



UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES
DEPARTEMENT DES MATHEMATIQUES



Master Mathématique et Application au Calcul Scientifique (MACS)

MEMOIRE DE FIN D'ETUDES

Pour l'obtention du Diplôme de Master Sciences et Techniques
(MST)

Segmentation d'image en statistique décisionnelle

Réalisé par : MESSAOUDI Mohammed

Encadré par: Pr. AMMOR Ouafae

Soutenu le 16/06/2016

Devant le jury composé de:

- | | |
|----------------------------|-------------|
| -Pr. AMMOR Ouafae | FST Fès |
| -Pr. ELHILALI ALAOUI Ahmed | FST Fès |
| - Pr. EL KHOUKHI Fatima | FLSH Meknès |
| -Pr. EZZAKI Fatima | FST Fès |

Année Universitaire 2015 / 2016

FACULTE DES SCIENCES ET TECHNIQUES FES – SAISS

☒ B.P. 2202 – Route d'Imouzzer – FES

Table des matières

Table des matières	2
Dédicaces	5
Remerciements	6
Introduction générale	7
1 Quelques notions de base sur la segmentation d'image	10
1.1 Introduction	10
1.2 Définition d'une image	11
1.3 Types d'image	12
1.3.1 Image binaire (noir ou blanc)	12
1.3.2 Image en niveaux de gris	12
1.3.3 Image couleur	12
1.3.4 Image à valeurs réelles	12
1.4 Caractéristique d'image	13
1.4.1 Pixel	13
1.4.2 La taille d'une image	13
1.4.3 Histogramme	13
1.5 Définition de segmentation d'image	14
1.6 Différentes approches de la segmentation d'image	14

1.6.1	La segmentation par contours	15
1.6.2	La segmentation par région	15
1.7	Segmentation d'image par classification des pixels	19
1.7.1	Algorithme K-means	19
1.7.2	Algorithme Estimation-Maximisation (EM)	20
1.8	Segmentation par seuillage	22
1.8.1	Définition du seuillage	22
1.9	Exemples d'applications	25
1.10	L'importance de la segmentation en statistique décisionnelle	25
1.10.1	Segmentation marketing de votre marché	25
2	Méthodes de classification d'image	29
2.1	Introduction	29
2.2	Méthodes supervisées	29
2.2.1	Méthodes bayésiennes	30
2.2.2	Algorithme des k plus proches voisins	31
2.2.3	Méthodes basées sur les réseaux de neurones artificiels	32
2.3	Méthodes non Supervisées	33
2.3.1	Algorithme des K-means	33
2.3.2	Algorithme des C-moyennes	34
2.3.3	Algorithme Fuzzy Cmeans (FCM)	35
2.4	Inconvénients de l'algorithme FCM	38
3	Les variantes de l'algorithme FCM	39
3.1	Introduction	39
3.2	Variante selon la fonction objectif	39
3.2.1	Algorithme d'Ahmed et al	39
3.2.2	Algorithme de Chen et Zhang	40
3.2.3	Algorithme de Yang et al	41
3.2.4	Algorithme de Li et Mukaidono	42

3.2.5	Algorithme Cmoynes possibiliste (PCM)	42
3.3	Variante selon la distance	43
3.3.1	Algorithme de Tsai et Lin	43
3.3.2	Algorithme FCM basé sur des fonctions noyau	44
3.4	Variante selon la fuzzification	46
3.4.1	L'algorithme Miin.Shen Yang (MSFCM)	46
3.4.2	l'algorithme de Mohamed Adel Alimi et al	47
4	Tests et résultats	48
4.1	Introduction	48
4.2	L'imagerie médicale	48
4.2.1	Imagerie par résonance magnétique (IRM)	49
4.2.2	L'observation du cerveau	49
4.3	Présentation des images médicales utilisées	50
4.4	Les critères utilisées	51
4.4.1	Coefficient de Partition (Vpc)	51
4.4.2	Entropie de Partition (Vpe)	51
4.4.3	Indice de FUKUYAMA et SUGENO (Vfs)	52
4.4.4	Indice de XIE et BENI (Vxb)	52
4.4.5	Critère de Max et Min-Max	52
4.5	L'application de l'algorithme K-means	53
4.6	Comparision entre l'algorithme FCM et ses variantes	55
	Conclusion générale	59
	Bibliographie	60

Dédicaces

Au mémoire de mon cher père

A ma très chère mère

A mes frères : Rachid, Mohcine et Moussa

A mes sœurs : Kaltom et Fatiha

A mes grands-parents, mes oncles, mes tantes, mes cousins,
cousines et toute ma grande famille.

A tous mes amis de l'Université et d'ailleurs.

Messaoudi Mohammed

Remerciements

Je remercie tout d'abord le bon dieu pour m'avoir donnée le courage et la santé pour accomplir ce travail.

Ce travail n'aurait pas pu aboutir à des résultats sans l'aide et les encouragements de plusieurs personnes que je remercie.

Mes vifs remerciements accompagnés de toute ma gratitude vont ensuite à mon encadreur **Mme. Ammor Ouafae** professeur à la Faculté des Sciences et Techniques de Fès. Je lui adresse un grand merci pour sa grande disponibilité, ses remarques et ses conseils tout au long de ce travail.

Je tiens à remercier également les membres de mon jury : **Pr. ELHILALI ALAOUI Ahmed, Pr. EZZAKI Fatima et Pr. EL KHOUKHI Fatima**, qui ont accepté d'évaluer mon travail.

Mon vifs remerciements s'adressent également à nos professeurs qui nous ont encouragés et nous ont aidés avec leurs remarques et observations durant la période de notre parcours universitaire.

Enfin, que tous ceux qui nous ont aidés et encouragés de près ou de loin dans la concrétisation de ce projet, trouvent ici ma gratitude et mes sincères remerciements.

Je tiens aussi à exprimer ma reconnaissance à toutes mes amies pour leur soutien inconditionnel.

Introduction générale

L'information visuelle est sans doute la plus riche des différentes sources d'information existantes dans notre vie quotidienne. L'extraction de cette information par des systèmes de traitement et d'analyse d'images, a suscité un intérêt sans cesse croissant. Le traitement d'images qui est un processus comprenant plusieurs étapes, est né de la nécessité de remplacer l'observateur humain par la machine. Il intervient dans bon nombre de domaines tels que la médecine, l'agriculture et l'industrie.

Une étape incontournable dans tout processus d'analyse d'images est celle de la segmentation. En fournissant une description compacte de l'image, plus exploitable que l'ensemble des pixels, la segmentation d'images permet de faciliter l'interprétation automatique d'une image de façon similaire à une interprétation humaine. En effet, elle a été inspirée du système de perception visuel humain qui utilise la notion de similarité et de différence afin de localiser et de délimiter les objets d'une scène. Elle peut être définie ainsi : la segmentation d'images est un traitement de bas niveau qui consiste à créer une partition de l'image I en sous-ensembles R_i , appelés régions tel qu'aucune région ne soit vide, l'intersection entre deux régions soit vide et l'ensemble des régions recouvre toute l'image. Une région est un ensemble de pixels connexes ayant des propriétés communes qui les différencient des pixels des régions voisines.

Cette tâche bien que relevée avec facilité par le système visuel humain, est en réalité complexe et reste un véritable défi pour la communauté du traitement d'images malgré plusieurs décennies de recherche. Ainsi, plusieurs méthodes de segmentation

ont été proposées dans la littérature, et qui peuvent être classées selon trois grandes approches : Approche contour, Approche région et Approche seuillage.

Nous nous intéressons dans ce mémoire à l'approche région de la segmentation d'images par classification non supervisée.

Les algorithmes de classification les plus répandus tendent à minimiser une fonction cout, dépendant de la distance de chaque pixel aux prototypes (centres) des classes. Le prototype d'une classe étant un point connu dont l'appartenance à la classe est garantie.

Parmi ces algorithmes, l'algorithme Fuzzy CMeans (FCM), proposé par Bezdek, est le plus connu et a été utilisé dans bon nombre de domaines à cause de sa simplicité.

Cependant, cet algorithme qui nécessite la connaissance du nombre de classes est jugé moins robuste face au bruit et son efficacité dépend fortement de l'étape d'initialisation des centres des classes car le processus itératif peut facilement fournir une solution localement optimale. De plus, il utilise généralement la distance euclidienne pour mesurer la similarité entre une observation et le centre d'une classe ce qui le rend utilisable que pour détecter des classes de forme sphérique.

Afin d'éviter ces inconvénients et donc améliorer les résultats de la classification, plusieurs modifications ont été apportées à l'algorithme standard et qui sont présentées comme des variantes du FCM. Les différentes variantes, auxquelles nous nous sommes intéressés dans ce mémoire, font face au problème du bruit.

Nous les avons classées, dans notre travail, en trois catégories :

- Dans la première, la distance euclidienne utilisée par le FCM standard, est remplacée par un autre type de distance afin de détecter des classes de formes quelconques.
- Dans la deuxième catégorie, l'information spatiale a été introduite dans la fonction objectif sous forme d'un terme supplémentaire dans le but de prendre le bruit en considération.
- La troisième catégorie modifie l'influence du facteur flou.

La plupart de ces méthodes ont été développées dans le cadre de la segmentation d'images en niveaux de gris ou en couleur. Leurs applications pour la segmentation d'images médicales n'ont pas fait l'objet d'études. Ainsi, afin de tester l'apport de ces méthodes

par rapport à l'algorithme FCM standard, nous avons effectué, dans ce mémoire, une étude comparative entre ce dernier avec certaines de ses variantes dans le cadre de la segmentation d'images médicales .

Le mémoire est principalement scindé en quatre chapitres.

Dans le premier chapitre, nous avons abordé les notions sur la segmentation d'image en niveaux de gris, nous décrirons les différentes approches.

Le deuxième chapitre donne un aperçu général sur la classification d'images.

Le troisième chapitre donne une Variante de l'algorithme FCM, Le problème du nombre de classes et celui de l'initialisation des centres des classes ne sera pas abordé. Seules les méthodes permettant d'éviter le problème du bruit et celui de la distance sont présentées.

Le quatrième chapitre est consacré à l'application de la segmentation d'images médicales basée sur l'algorithme FCM et quelques variantes de FCM.

Chapitre 1

Quelques notions de base sur la segmentation d'image

1.1 Introduction

L'analyse et l'interprétation des images sont actuellement très utilisées dans différents domaines allant de l'aide au diagnostic, en médecine, à la navigation autonome des véhicules en passant par la reconnaissance des visages, des empreintes, de l'iris, le contrôle de qualité des produits manufacturables (bois, tissus, verre, pièces mécaniques,.....etc).

Dans un système d'analyse d'images, la segmentation apparaît l'étape la plus importante car toutes les tâches ultérieures comme l'extraction de primitives, la détection d'une position d'un objet, ou la reconnaissance d'un objet dépendent fortement de la qualité de la segmentation.

Nous avons présenté dans ce chapitre quelques notions de base sur la segmentation d'images. nous commençons par les deux définitions : définition d'image et définition de segmentation d'image, nous présentons, ensuite, les différentes approches de la segmentation d'images : segmentation par région, segmentation par contours et segmentation par seuillage.

1.2 Définition d'une image

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

Elle peut être décrite sous la forme d'une fonction $p(x, y)$ de brillance analogique continue, définie dans un domaine borné, tel que x et y sont les coordonnées spatiales d'un point de l'image et est une fonction d'intensité lumineuse et de couleur. Sous cet aspect, l'image est inexploitable par la machine, ce qui nécessite sa numérisation [11].

Le terme d'image numérique désigne, dans son sens le plus général, toute image qui a été acquise, traitée et sauvegardée sous une forme codée représentable par des nombres (valeurs numériques).

Pratiquement toujours, la valeur d'un pixel est un mot binaire de longueur K bits, par conséquent un pixel peut prendre l'une des valeurs de l'intervalle $[0, 2^K - 1]$. La valeur K est appelée profondeur de l'image. L'ensemble de ces pixels est contenu dans un tableau à deux dimensions (une matrice) constituant l'image finalement obtenue.

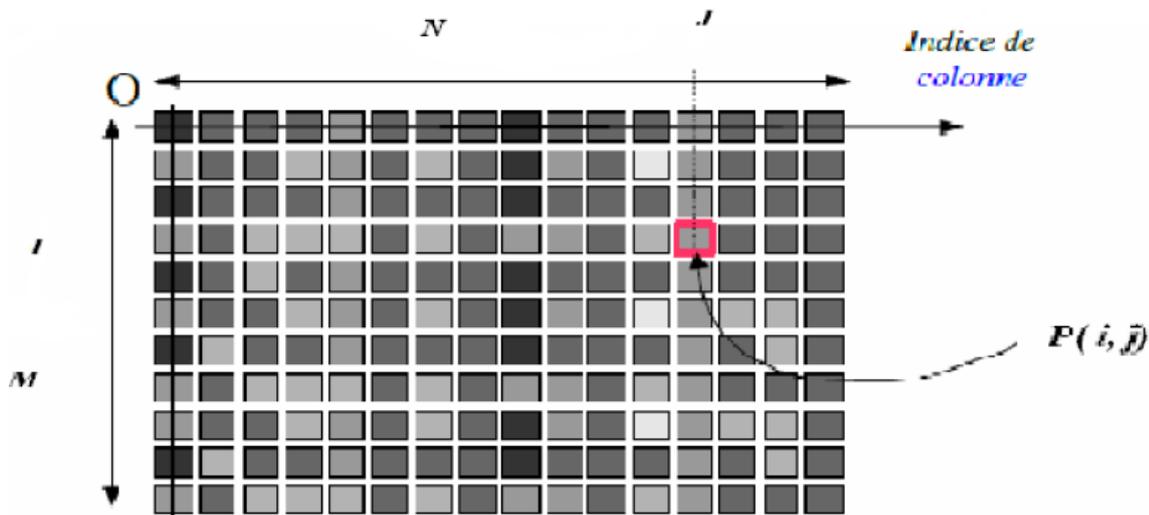


Figure 1. Représentation de la matrice (M, N)

Où $P(i, j)$ est le niveau de gris du pixel de coordonnées ligne i et colonne j dans l'image.

