



HAL
open science

Segmentation et détection d'objets par caractérisation multi-échelle

Enguerran Grandchamp, Philippe Marthon

► **To cite this version:**

Enguerran Grandchamp, Philippe Marthon. Segmentation et détection d'objets par caractérisation multi-échelle. 14ème Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA 2004), AFRIF; AFIA, Jan 2004, Toulouse, France. pp.1–10. hal-00602265

HAL Id: hal-00602265

<https://hal.science/hal-00602265v1>

Submitted on 5 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Segmentation et détection d'objets par caractérisation multi-échelle

Driving segmentation and recognition phases using multiscale characterization

E. Grandchamp¹

P. Marthon²

¹GRIMAAG UAG
Campus de Fouillole, 97157 Pointe-à-Pitre
enguerran.grandchamp@univ-ag.fr

²LIMA ENSEEIHT
2 rue Charles Camichel 31071 Toulouse
philippe.marthon@enseeiht.fr

Résumé

A mi chemin entre la segmentation et la reconnaissance de formes, cette étude propose de caractériser les points de rupture d'une image afin de guider ces deux phases.

La caractérisation est obtenue par une généralisation des exposants de Lipschitz dans le cas de structures complexes (hétérogènes, recouvrement, ...). On parle alors de chaînes de maxima. Cette notion intègre l'évolution de la valeur et l'évolution spatiale d'un maximum.

Ces chaînes de maxima sont obtenues par un suivi des maxima du module de l'image du gradient à différents niveaux de résolution. La décomposition multi-résolution est obtenue à partir d'une transformée en ondelettes de l'image.

Une classification des points de rupture est réalisée sur la base de ces chaînes de maxima.

Cette caractérisation est utilisée : lors de la phase de segmentation pour détecter des points contours d'un même objet et construire à partir de ces échantillons un contour fermé ; lors de la phase de reconnaissance pour identifier la nature des objets en fonction de la nature des transitions.

Mots Clef

ondelettes, segmentation, classification, reconnaissance de formes, exposants de Lipschitz, traitement d'images

Abstract

Half away between edge detection and object recognition this study aim to characterize singularities in order to guide this two phases.

The characterization is obtained with a generalization of the Lipschitz exponents for complex structures. This new characterization is called « maxima chain » and

integrates the notion of value and spatial repartition of the maxima.

The maxima chains are obtained with a multiscale decomposition using a Diadic Discrete Wavelet Transform (DDWT). We use the maxima of the gradient image computed with the wavelet details. The local maxima detect the position of irregular structures (edges) in the image (Canny algorithm).

Then we characterize each maxima using their evolution across the different scales.

A classification is made using the maxima chains.

We integrate the classification to eliminate unwanted maxima (from noise or other objects) to drive the search of closed contours. In fact, there is a totally different signature from one maxima to another depending on the parameters of the objects : nature, shape, size, gray scale, texture, ...

Keywords

wavelet transform, segmentation, pattern recognition, singularity

1. Introduction

Cette étude propose une extension des exposants de Lipschitz pour des signaux non réguliers. En effet, la plus part des objets présents dans une scène réelle ne peuvent pas être modélisés par un unique signal (un dirac, une sinusoïde, un échelon, ...) mais une combinaison de structures plus complexes. Nous proposons d'adapter dans ce contexte la caractérisation des signaux par les exposants de Lipschitz [1] afin de classer les points de rupture de l'image et ainsi séparer la scène en objets.

Cet article est organisé de la manière suivante. La section 2 présente l'analyse multi-échelle. La section 3 présente le principe de construction des chaînes de maxima. La section 4 présente l'extension des coefficients de Lipschitz (chaînes de maxima généralisées). La classification supervisée et non supervisée des chaînes de maxima est présentée en section 5. Cette classification est la base de l'aide à la segmentation et à la reconnaissance d'objets. Les sections 6 et 7 traitent de l'intégration des chaînes dans ces deux phases. La section 8 présente les principaux résultats d'une application. Les travaux en cours et les perspectives sont présentées dans la section 9. L'article présenté ici est essentiellement orienté vers l'étape de classification et son utilisation dans les post-traitements. Peu de détails seront donc donnés ici concernant les autres étapes.

2. Analyse multi-échelle

Nous allons présenter ici un résumé du principe de multi-résolution et de la décomposition en ondelettes, plus de précisions sont disponibles en [5] et [6]. La multi-résolution est obtenue en utilisant une transformation en ondelettes.

La transformation en ondelette est obtenue par projection orthogonal de l'image sur des espaces affines obtenus à partir de la base d'ondelette [6]. On parle d'espace d'approximation (V) et d'espace de détail (W). L'espace d'approximation de niveau i contient des signaux plus grossiers que l'espace d'approximation de niveau $i+1$. Plus i augmente, plus on s'intéresse aux basse fréquences du signal. L'espace de détail de niveau i est la différence d'information entre l'espace d'approximation de niveau $i-1$ et l'espace d'approximation de niveau i (il sert donc à stocker des hautes fréquences).

Les valeurs des images d'approximation et de détail sont calculées en utilisant un paramètre de dilatation/réduction σ et un paramètre de translation τ . Physiquement, σ permet d'analyser le signal à différentes fréquences et τ de parcourir le signal. $W\{f,\psi\}(\sigma,\tau)$ est la valeur du coefficient d'ondelette obtenue par application de l'ondelette ψ au signal f avec comme coefficient de dilatation σ et comme coefficient de translation τ .

$$(1) \quad W\{f,\psi\}(\sigma,\tau) = \langle f(t), \psi_{\sigma,\tau}(t) \rangle$$

avec $\psi_{\sigma,\tau}(t) = \frac{1}{\sqrt{|\sigma|}} \psi\left(\frac{t-\tau}{\sigma}\right)$ où $\psi(t)$ est appelée

l'ondelette mère.

L'ondelette utilisée ici est bi-orthogonale. Les filtres utilisés pour la décomposition et la reconstruction du signal à partir de la transformation en ondelettes ne sont donc pas identiques (contrairement aux ondelettes orthogonales de type Haar, Meyer ou Daubechies [6]).

Parmi les ondelettes bi-orthogonales, on trouve les fonctions *splines*.

Par ailleurs, nous utilisons une transformation en ondelettes discrète dyadique. Ceci signifie que le coefficient de dilatation utilisé est 2. Le coefficient de dilatation correspondant au niveau i est donc $\sigma_i = 2^i$. Ce sont donc les images d'approximations obtenues à partir de ces coefficients de dilatation qui vont composer la multi-résolution.

