



Faculté des Sciences
Département d'informatique

Spécialité : Informatique

*Option : Télédétection, Analyse et
Traitement Informatique des
données Spatiales*

MEMOIRE
Présenté par

Mr. Beghoura Mohamed Amine

Pour l'obtention du diplôme de Magister en Informatique

Thème

**Segmentation multi sources des Images Satellitaires par
l'algorithme Firefly**

Soutenue ledevant la commission d'examen composée de :

<u>Qualité</u>	<u>Nom et Prénoms</u>	<u>Grade</u>	<u>Etb d'origine</u>
Président	M. Benyettou Abdelkader	Professeur	Usto-MB
Rapporteur	M. Fizazi Hadria	M.conf. A	Usto-MB
Examineur	M. Benamrane Nacera	M.conf. A	Usto-MB
Examineur	M. Mekki Rachida	M.conf. A	Usto-MB
Examineur	M. Chouraqui Samira	M.conf. A	Usto-MB

Année universitaire : 2011/2012

RESUME

La classification non-supervisée des images consiste à les diviser en un ensemble de groupes homogènes. Cette opération s'avère très délicate surtout lorsqu'il s'agit de données complexes telles que celle de la télédétection. Dans notre travail, nous avons appliqué un nouveau méta heuristique bio-inspirée, appelée l'algorithme Firefly, à la segmentation des images satellitaires. L'algorithme est inspiré du comportement social des lucioles dans la nature. Nous avons introduit la notion de la logique floue et les distances noyaux sur cet algorithme afin d'améliorer les résultats de la segmentation. Les résultats obtenus ont montré l'efficacité de l'algorithme Firefly ainsi que les approches proposées.

Mots Clés: Classification Non-Supervisée, Algorithme Firefly, Logique Floue, Distances Gaussienne.

ABSTRACT

Unsupervised classification has to divide the images on a set of homogenous clusters. This operations face lot of difficulties, especially when it comes to a complex data like the remote sensing images. In our work, we applied a new bio-inspired meta heuristic, called the Firefly Algorithm, to solve the problem of remote sensing images segmentation. This algorithm is inspired from the social life of the fireflies in the nature. In order to enhance the segmentations resultants, we have introduced the fuzzy logic and kernel distance on the algorithm. The obtained results show the efficiency of the Firefly algorithm and the proposed approaches.

Keywords: Unsupervised Images Classification, Firefly Algorithm, Fuzzy Logic, Gaussian Distance.

Remerciements

Je remercie...

Dieu de m'avoir illuminé et guidé sur le chemin du bien...

*Mon remerciement le plus chaleureux va principalement à mon encadreur: Mme **FIZAZI Hadria** maître de conférences à l'USTOMB qui m'a orienté, guidé et soutenu sans hésitation ni relâche dans le choix et l'accomplissement de ce thème.*

*J'adresse mes remerciements les plus sincères à monsieur **BENJETTOU Abdelkader**, professeur à l'université USTO-MB et responsable du laboratoire de recherche « **SIMPA** », qui me fait l'honneur d'accepter d'évaluer et de juger ce modeste travail et d'en être président.*

*J'adresse mes remerciements les plus sincères à mademoiselle **BENAMRANE Nacera**, madame **MEKKI Rachida** et madame **CHOURAQUI Samira** maître de conférences A pour avoir accepté d'être membres de mon jury.*

Je remercie tous mes professeurs chacun par son nom pour leur patience et leur abnégation tout au long de mes études.

Enfin je remercie tout le cadre pédagogique et administratif qui sans eux ma mission ne se serait pas accomplie.

DEDICACE

Si je suis arrivé là c'est grâce à Dieu.

J'adresse un grand salut à toutes les personnes qui m'ont
soutenu.

TABLE DES MATIERES

Introduction Générales

Chapitre 1 : Télédétection

1. Introduction	1
2. Définition de la télédétection	2
3. Le processus de la Télédétection	3
4. Emission d'un Rayonnement Electromagnétique (REM)	4
4.1. Interaction rayonnement électromagnétique atmosphère	6
4.2. Interaction rayonnement électromagnétique cible	7
5. Enregistrement de l'énergie émise ou réfléchié par la cible	8
5.1 Capteurs	8
5.2 La résolution	8
5.3 Les plates formes	10
6. Transmission, réception et traitement des données	11
6.1. Transmission des données	11
6.2. Réception des données (Images Satellitaire)	12
6.3. Traitement des données	13
7. Interprétation des images:	15
8. Domaines d'application	15
9. Conclusion	18

Chapitre 2 : Segmentation des Images

1. Introduction	19
2. Qu'est-ce que la segmentation ?	20
2.1. Définition	20
3. Approches de Segmentation	21
4. La Classification Non-Supervisée	22
5. Mesure de Similarité	23
6. Mesures de Validité (Fitness)	24
7. Classification Non-Supervisée Floue	25
8. Segmentation des données imbriquées	27
9. Algorithmes de Classification	30
9.1. Les méthodes Classiques	31
9.1. 1. L'Algorithme k-means	31
9.1. 2. L'Algorithme k-means ISODATA	31
9.2. Les méthodes bio-inspirées	32
9.2.1. Particular Swarm Optimization	32
10. Conclusion	34

Chapitre 3 : Les Lucioles et l'Algorithme Firefly

1. Introduction	35
2. Les lucioles naturelles	36
3. L'Algorithme Firefly	37
4. Validation et Comparaisons	41
5. Conclusion	43

Chapitre 4 : Mise en Œuvre

1. Introduction	44
2. Modélisation du Problème	44
2.1 Organigramme de l'algorithme Firefly	47
2.2 Segmentation par Firefly	48
2.3 Segmentation par Firefly Flou	48
2.4 Segmentation par Firefly Gaussien	50
2.5 Segmentation par Firefly Flou Gaussien	51
3. Images Utilisées	52
3.1. Images Synthétiques:	52
3.2 Image Aérienne:	53
3.3 Images Satellitaires:	54
4. Matériel et Environnement Utilisé	56
5. Discussion des résultats	57
5.1 Images synthétiques	57
5.1.1. Image Synthétique 1	57
5.1.2. Image Synthétique 2	59
5.2. Image Aérienne	64
5.3. Images Satellitaires	67
6. Le Comportement de l'approche Firefly Flou Gaussienne	73
6.1 Comportement vis-à-vis la taille de la population	73
6.2 Comportement vis-à-vis le nombre d'itérations	74
7. Conclusion	75

Conclusion Générale	76
----------------------------	-----------

Liste des Figures.

- Figure 1** Le processus de la télédétection.
- Figure 2** Rayonnement électromagnétique
- Figure 3** Domaines spectraux
- Figure 4** Diffusion et absorption atmosphérique.
- Figure 5** Interaction REM cible
- Figure 6** Capteurs passifs et Capteurs Actifs
- Figure 7** Plates formes d'acquisition
- Figure 8** Image satellitaire « Oran » (LANDSAT 5 -TM)
- Figure 9** Rehaussement de contraste.
- Figure 10** Composition Colorée.
- Figure 11** Exemple d'identification d'objet constituant l'image satellite
- Figure 12** Epaisseur de la « couche d'ozone » sur l'Antarctique, le 16 septembre 2000.
- Figure 13** Mesurer la température de la surface de la mer
- Figure 14** Surveillance des catastrophes naturelles : les inondations du Gard (2002).
- Figure 15** Taxonomie des Approches de Segmentation des Images.
- Figure 16** Imprécision et incertitude de partitionnement
- Figure 17** données non-étiquetées dont les groupes sont non-linéairement séparables
- Figure 18** Séparation par l'algorithme K-means de données non-linéairement séparables
- Figure 19** Transformation d'un espace original vers un espace de dimension supérieur
- Figure 20** Projection dans un espace à dimension supérieure
- Figure 21** Une luciole naturelle
- Figure 22** Organigramme de l'algorithme des lucioles
- Figure 23** Représentation de la Fonction d'Ackley à deux variables avec un optimum global à (0, 0).
- Figure 24** Représentation de la fonction Forest de Yang proche du minimum (0, 0).
- Figure 25** Vecteur de position d'une luciole.
- Figure 26** Schéma général de l'application

- Figure 27** Problème de résolution spatiale des images Satellitaires
- Figure 28** Illustration du d'imprécision dans les Images Satellitaire
- Figure 30** Image Synthétique 1 (image légèrement bruitée)
- Figure 31** Image Synthétique 2 (Image considérablement bruitée)
- Figure 32** Image Aérienne
- Figure 33** Carte Thématique de l'image aérienne
- Figure 34** La région d'Oran
- Figure 35** Image Satellitaire Oran Ouest datée de 1993
- Figure 36** Image Satellitaire Oran Ouest datée de 2001
- Figure 37** Schéma UML de l'application
- Figure 38** Résultats visuels des testes avec différents paramètres de l'algorithme Firefly
- Figure 39** Segmentation Flou de l'image Synthétique 1
- Figure 40** Résultats Visuels de la segmentation de l'image synthétique 2 avec l'algorithme Firefly et $n = 4$.
- Figure 41** Résultats Visuels de la segmentation de l'image synthétique 2 avec l'algorithme Firefly et $n = 5$.
- Figure 42** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec $n = 4$
- Figure 43** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec $n = 5$
- Figure 44** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec $n = 4$
- Figure 45** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec $n = 5$
- Figure 46** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec $n = 4$
- Figure 47** Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec $n = 5$
- Figure 48** Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly.
- Figure 49** Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Flou.
- Figure 50** Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Gaussien.
- Figure 51** Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Flou Gaussien.

- Figure 52** Prétraitement de la première image satellitaire
- Figure 53** Résultat Segmentation des images Satellitaires par l'algorithme Firefly.
- Figure 54** Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Flou
- Figure 55** Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Gaussien
- Figure 56** Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Flou Gaussien
- Figure 57** Effet de la taille de la population sur la fonction de fitness

Liste des Tables.

Table 1	Comparaison des performances FFA, PSO, AG
Table 2	Evaluation de la segmentation d'image synthétique avec différents paramètres de l'algorithme Firefly.
Table 3	Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly avec $n = 4$ et $n = 5$.
Table 4	Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec $n = 4$ et $n = 5$.
Table 5	Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec $n = 4$ et $n = 5$.
Table 6	Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec $n = 4$ et $n = 5$.
Table 7	Résultats de la segmentation des images satellitaires avec l'algorithme Firefly
Table 8	Résultats Segmentation des Images Satellitaires avec l'algorithme Firefly Flou
Table 9	Résultats de la segmentation des images satellitaires avec l'algorithme Firefly Gaussien
Table 10	Résultats de la segmentation des images satellitaires avec l'algorithme Firefly Flou Gaussien
Table 11	Effet de la taille de la population sur la fonction de fitness
Table 12	Effet du nombre d'itérations sur la fonction de fitness

Introduction Générale

L'extraction de l'information à partir d'une image satellitaire reste toujours un problème pour la communauté des chercheurs. Ce type d'images est caractérisé par la grandeur de leur taille (images multi-spectrales / hyper-spectrales) et la variété des thèmes qu'elles contiennent. Les images satellitaires à l'état brute sont difficilement exploitables. Ceci est dû à l'influence du bruit et du contraste d'un coté et à la quantité et la diversité des informations qu'elles peuvent contenir. L'utilisation de ces images nécessite l'intervention du processus de la télédétection afin de faciliter l'interprétation de ces données.

La télédétection est un domaine de recherche qui vise l'ensemble des outils et des techniques permettant de traiter et d'analyser automatiquement les données enregistrées par les capteurs installés sur des satellites où autres plateformes. La télédétection se forme autour d'un processus constitué de plusieurs étapes de traitement, où la classification des données réside au cœur de ce processus. Cette étape est très importante à cause de son rôle qui consiste à partitionner les données en un ensemble de groupes. Dans la littérature, il existe deux types de classification : classification supervisée nécessitant une connaissance a priori des classes des données afin d'établir une étape d'apprentissage et classifier l'image par la suite. Le deuxième type est la classification non-supervisée. Ce type est appliqué dans le cas d'absence des connaissances a priori sur les classes de l'image.

Dans ce travail, nous nous intéressons à la classification non-supervisée des images satellitaires. Pour cela, nous allons utiliser une nouvelle méthode de classification. Cette méthode, appelée l'algorithme Firefly (où luciole en Français), est basée sur le comportement social des lucioles dans la nature. Cet

algorithme a été introduit par Xan-She Yan en 2007 à l'université de Cambridge. Les expérimentations établies par l'auteur de cet algorithme ont montré la supériorité des performances de cette méthode par rapport à d'autres algorithmes connus tel que l'algorithme PSO (Particular Swarm Optimization) et les algorithmes génétiques.

D'un autre coté, nous allons présenter trois nouvelles approches de segmentations basées sur cette nouvelle métaheuristique. Les approches proposées visent à résoudre deux problèmes souvent rencontrés lors de la classification des images satellitaires. Le premier problème est le problème de l'imprécision et d'incertitude. Dans une image satellitaire, un pixel représente une région géographique relativement large. Par conséquent, cette zone pourra contenir plus d'un seul type de terrain. Pour cela, nous proposons l'algorithme Firefly Flou pour la segmentation floue des images satellitaires. Le deuxième problème est le problème des classes imbriquées. Ce dernier peut être engendré par plusieurs facteurs tels que le bruit et la nature des objets captés. Comme résolution à ce deuxième problème, nous proposons l'algorithme Firefly Gaussien afin d'améliorer les résultats de classification. La troisième approche consiste à combiner l'algorithme Firefly Flou et Firefly Gaussien. Cette combinaison a montré des performances de segmentation.

Le premier chapitre de ce mémoire contient des généralités sur le processus de la télédétection. Dans le deuxième chapitre, nous présentons un état de l'art des méthodes de segmentation des images. Dans le troisième, nous présentons la théorie derrière l'algorithme Firefly. Le quatrième chapitre présente en détails la mise en œuvre de la segmentation des images par l'algorithme Firefly, ainsi que les expérimentations et les résultats obtenus par les approches proposées.

Chapitre 1

La Télédétection

1. Introduction

La surface de la terre subit constamment des changements rapides dus à l'urbanisation, l'industrie, et la globalisation. Les problèmes environnementaux tels que la sécheresse, la désertification, les raide marais et bien d'autres problèmes causent de plus en plus des troubles à l'activité humaine. L'étude et l'analyse de l'environnement fut donc importante pour l'être humain [1].

La prise de décision concernant l'environnement dépend de la qualité et de la fiabilité des informations et des connaissances disponibles à la porté des décideurs. Pour cela, l'observation de la terre est devenue un moyen efficace et nécessaire pour mieux comprendre la dynamique de la terre.

La télédétection (remote sensing) est un terme qui est apparu dans les années 60 aux USA pour désigner cette technique de détection à longue distance, permettant d'observer la terre à distance. Le terme de télédétection a été introduit officiellement dans la langue française en 1973 [7]. Le développement de cette technologie non photographique fit d'immenses progrès après la mise en orbite, en 1972, du premier satellite de cartographie, LANDSAT 1 [2][3].

La télédétection joue un rôle important dans l'étude de l'environnement terrestre au moyen des capteurs, à bord des plates-formes aériennes ou spatiales. Un rôle qui s'étend de l'interprétation visuelle jusqu'à l'extraction sophistiquée des informations grâce à des algorithmes d'analyse.

L'histoire de la télédétection peut être découpée en cinq grandes époques. La première commence de 1856, où un appareil photographique a été installé de façon fixe à bord d'un ballon, jusqu'à la première guerre mondiale pendant laquelle sont explorées les possibilités de la photographie aérienne pour la cartographie. La deuxième époque débute de la première guerre mondiale jusqu'à la fin des années 50. Dans cette deuxième époque, la photographie aérienne est devenue opérationnelle pour la cartographie, la recherche pétrolière, la surveillance de la végétation. La période qui commence en 1957 et s'achève

en 1972 est marqué par les débuts de l'exploration de l'espace. La première application opérationnelle de la télédétection spatiale apparaît dans les années 60 avec les satellites météorologiques de la série ESSA. La quatrième époque est marquée par le lancement en 1972 du satellite ERTS (rebaptisé ensuite Landsat 1), premier satellite de télédétection des ressources terrestres, ce qui a ouvert l'époque de la télédétection moderne. Depuis cette époque, la télédétection assiste à une évolution continue due à l'augmentation de la résolution spatiale et la diversification des capteurs. Les technologies et les méthodes de la télédétection se sont dramatiquement évoluées. Le spectre des capteurs installés sur les satellites et les dispositifs aériens a permis de fournir une suite d'information d'intérêt et d'une grande importance aux experts et spécialistes. L'accès facile aux archives des données de la télédétection, la continuité de baisse des prix de ces données, et l'amélioration des performances des satellites sont les facteurs essentiels à l'augmentation de l'importance de la télédétection.

Dans ce chapitre, nous allons présenter les notions de bases de la télédétection afin d'expliquer son processus et son intérêt dans différents domaines d'application tel que: l'agriculture, la foresterie, la géologie, la météorologie, la cartographie, etc.

2. Définition de la télédétection

La télédétection est la technique qui, par l'acquisition d'images, permet d'obtenir de l'information sur la surface de la Terre sans contact direct avec celle-ci [8]. La télédétection englobe tout le processus qui consiste à capter et à enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchi, à traiter et à analyser l'information, pour ensuite mettre en application cette information. Le terme télédétection est utilisé pour définir [4][5] :

« L'ensemble des connaissances et techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci »

3. Le processus de la Télédétection

La télédétection est un processus résultant de l'interaction entre trois éléments fondamentaux : une source d'énergie, une cible et un vecteur. La télédétection comprend les étapes suivantes [4]:

a. Source d'énergie ou d'éclairage: la source d'énergie est nécessaire pour la télédétection permettant l'éclairage en fournissant de l'énergie électromagnétique à la cible.

b. Les radiations et l'atmosphère: Au cours du déplacement de l'énergie de la source vers la cible, elle interagit avec l'atmosphère. Cette interaction se présente une deuxième fois lorsque l'énergie se réfléchit à partir de la cible vers le capteur après l'avoir éclairé.

c. Interaction avec la cible: une fois l'énergie rentre en contact avec la cible, elle interagit avec celle-ci selon les propriétés de la cible et aussi le type des radiations.

d. Enregistrement de l'énergie par les capteurs: une fois l'énergie réfléchit par la cible, des capteurs installés sur, des dispositifs aériens tels que les satellites, sont nécessaires pour enregistrer ces radiations électromagnétiques.

e. Transmission, réception, et traitement: l'énergie enregistrée par les capteurs doit être transmise, la plupart du temps sous forme électronique, vers une station de réception, où ces données vont être traitées.

f. Interprétation et l'analyse: l'image traitée sera par la suite interprétée, visuellement ou numériquement, afin d'extraire les informations sur la cible.

g. Application : à la fin, le processus de la télédétection se termine par l'utilisation des informations extraites, à partir des images, dans des applications afin de donnée l'état d'utilisation de la terre.

Ces septes éléments comprennent tous les aspects du processus de la télédétection. Le schéma suivant illustre davantage le processus de la télédétection :

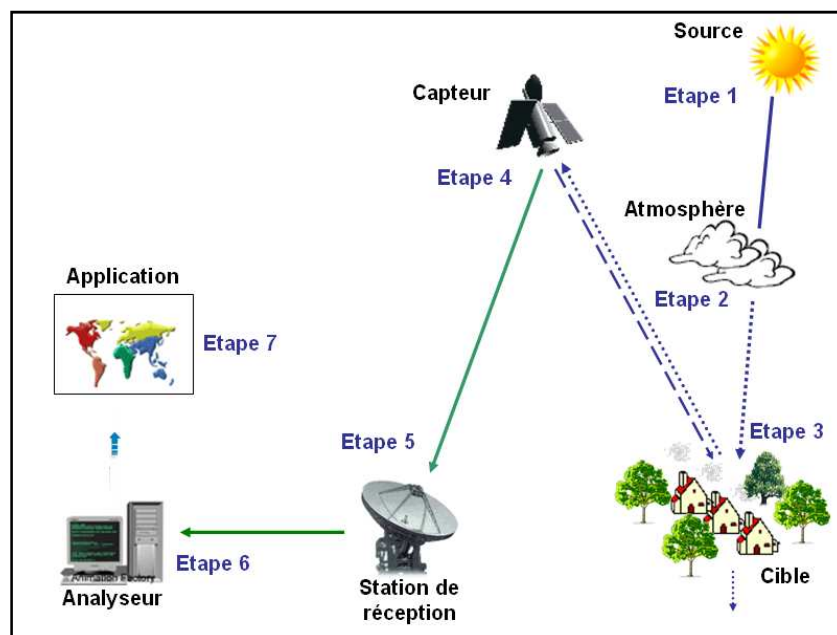


Figure 1 : Le processus de la télédétection.

4. Emission d'un Rayonnement Electromagnétique (REM)

Une source d'énergie est nécessaire dans le processus d'illumination de la cible dans le processus de la télédétection. Cet éclairage se présente sous forme de rayonnements électromagnétiques dotés des mêmes propriétés et ils se comportent suivant la théorie de base des ondes. Le rayonnement s'agit d'un champ électrique qui change de d'amplitude dans une direction perpendiculaire à la direction, et un champ magnétique orienté vers l'angle droit du champ électrique. Les deux rayonnements se déplacent à la vitesse de la lumière. La figure suivante représente deux caractéristiques des rayonnements électromagnétiques importantes à la compréhension de la télédétection, ces deux caractéristiques sont la fréquence et la longueur d'onde

- ♣ **Une longueur d'onde (λ)** : représente la longueur d'un cycle d'une onde de rayonnement. Elle correspond à la distance entre deux crêtes successives d'une onde.
- ♣ **Une fréquence (ν)** : La fréquence représente le nombre d'oscillations par unité de

temps. La fréquence est mesurée en Hertz (Hz) ou en multiples d'Hertz.

La relation entre la longueur d'onde et la fréquence peut être résumée par la formule suivante : $c = \lambda \cdot \nu$ [9, 10, 11] Ainsi, la longueur d'onde et la fréquence sont donc inversement proportionnelles.

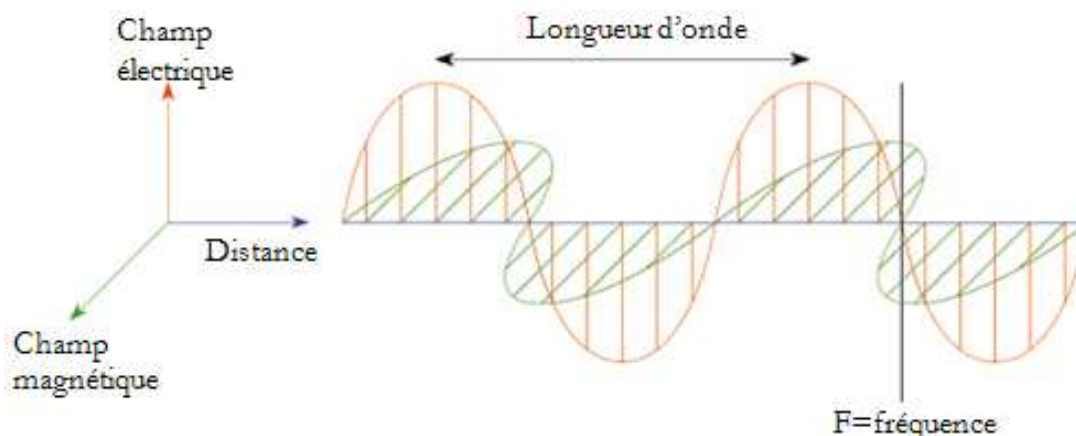


Figure 2 : Rayonnement électromagnétique

L'énergie constituée d'ondes électromagnétiques est considérée comme le vecteur porteur d'information relative à l'objet étudié. Les objets reflètent une partie de l'énergie qui les atteint. C'est généralement cette portion de lumière qui donne leurs couleurs aux objets.