1.3 Types d'image

1.3.1 Image binaire (noir ou blanc)

Les images binaires sont des images de profondeur $K=1$ bit, donc un pixel peut prendre l'une des valeurs : noir ou blanc (0 ou 1).

C'est typiquement le type d'images que l'on utilise pour scanner du texte quand celui-ci est composé d'une seule couleur.

1.3.2 Image en niveaux de gris

En général, les images en niveaux de gris sont des images de profondeur $k=8$ bits, donc chaque pixel peut prendre l'une des valeurs de l'intervalle $[0,255]$, où la valeur 0 représente la brillance minimale (le noir) et 255 la brillance maximale (le blanc). Ce type d'image est fréquemment utilisé pour reproduire des photos en noir et blanc ou du texte.

Dans plusieurs applications professionnelles de photographie et d'impression ainsi qu'en médecine et astronomie, 8 bits par pixel n'est pas suffisant, pour cela il existe d'autres types d'images en niveaux de gris de profondeur $K=12$, $K=14$ ou $K=16$ bits.

1.3.3 Image couleur

L'espace couleur est basé sur la synthèse additive des couleurs, c'est à dire que le mélange entre différentes couleurs (trois, quatre...) donne une couleur.

La plupart des images couleurs sont basées sur trois couleurs primaires : Rouge, Vert et Bleu (RVB) (RGB en anglais), et utilisent typiquement 8 bits pour chaque composante de couleur, donc chaque pixel nécessite $3 \times 8 = 24$ bits pour coder les trois composantes, et chaque composante de couleur peut prendre l'une des valeurs de l'intervalle $[0,255]$.

1.3.4 Image à valeurs réelles

Pour certains calculs sur les images, le résultat peut ne pas être entier, il est donc préférable de définir l'image de départ et l'image résultat comme des images à valeurs

réelles. En général, **une image à valeurs réelle** est telle que le niveau de gris est un réel compris entre 0.0 et 1.0. On a dans ce cas pour une image à niveaux de gris : $p(i, j) \in [0.0, 1.0]$. Pour une image couleur, la relation devient $p_R(i, j) \in [0.0, 1.0]$, $p_V(i, j) \in [0.0, 1.0]$, $p_B(i, j) \in [0.0, 1.0]$.

1.4 Caractéristique d'image

1.4.1 Pixel

Une image numérique est constituée d'un ensemble de points appelés pixels (abréviation de Picture Élément) pour former une image. Le pixel représente ainsi le plus petit élément constitutif d'une image numérique. L'ensemble de ces pixels est contenu dans un tableau à deux dimensions constituant l'image [1].

1.4.2 La taille d'une image

La taille d'une image est le nombre de pixels de cette image, la taille d'une image qui est représenté par (328×456) dont 328 est le nombre de lignes, et 456 est le nombre de colonnes, est égale à : $328 \times 456 = 149568$ pixels.

1.4.3 Histogramme

Un histogramme est un graphique statistique permettant de représenter la distribution des intensités des pixels d'une image, c'est-à-dire le nombre de pixels pour chaque intensité lumineuse. Par convention un histogramme représente le niveau d'intensité en abscisse en allant du plus foncé (à gauche) au plus clair (à droite).

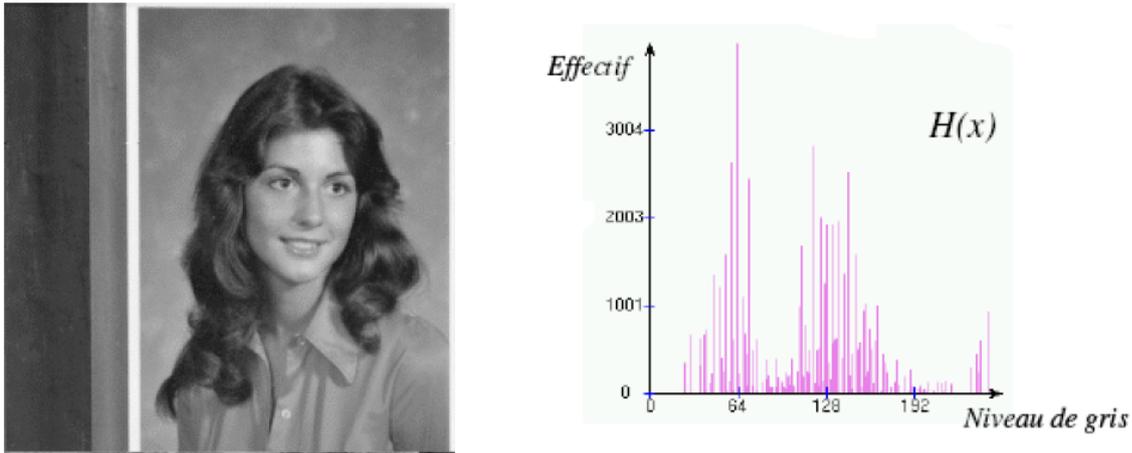


Figure 2. Représentation d'un histogramme d'une image

Avec : $H(x)$ est le nombre de pixels dont le niveau de gris est égal à x .

1.5 Définition de segmentation d'image

La segmentation est un traitement de bas-niveau qui consiste à effectuer une partition de l'image en régions homogènes par rapport à un ou plusieurs critères. Les régions obtenues se distinguent les unes des autres par des différences significatives selon ces mêmes critères. Après ces étapes, nous pouvons introduire un traitement sectoriel de différentes manières.

La segmentation consiste à extraire des points, des lignes ou des régions. Le choix d'une technique de la segmentation est liée à plusieurs facteurs comme : la nature de l'image, les conditions d'acquisition (bruit), les primitives à extraire (contour, texture,...).

La segmentation fait référence aux notions de similarité comme les perçoit, le système visuel humain et ceci donne naissance à deux approches couramment qualifiées d'approche (région) et d'approche (frontière).

1.6 Différentes approches de la segmentation d'image

Dans la littérature, plusieurs manières de catégoriser les méthodes de segmentation qui s'intègre généralement dans trois approches principales : approche par contours, approche par régions, et approche par seuillage.

Dans ce qui suit, on va présenter une description des différentes approches :

1.6.1 La segmentation par contours

Les méthodes basées contours sont parmi les méthodes les plus classiques en segmentation d'images. Ces méthodes supposent généralement un modèle a priori des discontinuités recherchées et opèrent de manière très localisée.

De façon général, un contour est défini comme étant la frontière entre deux régions. Les méthodes de segmentation basées sur l'approche contour ont donc pour objectif de trouver les lieux de fortes variations du niveau de gris. Nous pouvons citer les approches se basant sur l'opérateur de gradient, l'opérateur Laplacien et les différents filtres à savoir : le filtre de Sobel, Prewitt et Roberts ou bien des approches analytiques comme le filtre de Canny [7,9,16].

Mais ce genre de techniques est peu utilisable car il donne souvent des contours non fermés, bruités ou non détectés, une utilisation des propriétés des régions comprise entre ces contours pourrait nettement améliorer la détection de ce derniers.

1.6.2 La segmentation par région

Contrairement aux techniques d'extraction des contours, La segmentation en régions homogènes est basée sur les propriétés intrinsèques des régions. Le choix de ces propriétés détermine ce qu'on appelle (critère de segmentation). Pour segmenter l'image en régions, ces critères peuvent être la valeur de niveau de gris, de la couleur, de la texture, ou une combinaison de plusieurs informations.

D'une manière formelle nous pouvons définir la région par l'ensemble connexe de points répondants au même critère d'homogénéité. Le formalisme de la régionalisation a été introduit par Horowitz et Pavlidis, et est le fondement de base d'un grand nombre de techniques de segmentation en régions.

Une segmentation \mathbf{S} d'une image \mathbf{I} relativement à un prédicat \mathbf{Pred} [21] est une partition de \mathbf{I} en \mathbf{n} ensembles disjoints non vides R_1, R_2, \dots, R_n tels que :

1. $\forall i = 1, \dots, n; I = \bigcup_{i=0}^n R_i$

2. $\forall i = 1, \dots, n, R_i$ est constituée des pixels connexes.
3. $Pred(R_i) = vrai$ pour tout i .
4. $Pred(R_i \cap R_j) = faux$ pour tout i, j, R_i et R_j étant adjacents dans I .

La première condition implique que chaque point de l'image appartient à une seule région. La deuxième condition est une contrainte de connexité imposée aux régions. La troisième condition demande que les régions de l'image segmentée soient homogènes. La condition quatre exprime que les régions vérifiant le prédicat d'homogénéité, ont une taille maximale.

Généralement, l'approche région cherche à regrouper les pixels en régions homogènes. Elle se caractérise par la mesure d'uniformité des régions construites dans l'image. Ces régions sont construites en évaluant la similarité entre les pixels ou entre un pixel et ceux d'une même région. On distingue les méthodes par Croissance des régions, division des région, division-fusion et par classification.

1) Croissance des régions

Ce type de segmentation permet de sélectionner un pixel ou un ensemble de pixels de l'image, appelé **greme**, autour dequel on fait croître une région.

Les régions sont construites en ajoutant successivement à chaque **greme** les pixels qui lui sont connexes et qui vérifient un critère de similarité. La croissance s'arrête lorsque tous les pixels ont été taitées. La littérature en traitement d'image est riche en méthode de segmentation par croissance de région.

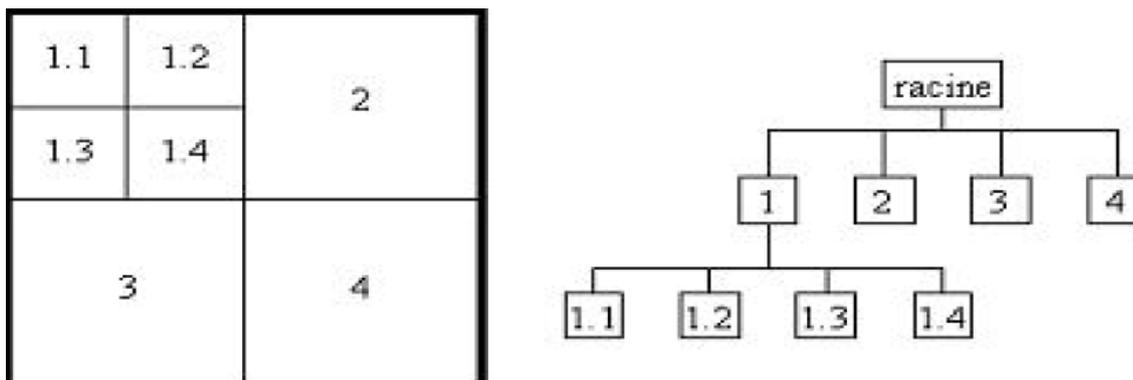
Tremeau et **Borel**[2] proposent un algorithme de segmentation qui combine une croissance de régions suivie d'un processus de fusion de régions. Cet algorithme procède par un balayage séquentiel de l'image et considère le premier pixel comme greme. Il tente alors de faire croître ce greme de plus longtemps possible en agrégeant des pixels voisins.

2) Division des régions

Les modèles de division de régions consistent à diviser l'image d'une manière récursive tant qu'un critère d'homogénéité sur les régions n'est pas vérifié. Initialement, l'image est divisée en zones. Chaque zone ainsi crée elle-même divisée si elle ne vérifie pas le critère

d'homogénéité. Ce processus de division se poursuit pour toutes les zones nouvellement créées et s'arrête que lorsque toutes les zones sont homogènes ou bien leurs tailles sont en dessous d'un seuil fixé préalablement. La division de l'image est généralement représentée par une structure géométrique bien particulière telle que la structure à arbre quaternaire (quadtree) ou la structure de voronoï.

La figure(3) montre le principe de la segmentation par division suivant la structure d'arbre quaternaire.



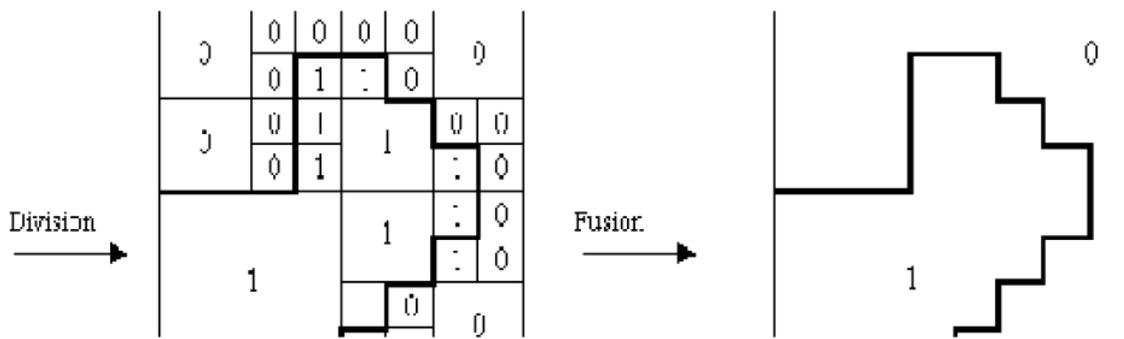
La figure(3) : le principe de la segmentation par division des régions.

Le critère d'homogénéité peut être évalué par la variance des niveaux de gris ou peut être défini à partir de l'allure l'histogramme de niveau de gris. Par exemple, l'histogramme de chaque région est analysé, si il est unimodal (région homogène), la région reste inchangée, sinon elle division prend fin quand tous les régions ont un histogramme unimodal.

3) Division-fusion

Proposées par Horowitz, elles réunissent l'ensemble des algorithmes utilisés dans les techniques précédemment citées (méthodes par fusion et méthodes par division). Le processus de segmentation se déroule en deux étapes : la division en blocs (régions) suivant un critère d'hétérogénéité est effectuée en premier lieu, puis, suit l'étape de fusion où toutes les régions adjacentes qui découlent de l'étape de division et vérifiant une heuristique d'homogénéité sont regroupées.

