

COMPARAISON DE DEUX CRITÈRES EN CLASSIFICATION ASCENDANTE HIÉRARCHIQUE SOUS CONTRAINTE DE CONTIGUÏTÉ. APPLICATION EN IMAGERIE NUMÉRIQUE

Israël-César LERMAN*, Kaddour BACHAR**

RÉSUMÉ

Nous analysons une algorithmique de classification ascendante hiérarchique sous contrainte de contiguïté par agrégation des voisins réciproques en la situant dans le contexte général des algorithmes rapides de classification ascendante hiérarchique. Surtout, nous la déclinons selon deux types de critères. Il s'agit d'une part, du critère de Ward de la variation de l'inertie expliquée et d'autre part, d'une famille paramétrée du critère VL de la vraisemblance du lien maximal. Le contexte applicatif est celui de la segmentation d'image. On souligne la nature linéaire de la complexité algorithmique que nous montrons expérimentalement. L'influence algorithmique de la notion de contiguïté retenue est mise en évidence. Une nouvelle stratégie mettant en oeuvre l'agrégation multiple dans la formation des classes montre tout son intérêt. On étudie aussi bien sur le plan théorique qu'expérimental la possibilité d'inversions compte tenu du type de critère utilisé. Nous terminons en proposant une analyse comparative des résultats sur des données réelles en imagerie satellitaire.

Mots-clés : Classification hiérarchique, graphe de contiguïté, agrégation multiple, complexité, inversion, traitement d'image.

ABSTRACT

We analyse an algorithm of ascendant hierarchical classification under contiguity constraint and using the aggregation principle of reciprocal nearest neighbors. This algorithm is situated in the general framework of quick ascendant hierarchical classification algorithms. Two cluster merging criteria are studied. The former is the classical inertia Ward criterion and the latter consists of the maximal likelihood linkage family criteria. A new contiguity version of this criterion proves its efficiency in image segmentation. One major feature of our algorithm is the linear nature of the computational complexity. New strategies concerning multiple aggregation in the class formation and contiguity notion are positively evaluated in terms of quality and efficiency. We establish mathematically and experimentally how the used criterion

* IRISA - Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cédex, lerman@irisa.fr.

** ESSCA, 1, Rue Lakanal BP 348, 49003 Angers Cédex 01, k.bachar@essca.asso.fr.

influences inversion possibility in the tree building. Finally, comparative results of both types of criteria in image segmentation on satellite pictures are discussed.

Keywords : Hierarchical classification, contiguity graph, multiple aggregation, complexity, inversion, image processing.

1. Introduction et position du problème

Notre étude se situe dans le cadre général de la classification ascendante hiérarchique sous contrainte, d'un ensemble \mathcal{O} d'objets. L'exemple typique est celui où \mathcal{O} est défini par un ensemble d'unités géographiques (e.g. des communes) décrites par un ensemble \mathcal{V} de variables statistiques [35]. On suppose qu'il existe par ailleurs sur \mathcal{O} une relation de voisinage. Elle sera dans notre cas une relation de contiguïté spatiale. Cette relation binaire définit un graphe discret G que nous dirons de contiguïté. La description statistique (au moyen de \mathcal{V}) permet de munir \mathcal{O} d'un indice de dissimilarité d . L'extension de ce dernier pour évaluer la dissimilarité entre deux parties disjointes de \mathcal{O} , permet la construction ascendante hiérarchique d'un arbre de classification sur \mathcal{O} . Chaque niveau de cet arbre détermine une partition (classification) sur l'ensemble \mathcal{O} . La partition d'un niveau k donné se déduit de celle du niveau $(k - 1)$ qui précède en fusionnant les paires de classes également les plus proches au sens de l'indice de dissimilarité retenu entre parties disjointes de \mathcal{O} . Notons ce dernier δ . Dans le cas où la contrainte de contiguïté doit être respectée à chaque pas de la construction ascendante, les classes produites à chacun des niveaux de l'arbre des classifications sont nécessairement connexes au sens du graphe G . Si ce dernier traduit une notion de contiguïté spatiale, chacune des classes détermine une zone spatialement connexe. Un problème important consiste en la détermination de niveaux «intéressants» de l'arbre des classifications [16, 17].

Nous sommes concernés dans notre application par la segmentation d'images numérisées. Dans ce cas, l'ensemble \mathcal{O} est défini par l'ensemble des pixels d'une image rectangulaire. Deux notions de contiguïté entre classes que nous notons *Cont1* et *Cont2* seront prises en compte. Elles sont définies à partir de deux graphes de contiguïté sur l'ensemble des pixels de l'image, notés respectivement G_8 et G_4 . Pour G_8 (resp. G_4) chaque pixel est relié par une arête à l'ensemble des pixels qui lui sont spatialement adjacents, soit horizontalement, soit verticalement, soit diagonalement (resp. soit horizontalement, soit verticalement). D'autre part, dans notre application, les pixels sont supposés décrits au moyen de variables numériques. Précisons déjà que dans nos expériences il s'agit de la seule variable *luminance* (voir paragraphes 3, 4 et 5). Précisément et de façon générale, en considérant une description par des variables numériques de l'ensemble des objets, ou, plus généralement, une représentation euclidienne, deux types d'indices δ entre classes ou parties disjointes de \mathcal{O} seront considérés. Le premier est celui classique (critère de Ward [37]) de la perte d'inertie expliquée dans la fusion des deux classes définies par les deux ensembles disjoints à comparer. Notons D ce premier indice. Le second indice est fourni par un élément générique de la famille de