La décomposition du rayonnement électromagnétique en termes de fréquence (ν), d'énergie (e) ou encore de longueur d'onde (λ) forme le spectre électromagnétique. [9] [11] [2]

Le spectre électromagnétique est une décomposition du rayonnement électromagnétique en différentes bandes. Chacune de ses bandes spectrales peut être utilisée dans un domaine d'application. Cependant, les bandes les plus utilisées en télédétection sont:

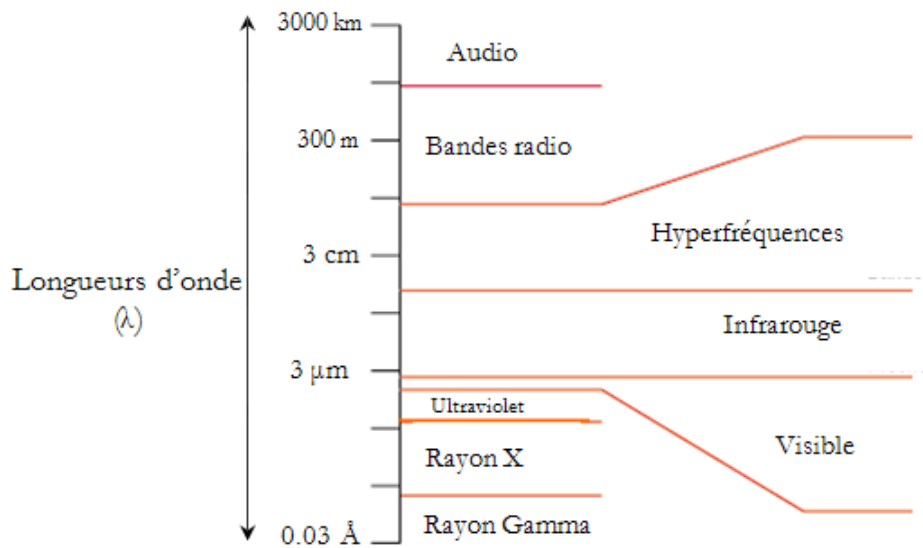


Figure 3 : Domaines spectraux

4.1. Interaction rayonnement électromagnétique atmosphère

La propagation du rayonnement électromagnétique est perturbée au niveau de l'atmosphère par l'effet des mécanismes d'absorption et de diffusion :

- ♣ **L'absorption atmosphérique:** Ce phénomène survient suite à cause de l'absorption de différentes longueurs d'onde par les molécules atmosphériques, à l'exemple des molécules d'ozone qui absorbent les ultraviolets.
- ♣ **La diffusion atmosphérique:** est le résultat de la déviation du rayonnement de sa trajectoire initiale par les particules. Les particules responsables de ce mécanisme peuvent être la poussière, le pollen ou encore des molécules gazeuses comme l'azote, l'oxygène ... etc.

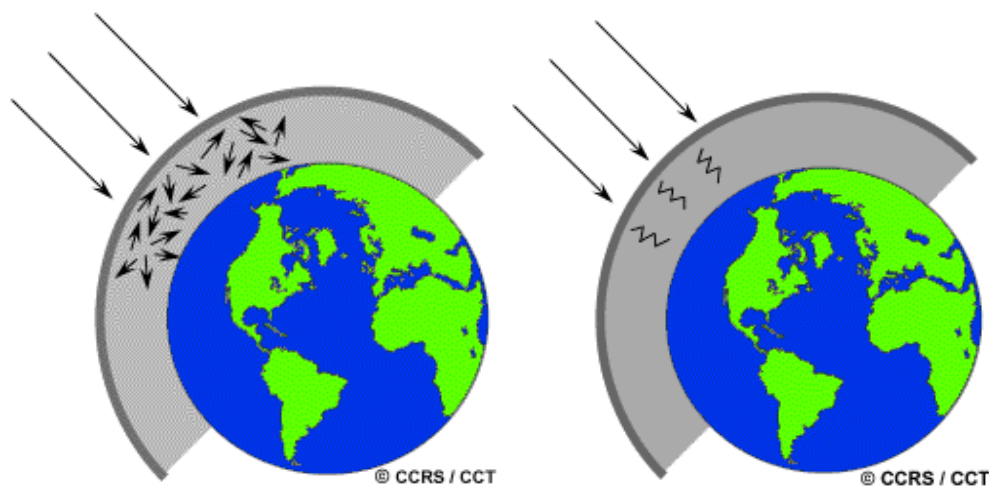


Figure 4 : Diffusion et absorption atmosphérique.

4.2. Interaction rayonnement électromagnétique cible

Le rayonnement électromagnétique non absorbé ou diffusé par l'atmosphère atteint et interagit avec la cible, (étape 3), subit à son tour des changements causés par les caractéristiques du rayonnement incident et les propriétés de la surface de la cible. Les trois phénomènes suivants se produisent: [12]

- ♣ **L'absorption:** L'absorption se produit lorsque l'énergie du rayonnement est absorbée par la cible, ce qui modifie l'énergie interne de l'objet en augmentant sa température. L'absorptance $\alpha \in [0,1]$ ou le coefficient d'absorption, est défini par le rapport entre l'énergie absorbée et l'énergie reçue.
- ♣ **La transmission:** La transmission se produit lorsque l'énergie du rayonnement passe à travers la cible. La transmission $\tau \in [0,1]$ ou coefficient de transmission représente le rapport entre l'énergie transmise et l'énergie reçue.
- ♣ **La réflexion:** La réflexion se produit lorsque la cible redirige l'énergie du rayonnement. La réflexion $\rho \in [0,1]$ ou coefficient de réflexion est donné par le rapport entre l'énergie réfléchie et l'énergie reçue.

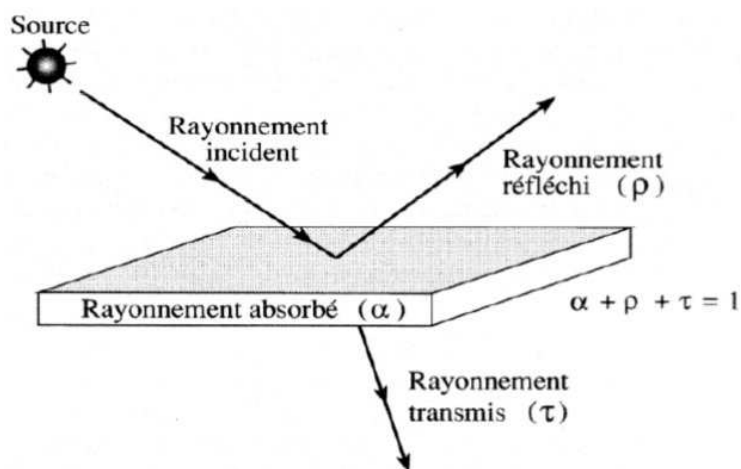


Figure 5: Interaction REM cible

5. Enregistrement de l'énergie émise ou réfléchie par la cible

L'enregistrement de l'énergie réfléchie ou émise par la cible, (étape 4) nécessite l'utilisation d'un système d'acquisition. Ce dernier englobe, des instruments de mesure que sont les capteurs ainsi que des plates-formes ou satellites sur lesquelles ils sont installés.

5.1 Capteurs

Pour si longtemps, le soleil a été considéré comme la seule source d'énergie et de rayonnement. Le soleil représente une source d'énergie idéale pour la télédétection, les capteurs qui mesurent cette énergie naturellement sont appelés des capteurs passifs. Les capteurs passifs sont utilisables que par la présence de l'énergie solaire, ce qui fait, ils deviennent inactifs la nuit, lorsque l'énergie du soleil ne sera pas émise. Par contre une énergie naturellement émit (comme l'infrarouge thermique) peut être détectée durant la nuit. D'un autre coté, les capteurs actifs sont indépendants de l'énergie produite par le soleil. Les capteurs produisent leur propre énergie pour illumination de la cible.

5.2 La résolution

La résolution est une mesure de la capacité d'un système optique de séparer des

signaux proches spatialement et/ou spectralement. La capacité de mesurer une information par télédétection exige la considération prudente de quatre types de résolution : radiométrique, spatiale, spectrale et temporelle [6].

La résolution spatiale : La résolution spatiale est la surface minimale que peut distinguer un capteur. Cette résolution s'exprime en mètre ou kilomètre et mesure le côté d'un pixel. Les images satellitaires sur lesquels seuls les grands éléments sont visibles ont une résolution spatiale grossière ou basse. Les images à résolution fine ou élevée permettent l'identification d'éléments de plus petites dimensions.

La résolution spectrale : La résolution spectrale d'un capteur représente l'intervalle de longueur d'onde dans lequel le capteur a la capacité d'enregistrer l'énergie reçue.

La résolution radiométrique : Cette résolution est définie par le seuil de sensibilité d'un capteur. Elle représente la plus faible intensité radiométrique (dans chaque bande spectrale) que le capteur est capable de détecter. Ainsi, plus la résolution radiométrique est grande plus la qualité thématique des images est bonne.

La résolution temporelle : La résolution temporelle représente le temps que prend le satellite pour effectuer un cycle orbital complet, et donc pour observer de nouveau la même scène à partir du même point dans l'espace. On utilise ce type de résolution pour le suivi d'un développement de la végétation ... etc., ou en de source humaine comme le développement urbain, la déforestation ... etc.

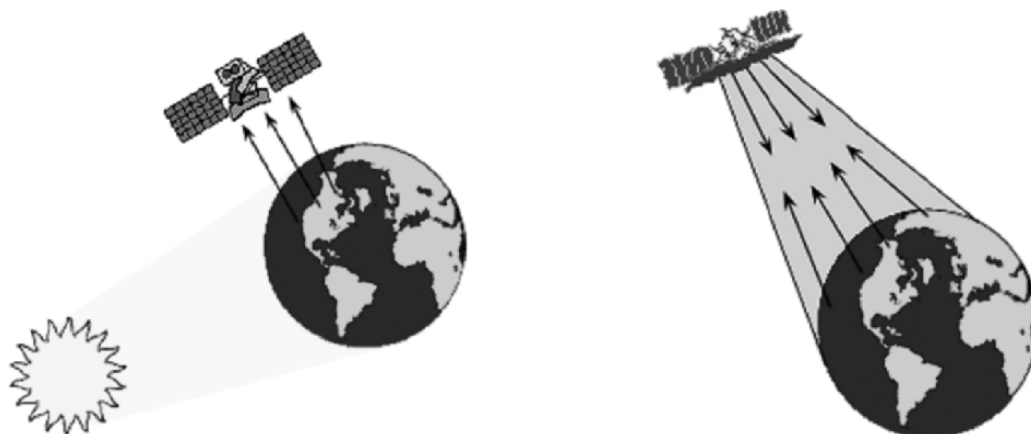


Figure 6 : Capteurs passifs et Capteurs Actifs

5.3 Les plates formes

Parmi les plates formes les plus utilisées pour l'acquisition de l'énergie réfléchie, on a les avions et les satellites.

- ✦ **Les avions:** Les avions volent à une altitude relativement faible (quelques kilomètres seulement) ils ne peuvent par conséquent prendre en photo (du point de vue de la télédétection) que des portions de territoire limitées avec leurs détails.
- ✦ **Les satellites:** Ils se déplacent en suivant une orbite. Généralement, l'orbite a une forme quasi-circulaire. Les satellites sont des plates-formes utilisées communément en télédétection; ils véhiculent des capteurs extrêmement variés et souvent spécialisés dans l'observation de la météo, des paysages ou des catastrophes naturelles, ...etc.



Figure 7 : Plates formes d'acquisition

6. Transmission, réception et traitement des données

Les données acquises par un capteur aéroporté sont recueillies une fois que l'avion est de retour au sol. Elles peuvent être, par la suite, traitées et remise à l'utilisateur. [2]

Cependant, les données satellitaires doivent être transmises électroniquement, réceptionnées puis enfin traitées (étape 5) afin d'être interprétées et utilisées dans différents domaines d'application.

6.1. Transmission des données

Les données satellitaires peuvent être transmises, électroniquement, par le biais de trois différentes méthodes : [WEB 01]

A. Dans le cas où le satellite se situe dans le cercle de réception de la station, les données peuvent être transmises directement à une station de réception terrestre.

B. Si le satellite n'est pas dans le cercle de réception d'une station, les données peuvent être stockées momentanément à bord du satellite.

C. Les données peuvent être transmises par le biais de satellites de communications qui sont en orbite géostationnaire autour de la Terre. Les données sont transmises de satellite à satellite jusqu'à ce qu'on puisse les retransmettre à la station.

6.2. Réception des données (Images Satellitaire)

Les données de la télédétection sont reçues sous forme d'images satellitaire. Une image satellitaire est une représentation graphique, constituée de nombreux carrés appelés pixels. Le pixel représente la plus petite unité figurant sur une image satellitaire, quelle que soit la longueur d'onde ou le dispositif de télédétection qui a été utilisée pour enregistrer l'énergie électromagnétique [2].

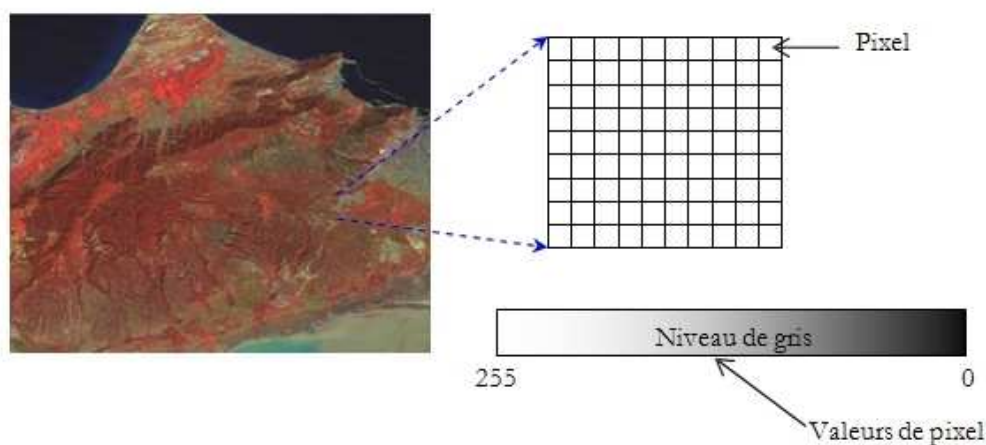


Figure 8 : Image satellitaire « Oran » (LANDSAT 5 -TM)

Afin d'obtenir des informations différentes, les capteurs observent la terre dans plusieurs bandes de longueurs d'onde, encore appelées "canaux". On dit qu'ils sont multi spectraux.

Chaque pixel d'une image a une valeur. Cette valeur correspond à l'intensité du rayonnement réfléchi par l'objet observé dans la gamme de longueur d'ondes auxquelles le capteur est sensible. Le nombre maximum de niveaux d'intensité disponibles dépend du nombre de bits utilisés pour représenter l'intensité enregistrée.

Mathématiquement c'est une fonction $F(x, y)$ à deux variables entières x et y qui sont respectivement la ligne et la colonne, fournissant une ou plusieurs valeurs entières représentant le niveau de gris du pixel dans les différents canaux.

6.3. Traitement des données

Les images satellitaires à l'état brut sont difficilement exploitables. Afin d'améliorer leurs qualités, différents traitements numériques ont été mis au point tel que les prétraitements, le rehaussement de contraste, la visualisation de l'image où encore la classification.

Les prétraitements: Les prétraitements sont des opérations effectuées sur les images en amont de toute analyse ou extraction d'information. Cette opération regroupe : [WEB 01][WEB 03]

- ♣ **Les corrections radiométriques:** Le vieillissement ou le dérèglement des capteurs, des défauts de conception ou encore des perturbations atmosphériques peuvent affecter la qualité des images satellitaires. Ainsi, des corrections radiométriques sont nécessaires. Ces corrections, consistent à réaffecter à chaque pixel une valeur radiométrique au plus proche possible de celle mesurée sur le terrain.
- ♣ **Les corrections géométriques:** Les distorsions géométriques sont généralement dues à l'environnement tel que les courbures de la terre, les variations d'altitude au sol...etc. ou encore les erreurs dues au mouvement de la plate-forme lors de l'enregistrement de la scène. Le processus de correction géométrique consiste à identifier différentes coordonnées de l'image (ligne, colonne) et à les assortir à leurs coordonnées au sol (latitude, longitude).
- ♣ **Le rehaussement de contraste:** Le rehaussement de contraste a pour but d'augmenter la distinction entre les différents éléments de la scène par un étirement de l'histogramme de façon à ce que la mesure la plus forte soit codée à $2^{\text{nombre de bits par pixel}-1}$ et la plus faible à 0.

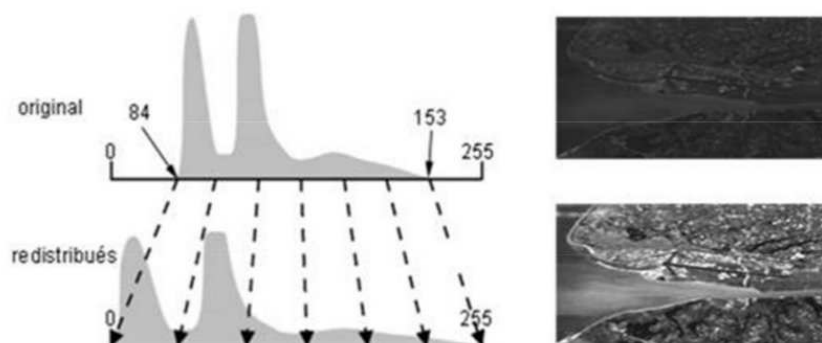


Figure 9 : Rehaussement de contraste.

- ✎ **La visualisation de l'image:** La visualisation de l'image a pour but de combiner plusieurs bandes de données multi spectrales ou multi temporelles pour former une "nouvelle" image qui montre plus clairement certains éléments de la scène.

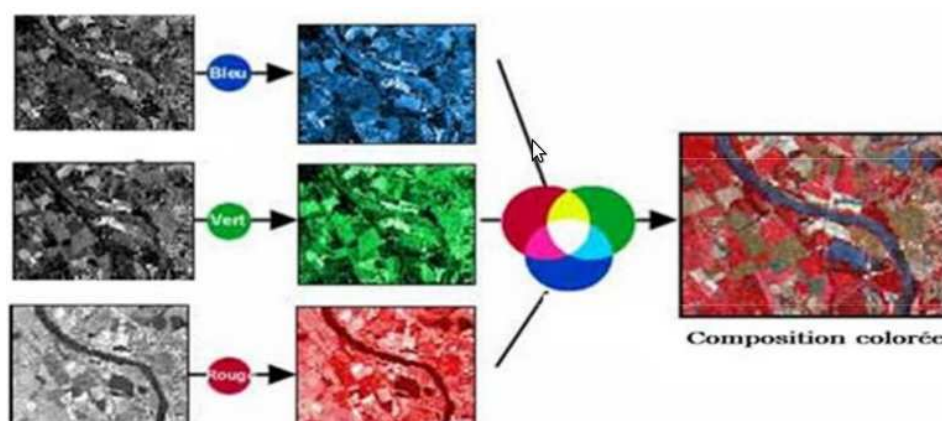


Figure 10 : Composition Colorée.

Classification: La classification consiste à regrouper les pixels selon leur ressemblance spectrale ou à déterminer les contours d'un groupe de pixels pour former des unités spatiales interprétables en termes de classes ou catégories thématiques. Ce processus est dénommé classification et les algorithmes associés sont les classificateurs. Une classe est définie comme un ensemble d'objets (ou d'individus) ayant des caractères communs. Dans l'espace spectral, la classe forme un agrégat de pixels repéré par son centre de gravité désigné par la suite par le vecteur moyen m et sa variance σ^2 ou, mieux,

dans le cas multi-bandes, par la matrice de variance-covariance Σ .

- ⤴ **Classe thématique:** catégorie ou classe d'occupation du sol défini selon la nature de la surface du territoire ou de son utilisation. Le terme « catégorie » paraît plus exact, car il n'implique aucune notion hiérarchique contrairement au terme de classe. Dans la pratique, c'est ce dernier qui est le plus usité en classification.
- ⤴ **Classe spectrale:** groupement(s) de pixels présentant des propriétés spectrales similaires.

7. Interprétation des images:

L'interprétation et l'analyse des images de télédétection, (étape 6) a pour but d'identifier et de mesurer différents objets constituant la scène étudiée, et ce, en se basant sur une combinaison de différentes caractéristiques tel que : [WEB 01]

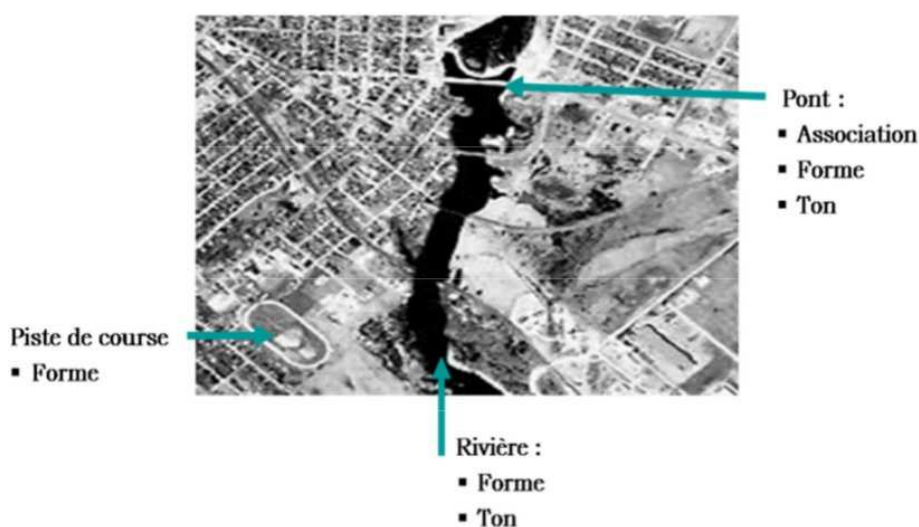


Figure 11 : Exemple d'identification d'objet constituant l'image satellite

8. Domaines d'application

La télédétection est utilisée dans plusieurs domaines scientifiques. Parmi les quels nous citons [7]:

Météorologie (climatologie) :

Le premier grand domaine d'application de la télédétection a été l'étude de l'atmosphère (météorologie et climatologie). L'intérêt de la télédétection dans ce domaine est d'assurer une couverture globale et très fréquemment répétée de la planète entière. Les capteurs utilisés permettent d'observer les nuages et leur déplacement, de mesurer des températures ou le contenu en vapeur d'eau de l'atmosphère.

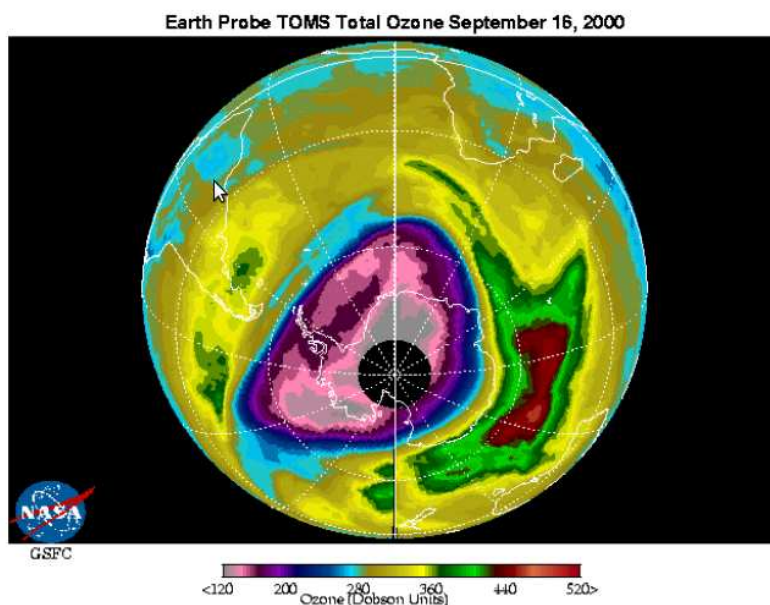


Figure 12 : Epaisseur de la « couche d’ozone » sur l’Antarctique, le 16 septembre 2000.

Océanographie:

En océanographie, la télédétection offre l'avantage de permettre une vision synoptique de vastes régions qu'il est impossible d'obtenir par les moyens traditionnels (bateaux). Parmi les applications océanographiques de la télédétection, citons enfin l'étude des glaces de mer en régions polaires.

