Eyrolles 1987
- [2] : MICROSOFT- Encyclopédie, ENCARTA 2006
- [3] : M.HADALLAH- Codage des images fixes par une méthode hybride basée sur la QV et les approximations fractales, PFE USTHB 1997
- [4] : R.C.GONZALES P.WINTZ-Digital Image Processing -Addison Wesley 1977
- [5] : M.KUNT-Traitement numérique des images. Vol.2 1993
- [6] : R.M.GRAY- Vector Quantization-IEEE ASSP magazine - Avril 1991
- [7] : K.TABARI S.TAGMA-Compression d'images animées à très faible débit par la géométrie des fractales PFE USTHB 1996
- [8] : Jacky DESACHY Université des Antilles et de la Guyane Pointe à Pitre janvier 2001
- [9] : S.DEKKAR- Compression des images fixes par la méthode des fractales. PFE INI 1996
- [10] : A.d'HARDANCOURT- Fou du multimédia. Sybex 1995
- [11] : CHRISTOPHE FIORIO -Analyse d'image LIRMM UMR CNRS-UMII Novembre2005
- [12] : L.MOKDAD -Compression d'images pour la visiophonie. PFE INI 92
- [13] : Hunt, E. B., Martin, j., Stone P. J., *Experiments In Induction*. Academic Press, 1966.
- [14] : Bouaza Ben Halima Nahal Abdelaziz-Segmentation d'images aériennes et Satéllitaire 2003
- [15] : Mekki Mokhtar Amina Rahim Mehdi Segmentation d'image satellitaire juin 2007

- [16] : Kodratoff, Y., *Technical and scientific issues of KDD, Algorithm Learning Theory*. LNAI,1994.
- [17] : Taleb Zouggar, S., *Apprentissage automatique neuro-symbolique, Une nouvelle technique hybride pour la sélection et la segmentation de données*. Mémoire de Magister, Université d'Oran, Algérie, 2005.
- [18] : Atmani, B., Beldjilali, B., *Knowledge Discovery in Database : Induction Graph and Cellular Automaton*, Computing and Informatics Journal, Vol. 26, N±2, 2007.
- [19] : Fisher, R.A, *The use of multiple measurements in taxonomic problems*. Ann, Eugenics, 1936.
- [20] : Michie, D., Spiegelhalter, D.J., Taylor, C.C., *Machine Learning, Neural and Statistical Classification*. 1994.
- [21] : Preux, P., *Fouille de données, Notes de cours*. Université de Lille, 2005.
- [22] : Cover, T. M., Hart P. E., *Nearset neighbor pattern classification*. IEEE Trans. Info, Theory, 13, pages 21-27, 1967.
- [23] : Brostaux, Y., *Etude du classement par forêts aléatoires d'échantillons perturbés à forte structure d'interaction*. Université Gembloux, Belgique, 2005.
- [24] : Warren McCulloch & Walter Pitts, *A Logical Calculus of Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics 5:115-133, 1943.
- [25] : Lefébure, R., Venturi, G., *Data mining, Gestion de la relation client, Personnalisation des sites web*. Eyrolles, 2001.
- [26] : Hopfield, J.J., *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the National Academy of Sciences USA.
- [27] : Kohonen, T., *Self orgainsation and associtive memories*. Springer series in information sciences. Springer Verlag, 1984.
- [28] : Rumelhart, D. E., Hinton, G. E., Williams, R. J., *Learning Internal Representations by Error Propagation. Parallel Distrubuted Processing*, MIT press, 1, 8, 1986.