La figure (4) illustre le principe de la **Division-fusion**. Les résultats dépendent de l'ordre dans lequel les régions sont examinées pour être éventuellement regroupées.



La figure(4) : le principe de la segmentation par division-fusion.

4) La segmentation basée sur la classification

Cette méthode consiste à regrouper et à classer les pixels d'une image en classes en fonction de leurs propriétés. A chaque point de l'image est associé un vecteur d'attributs. La classification est alors effectuée sur ces vecteurs d'attributs de façon à aboutir à un nombre restreint de régions homogènes au sein de l'image.

Donc la classification est définie comme une procédure dans laquelle les pixels similaires d'une image sont identifiés et regroupés dans une même classe. Il existe deux grandes tendances :

- **La classification non supervisée :** Elle vise à séparer automatiquement l'image en clusters sans aucune connaissance a priori sur les classes. elle se base sur une mesure de distance entre les vecteurs d'attributs. Les algorithmes les plus fréquemment cités dans la littérature pour cette catégorie sont K-means, Isodata, et Fuzzy c-means...
- **La classification supervisée :** Elle s'opère à partir de la connaissance de chacune des classes définies par une approche probabiliste. Elle se base sur l'apprentissage de propriétés discriminantes sur un échantillon de données déjà classées. Les algorithmes de cette catégorie sont Estimation-Maximisation (EM), Algorithme des k plus proches voisins et Réseaux de neurone...

1.7 Segmentation d'image par classification des pixels

La classification peut se faire de deux manières : La première suppose l'existence de certains pixels dont l'appartenance aux classes est connue à priori (supervisée), elle est très peu utilisée en segmentation car elle nécessite l'intervention de l'utilisateur (par exemple : **l'algorithme d'Estimation-Maximisation (EM)**).

La seconde dite non supervisée, vise à regrouper automatiquement des pixels de l'image en classes sans aucune connaissance préalable sur l'appartenance des pixels (par exemple : **K-means et algorithme Fuzzy-means**).

1.7.1 Algorithme K-means

C'est l'un des algorithmes les plus connus en classification non supervisée. Il vise à produire un partitionnement des pixels de manière à ce que les pixels d'une même classe soit semblables et les pixels issus de deux classes différentes soient dissemblables.

L'idée principale est de définir K centroïdes, un pour chaque classe. Chaque classe C_k est ainsi caractérisée par centre noté μ_k et le nombre d'éléments N_k .

L'algorithme K-means dans sa formulation originale cherche à minimiser une fonction de coût global définie :

$$J = \sum_{k=1}^K \sum_{(x,y) \in C_k} (f(x,y) - \mu_k)^2$$

où $f(x,y)$ représente le niveau de gris du pixels de coordonnées (x,y) .

Il se déroule selon les étapes suivantes :

1. Initialisation de chaque centre μ_k .
2. Pour chaque pixels (x,y) , calculer la distance $d(f(x,y), \mu_k)$ aux différents centres de classes μ_k , et affecter à la classe la plus proche $C_l = \arg \min_k d(f(x,y), \mu_k)$ avec $d(f(x,y), \mu_k) = |f(x,y) - \mu_k|$.
3. Mise à jour de nombre de pixels et des centres des classes ; $\mu_k = \frac{\sum_{(x,y)} f(x,y)}{N_k}$
4. Arrêt si $N_k = N_{k+1}, \forall (x,y) \in C_k$
sinon retour à l'étape 2.

Le principal inconvénient de cette méthode est que la classification finale dépend du choix de la partition initiale.

Le minimum global n'est pas obligatoirement atteint, on est seulement certain d'obtenir la meilleure partition à partir de la partition de départ choisie. Une version floue de l'algorithme K-means appelée Fuzzy C-means est également très populaire. Cet algorithme nécessite la connaissance préalable du nombre de classe et génère les classes par un processus itératif en minimisant une fonction objective. Il permet d'obtenir une partition floue de l'image en donnant à chaque pixels un degré d'appartenance (compris entre 0 et 1) à une classe donnée.

1.7.2 Algorithme Estimation-Maximisation (EM)

Le principe de ces méthodes consiste à estimer la fonction de densité de probabilité (f.d.p) sous jacente à l'ensemble des données à la classer et assimiler chaque mode de cette fonction à une classe.

Sous l'hypothèse paramétrique, ces méthodes consistent à fixer, a priori, un modèle aux fonction de densités de probabilités conditionnelles de chaque classe. La fonction densité de probabilité en un point est alors composée d'un mélange de K composantes ou fonctions de densité de probabilité conditionnelle pondérés par leurs probabilités à priori. Les paramètres du modèle relatifs à chaque classe et les probabilités a priori des classes constituent les paramètres du mélange que l'on cherche à identifier à partir de l'ensemble des observations à analyser.

Dans l'algorithme EM, la densité de probabilité $f(x_i)$ en un point x_i est décrite par modèle de mélange. Le principe consiste à décomposer cette densité en une somme de K composantes $f(x_i/\theta_k)$ conditionnellement aux paramètres θ_k correspondant aux K classe.

Il s'agit alors d'estimer les paramètres θ_k ($k=1,2,\dots,K$) à partir d'un échantillon X . Ces densités de probabilités $f(x_i/\theta_k)$ peuvent aller du modèle le plus simple aux distribution les plus complexes. Les proportions P_k entre les différentes composantes représentent les probabilités a priori des différentes classes notées $f(C_k)$. En général, ces

proportions sont également inconnues et doit être estimées sous les contraintes :

$$P_k \in [0, 1] \text{ et } \sum_{k=1}^K P_k = 1 \quad (1.1)$$

Soit $\theta = [P_1, P_2, \dots, P_K, \theta_1, \theta_2, \dots, \theta_K]$, le vecteur de paramètres à estimer. La fonction densité de probabilité $f(x_i)$ en un point x_i est donnée par la relation suivante :

$$f(x_i/\theta) = \sum_{k=1}^K P_k f(x_i/\theta_k), i = 0, \dots, L - 1 \quad (1.2)$$

L'estimation des paramètres d'un mélange est effectuée suivant l'algorithme Estimation-Maximisation (EM) qui est basé sur la maximisation de la loi de vraisemblance.

La loi de vraisemblance d'un ensemble d'échantillons X_n d'une variable aléatoire X relativement au modèle de paramètre θ s'écrit :

$$L(\theta) = f(X/\theta) \quad (1.3)$$

sous l'hypothèse que les données de l'ensemble d'apprentissage X_n sont des réalisations indépendantes du vecteur aléatoire X , la loi de vraisemblance se réécrit en produit de probabilités :

$$L(\theta) = \prod_{i=1}^n f(X_i/\theta) \quad (1.4)$$

Dans le cas de modèles de mélange, cette équation se met sous la forme :

$$L(\theta) = \prod_{i=1}^n \sum_{k=1}^K P_k f(X_i/\theta_k) \quad (1.5)$$

$$\log(L(\theta)) = \sum_{i=1}^n \log \sum_{k=1}^K P_k f(X_i/\theta_k) \quad (1.6)$$

La solution de cette équation par la méthode d'estimation du maximum de vraisemblance équivaut à la recherche des racines de l'équation suivante :

$$\frac{\partial \log(L(\theta))}{\partial \theta} = 0 \quad (1.7)$$

Dans le cas des mélanges gaussiens, les θ_k représentent les moyennes et les matrices de covariance de la $k^{ième}$ classe. Les paramètres qui annulent des dérivées sont donnés par

les équations suivantes pour $k = 1, \dots, K$.

$$\begin{aligned}
 P_k &= \frac{1}{n} \sum_{i=1}^n f(C_k/X_i) \\
 \bar{X}_k &= \frac{\sum_{i=1}^n f(C_k/X_i) X_i}{\sum_{i=1}^n f(C_k/X_i)} \\
 \sum_k &= \frac{\sum_{i=1}^n f(C_k/X_i) (X_i - \bar{X}_k)(X_i - \bar{X}_k)^T}{\sum_{i=1}^n f(C_k/X_i)}
 \end{aligned} \tag{1.8}$$

Où $f(C_k/X_i)$ représente l'estimation de la probabilité à posteriori d'être en présence d'une observation X_i de la classe C_k . Il obtenue par la formule de Bayes :

$$f(C_k/X_i) = \frac{f(X_i/C_k)P_k}{\sum_{j=1}^K f(X_i/C_j)P_j} \tag{1.9}$$

1.8 Segmentation par seuillage

1.8.1 Définition du seuillage

Le seuillage (thresholding en Anglais) représente un outil largement utilisé dans la segmentation d'image pour extraire des objets de leurs fonds en fonction d'un seuil. Tout problème de seuillage consiste alors à rechercher la valeur du seuil. La plupart des méthodes de seuillage déterminent le seuil en optimisant une fonction objective.

On distingue le Seuillage de base (simple) (2 classes) où le résultat du seuillage est une image binaire (0 : 1, parfois en 0 :255 pour l'affichage) (Figure 5), et le multi-seuillage (multi-level thresholding en Anglais) qui est utile quand on a affaire à des images qui contiennent plusieurs objets ayant des luminances différentes. Pour extraire ces objets, plusieurs seuils sont nécessaires. Le résultat du seuillage est une image avec $n + 1$ classes pour n seuils (Figure 6).

$$G(x, y) = \begin{cases} 1 & \text{si } f(x, y) \geq T \\ 0 & \text{si } f(x, y) \leq T \end{cases}$$

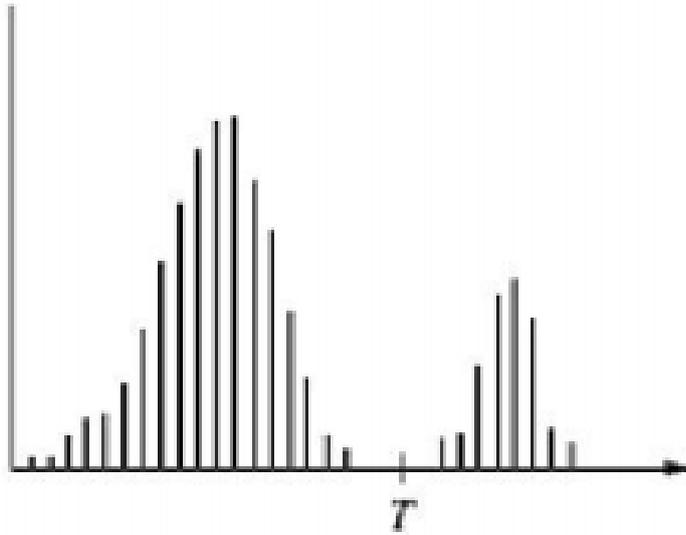


Figure 5 : Seuillage simple d'un histogramme.

$$G(x, y) = \begin{cases} 2 & \text{si } f(x, y) \geq T_2 \\ 1 & \text{si } T_1 \leq f(x, y) < T_2 \\ 0 & \text{si } f(x, y) < T_1 \end{cases}$$

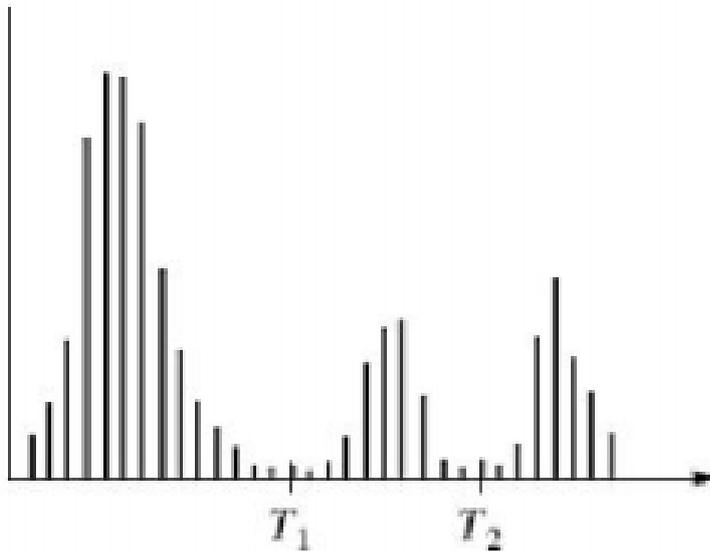


Figure 6 : Seuillage multiple d'un histogramme.

La segmentation par seuillage d'histogramme constitue un cas particulier de la segmentation par classification. Elle permet de répartir les pixels en classes en fonction

de leurs niveaux de gris. Les classes sont alors délimitées par des seuils.

Les méthodes de seuillage peuvent être réparties en deux catégories :

1. **Les méthodes non paramétriques** : qui ont la réputation d'être robustes; elles sont basées sur l'optimisation d'un ou plusieurs critères comme, par exemple, maximiser la variance interclasse. La plus connue d'entre elles est la méthode d'Otsu.
2. **Les méthodes paramétriques** : qui sont basées sur l'hypothèse que les densités de probabilité des niveaux de gris des différentes classes sont gaussiennes, une approximation de l'histogramme de l'image à traiter par une combinaison linéaire de gaussiennes est toujours effectuée avant l'opération de seuillage, ou les seuils optimaux sont à l'intersection de ces dernières.