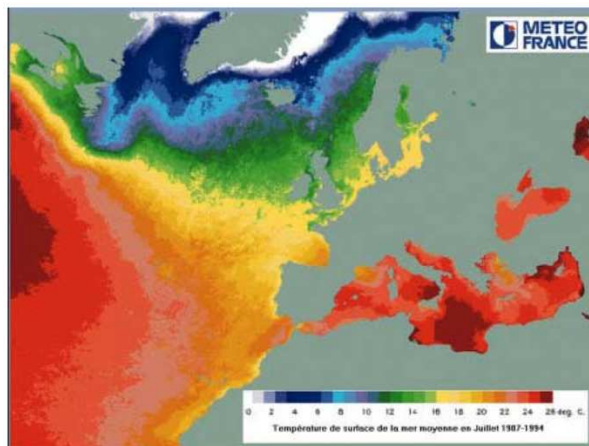


Figure 13 : Mesurer la température de la surface de la mer

Applications terrestres:

Les applications terrestres de la télédétection sont extrêmement variées. La télédétection s'avère utile dans plusieurs applications terrestres tel que: cartographie, géologie et prospection minière, mais aussi surveillance des cultures ou du couvert forestier, urbanisme, aménagement, génie civil, etc... Le traitement de l'imagerie satellitaire numérique est une discipline en constant développement, et la baisse du coût des matériels informatiques a entraîné une augmentation rapide du nombre des utilisateurs.

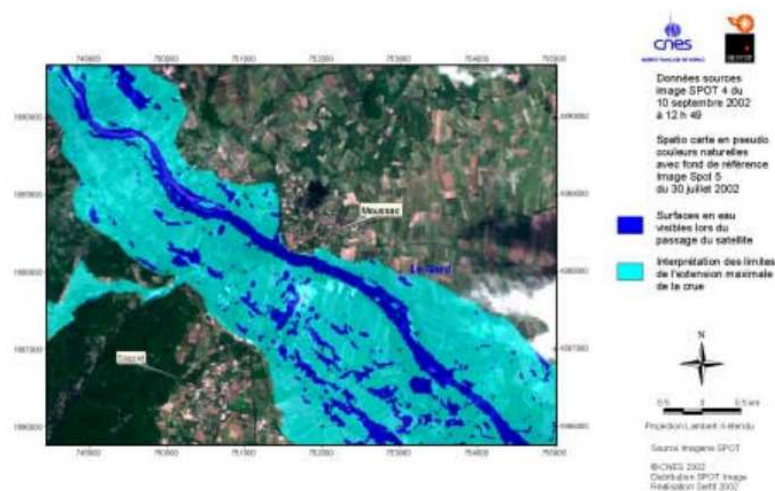


Figure 14 : Surveillance des catastrophes naturelles : les inondations du Gard (2002).

9. Conclusion

La télédétection a été investie dans de divers domaines scientifiques. Elle est considérée comme un moyen scientifique d'avant-garde permettant de fournir des informations de plus en plus précises et précieuses sur les objets, les surfaces et les profondeurs les plus inexplorés de notre univers. Les traitements effectués lors du processus de la télédétection affectent la qualité des informations induites. Notre travail vise l'étape de la classification des images satellitaires, dont le but est d'améliorer la qualité de cette opération fondamentale pour la télédétection.

Chapitre 2

La Segmentation des Images

1. Introduction

La segmentation des images est au cœur du processus de la télédétection [35]. Bien souvent, les résultats de cette étape sont utilisés comme base aux applications d'interprétations. Cette étape est très importante pour l'extraction automatique des caractéristiques géographiques à partir des images satellitaires.

La classification figure parmi les méthodes permettant de partitionner l'image en un ensemble de régions homogènes, suivant les caractéristiques de l'image (par exemple, l'intensité des pixels [36]). Dans la littérature, Nous distinguons deux types de classification : supervisée (analyse discriminante) et non-supervisée (Clustering) [37]. Il est nécessaire de connaître la différence entre ces deux types. La classification supervisée est un processus qui nécessite une étape d'apprentissage. En se basant sur une collection de données étiquetées, où à chaque donnée d'apprentissage est associé la classe dont elle appartient, l'apprentissage consiste à construire un modèle descriptif des différentes classes. Ce modèle servira comme base à la classification des nouvelles données non-étiquetées [38]. Dans le cas des images satellitaires, les classes correspondent aux différents types de terrains (Par exemple : Mer, Forêt, Sable, Urbain, etc.). Contrairement à la classification supervisée, la classification non-supervisée consiste à grouper une collection de données non-étiquetées en un ensemble de groupes homogènes, selon des critères de similarité/dissimilarité entre ces données. Donc, la classification non-supervisée examine les pixels de l'image afin des les rassembler en un certain nombre de groupes (clusters). Les pixels du même groupe montrent les mêmes caractéristiques (par exemple, les valeurs spectrales), et donc, peuvent représenter un seul type de terrain [39]. La classification non-supervisée est utilisée dans les cas d'absence des informations à priori sur les types de terrains de la région.

Dans ce chapitre, nous allons présenter un état de l'art des méthodes de segmentations et poser les problématiques liées à la segmentation des données, telles que les données satellitaires.

2. Qu'est-ce que la segmentation ?

Il n'est pas aisé de trouver une seule définition de la segmentation car cette tâche est souvent confondue avec la classification ou l'étiquetage, on peut néanmoins essayer d'en trouver une afin de fixer le cadre dans lequel nous écrirons et ainsi faire disparaître les ambiguïtés. Segmenter une image signifie trouver ses régions homogènes et ses contours. Ces régions et contours sont supposés être pertinents, c'est-à-dire que les régions doivent correspondre aux parties significatives des objets du monde réel, et les contours à leurs contours apparents. Arabie et al. [42] a défini la segmentation comme:

« L'ensemble des méthodes conçues pour identifier les groupes d'objets homogènes »

Tandis que Everitt [43] a défini la notion du groupe par :

« Un groupe est un ensemble d'entités semblables. Cependant, les entités des différents groupes ne sont pas semblables »

2.1. Définition

Soit X le domaine de l'image et f la fonction qui associe à chaque pixel une valeur $f(x,y)$. Si nous définissons un prédicat P sur l'ensemble des parties de X , la segmentation de X est définie comme une partition de X en n sous-ensemble $\{R_1, \dots, R_n\}$ tels que :

1. $X = \bigcup_{i=1}^n R_i$
2. $\forall i \in \{1, \dots, n\}$ R_i est connexe
3. $\forall i \in \{1, \dots, n\}$ $P(R_i) = \text{vrai}$
4. $\forall i, j \in \{1, \dots, n\}^2$ R_i est adjacent à R_j et $i \neq j \Rightarrow P(R_i \cup R_j) = \text{faux}$

Où \bigcup représente une union d'ensemble disjoints.

Le prédicat P est utilisé pour tester l'homogénéité des ensembles R_i . Ces sous-ensembles constituent les régions de l'image. Une segmentation de l'image est donc sa décomposition en un ensemble de régions homogènes, le critère d'homogénéité P restant à déterminer.

3. Approches de Segmentation

De très nombreuses méthodes de segmentation existent permettant de résoudre autant le problème de la segmentation des images. D'après [44], nous comptons plus d'un millier d'approches distinctes. La taxonomie suivante illustre les approches de segmentation :

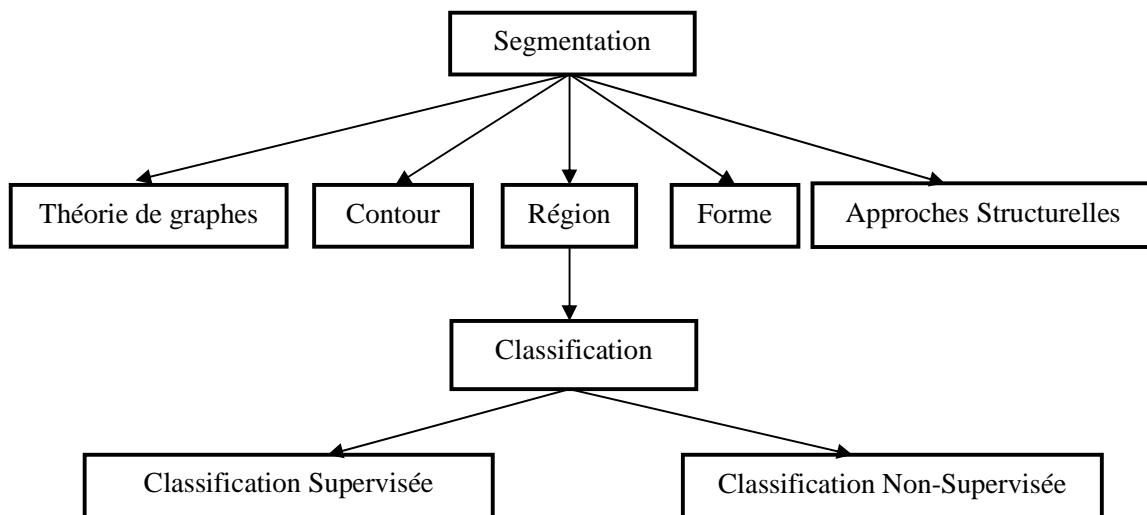


Figure 15 : Taxonomie des Approches de Segmentation des Images.

Approches Contours (où frontières) : nous considérons que les primitives à extraire sont les lignes de contrastes séparant les régions de niveaux de gris différents et relativement homogènes.

Approches structurelles : Dans ces approches, nous trouvons la segmentation par morphologie mathématique [46] et la segmentation par lignes de partage des Eux [47].

Approches basées sur les formes : ces approches tendent à rechercher des régions qui dérivent d'une forme donnée à priori.

Approches basées sur Théorie des Graphes : Les approches basées sur la théorie des graphes consistent à créer un graphe à partir de l'image selon des procédés assez simples et de travailler sur ces graphes par la suite, à l'exemple de la méthode qui utilise les hypergraphes [45].

Approches basées sur les régions : L'approche région de la segmentation utilise des techniques d'identification et de localisation d'ensembles connexes de pixels. Parmi ces méthodes, nous trouvons les méthodes de classification, que nous pouvons séparer en classification supervisée et non-supervisée.

Dans ce travail, nous nous intéressons à la segmentation des images par des méthodes de classification non-supervisée.

4. La Classification Non-Supervisée

Considérons $\mathbf{P} = \{P_1, P_2, \dots, P_n\}$ un ensemble de n vecteurs de d dimensions. Ces vecteurs peuvent être représentés par la matrice $Z_{n \times d}$, où $z_{i,j}$ correspond à la $j^{\text{ème}}$ valeur du vecteur Z_i dans l'ensemble \mathbf{P} et $i = 1, 2, \dots, n$. A partir de $Z_{n \times d}$, un algorithme de classification non-supervisée vise à trouver un partitionnement $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$ tel que la similarité entre les composantes de chaque groupe C_i est maximale tant dis que les composantes des différents groupes se diffèrent le plus que possible. Le partitionnement doit maintenir les propriétés suivantes :

1. Chaque groupe doit contenir au moins un élément. $C_i \neq \phi$
2. Deux groupes différents ne doivent pas contenir un élément en commun. Donc, $C_i \cap C_j = \phi$.
3. Chaque vecteur doit être affectée à un groupe. $\bigcup_{i=1}^k C_i = P$

Depuis que l'ensemble des vecteurs peut être partitionné différemment tout en conservant les propriétés précédentes, une fonction de fitness est définie pour la mesure de la qualité du partitionnement. Ce qui fait, le problème se transforme en un problème de trouver un partitionnement C^* optimal ou proche de l'optimal comparé avec le reste de solutions possible C . Formellement :

$$\text{Optimiser } f(Z_{n \times d}, c)$$

La fonction f est la fonction de fitness qui qualifie la qualité du partitionnement en se basant sur des mesures de similarité. Cependant, Brucker [48] a illustré que le problème de partitionnement est un problème NP-Complet lorsque le nombre de groupes dépasse 3.

5. Mesure de Similarité

Dans la littérature, il existe deux types de mesure. Le premier type est connu par les mesures de dissemblance (DM), où dissimilarity en anglais. Ce type prend une valeur large lorsqu'il s'agit de deux vecteurs différents. Cependant, il existe un second type, qui est les mesures de similarités (SM). Ce type de mesure tend à avoir une grande valeur lorsque les vecteurs sont similaires.

Parmi les mesures de dissemblance les plus connues, nous trouvons la distance Euclidienne [49]. Cette distance entre deux vecteur \vec{Z}_u et \vec{Z}_v est calculée par :

$$d(\vec{Z}_u, \vec{Z}_v) = \sqrt{\sum_{i=1}^d (z_{u,i} - z_{v,i})^2} = \|\vec{Z}_u - \vec{Z}_v\|$$

Parmi les mesures de similarité les plus connues, nous trouvons la distance cosinus [50]. Cette distance mesure la différence d'angle entre deux vecteurs de données. Elle est calculée par :

$$\vec{Z}_i^T \cdot \vec{Z}_j = \frac{\sum_{i=1}^d Z_{u,i} \cdot Z_{v,i}}{\|\vec{Z}_u\| \|\vec{Z}_v\|}$$

6. Mesures de Validité (Fitness)

Une classe de fonctions mathématiques/statistique basées sur les mesures de similarité et dissemblance est employée pour juger la qualité des partitionnements établis par les algorithmes. Cette classe de fonctions est appelée indices de validité. Les indices de validité servent dans deux situations. La première, pour déterminer un nombre optimal de groupes. La deuxième, ils servent à déterminer le meilleur partitionnement des données [51]. L'indice de validité doit prendre en considération les aspects suivants :

1. **Cohésion** : Les éléments d'un groupe doivent être similaires les un aux autre au maximum.
2. **Séparation** : Les groupes doivent être bien séparés.

Dans la littérature, il existe plusieurs indices bien connus tel que L'Indice de Dunn (DI) [52], l'indice Calinski-Harabasz [53], l'indice Davis-Bouldin (DB) [54] et l'indice CS [55]. Dans ce qui suit, nous présentons les formules de calcul de quelques indices:

L'Indice de Dunn (DI): L'indice vise à maximiser les distances intergroupes tout en minimisant les distances intra-groupes. L'indice est défini comme :

$$D = \min_{i=1,2,\dots,k} \left\{ \min_{j=i+1,\dots,k} \left(\frac{d(C_i, C_j)}{\max_{p=1,2,\dots,k} \text{diam}(C_p)} \right) \right\}$$

Où

$$d(C_i, C_j) = \min_{\bar{Z}_r \in C_i, \bar{Z}_s \in C_j} \{d(\bar{Z}_r, \bar{Z}_s)\} \text{ and } \text{diam}(C_p) = \max_{\bar{Z}_u, \bar{Z}_v \in C_p} \{d(\bar{Z}_u, \bar{Z}_v)\}.$$

L'Indice Davis-Bouldin (DB) : Cette fonction représente la portion de la somme des dispersions intergroupes par la séparation entre les groupes. La dispersion dans le $i^{\text{ème}}$ groupe est définie par :

$$S_{i,q} = \left[\frac{1}{N_i} \sum_{\bar{x} \in C_i} \|\bar{Z} - \bar{m}_i\|_2^r \right]^{1/r}$$

La distance entre un groupe i et un groupe j :

$$d_{ij,t} = \left\{ \sum_{p=1}^d |m_{i,p} - m_{j,p}|^t \right\}^{1/t} = \|\bar{m}_i - \bar{m}_j\|_t$$

Où m_i et le centre du $i^{\text{ème}}$ groupe, r and t sont des entiers et q, t peuvent être sélectionnée indépendamment. N_i le nombre des éléments dans $i^{\text{ème}}$ groupe C_i . $R_{i,rt}$ est défini comme :

$$R_{i,rt} = \max_{j \in K, j \neq i} \left\{ \frac{S_{i,r} + S_{j,r}}{d_{ij,t}} \right\}$$

Ce qui fait, l'indice DB est défini par :

$$DB(k) = \frac{1}{k} \sum_{i=1}^k R_{i,rt}$$

Une valeur minimale de DB indique un partitionnement optimal.

7. Classification Non-Supervisée Floue

Comme il a été défini auparavant, La classification non-supervisée consiste à affecter les éléments à un et un seul groupe. Sauf que dans la réalité, il y a des situations où les données n'appartiennent pas qu'à un seul groupe. Ces situations ne peuvent pas être résolues par une classification exacte déterministe. Contrairement à la classification non-supervisée floue, les vecteurs peuvent appartenir à plus d'un groupe. Une valeur d'appartenance indique le degré d'appartenance d'un élément à chaque groupe [56, 57]. Considérons l'exemple suivant :

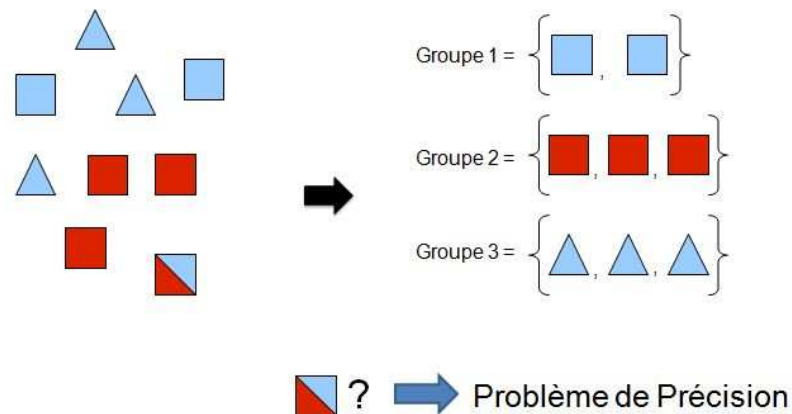


Figure 16 : Imprécision et incertitude de partitionnement

L'exemple montre le cas de partitionnement d'un ensemble d'objets selon leurs formes et leurs couleurs. Nous voyons que le cas du carré rouge et bleu a fait l'exception du partitionnement. Cette situation illustre le cas d'imprécision et d'incertitude liés à l'appartenance de l'objet à plus d'un groupe.

La solution à ce problème réside dans l'introduction de la logique floue. Le degré d'appartenance permet de représenter l'appartenance de l'objet aux différents groupes. Ce qui implique l'introduction d'une matrice de partitionnement $\mathbf{U} = [u_{ij}]_{k \times n}$, où u_{ij} indique l'appartenance du $i^{\text{ème}}$ élément aux $j^{\text{ème}}$ groupe. Le partitionnement flou des données suit les conditions suivantes:

$$0 < \sum_{j=1}^n u_{ij} < n, \quad \sum_{i=1}^k u_{ij} = 1 \quad \text{et} \quad \sum_{i=1}^k \sum_{j=1}^n u_{ij} = n$$

Pour $i = 1, 2, \dots, k$ et $j = 1, 2, \dots, n$.

De même que la classification non-supervisée déterministe, il existe plusieurs indices de validité permettant d'évaluer la qualité des partitionnements flou. Parmi les plus connus, nous citons l'indice J_m et l'indice XB [58, 59].

L'indice J_m a pour but de calculer la variance totale des groupes, cette valeur représente la compacité des groupes. L'indice J_m est donné par l'équation suivante :

$$J_m = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^2 \|\vec{v}_i - \vec{c}_j\|^2$$

La mesure XB calcule le rapport entre la variance totale et la séparation minimale entre les groupes :

$$XB = \frac{\sum_{i=1}^n \sum_{j=1}^k u_{ij}^2 \|\vec{v}_i - \vec{c}_j\|^2}{n \min_{i \neq j} \|\vec{c}_i - \vec{c}_j\|}$$

Sachant que, u_{ij} représente le degré d'appartenance d'un pixel i au groupe j , c_j représente le centre du groupe j , on a :

$$u_{ij} = \sum_{m=1}^k \left(\frac{\|\vec{v}_i - \vec{c}_j\|}{\|\vec{v}_i - \vec{c}_m\|} \right)^{-2}$$

Une valeur minimale des indices J_m et XB indique un partitionnement optimale.

8. Segmentation des données imbriquées

Les distances classiques, telle que la distance euclidienne, employées dans les méthodes de partitionnement, sont efficace lorsqu'il s'agit de la segmentation d'un ensemble de données où les groupes sont linéairement séparables. Cependant, ces distances causent plusieurs erreurs de partitionnement lorsqu'il s'agit de données complexes, où les groupes ne sont pas linéairement séparables. La figure suivante illustre un exemple dans le quel les groupes ne sont pas linéairement séparables :

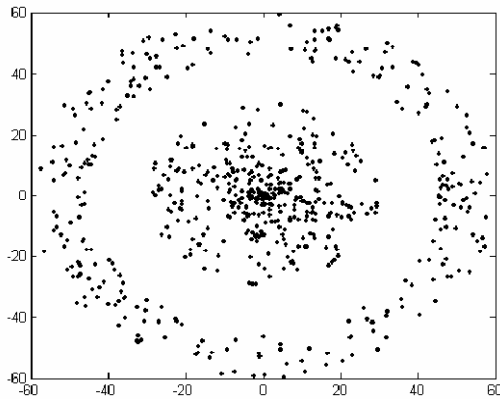


Figure 17 : données non-étiquetées dont les groupes sont non-linéairement séparables

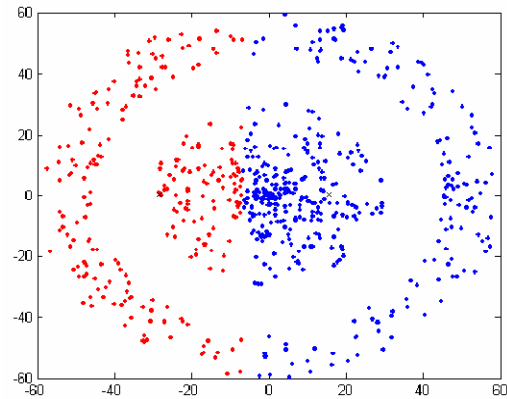


Figure 18 : Séparation par l'algorithme K-means de données non-linéairement séparables

A partir des deux figures précédentes, nous constatons que le partitionnement utilisant la distance euclidienne a échoué à séparer les données. Afin de résoudre ce problème, la première étape consiste à projeter les données dans un espace de dimensionnalité supérieure grâce à des fonctions noyau. Une fonction noyau mesure la distance entre deux points de données par une transformation implicite dans un espace à dimension supérieure où ces données seront linéairement séparables [60]. Ceci permet de conserver la structure des groupes ainsi que la simplifier [60, 61].

Plusieurs méthodes noyau ont été proposées récemment à l'exemple des séparateurs à vaste marge (SVM), où Support Vector Machine en anglais, dans le cas de classification supervisée [62, 63]. Des versions noyau des algorithmes k-means et fuzzy c-means ont été proposées dans [64] et [63].

Dans le but de résoudre le problème des données imbriquées, les mesures de validité sont reformulées afin de s'adapter à ce genre de difficultés. Cette reformulation est établie par introduire une fonction de transformation de l'espace de données initial vers un espace de dimension supérieure.