L'image I (équivalente à l'approximation de niveau 0 V_0) est donc décomposée en n niveaux d'approximation ($\{V_i\}$, $1 \leq i \leq n$) et de détails ($\{W^x_i, W^y_i\}$, $1 \leq i \leq n$) :

$$(2) \quad I = V_1 + W^x_1 + W^y_1$$

$$(3) \quad V_i = W^x_{i+1} + W^y_{i+1} + V_{i+1}, \quad 0 \leq i \leq n-1$$

La valeur de n dépend de la taille de l'image. On considère par la suite des images I de 2^n lignes et 2^n colonnes.

$$(4) \quad I = \{ (x_{ij}, y_{ij}) \mid i \in [0, 2^n - 1], j \in [0, 2^n - 1] \}$$

3. Construction des chaînes de maxima

Après la décomposition multi-échelle, on dispose d'un ensemble d'images d'approximation et d'images de détails.

A partir de cette décomposition on calcule le module des images de détails pour chacun des niveaux

$$(5) \quad (|W^x_{i+1}|^2 + |W^y_{i+1}|^2)^{1/2}, \quad 0 \leq i \leq n-1$$

On extrait ensuite les maxima locaux de ces modules. Un maximum étant un pixel de valeur supérieure aux deux pixels adjacents (précédente et suivante) dans la direction donnée par $\arctan(W^x / W^y)$. On note W^m_i l'ensemble des maxima locaux au niveau i .

On note $W^m = \{ W^m_i \mid 0 \leq i \leq n-1 \}$

Le module des coefficients de détails est en relation directe avec le gradient de l'image. Les maxima locaux correspondent donc aux zones de forte transition de l'image.

On dispose maintenant d'un ensemble de maxima à différents niveaux de résolution et l'objectif est, partant d'un maximum appartenant au premier niveau de résolution (niveau le plus fin), de trouver ses différents correspondants à travers les autres niveaux.

Le suivi des maxima est basé sur le fait que les maxima du gradient de l'image originale se retrouvent à travers tous les niveaux de la décomposition.

Soit $M_0(x_0, y_0)$ le représentant d'un maximum sur l'image de niveau 0 de la décomposition multi-échelle.

On recherche le successeur $M_1(x_1, y_1)$ de M_0 dans le niveau 1 de la décomposition.

La liste des maxima ainsi obtenue en parcourant tous les niveaux est appelée chaîne de maxima.

La recherche est limitée [1] au niveau 1 au voisinage défini par un cône d'influence centré aux coordonnées (x_0, y_0) .

L'expression du cône d'influence est fonction de l'échelle et de l'ondelette utilisée. Pour l'ondelette utilisée ici le cône d'influence a l'expression suivante :

$$(6) \quad |x_0 - x| < \frac{1}{2} \sigma \Delta x \text{ et } |y_0 - y| < \frac{1}{2} \sigma \Delta y$$

σ étant le coefficient de dilatation (facteur d'échelle).

La recherche du successeur sera effectuée sur plusieurs critères. En effet, plusieurs configurations peuvent se présenter. Si un seul maximum est présent, le choix est facilité. Si plusieurs maxima se trouvent dans le cône d'influence, une ambiguïté subsiste quant au choix du bon candidat. Inversement, si aucun maximum n'est présent, la chaîne s'arrête au niveau correspondant.

Ces différents comportements, conjugués avec des intersections non nulles des différents cônes d'influences, conduit à des comportements hétérogènes des chaînes de maxima (fusion, maintien, disparition). Cette étape nécessite une part plus importante d'heuristiques basées sur divers critères.

Un des critères les plus utilisés est de prendre comme successeur le maximum le plus proche du centre du cône. Mais ce critère ne permet de faire le choix le plus judicieux que si il n'y a pas d'autres maxima qui interagissent de manière significative avec le maximum courant.

A ce premier critère on peut ajouter le fait que la valeur du maximum ne doit pas changer de manière trop significative [1]. Ce critère donne de bons résultats pour des maxima proches géographiquement mais n'ayant pas des valeurs du même ordre de grandeur. En effet, même si il y a une interaction entre ces maxima, leur valeur permet parfaitement de les distinguer.

Dans le cas le plus général, les maxima interagissent entre eux et ont des valeurs voisines (surtout si ils appartiennent au même contour). Il faut donc faire intervenir les deux critères afin de tenir compte de ces différentes contraintes. On peut par exemple prendre le problème comme un problème de minimisation de contrainte avec comme fonction à minimiser la fonction suivante :

$$(7) \quad \alpha \left(\frac{d}{d_{\max}} \right)^2 + \beta \left(\frac{V}{V_{\max}} \right)^2$$

d_{\max} étant la plus grande distance entre deux maxima du cône d'influence et V_{\max} la plus grande différence de valeur entre deux maxima du cône d'influence.

La difficulté est maintenant d'estimer les paramètres α et β . En effet, on ne peut pas calculer leur valeur théorique et il faut donc la fixer empiriquement.

Sans ces deux paramètres on serait confronté à différents problèmes :

si $[(x, y), V]$ sont les coordonnées et la valeur d'un maximum M de niveau k , la résolution de cette équation peut conduire à deux solutions : $S_1 = [(x, y), V']$ et $S_2 = [(x', y'), V]$ tel que : $V' = V_{\max}$ et distance de S_2 à M égale à d_{\max} .

Ainsi on aurait le choix entre : une solution de même coordonnée que le maximum mais de niveau de gris éloigné et une solution éloigné mais de même niveau de gris.

Avec les coefficients α et β , on peut faire varier l'importance que l'on accorde à ces différents informations. D'autre part, pour pouvoir faire un choix plus judicieux, il faudrait faire intervenir le comportement des maxima voisins.

La valeur de ces deux paramètres va dépendre des images traitées et du type de chaîne que l'on souhaite faire ressortir.

Pour illustrer les différents comportements possibles d'un chaîne de maxima, la Figure 1. présente les cas de fusion (de 3 et 5 chaînes), de disparition (au niveau 3 après fusion) et de maintien (jusqu'au niveau 4). On s'intéresse ici uniquement à l'évolution des chaînes la valeur des maxima n'ayant que peu d'importance.