Principe des méthodes paramétriques

Les méthodes de seuillage paramétrique supposent l'histogramme ($h(i)$) peut être approximé de probabilité (f.d.p) ($P_a(i)$) dont le modèle est connu à priori tel que :

$$P_a(i) = \sum_{k=1}^K P_k P(i, \theta_k) \quad (1.10)$$

Où $P(i, \theta_k)$ est la fonction de probabilité de la classe C_k ayant les paramètres statistiques θ_k et P_k la probabilité à priori de la classe C_k et qui peut être estimé par le nombre de pixels de la classe C_k sur le nombre total N de pixels.

$$P_k = \frac{h(i)}{N} \Big|_{i \in C_k}$$

Le problème du seuillage revient donc à estimer les probabilités à priori P_k et les paramètres statistiques θ_k de chaque f.d.p ou les seuils t_k en minimisant l'erreur quadratique moyenne suivante :

$$J(T) \approx J(\theta) = \frac{1}{L} \sum_{i=0}^{L-1} (P(i) - P_a(i))^2 \quad (1.11)$$

Souvent, les fonctions de densités de probabilités sont supposées suivre un modèle gaussienne de moyenne μ_k et variance σ_k tel que :

$$P(i, \theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\left(\frac{i - \mu_k}{\sigma_k}\right)^2\right) \quad (1.12)$$

P_k, μ_k et σ_k peuvent être estimés par :

$$P_k = \frac{1}{N} \sum_{i=t_{k-1}}^{t_k-1} h(i) ; \quad \mu_k = \frac{1}{N} \sum_{i=t_{k-1}}^{t_k-1} ih(i) ; \quad \sigma_k = \frac{1}{N} \sum_{i=t_{k-1}}^{t_k-1} (i - \mu_k)^2 h(i) \quad (1.13)$$

Le vecteur paramètre recherché est défini par :

$$\Theta = \{P_k, \mu_k, \sigma_k\} \quad \text{avec} \quad k = 1, \dots, K.$$

La minimisation de la fonctionnelle $J(\theta)$ ou $J(T)$ peut être résolue par l'algorithme d'estimation-maximisation ou les méthodes d'optimisation métaheuristique.

1.9 Exemples d'applications

La classification et la segmentation ont été utilisées dans de nombreuses applications dont certaines concernent : l'étude du sommeil ,le diagnostic médical ,le traitement d'images aériennes, la segmentation d'images RMN (Résonance Magnétique Nucléaire), le traitement des eaux usées ,l'étude du climat, les domaines de l'agriculture, l'astronomie, la chimie, la géologie, le traitement d'image, la reconnaissance de la maturité de squelettes à partir de radiographies aux rayons X, la reconnaissance de la parole, la reconnaissance des caractères écrits, le traitement automatique des documents,... etc.

1.10 L'importance de la segmentation en statistique décisionnelle

Dans cette section, nous avons intéressé par la segmentation marketing de votre marché comme exemple.

1.10.1 Segmentation marketing de votre marché

La **segmentation marketing d'un marché** peut se définir comme l'identification de sous-ensembles d'acheteurs partageant des besoins et des comportements d'achats similaires.

La Terre se compose de milliards d'acheteurs ayant leurs propres ensembles de besoins et comportements. La segmentation vise à réaliser des groupes d'acheteurs, aux besoins et comportements similaires. Un tel group est nommé (Segment).

1. La segmentation produit ne doit pas être confondue avec la segmentation clients

On peut découper un marché par circuits de distribution, par catégories de produits et par segments de clientèle. Par segmentation marketing, on entend la segmentation de clientèle. Une entreprise segmente pour mieux répondre aux attentes des clients, pour exploiter de nouvelles opportunités de marché et pour concentrer ses forces et son lieu de bataille.

2. La démarche de la segmentation clients ou segmentation marketing

- Segmenter un public consiste à le découper en groupes homogènes en fonction de critères déterminés, chacun de ces groupes étant distinct l'un de l'autre et pouvant être choisi comme cible d'une action marketing (toute segmentation doit être opérationnelle).
- Le terme de segmentation peut désigner à la fois le travail d'analyse et de découpage d'un marché et la politique qui vise à adapter son offre à certains segments. Par souci de clarté, on devrait réserver le terme de segmentation à l'analyse et celui de ciblage à la politique marketing d'adaptation à tout ou partie des segments définis.

3. Les principaux critères possibles de segmentation

Il existe de nombreux critères possibles de segmentation, qu'on peut regrouper en quatre catégories principales :

1. Les critères démographiques (sexe, âge, caractéristiques physiques, taille du foyer, composition de la famille...), géographiques (régions multinationales, régions nationales, taille de l'agglomération, climat, géodémographie...), sociaux et économiques (revenus, niveau d'instruction, profession, religion...).

2. Les critères de personnalité et de style de vie.
3. Les critères comportementaux : segmentation en fonction du statut d'utilisateur et de sa fidélité, segmentation du marché en fonction des quantités consommées, segmentation RFM (Récence/Fréquence/Montant), rentabilité des clients, modes de consommation, rôle dans le processus d'achat, segmentation en fonction des situations d'utilisation et du degré d'implication.
4. Les critères d'avantages recherchés

On recourt également souvent à la segmentation multicritères (croisement de plusieurs des critères ci-dessus).

4. Il existe deux grandes méthodes de découpage d'une population

1. La segmentation à partir de critères choisis a priori, qui vont permettre de découper et d'étudier les segments obtenus (segmentation proprement dite). Le choix a priori de critères peut être adapté lorsque les critères de segmentation sont simples et paraissent pertinents, comme c'est souvent le cas avec les variables sociodémographiques (exemple : pour le marché du jouet, les critères d'âge, de sexe, de niveau de revenus des parents...). Cette méthode cesse d'être applicable dès que le nombre de critères retenus simultanément devient élevé, générant ensuite un nombre de segments trop importants.
2. Les méthodes de typologie, ou segmentation statistique, consistent à regrouper les individus en fonction de leur proximité, mesurée par exemple par leurs réponses à un questionnaire ou par des caractéristiques observées. Il revient à l'analyste de savoir quel est le bon niveau de regroupement et donc de découpage de cette population, c'est-à-dire le nombre de segments auquel on s'arrête dans le traitement statistique. Il faudra ensuite décrire les segments obtenus par souci d'opérationnalité.

5. Bien segmenter signifie qu'on respecte quatre grandes conditions

1. Les conditions de pertinence par rapport au public : les segments obtenus doivent traduire des attitudes et des comportements différents.

2. Les conditions d'évaluation : les segments obtenus doivent être mesurables en termes de taille et de potentiel.
3. Les conditions d'opérationnalité : les segments obtenus doivent être atteignables, c'est-à-dire pouvoir se prêter à des politiques marketing adaptées de l'entreprise.
4. Les conditions de substantialité : les segments doivent être suffisamment substantiels pour justifier une adaptation des politiques de l'entreprise.

Chapitre 2

Méthodes de classification d'image

2.1 Introduction

La classification est une notion qui intervient fréquemment dans la vie courante. En effet, il est souhaitable de regrouper les éléments d'un ensemble hétérogène en un nombre restreint de classes, les plus homogènes possibles. Son application a joué un rôle très important pour résoudre plusieurs problèmes en reconnaissance des formes, imagerie, segmentation d'images couleur, data mining, et dans différents domaines comme la médecine, la psychologie, la biologie, etc....

Dans ce chapitre, nous avons donné un aperçu sur les différentes approches utilisées pour la classification : supervisées et non supervisées.

2.2 Méthodes supervisées

Ces sont des méthodes dans lesquelles les classes sont connues a priori avant d'effectuer l'opération d'identification des éléments de l'image. Elles demandent une phase d'apprentissage sur l'échantillon représentatif dans le but d'apprendre les caractéristiques de chaque classe et une autre phase pour décider l'appartenance d'un individu à telle ou telle classe.

Dans le cas qui nous intéresse, les données segmentées de l'ensemble d'apprentissage

proviennent d'un étiquetage manuel des images ou des régions d'intérêt en C classes de tissus $C_1, C_2, \dots, C_{c-1}, C_c$ par un ou plusieurs experts. Chaque classe C_i se voit donc affecter un ensemble d'apprentissage E_i , et les données de l'ensemble de test sont segmentées en fonction des E_i .

Puisque la structure anatomique d'un cerveau est différente d'un patient à l'autre, l'étiquetage doit être renouvelé pour chaque patient ou groupe de patients analysé, ce qui représente une tâche longue et fastidieuse pour les spécialistes.

2.2.1 Méthodes bayésiennes

La segmentation bayésienne consiste à calculer, pour chaque vecteur forme X_j , la probabilité conditionnelle $P(C_i/X_j)$ pour chacune des C classes ($C_1, C_2, \dots, C_{c-1}, C_c$) à l'aide de la règle Bayes :

$$\forall i \in \{1, \dots, c\} \quad P(C_i/X_j) = \frac{P(C_i) \cdot P(X_j/C_i)}{\sum_{k=1}^c P(C_k) \cdot P(X_j/C_k)} \quad (2.1)$$

La segmentation bayésienne admet :

- ★ Soit une approche paramétrique, dans laquelle l'intensité d'un pixel est considérée comme la combinaison linéaire de probabilités d'appartenance attachées à chaque classe (en effectuant l'hypothèse que les données sont conformes à des distributions paramétriques). La probabilité conditionnelle $P(X_j/C_i)$ est modélisée par une fonction dépendant d'un vecteur de paramètre θ . Le problème est alors pour chaque classe C_i d'estimer le meilleur paramètre θ connaissant l'ensemble d'apprentissage E_i de cette classe.
- ★ Soit une approche non paramétrique, dans ce type d'approche, les probabilités conditionnelles $P(X_j/C_i)$ sont supposées quelconques. La forme non paramétrique doit permettre de rendre compte le plus fidèlement possible de la réelle distribution statistique des niveaux de gris dans l'image.

2.2.2 Algorithme des k plus proches voisins

L'algorithme des k plus proches voisins (KPPV) (kNearestNeighbors (kNN) en anglais) est une méthode non paramétrique et supervisée de classification introduite dans [17]. Elle est largement utilisée en classification d'une manière générale et en segmentation d'images en particulier. Elle repose sur un principe simple et intuitif de regroupement d'individus en fonction de leur voisinage.

L'algorithme de K Plus Proche Voisin se base essentiellement sur les deux éléments principaux suivants :

1. le nombre de cas les plus proches (K) à utiliser et une métrique pour mesurer le plus proche voisin.
2. La valeur de K est spécifiée à chaque utilisation de l'algorithme puisqu'il détermine le nombre de cas existants qui sont considérés pour prédire un nouveau cas.

Le K Plus Proche Voisin est basé sur le concept de distances. Une métrique est nécessaire pour déterminer les distances, cette dernière est à la fois importante car le choix de métrique influe beaucoup sur la qualité des prédictions et arbitraire du fait qu'il n'existe pas de définition préalable sur ce qui constitue une bonne métrique.

La méthode des k plus proches voisins repose sur le regroupement des pixels en fonction de leur voisinage : chaque point est affecté à la classe la plus représentée parmi ses k plus proches voisins. Cette méthode requiert l'établissement d'une règle de distance et la détermination du nombre de voisins à prendre en considération, ainsi qu'un ensemble d'apprentissage représentant les différentes classes.

L'algorithme de KPPV est :

1. Choisir nombre de voisins (un entier k compris entre 1 et n , le nombre maximum de voisins).
2. Calculer les distances $d(v, v_i)$, $i = 1, \dots, n$ (où v_i est un des n pixels de l'image).
3. Retenir les K pixels pour lesquels ces distances sont les plus petites.
4. Compter le nombre de fois k_1, k_2, \dots, k_m que ces k pixels apparaissent dans chacune des classes.
5. Attribuer v à la classe la plus représentée dans son entourage.

2.2.3 Méthodes basées sur les réseaux de neurones artificiels

Un réseau de neurones est constitué d'un certain nombre d'unités élémentaires (neurones) fortement interconnectés (neurones) fonctionnant en parallèle et organisée en couches avec un flux d'informations à anticipation. L'architecture du perceptron multicouches est organisée de la manière suivante : les signaux flux successivement dans les différentes couches de l'entrée à la couche de sortie. Les couches intermédiaires sont appelées couches cachées. Pour chaque couche, chaque unité élémentaire calcule un produit scalaire entre un vecteur de poids et le vecteur de sortie donné par la couche précédente. Une fonction de transfert est alors réalisée à la suite de faire une entrée pour la couche suivante. Les poids des connexions sont ajustés au cours du processus d'apprentissage pour obtenir la relation d'entrée/sortie désirée du réseau. De nombreux modèles de réseaux existent (réseaux de Hopfield, perceptrons multicouche, etc.).

Les méthodes de segmentation supervisée offrent l'avantage d'être plus rapides et plus reproductibles que les méthodes non supervisées. Toutefois, leur inconvénient est de rester très dépendant de la base d'apprentissage. C'est pourquoi il est intéressant de développer des algorithmes entièrement automatiques "non supervisée", qui présenteront l'avantage d'offrir un résultat reproductible et indépendant des actions de l'opérateur.

2.3 Méthodes non Supervisées

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent **aucun apprentissage** et par là même aucune tâche préalable d'étiquetage manuel. Elles ont pour but de découper l'espace d'individus (pixels) en zones homogènes selon un critère de ressemblance (critère de proximité de leurs vecteurs d'attributs dans l'espace de représentation entre les individus).

Nous pouvons donc citer des algorithmes de classification pixels non-supervisés comme les **K-means** et **C-moyennes** proposée par **Mac Queen** et **C-moyennes floues**.

2.3.1 Algorithme des K-means

L'algorithme k-means est proposé par McQueen en 1967[6], un des plus simples algorithmes d'apprentissage non supervisé , appelée algorithme des centres mobiles [3] [4] , il attribue chaque point dans un cluster dont le centre (centroïde) est le plus proche. Le centre est la moyenne de tous les points dans le cluster , ses coordonnées sont la moyenne arithmétique pour chaque dimension séparément de tous les points dans le cluster c'est-à-dire chaque cluster est représentée par son centre de gravité.

L'algorithme K-means peut être résumé comme suit :

Entrée :

Ensemble de N données, noté par X.

Nombre de groupes souhaité, noté par K.

Sortie :

Une partition de K groupes C_1, C_2, \dots, C_k .

Début :

1. Initialisation aléatoire des centres C_k ;

Répéter

2. Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche ;

$$X_i \in C_k \text{ si } \forall j |X_i - b_k| = \min_j |X_i - b_j|$$

Avec b_k le centre de la classe K ;

3. Représentation : Calculer les centres associe à la nouvelle partition ;

$$b_k = \frac{1}{N} \sum_{X_i \in C_k} X_i$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

Fin.