Supposons une collection de données Z dans un espace de d -dimension et considérons une fonction de transformation non-linéaire de l'espace de recherche initial vers un espace de dimension supérieure H , tel que :

$$\varphi : \mathcal{X}^d \rightarrow H, \bar{Z}_i \rightarrow \varphi(\bar{Z}_i)$$

Où $\bar{Z}_i = [z_{i,1}, z_{i,2}, \dots, z_{i,d}]^T$ et $\varphi(\bar{Z}_i) = [\varphi_1(\bar{Z}_i), \varphi_2(\bar{Z}_i), \dots, \varphi_H(\bar{Z}_i)]^T$

Cette transformation est illustrée dans la figure suivante :

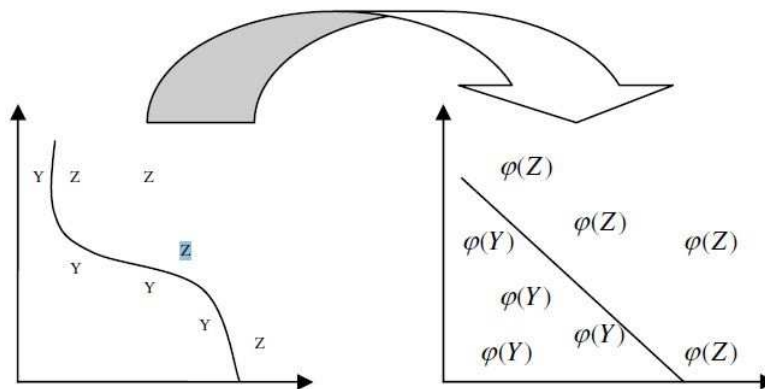


Figure 19 : Transformation d'un espace original vers un espace de dimension supérieur

Revenons à l'exemple précédent et considérons la transformation suivante :

$$\varphi : \mathbb{R}^2 \rightarrow H = \mathbb{R}^3 \quad \text{et} \quad [z_{i,1}, z_{i,2}]^T \rightarrow [z_{i,1}, z_{i,2}, z_{i,1}^2 + z_{i,2}^2]^T$$

Par conséquent, les deux cercles deviendront linéairement séparables comme le montre la figure suivante :

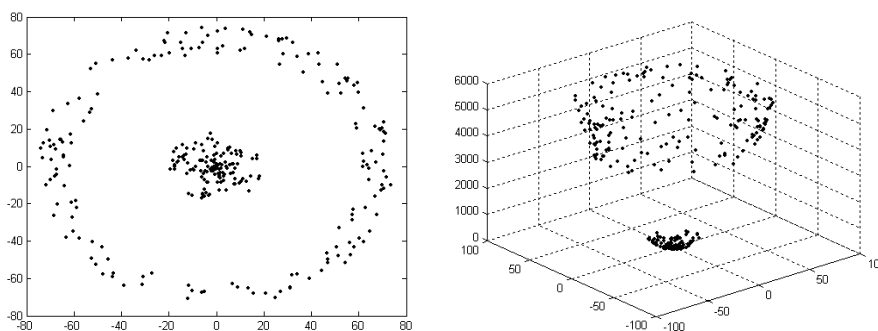


Figure 20 : Projection dans un espace à dimension supérieure

La littérature des fonctions noyaux contient plusieurs fonctions [65] à l'exemple des fonctions :

La fonction Linéaire :

$$K(\bar{Z}_i, \bar{Z}_j) = \bar{Z}_i^T \cdot \bar{Z}_j$$

La fonction Polynomial avec degré :

$$K(\bar{Z}_i, \bar{Z}_j) = (1 + \bar{Z}_i^T \cdot \bar{Z}_j)^p, p \in \mathbb{N}$$

La fonction Gaussienne :

$$K(\bar{Z}_i, \bar{Z}_j) = \exp\left(-\frac{\|\bar{Z}_i - \bar{Z}_j\|^2}{2\sigma^2}\right), \sigma \in \mathfrak{R}$$

Parmi les différentes fonctions noyaux existantes, nous avons choisi de formuler les indices de validité avec la fonction gaussienne dû à ses performances de classification qu'elle a montré par rapport au reste des fonctions [61, 66, 67]. Ainsi, la version noyau de l'indice Jm et XB sont [68]:

$$J_q^\phi = 2^* \sum_{j=1}^n \left\{ \sum_{p=1}^k [1 - K(\bar{Z}_j, \bar{m}_p)]^{1/q} \right\}^{1-q} \quad XB_q^\phi = \frac{J_q^\phi}{n^* \min_{i \neq j} 2^* [1 - K(\bar{m}_i, \bar{m}_j)]}$$

Et

$$u_{ij}^\phi = \frac{[1 - K(\bar{Z}_j, \bar{m}_i)]^{1/q}}{\sum_{p=1}^k [1 - K(\bar{Z}_j, \bar{m}_p)]^{1/q}}$$

9. Algorithmes de Classification

Ces dernières années, plusieurs méthodes de classification non-supervisées sont apparues. L'ensemble de ces méthodes peut être divisé en deux grandes classes, des méthodes classiques et des méthodes bio-inspirées. L'inconvénient des méthodes classiques réside dans leur gourmandise aux ressources matérielles et au temps de calcul, ceci est dû à leur complexité. Ce qui fait, une nouvelle catégorie d'algorithmes est apparue qui est la catégorie des algorithmes bio-inspirés. Ces algorithmes reproduisent le comportement des systèmes naturels afin de résoudre les problèmes.

9.1. Les méthodes Classiques

Parmi les méthodes classiques les plus connues nous citons l'algorithme k-means, et ISODATA :

9.1. 1. L'Algorithme k-means :

Le premier algorithme de classification non-supervisée est le k-means. Il a été proposé par MacQueen en 1967 [69].

Pseudo Code

Entrée : un ensemble d'entrée P d'objets et le nombre de classe k

Sortie : Partitionnement de P en k groupes

Début :

Générer aléatoirement k centres de gravités initiaux

Tant que (Condition d'arrêt = faux)

Pour i allant de 1 à Nombre d'objets :

 Affecter P_i à la classe C_k dont le centre de gravité est le plus proche

 Calculer les nouveaux centres de gravité des classes

Fin Pour

Fin Tant que

Fin

L'inconvénient de cet algorithme réside dans le recalcul des centres de gravités après chaque réaffectation.

9.1. 2. L'Algorithme k-means ISODATA:

L'Algorithme k-means ISODATA est une version modifiée du k-means qui permet de pallier le problème du nombre de classe. En effet, au lieu de donner le nombre de classes en paramètre, l'utilisateur donnera une borne inférieure et une borne supérieure du nombre de classes.

Pseudo Code

Entrée : un ensemble P d'objets le nombre minimal et maximal des groupes.

Sortie : partitionnement en k groupes.

Début :

Pour k allant de k_{min} à k_{max} :

 Exécuter k-means(k).

Fin Pour

 Donner le partitionnement correspondant au nombre de classe ayant la plus petite erreur quadratique.

Fin

9.2. Les méthodes bio-inspirées

Les algorithmes bio-inspirés sont devenus de plus en plus efficaces à la résolution des problèmes de l'optimisation combinatoire [70, 71]. Par conséquent, les méthodes bio-inspirées ont montré de meilleurs résultats par rapport aux méthodes de classification classiques. Cette supériorité est exprimée par la qualité des résultats obtenus ainsi que le temps de calcul nécessaire à la résolution du problème. L'algorithme de l'optimisation par essaim particulaire, où Particular Swarm Optimisation (PSO) en anglais, est un exemple des algorithmes bio-inspirés les plus connus. Développé par Kennedy et Eberhart en 1995 [73], l'algorithme PSO est inspiré du comportement des essaims de poissons et des oiseaux dans la nature. Cette algorithme a été appliqué avec succès dans à la segmentation des images dans plusieurs travaux à l'exemple de [73]. Les algorithmes génétiques [75] représentent une autre méthode d'optimisation et qui a été appliqué sur le problème de la classification des images [74].

Nous notons l'existence d'une diversité d'algorithmes bio-inspirés. Dans cette partie, nous allons introduire le principe de fonctionnement de l'algorithme PSO.

9.2.1. Particular Swarm Optimization

Certains groupes d'animaux (certaines espèces de poissons, des oiseaux, insectes sociaux, etc.) avaient un comportement très complexe alors que les individus qui les composent n'ont accès qu'à des informations limitées et réagissent selon des schémas

assez simple. L'observation des groupes en essaim a permis l'élaboration de la métaheuristique PSO. Cette métaheuristique a été proposée initialement pour l'optimisation des fonctions continues sur \mathbb{R}^d .

Le principe de l'algorithme PSO est de générer aléatoirement N_p particules et associer à chacune d'elle une vitesse V_i appartenant également à l'espace de recherche. L'algorithme est déroulé pour N_b itérations et pour chaque itération t les particules se déplacent dans l'espace de recherche suivante la formule :

$$V_{id}(t+1) = \omega V_{id}(t) + C_1 \phi_1 (P_{lid} - X_{id}(t)) + C_2 \phi_2 (P_{gd} - X_{id}(t))$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1)$$

Où ϕ_1 et ϕ_2 sont des nombres réels aléatoires inférieurs à ϕ_{\max} qui est un paramètre de l'algorithme. ω est le poids de l'inertie. C_1 et C_2 sont des constantes d'accélération et P_{li} est la meilleure position trouvée par la particule i jusqu'à présent et P_g la meilleure solution globale trouvée par une particule jusqu'à présent.

Pseudo Code de classification

Début :

Initialiser toutes les particules avec K centres de gravités aléatoires.

Pour $nb = 1$ à $\max nb$ **faire**

Pour toutes les particules i **faire**

Pour tous les objets o **faire**

 Calculer la distance entre o et tous les centres de gravité

 Affecter o à la classe dont le centre est le plus proche

Fin pour

 Calculer la fonction de fitness f

Fin pour

 Calculer la meilleure position locale pour chaque particule ainsi que la meilleure position globale

 Mettre à jour la position et la vitesse des particules selon les formules du PSO

Fin pour

Fin

10. Conclusion

Dans ce chapitre nous avons présenté un état de l'art des différentes approches de la segmentation des images. Nous avons présenté en détails, l'approche de segmentation basée sur les méthodes de classification non-supervisée. Nous avons aussi cité deux problématiques majeures que nous rencontrons souvent lors de la segmentation des images. Ainsi que quelques méthodes de classification non-supervisée. Dans le chapitre suivant, nous allons présenter la classification non-supervisée par une nouvelle métaheuristique d'optimisation, qui est l'algorithme Firefly. Nous allons aussi présenter nos différentes approches pour la résolution des problèmes liés à la segmentation des images satellitaires.

Chapitre 3

Les Lucioles et L'Algorithme Firefly

1. Introduction

Les méta heuristiques bio-inspirées sont devenues de plus en plus intéressantes grâce à leur efficacité vis-à-vis de la résolution des problèmes de l'optimisation combinatoire [14, 15, 16, 17, 18, 19]. Pendant les deux dernières décennies, Les chercheurs ont commencé à s'intéresser de plus en plus à la manière dont la nature résout les problèmes [20], Ce qui fait, de nouvelles méta heuristiques se sont apparues presque chaque année [13].

Les méta heuristiques sont utilisées afin de résoudre les problèmes d'optimisation combinatoire. En principe, ce type de problème apparaît simple à résoudre du fait que nous connaissons presque souvent comment calculer les différentes solutions possibles. La simulation de chaque solution peut mener à la sélection de la meilleure (l'optimale). L'ensemble de ces solutions est appelé l'espace de recherche. Ce qui fait, le problème d'optimisation consiste à trouver la meilleure solution parmi cet espace. Pouvoir tester toutes les solutions possibles mène à ce que le problème ne sera pas difficile à résoudre. Cependant, lorsque l'espace de recherche devient très large, l'énumération de toutes les solutions ne serait pas possible à cause de la très longue durée que le processus de recherche prendra. Les méta heuristiques représentent l'ensemble des méthodes ayant la possibilité de résoudre ce type de problèmes.

Parmi les algorithmes bio-inspirés les plus connus on trouve les algorithmes évolutifs, inspirés de l'évolution biologique. Un autre exemple d'algorithmes bio-inspirés est les réseaux de neurones, où leur principe de fonctionnement est inspiré du mécanisme de fonctionnement des réseaux de neurones biologiques. Nous citons aussi un des plus connu des algorithmes bio-inspirés, qui est l'algorithme d'optimisation par essaims particuliers (OEP), où Particular Swarm Optimisation (PSO) en Anglais [2, 3, 18]. Introduit par Kennedy et Eberhart en 1995 [21]. Le principe de l'algorithme PSO se base sur le comportement collectif des essaims, tel que les essaims des oiseaux ou de poissons, dans la nature.

L'algorithme Firefly est un algorithme évolutif. Dans ce chapitre nous allons présenter cette méta heuristique bio-inspirée. Pour cela, nous allons commencer par introduire le côté biologique de l'algorithme. Nous allons par la suite décrire le mécanisme de l'algorithme et présenter quelques comparaisons effectuées par l'auteur de cet algorithme avec l'algorithme PSO et les Algorithmes Génétiques.

2. Les lucioles naturelles

Les lucioles (lampyridae), ou Fireflies(en anglais), sont des insectes qui appartiennent à la famille des abeilles. Les lucioles sont dotées d'ailes, leur nom vient du phénomène de bioluminescence. Les lucioles produisent une « lumière froide », sans fréquences infrarouges ou ultra-violetes [26]. Cette lumière chimique est générée à partir de l'abdomen inférieur du corps de ces insectes. La couleur de cette lumière peut être jaune, verte ou rouge pale, avec une longueur d'onde entre 510 à 670 nanomètres. [23]



Figure 21 : Une luciole naturelle

Environ 2000 espèces de lucioles se trouvent dans les régions tropicales. La plupart dans les zones humides couvertes de bois ou les larves de ces insectes trouvent des sources de nourriture. Ces larves, appelées « Glowworms », sont aussi capables d'émettre de la lumière. Pour la plupart des espèces de cet insecte, le male et la femelle sont capables de voler, à part quelques espèces où les femelles sont incapables de le faire. [22]

Les lucioles hibernent lors de l'hiver à l'étape de larve. Après plusieurs semaines d'alimentation, elles émergent en tant qu'adultes. Les larves de la plupart des espèces de cet insecte sont des prédateurs qui se nourrissent sur les autres larves, les escargots et les limaces. Tandis que les autres se nourrissent sur le pollen ou le nectar des plantes. La plupart de ces insectes sont toxique aux autres prédateurs. [24]

La production de la lumière dans ces insectes est due à un type de réaction chimique appelé bioluminescence. Ce processus s'effectue dans un organe spécial, au niveau du corps de la luciole, qui est l'abdomen inférieur de la luciole. L'enzyme Luciférase réagit sur le Luciférien, dans la présence du magnésium, l'ATP et l'oxygène pour produire la lumière.

Pendant l'étape de larve, la bioluminescence sert comme un signal d'alerte aux autres prédateurs, afin de protéger la luciole de ces derniers. Tandis que ce phénomène a une autre utilisation lors de l'âge adulte de cet insecte, les éclats lumineux représentent un moyen de sélection de partenaires [25]. Des espèces de lucioles sont distinguées grâce à leur modèle spécifique des flashes de lumière que les mâles diffusent dans le but de trouver des femelles. D'autre part, il existe une espèce dont les femelles reproduisent les signaux des autres espèces afin d'attirer les mâles de ces espèces comme proies.

L'intensité de la lumière à un point de distance r à partir d'une source de lumière obéit à l'inverse de la loi carrée inverse. Ceci consiste à dire que l'intensité de la lumière I diminue avec la croissance de la distance r , dont le terme est $I \propto 1/r^2$, dû à l'absorption de la lumière par l'air. L'absorption de la lumière par l'air rend les lucioles visibles que pour une distance limitée [31].

3. L'Algorithme Firefly

L'algorithme Firefly est une méta heuristique, bio-inspirée, introduite par Dr Xin-She Yan à l'université Cambridge en 2007. L'algorithme est basé sur le principe d'attraction entre les lucioles et simule le comportement d'un essaim de lucioles dans la nature, ce qui lui donne beaucoup de similarités avec d'autres méta heuristiques basées sur

l'intelligence collective du groupe, tel que l'algorithme PSO (Particle Swarm Optimisation), l'algorithme d'optimisation par colonies d'abeilles (ABC), et l'algorithme des bactéries de fourrages (BFA) [28, 29]. Selon des bibliographies récentes, les performances de l'algorithme Firefly dans la résolution des problèmes d'optimisation dépassent celles des autres algorithmes, tel que les algorithmes génétiques. Ceci a été justifié par des recherches récentes, où les performances de cet algorithme ont été comparées avec celles de quelques algorithmes connus [8, 27, 30].

L'algorithme prend en considération les trois points suivant [27] [28]:

1. Toutes les lucioles sont unisexe, ce qui fait l'attraction entre celles-ci n'est pas en fonction de leur sexe.
2. L'attraction est proportionnelle à leurs luminosités, donc pour deux lucioles, la moins lumineuse se déplacera vers la plus lumineuse. Si aucune luciole n'est lumineuse qu'une luciole particulière, cette dernière se déplacera aléatoirement.
3. La luminosité des lucioles est déterminée en fonction d'une fonction objective (à optimiser)

En se basant sur ces trois règles, l'algorithme Firefly se présente comme suit :

```

Définir une fonction objective  $f(x)$ ,  $x = (x_1, \dots, x_d)^T$ 
Générer une population de lucioles  $x_i$  ( $i = 1, 2, \dots, n$ )
Définir l'intensité de lumière  $I$  à un point  $x_i$  par la fonction objective  $f(x_i)$ 
Déterminer le coefficient d'absorption  $\gamma$ 
Tant que ( $t < \text{Max Génération}$ )
  Pour  $i = 1$  jusqu'à  $n$ 
    Pour  $j = 1$  jusqu'à  $n$ 
      Si ( $I_i < I_j$ )
        Déplacer la luciole  $j$  vers la luciole  $i$ 
      Fin Si
      Varier l'attraction en fonction de la distance  $r$  via  $\exp[-\gamma r]$ 
      Evaluation des nouvelles solutions et mettre à jour l'intensité de lumière
    Fin Pour  $j$ 
  Fin Pour  $i$ 
  Classer les lucioles et trouver la meilleure solution
Fin Tant que
Visualiser les résultats

```

L'algorithme Firefly peut être illustré davantage dans l'organigramme suivant :

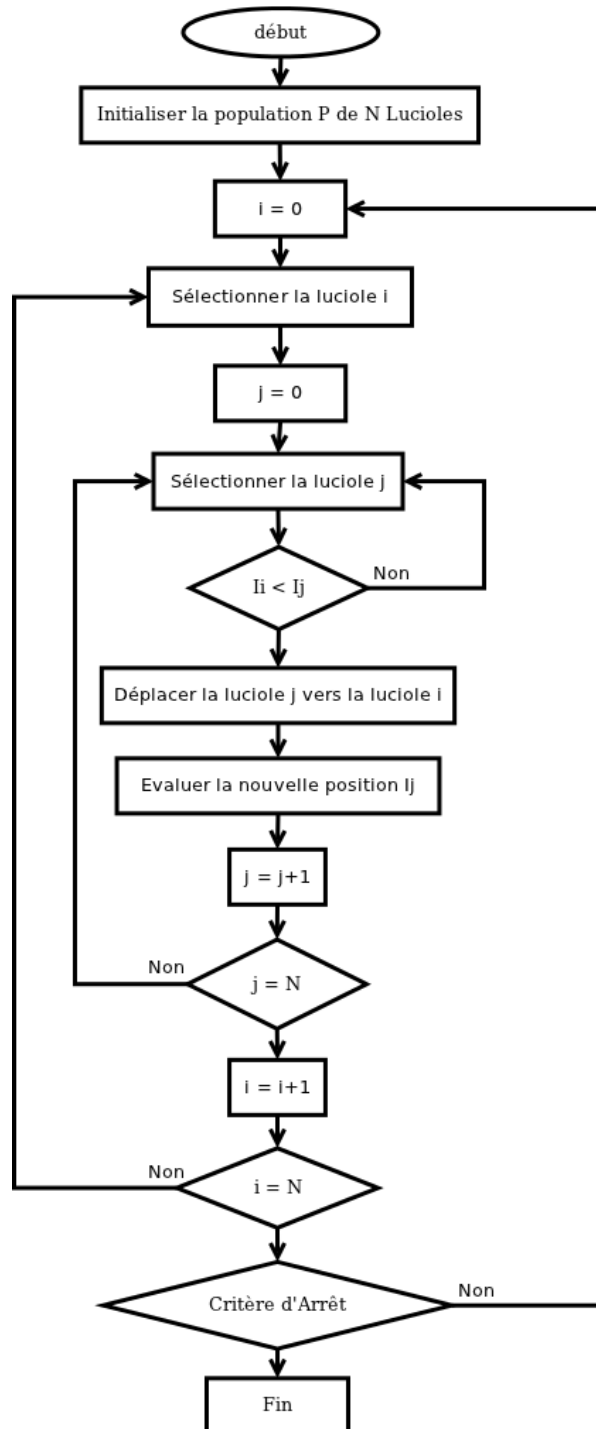


Figure 22 : Organigramme de l'algorithme des lucioles

L'algorithme Firefly est formulé avec deux choses importantes : La variation de l'intensité de la lumière et la formulation de l'attraction. Pour simplifier, l'attraction des lucioles est déterminée en fonction de la luminosité, où la luminosité est déterminée avec la fonction objective.

Dans le cas d'un problème de minimisation, la luminosité I d'une luciole à une position x peut être définie comme $I(x) \propto f(x)^{-1}$. Cependant, l'attraction β est relative à la position des autres lucioles. Par conséquent, elle varie en fonction de la distance r_{ij} entre la luciole i et la luciole j . D'un autre côté, l'intensité de la lumière diminue avec la croissance de la distance par rapport à la source. Ce qui fait que l'attraction peut varier selon le degré d'absorption. Pour simplifier, l'intensité de la lumière $I(r)$ varie en fonction de la loi $I(r) = I_s / r^2$ où I_s est l'intensité à la source. Pour une valeur constante de γ , l'intensité varie en fonction de la distance r , ce qui donne $I = I_0 e^{-\gamma r}$, où I_0 est l'intensité de la lumière de la source. La combine des deux effets de la loi carrée inverse et l'absorption peut être approximer avec la formule Gaussienne suivante [31]:

$$I(r) = I_0 e^{-\gamma r^2}$$

Parfois il est nécessaire d'utiliser une fonction mono tonique décroissante. Dans ce cas, nous pouvons utiliser l'approximation :

$$I(r) = \frac{I_0}{1 + \gamma r^2}$$

Sachant que l'attraction d'une luciole est proportionnelle à l'intensité des lucioles adjacentes, La formule de cette attractivité β d'une luciole peut être définie comme :

$$\beta(r) = \beta_0 e^{-\gamma r^2}$$

Où β_0 est l'attraction à $r = 0$. Pour généraliser, le calcul de $\beta(r)$ est défini par :

$$\beta(r) = \beta_0 e^{-\gamma r^m}, \quad (m \geq 1)$$

D'autre part, la distance entre deux lucioles i et j à des positions x_i et x_j est définie par la distance Cartésienne suivante :

$$r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2},$$

Où $x_{i,k}$ représente le $k^{\text{ème}}$ composant spatiale de la coordonné x_i de la luciole i . Dans un repaire 2-D, la formule r_{ij} devient :

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

Le mouvement d'une luciole i attirée par une autre lucioles j (plus lumineuse que i) est déterminé par :

$$\mathbf{x}_i = \mathbf{x}_i + \beta_0 e^{-\gamma r_{ij}^2} (\mathbf{x}_j - \mathbf{x}_i) + \alpha \left(\text{rand} - \frac{1}{2} \right)$$

Le second terme dans l'équation est dû à l'attraction. Tant dis que le troisième terme rajoute de l'aléatoire à l'équation, où α est aléatoire, *rand* une fonction de génération de nombre aléatoire uniforme dans l'intervalle $[0, 1]$. Le paramètre γ caractérise la variation de l'attractivité, sa valeur est cruciale dans la détermination de la vitesse de convergence et le comportement de l'algorithme.

4. Validation et Comparaisons

Dans cette partie, nous citons quelques résultats obtenus par l'auteur de l'algorithme Firefly. Le document [28] présente des résultats d'application de l'algorithme Firefly sur deux fonctions mathématiques. La première fonction est la fonction d'Ackley :

$$f(\mathbf{x}) = -20 \exp \left[-\frac{1}{5} \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right] - \exp \left[\frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i) \right] + 20 + e$$

Cette fonction a un optimum global $f_* = 0$ à $(0, 0, \dots, 0)$. La figure 23 illustre cette fonction à 2-Dimensions. L'optimum est atteint après 200 évaluations avec 40 lucioles et 5 itérations.

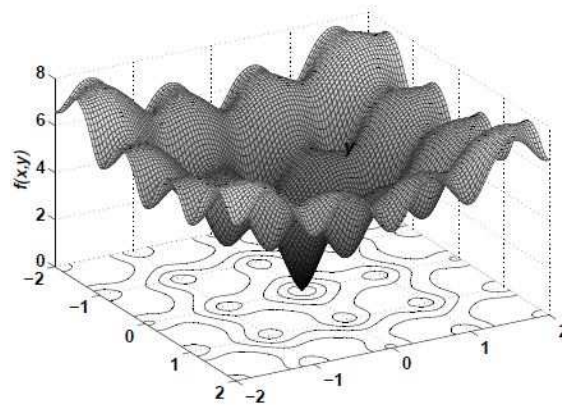


Figure 23 : Représentation de la Fonction d'Ackley à deux variables avec un optimum global à (0, 0).