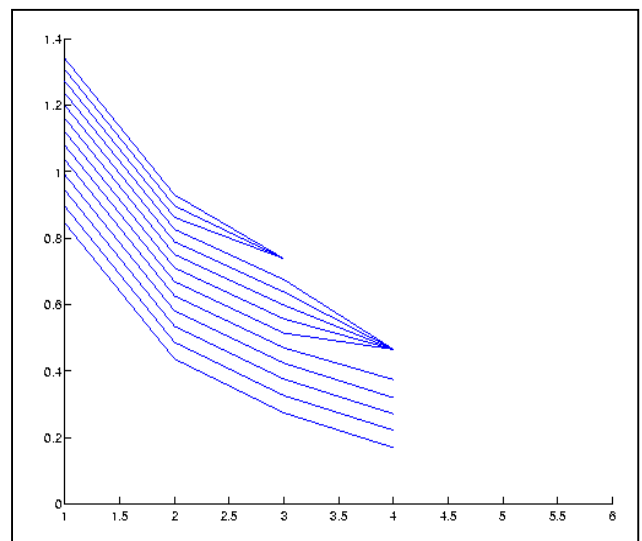


Figure 1. Chaîne de maxima : valeur des maxima (ordonnée) / niveau de résolution (abscisse)

4. Extension des coefficients de Lipschitz

Habituellement, on mesure la régularité d'un signal à l'aide des coefficients de Lipschitz. Cette caractérisation permet d'affecter une valeur, le coefficient de Lipschitz ou exposant de Hölder selon la littérature, à une rupture du signal.

Théoriquement, et pour un signal pur, la représentation graphique de la chaîne de maxima permet d'estimer le coefficient de Lipschitz. L'enchaînement des maxima doit en effet approcher une fonction affine dont le coefficient directeur est l'exposant de Lipschitz. La chaîne de maxima permet donc de calculer numériquement le coefficient de Lipschitz. Par exemple, la Figure 2. présente la chaîne de maxima obtenue pour un point de rupture d'un dirac (coefficient de Lipschitz de -1 à gauche) et d'un échelon (coefficient de Lipschitz de 0 à droite).

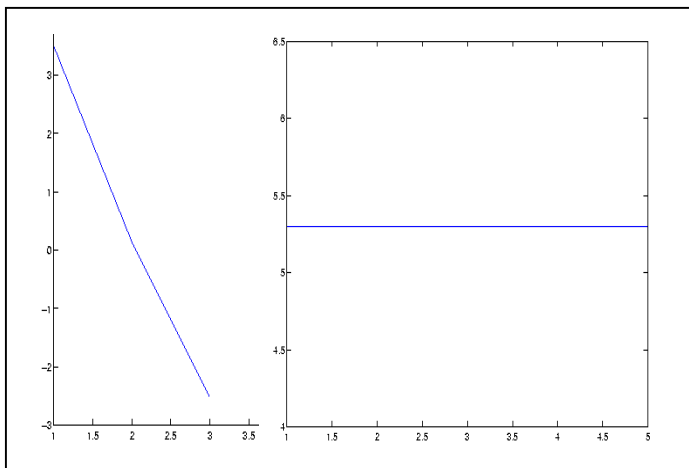


Figure 2. Chaînes de maxima classiques (valeur des maxima en ordonnée et niveau de résolution en abscisse) : a) Dirac, b) échelon

Nous étendons la notion d'exposant de Lipschitz au cas de signaux perturbés par la proximité d'autres signaux ou d'autres maxima appartenant au même signal. C'est le cas le plus fréquent dans une scène réelle composée de plusieurs objets ou d'un arrière plan non uniforme.

Dans ce cas, la chaîne de maxima n'est plus nécessairement une fonction affine et l'évaluation du coefficient de Lipschitz s'en trouve compromise.

La Figure 3. illustre le cas de chaînes de maxima non affines. On ne peut donc pas approcher ces courbes par des droites et donc extraire un coefficient de Lipschitz. Nous proposons donc d'étendre la caractérisation d'une rupture au cas de transitions complexes faisant intervenir plusieurs signaux.

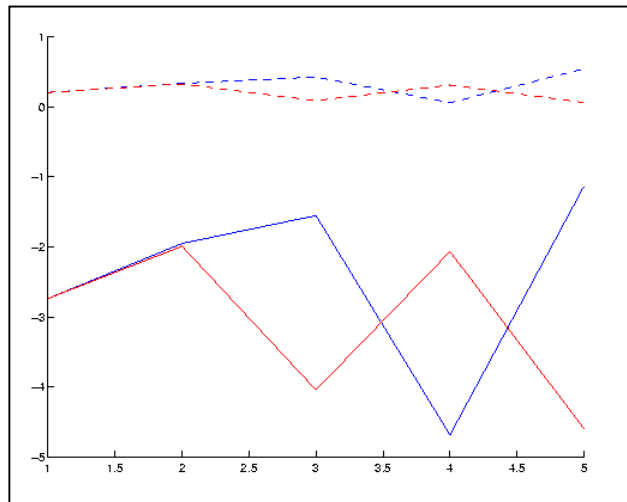


Figure 3. Chaînes de maxima non standard (valeur des maxima en ordonnée, niveau de résolution en abscisse)

L'extension du coefficient de Lipschitz se traduit par la prise en compte de l'évolution complète du maximum à travers les niveaux. La signature d'une transition n'est plus réduite au coefficient directeur de la fonction affine approchant la chaîne mais

- à la forme générale de la courbe (tableau de variation de la fonction),
- à l'interaction avec les autres chaînes (fusion, arrêt maintien, absorption, ...),
- aux valeurs prises par le maximum (valeur minimale, maximale, moyenne, ...),
- aux informations géographiques sur le maximum considéré (position initiale et chemin parcouru sur l'image au travers des niveaux).

Une analyse plus fine de la chaîne de maxima est donc réalisée.

La construction des chaînes de maxima est également révisée afin de tenir compte des autres chaînes présentes dans le voisinage d'un maximum.

En effet, cette étape est nécessaire lors du choix d'un successeur parmi plusieurs candidats. Une algorithmique d'optimisation (non présentée dans le cadre de cet article) est nécessaire pour affecter les candidats aux différentes chaînes.

Les techniques d'optimisation utilisées font appel aux méta-heuristiques (Algorithme Génétique [15], Recherche Tabou [16][18][19], Recuit Simulé [20], Recherche dispersée [17], ...).

Par le choix des heuristiques, il est possible de favoriser l'apparition de chaînes de maxima fusionnant avec d'autres ou au contraire de les éviter.