2.3.2 Algorithme des C-moyennes

L'algorithme des C-moyennes (C-means) (centres mobiles) est l'un des algorithmes de clustering les plus connus, il a été mis au point par Mac Queen à la fin des années soixante [9]. L'algorithme des C-moyennes (C-means) [6] réalise un partitionnement automatique des pixels de l'image en C classes homogènes.

Soit $X = \{X_j\}$, $j = 1, 2, \dots, N$ un sous-ensemble d'un espace N -dimensionnel. Une partition de X en C classes peut être représentée par des sous groupes mutuellement disjoints C_1, C_2, \dots, C_C , tel que :

$$X = \bigcup_{i=1}^C C_i$$

Chaque vecteur X_j est assigné à une et seulement une classe qui lui est la plus proche.

La fonction objectif correspondante s'exprime sous la forme

$$J(B, U, X) = \sum_{i=1}^C \sum_{j=1}^N U_{ij} d^2(X_j, b_i) \quad (2.2)$$

Ou : b_i : désigne le prototype (centre) de cette classe.

d : mesure l'inverse de la similarité entre le prototype et le vecteur de données.

U_{ij} : est un indicateur binaire désignant l'appartenance ou non du vecteur X_j

à la classe C_i tel que :

$$U(X_j, b_i) = \begin{cases} 1 & \text{si } d^2(X_j, b_i) = \min\{d(X_j, b_k)\} \text{ pour } k = 1, \dots, C \\ 0 & \text{sinon} \end{cases}$$

où

$$b_i = \frac{\sum_{j=1}^N U_{ij} X_j}{\sum_{j=1}^N U_{ij}}$$

L'optimisation se fait de manière itérative.

Les étapes de cette classification pouvant être résumé comme suit :

1. Initialiser l'ensemble des centres avec des éléments, soit pris au hasard de l'ensemble de données, soit d'après des connaissances à priori.
2. Assigner chaque point au centre le plus proche.
3. Calculer les nouveaux centres des classes construites. Ils sont calculés par la moyenne des éléments de la classe.
4. Générer une nouvelle partition de l'ensemble de données en assignant chaque point au centre le plus proche.
5. S'il y a changement dans la partition, réitérer depuis l'étape 3.

2.3.3 Algorithme Fuzzy Cmeans (FCM)

Le Fuzzy Cmeans est un algorithme de classification non supervisée, issu de l'algorithme Kmeans. Son apport par rapport à ce dernier est l'introduction de la notion du flou, afin de prendre en compte l'imprécision des données [8]. Développé par Bezdek en 1981 à la suite des travaux de Dunn [12], l'algorithme FCM est un algorithme de réallocation floue, dans lequel les classes sont représentées par des prototypes (centres de gravité). Son application fournit donc pour chaque observation à classifier un degré d'appartenance (compris entre 0 et 1) à chaque classe, produisant ainsi une partition floue. Comme pour la plus part des autres algorithmes de classification par partition, le FCM est basé sur la minimisation d'un critère en suivant un processus itératif.

a) Principe

Soit $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$ un ensemble de N observations à classifier en K classes où chaque observation x_i ($i=1,2, \dots, N$) peut être représentée par un ensemble de D attributs : $x_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iD})^T$.

Les K classes sont représentées par un vecteur des centres des classes $V = \{v_1, v_2, \dots, v_i, \dots, v_K\}$ où $v_k = (v_{k1}, v_{k2}, \dots, v_{kj}, \dots, v_{kD})^T$ est le centre de la classe k .

Chaque observation x_i est définie par son degré d'appartenance u_{ki} à la classe k tel que $u_{ki} \in [0, 1]$. On peut alors définir une matrice de partition $U = [u_{ki}]$ de dimension $(K \times N)$.

L'algorithme FCM consiste à partitionner les N observations en K classes de manière à minimiser la similarité des observations à l'intérieur de chaque classe. Il se traduit par la minimisation de la fonction objectif suivante :

$$J_{FCM}(V, U, X) = \sum_{k=1}^K \sum_{i=1}^N (U_{ki})^m d^2(x_i, v_k) \quad (2.3)$$

Sous les contraintes suivantes :

$$0 < \sum_{i=1}^N U_{ki} < N \quad (2.4)$$

$$\sum_{k=1}^K U_{ki} = 1 \quad (2.5)$$

m : est le facteur de fuzzification ou facteur flou tel que $1 < m < \infty$.

$d(x_i, v_k) = \|x_i - v_k\|$ est la distance séparant l'observation i et le centre de la classe k .

Généralement la distance utilisée est euclidienne

$$d(x_i, v_k) = \sqrt{\sum_{j=1}^D (x_{ij} - v_{kj})^2} \quad (2.6)$$

La première contrainte assure qu'aucune classe ne doit être vide et la seconde est une contrainte de normalisation qui assure que la somme des degrés d'appartenance de chaque observation à toutes les classes soit égale à 1.

Le partitionnement flou est réalisé par une optimisation itérative de la fonction objectif donnée par l'équation (2.3), avec une mise à jour des degrés d'appartenance u_{ki} et des centres des classes comme dans le cas de l'algorithme K-means.

Les formules de mise à jour sont obtenues par l'introduction d'un multiplicateur de Lagrange λ associé à la contrainte de normalisation donnée dans l'équation (2.5).

En appliquant le Lagrangien par rapport à x_i , on aura :

$$L(x_i) = \sum_{k=1}^K u_{ki}^m \|x_i - v_k\|^2 + \lambda \left(\sum_{k=1}^K u_{ki} - 1 \right), \quad \lambda > 0 \quad (2.7)$$

La minimisation du Lagrangien par rapport aux degrés d'appartenance u_{ki} et au coefficient de Lagrange λ nous donne :

$$\frac{\partial L(x_i)}{\partial u_{ki}} = 0 \Rightarrow m(u_{ki}^{m-1})\|x_i - v_k\|^2 + \lambda = 0 \quad (2.8)$$

$$\frac{\partial L(x_i)}{\partial \lambda} = 0 \Rightarrow \sum_{k=1}^K u_{ki} - 1 = 0 \Rightarrow \sum_{k=1}^K u_{ki} = 1 \quad (2.9)$$

A partir de l'équation (2.8), on déduit :

$$u_{ki} = \left(\frac{-\lambda}{m}\right)^{\frac{1}{m-1}} \left(\frac{1}{\|x_i - v_k\|^2}\right)^{\frac{1}{m-1}} \quad (2.10)$$

et à partir de l'équation (2.9) et (2.10), on obtient :

$$\left(\frac{-\lambda}{m}\right)^{\frac{1}{m-1}} \sum_{l=1}^K \left(\frac{1}{\|x_i - v_l\|^2}\right)^{\frac{1}{m-1}} = 1 \iff \left(\frac{-\lambda}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{l=1}^K \left(\frac{1}{\|x_i - v_l\|^2}\right)^{\frac{1}{m-1}}} \iff \left(\frac{-\lambda}{m}\right) = \frac{1}{\sum_{l=1}^K \left(\frac{1}{\|x_i - v_l\|^2}\right)} \quad (2.11)$$

En remplaçant l'équation (2.11) dans l'équation (2.10), on aboutit à :

$$u_{ki} = \left(\frac{1}{\sum_{l=1}^K \left(\frac{1}{\|x_i - v_l\|^2}\right)}\right)^{\frac{1}{m-1}} \left(\frac{1}{\|x_i - v_k\|^2}\right)^{\frac{1}{m-1}} \iff u_{ki} = \left(\frac{1}{\sum_{l=1}^K \left(\frac{\|x_i - v_k\|^2}{\|x_i - v_l\|^2}\right)^{\frac{1}{m-1}}}\right)$$

D'où la formule de mise à jour des degrés d'appartenance :

$$u_{ki} = \sum_{l=1}^K \left(\frac{\|x_i - v_k\|^2}{\|x_i - v_l\|^2}\right)^{\frac{-1}{m-1}} \quad (2.12)$$

La minimisation du Lagrangien par rapport à la variable v_k représentant le centre de la classe k , s'écrit comme suit :

$$\frac{\partial L(x_i)}{\partial v_k} = 0 \iff \sum_{i=1}^N 2\|x_i - v_k\|u_{ki}^m = 0 \iff \sum_{i=1}^N u_{ki}^m x_i - \sum_{i=1}^N u_{ki}^m v_k = 0$$

D'où la formule de mise à jour des centres de classes :

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m} \quad (2.13)$$

L'algorithme FCM peut être résumé comme suit :

Algorithme général d'exécution du FCM

Étape 1 : Fixer les paramètres :

- Le nombre de classes K
- Le degré flou m , généralement pris égal à 2
- Choisir ϵ pour le test d'arrêt.

Étape 2 : Initialiser les centres des K classes de manière aléatoire.

Étape 3 : Initialiser le compteur $t = 0$.

Étape 4 : Mettre à jour la matrice U des degrés d'appartenance par la relation

$$u_{ki} = \sum_{l=1}^K \left(\frac{\|x_i - v_k\|}{\|x_i - v_l\|} \right)^{\frac{-2}{m-1}}$$

Étape 5 : Mettre à jour le vecteur V des centres des classes par la formule

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m}$$

Étape 6 : Comparer V^{t+1} à V^t si $|V^{t+1} - V^t| < \epsilon$ alors arrêt sinon $t = t + 1$ et retourner à l'étape 4.

2.4 Inconvénients de l'algorithme FCM

L'algorithme FCM a été largement étudié et a été utilisé dans bon nombre de domaines (segmentation d'images médicales, géologiques et satellitaires). Cependant cet algorithme qui nécessite la connaissance du nombre de classes, n'est pas robuste face aux bruits introduits par l'imprécision des attributs et son efficacité dépend fortement de l'étape d'initialisation des centres des classes car le processus itératif peut facilement fournir une solution localement optimale. De plus, il est basé sur la distance euclidienne lors de la mesure de similarité entre une observation et le centre d'une classe ce qui le rend utilisable que pour détecter des classes de forme sphérique. Afin d'éviter ces inconvénients donc améliorer les résultats de la classification, plusieurs modifications ont été apportées à l'algorithme standard et qui sont présentées comme des variantes du FCM.

Chapitre 3

Les variantes de l'algorithme FCM

3.1 Introduction

De nombreuses variantes ont été ainsi proposées, celles-ci consistent soit à changer la fonctionnelle à minimiser, soit à définir une autre distance ou soit encore à modifier l'influence du facteur flou. Nous les avons ainsi organisées en trois catégories :

- Variante selon la fonction objectif.
- Variante selon la distance.
- Variante selon la fuzzification.

3.2 Variante selon la fonction objectif

D'autres variantes du FCM agissent directement sur la fonction objectif afin d'améliorer les résultats de la classification notamment en présence du bruit. Celles-ci introduisent généralement un autre terme dans la fonction objectif de l'algorithme de FCM.

3.2.1 Algorithme d'Ahmed et al

Dans le cadre de la segmentation d'images en niveaux de gris, Ahmed et al [13] ont proposé une modification de l'algorithme de FCM standard en introduisant un terme qui fait intervenir l'influence des pixels voisins. Ce terme représente une contrainte spatiale

exercée par les pixels voisins sur chaque pixel courant en tenant compte de leurs distances au même centre de la classe. Si un pixel est proche d'un centre d'une classe et ses voisins affichent également une proximité pour ce même centre, alors son degré d'appartenance à cette classe va augmenter. Donc un pixel est contraint de suivre son voisinage, ce qui produit un effet de régularisation.

La fonction objectif proposée est :

$$J_A(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \|x_i - v_k\|^2 + \frac{\alpha}{\text{card}(N_i)} \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \left(\sum_{r \in N_i} \|x_r - v_k\|^2 \right) \quad (3.1)$$

N_i : représente l'ensemble des pixels spatialement voisins du pixel i et α un paramètre réel qui contrôle l'influence du terme de voisinage.

Chaque pixel i respecte la contrainte $\sum_{k=1}^K u_{ki} = 1$. La minimisation de la fonction objectif J_A est résolue en appliquant également le principe du Lagrangien et en annulant ses dérivées par rapport aux inconnues u_{ki} et v_k . Cette minimisation nous conduit à :

$$u_{ki} = \frac{(\|x_i - v_k\|^2 + \frac{\alpha}{\text{card}(N_i)} \sum_{r \in N_i} \|x_r - v_k\|^2)^{\frac{-1}{m-1}}}{\sum_{l=1}^K (\|x_i - v_l\|^2 + \frac{\alpha}{\text{card}(N_i)} \sum_{r \in N_i} \|x_r - v_l\|^2)^{\frac{-1}{m-1}}} \quad (3.2)$$

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m (x_i + \frac{\alpha}{\text{card}(N_i)} \sum_{r \in N_i} x_r)}{(1 + \alpha) \sum_{i=1}^N u_{ki}^m} \quad (3.3)$$

3.2.2 Algorithme de Chen et Zhang

Chen et Zhang [20] ont proposé une extension de la technique d'Ahmed et al, en simplifiant le terme de voisinage afin de réduire le temps de calcul. Leur fonction objectif s'exprime comme :

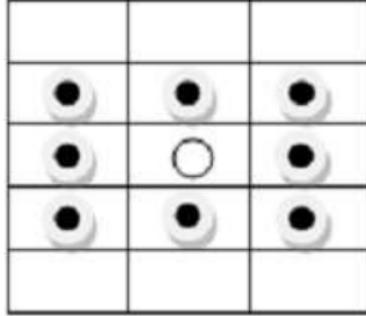
$$J_{CZ}(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \|x_i - v_k\|^2 + \alpha \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \|\bar{x}_i - v_k\|^2 \quad (3.4)$$

\bar{x}_i : représente la moyenne des pixels voisins qui se trouve à l'intérieur d'une fenêtre de voisinage centré sur i . La minimisation de la fonction objectif J_{CZ} suivant u_{ki} et v_k peut être tiré selon le même principe de la technique des multiplicateurs de Lagrange, en utilisant les équations suivantes :

$$u_{ki} = \frac{(\|x_i - v_k\|^2 + \alpha \|\bar{x}_i - v_k\|^2)^{\frac{-1}{m-1}}}{\sum_{l=1}^K (\|x_i - v_l\|^2 + \alpha \|\bar{x}_i - v_l\|^2)^{\frac{-1}{m-1}}} \quad (3.5)$$

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m (x_i + \alpha \bar{x}_i)}{(1 + \alpha) \sum_{i=1}^N u_{ki}^m} \quad (3.6)$$

Cet algorithme permet non seulement de réduire le temps d'exécution mais aussi d'être très performante pour segmenter des images contaminées par un bruit gaussien. Cependant, il s'est avéré moins robuste face à la présence du bruit de type poivre et sel. Afin de résoudre ce problème, les auteurs ont introduit une modification dans laquelle ils ont remplacé la moyenne des pixels voisins par leur valeur médiane.