La deuxième fonction est la fonction Forest (Forêt en français) [32]. Pareil que la fonction d'Ackley, cette fonction a un minimum global à $f_* = 0$ à $(0, 0, \dots, 0)$, sauf que celle-ci a une forme différente que la fonction d'Ackley où sa dérivée n'est pas bien définie au point optimal $(0, 0, \dots, 0)$:

$$f(\mathbf{x}) = \left(\sum_{i=1}^d |x_i| \right) \exp \left[- \sum_{i=1}^d \sin(x_i^2) \right], \quad -2\pi \leq x_i \leq 2\pi$$

La figure suivante illustre une représentation 3D de la fonction :

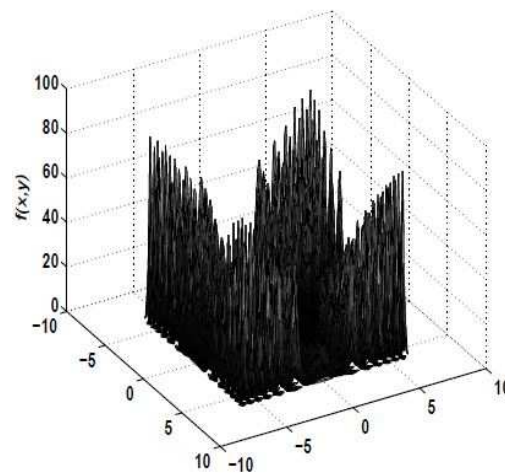


Figure 24 : Représentation de la fonction Forest de Yang proche du minimum (0, 0).

L'algorithme Firefly a été comparé avec deux algorithmes connus : PSO standard [17] et les algorithmes Génétiques (AG). Les tests ont été appliqués sur plusieurs fonctions de

validation [33, 34] (parmi ces fonctions, les deux fonctions d'Ackley et Forest). Les expériences ont été effectuées en prenant une population $n = 40$ pour tous les algorithmes. Tant dis qu'une probabilité de croisement égale à 0.95 et 0.05 comme probabilité de mutation pour l'AG. Chaque algorithme a été déroulé 100 fois. Le critère d'arrêt a été fixé à $\varepsilon \leq 10^{-5}$. Les résultats obtenus sont illustrés dans la table suivante:

Fonction/Algorithmes	GA	PSO	FA
Michalewicz (d=16)	89325 ± 7914 (95%)	6922 ± 537 (98%)	2889 ± 719 (100%)
Rosenbrock (d=16)	55723 ± 8901 (90%)	32756 ± 5325 (98%)	6040 ± 535 (100%)
De Jong (d=256)	25412 ± 1237 (100%)	17040 ± 1123 (100%)	5657 ± 730 (100%)
Schwefel (d=128)	227329 ± 7572 (95%)	14522 ± 1275 (97%)	7923 ± 524 (100%)
Ackley (d=128)	32720 ± 3327 (90%)	23407 ± 4325 (92%)	4392 ± 2710 (100%)
Rastrigin	110523 ± 5199 (77%)	79491 ± 3715 (90%)	12075 ± 3750 (100%)
Easom	19239 ± 3307 (92%)	17273 ± 2929 (90%)	6082 ± 1690 (100%)
Griewank	70925 ± 7652 (90%)	55970 ± 4223 (92%)	10790 ± 2977 (100%)
Yang	37079 ± 8920 (88%)	19725 ± 3204 (98%)	5152 ± 2493 (100%)
Shubert (18 minima)	54077 ± 4997 (89%)	23992 ± 3755 (92%)	9925 ± 2504 (100%)

Table 1 : Comparaison des performances FFA, PSO, AG

D'après les résultats obtenus, nous remarquons que l'algorithme Firefly a montré plus de performances. La désignation 6922 ± 537 (98%) signifie que le taux de réussite est égal à 98%.

5. Conclusion

Dans ce chapitre, nous avons présenté la vie des lucioles dans la nature, ainsi que l'inspiration de l'algorithme Firefly à partir du phénomène d'attraction entre ces insectes dans la nature. Les expérimentations établies par l'auteur de ce nouvel algorithme ont montré l'efficacité de cette méthode vis-à-vis le problème de l'optimisation. Dans le Chapitre qui, nous allons présenter l'utilisation de cet algorithme dans le problème de la segmentation non-supervisée des images satellitaires.

Chapitre 4

Mise en Œuvre

1. Introduction

L'algorithme des lucioles est une méta heuristique évolutive conçue pour résoudre les problèmes d'optimisation combinatoire, tel qu'il a été déjà présenté dans le chapitre précédent. Dans notre approche, nous avons appliqué cet algorithme à la segmentation des images satellitaires, afin de bien pouvoir évaluer les résultats nous avons appliqué l'algorithme d'abord sur des images synthétiques dont on peut distinguer les classes à l'œil nu. Des expérimentations ont aussi été effectuées sur des images aériennes dont les régions sont bien connues et limitées à l'aide des cartes thématiques.

2. Modélisation du Problème

Avant d'appliquer l'algorithme des lucioles à la segmentation des images, il est nécessaire de l'adapter en considérant des étapes de prétraitement, un critère d'arrêt, la structure de lucioles, une fonction objective, et une métrique de mesure de similarité. Les étapes suivantes illustrent le Schéma général de l'application :

a) Prétraitement

Initialement, nous disposons de trois fichiers (un fichier pour chaque canal). Le fichier contient les valeurs radiométriques brutes de l'image. Afin d'augmenter la distinction entre les classes nous avons effectué un rehaussement de contraste par un étalement de l'histogramme de l'image.

b) Composition Colorée :

Le but est de visualiser les données en affectant une couleur à chaque canal de l'image. Dans les images satellitaires, nous avons associé :

- La couleur **bleue** au canal TM1.
- La couleur **verte** au canal TM3.
- La couleur **rouge** au canal TM4.

c) Fixation des paramètres de l'algorithme

L'algorithme des lucioles nécessite la fixation préalable des paramètres cités auparavant, en effet, le choix de la configuration de l'algorithme joue un rôle important dans l'obtention d'une meilleure segmentation des images. Les paramètres tels que la taille de la population, la vitesse d'attraction, le facteur de l'absorption de la lumière par l'air, et le nombre d'itérations sont décisifs.

d) Population et Représentation des Lucioles

L'algorithme Firefly est un algorithme à base de population. En se basant sur d'autres travaux, nous avons associé à chaque luciole dans cette population un vecteur de position dans l'espace de recherche. Une luciole '*i*' va contenir pour chaque groupe '*j*' un vecteur C_{ij} . Tel que, C_{ij} représente le centre le centre du groupe *j* dans la luciole *i*. Par conséquent, la position de la luciole peut être représentée par $x_i = \{C_{i0}, C_{i1}, \dots, C_{ik}\}$, où *k* est le nombre de groupe. Le schéma suivant illustre graphiquement cette représentation :

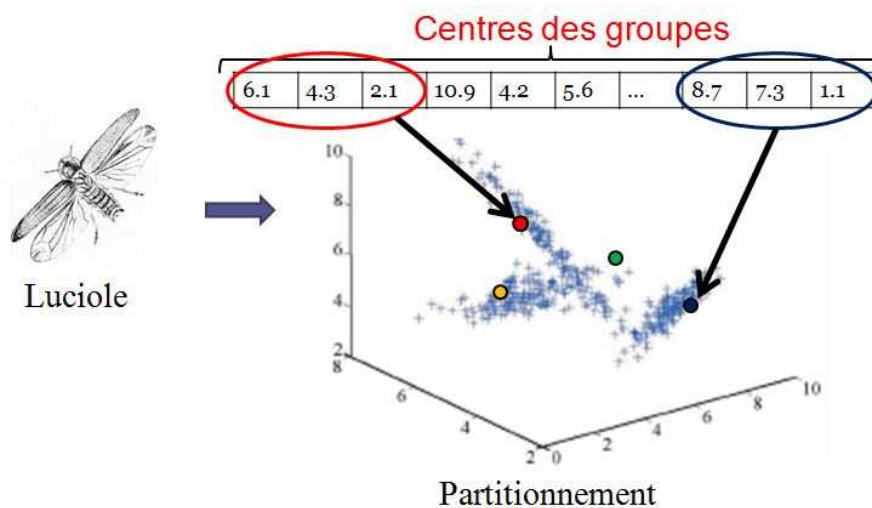


Figure 25 : Vecteur de position d'une luciole.

e) Evaluation des individus

A chaque luciole dans la population, nous associons une valeur indiquant l'intensité de sa lumière, cette valeur dépend de la qualité de la position de cette luciole. Dans nos approches, nous avons utilisé les indices E (erreur quadratique), Jm , XB et DB comme indices de validité afin de déterminer la qualité des lucioles.

Par conséquent, le problème de la segmentation des images est considéré comme le problème de trouver les valeurs des centres de groupes de telle sorte à optimiser les indices de validité, en prenant du fait que les valeurs des centres des groupes doivent être comprises dans l'intervalle $[0, 255]$.

La meilleure segmentation est présentée étant la luciole ayant une intensité de lumière optimale dans l'ensemble des lucioles.

f) La mesure de distances

Dans ce travail, nous avons utilisé deux types de distances :

1. Distance Euclidienne
2. Distance Gaussienne

g) Le Critère d'arrêt

Le critère d'arrêt peut être un nombre d'itération fixe au préalable, une valeur de fonction objective atteinte, ou encore la stagnation du mouvement des lucioles.

2.1 Organigramme de l'algorithme Firefly

L'organigramme ci-dessus illustre et organise les étapes de notre application :

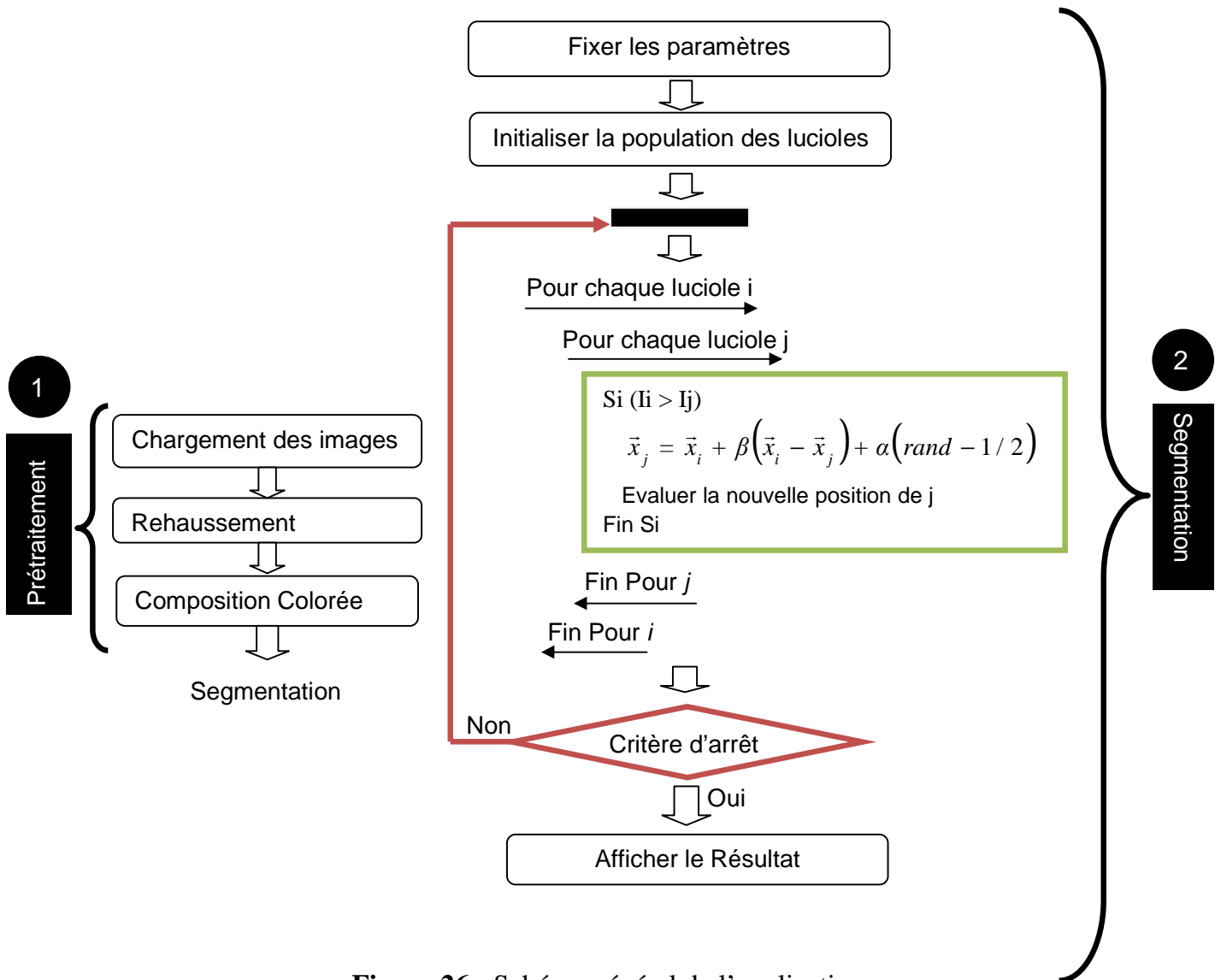


Figure 26 : Schéma général de l'application

Dans ce qui suit, nous expérimentons et nous discutons la résolution du problème de la segmentation des images avec l'algorithme des lucioles afin d'aboutir à une meilleure classification.

2.2 Segmentation par Firefly

Dans ce qui suit, nous présentons le pseudo code de la segmentation des images par classification non supervisée des pixels :

Pseudo code de l'algorithme Firefly

Initialiser la population et l'ensemble des paramètres de l'algorithme

Pour ($t <$ nombre maximal d'itérations)

 Pour chaque solution x_i Faire

 Pour chaque solution x_j Faire

 Si $f(x_i) < f(x_j)$ Alors

- Déplacer x_j vers x_i .

 Pour chaque pixel dans l'image Faire

- Calculer la distance entre le pixel et les centres dans x_j
- Assigner le pixel au groupe avec une distance minimal

 Fin Pour

- Calculer la qualité de la segmentation avec la solution x_j .
- Définir la valeur de l'indice de validité comme une valeur de Fitness à x_j .

 Fin Si

 Fin Pour

Fin Pour

Fin Pour

Visualiser les résultats

2.3. Segmentation par Firefly Flou

Un nombre important des erreurs de classification des images satellitaires est dû à l'affinité des pixels de l'image à plusieurs groupes. Dans une image satellitaire, un pixel représente une région géographique relativement large, ce qui veut dire que cette région

peut contenir plus d'un seul type de terrain. La figure suivante illustre le problème d'imprécision et d'incertitude lié aux images satellitaires :

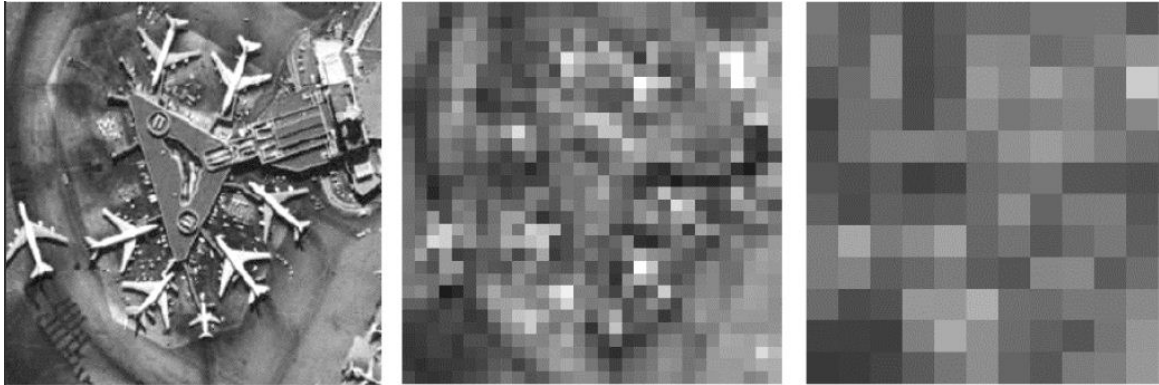


Figure 27. Problème de résolution spatiale des images Satellitaires (à gauche une image à haute résolution, au milieu la même image en résolution moyenne, à droite l'image en basse résolution)

Ce problème est engendré par la résolution spatiale des images. Dans la figure suivante, nous illustrons la problématique avec un schéma :

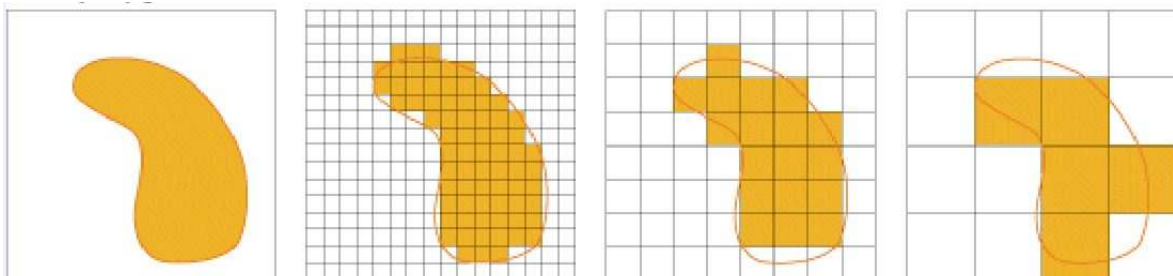


Figure 28. Illustration du d'imprécision dans les Images Satellitaire

A droite de la figure la région à prendre en image. A gauche un schéma illustre l'image mais à basse résolution où nous voyons que les pixels (représentés par des carrés) ne représente pas que l'objet. Pour cela, nous présentons l'algorithme Firefly Flou pour la segmentation des images satellitaires comme suit :

Pseudo code de l'Algorithme Firefly Flou

Initialiser la population et l'ensemble des paramètres de l'algorithme

Pour (t < nombre maximal d'itérations)

 Pour chaque solution x_i Faire

 Pour chaque solution x_j Faire

 Si $f(x_i) < f(x_j)$ Alors

- Déplacer x_j vers x_i .

 Pour chaque pixel dans l'image Faire

- Calculer la distance entre le pixel et les centres dans x_j
- Assigner le pixel au groupe avec une distance minimal
- *Calculer le degré d'appartenance du Pixel*

 Fin Pour

- Calculer la qualité de la segmentation avec la solution x_j .
- Définir la valeur de l'indice de validité comme une valeur de Fitness $f(x_j)$.

 Fin Si

 Fin Pour

Fin Pour

Fin Pour

Visualiser les résultats

2.4 Segmentation par Firefly Gaussien

De plus au problème d'imprécision, la classification des données satellitaires a souvent des problèmes de classes imbriqués dû à la complexité de ce type de données. Dans nos expérimentations, nous avons un cas où les classes de l'image sont imbriquées.

Pseudo code de l'Algorithme Firefly Gaussien:

Initialiser la population et l'ensemble des paramètres de l'algorithme

Pour (t < nombre maximal d'itérations)

 Pour chaque solution x_i Faire

 Pour chaque solution x_j Faire

 Si $f(x_i) < f(x_j)$ Alors

- Déplacer x_j vers x_i .

 Pour chaque pixel dans l'image Faire

- Calculer la distance Gaussienne entre le pixel et les centres dans x_j
- Assigner le pixel au groupe avec une distance minimal

 Fin Pour

- **Calculer la qualité de la segmentation avec la solution x_j en introduisant la distance gaussienne sur l'indice de validité.**
- Définir la valeur de l'indice de validité comme une valeur de Fitness à x_j .

 Fin Si

 Fin Pour

Fin Pour

Fin Pour

Visualiser les résultats

2.5. Segmentation par Firefly Flou Gaussien

Après avoir présenté une approche floue et une approche gaussienne de l'algorithme Firefly, nous avons combiné les deux approches.

Pseudo code de l'Algorithme Firefly Flou Gaussien:

Initialiser la population et l'ensemble des paramètres de l'algorithme

Pour (t < nombre maximal d'itérations)

 Pour chaque solution x_i Faire

 Pour chaque solution x_j Faire

 Si $f(x_i) < f(x_j)$ Alors

- Déplacer x_j vers x_i .

 Pour chaque pixel dans l'image Faire

- Calculer la distance Gaussien entre le pixel et les centres dans x_j
- Assigner le pixel au groupe avec une distance minimal
- *Calculer le degré d'appartenance du Pixel*

 Fin Pour

- **Calculer la qualité de la segmentation avec la solution x_j en introduisant la distance gaussienne sur l'indice de validité.**
- Définir la valeur de l'indice de validité comme une valeur de Fitness à $x_j f(x_j)$.

 Fin Si

 Fin Pour

Fin Pour

Fin Pour

Visualiser les résultats.

3. Images Utilisées :

Afin d'évaluer nos approches, nous les avons appliqués à la segmentation des différents types d'images. Nous présentons les résultats obtenus sur des images synthétiques, satellitaires et aériennes. Les images utilisées sont présentées comme suit :

3.1. Images Synthétiques:

En premier lieu, nous avons appliqué nos approches sur des images synthétiques. L'utilisation de ce type d'images a pour but de pouvoir facilement évaluer le fonctionnement de l'algorithme et nos variantes proposées, et aussi de bien comparer à l'œil les résultats obtenus. Les images synthétiques utilisées sont illustrées dans les figures suivantes:

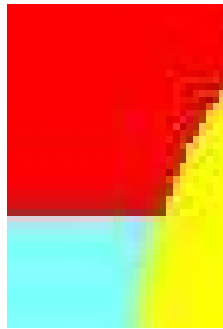


Figure 30 : Image Synthétique 1 (image légèrement bruitée)

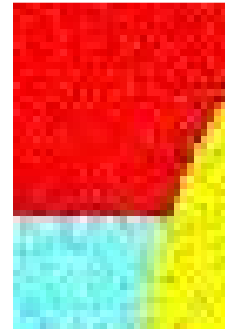


Figure 31: Image Synthétique 2 (Image considérablement bruitée)

Nous avons créé nous mêmes les deux images synthétiques proposées avec un outil de dessin. Les images comportent trois régions de couleurs différentes, dont la différence est facilement distinguable à l'œil. Les images présentent un bruit beaucoup plus remarquable sur la deuxième image synthétique que sur la première, ainsi que des zones floues sur les frontières des régions.

3.2 Image Aérienne:

L'image aérienne utilisée couvre des fermes d'agriculture à l'Est de Nebraska aux USA. Prise lors de l'été, pendant la saison de croissance. L'image contient de champs de Maïs, de blé, de Soja, Ainsi que des arbres et une petite ferme.



Figure 32: Image Aérienne

Nous avons obtenu l'image depuis l'internet. Le choix de cette image est dû à la disponibilité d'une carte thématique Figure 33, présentée dans la figure suivante:

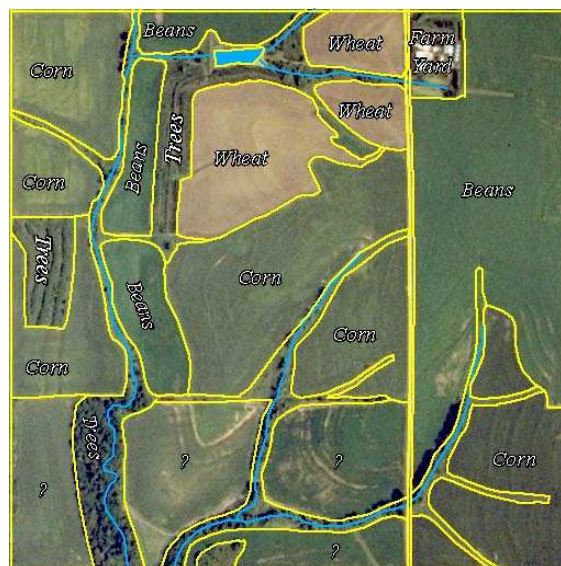


Figure 33: Carte Thématique de l'image aérienne

3.3 Images Satellitaires:

Les images satellitaires que nous avons utilisées représentent la région d'Oran (Algérie), à l'ouest d'Alger. La spécificité de la région consiste à ses variété de thèmes (zones urbaines, eau, céréaliculture, ressac, etc.). Le relief de la région engendre des confusions

au niveau des images satellitaires de cette zone. Ces confusions sont intéressantes pour la problématique posée et l'évaluation de l'efficacité des méthodes de reconnaissance.