Nous adopterons les notations suivantes dans la suite du document :

- Chaîne de maxima associée au maximum m de niveau 0 ($m \in D^m_0$)
 (8) $C^m = (m_0, m_1, \dots, m_n)$
 avec $m_i = (v^m_i, x^m_i, y^m_i) \in D^m_i$ si il existe un représentant de m au niveau i (v^m_i : valeur du maximum au niveau i , (x^m_i, y^m_i) : coordonnées du maximum au niveau i).
 ou $m_i = \emptyset$ sinon
- Ensemble des chaînes de maxima
 (9) $S = \{ C^m \mid m \in D^m_0 \}$
- Ensemble des classes : C
- Maxima échantillons d'une classe c : m_c

5. Classification des chaînes de maxima

Après la phase de construction des chaînes de maxima, nous disposons d'un ensemble S hétérogène qu'il faut classer afin d'extraire et localiser les structures d'intérêt.

On utilise toutes les caractéristiques des chaînes de maxima décrites dans la section précédente (forme, interaction, valeur, localisation). Nous décrivons ici uniquement les grands principes de la classification sans détailler les techniques sous jacentes mises en place.

Actuellement, la classification fonctionne en mode supervisé. Un mode non supervisé est à l'étude mais n'a pas encore été mis en œuvre de manière concluante et n'est donc pas présenté dans le cadre de cet article.

Classification supervisée

Dans le cadre de la classification supervisée, les classes que l'on souhaite obtenir peuvent être représentées par une ou plusieurs des caractéristiques générales d'une chaîne de maxima présentée dans la section précédente. On distingue ainsi trois types de classifications des chaînes de maxima.

La classification par comportement

Lors d'une classification par comportement, trois classes sont définies :

- **La classe de maintien (CM)** : l'ensemble des chaînes regroupées dans cette classe ont un représentant à chacun des niveaux de résolution. Elles ne rentrent pas en interaction directe (fusion ou absorption) avec d'autres chaînes. Cette classe est utilisée pour regrouper des ruptures fortes et/ou isolées de l'image.

(10)

$$CM = \{ C^m = (m_0, m_1, \dots, m_n) \mid m \in D^m_0, \forall i \in [0, n] m_i \neq \emptyset \text{ et } \forall C^p = (p_0, \dots, p_n) \in S \setminus C^m, p_i \neq m_i \}$$

- **La classe d'arrêt (CA)** : paramétrée par le niveau d'arrêt k , elle regroupe l'ensemble des

chaînes de maxima s'arrêtant au niveau k . Cette classe regroupe les ruptures faibles et isolées de l'image.

(11)

$$CA = \{ C^m = (m_0, m_1, \dots, m_n) \mid m \in D^m_0, \exists i \in [0, n], \forall k \in [i, n] m_k = \emptyset \}$$

- **La classe de fusion (CF)** : cette classe est la plus complexe, elle regroupe les chaînes de maxima qui vont interagir à un ou plusieurs niveaux de résolution. De manière générale, lorsque deux chaînes de maxima fusionnent à un niveau k , elles restent confondues jusqu'au dernier niveau n . Ceci est lié à la méthode de construction basique des chaînes de maxima qui ne prend en compte que le maximum courant pour la recherche du successeur. Néanmoins, les techniques d'optimisation citées dans la section 4 permettent de prendre en compte la totalité des maxima et ainsi de définir des schémas d'évolution des chaînes qui permettent à deux chaînes de se séparer après une fusion. Enfin, il faut noter que toutes les chaînes qui fusionnent ne sont pas nécessairement identiques, il peut y avoir fusion homogène ou hétérogène. Dans le cas d'une fusion hétérogène, il y a en fait absorption d'une rupture de l'image par une autre rupture de plus grande importance. En tenant compte de toutes ces considérations, on peut dériver la classe fusion en plusieurs classes

(12)

$$CF = \{ C^m = (m_0, m_1, \dots, m_n) \mid m \in D^m_0, \exists i \in [0, n], \exists C^p = (p_0, \dots, p_n) \in S \setminus C^m, p_i = m_i \}$$

Si par ailleurs : $\forall k \in [i, n] p_k = m_k$, il y a fusion totale (absorption), sinon il y a fusion partielle.

Ce type de classification est le plus complexe. Il faut noter que les classes CA et CF ne sont pas disjointes. Lors de la classification, une priorité peut être donnée à aux classes pour l'affectation des chaînes ou alors il y peut y avoir duplication de celles-ci. Le choix dépend de l'application. Par exemple si l'on recherche en priorité des ruptures faibles, on classera une chaînes fusionnant et s'arrêtant à un certain niveau dans la classe CA plutôt que dans la classe CF .

$$(13) \quad CA \cap CF \neq \emptyset$$

La classification par localisation (CL)

Lors de la classification par localisation, on regroupe les chaînes de maxima qui évoluent dans un même voisinage. Le nombre de classes de ce type dépend de la taille des voisinages choisis. Pour un voisinage restreint (4 ou 8 connexité), la classification par localisation consiste en un regroupement des ruptures par continuité.

Les classes obtenues regroupent donc des chaînes de ruptures localisées dans une même zone de l'image et donc appartenant potentiellement au même objet. Soit $V(x,y)$ un voisinage autour du point (x,y) .

$$(14) \quad CL(V(x,y)) = \{ C^m = (m_0, m_1, \dots, m_n) \mid m \in D^m_0, \forall i \in [0,n] m_i = (v_i, x_i, y_i), (x_i, y_i) \in V(x,y) \}$$

Si (x,y) parcourt tous les points de l'images I (équation (4)) alors les classes de localisation ne sont pas nécessairement disjointes et ce quel que soit le voisinage V choisi (différent du singleton (x,y)).

En effet,

$$(15) \quad \begin{aligned} \forall V(x,y) \neq \{(x,y)\}, \forall (x1,y1) \neq (x,y) \in V(x,y), \\ V(x1,y1) \cap V(x,y) \neq \emptyset \\ V(x1,y1) \cap V(x,y) \supset \{(x,y), (x1,y1)\} \end{aligned}$$

On peut donc trouver des chaînes de maxima appartenant à plusieurs classes. L'obtention d'une partition de l'image à partir des classes n'est pas une nécessité et les chaînes peuvent être présentes dans plusieurs classes.

