

# Segmentation d'images par la Classification floue basée sur la programmation DC et DCA

Le Thi Hoai An\*, Le Hoai Minh\*,  
Nguyen Trong Phuc\*, Pham Dinh Tao\*\*

\* LITA – UFR MIM  
Université Paul Verlaine – Metz  
Ile du Saulcy  
57045 Metz Cedex 01, France  
{lethi,lehoai,nguyen}@univ-metz.fr

\*\* Equipe Modélisation, Optimisation et Recherche Opérationnelle,  
INSA de Rouen  
Place Emile Blondel BP 08  
76131 Mont-Saint-Aignan Cédex  
pham@insa-rouen.fr

**Résumé.** Dans ce papier, nous proposons une nouvelle méthode de segmentation d'image via la classification floue, et basée sur la programmation DC (Difference of Convex functions) et DCA (DC Algorithm). En exploitant un schéma de DCA simple et robuste pour la classification floue, nous proposons un nouveau modèle par l'introduction des informations spatiales au modèle FCM standard. DCA appliqué au nouveau modèle est de la même forme que DCA pour le modèle FCM standard, et avec les informations spatiales, notre algorithme améliore nettement la segmentation des images bruitées. Pour trouver un bon point initial de l'algorithme, nous proposons une procédure alternative de DCA et FCM. Les simulations numériques sur plusieurs imageries médicales, qui sont en très haute résolution issue d'un signal, montrent l'efficacité de notre approche par rapport aux méthodes standards sur le temps de calcul et la qualité des solutions.

## 1 Introduction

La segmentation d'image joue un rôle important dans une variété d'applications telles que la médecine, la géologie, la biométrie et la bureautique. Le premier champs d'application est le traitement de l'imagerie médicale qui est fondé sur des images de l'intérieur du corps d'un patient (échographies, radiographies,...). En effet, on peut repérer sur ces images la présence d'anomalies ce qui permet de détecter certaines maladies. Par exemple, la détection de micro-calcifications dans une mammographie peut relever la présence d'un cancer du sein. La géologie applique des cartes dont les

## Segmentation d'images par la Classification floue basée sur DCA

couleurs peuvent représenter la densité de population, le climat. En plus, le champs d'applications de la segmentation d'image peut s'élargir dans bien d'autres domaines tels que : le suivi de forme dans des documents vidéo, la détection d'objets stellaires dans des images astronomiques, la détection de fronts dans les images satellites pour l'assimilation de données en météologie.

La segmentation d'image est une étape primaire dans la plupart des applications de la vision d'ordinateur à l'analyse d'image. C'est une des tâches les plus importantes de la phase de pré-traitement. L'identification d'objets réels, de pseudo-objets et d'ombres, ou la recherche de tout élément d'intérêt présent dans l'image, tous nécessitent une forme de segmentation. La segmentation se définit comme un processus qui consiste à découper une image en régions connexes présentant une homogénéité selon un certain critère, comme par exemple les critères de texture et/ou de couleur. Ce processus est connu par sa complexité en raison de la subjectivité de la définition de régions connexes, et de l'objectif qu'on veut atteindre de notre image segmentée. Par exemple, dans un mur de brique, doit-on considérer que chacune des briques forme une région autonome ?

Dans ce domaine, de nombreuses méthodes basées sur différentes approches, basées sur le contour, la région ou la texture, ont été développées au cours de ces dernières années (Pal et Pal, 1993). On peut confirmer qu'aucune méthode ne semble prévaloir sur les autres, chacune ayant son domaine de prédilection. En absence de méthode universelle, il est classique de retrouver les différentes approches classifiées en quatre thèmes : clustering, approches contours, approches régions et méthodes hybrides. Rajapakse et al. (2004) a classifié les différentes méthodes en quatre catégories :

- Les méthodes classiques telles que le seuillage, la croissance de région, la segmentation basée sur les contours. La première méthode consiste à déterminer le seuil à appliquer à l'image : le seuillage permet de sélectionner les parties de l'image qui intéressent l'opérateur. La deuxième méthode propose à faire croître chaque région autour d'un pixel de départ dont l'agglomération n'exploite aucune connaissance a priori de l'image ou du bruit et la décision d'intégrer à la région un pixel voisin repose seulement sur un critère d'homogénéité imposé à la zone en croissance. Et la troisième méthode s'intéresse aux contours des objets extraits de l'image. La plupart du temps, ces contours sont morcelés et peu précis, il faut donc utiliser des techniques de reconstruction de contours par interpolation ou connaître a priori la forme de l'objet recherché.
- Les méthodes statistiques telles que la segmentation bayésienne ou la segmentation au sens du maximum de vraisemblance sont basées sur les développements des chaînes de Markov. Dans Bouman et Liu (1998), on peut trouver un algorithme de segmentation non supervisé "Multiple Resolution Segmentation" (MRS), formulé dans le contexte bayésien. La méthode utilise un modèle AR-2D causal (Gaussian Autoregressive) : l'image observée est considérée comme le mélange de champs aléatoires statistiquement homogènes. Le champ des classes est modélisé par un champ aléatoire markovien. La segmentation optimale est définie au sens du Maximum a Posteriori (MAP) et estimée grâce à un algorithme de minimisation locale (Iterated Conditional Mode).
- Les méthodes de réseaux de neurones : Une des stratégies possibles pour la seg-

mentation est celle de la classification de pixels. Quelques exemples de segmentation d'images colorées par réseaux de neurones ont été publiés récemment. Dans Campadelli et al. (1997), les réseaux utilisés étaient des réseaux de Hopfield, configurés à l'aide de l'histogramme des couleurs. Dans Littman et Ritter (1997), les auteurs utilisent les réseaux auto-organisés du type Kohonen pour la segmentation.

- Les méthodes de *clustering* flou : ces techniques permettent d'obtenir une partition floue de l'image en donnant à chaque pixel de l'image un degré d'appartenance à une région donnée.

Un inconvénient du modèle standard de FCM dans la segmentation d'image est de ne pas tenir compte de l'information spatiale qui est une relation entre le pixel et ses voisinages. Pourtant, cette information rend l'algorithme très sensible au bruit et à d'autres objets façonnés dans l'image. En fait, cette relation est une des caractéristiques importantes d'une image car les voisinages possèdent souvent les valeurs semblables, et la probabilité qu'ils appartiennent à la même partition est très élevée. Par ailleurs, si nous considérons une image bruitée, le FCM n'est pas une méthode adaptée pour surmonter ce problème. Récemment, de nombreux chercheurs ont ajouté l'information spatiale à l'algorithme original de FCM pour améliorer l'efficacité de la segmentation d'image (Pham, 2002; Zhang et Chen, 2004; Hung et al., 2006).