Figure 34: La région d'Oran

Deux images satellitaires, prises à deux dates différentes et par deux satellites différents, de cette région sont utilisées dans nos expérimentations. La figure 35 représente une image satellitaire de la région Oran Ouest prise par le satellite LANDSAT7 TM (Thematic Mapper) du 15 Mars 1993, à partir des trois canaux TM1, TM3 et TM4. La figure 36 présente une image de la même région prise en 2001 par le satellite SPOT4 de résolution 20m, à partir des trois canaux spectraux (Proche infrarouge, Rouge, Vert).

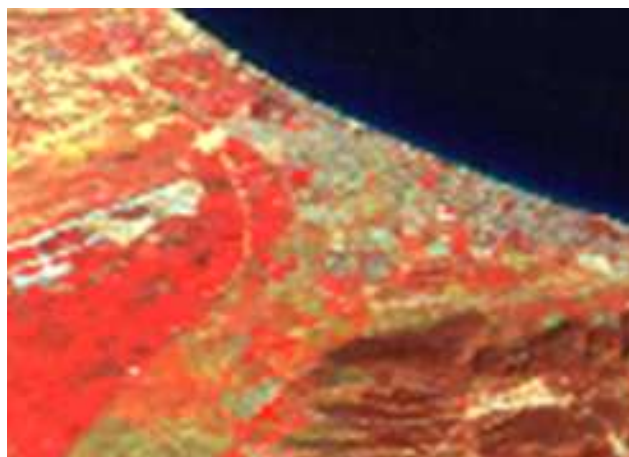


Figure 35: Image Satellitaire Oran Ouest datée de 1993

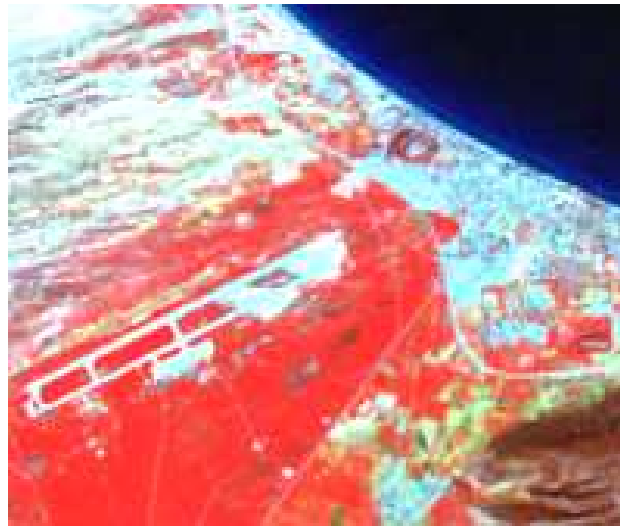


Figure 36: Image Satellitaire Oran Ouest datée de 2001

4. Matériel et Environnement Utilisé :

Notre application a été réalisée avec le langage de programmation Java. Le choix de ce langage a été fait pour sa flexibilité, son efficacité, ainsi que sa facilité pour manipuler les données. Le programme a été exécuté sur un micro-ordinateur ayant les caractéristiques suivantes :

- Microprocesseur Intel Core 2 Duo (1.8 GHz).
- 2Go de Mémoire RAM.
- Système d'exploitation Linux (ubuntu).

L'exécution de notre application a été exécutée aussi sur le système d'exploitation Windows (Windows Seven).

Le code source a été réalisé d'une façon modulaire offrant plus de flexibilité et extensibilité au programme. Nous avons respecté les notions de l'orienté objet, et suivi des design patterns tel que le modèle singleton, et le modèle observer. Ci-dessus le diagramme de classe illustrant l'architecture de notre code source:

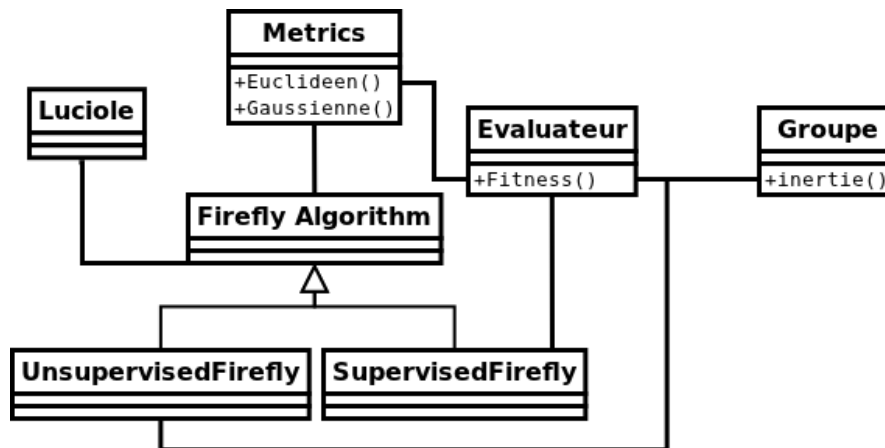


Figure 37: Schéma UML de l'application

5. Discussion des résultats

Dans nos expérimentations, nous appliquons l'algorithme Firefly et nos variantes proposées de cet algorithme sur les images de tests présentées auparavant. Dans chaque essai, nous appliquons une approche de l'algorithme et nous varions les indices de validité afin de comparer les résultats en prenant compte de l'effet de ces indices sur la qualité de la segmentation. Les résultats obtenus sont présentés dans les sections qui suivent :

5.1 Images synthétiques

Dans les essais sur les images synthétiques, Nous testons le fonctionnement et l'efficacité de l'algorithme Firefly et les variantes proposées de cet algorithme sur le problème de la segmentation non-supervisée des images synthétiques bruitées. Nous appliquons l'algorithme en variant ces paramètres, ainsi que le nombre maximal des classes de l'image. Les tests sont appliqués avec un nombre de groupe égal à 3, 4 et 5. Les résultats obtenus sont présentés dans ce qui suit.

5.1.1. Image Synthétique 1

Nous avons utilisé l'image synthétique 1 qui représente une image simple. Les régions de cette image peuvent être facilement séparées visuellement.

Essai 1 : L'algorithme Firefly

Dans cet essai, Nous appliquons l'algorithme Firefly avec variation de la taille de la population n , le nombre d'itération t et le paramètre γ de l'algorithme. Nous avons choisi l'indice Jm pour évaluer la qualité des segmentations. L'algorithme Firefly est déroulé 10 fois pour chaque configuration des paramètres. Les meilleurs résultats obtenus sont illustrés dans la table suivante:

Test	k	n	t	γ	β	Jm	Temps (ms)
1	3	2	5	1	1	2.67×10^7	15
2	3	5	5	1	1	1.79×10^7	141
3	3	5	10	1	1	1.42×10^7	249
4	3	5	10	0.5	1	1.30×10^7	234
5	3	5	10	0.05	1	1.14×10^7	249
6	3	5	10	0.005	1	1.36×10^6	234

Table 2. Evaluation de la segmentation d'image synthétique avec différents paramètres de l'algorithme Firefly.

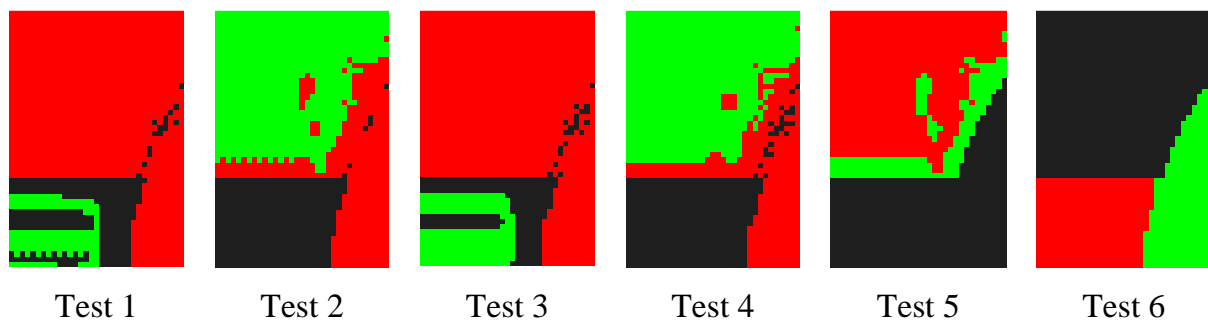


Figure 38. Résultats visuels des testes avec différents paramètres de l'algorithme Firefly

Discussion

Le 6^{ème} test représente le meilleur résultat obtenu par l'algorithme Firefly. La valeur de l'indice Jm est la valeur la plus minimale par rapport au reste des résultats. Le résultat visuel est satisfaisant, nous remarquons que l'algorithme a bien séparé les régions de l'image.

D'après cet essai, nous concluons que l'algorithme Firefly peut être utilisé dans la segmentation des images.

Essai 2 : L'Algorithme Firefly Flou

Dans cet essai nous appliquons l'approche Firefly Flou sur l'image synthétique 1. Le but est de montrer l'apporte de cette approche en montrant le degré d'appartenance des pixels. L'image de la figure 39 montre une segmentation floue de l'image synthétique 1. Les pixels ayant un degré d'appartenance inférieur à 0.8 sont colorés en noir.

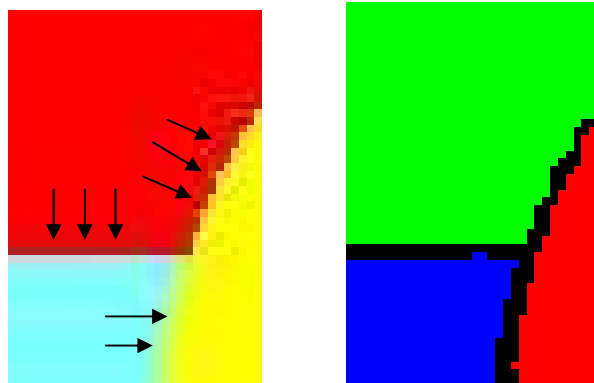


Figure 39 : Segmentation Floue de l'image Synthétique 1

Nous remarquons que l'algorithme a détecté les zones floues sur l'image synthétique 1.

5.1.2. Image Synthétique 2

Dans cette section, nous présentons l'application de l'algorithme Firefly et les variantes proposées à la segmentation de la deuxième image synthétique figure 31. L'objectif est de tester la robustesse de la méthode vis-à-vis le bruit dans les images, et observer le comportement de nos approches concernant un nombre élevé de groupes par rapport au nombre réel que l'image contient. Dans chaque essai, nous utilisons une approche avec plusieurs indices de validité et nous augmentant le nombre de groupes de 4 à 5. Nous avons déroulé les algorithmes avec $n = 10$, $t = 20$, $\gamma = 0.005$ et $\beta = 1$. Les résultats obtenus se présentent comme suit :

Essai 1: L'Algorithme Firefly

L'indice	n = 4	n = 5
E	1184.06	1475.52
Jm	1.914×10^6	2.7108×10^6
XB	45.11	135.45
DB	0.5319	1.3051

Table 3. Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly avec $n = 4$ et $n = 5$.

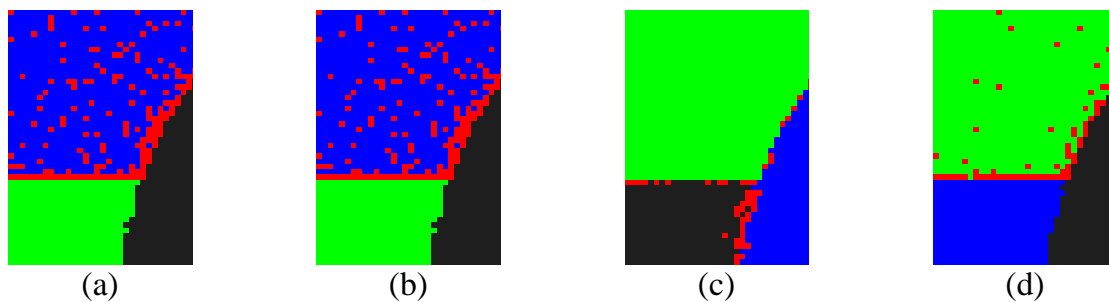


Figure 40. Résultats Visuels de la segmentation de l'image synthétique 2 avec l'algorithme Firefly et $n = 4$.

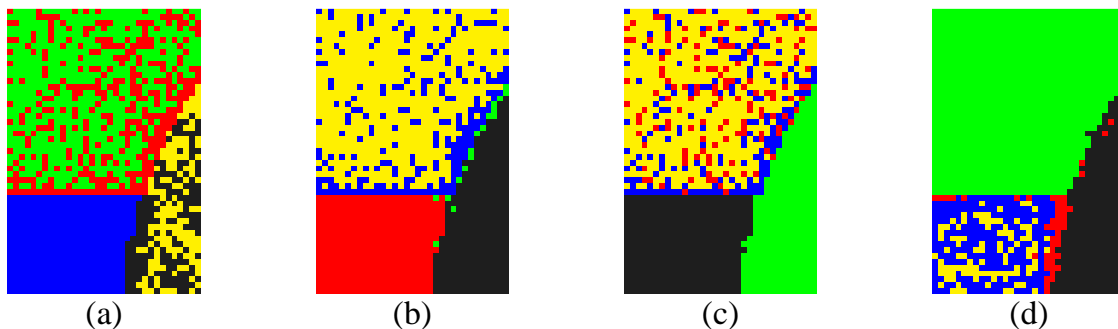


Figure 41. Résultats Visuels de la segmentation de l'image synthétique 2 avec l'algorithme Firefly et $n = 5$.

Discussion

Nous remarquons que tous les résultats contiennent des pixels mal segmentés. Le nombre de ces pixels est beaucoup plus considérable dans les tests avec les indices E et DB . Le nombre de pixels mal classés augmente aussi avec l'augmentation du nombre des groupes de 4 à 5. Dans les essais qui suivent, nous allons utiliser les indices Jm et XB .

Essai 2: L'Algorithme Firefly Flou

L'indice	n = 4	n = 5
Jm	1.853×10^6	1.358×10^6
XB	5.90	9.5914

Table 4. Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec n = 4 et n = 5.

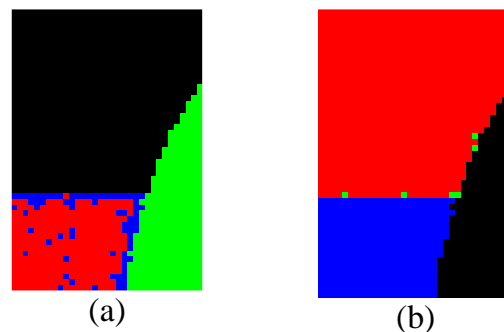


Figure 42. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec n = 4

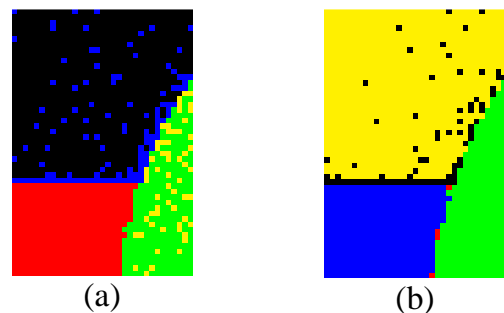


Figure 43. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou avec n = 5

Discussion

La segmentation de l'image synthétique 2 avec l'algorithme Firefly flou a permis d'améliorer les résultats obtenus. La différence par rapport à la segmentation avec l'algorithme Firefly est remarquable à l'œil.

L'indice XB nous a donné de meilleurs résultats dans cet essai. Nous remarquons qu'il y a moins de pixels mal classés dans le cas de n = 4 et n = 5. Ceci, nous laisse conclure que l'introduction de la logique flou sur l'algorithme Firefly a donné plus de robustesse à cet algorithme vis-à-vis le bruit et le nombre élevé de groupes.

Essai 3: L'Algorithme Firefly Gaussien

Nous avons pris γ de la distance gaussienne égale à 1000 dans les expérimentations avec l'algorithme Firefly gaussien.

L'indice	n = 4	n = 5
Jm	0.0058	0.0064
XB	0.00446	0.00490

Table 5. Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec n = 4 et n = 5.

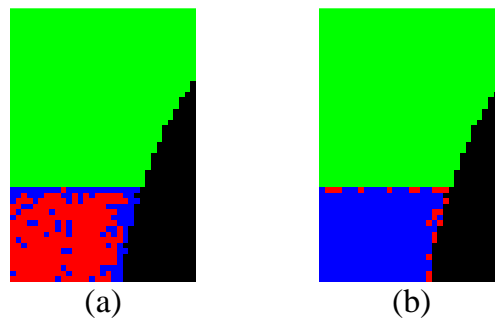


Figure 44. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec n = 4

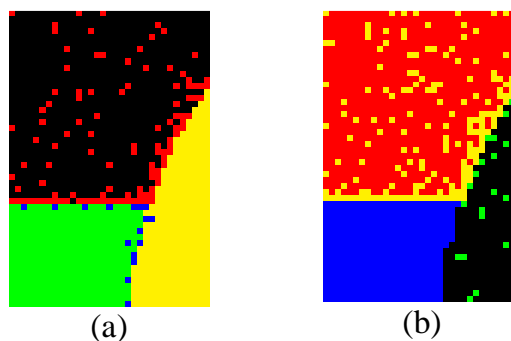


Figure 45. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Gaussien avec n = 5

Discussion

Les résultats obtenus par l'approche gaussienne de l'algorithme Firefly sont légèrement inférieurs à ceux obtenus par l'algorithme Firefly Flou. Par contre, l'algorithme Firefly Gaussien a montré une stabilité dans la segmentation. Nous avons remarqué de petits changements dans la valeur de l'indice de validité d'un teste à un autre.

Essai 4: L'Algorithme Firefly Flou Gaussien

Nous avons pris γ de la distance gaussienne égale à 1000 dans les expérimentations avec l'algorithme Firefly gaussien.

L'indice	n = 4	n = 5
Jm	0.0051	0.0044
XB	0.000017	0.000021

Table 6. Résultats de segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec n = 4 et n = 5.

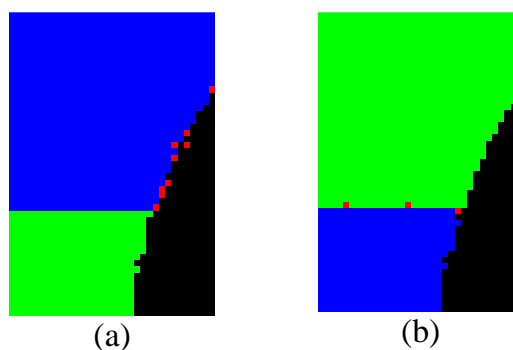


Figure 46. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec n = 4

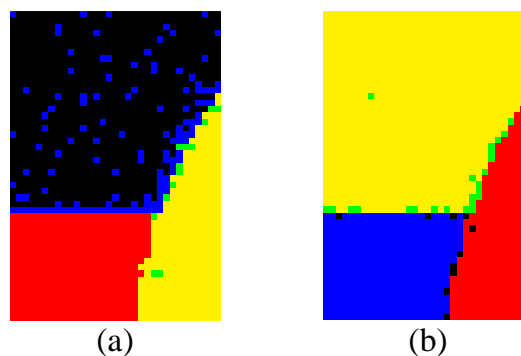


Figure 47. Résultats visuels de la segmentation de l'image synthétique 2 par l'algorithme Firefly Flou Gaussien avec n = 5

Discussion

Nous remarquons que l'indice XB a donné les meilleurs résultats. Comparer aux testes précédents, L'approche Firefly Flou Gaussienne a montré une efficacité supérieure aux approches appliquées précédemment. Ceci peut être vérifié à l'œil, où nous voyons qu'il

y a très peu de pixel mal classé sur l'image segmentée en 4 groupes figure 46 (b), ainsi que sur le résultat à 5 groupes figure 47 (b).

5.2. Image Aérienne

Dans ce qui suit, nous allons utiliser l'algorithme Firefly et les approches proposées à la segmentation de l'image aérienne de la figure 32. Les algorithmes sont appliqués afin de segmenter l'image en 6 groupes (Corn 1, Corn 2, Beans, Wheat, Farm, Trees). Les résultats sont comparés avec la carte thématique du terrain (figure 33). Les algorithmes sont appliqué avec $n = 10$, $t = 30$, $\gamma = 0.005$ et $\beta = 1$.

Essai 1: L'Algorithme Firefly

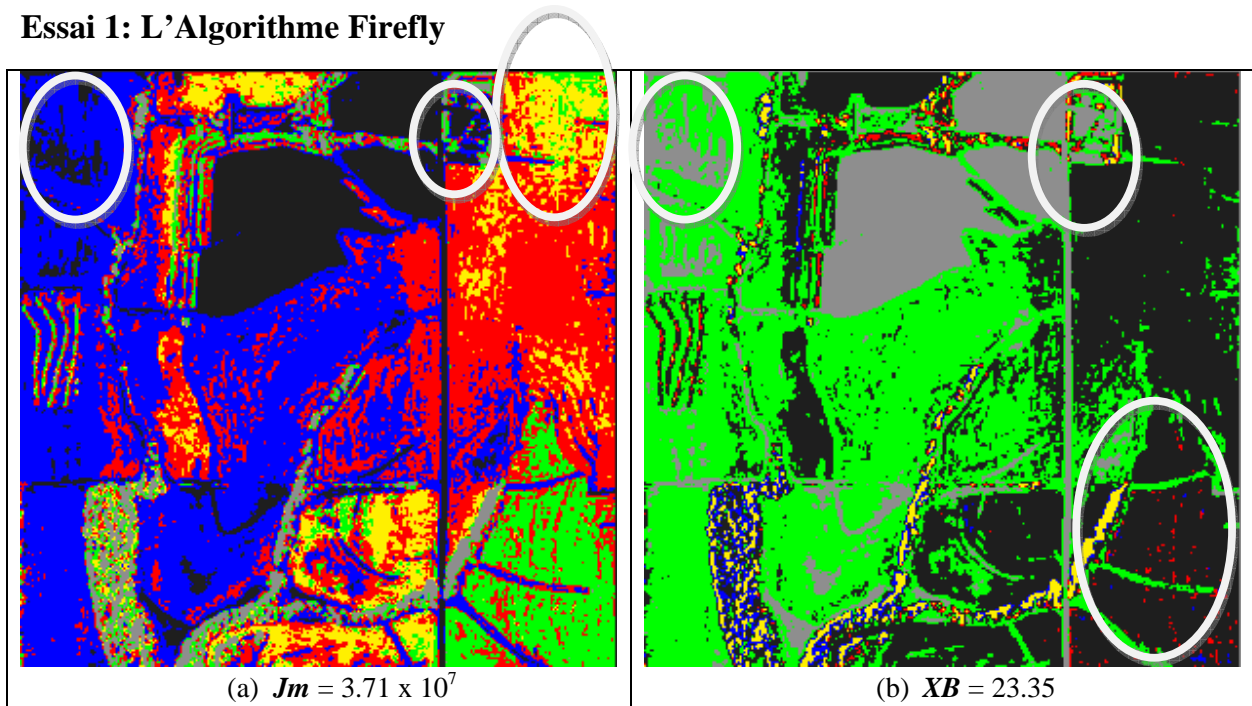


Figure 48 Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly.

Discussion

La figure 48(a) représente le meilleur résultat obtenu. L'algorithme Firefly appliqué avec l'indice Jm a donné de meilleurs résultats comparé à l'indice XB . Nous remarquons des confusions au niveau de la classe Beans et Corn sur la figure 48 (a). Tant dis que nous remarquons plus de confusion au niveau de la classe dans le résultat obtenu pas

l'algorithme Firefly et l'indice Jm , illustré dans la figure 48 (b). La classe Farm (Urbain) n'a pas été segmentée.