La classification par valeur (CV)

Dans ce dernier mode de classification, les regroupements sont faits en tenant compte de la valeur des maxima aux différents niveaux de résolution. Le critère d'affectation est donc une fonction sommant les erreurs commises aux différents niveaux par rapport à un échantillon référent. On cherche à minimiser cette fonction critère.

Soit $C^{mec} = (me_{c0}, me_{c1}, \dots, me_{cn})$ la chaîne de maxima de l'échantillon de la classe c .

Soit $C^{m1} = (m1_0, m1_1, \dots, m1_n)$ une première chaîne à classer.

La fonction critère à minimiser est alors

$$(16) \quad F(mec) = \sum_{i=0}^n |v_{mec}^i - v_{m1}^i|$$

pour c parcourant l'ensemble des classes.

Toutefois, l'affectation d'une chaîne à la classe permettant de minimiser F n'est pas systématique. En effet, $F(mec)$ doit être également inférieur à un seuil fixé.

Le problème d'optimisation est donc un problème de minimisation avec contrainte défini comme suit :

$$(17) \quad \left\{ \begin{aligned} \text{Min } (F(mec) = \sum_{i=0}^n |v_{mec}^i - v_{m1}^i|) \\ c \in C \\ F(mec) < S \end{aligned} \right.$$

Les différents modes de classification présentés dans cette première partie de section peuvent être utilisés indépendamment les uns des autres mais les meilleurs résultats sont obtenus en intégrant les trois modes. L'intégration peut être simultanée ou en cascade.

En simultanée, une fonction d'agrégation pondérée est définie afin de prendre en compte l'ensemble des critères.

La recherche des pondérations est souvent délicate et dépend souvent des images à traiter. Nous avons donc décidé d'écartier ce mode de classification et de mettre en œuvre l'intégration en cascade. Dans ce dernier mode, on choisit un ordre d'enchaînement des filtres de classification. On assiste dans ce cas à une multiplication des classes afin de respecter les caractéristiques de chacun des modes (une classe d'arrêt, de fusion et de maintien par voisinage par exemple) mais également à l'homogénéisation des classes ce qui améliore les post-traitements (segmentation et reconnaissance de formes).

La Figure 4. donne le résultat d'une classification obtenue dans un premier temps par comportement : dans ce cas toutes les chaînes de la Figure 4. font partie de la même classe (même tableau de variation). On applique ensuite une classification par valeur qui permet de séparer cette classe en deux sous classes (partie inférieure et partie supérieure de la classe).

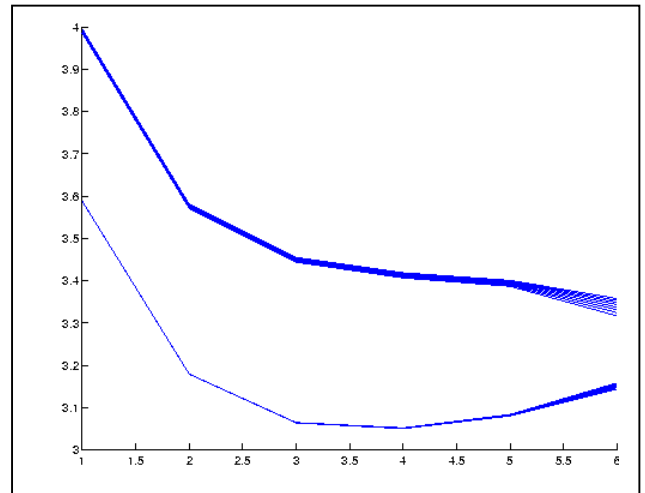


Figure 4. Classification par comportement puis par valeur

On peut ainsi regrouper des chaînes de maxima ayant des valeurs et des comportements proches, dans des zones bien localisées. Ceci permet de préparer les phases de segmentation et de reconnaissance de forme présentées dans la suite du document en proposant des classes contenant des points de rupture de même nature.

Dans le cas le plus général de la classification supervisée, on dispose d'un ou plusieurs échantillons de chaque classe et nous appliquons une technique de classification par apprentissage [21][22].

L'espace de représentation des chaînes de maxima est alors un espace à trois dimensions : les coordonnées x et y sur les deux premiers axes et la valeur du maximum sur le troisième. Une chaîne de maxima est donc un polygone ouvert ayant un maximum de n points dans un espace de dimension 3.

La classification consiste donc à définir des surfaces séparatrices (hyperplans) entre les classes. Les classes n'étant pas nécessairement disjointes, on doit résoudre un problème d'optimisation classique par des techniques de descente de gradient (perceptron, relaxation, ... [21][22]). D'autres techniques d'optimisation (méthodes heuristiques) sont également à l'étude afin d'accélérer les temps de calcul.

6. Phase de segmentation

Dans cette section, on considère qu'un maximum de l'image du module des coefficients de détail est un point de forte transition dans l'image.

Pendant la phase de segmentation, nous devons extraire uniquement les transitions correspondant à un contour d'objet (écarter le bruit, ..). Par ailleurs, nous souhaitons obtenir des contours fermés pour ces mêmes objets.

Pour guider cette phase de segmentation, nous allons utiliser la classification précédente afin de rejeter les transitions non désirées.

Nous allons donc intégrer tous les aspects de la classification : la classification par localisation nous permettra de satisfaire la continuité des points contours. La classification par comportement et valeur nous permettra de satisfaire l'appartenance des points contours à un même type de transition (donc potentiellement à un même objet).

La difficulté majeure provient des ruptures de l'objet qui interagissent avec d'autres objets. Ces ruptures appartiennent au contour de l'objet mais peuvent être rejetées au niveau de la classification en raison de la modification de leur évolution au travers des niveaux (fusion avec d'autres chaînes, ...). Elles vont conduire à l'apparition de discontinuités dans les contours des objets au niveau des zones d'interaction.

Nous avons fait le choix de fixer des seuils de tolérance très faible pour l'affectation des chaînes de maxima aux différentes classes. Ceci, couplé aux nombreuses interactions présentes dans une scène réelle, a pour conséquence de fortement réduire le nombre de ruptures regroupées dans la même classe. Beaucoup de chaînes de maxima ne seront pas classées et seront donc rejetées par le classifieur.