Le modèle de la classification dans la segmentation d'image est, en général, un problème de très grande dimension pour lequel, la recherche des méthodes efficaces est toujours d'actualité. Le but de notre travail est double. Premièrement, nous proposons une nouvelle méthode de segmentation d'image via le modèle de FCM basée sur la programmation DC et DCA. Deuxièmement, pour la segmentation d'image bruitée, nous considérons un modèle adaptatif de FCM (appelé FCM-Spatial) qui incorpore l'information spatiale à la fonction de clustering.

La programmation DC et DCA ont été appliquées avec succès à de nombreux problèmes d'optimisation non convexe différentiable ou non de grande dimension dans différents domaines des sciences appliquées, en particulier aux problèmes du data mining (voir par exemple LeThi et al. (2006, 2005); Liu et al. (2003); Neumann et al. (2004); Weber et al. (2005)). Nous étudions dans ce travail, en premier lieu, un schéma DCA appliqué au modèle FCM qui est simple et efficace. Nous utilisons ensuite cet algorithme pour la résolution du nouveau modèle FCM-Spatial qui est de même forme que FCM standard (mais le nombre de variables est double). Pour calculer le bon point initial et accélérer la convergence de DCA, nous proposons une procédure alternative de FCM-DCA qui combine le DCA avec l'algorithme classique de FCM. Les résultats expérimentaux sur plusieurs images bruitées ont illustré l'efficacité de l'algorithme proposé et sa supériorité par rapport à l'algorithme standard de FCM sur le temps de calculs et la qualité des solutions. Par ailleurs, avec le modèle spatial, notre algorithme réduit considérablement l'effet du bruit.

Le papier est organisé de la façon suivante. Dans la deuxième section, nous présentons la formulation du problème FCM. La résolution de ce problème par la programmation DC et DCA est étudiée dans la troisième section. Enfin, la segmentation d'image par la classification floue est présentée dans la dernière section.

## 2 Fomulation du problème FCM

Soit  $X := \{x_1, x_2, \dots, x_n\}$  l'ensemble de  $n$  points à classer. Chaque point  $x_i$  est un vecteur dans l'espace  $\mathbb{R}^p$ . Nous avons à classer ces  $n$  points dans  $c$  ( $2 \leq c \leq n$ ) classes différentes.

Considérons une matrice de pourcentages  $U$  de taille  $(c \times n)$  dont chaque élément  $u_{i,k}$  définit le pourcentage d'appartenance d'un point  $x_k$  à la classe  $C_i$ . Il est clair que

$$u_{i,k} \in [0, 1] \text{ pour } i = 1 \dots c, k = 1 \dots n ; \sum_{i=1}^c u_{i,k} = 1, \text{ pour } k = 1 \dots n. \quad (1)$$

Si la matrice de pourcentages  $U$  est déterminée, on en déduit la classification selon la règle suivante : le point  $x_k$  (pour  $k = 1, \dots, n$ ) est classé dans la classe  $C_i$  (pour  $i = 1, \dots, c$ ) si et seulement si

$$u_{i,k} = \max\{u_{j,k} : j \in \{1, \dots, c\}\}.$$

Considérons la fonction  $J_m$  définie par :

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{i,k}^m \|x_k - v_i\|^2, \quad (2)$$

où  $V$  est une  $(c \times p)$  - matrice dont chaque ligne  $v_i$  correspond au centre de la classe  $C_i$ .  $m \geq 2$  est un paramètre entier qui définit le degré de flou du modèle.

La tâche de chercher une classification revient ainsi à celle de trouver la matrice de pourcentages  $U$  et les centres  $v_i$ . Le modèle mathématique de FCM s'écrit ainsi :

$$\begin{cases} \min J_m(U, V) \\ \text{s.t. } u_{i,k} \in [0, 1] \text{ for } i = 1, \dots, c, k = 1, \dots, n \\ \sum_{i=1}^c u_{i,k} = 1, k = 1, \dots, n. \end{cases} \quad (3)$$

Soient  $x_{k,j}$  la  $j^{\text{ème}}$  composante,  $j = 1, \dots, p$ , du vecteur  $x_k$  et

$$\alpha_j := \min_{k=1, \dots, n} x_{k,j}, \beta_j := \max_{k=1, \dots, n} x_{k,j}.$$

Il est clair que  $v_i \in \Pi_{j=1}^p [\alpha_j, \beta_j]$ .

Pour chaque  $k \in \{1, \dots, n\}$ , soit  $\Delta_k$  le  $(c-1)$ -simplexe dans  $\mathbb{R}^c$  défini par

$$\Delta_k := \left\{ U^{(k)} := (u_{i,k})_i \in \mathbb{R}_+^c : \sum_{i=1}^c u_{i,k} = 1 \right\}.$$

Le problème FCM est reformulé comme :

$$\begin{cases} \min J_m(U, V) \\ \text{s.t. } U \in \Delta := \Pi_{k=1}^n \Delta_k, V \in \mathcal{T} := \Pi_{j=1}^p [\alpha_j, \beta_j] \times \dots \times \Pi_{j=1}^p [\alpha_j, \beta_j] \end{cases} \quad (4)$$

L'algorithme FCM proposé par Bezdek (1981) se décrit comme suit :

---

**Algorithme 1 : FCM**


---

**Initialisation :**

- Choisir le nombre de classes  $c$ .
- Initialiser la matrice des pourcentages  $U$  ainsi que les centres  $v_i$  aléatoirement.

**Répéter**

- Calculer les centres  $v_i$  suivant l'équation :

$$v_i = \sum_{k=1}^n u_{ik}^m x_k / \sum_{k=1}^n u_{ik}^m \quad \forall i = 1, 2, \dots, c.$$

- Calculer la matrice des pourcentages  $U$  :

$$u_{ik} = \left[ \sum_{j=1}^c \frac{\|x_k - v_i\|^{2/(m-1)}}{\|x_k - v_j\|^{2/(m-1)}} \right]^{-1} \quad \forall i = 1, 2, \dots, c \quad \text{et} \quad \forall k = 1, 2, \dots, n.$$

**Jusqu'à** [la modification de solution  $(U, V)$  entre les deux itérations successives est suffisamment petite].

---

On voit que FCM est un problème d'optimisation non convexe dont la résolution sera détaillée dans la suite.

### 3 La programmation DC et DCA pour la résolution de FCM

Pour faciliter la compréhension de notre méthode, nous présentons, en premier lieu de cette section, une brève description de la programmation DC et DCA.