La figure(1) : Voisinage d'un pixel.

Analogie à l'application du FCM avec un vecteur d'attributs {niveau de gris, moyenne} (ou {niveau de gris, médiane}).

3.2.3 Algorithme de Yang et al

La fonction objectif proposée par Yang et al [22] dans le cadre de la segmentation d'images en niveaux de gris est donnée par la relation suivante :

$$J_y(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^2 \|x_i - v_k\|^2 + \sum_{i=1}^N d_i \sum_{k=1}^K u_{ki} (1 - u_{ki}) \quad (3.7)$$

d_i désigne un paramètre de pondération de chaque pixel. Il peut être déterminé par la relation suivante :

$$d_i = \min\{d_{ki}^2 / k \in \{1, 2, \dots, K\}\} - \eta \quad (3.8)$$

η est un paramètre strictement positif ($\eta > 0$). Une valeur égale à 0.01 a été utilisée par les auteurs.

La minimisation de J_y par rapport à u_{ki} conduit à :

$$u_{ki} = \frac{1}{\sum_{l=1}^K \left(\frac{d_{ik}^2 - d_i}{d_{il}^2 - d_i} \right)} \quad (3.9)$$

La formule de mise à jour des centres de classes reste identique à celle de l'algorithme FCM classique avec $m = 2$.

3.2.4 Algorithme de Li et Mukaidono

Li et Mukaidono [19] proposent de rajouter à la fonction objectif de l'algorithme FCM un deuxième terme caractérisé par l'entropie du degré d'appartenance comme suit :

$$J_{LM}(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki} \|x_i - v_k\|^2 + \delta \sum_{k=1}^K \sum_{i=1}^N u_{ki} \log u_{ki} \quad (3.10)$$

δ est un coefficient réel strictement positif ($\delta > 0$).

La mise à jour des degrés d'appartenance est réalisée par la formule définie ci-dessous :

$$U_{ki} = \frac{\exp\left(-\frac{\|x_i - v_k\|^2}{\delta}\right)}{\sum_{l=1}^K \left(-\frac{\|x_i - v_l\|^2}{\delta}\right)} \quad (3.11)$$

Et les centres des classes sont déterminés comme dans l'algorithme FCM classique mais avec un facteur flou $m = 1$:

$$v_k = \frac{\sum_{i=1}^N u_{ki} x_i}{\sum_{i=1}^N u_{ki}}$$

3.2.5 Algorithme Cmoyennes possibiliste (PCM)

L'algorithme PCM proposé par Krishnapuram et Keller [18] introduit la notion de possibilité très voisine du flou. Le degré d'appartenance d'une observation à une classe est considéré comme une possibilité d'appartenir à une classe. Contrairement à l'algorithme FCM, la somme des degrés d'appartenance d'un pixel à toutes les classes n'est pas égale à 1. La fonction objectif ainsi modifiée est :

$$J_{PCM}(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m d_{ik}^2 + \sum_{k=1}^K \eta_k \sum_{i=1}^N (1 - u_{ki})^m \quad (3.12)$$

$$\sum_{k=1}^K u_{ki} > 1, \quad \forall i \in \{1, 2, \dots, N\}$$

η_k est un nombre positif à déterminer.

Le premier terme de l'équation (3.12) est identique à celui de FCM, le second force u_{ki} à la plus grande valeur possible.

En dérivant J_{PCM} par rapport à u_{ki}, v_k et en annulant cette dérivée, on obtient l'équation de mise à jour des degrés d'appartenance et les centres :

$$u_{ik} = \frac{1}{1 + \left(\frac{d^2(x_i, v_k)}{\eta_k}\right)^{\frac{1}{m-1}}} \quad (3.13)$$

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m} \quad (3.14)$$

Quant au paramètre η_k , il est estimé à chaque itération en utilisant l'équation suivante :

$$\eta_k = \frac{\sum_{i=1}^N u_{ki}^m d_{ik}^2}{\sum_{i=1}^N u_{ki}^m} \quad (3.15)$$

Les centres de classes sont mis à jour comme dans le FCM. Dans cet algorithme, les auteurs ont fixé le facteur flou à 1,5 et non pas 2 comme dans la plus part des cas.

3.3 Variante selon la distance

La version standard de FCM décrite précédemment utilise la distance euclidienne. Cependant cette distance donne de bons résultats que lorsque les classes sont de forme sphérique et ayant la même taille, ou lorsqu'elles sont bien séparées. Or en réalité les classes peuvent avoir des formes géométriques quelconques et des tailles différentes. Ainsi, plusieurs variantes du FCM utilisant d'autres types de distances ont été proposées.

3.3.1 Algorithme de Tsai et Lin

Afin d'améliorer l'efficacité de l'algorithme FCM, une nouvelle mesure de distance différente de celle de Mahalanobis et qui prend en compte la variation des distances dans chaque classe a été proposée par Tsai et Lin [5]. Cette distance s'exprime comme :

$$\hat{d}_{ik}^2 = \frac{\|x_i - v_k\|^2}{\sigma_k}, \quad i = 1, \dots, N, \quad k = 1, \dots, K \quad (3.16)$$

Où σ_k est la moyenne pondérée des distances dans la classe k, définie par :

$$\sigma_k = \left(\frac{\sum_{i=1}^N u_{ki}^m \|x_i - v_k\|^2}{\sum_{i=1}^N u_{ki}^m} \right)^{\frac{1}{2}} \quad (3.17)$$

La fonction objectif ainsi proposée est alors :

$$J_{TL}(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \frac{\|x_i - v_k\|^2}{\sigma_k} \quad (3.18)$$

Les équations des degrés d'appartenance et des centres des classes sont :

$$u_{ki} = \frac{1}{\sum_{l=1}^K \left(\frac{\hat{d}_{ik}}{\hat{d}_{il}} \right)^{\frac{2}{m-1}}}, \quad m \neq 1 \quad (3.19)$$

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m}, \quad \forall k \quad (3.20)$$

Cette distance a été appliquée sur l'algorithme FCM classique pour la classification dans l'espace de données et l'algorithme FCM basé sur les fonctions noyau. Les deux versions de FCM utilisant cette distance ont été appliquées à la segmentation d'images couleurs.

3.3.2 Algorithme FCM basé sur des fonctions noyau

a) Fonction noyau

Les fonctions noyaux sont devenues, durant ces dernières années, très populaires et largement utilisées en classification de données. L'idée générale, est de supposer que l'espace des données peut être plongé dans un espace de plus haute dimension (via une transformation non linéaire φ) dans lequel les données peuvent être séparées linéairement. La fonction noyau s'applique à tout algorithme pouvant s'exprimer sous forme d'une expression basée sur des produits scalaires entre les observations à classer.

L'ensemble d'observations, $X = \{x_1, x_2, \dots, x_N\}$ à classer peuvent être présentées dans un espace D de dimensions \mathcal{R}^D , tel que $x_i = \{x_{i1}, x_{i2}, \dots, x_{iD}\}$. Soit φ une fonction de transformation non linéaire permettant de projeter les observations initiales à classer vers un espace de plus grande dimension dit \mathcal{H} espace d'attributs, tel que : $\varphi : \mathcal{R}^D \mapsto \mathcal{H}$

L'idée est de supposer qu'on peut calculer le produit scalaire $\langle \varphi(x_i), \varphi(x_n) \rangle$, directement

sans jamais avoir à calculer explicitement $\varphi(x)$, en choisissant une fonction noyau O , telle que :

$$O(x_i, x_n) = \langle \varphi(x_i), \varphi(x_n) \rangle$$

Plusieurs fonctions noyaux sont proposées, Parmi elles on peut citer :

Noyau linéaire : $O(x_i, x_n) = x_i \cdot x_n$

Noyau Gaussien : $O(x_i, x_n) = \exp(-\frac{\|x_i - x_n\|^2}{2\sigma^2})$ où $\sigma > 0$.

Noyau de base radiale : $O(x_i, x_n) = \exp(-\frac{\|x_i - x_n\|^b}{\sigma^2})$ où $\sigma, a, b > 0$. Pour $a = 1$ et $b = 2$ ce noyau devient un noyau Gaussien.

b) La version noyau de l'algorithme FCM

L'exploitation de la fonction noyau dans l'algorithme de Fuzzy Cmeans, a donné naissance à une version dite Kernel FCM [15]. Il s'agit dans cet algorithme d'appliquer la transformation non linéaire sur les données à classer afin de les projeter dans un espace de plus grande dimension dans lequel les classes deviennent linéairement séparables. L'algorithme KFCM modifie la fonction objectif du FCM via la transformation φ , comme suit :

$$J_{KFCM}(U, V) = \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m \|\varphi(x_i) - \varphi(v_k)\|^2 \quad (3.21)$$

Où $\|\varphi(x_i) - \varphi(v_k)\|^2$ est le carré de la distance entre $\varphi(x_i)$ et $\varphi(v_k)$.

La distance dans l'espace d'attributs entre les transformées d'une observation $\varphi(x_i)$ et d'un centre d'une classe $\varphi(v_k)$ est donnée par :

$$\varphi_{d_{ik}^2} = \|\varphi(x_i) - \varphi(v_k)\|^2 = [\varphi(x_i) - \varphi(v_k)][\varphi(x_i) - \varphi(v_k)] = O(x_i, x_i) - 2O(x_i, v_k) + O(v_k, v_k)$$

En utilisant le noyau gaussien à base radiale, c'est-à-dire $O(x, y) = \exp(-\frac{\|x-y\|^2}{\sigma^2})$ on en déduit que $O(x, x) = 1$ et $\varphi_{d_{ik}^2} = 2(1 - O(x_i, v_k))$.

La fonction objectif de l'équation (3.20) devient donc :

$$J_{KFCM}(U, V) = 2 \sum_{k=1}^K \sum_{i=1}^N u_{ki}^m (1 - O(x_i, v_k)) \quad (3.22)$$

D'une manière similaire à l'algorithme FCM, la fonction objectif J_{KFCM} peut être minimisée sous la contrainte (2.3). Plus précisément en prenant les dérivées premières

du Lagrangien de J_{KFCM} par rapport à u_{ki} et v_k égales à zéros, on aboutit à La matrice de partition floue définie par son élément général :

$$u_{ki} = \sum_{l=1}^K \frac{(1 - O(x_i, v_k))^{\frac{1}{m-1}}}{(1 - O(x_i, v_l))^{\frac{1}{m-1}}} \quad (3.23)$$

Et aux centres des classes v_k définissant le vecteur V dont le k élément est :

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m O(x_i, v_k) x_i}{\sum_{i=1}^N u_{ki}^m O(x_i, v_k)} \quad (3.24)$$

3.4 Variante selon la fuzzification

Une autre classe de variantes du FCM concerne la modification de l'influence du facteur flou m .

3.4.1 L'algorithme Miin.Shen Yang (MSFCM)

Suppressed Fuzzy C-Means (S-FCM) proposé par Fan et al.(2003), est un algorithme dérivé de FCM visant l'amélioration de la qualité de la segmentation pour un volume important de données. L'algorithme en question ajoute un nouveau paramètre α . La seule modification apportée à l'algorithme de base (FCM) est d'ajouter une nouvelle étape, après le calcul de la matrice U . Cette nouvelle étape est : si le point x_j appartient à une classe p et si son degré d'appartenance à celle-ci est le plus large (si on a plusieurs degrés alors choisir un au hasard.) alors modifie U_{pj} par :

$$U_{pj} = 1 - \alpha \sum_{i \neq p} U_{ij}, \quad U_{ij} = \alpha U_{ij} \quad (3.25)$$

L'algorithme Miin.Shen Yang et al [14] ont proposé un nouvel algorithme appelé MSFCM (Modified S-FCM) pour choisir la valeur de α la plus adéquate en se basant sur la capacité d'apprentissage de la machine sur la base d'autres expériences précédentes. L'apprentissage du paramètre se base sur la potentialité de séparation entre les classes et sur sa mise à jour à chaque itération.

L'algorithme MS-FCM se résume ainsi :

1. Fixer le nombre de classe tel que $1 < c < n$; n étant le nombre de pixels. Fixer une valeur de m telle que $m > 1$. Fixer le seuil de stabilité ϵ
2. Initialiser les c centroïdes.
3. Initialiser le compteur $t = 0$.
4. Calculer α à partir de v_i :

$$\alpha = \exp\left(-\min_{i \neq k} \frac{\|v_i - v_k\|^2}{\beta}\right) \quad (3.26)$$

Avec $\beta = \frac{\sum_{j=1}^n \|x_j - \bar{x}\|^2}{n}$ et $\bar{x} = \frac{\sum_{j=1}^n x_j}{n}$

5. Calculer u_{ij}^t , en utilisant l'équation suivant :

$$u_{ij} = \sum_{l=1}^K \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_l\|^2} \right)^{\frac{-2}{m-1}}$$

6. Modifier u_{ij}^t , en utilisant l'équation (3.25).
7. Mettre à jour les c centroïdes de classes V^t à l'aide de u_{ij} modifié, en utilisant l'équation suivant :

$$v_k = \frac{\sum_{i=1}^N u_{ki}^m x_i}{\sum_{i=1}^N u_{ki}^m}$$

8. Calculer $E^t = |V^t - V^{t-1}|$, si $E^t < \epsilon$, arrêter ; sinon $t = t + 1$.

3.4.2 l'algorithme de Mohamed Adel Alimi et al

Mohamed Adel Alimi et al [10] ont proposé un nouvel algorithme appelé Robust Modified Suppressed Fuzzy C-means (RMS-FCM). Cette algorithme est le même que MS-FCM, la seule différence est la modification du paramètre α qui diffère d'un point à un autre.