Complexité de l'image (densité des objets)	Maxima correctement classés	maxima mal classés	maxima rejetés
Simple	74,8 %	0,1 %	25,1 %
Moyenne	27,2 %	0,3 %	72,5 %
Complexe	5 %	0,3 %	94,7 %

Tableau 1. Résultats de la classification avec un seuil de tolérance faible

Le Tableau 1 présente les résultats moyens obtenus sur des images de synthèse (3 images pour chaque catégorie). On constate que le pourcentage de chaînes correctement classées est faible mais également que le pourcentage de chaînes mal classées est quasiment nul. La classification nous permet donc d'obtenir des maxima dont l'affectation n'est pas à remettre en cause.

Ceci n'est plus le cas si on augmente le seuil de tolérance lors de la classification (Tableau 2).

Complexité de l'image (densité des objets)	Maxima correctement classés	maxima mal classés	maxima rejetés
simple	82,4 %	15,8 %	1,8 %
moyenne	37,6 %	45,9 %	17,5 %
complexe	25,1 %	55,2 %	19,7 %

Tableau 2. Résultat de la classification avec un seuil de tolérance élevé

L'extraction de maxima de référence est un point important sur lequel nous allons nous appuyer dans la suite du processus de segmentation.

L'algorithme de segmentation que nous avons mis en place est un algorithme en deux passes. Dans un premier temps, la classification précédente nous permet d'obtenir une segmentation labellisée partielle de l'image (un label par classe). Le type de transition que l'on souhaite extraire de l'image conduit à la définition des classes correspondantes (par la définition des échantillons). Ensuite, les chaînes de maxima correspondantes sont extraites et donc également les transitions de l'image. On obtient donc des contours des objets avec des discontinuités (ruptures non classées de l'image).

Nous partons du principe que le contour de tous les objets est fermé. A partir de ces maxima isolés, nous allons donc essayer de reconstruire les contours fermés des objets en intégrant progressivement les chaînes de maxima précédemment rejetées. L'Algorithme 1 présente l'approche retenue pour cette phase.

Un algorithme d'optimisation combinatoire est mis en œuvre afin de sélectionner les maxima adéquats parmi les chaînes non classées. C'est une approche locale qui a été retenue de manière à construire le contour à partir des maxima déjà existant. Par ailleurs, lors de cette progression locale il arrive que certains choix soient à remettre en cause. Il faut donc mettre en œuvre une technique de retour permettant de corriger les erreurs et d'explorer d'autres maxima.

En tenant compte de ces éléments, nous avons choisis la recherche tabou pour rechercher le contour fermé optimal. Le critère de sélection des maxima intègre la distance entre celui-ci et les maxima de référence obtenus

après classification (afin de forcer la fermeture des contours en passant par les maxima de référence) ainsi que les caractéristiques des chaînes (comportement, valeur, ...).

A chaque réinitialisation de la recherche (dépassement du nombre maximal d'itérations), le choix du maximum de départ ainsi que des maxima intégrés dans le contour dépend des précédentes recherches (on ne réitère pas les erreurs commises). C'est au travers de la notion de mémoire (court et long terme, fréquentielle, qualitative, ...) de la recherche Tabou que les informations nécessaires sont stockées.

```

Nmax : Nombre maximal d'itérations
Classification : en cascade des chaînes de maxima par
comportement, valeur puis localisation
    S = ensemble des chaînes de maxima
    C = ensemble des classes
    Srej = ensemble des chaînes rejetées
// Les chaînes d'une classe sont considérées comme des
représentants d'un même objet
V(m) = voisinage du maxima m
Pour chaque c ∈ C // Construction du contour fermé
    Tantque Contour ouvert
        NbIt = 0
        max = un maxima de niveau 0 choisi parmi les
                chaînes de c
    Tantque Contour ouvert et (NbIt < Nmax)
        Chercher Cm = (m0, m1, ..., mn) tel
        que : Cm ∈ Srej, m0 ∈ V(max)
        Si Cm existe alors
            max ← m0
        Sinon
            Elargir le voisinage V(max)
        Fsi
        NbIt ← NbIt + 1
    Ftantque
Ftantque
Fpour
    
```

Algorithme 1. Segmentation

7. Phase de reconnaissance de forme

Au terme de la phase précédente, on obtient un ensemble de classes (C) telles que les représentants d'une classe forment un contour fermé.

Le contour fermé (C_f) est construit de la manière suivante :

$$(18) \quad \forall c \in C, C_f(c) = \{m_{i0}\}, C^{mi} = (m_{i0}, m_{i1}, \dots, m_{in}) \in c$$

L'objectif de cette étape est de caractériser les contours afin de reconnaître les objets qu'ils délimitent.

Une base de connaissance est utilisée pour faire cette reconnaissance.

La base regroupe des informations sur la forme du contour et sur la nature des transitions qui composent

l'objet. Elle est construite à partir d'objets isolés (donc sans interaction avec d'autres objets).

La phase de reconnaissance consiste actuellement à comparer les contours obtenus en fin de phase précédente avec les modèles de la base.

Un contour est reconnu comme étant un objet si la distance entre celui-ci et le modèle est inférieure à un seuil fixé.

Actuellement la distance utilisée est une distance euclidienne classique basée sur une comparaison des enveloppes convexes du modèle et de la forme à reconnaître.

D'autres distances sont envisageables afin notamment de reconnaître des objets ayant subi des déformations (rotation, dilatation/réduction, ...).

8. Application et résultats

Les techniques décrites dans cet article ont été appliquées à des images optiques et des images radars.

La présence de bruit multiplicatif (le speckle) sur les images radars engendre beaucoup de maxima dans l'image du gradient et masque partiellement les maxima engendrés par les objets de la scène.

Néanmoins, on constate qu'un certain nombre de ces maxima disparaît dès les premiers niveaux de résolution.

On peut donc appliquer un premier filtre sur les chaînes de maxima de manière à rejeter les chaînes n'ayant pas de représentant au delà d'un certain niveau ou fusionnant avec d'autres chaînes.

Des adaptations sont donc à envisager pour une application concluante aux images radars.

Nous illustrerons donc les différentes techniques décrites dans cet article avec des images optiques et plus précisément des images de synthèse dont nous connaissons exactement la composition.

La Figure 5. représente l'image de synthèse utilisée.