#### 3.1 Introduction à la programmation DC et DCA

La programmation DC joue un rôle central en programmation non convexe (différentiable ou non) car la quasi totalité des problèmes d'optimisation de la vie courante est de nature DC. Elle connaît des développements spectaculaires au cours de cette dernière décennie. DCA est une méthode de descente (de type primal-dual sans recherche linéaire) pour la résolution d'un programme DC de la forme

$$\alpha := \inf \{f(x) := g(x) - h(x) : x \in \mathbb{R}^p\}, \quad (5)$$

où  $g, h$  sont les fonctions convexes semi-continues inférieurement et propres sur  $\mathbb{R}^p$ . Une telle fonction  $f$  est appelée fonction DC, et les fonctions convexes  $g$  et  $h$  les composantes DC de  $f$ . Il est à noter que la minimisation d'une fonction DC sur un ensemble convexe fermé  $C$  de  $\mathbb{R}^p$  se ramène à un problème de type (5) car la contrainte  $x \in C$  peut être incorporée dans la fonction objective à l'aide de la fonction indicatrice

## Segmentation d'images par la Classification floue basée sur DCA

$\chi_C$  définie par  $\chi_C = 0$  si  $x \in C$ ,  $+\infty$  sinon. Lorsqu'une de ses composantes DC est polyédrale la fonction  $f$  est dite DC polyédrale et le programme DC correspondant DC polyédral. La programmation DC polyédrale joue un rôle crucial en programmation non convexe.

La conjugaison d'une fonction convexe  $g$ , notée  $g^*$  est définie par

$$g^*(y) := \sup\{\langle x, y \rangle - g(x) : x \in \mathbb{R}^p\}.$$

La dualité DC est définie via la conjugaison des composantes DC et le programme dual de (5) est donné par (ici l'espace dual de  $\mathbb{R}^p$  est identifié à lui-même) :

$$\alpha_D := \inf\{h^*(y) - g^*(y) : y \in \mathbb{R}^p\}. \quad (6)$$

Puisque chaque fonction  $h \in \Gamma_0(\mathbb{R}^p)$  est caractérisée comme le supremum d'une famille finie des fonctions affines, c.à.d.

$$h(x) := \sup\{\langle x, y \rangle - h^*(y) : y \in \mathbb{R}^p\},$$

on a

$$\alpha = \inf\{g(x) - \sup\{\langle x, y \rangle - h^*(y) : y \in \mathbb{R}^p\} : x \in \mathbb{R}^p\} = \inf\{\alpha(y) : y \in \mathbb{R}^p\},$$

où

$$\alpha(y) := \inf\{g(x) - [\langle x, y \rangle - h^*(y)] : x \in \mathbb{R}^p\} \quad (P_y).$$

Il est clair que  $(P_y)$  est un programme convexe et

$$\alpha(y) = h^*(y) - g^*(y) \text{ si } y \in \text{dom } h^*, \text{ et } +\infty \text{ sinon.} \quad (7)$$

Par suite

$$\alpha = \inf\{h^*(y) - g^*(y) : y \in \text{dom } h^*\}.$$

Finalement on obtient, avec la convention naturelle  $+\infty - (+\infty) = +\infty$  :

$$\alpha = \alpha_D := \inf\{h^*(y) - g^*(y) : y \in \mathbb{R}^p\}.$$

On observe ainsi la symétrie parfaite entre les programmes DC primal et dual : le dual de (6) est exactement (5).

Le transport des solutions optimales globales entre l'ensemble des solutions optimales  $\mathcal{P}$  de (5) et celui de (6) noté  $\mathcal{D}$  s'exprime de la manière suivante (LeThi et PhamDinh, 2005; PhamDinh et LeThi, 1997) :

$$\cup\{\partial h(x^*) : x^* \in \mathcal{P}\} \subset \mathcal{D} \text{ et } \cup\{\partial g^*(y^*) : y^* \in \mathcal{D}\} \subset \mathcal{P}. \quad (8)$$

La relation (8) indique que la résolution d'un programme DC implique celle de son dual. D'autre part, ce transport reste valable entre les ensembles des solutions locales de (5) et (6) sous certaines hypothèses techniques.

En analyse convexe,

$$\partial h(x^0) := \{y \in \mathbb{R}^p : h(x) \geq h(x^0) + \langle x - x^0, y \rangle, \forall x \in \mathbb{R}^p\}$$

est appelé le sous-différentiel de  $h$  au point  $x^0$ . Tout élément de  $\partial h(x^0)$  est appelé gradient de  $h$  en  $x^0$ . Le sous-différentiel  $\partial h(x^0)$  est une partie convexe fermée qui coïncide avec le gradient  $\nabla h(x^0)$  si et seulement si  $h$  est différentiable en  $x^0$ . Pour un  $\epsilon > 0$ , le  $\epsilon$ -sous-différentiel de  $h$  est défini par

$$\partial_\epsilon h(x^0) := \{y \in \mathbb{R}^p : h(x) \geq h(x^0) + \langle x - x^0, y \rangle - \epsilon, \forall x \in \mathbb{R}^p\}.$$

L'égalité des valeurs optimales des programmes primal et dual (5) et (6) peut être traduite de manière équivalente par

$$\mathcal{P} = \{x^* : \partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*), \forall \epsilon > 0\}.$$

Mais sauf des cas très rares, cette condition d'optimalité globale est impraticable. Nous nous intéressons dès lors aux conditions d'optimalité locale pour les programmes DC (voir LeThi (1997); LeThi et PhamDinh (2005); PhamDinh et LeThi (1997, 1998) et références incluses) :

$$\partial h(x^*) \subset \partial g(x^*), \quad (9)$$

et

$$\partial h(x^*) \cap \partial g(x^*) \neq \emptyset. \quad (10)$$

(Un tel point  $x^*$  vérifiant (10) est appelé *point critique* de  $g - h$ ).

La condition nécessaire d'optimalité locale (9) est également suffisante dans plusieurs cas rencontrés en pratique - par exemple, quand la fonction objectif  $f := g - h$  est DC polyédrale avec  $h$  polyédrale, ou quand  $f$  est localement convexe en  $x^*$ .