- [29] : Atmani, B., Beldjilali, B. : *Réseau de Neurones pour des Systèmes d'Inférences à base de Règles (RNSIR)*. 5ème Conférence Maghrébine sur le Génie Logiciel et l'Intelligence Artificielle. Tunis, 1998.
- [30] : Atmani, B., Beldjilali, B., *Neural Network for the Selection and the Elimination of the Exogenous Variables and the non applicable Individuals*, The Second International Workshop on Advanced Computation For Engineering Applications (ACEA03), Electronics Research Institute (ERI), Cairo, Egypt, December 21-22, 2003.
- [31] : Mejia, C. E., *Architectures Neuronales pour l'approximation des fonctions de transfert, application à la télédetection*. Thèse de Doctorat en Sciences, 1992.
- [32] : Rosenblatt, F., *The perceptron: a probabilistic model for information storage and organisation of the brain*. Psychological Review, 65, pages 386-407, 1958.
- [33] : Widrow, B., Hoff, M. E., *Adaptative Swithing Circuits*. IRE WESCON Convention Record, 4, pages 96-104, 1960.
- [34] : Sonquist, J. A., Morgan, J. N., *The detection of interaction effects*. An Arbor: Institute for social research. University of Michigan, 1963.
- [35] : Hunt, E. B., Martin, j., Stone P. J., *Experiments In Induction*. Academic Press, 1966.
- [36] : Quinlan, J.R., *Learning efficient classification procedures and their applications te chess endgames*. In *Machine Learning: An Artificial Intelligence Approach*, Volume 1. Morgan Kofmann, 1983.
- [37] : Cetnik, B., Kononenko, I., Brackto, I., *Assistant 86: A knowledge-elicitation tool for sophisticated users*. *Progress in Machine Learning*, 2nd European Working Session on Learning, 1987.
- [38] : Cheng, J., Fayyad, Irani K. B., *Improved Decision Trees: A Generalized Version of ID3*, Proceedings of the 5th Internationnal Conference on Machine Learning, pages 100-105, 1988..
- [39] : Manago, M., *intégration de techniques symboliques te numériques en apprentissage*. In *Induction symbolique et numérique à partir des données*. Cépadues, pages 125-149, 1991.

- [40] : Fayyad, K., *The Attribute Selection Problem in Decision Tree Generation*. Proceedings of the 10th National Conference on Artificial Intelligence, 1992.
- [41] : Quinlan, J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.
- [42] : Ramdani, M., *Système d'induction formelle à base de connaissances imprécises*. Thèse de Doctorat, université de Paris VI, 1994.
- [43] : Bouroche, J.P., Tenenhaus, N., *Quelques méthodes de segmentation*. RIRO, 1970.
- [44] : Breiman, L., Friedman, J.H., Olshen, R.A., *Classification and Regression Trees*. Chapman & Hall, 1984
- [45] : Zighed, D.A., *Méthodes et outils pour les processus d'interrogation non arborescents*, Thèse de Doctorat, Université Lyon 1, 1985.
- [46] : Zighed, D.A. and R. Rakotomalala, *Graphs of induction, Training and Data Mining*, Hermes Science Publication, 21-23, 2000.
- [47] : Michalski R. S., Stepp, R. E., Diday, E., *A recent advance in data analysis: clustering objects into classes characterized by conjunctive concepts*. In *Progress in Pattern Recognition*. Kanal and Rosenfeld, pages 33-56, 1981.
- [48] : Michalski, R. S., *Theory and methodology of inductive learning*. In *Machine Learning: an artificial intelligence approach*, vol 1, Morgan Kaufmann, 1983.
- [49] : Clark, P., Niblett T., *The CN2 induction algorithm*, *Machine Learning*. 3, pages 261-283, 1989.
- [50] : Sebag, M., *2nd Order Understandability of Disjunctive Version Spaces*, Workshops of the 14th International Joint Conference on Artificial Intelligence. 1995.
- [51] : Venturini, G., *SIA: a supervised inductive algorithm with genetic search for learning attributes based concepts*. Proceedings of the European Conference on Machine Learning, pages
- [52] : Venturini, G., *Apprentissage Adaptatif et Apprentissage Supervisé par Algorithme Génétique*. Thèse de doctorat. Université de Paris-Sud, 1994.

[53] : Holland, J., *Adaptation in natural and artificial system*. The University of Michigan Press, Ann Arbor, 1975.

Holland, J., *Genetic algorithms and classifier systems: foundations and future directions*. Proceeding of the Second International Conference on Genetic Algorithms, pages 82-89, 1987.

[54] : Holland, J., *Genetic algorithms and classifier systems: foundations and future directions*. Proceeding of the Second International Conference on Genetic Algorithms, pages 82-89, 1987.

[55] : Laurini R., Thompson D.: *Fundamentals of Spatial Information Systems*, Academic Press, London, UK, 680 p, 3rd printing, 1994.

[56] : Zeitouni K., *Etude de l'application du Data Mining à l'analyse spatiale du risque d'accidents routiers par l'exploration des bases de données en accidentologie*, Rapport de contrat PRISM -INRETS, 1998.

[57] : Scholl M. et al. *SGBD Géographiques, Spécificités*, International Thomson Publishing, Paris, France 185 p, 1996.

[58] : Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. *Geographic Information Systems*, Volume 1, Wiley, 1999.