Essai 2: L'Algorithme Firefly Flou

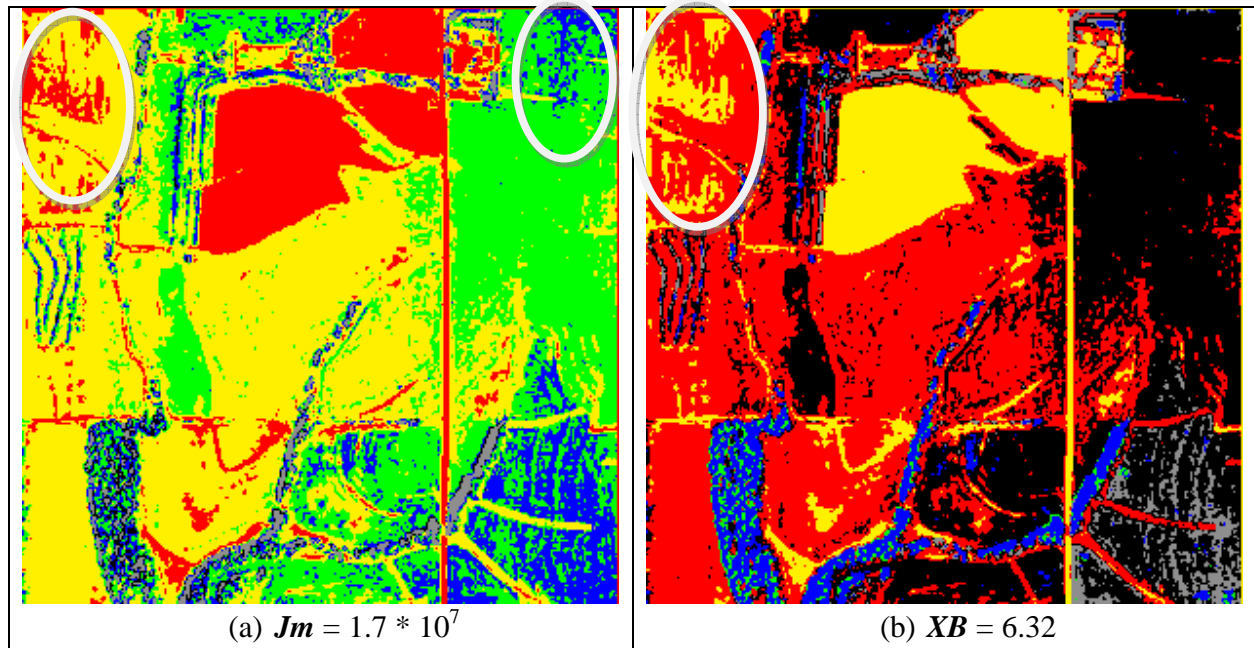


Figure 49 Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Flou.

Discussion

Nous remarquons que l'algorithme Firefly flou a donné des résultats supérieurs à l'algorithme Firefly. Ceci est justifié par l'absence par le nombre de confusions obtenues dans les résultats. A l'œil, nous voyons que l'algorithme a toujours ignoré la classe urbaine (Farm). Tant dis qu'il y a moins de confusion au niveau de la casse Corn. Nous remarquons que les résultats obtenus par l'indice XB ont été meilleur par rapport à ceux obtenus dans l'essai précédent. Nous remarquons aussi que l'indice Jm a donné de meilleurs résultats comparé à l'indice XB.

Essai 3: L'Algorithme Firefly Gaussien

Dans cet essai nous avons fixé le paramètre γ de la distance gaussienne à 50.

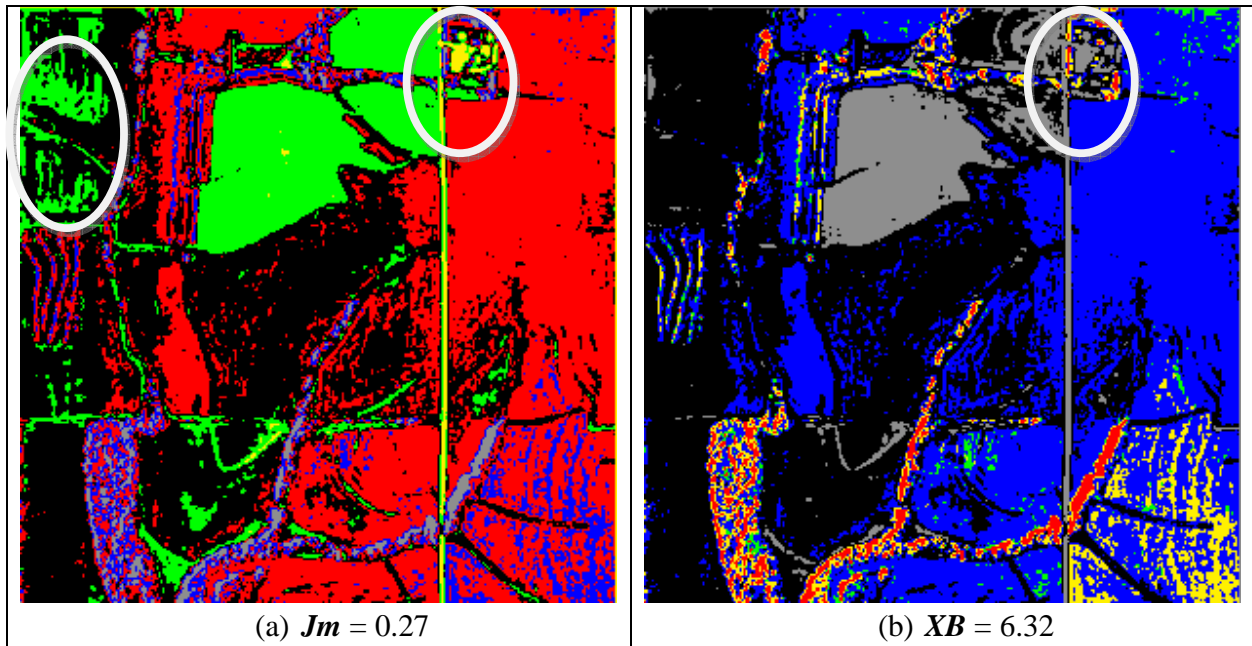


Figure 50. Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Gaussien.

Discussion

Dans les résultats de la segmentation avec l'algorithme Firefly Gaussien, nous notons que l'algorithme a détecté la classe Farm (urbain) en l'application avec l'indice Jm (figure a). Comparés aux résultats précédents, les résultats de l'algorithme Firefly flou gaussien sont légèrement supérieurs à ceux obtenus par l'algorithme Firefly Flou.

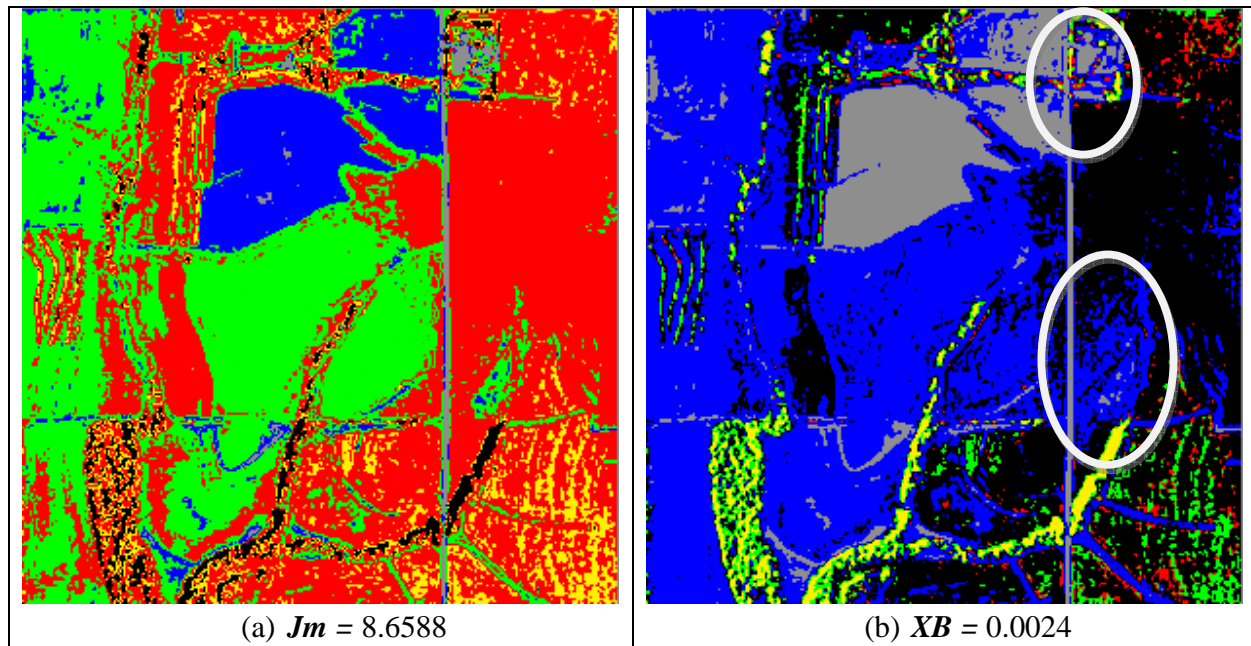
Essai 4: L'Algorithme Firefly Flou Gaussien

Figure 51. Résultats de la segmentation de l'image aérienne avec l'algorithme Firefly Flou Gaussien.

Discussion

Les résultats de cet essai expriment l'efficacité de l'approche Firefly Flou Gaussien. Nous remarquons que le nombre de confusions a considérablement diminué par rapport aux essais précédents. De plus, nous observons que l'approche Firefly Flou Gaussien utilisée avec l'indice Jm a donné de meilleurs résultats par rapport à ceux obtenus par l'indice XB . Les classes dans la figure (a) ont été bien séparées, ou nous voyons que l'algorithme a pris en considération la classe Farm aussi bien que les autres classes de l'image.

5.3. Images Satellitaires

Pour mieux évaluer la performance de nos approches, dans cette section, nous allons l'appliquer sur les deux images satellitaire figure 35 et figure 36 de la région d'ORAN.

L'image satellitaire doit subir un prétraitement, cependant nous avons commencé notre application par le chargement des trois images correspondantes aux trois canaux. Nous effectuons un rehaussement de contraste puis une composition colorée en associant les couleurs bleu, vert et rouge aux trois canaux :

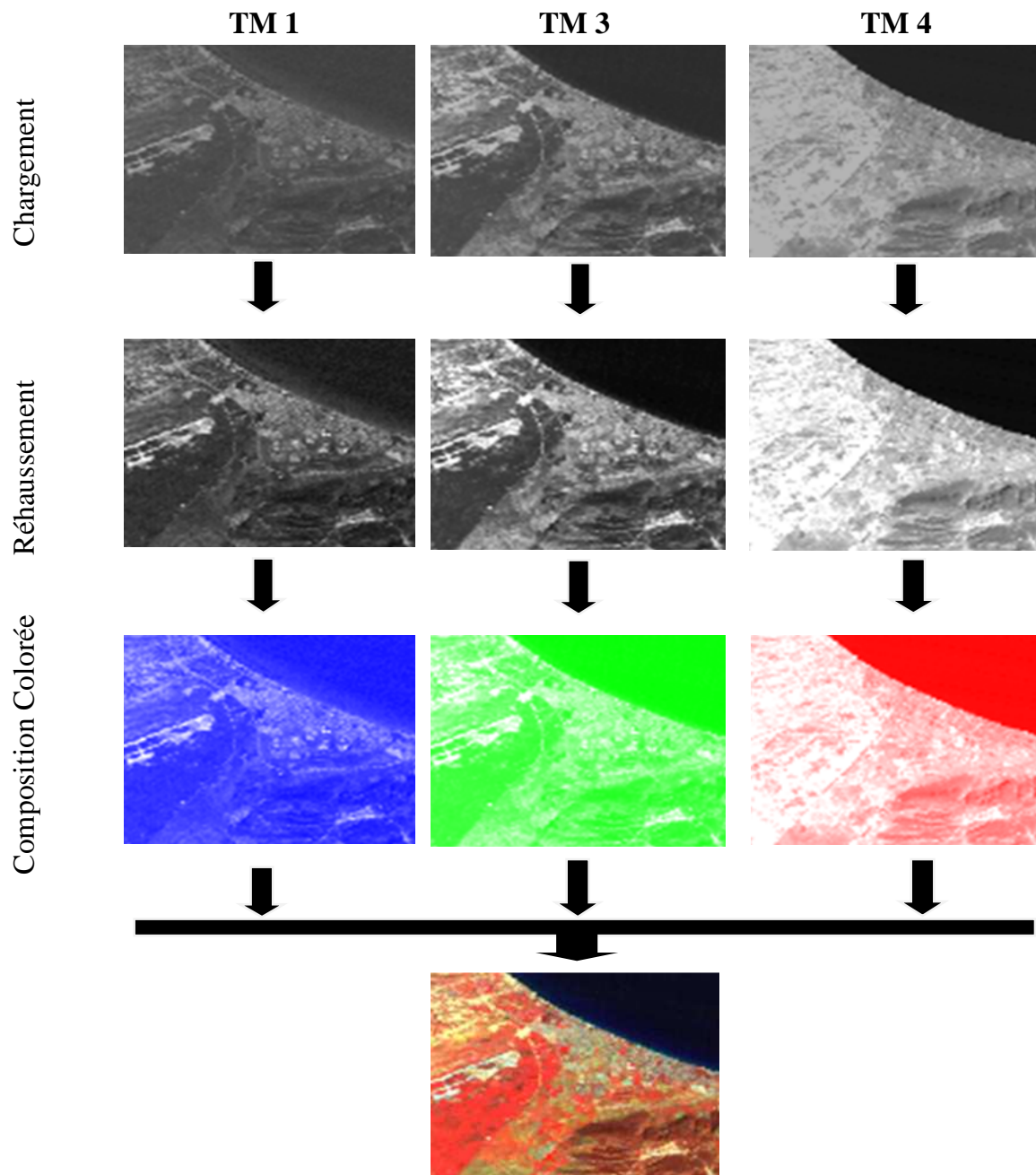


Figure 52 : Prétraitement de la première image satellitaire

Après l'obtention de l'image prétraitée, nous passons à l'application des algorithmes. De la même façon, nous allons d'abord choisir les paramètres de l'algorithme Firefly, ensuite nous allons comparer l'algorithme avec les approches que nous avons proposé. Dans ce qui suit, nous discutons les résultats obtenus dans chaque essai.

Note : Dans les essais qui suivent, nous avons utilisé les paramètres suivant : $n = 20$, $t = 30$, $\gamma = 0.0005$ et $\beta = 1$.

Essai 1. Algorithme Firefly

L'indice	Jm	XB
<i>Image Sat 1</i>	$3.076 * 10^7$	95.9921
<i>Image Sat 2</i>	$3.572 * 10^7$	61.3394

Table 7 : Résultats de la segmentation des images satellitaires avec l'algorithme Firefly

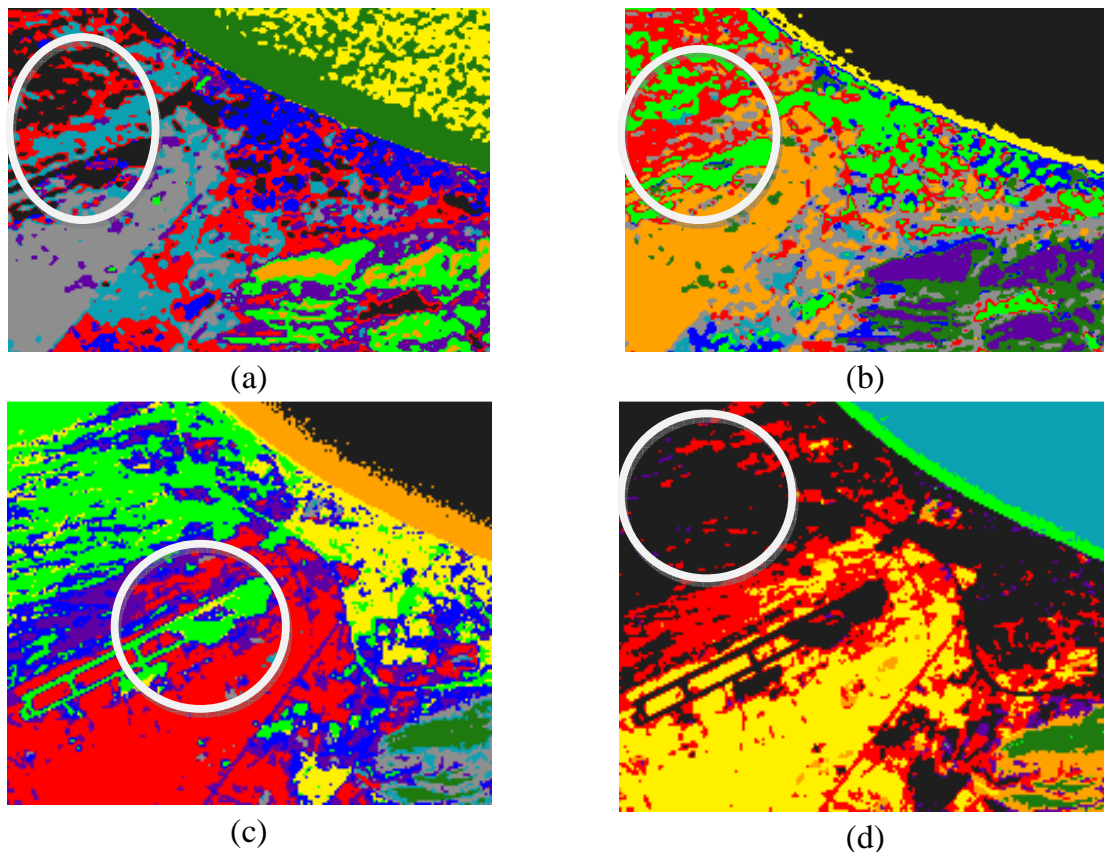


Figure 53 : Résultat Segmentation des images Satellitaires par l'algorithme Firefly.

Discussion

Dans cet essai, nous avons déployé l'algorithme Firefly à la segmentation des images satellitaires. Nous remarquons que l'algorithme a bien fonctionné sur ce type d'images. Par contre, nous observons l'occurrence de quelques confusions entre des types de terrains dans les résultats, tel que l'urbain et le sol nue.

Essai 2. Algorithme Firefly Flou

L'indice	Jm	XB
<i>Image Sat 1</i>	1.338×10^7	9.8582
<i>Image Sat 2</i>	1.560×10^7	10.5263

Table 8. Résultats Segmentation des Images Satellitaires avec l'algorithme Firefly Flou

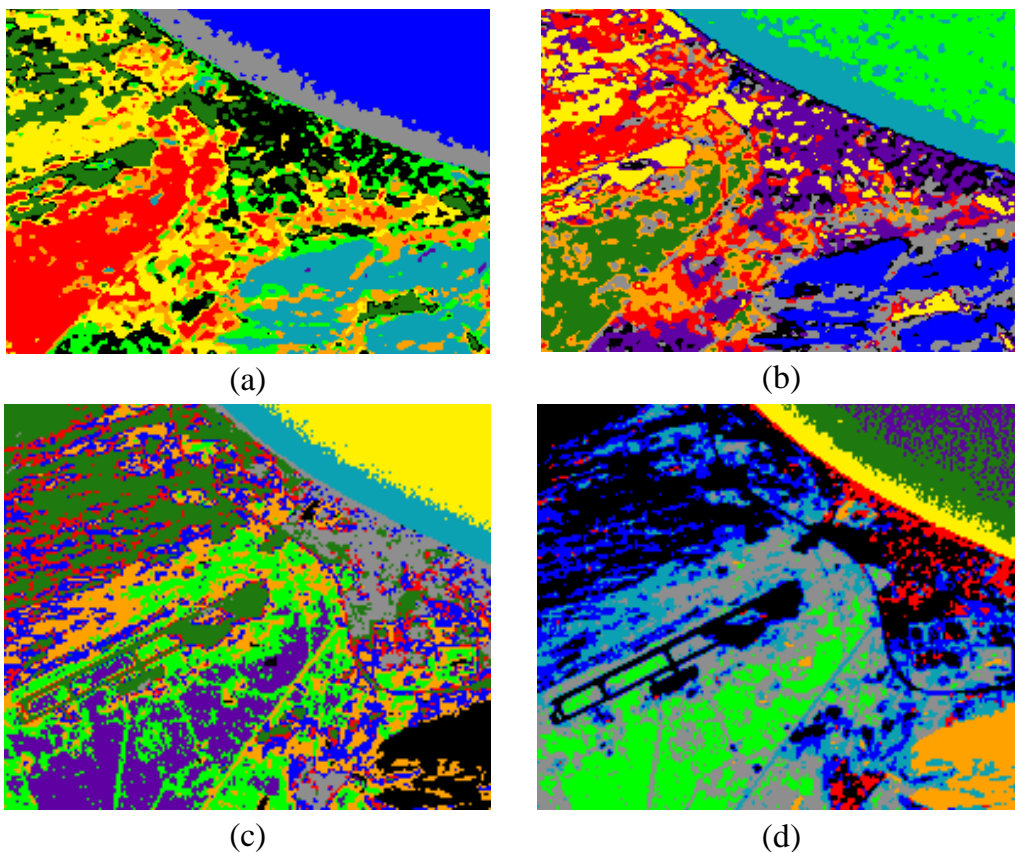


Figure 54 : Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Flou

Discussion

La segmentation des deux images satellitaires avec l'algorithme Firefly flou a donné de résultats légèrement supérieurs à ceux obtenus par l'algorithme Firefly. Ceci est exprimé par la diminution du nombre de confusions sur les images par rapport aux résultats de l'essai avec l'algorithme Firefly. La confusion entre la classe urbain et sable/sol nu a diminuée.

Essai 3. Algorithme Firefly Gaussien

Dans cet essai nous avons fixé le paramètre γ de la distance gaussienne à 1000.

L'indice	Jm	XB
<i>Image Sat 1</i>	0.0285	0.0472
<i>Image Sat 2</i>	0.0067	0.0032

Table 9. Résultats de la segmentation des images satellitaires avec l'algorithme Firefly Gaussien

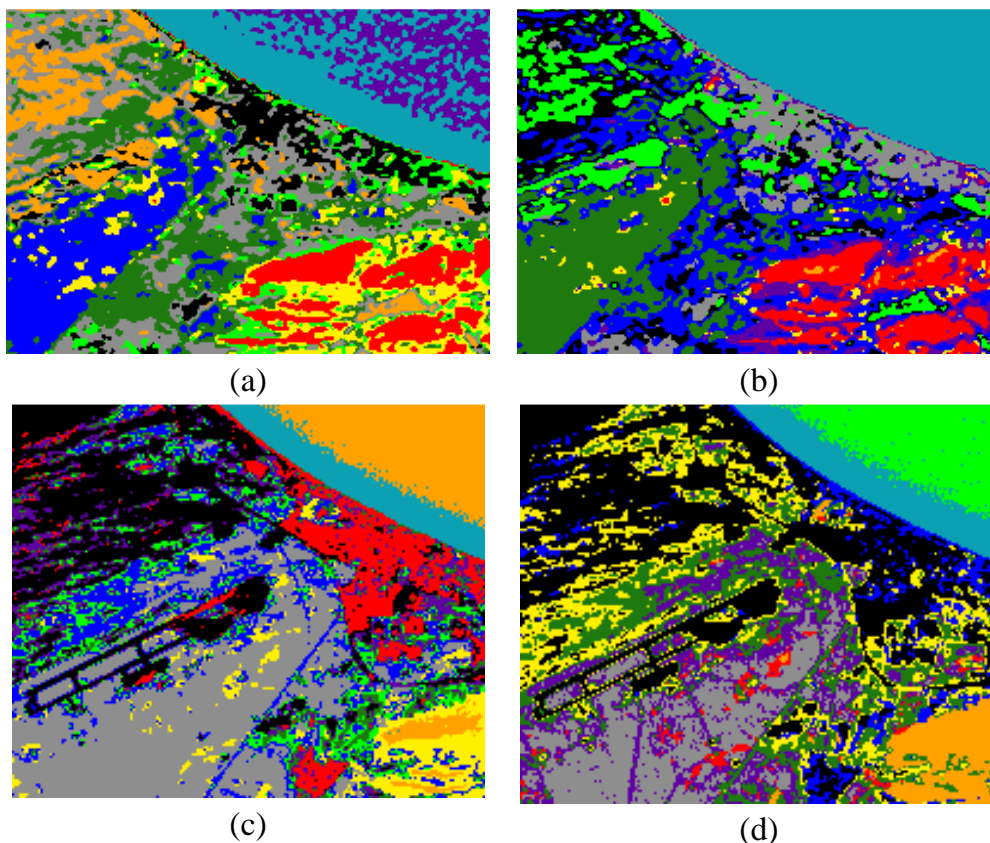


Figure 55 : Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Gaussien

Discussion

L'application de l'algorithme Firefly gaussien a permis l'amélioration de la qualité des résultats. Les résultats obtenus sont légèrement supérieurs à ceux obtenus par l'algorithme Firefly Flou.