Basé sur les conditions d'optimalité locale et la dualité DC, DCA consiste en la construction de deux suites  $\{x^k\}$  et  $\{y^k\}$ , candidates respectives aux solutions des problèmes primal et dual que l'on améliore à chaque itération (les deux suites  $\{g(x^k) - h(x^k)\}$  et  $\{h^*(y^k) - g^*(y^k)\}$  sont décroissantes) et qui convergent vers des solutions primale et duale  $x^*$  et  $y^*$  vérifiant des conditions d'optimalité locale. Le schéma général de DCA prend la forme :

$$y^k \in \partial h(x^k); \quad x^{k+1} \in \partial g^*(y^k). \quad (11)$$

La première interprétation de DCA est simple : à chaque itération on remplace dans le programme DC primal la deuxième composante DC  $h$  par sa minorante affine  $h_k(x) := h(x^k) + \langle x - x^k, y^k \rangle$  au voisinage de  $x^k$  pour obtenir le programme convexe suivant

$$\inf \{g(x) - h_k(x) : x \in \mathbb{R}^p\} \quad (12)$$

dont l'ensemble des solutions optimales n'est autre que  $\partial g^*(y^k)$ .

De manière analogue, la deuxième composante DC  $g^*$  du programme DC dual (6) est remplacée par sa minorante affine  $(g^*)_k(y) := g^*(y^k) + \langle y - y^k, x^{k+1} \rangle$  au voisinage de  $y^k$  pour donner naissance au programme convexe

$$\inf \{h^*(y) - (g^*)_k(y) : y \in \mathbb{R}^p\} \quad (13)$$

dont  $\partial h(x^{k+1})$  est l'ensemble des solutions optimales. DCA opère ainsi une double linéarisation à l'aide des sous-gradients de  $h$  et  $g^*$ . Il est à noter que DCA travaille

avec les composantes DC  $g$  et  $h$  et non pas avec la fonction  $f$  elle-même. Chaque décomposition DC de  $f$  donne naissance à un DCA. Pour un programme DC donné, la question de décomposition DC optimale reste ouverte, en pratique on cherche des décompositions DC bien adaptées à la structure spécifique du programme DC étudié pour lesquelles les suites  $\{x^k\}$  et  $\{y^k\}$  sont faciles à calculer, (si possible) explicites pour que les DCA correspondants soient moins coûteux en temps et par conséquent capables de supporter de très grandes dimensions.

Soient  $C$  (resp.  $D$ ) l'ensemble convexe qui contient la suite  $\{x^k\}$  (resp.  $\{y^k\}$ ) et  $\rho(g, C)$  (ou  $\rho(g)$  si  $C = \mathbb{R}^p$ ) défini par

$$\rho(g, C) = \sup \left\{ \rho \geq 0 : g - \frac{\rho}{2} \|\cdot\|^2 \text{ soit convexe sur } C \right\}.$$

DCA est une méthode de descente sans recherche linéaire, qui possède les propriétés suivantes :

i) Les suites  $\{g(x^k) - h(x^k)\}$  et  $\{h^*(y^k) - g^*(y^k)\}$  sont décroissantes et

- $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$  ssi  $y^k \in \partial g(x^k) \cap \partial h(x^k)$ ,  $y^k \in \partial g(x^{k+1}) \cap \partial h(x^{k+1})$  et  $[\rho(g, C) + \rho(h, C)] \|x^{k+1} - x^k\| = 0$ . De plus, si  $g$  ou  $h$  est strictement convexe sur  $C$  alors  $x^k = x^{k+1}$ .

Dans ce cas DCA se termine à l'itération  $k$  (convergence finie de DCA).

- $h^*(y^{k+1}) - g^*(y^{k+1}) = h^*(y^k) - g^*(y^k)$  ssi  $x^{k+1} \in \partial g^*(y^k) \cap \partial h^*(y^k)$ ,  $x^{k+1} \in \partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1})$  et  $[\rho(g^*, D) + \rho(h^*, D)] \|y^{k+1} - y^k\| = 0$ . De plus, si  $g^*$  ou  $h^*$  est strictement convexe sur  $D$  alors  $y^{k+1} = y^k$ .

Dans ce cas DCA se termine à l'itération  $k$  (convergence finie de DCA).

ii) Si  $\rho(g, C) + \rho(h, C) > 0$  (resp.  $\rho(g^*, D) + \rho(h^*, D) > 0$ ) alors la série  $\{\|x^{k+1} - x^k\|^2\}$  (resp.  $\{\|y^{k+1} - y^k\|^2\}$ ) converge.

iii) Si la valeur optimale  $\alpha$  du problème (5) est finie et deux suites  $\{x^k\}$  et  $\{y^k\}$  sont bornées alors toute valeur d'adhérence  $\tilde{x}$  (resp.  $\tilde{y}$ ) de la suite  $\{x^k\}$  (resp.  $\{y^k\}$ ) est le point critique de  $g - h$  (resp.  $h^* - g^*$ ).

iv) DCA a la convergence linéaire pour les programmes DC généraux.

v) DCA a la convergence finie pour les programmes DC polyédraux.

Pour une étude complète de la programmation DC et DCA, se référer à LeThi (1997); LeThi et PhamDinh (2005); PhamDinh et LeThi (1997, 1998) et références incluses. Il est à noter que la recherche d'une décomposition DC adéquate et celle d'un bon point initial sont deux tâches importantes dans la résolution d'un programme non convexe par DCA car elles conditionnent la réussite du résultant DCA.

### 3.2 Résolution du problème FCM

Dans toute la suite nous utilisons la présentation matricielle qui nous semble plus commode, sachant que l'on peut identifier une matrice et un vecteur (par ligne ou par colonne).

Nous cherchons tout d'abord une décomposition DC de la fonction objectif de (4). En appliquant la formule :

$$2f_1f_2 = (f_1 + f_2)^2 - (f_1^2 + f_2^2)$$



nous obtenons :

$$\begin{aligned}
 J_m(U, V) &= \sum_{k=1}^n \sum_{i=1}^c u_{i,k}^m \|x_k - v_i\|^2 \\
 &= \frac{1}{2} \sum_{k=1}^n \sum_{i=1}^c (u_{i,k}^m + \|x_k - v_i\|^2)^2 - \frac{1}{2} ((u_{i,k}^{2m} + \|x_k - v_i\|^4)) \\
 &= G(U, V) - H(U, V),
 \end{aligned}$$

avec

$$\begin{aligned}
 G(U, V) &= \frac{1}{2} \sum_{k=1}^n \sum_{i=1}^c (u_{i,k}^m + \|x_k - v_i\|^2)^2 + \chi_K(U, V); \\
 H(U, V) &= \frac{1}{2} \sum_{k=1}^n \sum_{i=1}^c ((u_{i,k}^{2m} + \|x_k - v_i\|^4).
 \end{aligned}$$

En tenant compte du fait que la fonction  $\theta(x) = f(x)^p$  est convexe pour  $p \geq 1$  si  $f$  est convexe non négative, on démontre facilement que  $G(U, V)$  et  $H(U, V)$  sont convexes, par suite  $J_m$  est une fonction DC.