[59] : Han J., Koperski K., and Stefanovic N.: *GeoMiner: A System Prototype for Spatial Data Mining*, *Proc. 1997 ACM-SIGMOD Int'l Conf. on Management of Data (SIGMOD'97)*, Tucson, Arizona, System prototype demonstration, 1997.

[60] : Zeitouni K., Yeh L., *Les bases de données spatiales et le dataming spatial*, *Revue internationale de géomatique*, Numéro spécial "Data mining spatial", Vol. 9, N° 4 (99). *Journées sur le Data Mining spatial et l'analyse du risque*, Versailles, 2000.

[61] : Fayyad, K. et al., *Advances in Knowledge Discovery and Data Mining*, AAAI Press / MIT Press 1996.

[62] : Zeitouni K., *A Survey on Spatial Data Mining Methods Databases and Statistics Point of Views*, IRMA 2000, Information Resources Management Association International Conference, Data Warehousing and Mining Track, Anchorage, Alaska, USA, 2000.

- [63] : Getis, A., Fisher, M., *Spatial analysis – spatial statistics, behavioural modelling and neurocomputing*, Berlin, Springer, 1997.
- [64] : Fotheringham, S., and Rogerson, P.: *Spatial Analysis and GIS*, 1995.
- [65] : Matheron, G., *Principles of geostatistics*. Economic Geology, pp.1246-1266, 1963.
- [66] : Andrienko, N., Andrienko, G., *Interactive Maps for Visual Data Exploration*, International Journal of Geographical Information Sciences 13 (4), pp. 355-374, 1999.
- [67] : Anselin, L., *Local indicators of spatial association*, LISA, Geographical Analysis, 27, 2, pp. 93-115, 1995.
- [68] : Kraak, M.J. and MacEachren, A.M., *Visualisation for exploration of spatial data*. International Journal of Geographical Information Sciences 13 (4), pp. 285-287, 1999.
- [69] : Kraak, M.J., *Visualizing spatial distributions*. Chapter 11 in Longley, P., M. Goodchild, D. Maguire & D. Rhind (editors) *Geographical information systems: principles, techniques, management and applications*. New York: J. Wiley & Sons pp.157-173, 1999.
- [70] : Openshaw, S., Charlton, M., Wymer C. and Craft, *A mark 1 geographical analysis machine for the automated analysis of point data sets*, International Journal of Geographical Information Systems, Vol. 1 (4), pp. 335-358 1987.
- [71] : Lebart L. et al, *Statistique exploratoire multidimensionnelle*, Editions Dunod, Paris, 2^o édition, 1997.
- [72] : Han J., Cai Y. and Cerone N., *Knowledge Discovery in Databases; An Attribute-Oriented Approach*, *Proceedings of the 18th VLDB Conference*, Vancouver, B.C pp. 547-559. 1992.
- [73] : Han J., et al, *A System for Mining Knowledge in Large Relational Databases*, Proc. 1996 Int. Conf. on Data Mining and Knowledge Discovery (KDD'96) , Portland, Oregon, pp. 250-255, 1996.
- [74] : Ester, M., Kriegel, H.-P., Sander J.: *Spatial Data Mining: A Database Approach*, *Proceedings of the 5th Symposium on Spatial Databases*, Berlin, Germany, 1997.