Essai 4. Algorithme Firefly Flou Gaussien

Dans cet essai nous avons fixé le paramètre γ de la distance gaussienne à 1000.

L'indice	Jm	XB
<i>Image Sat 1</i>	0.031	$1.1635 * 10^{-3}$
<i>Image Sat 2</i>	0.055	$5.1638 * 10^{-4}$

Table 10. Résultats de la segmentation des images satellitaires avec l'algorithme Firefly Flou Gaussien

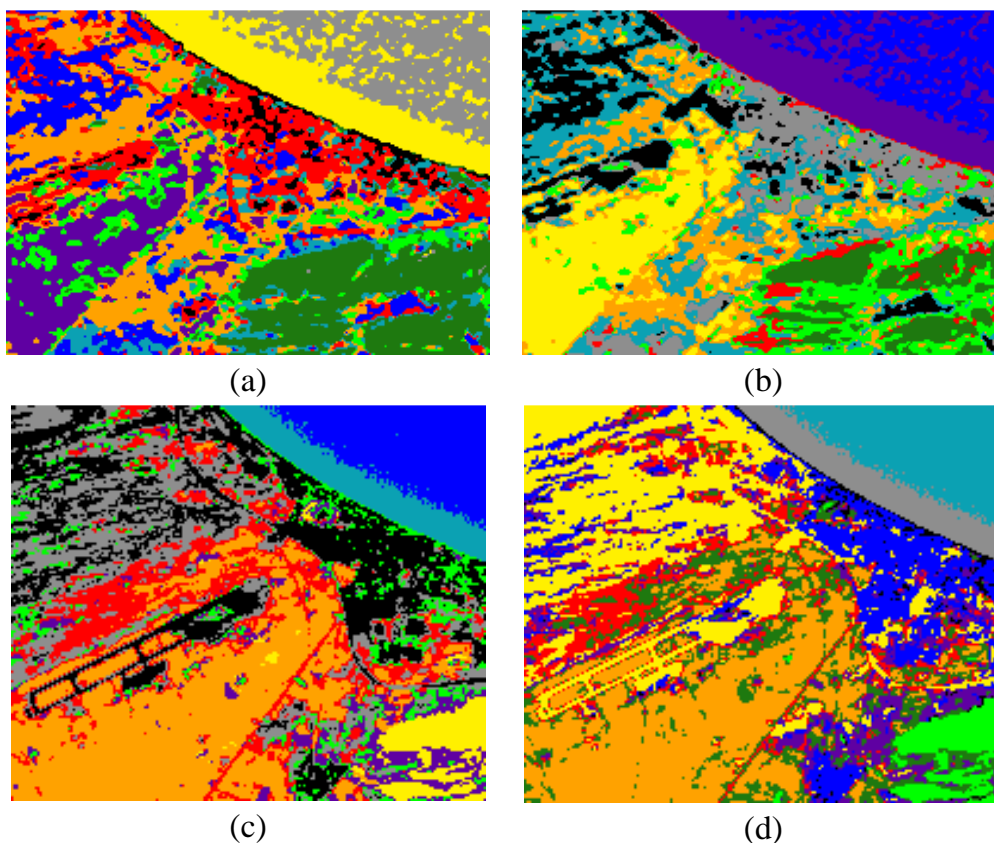


Figure 56: Résultats Visuels de la segmentation des images Satellitaires avec l'algorithme Firefly Flou Gaussien

Discussion

Les résultats obtenus par l'algorithme Firefly Flou Gaussien dans cet essai confirme l'efficacité de cette approche. Nous voyons que les régions sur l'image ont été bien séparées. La segmentation de l'image vérifie la réalité du terrain.

6. Le Comportement de l'approche Firefly Flou Gaussienne

Dans cette partie, nous allons étudier le comportement de l'approche proposée vis-à-vis le changement des paramètres. Nous allons varier la taille de la population, le nombre d'itération, les paramètres de l'algorithme Firefly Flou afin de voir la convergence de cet algorithme:

6.1 Comportement vis-à-vis la taille de la population :

Nous allons fixer les paramètres de l'algorithme à et varier la taille de population. Les paramètres choisis sont: nombre d'itérations $t = 30$, $\gamma = 0.0005$, $\beta = 1$ et γ de la distance gaussienne à 1000 :

Taille de la population	L'indice J_m					
	Essai 1	Essai 2	Essai 3	Essai 4	Essai 5	Essai 6
$n = 2$	0.1385	0.1765	0.1707	0.1593	0.1923	0.1315
$n = 5$	0.07192	0.09647	0.09326	0.07989	0.1182	0.08107
$n = 10$	0.07655	0.0496	0.09619	0.06726	0.0625	0.05776
$n = 20$	0.05962	0.0456	0.0549	0.06488	0.0600	0.04775

Table 11. Effet de la taille de la population sur la fonction de fitness

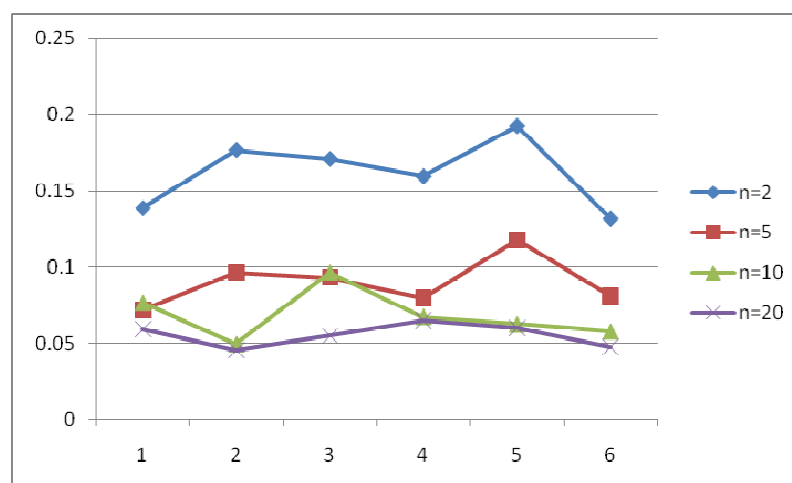


Figure 57. Effet de la taille de la population sur la fonction de fitness

Nous remarquons que la valeur de la fonction de fitness s'est stabilisé dans l'intervalle $[0.045, 0.060]$ pour une taille de population égale à 20.

6.2 Comportement vis-à-vis le nombre d'itérations :

Nous allons fixer les paramètres de l'algorithme à et varier le nombre d'itérations. Les paramètres choisis sont: nombre d'itérations $n = 10$, $\gamma = 0.0005$, $\beta = 1$ et γ de la distance gaussienne à 1000 :

Nombre d'itérations	L'indice J_m					
	Essai 1	Essai 2	Essai 3	Essai 4	Essai 5	Essai 6
$t = 1$	0.1018	0.1143	0.1137	0.1030	0.0825	0.0863
$t = 2$	0.0837	0.1023	0.0730	0.1326	0.1059	0.0954
$t = 5$	0.0919	0.1427	0.0983	0.1177	0.0851	0.1270
$t = 10$	0.0868	0.0802	0.0904	0.0808	0.0790	0.1036
$t = 20$	0.0830	0.0795	0.0772	0.0785	0.0800	0.1106
$t = 30$	0.0580	0.0507	0.0494	0.0604	0.0759	0.0570

Table 12. Effet du nombre d'itérations sur la fonction de fitness

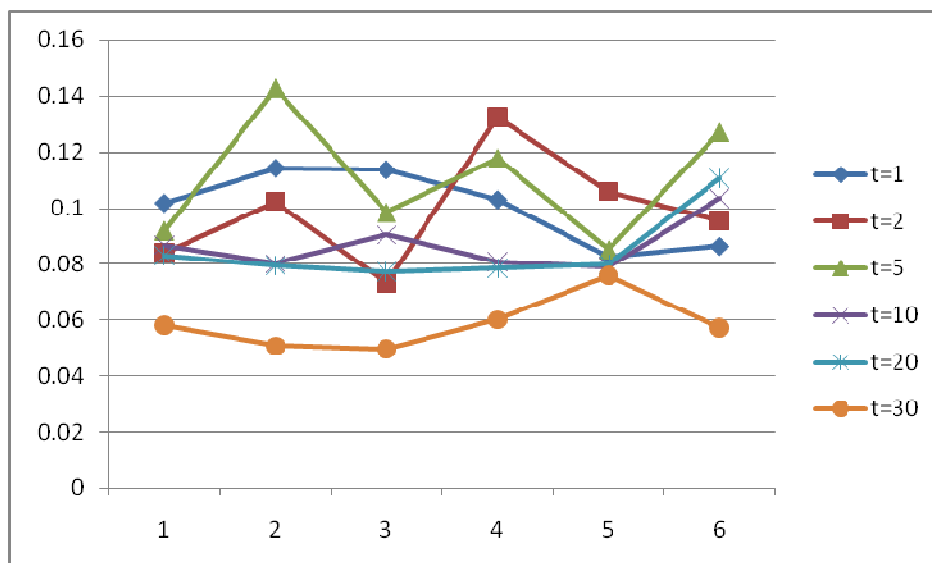


Figure 58. Effet du nombre d'itérations sur la fonction de fitness

Nous remarquons que la fonction de fitness s'est stabilisée pour un nombre d'itérations supérieur à 30.

7. Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode de segmentation des images satellitaires. De plus, Nous avons proposé trois approches de segmentation des images basées sur le nouvel algorithme Firefly: une approche Firefly floue pour résoudre le problème d'imprécision et d'incertitude dans les données, une approche Firefly Gaussienne pour résoudre le problème des classes imbriquées, une approche Firefly Flou Gaussienne. Ces approches ont été appliquées sur des images synthétiques, aériennes, et satellitaires. Les résultats obtenus ont montré l'efficacité de ces approches. Cependant, nous avons remarqué montrent la supériorité de l'approche Firefly Flou Gaussienne sur le reste des approches.

Conclusion Générale

Dans notre travail, nous avons abordé le problème de la classification non supervisée des images par une nouvelle méthode évolutive récente d'optimisation combinatoire. Cette méthode reproduit le phénomène des éclats de lumière générés par les lucioles dans la nature. De surcroît, nous avons établi des segmentations des images par trois approches basées sur cet algorithme. Les approches ont été appliquées sur des images synthétiques bruitées et facilement distinguable par l'œil, une image aérienne dotée d'une carte thématique qui nous a permis de valider les résultats obtenus et des images satellitaires de la région d'Oran.

Les expérimentations que nous avons effectuées ont bien montré l'efficacité de l'algorithme Firefly. De surcroît, nous avons constaté que l'intégration de la logique floue et la notion des distances noyaux sur l'algorithme Firefly a permis d'améliorer les résultats de la segmentation.

D'un autre côté, nous avons constaté que l'algorithme Firefly est influencé par les facteurs tel que la taille de la population, le nombre d'itération et beaucoup plus la fonction de fitness. Nous avons vue la différence entre l'utilisation de cet algorithme avec plusieurs indices de validité.

Comme perspectives, nous proposons d'établir une segmentation multiobjectives afin de minimiser l'influence des indices de validité sur la qualité du partitionnement. Nous proposons d'intégrer de nouvelles sources de données telles que le modèle numérique du terrain, les données radar, les textures, etc.

Références.

- [1] Shunlin Liang, Recent Advances in Land Remote Sensing: An Overview, University of Maryland, Springer Science, Business Media B.V, 2008.
- [2] Amghar Y., « Classification des données de télédétection par une approche neuronale RBF-SGNG » : mémoire de projet de fin d'étude. Université des sciences et de la technologie d'Oran Mohamed Boudiaf, Oran, Algérie, 2006.
- [3] Hannane Mokhtar Amir, Hadria Fizazi. « Conception d'un système hybride entre la Métaheuristique Electromagnétique et l'algorithme de colonie de fourmis pour la classification des images satellitaires » : mémoire de post-graduation (magister). Université des sciences et de la technologie d'Oran Mohamed Boudiaf, Oran, Algérie, 2010.
- [4] Centre canadien de télédétection, « Notions fondamentales de télédétection : Un cours tutoriel du Centre canadien de télédétection »
- [5] Télédétection de l'environnement dans l'espace francophone. Presse de l'université du Québec. 1994.
- [6] K. Tadjerouni .cours de télédétection. Cycle ingénieur 5ème année CNTS 2003.
- [7] Claude Kergomard, La TELEDETECTION AERO-SPATIALE: Une Introduction, Ecole Normale Supérieure Paris .
- [8] GARETH REES, The Remote Sensing Data Book , Cambridge University Press 1999
- [9] http://cct.rncan.gc.ca/resource/tutor/fundam/chapter1/04_f.php
- [10] http://iarivo.cirad.fr/doc/dr/big_3.pdf
- [11] <http://www.ipt.univ-paris8.fr/vgodard/enseigne/teled2/framtele.htm>
- [12] http://perso-sdt.univ-brest.fr/~delacour/COURS/M2 10.1/Cours1_Teldet_REM/

- [13] Xin-She Yang. Metaheuristic optimization. http://www.scholarpedia.org/article/Metaheuristic_Optimization, December 2010.
- [14] Bonabeau E., Dorigo M., Theraulaz G., Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, (1999)
- [15] Deb. K., Optimisation for Engineering Design, Prentice-Hall, New Delhi, (1995).
- [16] Goldberg D. E., Genetic Algorithms in Search, Optimisation and Machine Learning, Reading, Mass.: Addison Wesley (1989).
- [17] Kennedy J., Eberhart R., Shi Y.: Swarm intelligence, Academic Press, (2001).
- [18] Yang X. S., Nature-Inspired Metaheuristic Algorithms, Luniver Press, (2008).
- [19] Yang X. S., Biology-derived algorithms in engineering optimization (Chapter 32), in Handbook of Bioinspired Algorithms and Applications (eds Olariu & Zomaya), Chapman & Hall / CRC (2005).
- [20] Xin-She Yang. Nature-Inspired Metaheuristic Algorithms. Luniver Press, 2008.
- [21] Kennedy J. and Eberhart R. C.: Particle swarm optimization. Proc. of IEEE International Conference on Neural Networks, Piscataway, NJ. pp. 1942-1948 (1995).
- [22] <http://www.sciencedaily.com/releases/2011/04/110405093656.htm>
- [23] http://www.cirrusimage.com/beetle_firefly_Photuris_lucicrescens.htm
- [24] Eisner, Thomas; Wiemer, David; Haynes, Leroy; Meinwald, Jerrold (1978), Lucibufagins: Defensive steroids from the fireflies *Photinus ignitus* and *P. marginellus* (Coleoptera: Lampyridae), The National Academy of Sciences of the USA
- [25] Stanger-Hall, K.F.; Lloyd, J.E.; Hillis, D.M. (2007), "Phylogeny of North American fireflies (Coleoptera: Lampyridae): implications for the evolution of light signals", *Molecular Phylogenetics and Evolution* 45 (1): 33-49, doi:10.1016/j.ympev.2007.05.013, PMID 17644427
- [26] Encyclopædia Britannica. Firefly. <http://www.britannica.com/EBchecked/topic/207935/firefly>, May 2011.

- [27] Yang, X. S., “Firefly Algorithm, Stochastic Test Functions and Design Optimisation” , Int. J. Bio-Inspired Computation, Vol. 2, No. 2, 2010, pp.78--84
- [28] X.-S. Yang, “Firefly Algorithm, Levy Flights and Global Optimization,” Research and Development in Intelligent Systems XXVI (Eds M. Bramer, R. Ellis, Petridis), Springer London, 2010, pp. 209-218.
- [29] S. Lukasik and S. Zak, “Firefly algorithm for continuous constrained optimization tasks” in Proceedings of the International Conference on Computer and Computational Intelligence (ICCCI ‘09), N. T. Nguyen, R. Kowalczyk, and S.-M. Chen, Eds., vol. 5796 of LNAI, pp. 97–106, Springer, Wroclaw, Poland, October 2009.
- [30] X. S. Yang, “Firefly algorithms for multimodal optimization,” in Proceedings of the Stochastic Algorithms: Foundations and Applications (SAGA ‘09), vol. 5792 of Lecture Notes in Computing Sciences, pp. 178-178, Springer, Sapporo, Japan, October 2009.
- [31] Xin-She Yang. Firefly algorithm for multimodal optimization. Stochastic Algorithms: Foundation and Application, 5th, 2009.
- [32] Yang, X. S.: Engineering Optimization: An Introduction with Metaheuristic Applications, Wiley & Sons, New Jersey, (2010).
- [33] Shilane, D., Martikainen, J., Dudoit, S., Ovaska, S. J.: A general framework for statistical performance comparison of evolutionary computation algorithms, Information Sciences: an Int. Journal, 178, 2870-2879 (2008).
- [34] Baeck, T., Fogel, D. B., Michalewicz, Z.: Handbook of Evolutionary Computation, Taylor & Francis, (1997).
- [35] Wilkinsom G G.: Results and implications of a study of Fifteen years of satellite image classification experiments. IEEE T Geosci Remote Sens, 43(3): 433-440, 2005
- [36] Jain, A.K., Murty M.N., and Flynn P.J.: "Data Clustering: A Review, ACM Computing Surveys", Vol 31, No. 3, 264-323, 1999.
- [37] Björn Waske, Mingmin Chi, Jón Atli Benediktsson, Sebastian van der Linden, Benjamin Koetz: "Algorithms and Applications For Land Cover Classification – A Review" D. Li et al. (eds.), Geospatial Technology for Earth Observation, 2009.
- [38] Jain, A.K., Murty, M.N., Flynn, P.J.: "Data clustering: a review.", ACM Computing Surveys 31(3), 264-323, 1999.
- [39] T Lillesand, R Kiefer: "Remote Sensing and Image Interpretation", John Wiley & Sons Publishing, 1994.
- [40] S. Horowitz et T. Pavlidis : Picture segmentation by a directed split-and-merge procedure. Rapport de Recherche, Departement of Electrical Engineering, Princeton University, 1975.

- [41] S. Horowitz et T. Pavlidis : Picture segmentation by a tree traversal algorithm. *Journal of The Association for Computing Machinery*, 23(3):368-388, avril 1976.
- [42] Arabie, P., Hubert, L.J., De Soete, G. (eds.): *Clustering and Classification*, River Edge. World Scientific Publishing, Singapore (1996)
- [43] Everitt, B.S.: *Cluster Analysis*, 3rd edn. Halsted Press (1993)
- [44] Jean SERRA : A lattice approach to image segmentation. *Journal of Mathematical Imaging and Vision*, 24(1):83-130, 2006.
- [45] C. Berge : *Graphes et hypergraphes*. Dunod, 1969.
- [46] G. Matheron : *Random Sets and integral geometry*. John Wiley & Sons Inc, 1975.
- [47] H. Digabel et C. Lantujoul : Iterative algorithm. *Quantitative analysis of microstructures in materials sciences*, 1:85-99, 1978.
- [48] Brucker, P.: On the complexity of clustering problems. In: Beckmenn, M., Kunzi, H.P. (eds.) *Optimization and Operations Research. Lecture Notes in Economics and Mathematical Systems*, vol. 157, pp. 45-54. Springer, Berlin (1978)
- [49] Jean-Luc Verley, *Espaces métriques*, dans *Dictionnaire des mathématiques; algèbre, analyse, géométrie*, éd. Albin Michel, p.651
- [50] P.-N. Tan, M. Steinbach & V. Kumar, "Introduction to Data Mining", , Addison-Wesley (2005), ISBN 0-321-32136-7, chapter 8; page 500
- [51] Halkidi, M., Vazirgiannis, M.: Clustering validity assessment: Finding the optimal partitioning of a data set. In: *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM 2001)*, San Jose, California, USA, pp. 187-194 (2001)
- [52] Dunn, J.C.: Well Separated Clusters and Optimal Fuzzy Partitions. *J. Cybern.* 4, 95-104 (1974)
- [53] Calinski, R.B., Harabasz, J.: Adendrite Method for Cluster Analysis. *Commun. Statistics*, 1-27 (1974)
- [54] Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1, 224-227 (1979)
- [55] Chou, C.H., Su, M.C., Lai, E.: A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications* 7(2), 205-220 (2004)

- [56] D. Pham, Spatial Models for Fuzzy Clustering, *Computer Vision and Image Understanding*, Vol. 84, No. 2, 2001, pp. 285-297.
- [57] J. Bezdek, L. Hall, and L. Clarke, Review of MR Image Segmentation Techniques Using Pattern Recognition, *Medical Physics*, Vol. 20, No. 4, 1993, pp. 1033-1048.
- [58] Pal NR, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. *IEEE Trans Fuzzy Syst* 3:370-379
- [59] Kuo-Lung Wu a, Miin-Shen Yang, A cluster validity index for fuzzy clustering, *Pattern Recognition Letters* 26 1275-1291, 2005.
- [60] Muller, K.R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An introduction to kernelbased learning algorithms. *IEEE Transactions on Neural Networks* 12(2), 181-202 (2001)
- [61] Girolami, M.: Mercer kernel-based clustering in feature space. *IEEE Transactions on Neural Networks* 13(3), 780-784 (2002)
- [62] Scholkopf, B., Smola, A.J.: *Learning with Kernels*. MIT Press, Cambridge (2002)
- [63] Zhang, R., Rudnick, A.I.: A large scale clustering scheme for kernel k-means. In: *The Sixteenth International Conference on Pattern Recognition*, pp. 289-292 (2002)
- [64] Zhang, D.Q., Chen, S.C.: Clustering incomplete data using kernel-based fuzzy cmeans algorithm. *Neural Process Letters* 18, 155-162 (2003)
- [65] Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
- [66] Pirooznia, M., Deng, Y.: SVM Classifier “a comprehensive java interface for support vector machine classification of microarray data. In: *Proc. of Symposium of Computations in Bioinformatics and Bioscience (SCBB 2006)*, Hangzhou, China (2006)
- [67] Mezura-Montes, E., Velázquez-Reyes, J., Coello, C.A.C.: A comparative study of differential evolution variants for global optimization. In: *Genetic and Evolutionary Computation Conference (GECCO 2006)*, pp. 485-492 (2006)
- [68] Pal, N.R., Bezdek, J.C.: On cluster validity for the fuzzy c means model. *IEEE Transactions on Fuzzy Systems* 3(3), 370-379 (1995)
- [69] J. B. MacQueen "Some methods for classification analysis of multivariate observations." *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Pages 281-297. University of California Press. 1967
- [70] Baeck, T., Fogel, D. B., Michalewicz, Z.: *Handbook of Evolutionary Com- putation*, Taylor & Francis, (1997).

- [71] Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, (1999)
- [72] Kennedy, J. and Eberhart, R. C.: Particle swarm optimization. *Proc. of IEEE International Conference on Neural Networks*, Piscataway, NJ. pp. 1942-1948 (1995).
- [73] Omran M. Particle Swarm Optimization methods for pattern Recognition and Image Processing. Dissertation for the Doctoral Degree. University of Pretoria, 2005
- [74] S. Bandyopadhyay and S. K. Pal, "Pixel classification using variable string genetic algorithms with chromosome differentiation" *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 303-308, 2001.
- [75] J. H. Holland. "Adaptation in natural and artificial systems ". Ann Arbor: University of Michigan Press. 1975.

Contributions

JDLIO'11 : « Segmentation des Images Satellitaires par l'Algorithme Firefly Flou », 1ères Journées Doctorales du Laboratoire d'Informatique d'Oran, Oran, Algerie, 2011.

COSI'12 : « An Enhanced Bio-Inspired Firefly Algorithm for Remote Sensing Images Segmentation», Colloque sur l'Optimisation et les Systèmes d'Information, 12-15 Mai 2012, Tlemcen, Algérie