La résolution de FCM par DCA revient aux calculs de sous-différentiels de  $H$  et de  $G^*$ .

**Calcul de  $(Y^l, Z^l) \in \partial H(U^l, V^l)$**

En utilisant les règles de base de calcul de sous-gradient, nous avons :

$$\begin{aligned}
 \partial H(U, V) &= (\partial_U H(U, V), \partial_V H(U, V)) \\
 &= (m \cdot \text{Exp}(2m - 1), 2 \sum_{k=1}^n (\|x_k - v_i\|^2 (v_i - x_k))_{i=1..c})
 \end{aligned} \tag{14}$$

$(\text{Exp}(2m - 1))$  représente la matrice dont chaque élément est  $(\text{Exp}(2m - 1))_{i,k} := u_{i,k}^{2m-1}$ ,  $i = 1, \dots, c$ ,  $k = 1, \dots, n$ .

**Calcul de  $(U^{l+1}, V^{l+1}) \in \partial G^*(Y^l, Z^l)$**

Rappelons que  $(U^{l+1}, V^{l+1}) \in \partial G^*(Y^l, Z^l)$  si et seulement si  $(U^{l+1}, V^{l+1})$  est une solution du problème convexe suivant :

$$\min \{ G(U, V) - \langle (U, V), (Y^l, Z^l) \rangle : (U, V) \in \Delta \times \mathcal{T} \} \tag{15}$$

qui est solvable par n'importe quel algorithme pour la programmation convexe. Dans l'implémentation de notre algorithme nous utilisons la méthode de Gradient Projeté (Polyak, 1987), vu que la projection d'un point sur un simplexe ou sur un rectangle est explicitement déterminée.

La méthode de Gradient Projeté pour la résolution de (15) se décrit comme suit :

### Algorithme GRP (pour résoudre (15))

---

#### Algorithme 2 : GRP

---

1. Soient  $r = 1$ ,  $(U^r, V^r) := (U^l, V^l)$ , une suite  $\{\lambda_r\}$  telle que :  
 $\lim_{r \rightarrow +\infty} \lambda_r = 0$ ,  $\sum_{r=1}^{+\infty} \lambda_r = +\infty$ .
2. Calculer  $(U^{r+1}, V^{r+1})$  de la manière suivante (Proj denote la projection) :  
la  $k^{ième}$  colonne de la matrice  $U^{r+1}$  est

$$U^{r+1(k)} := \text{Proj}_{\Delta_k} \left( U^{r(k)} - \lambda_r \frac{\theta^r}{\|\theta^r\|} \right),$$

où  $\theta^r$  est la  $k^{ième}$  colonne du  $\nabla_U G(U^r, V^r) - Y^l$ , et la  $i^{ième}$  ligne de la matrice  $V^{r+1}$  est ( $T := \Pi_{j=1}^p [\alpha_j, \beta_j]$ )

$$V_i^{r+1} := \text{Proj}_T \left( V_i^r - \lambda_r \frac{[\nabla_V G(U^r, V^r) - Z^l]_i}{\|[\nabla_V G(U^r, V^r) - Z^l]_i\|} \right).$$

3. Si  $\|(U^{r+1}, V^{r+1}) - (U^r, V^r)\| \leq \epsilon$  alors arrêter,  
sinon remplacer  $r$  par  $r + 1$  et aller à l'étape 2.
- 

Finalement nous pouvons décrire le schéma DCA pour la résolution de (4) comme suivant :

---

#### Algorithme 3 : Fuzzy-DCA

---

**Initialisation :**

- Choisir  $U^0 \in \mathbb{R}^{c,n}$  et  $V^0 \in \mathbb{R}^{c,p}$ ;
- Choisir une tolérance  $\epsilon > 0$ .

**Répéter**  $l = 0, 1, 2, \dots$

- Calculer  $(Y^l, Z^l) \in \partial H(U^l, V^l)$  à l'aide de (14).
- Calculer  $(U^{l+1}, V^{l+1}) \in \partial G^*(Y^l, Z^l)$  en résolvant le problème (14) par l'Algorithme GRP.

**Jusqu'à**  $\|(U^{l+1}, V^{l+1}) - (U^l, V^l)\| \leq \epsilon$ .

---

### 3.3 La recherche d'un bon point initial pour DCA par une procédure alternative FCM-DCA

La recherche d'un bon point initial joue un rôle crucial dans la résolution d'une programmation DC par DCA. Elle dépend de la structure du problème considéré et elle peut être effectuée par, par exemple, une méthode heuristique bien adaptée au problème. D'une manière générale, un bon point initial pour le DCA ne doit pas être un minimum local, parce qu'à partir d'un tel point, le DCA est stationnaire. En plus, nous observons qu'à partir de n'importe quel point n'étant pas minimum local, la fonction objective diminue rapidement durant quelques premières itérations

de DCA. Nous avons la même remarque pour l'algorithme standard FCM. C'est pourquoi, nous proposons une procédure alternative de FCM-DCA pour le problème (4).

---

**Algorithme 4 : FCM-DCA**


---

*Initialisation* : Initialise la matrice de partition  $U^o$  ainsi que les centres  $V^o$  aléatoirement. Soit *maxiter* un nombre entier positif.

*Répéter*

- Une itération de FCM : Calculer les centres  $V^l = (v_1^l, v_2^l, \dots, v_c^l)$  par la formule

$$v_i^l = \sum_{k=1}^n u_{ik}^m x_k / \sum_{k=1}^n u_{ik}^m \quad \forall i = 1, \dots, c.$$

Calculer  $U^l$  par

$$u_{ik}^l = \left[ \sum_{j=1}^c \frac{\|x_k - v_i\|^{2/(m-1)}}{\|x_k - v_j\|^{2/(m-1)}} \right]^{-1}.$$

- Une itération de DCA : exécuter une itération de **Fuzzy-DCA** avec le point initial  $(U^l, Z^l)$  pour obtenir  $(U^{l+1}, V^{l+1})$ .
  - $l \leftarrow l + 1$
  - Jusqu'à  $l = \text{maxiter}$ .
- 

Si nous utilisons l'algorithme combiné de FCM-DCA jusqu'à sa convergence, peut-être l'efficacité de DCA n'est pas bien exploitée. Pour remédier à cette situation, nous proposons un algorithme en deux phases. Dans la première phase, nous exécutons quelques itérations de l'algorithme combiné de FCM-DCA pour trouver un bon point initial. Et dans la deuxième, à partir du point trouvé, nous appliquons DCA jusqu'à sa convergence. Comme on verra dans la suite, parmi différentes versions de DCA, cet algorithme en deux phases est la meilleure option.