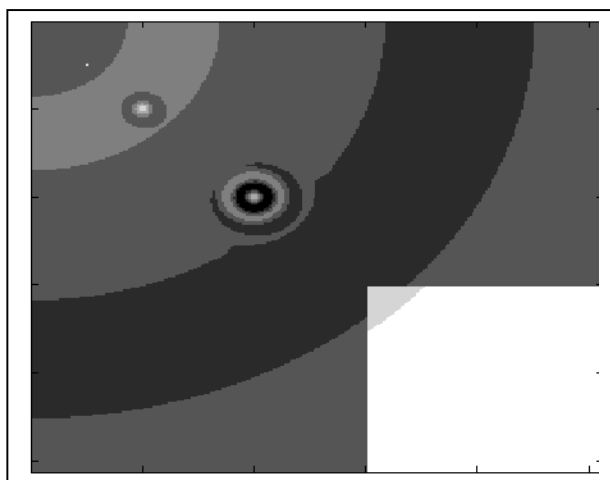


Figure 5. Image de synthèse

L'image est construite à partir de l'expression d'un signal 2D comportant les éléments suivants :

- Un Dirac : $f_1(x, y) = \delta_{x_1, y_1}(x, y)$

- Un chapeau mexicain :

$$f_2(x, y) = \alpha e^{-\left[\frac{(x-x_2)^2}{\varepsilon} + \frac{(y-y_2)^2}{\varepsilon}\right]} * (-1 + \sigma(x-x_2)^2 + \sigma(y-y_2)^2)$$

- Une sinusoïde amortie :

$$f_3(x, y) = \beta e^{-\left[\frac{(x-x_3)^2}{\tau} + \frac{(y-y_3)^2}{\tau}\right]} * \cos(w_1 * ((x-x_5)^2 + (y-y_5)^2)^{1/2})$$

- Un échelon :

$$f_4(x, y) = \begin{cases} x > x_4 \wedge y > y_4 \rightarrow f_4(x, y) = \mu \\ \text{sin on} \rightarrow f_4(x, y) = 0 \end{cases}$$

- Une sinusoïde basse fréquence :

$$f_5(x, y) = \sin(w_2 * (x^2 + y^2)^{1/2})$$

L'expression globale du signal 2D est :

$$f(x, y) = f_1(x, y) + f_2(x, y) + f_3(x, y) + f_4(x, y) + f_5(x, y)$$

Dans cette application nous nous intéressons aux structures les plus importantes de l'image. Les classes définies vont donc conduire à rejeter les classes de fusion et d'arrêt.

Par ailleurs, la phase de reconnaissance de forme vise à reconnaître les signaux f_3 et f_4 plutôt que des objets réels.

La Figure 6. présente le résultat de quatre classes obtenu après une classification par comportement et valeur avec un seuil de tolérance moyen. En haut à gauche (C1) on retrouve le dirac (f_1), en haut à droite (C2) le chapeau mexicain (f_2), en bas à gauche (C3) la sinusoïde amortie (f_3) et en bas à droite (C4) l'échelon (f_4).

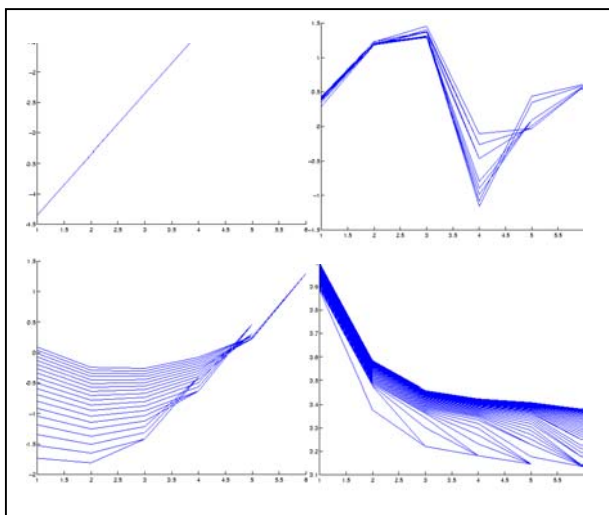


Figure 6. Classification

L'augmentation volontaire du seuil de tolérance génère la dérive constatée sur les classes par rapport à la classification présentée à la Figure 4. Dans notre cas d'application, ceci est sans incidence mais un seuil de tolérance plus faible est préférable dans la plus part des cas.

La Figure 7. représente les maxima de niveau 0 obtenus après classification.

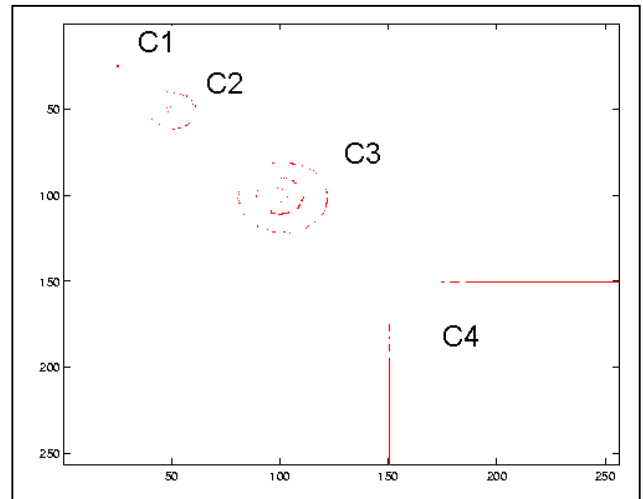


Figure 7. Maxima de niveau 0 après classification

On constate que les classes correspondant au dirac (C1) et à l'échelon (C4) correspondent bien aux signaux initiaux. Par contre, des difficultés apparaissent pour le chapeau mexicain (C2) et la sinusoïde amortie (C3).

La classification n'a fait intervenir que le comportement et la valeur des maxima. Par ailleurs, on constate que les maxima regroupés dans une même classe correspondent bien au même type de signal. Les maxima de la classe C2 sont bien des représentants du signal f_2 , et les maxima de la classe C3 sont bien des représentants du signal f_3 . Mais ces différents représentants appartiennent à des lobes différents des signaux.

Une séparation des classes C2 et C3 est donc nécessaire pour fermer correctement les contours lors de la phase de segmentation. Cette séparation s'effectue par localisation avec un seuil faible pour l'affectation des chaînes de maxima aux différentes classes. Cette classification est non supervisée et ne concerne que les classes C2 et C3. Elle permet de décomposer la classe C2 en C2.1 et C2.2 et la classe C3 en C3.1, C3.2 et C3.3 (Figure 8.)

Une autre approche aurait été de définir directement une classe par lobe des signaux f_2 et f_3 mais cette approche aurait nécessité l'intervention d'un opérateur pour définir les échantillons.