- [75] : Wang, W., Yang, J. and Muntz, R., *STING+: An approach to active spatial data mining*, Proceedings of the Fifteenth International Conference on Data Engineering, Sydney, Australia. IEEE Computer Society. pp. 116-12, 1999.
- [76] : Roddick, J.F, Spiliopoulou, M., *A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research*, ACM SIGKDD Explorations, volume 1, Issue 1 1999.
- [77] : Lu W., Han J. and Ooi B. C., *Discovery of General Knowledge in Large Spatial Databases*, Proc. of 1993 Far East Workshop on Geographic Information Systems (FEGIS'93), Singapore, pp. 275-289, 1993.
- [78] : Ester M., Kriegel H., Sander J., Xu X., *Clustering for Mining in Large Spatial Databases, Special Issue on Data Mining*, KI-Journal, ScienTec Publishing, No. 1, 1998.
- [79] : Han J., Stefanovic N., Koperski K., *Selective Materialization: An Efficient Method for Spatial Data Cube Construction*, Proc. 1998 Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'98) , Melbourne, Australia, 1998.
- [80] : Benali, H., Escofier, B., *Analyse factorielle lissée et analyse factorielle des différences locales*, Revue Statistique Appliquée, XXXVIII (2), pp 55-76, 1990.
- [81] : Ester M., Kriegel H., Sander J., Xu X., *Clustering for Mining in Large Spatial Databases, Special Issue on Data Mining*, KI-Journal, ScienTec Publishing, No. 1, 1998.
- [82] : Ester M., Kriegel H., Sander J., Xu X., *Clustering for Mining in Large Spatial Databases, Special Issue on Data Mining*, KI-Journal, ScienTec Publishing, No. 1, 1998.
- [83] : Agrawal, R., Imielinski, T., Swami, A., *Association Rules between sets of items in large databases*. Proceedings of the ACM SIGMOD. Washington, DC, pp. 207-216, 1993.
- [84] : Koperski, K., and Han J., *Discovery of Spatial Association Rules in Geographic Information Databases, Advances in Spatial Databases (SSD'95)*, pp. 47-66, Portland, ME, 1995.
- [85] : Quinlan J.R., *Induction of Decision Trees, Machine Learning*. vol.1, pp.81-106, 1986.
- [86] : Koperski, K., Han J., and Stefanovic N., *An Efficient Two-Step Method for Classification of Spatial Data*, Proc. International Symposium on Spatial Data Handling (SDH'98) , pp. 45-54, Vancouver, Canada, 1998.

- [87] : [https://math.univ.angers.fr/~labatte/enseignement%20UFR/master%20MIM/classificati
onsupervisee.pdf](https://math.univ.angers.fr/~labatte/enseignement%20UFR/master%20MIM/classificati
onsupervisee.pdf)
- [88] : J.-P. Cocquerez And S. Philipp, 1995 , *Analyse d'Images : filtrage et segmentation*,
Masson,
- [89] : M. Meliani 2012''*Segmentation d'Image par Coopération Régions-Contours''*. Thèse de
magister, Ecole nationale Supérieure en Informatique, Oued-Smar Alger, Ecole Doctorale
STIC.
- [90] : L. Lechlek, M. Ghougal. 2012''*la segmentation d'image médicale par la méthode
d'ensemble de niveaux (level sets)''*. Thèse de Master, Université Larbi Ben M'hidi
Oum El Bouaghi.
- [91] : Glover F., " Future paths for Integer Programming and Links to Artificial Intelligence",
Computers and Operations Research, 5:533-549, 1986.
- [92] : J. Pictet. *Dépasser l'évaluation environnementale, procédure d'étude et d'insertion
dans la décision globale*. Collection Meta, Presses polytechnique et universitaires Romandes,
Volume 1015, Lausanne, suisse, 1990
- [93] : Hamdadou D., *Un Modèle pour la prise de décision en Aménagement du Territoire :
Une Approche Multicritère et une Approche de Négociation*. Thèse de doctorat en
Informatique, université d'Oran, 2008.
- [94] : Bezdek, J. C., *Un theorem de convergence pour l'algorithm de clustering Fuzzy*, IEEE
Trans. Pattern Anal. Machine Intell. 2, 1980.
- [95] : Hamdadou D., Slimani Y., Beldjilali B., *Proposal for a Spatial Decision Support
System in Territory Planning : A Hybrid Approach : Linear Programming, Fuzzy
Measurement, Multicriterion Methods and GIS*, ACIT'2007, the International Arab
Conference on Information Technology, Syria, pp 54-60, 2007.
- [96] : Hafiane A., Zavidovique, B. *FCM with Spatial and Multiresolution Constraints for
Image Segmentation* , Springer Berlin / Heidelberg, 2005.

[97] : Maulik Ujjwal, Sanghamitra Bandyopadhyay, *Partitionnement par Fuzzy à l'aide d'une longueur variable réelle et codé algorithme génétique pour la classification de Pixel*. IEEE Transactions sur les géosciences et la télédétection, Vol. 5, No. 2, Avril 2008.

[98] : Guo Chen, Yu Jie, Peihuang et al, *Classification des images satellitales Basé sur l'amélioration Fuzzy C-Means*, Geo-Information Science spatiale 11 (2) :90-94 Juin 2008.

[99] : Gonzalez, Rafael C., Woods, Richard E., *Thresholding*. Dans *traitement d'images*, pp. 595–611. Pearson Education. ISBN 81-7808-629-8. 2002.

[100] : Chorfi Amina et Belghit Mohamed taher, *Memoire de Master, Classification des images satellitaires par Régions*, université de larbi tébessi - Tébessa