## 4 Segmentation des images par la classification floue via DCA

Nous présentons dans cette section une application de notre algorithme Fuzzy-DCA à la segmentation d'image.

### 4.1 Modèle de FCM avec l'information spatiale

Dans la segmentation d'image par le modèle standard de FCM, chaque pixel  $x_k \in \mathbb{R}^p$  représente les données multispectrales. Cependant, comme mentionné précédemment, une des caractéristiques importantes d'une image est que les voisinages du pixel possèdent les valeurs semblables, le rapport spatial est donc intéressant pour la segmentation d'image. L'information spatiale est la relation entre le pixel et ses voi-

sinages. Il y a différentes manières d'incorporer l'information spatiale (voir la figure 1).

Dans notre cadre, nous considérons l'information spatiale de  $x_k$  comme une valeur moyenne de ses voisinages  $3 \times 3$ , et chaque point  $x_k$  dans (4) a deux groupes de valeurs : les valeurs du pixel et les valeurs moyennes de ses voisinages  $3 \times 3$ .

Soit  $N_k$  les voisinages  $3 \times 3$  du pixel  $x_k$ . Les données entrées  $x_k$  dans notre modèle spatial de FCM sont  $x_k = (x_{k1}, x_{k2})$  où  $x_{k1}$  représente les valeurs du pixel de  $k^{th}$  de l'image et  $x_{k2} = (x_{k1} + \sum_{i \in N_k} x_{i1})/9$ . D'où, le nombre de variables  $U$  n'est pas changé et  $V$  devient une matrice de  $c \times 2p$  dont la  $i^{eme}$  ligne est,  $v_i \in \mathbb{R}^{2p}$ , le centre de la région  $C_i$ .

Le modèle spatial de FCM dans notre approche n'est pas tellement différent par rapport à son modèle standard FCM (4) sauf le fait que chaque  $x_k \in \mathbb{R}^d$  est remplacé par  $x_k = (x_{k1}, x_{k2}) \in \mathbb{R}^{2p}$ . Par conséquent, n'importe quel algorithme pour le modèle standard de FCM (4) peut être appliqué au modèle spatial de FCM. Au point de vue numérique, le problème spatial de FCM est plus difficile car le nombre de variables  $V$  a doublé.

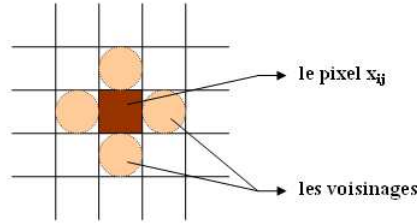


FIG. 1 – Le pixel et ses 4 voisinages

## 4.2 Expériences numériques

Nous testons nos algorithmes sur deux types d'images : une image originale où les contours et les régions sont parfaitement localisés, et une image bruitée avec un bruit gaussien. Tous les tests ont été réalisés sur un ordinateur de Pentium[R] 4 CPU 3.00GHz 1.00Go RAM.

Dans la première expérimentation nous comparons la performance de l'algorithme de FCM (Bezdek, 1981) (**Algorithme 1**) et nos deux algorithmes : **Algorithme 3** avec la procédure de recherche d'un point initial par **Algorithme 4** (noté **Algorithme 4-3**) et **Algorithme 4** (procédure alternative FCM-DCA) sur les modèles sans ou avec l'information spatiale. Les résultats sont reportés dans les figures 2.

Dans la deuxième expérimentation nous comparons la performance de **FCM (Algorithme 1)** et les deux variants de DCA : **Algorithme 3** sans la procédure de recherche d'un point initial et **Algorithme 4-3**. Nous utilisons les mêmes paramètres initiaux pour tous les algorithmes. Les résultats sont présentés dans le tableau 1. Nous utilisons les notations suivantes :

- *Taille* : la taille de l'image.
- *c* : le nombre de régions de l'image.
- *iter* : le nombre d'itérations de l'algorithme.
- *Temps* : le temps de calcul de l'algorithme en secondes.

### 4.3 Commentaires

À partir des résultats expérimentaux, nous constatons que :

- Dans plusieurs images, notre algorithme donne une segmentation presque parfaite. En plus, sans information spatiale, nos **Algorithme 4** et **Algorithme 3** peuvent surmonter la segmentation d'image bruitée dans certains cas.
- Avec l'information spatiale, **Algorithme 3** fonctionne bien sur toutes les images bruitées. Il peut supprimer les bruits de manière efficace. Ainsi les deux algorithmes DCA apportent l'image de meilleure qualité par rapport à l'algorithme FCM.
- Dans la plupart des cas, les deux variantes de DCA sont plus rapides que FCM (**Algorithme 1**). Par ailleurs **Algorithme 4-3** permet d'avoir le gain de calcul le plus important.

## 5 Conclusion

Dans ce travail, nous proposons une nouvelle méthode de segmentation d'images via le modèle de FCM en utilisant un nouveau et robuste algorithme basé sur la programmation DC et DCA. Le modèle de FCM a été reformulé comme une programmation DC sur laquelle un schéma simple et rapide de DCA a été appliqué. La procédure alternative de FCM-DCA est efficace pour trouver un bon point initial de DCA et pour accélérer sa convergence. L'algorithme en deux phases de DCA peut alors être appliqué dans le problème de la classification de grande dimension. D'autre part, l'utilisation d'information spatiale pour la segmentation d'images bruitées semble être efficace. Les expériences numériques préliminaires prouvent que les algorithmes proposés sont prometteurs pour la segmentation d'images bruitées.

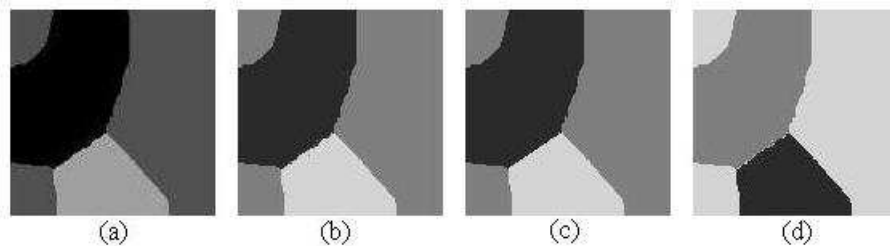


FIG. 2 – L'image originale et les résultats de segmentation ( $c=3$ ).