Le nouveau paramètre qui va remplacer α (3.26) par un poids ou chaque pixel de l'ensemble des données a une valeur par rapport à la classe qu'elle appartienne est calculer par :

$$\alpha_{ij} = \exp\left(\frac{-\|x_j - v_i\|^2}{\sum_{j=1}^n \|x_j - \bar{x}\|^2} \times \frac{n}{c}\right) \quad (3.27)$$

Dans la suite, on va appliquer la même chose qui faire dans l'algorithme MS-FCM.

Chapitre 4

Tests et résultats

4.1 Introduction

La segmentation est une étape indispensable dans de nombreuses chaînes de traitement de plusieurs domaines fondamentaux de la recherche clinique. La qualité de l'interprétation d'une image dépend fortement de celle de la segmentation qui est une étape de base du traitement d'une image. Parmi les images qu'on a utilisées dans ce mémoire les images médicales. L'information apportée par l'imagerie médicale est d'un apport considérable en matière de diagnostic.

Dans ce chapitre nous appliquons les méthodes étudiées dans les chapitres précédents. Cependant avant d'appliquer ces méthodes nous présentons une brève description des images médicales utilisées.

4.2 L'imagerie médicale

L'imagerie médicale est un ensemble de techniques consistant à mettre en image différentes régions ou différents organes de l'organisme. Il existe plusieurs types d'imageries médicales qui sont plus ou moins adaptées en fonction des zones à étudier. On distingue notamment la radiologie, qui utilise les rayons X pour explorer les os, les poumons, l'abdomen ou système digestif, l'échographie, qui se sert des ultrasons pour explorer les

organes pleins et qui permet d'observer le fœtus au cours de la grossesse ; l'imagerie par résonance magnétique (IRM) qui permet d'obtenir des images en coupes du corps : le scanner qui donne également des images en coupe et qui permet de réaliser la représentation en 3D.

4.2.1 Imagerie par résonance magnétique (IRM)

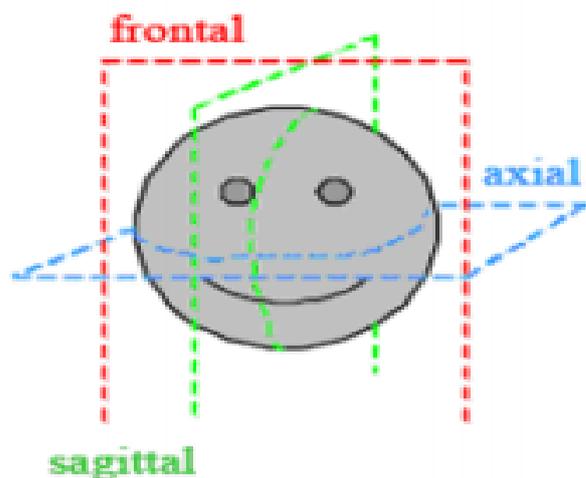
L'imagerie par résonance magnétique nucléaire (IRM) est une technique d'imagerie médicale d'apparition récente, non invasive et sans effets secondaires permettant d'avoir une vue 2D ou 3D d'une partie du corps, notamment du cerveau. Elle est basée sur le phénomène physique de résonance magnétique nucléaire.

Le principe consiste à mesurer l'aimantation des tissus biologiques et à reconstruire une image à partir de ces aimantations. Cette dernière est induite par l'hydrogène, qui s'y trouve en abondance dans le corps humain.

4.2.2 L'observation du cerveau

Le cerveau est la partie la plus volumineuse du système nerveux central.

L'observation des coupes axiale, frontale et sagittale sont des coupes du cerveau approximativement parallèles, respectivement, au plan qui comprend nez et oreilles, au plan du visage et plan de symétrie de la tête. Ces coupes sont orthogonales deux à deux (Voir la figure (1)).



Figure(1) : Plan de coupe en imagerie médicale

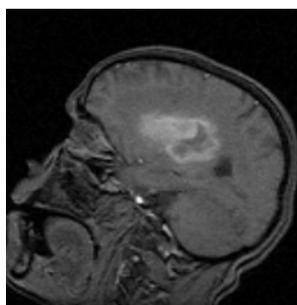
4.3 Présentation des images médicales utilisées

Nous présentons quelque information sur le diagnostic de ces images médicales nécessaire pour savoir la qualité de la classification.

Dans cette chapitre, nous travaillons sur l'image médicale d'un cerveau humain avec les différents coupes : axiale, frontale et sagittale (Voir la Figure (2)).



(a)



(b)



(c)

Figure (2) : Les trois coupes du cerveau : (a) : coupe axiale, (b) :coupe sagittale et (c) : coupe frontale.

4.4 Les critères utilisées

4.4.1 Coefficient de Partition (Vpc)

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

$$V_{pc}(U, C) = \frac{1}{N} \sum_{k=1}^C \sum_{i=1}^N U_{ik}^2 \quad \text{à maximiser}$$

Ce coefficient est borné, $V_{pc} \in [\frac{1}{C}, 1]$ Cette gamme a été obtenue à partir de deux cas extrêmes. Le premier se produit lorsque les degrés d'appartenance de chaque élément sont soit égaux à 0 soit à 1, un élément appartenant donc à une seule classe. Dans tel cas, la valeur Vpc est égale à 1, indiquant que toutes les classes sont bien délimitées. Dans le second cas, chaque élément appartient à toutes les classes avec le même degré d'appartenance et Vpc indiquant que la classification est la plus floue possible. Par conséquent, avec l'augmentation de la qualité de la classification, la valeur de Vpc augmente également.

4.4.2 Entropie de Partition (Vpe)

L'entropie de partition, Vpe ne dépendant également que des degrés d'appartenance, a été définie toujours par Bezdek comme :

$$V_{pe}(U, C) = -\frac{1}{N} \sum_{k=1}^C \sum_{i=1}^N U_{ik} \log_a(U_{ik}) \quad \text{à minimiser.}$$

Où a est la base du logarithme. Cette mesure est également bornée, $V_{pe} \in [0, \log_a(C)]$. D'une façon similaire, il y a deux cas extrêmes qui forment la gamme de valeurs possibles pour Vpe. Lorsque les classes sont bien séparées, la valeur de Vpe est proche du zéro, en revanche, lorsque la classification est plus floue, la valeur de H s'approche du $\log_a(C)$.

4.4.3 Indice de FUKUYAMA et SUGENO (Vfs)

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

$$Vfs(U, C) = \sum_{k=1}^N \sum_{i=1}^C U_{ik}^m (\|x_k - v_i\|^2 - \|v_i - \bar{v}\|^2) \text{ à minimiser}$$

Avec : $\bar{v} = \frac{1}{C} \sum_i v_i$ Le premier terme combine les degrés d'appartenances aux classes avec la compacité géométrique de la représentation des données x_i par les centre de classe v_i . Le second terme combine les degrés d'appartenances avec la distance entre le prototype k et la moyenne globale des données.

4.4.4 Indice de XIE et BENI (Vxb)

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

$$Vxb(U, C) = \frac{\sum_{k=1}^N \sum_{i=1}^C U_{ik}^m \|x_k - v_i\|^2}{N \times \min_{i,j,i \neq j} \|v_i - v_j\|^2} \text{ à minimiser}$$

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

4.4.5 Critère de Max et Min-Max

– *Max* :

$$MAX(U, C) = \frac{1}{N} \sum_k \max_i U_{ik} \text{ à minimiser}$$

Ce paramètre représente l'appartenance moyenne d'un point à la classe dont il est le plus proche.

– *Min-Max* :

$$MAXMIN(U, C) = \frac{1}{N} \sum_k \frac{\min_i U_{ik}}{\max_i U_{ik}} \text{ à minimiser.}$$

Ce paramètre donne une information quant à la forte appartenance d'un point à la classe dont il est le plus proche, et la faible appartenance de ce point à la classe dont il est le plus loin.

4.5 L'application de l'algorithme K-means

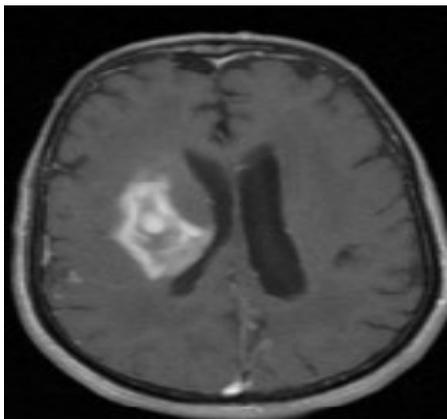
Nous avons choisi de débiter par l'algorithme k-means car c'est le classificateur non supervisé le plus simple et le plus utilisé.

Au début, nous allons commencer l'algorithme k-means en utilisant la fonction `wssplot()` dans R pour trouver la valeur k optimal des images (figure (2).(a); figure (2).(b); figure (2).(c)).

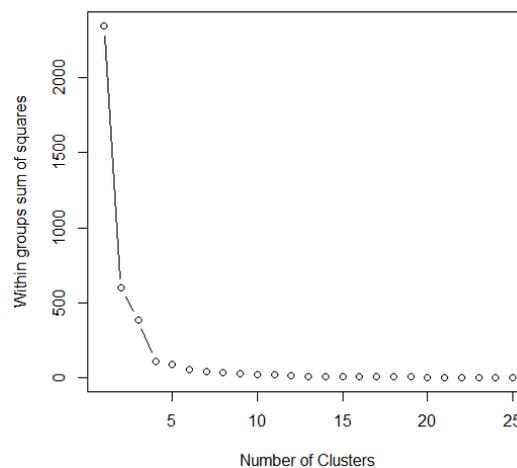
```
#La fonction WSSPLOT
wssplot <- function(data, nc=15, seed=1234){
  wss <- (nrow(data)-1)*sum(apply(data,2,var))
  for (i in 2:nc){
    set.seed(seed)
    wss[i] <- sum(kmeans(data, centers=i)$withinss)}
  plot(1:nc, wss, type="b", xlab="Number of Clusters",
       ylab="Within groups sum of squares")}

wssplot(img_RGB[c(3,4,5)],25)
```

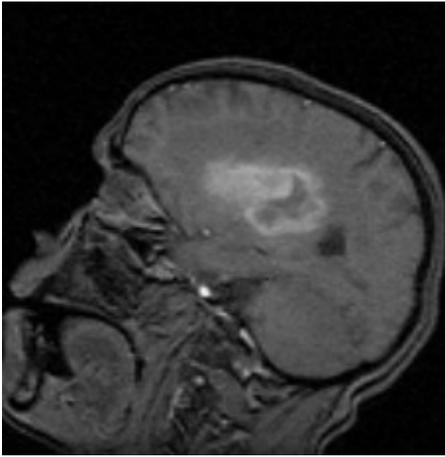
La fonction `wssplot()` sur R



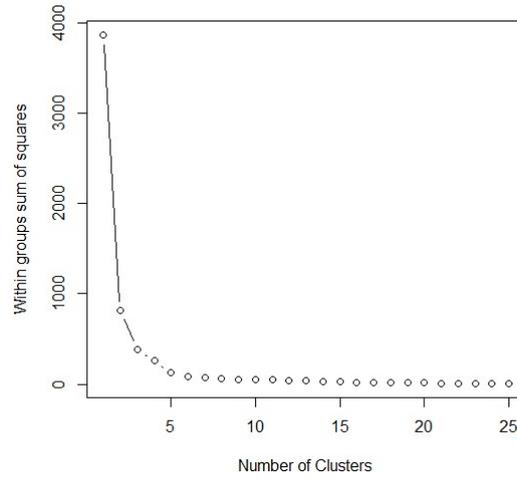
(a)



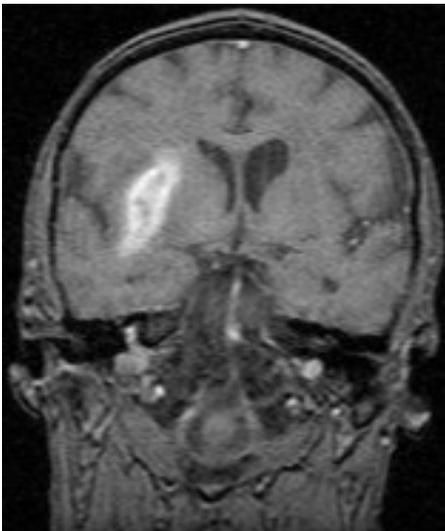
Résultat de l'image (a)



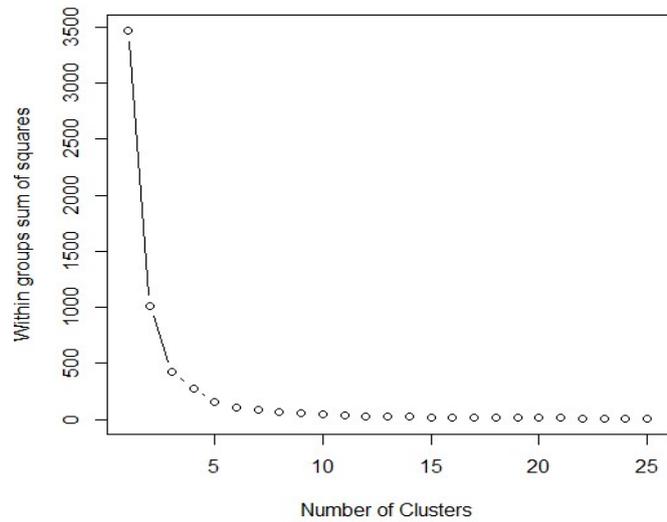
(b)



Résultat de l'image(b)



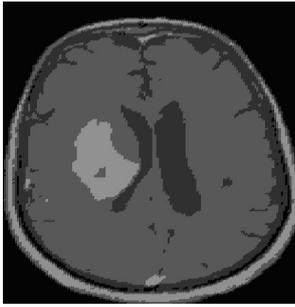
(c)



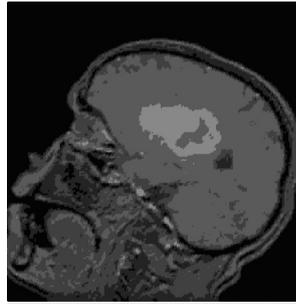
Résultat de l'image (c)

A partir les tarces de WSS en haut, nous pouvons voir que la prise de nombre de classe $k=4$ devrait être le nombre optimal des classes pour l'image (a) et $k=5$ devrait être le nombre optimal des classes pour les deux images (b) et (c).