Après la phase de segmentation (intégration progressive des chaînes précédemment rejetées), on obtient la classification suivante (Figure 8.).

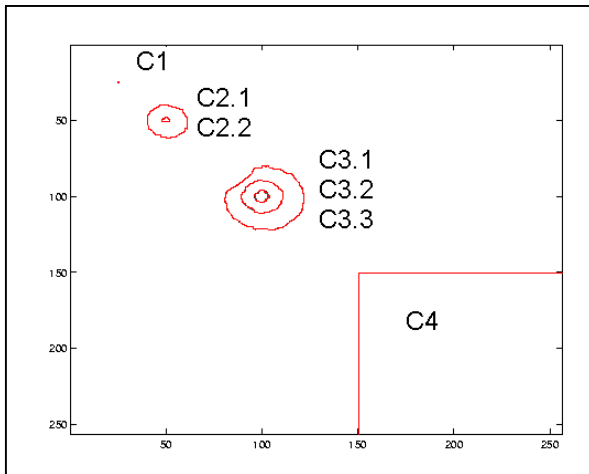


Figure 8. Résultat de la segmentation (fermeture des contours)

On constate que les contours sont correctement fermés pour les classes $C1$, $C2.1$, $C2.2$, $C3.1$, $C3.2$, $C3.3$ et $C4$. Si on veut retrouver les contours fermés correspondant au signaux f_2 et f_3 initialement représentés par les classes $C2$ et $C3$, il suffit de fusionner les classes $C2.1$ et $C2.2$ et les classes $C3.1$, $C3.2$ et $C3.3$.

9. Conclusion et perspectives

Actuellement les résultats de la classification sont satisfaisant et permettent de segmenter correctement des images de synthèse ou optique dès lors que la modélisation des signaux présents (types de transition) et/ou des interactions entre objets est correctement réalisée. Cette étape de construction des échantillons est importante mais non contraignante dès lors que l'on dispose d'une base de donnée regroupant les chaînes de maxima de différents types de transitions dans diverses situations (isolement, proximité d'autres types de transitions, ...). La construction d'une telle base est en cours.

Par ailleurs, lors de la construction des chaînes de maxima, l'algorithme d'optimisation combinatoire permettant d'affecter correctement les différents représentants à travers les niveaux de résolution peut être amélioré en intégrant des connaissances a priori sur le type de transitions cherché. Une étude est en cours sur cette partie.

Des améliorations peuvent également être envisagées pour les phases de segmentation et de reconnaissance de forme.

Pour la phase de segmentation, la recherche de contours fermés peut être abordée par une approche plus globale avec par exemple des contours actifs [23]. Ceci permettrait éventuellement de fermer certains contours lorsque les maxima obtenus après classification sont trop dispersés.

Pour la phase de reconnaissance de forme, le type de forme à reconnaître dépend de l'application. Néanmoins, des adaptations sont à envisager pour autoriser la

reconnaissance d'objets déformés (rotation, agrandissement/réduction) par rapport au modèle.

Annexe : remerciements

Ce travail est basé sur une étude réalisée par les auteurs en 1998. L'objectif était alors de détecter et d'isoler les singularités d'un signal dans des images radars (isolement par rapport au bruit). Cette précédente étude a été réalisée à l'ENSEEIH et en collaboration avec le CNES (Frédéric Adragnat) et le CESBIO (Armand Lopes, Roger Fjortoft) de Toulouse.

Bibliographie

- [1] Stephane G. Mallat and Hwang, "Singularity Detection And Processing With Wavelets", IEEE Trans. Inform. Theory, Vol 38, number 2, p 617-643, June 1992
- [2] Stephane G. Mallat and S. Zhong, "Characterization of signals from multiscale edges", IEEE Trans. Patt. Anal. Mach. Intell., vol 14, number 7, p 710-732, July 1992
- [3] Stephane G. Mallat, "A theory for multiresolution signal decomposition : the wavelet representation", IEEE Trans. Patt. Anal. Mach. Intell., vol 11, number 7, 1989
- [4] A. Enis, Cetin and Rashid Ansari, "Signal Recovery from Wavelet Transform Maxima", IEEE Trans. Signal Processing, vol 42, number 1, 1994
- [5] Enguerran Grandchamp, "Outils d'aide à la segmentation d'image : classification de ruptures par décomposition en ondelettes", INP-ENSEEIH, 1998
- [6] Yves Meyer, Stéphane Jaffard et Olivier Rioul, "L'analyse par ondelettes". Pour la science, September 1987
- [7] Ilya Polyak, and al, "Robust knot detection and spline approximation using wavelet transform extrema and multi-target tracking", TFT, 1994
- [8] Enguerran Grandchamp, Philippe Marthon, "Driving Segmentation and Recognition Phases using multiscale characterization", IGARSS 2003
- [9] Stephane G. Mallat, "A Wavelet Tour of Signal Processing", Academic Press, 1999
- [10] H.B. Nugraha and A.Z.R. Langi, "A wavelet-based Measurement of fractal dimensions of a 1D signal", APCCAS, 2002
- [11] Roger Fjortoft, "Analyse multiéchelle d'images SAR : étude bibliographique", 1997
- [12] J.J. Staal, "Characterizing the irregularity of measurements by means of the wavelet transform", 1995.
- [13] E. V. Hoekstra, "Multiscale Analysis Of Seismic Data By The Wavelet Transform", 1996
- [14] Enguerran Grandchamp, "Outils d'aide à la segmentation d'image : détection de contours sûrs par décomposition en ondelettes", rapport de stage, 1998
- [15] D. E. Golberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Adison-Wesley, 1989
- [16] Fred Glover, "Tabu Search",
- [17] Fred Glover, "Scatter Search And Star-Paths : Beyond The Genetic Metaphor", 1993
- [18] Fred Glover, "Tabu Search Part I", ORSA Journal On Computing, 1989
- [19] Fred Glover, "Tabu Search Part II", ORSA Journal On Computing, 1990
- [20] D. S. Johnson and al, "Optimization By Simulated Annealing : An Experimental Evaluation; Part I, Part II", Operational Research, 1989
- [21] J. Héroult, C. Jutten, « Réseaux neuronaux », Hermes, 1993
- [22] R. Rojas, « Neural Network », Springer, 1996
- [23] Michael Kass, Andrew Witkin et Demetri Terzopoulos « Snakes : Active Contour Models », International Journal of Computer Vision, vol. 1, n. 4, pp. 321-331, 1987