(a) : L'image originale (b) : Le résultat de **Algorithme 1** (c) : Le résultat de **Algorithme 4** (d) : Le résultat de **Algorithme 4-3**.

## Segmentation d'images par la Classification floue basée sur DCA

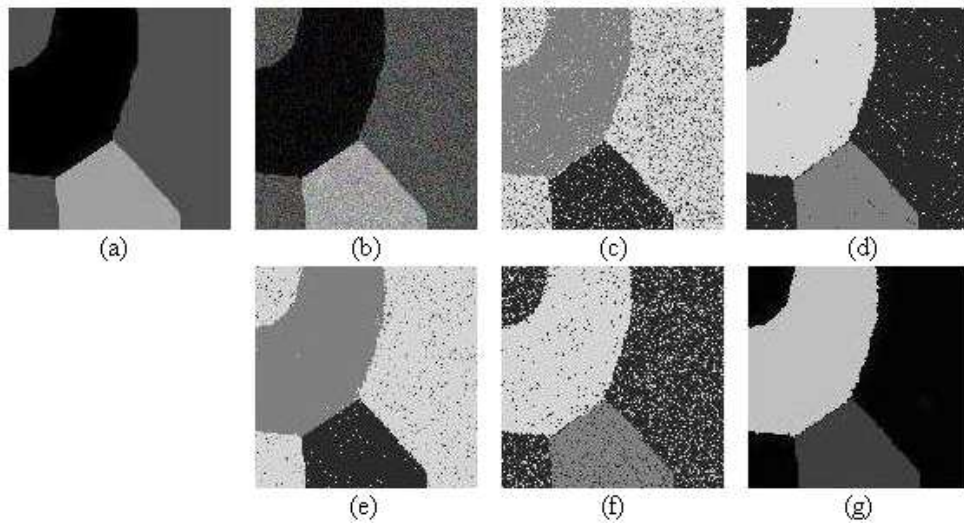


FIG. 3 – L'image originale avec le bruit et les résultats de segmentation ( $c=3$ ).  
 (a) : L'image originale (b) : L'image originale avec le bruit Gaussien (c) : Le résultat de **Algorithme 1** (d) : Le résultat de **Algorithme 1** avec l'information spatiale. (e) : Le résultat de **Algorithme 4** (f) : Le résultat de **Algorithme 4-3** (g) : Le résultat de **Algorithme 4-3** avec l'information spatiale.

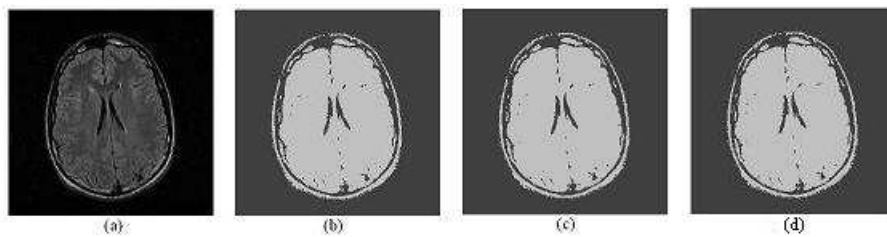


FIG. 4 – L'image médicale originale et les résultats de segmentation ( $c=2$ ).  
 (a) : L'image originale (b) : Le résultat de **Algorithme 1** (c) : Le résultat de **Algorithme 4** (d) : Le résultat de **Algorithme 4-3**.

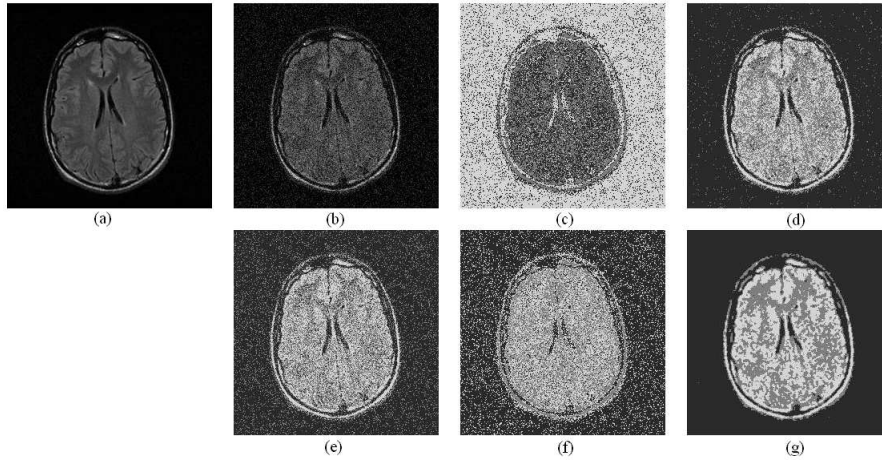


FIG. 5 – *L'image médicale avec le bruit Gaussien et les résultats de segmentation ( $c=3$ ).*  
 (a) : L'image originale (b) : L'image originale avec le bruit Gaussien (c) : Le résultat de **Algorithme 1** (d) : Le résultat de **Algorithme 1** avec l'information spatiale. (e) : Le résultat de **Algorithme 4** (f) : Le résultat de **Algorithme 4-3** (g) : Le résultat de **Algorithme 4-3** avec l'information spatiale.

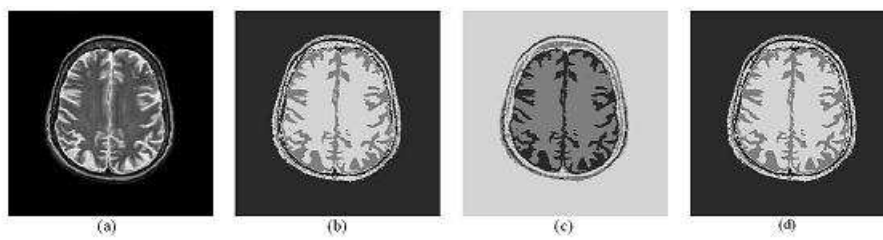


FIG. 6 – *L'image médical originale et les résultats de segmentation ( $c=3$ ).*  
 (a) : L'image originale (b) : Le résultat de **Algorithme 1** (c) : Le résultat de **Algorithme 4** (d) : Le résultat de **Algorithme 4-3**.