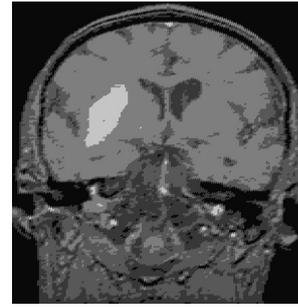
En suite, nous appliquons les résultats trouver dans l'algorithme k-means pour les tois images. Nous obtenons les résultats suivant :



a) Segmentation pour $k=4$



b) Segmentation pour $k=5$



c) Segmentation pour $k=5$

4.6 Comparision entre l'algorithme FCM et ses variantes

L'étude que nous menons est faite avec le logiciel Matlab, Les paramètres de test pour les différents algorithmes, nous choisissons les valeurs suivantes : $m = 2$, Nombre maximal d'itérations=300, $\epsilon = 1e - 5$.

Dans cette section, nous comparisons entre l'algorithme FCM et quelques variantes de FCM comme : l'algorithme TSAI et LIN, l'algorithme MSFCM et l'algorithme PCM. Nous avons utilisé les critères de validité, le nombre de itération et le temps d'excutioin à chaque algorithme pour la comparisation ; puis nous regardons l'algorithme qui donner la meilleur résultat.

Maintenant, nous utilisons les résultat qui trouver par l'algorithme k-means dans la section précédent pour les trois images.

Nous allons faire le test sur les trois vaiantes de l'algorithme FCM (Tsai et Lin, MSFCM et PCM) pour chaque image ; après on va noté les résultats suivantes :

♥ **Pour l'image (a) :**

Image	Algorithme	Vpc	Vpe	MAX	MIN_MAX	Vxb	Vfs
	MSFCM	0,64061291	0,59728458	0,76317361	0,04818035	0,943385872	-697300016
	TSAI ET LIN	0,91702376	0,12700164	0,94789604	0,00274691	0,028406629	-1482540810
a)	PCM	0,00219892	0,04988783	0,02508628	0,00483381	3,50E-07	-1637368938
	FCM	0,93410928	0,09792646	0,9570251	0,00483179	0,001858076	-1631784385

Figure (3) : Le résultat obtenu pour l'image (a)

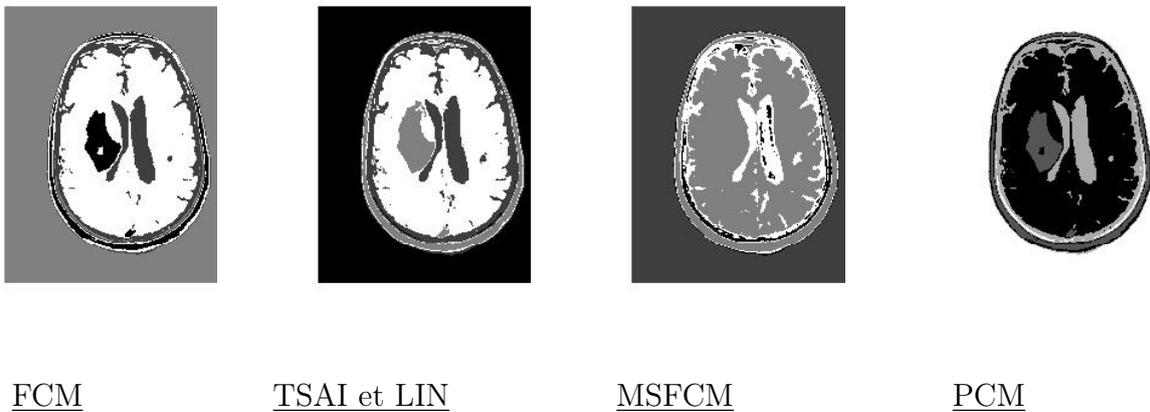
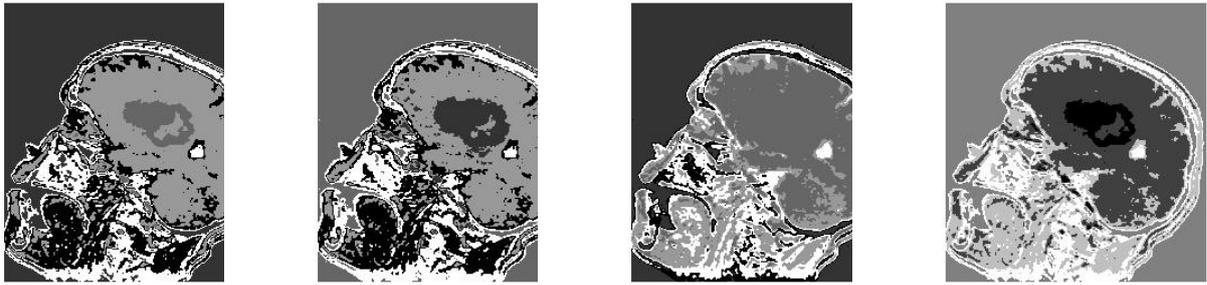


Figure (4) : Les résultats de segmentation d'image (a).

♡ Pour l'image (b) :

IMAGE	Algorithme	Vpc	Vpe	MAX	MIN_MAX	Vxb	Vfs
	MSFCM	0,80944841	0,23917426	0,87266393	0,00443409	0,066010915	-1545791244
b)	TSAI ET LIN	0,54971696	0,65000843	0,67962555	0,04177907	0,609404799	-987807268
	PCM	0,01022372	0,03225251	0,02872985	0,00556979	1,26E-07	-1761216715
	FCM	0,83560681	0,20528801	0,88911569	0,00556698	0,011736092	-1755470188

Figure (5) : Le résultat obtenu pour l'image (b)



FCM

TSAI et LIN

MSFCM

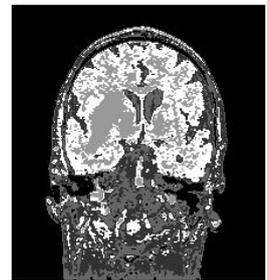
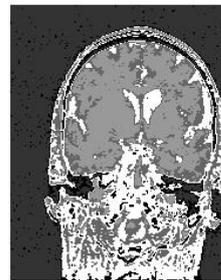
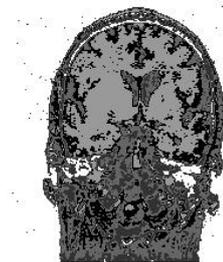
PCM

Figure (6) : Les résultats de segmentation d'image (b).

♡ Pour l'image (c) :

IMAGE	Algorithme	Vpc	Vpe	MAX	MIN_MAX	Vxb	Vfs
	MSFCM	0,63165133	0,52907368	0,75344648	0,02291725	0,563144082	-1662232089
c)	TSAI ET LIN	0,7806923	0,25370151	0,85145449	0,00407756	0,036994024	-1736301510
	PCM	0,00831324	0,02605797	0,02320137	0,00692729	6,50E-08	-1667411387
	FCM	0,83801913	0,20126776	0,88943783	0,00692427	0,004651267	-1658451327

Figure (7) : Le résultat obtenu pour l'image (c)



FCM

TSAI et LIN

MSFCM

PCM

Figure (8) : Les résultats de segmentation d'image (c).

A partir des figures : (3), (5) et (7), nous pouvons voir que les tests effectués sur les trois images montrent que la plupart des critères utilisés prennent l'algorithme PCM plusieurs fois parmi les autres algorithmes.

Ensuite nous avons calculé le nombre d'itération et le temps d'exécution pour les trois variantes de FCM (voir la figure (9)).

IMAGE	Algorithme	Nombre d'iteration	Temps d'exécution(seconds)
a)	MSFCM	48	26.640
	TSAI ET LIN	30	30.425
	PCM	33	5.412
	FCM	42	3.963
b)	MSFCM	120	34.458
	TSAI ET LIN	189	141.065
	PCM	83	17.133
	FCM	79	9.739
c)	MSFCM	84	56.345
	TSAI ET LIN	81	44.608
	PCM	213	38.760
	FCM	203	21.877

figure (9) : Présente le nombre d'itérations et le temps d'exécution pour chaque algorithme

D'après les figures (9), Nous déduisons que l'algorithme PCM reste le meilleur parmi les autres algorithmes avec un nombre d'itération minimal et le temps d'exécution plus court.

Conclusion générale

Le travail présenté dans ce mémoire concerne le domaine du traitement d'image et précisément celui de la segmentation et classification d'image basée sur l'algorithme Fuzzy C-means et ses variantes.

Nous avons présenté dans ce travail quelques notions sur la segmentation d'image. Il ressort que le domaine de la segmentation est tellement riche et diversifié qu'il est difficile de donner une étude exhaustive.

Celles-ci se déclinent en trois grandes approches. Nous nous sommes particulièrement intéressés à l'approche région dans laquelle des méthodes sont basées sur la classification et plus précisément de segmenter l'image IRM par les deux algorithmes K-means et fuzzy C-means.

Aussi, nous avons donné un aperçu sur les différentes approches utilisées pour la classification : supervisées et non supervisées. Nous avons vu que la segmentation (clustering) permet de regrouper des objets (individus ou variables) en un nombre limité de groupes ou de classes (segmentes, ou cluster).

Dans ce mémoire aussi, nous avons étudié un nombre exhaustif de ces variantes qui ont été proposées dans la littérature pour remédier au problème de la sensibilité au bruit de l'algorithme FCM classique. Celles-ci tentent soit à changer la fonctionnelle à minimiser, soit à définir une autre métrique ou soit encore à modifier l'influence du facteur flou m . Parmi ces variantes : TSAI et LIN, PCM, MSFCM.

Nous avons relevé dans la littérature plusieurs méthodes modifiant l'algorithme FCM.

Trois d'entre elles seulement ont été implémentées et testées. Il serait donc souhaitable d'étendre notre étude à d'autres variantes comme celles qui agissent sur le facteur flou ou celles qui utilisent un autre type de distance ou combiner des méthodes appartenant à des catégories différentes.

Bibliographie

- [1] Amir NAKIB, *Conception de métaheuristiques d'optimisation pour la segmentation d'images. Application à des images biomédicales*, thèse de doctorat De l'université Paris , 12-val de marne, 2007.
- [2] A. Tremeau et N. Borel, *A Region Growing and Merging Algorithm to Color Segmentation*. Dans *Pattern Recognition*, vol. 30, no. 7, pp. 1191-1204, 1997.
- [3] Benzécri J.P. L'analyse des données. Dunod, Paris, 197.
- [4] Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., and Ralambondrainy, H. Classification automatique des données, environnement statistique et informatique. DUNOD informatique. 1989.
- [5] D. M. Tsai et C. C. Lin, fuzzy c-means based clustering for linearly and non linearly separable data. *Pattern recognition* 44, 1750-1760, 2011.
- [6] J. B. MacQueen (1967), *Some Methods for classification and Analysis of Multivariate Observations* , Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability , Berkeley, University of California Press, no 1, pp281-297.1967.
- [7] J. Canny, *computational approach to edge detection*, IEEE Trans. On Pattern Analysis and Machine Intelligence , vol. 8, n°6, p. 679-698, novembre 1986.
- [8] J. C. Bezdek, *pattern recognition with fuzzy objective function*. Plenum press, New York, 1981.

- [9] Kass, M., Witkin, A. et Terzopoulos, D. (1987). *Snakes : active contour models*. International Journal of Computer Vision,1(4) ,321 – 331.
- [10] Mohamed Adel Alimi, Fatma Kareem et Mohamed Fadhel Saad, Classification des Images IRM par Robust Modified Suppressed Fuzzy C-means, 2006.
- [11] M. ANDRE, *Introduction aux techniques de traitement d'images* , Eyrolles 1987.
- [12] M. F. Saad, F. Kareem et M. A. Alimi, Classification des images IRM.
- [13] M. N. Ahmed, S. M. Yamani, N. Mohammed, A. Farag et T. Moriarty, A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans. On medical imaging, 21(3), pp. 193-199, 2002.
- [14] M Yang W.L. Hung and D. Chen. Parameter selection for suppressed fuzzy c-means with an application to mri segmentation. Pattern Recognition Letters, 2005.
- [15] P. Kaur, A. k. Soni et A. Gosain, Robust kernalized approach to clustering by incorporating po new distance measure. Engineering App. of art. Int. Elsevier 2012. pp. 833-847.
- [16] R. Deriche, *Using Canny's criteria to derive a recursively implemented optimal edge detector* , International Journal of Computer Vision, p. 167-187, 1987.
- [17] R. DUDA, P. HART, *Pattern Classification and Scene Analysis*. New York, 1973.3796-4/1997. IEEE.
- [18] R. Krishnapuram et J. Keller, A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. 1 (2), 98-110.
- [19] R. Li et M. Mukaidono, A maximum entropy approach to fuzzy clustering estimation. IEEE, International Conference on Fuzzy Systems, 1995.
- [20] S. Chen et D. Zhang, Robust image segmentation using fuzzy c-means with spatial constraints based on new kernel-induced distance measure. IEEE Trans. Trans. Syst. Man. Cybern, vol. 34, pp. 1907-1916, 2004.
- [21] S.L. Horowitz et T.PAVLIDIS, *Picture segmentation by traversal algorithm*, J. ACM, 23, 368-388, 1976.

- [22] Z. Yang, F-L. Chung et W. Shitong, Robust fuzzy clustering based image segmentation. Applied soft computing, 2008.