## Segmentation d'images par la Classification floue basée sur DCA

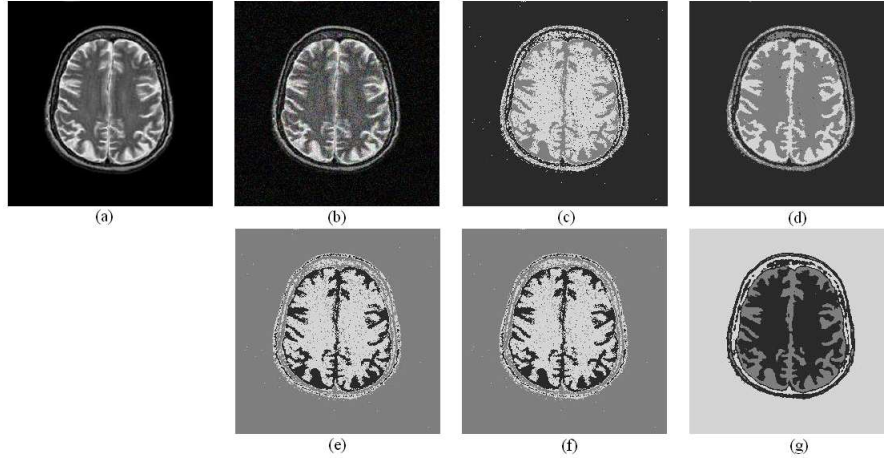


FIG. 7 – L'image médicale avec le bruit Gaussien et les résultats de segmentation ( $c=3$ ).  
 (a) : L'image originale (b) : L'image originale avec le bruit Gaussien (c) : Le résultat de **Algorithme 1** (d) : Le résultat de **Algorithme 1** avec l'information spatiale. (e) : Le résultat de **Algorithme 4** (f) : Le résultat de **Algorithme 4-3** (g) : Le résultat de **Algorithme 4-3** avec l'information spatiale.

TAB. 1 – Résultats comparatifs de **Algorithme 1**, **Algorithme 3**, et **Algorithme 4-3**.

Données		Algorithme 1 (FCM)			Algorithme 3		Algorithme 4-3	
$N^{\circ}$	Taille	$c$	$iter$	Temps	$iter$	Temps	$iter$	Temps
1	$128^2$	2	24	1.453	16	1.312	10	1.219
2	$128^2$	2	17	1.003	12	0.985	2	0.765
3	$256^2$	3	36	15.340	24	13.297	2	10.176
4	$256^2$	3	75	31.281	57	30.843	12	26.915
5	$256^2$	3	39	15.750	27	14.687	14	13.125
6	$256^2$	5	91	84.969	75	86.969	78	61.500
7	$256^2$	3	73	31.094	62	34.286	21	24.188
8	$256^2$	3	78	34.512	52	32.162	13	29.182
9	$512^2$	3	49	92.076	41	102.589	46	74.586
10	$512^2$	5	246	915.095	196	897.043	86	691.854

## Références

- Bezdek (1981). Pattern recognition with fuzzy objective function algorithm. *New York, NY. Plenum Press 1.*
- Bouman et Liu (1998). Segmentation of textured images using a multiple resolution



- approach. *Proc. IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, New York, NY, April 11-14, 1124–1127.
- Campadelli, Medici, et Schettini (1997). Color image segmentation using hopfield networks. *Image and Vision Computing*, Vol.15, N° 3, 161–166.
- Hung, Yang, et Chen (2006). Parameter selection for suppressed fuzzy c-means with an application to mri segmentation. *Pattern Recognition Letters*, Vol.27, 424–438.
- LeThi, H. A. (1997). Contribution à l'optimisation non convexe et l'optimisation globale: Théorie, algorithmes et applications. *Habilitation à Diriger des Recherches, Université de Rouen*.
- LeThi, H. A., T. Belghiti, et T. PhamDinh (2006). A new efficient algorithm based on DC programming and DCA for clustering. *In Press, Available July 2006, Journal of Global Optimization*.
- LeThi, H. A., M. LeHoai, et T. PhamDinh (2005). Optimization based DC programming and DCA for hierarchical clustering. *In Press, Available online June 2006, European Journal of Operational Research*.
- LeThi, H. A. et T. PhamDinh (2005). The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research* 133, 23–46.
- Littman et Ritter (1997). Adaptative color segmentation - a comparison of neural and statistical methods. *IEEE Transactions on Neural Networks*, Vol.8, N°irc.1, 175–185.
- Liu, Y., X. SHEN, et Hani (2003). Multicategory  $\psi$ -learning and support vector machine: Computational tools. *Journal of Computational and Graphical Statistics* 14, 219–236.
- Neumann, J., C. Schnörr, et G. Steidl (2004). Svm-based feature selection by direct objective minimisation. *Pattern Recognition, Proc. of 26th DAGM Symposium* 3175, 212 – 219.
- Pal, J. P. et S. K. Pal (1993). A review on image segmentation techniques. *Pattern Recognition*, Vol.26, 1277–1294.
- Pham (2002). Fuzzy clustering with spatial constraints. *Proc. IEEE Intern. Conf. on Image Processing, New York, USA*.
- PhamDinh, T. et H. A. LeThi (1997). Convex analysis approach to DC programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica, dedicated to Professor Hoang Tuy on the occasion of his 70th birthday* 22, 289–355.
- PhamDinh, T. et H. A. LeThi (1998). DC optimization algorithms for solving the trust region subproblem. *SIAM J. Optimization* 8, 476–505.
- Polyak, T. (1987). Introduction to optimization. *Inc., Publications Division*.
- Rajapakse, Giedd, et Rapoport (2004). Statistical approach to segmentation of single-channel cerebral mr images. *IEEE Trans. On Medical Imaging* 16.
- Weber, S., T. Schüle, et C. Schnörr. (2005). Prior learning and convex-concave regularization of binary tomography. *Electr. Notes in Discr. Math* 20, 313–327.

Segmentation d'images par la Classification floue basée sur DCA

Zhang et Chen (2004). A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, Vol.32, 37–50.

## Summary

We present a fast and robust algorithm for image segmentation problems via Fuzzy C-Means (FCM) clustering model. Our approach is based on DC (Difference of Convex functions) programming and DCA (DC Algorithms) that have been successfully applied in a lot of various fields of Applied Sciences, including Machine Learning. In an elegant way, the FCM model is reformulated as a DC program for which a very simple DCA scheme is investigated. For accelerating the DCA, an alternative FCM-DCA procedure is developed. Moreover, in the case of noisy images, we propose a new model that incorporates spatial information into the membership function for clustering. Experimental results on noisy images have illustrated the effectiveness of the proposed algorithm and its superiority with respect to the standard FCM algorithm in both running-time and quality of solutions.