critères que nous notons $\Delta(\epsilon, \pi)$ de la Vraisemblance du Lien maximal [3, 4, 16, 20, 23]. Ces critères seront dits de type *VL*. D'ailleurs, un des objectifs majeurs de cette étude consiste à évaluer le comportement de cette famille, relativement au critère classique de l'inertie expliquée, dans le cadre de la classification ascendante hiérarchique « accélérée » sous contrainte de contiguïté spatiale de l'ensemble des pixels d'une image.

En effet, compte tenu de l'importance de la taille de l'ensemble traité (plusieurs centaines de milliers de points), il importe que l'algorithmique utilise les principes d'accélération de la classification ascendante hiérarchique. Il s'agira surtout du principe de la fusion des paires de classes plus proches voisines réciproques l'une de l'autre [8, 12]. Nous appelons une telle algorithmique une *CAHCVR* : Classification Ascendante Hiérarchique sous Contrainte procédant par agrégations successives des paires de classes Voisines Réciproques, au sens d'un indice δ de dissimilarité entre classes [3]. Comme nous l'avons mentionné ci-dessus, lorsque la contrainte est de contiguïté spatiale dans le cadre de la classification d'un ensemble d'unités spatiales, ce sont des classes spatialement connexes qui émergent à chaque étape du déroulement de l'algorithme de classification. Dans le cas de notre application en imagerie, nous souhaitons obtenir une « bonne » segmentation en zones connexes à partir de la partition déterminée par un niveau « intéressant » de l'arbre des classifications (voir paragraphe 5). C'est au paragraphe 2 que nous présenterons d'une façon spécifique l'algorithme de classification en faisant valoir, par rapport à l'agrégation binaire, celle multiple, où plusieurs paires de classes réalisent « en même temps », la propriété d'être plus proches voisines réciproques l'une de l'autre. Une telle circonstance d'exaequos dans la dissimilarité entre paires de classes est en effet loin d'être exceptionnelle. Elle se produit dans le cas où la description est limitée (e.g. quelques variables qualitatives ayant de faibles nombres de modalités) et où la taille de l'ensemble à organiser est très importante. Dans notre application en imagerie numérique, c'est la seule variable *luminance* qui a été considérée. L'intervalle de variation retenu est $[0, 1, \dots, l, \dots, 255]$. Ainsi, il comprend 256 valeurs entières. La valeur absolue de la différence entre deux quelconques des valeurs est comprise entre 0 et 255 ; d'autre part, toute fonction de dissimilarité est une fonction strictement croissante de cette valeur absolue. Ainsi, il y a au plus 256 valeurs distinctes de l'indice de dissimilarité ; lesquelles se répartissent sur l'ensemble des arêtes du graphe de contiguïté spatiale. En prenant le cas de G_4 (qui comprend près de deux fois moins d'arêtes que G_8) et en supposant une distribution uniforme de l'indice de dissimilarité, chaque valeur de ce dernier se répète près de 4000 fois pour une image 512×512 . En effet, l'ordre de grandeur du nombre d'arêtes de G_4 est 4 fois le nombre de pixels ; soit, 10^6 . De plus, on peut penser que compte tenu de la classifiabilité de l'image en zones connexes, la concentration des valeurs de la *luminance* sur le graphe de contiguïté spatiale est très sensiblement plus forte qu'elle n'aurait été dans le cas d'une distribution uniforme.

Dans ce même paragraphe 2 nous aborderons d'autres aspects algorithmiques ou métriques introduits pour réduire la complexité de la classification ascendante hiérarchique en général. Plus particulièrement, un résultat très impor-

tant concerne le caractère *linéaire* en moyenne de la complexité algorithmique de *CAHCVR*. Ce résultat sera juste évoqué et illustré sur le plan expérimental au paragraphe 5. Le développement de sa démonstration mathématique [2, 24] fera l’objet d’une future proposition d’article.

Au paragraphe 3, nous présenterons la famille paramétrée des critères de la vraisemblance du lien maximal en tenant compte d’un contexte applicatif où une contrainte de contiguïté est à respecter dans la fusion de classes connexes. Nous montrerons la construction de cette famille en partant d’une description de l’ensemble \mathcal{O} de n objets par un ensemble \mathcal{V} de p variables numériques. Cependant, il est entendu, qu’à la différence du critère de l’inertie expliquée, un critère de type *VL* (Vraisemblance du Lien) peut être utilisé quelle que soit la nature mathématique du tableau des données et l’hétérogénéité des variables de description ; ceci, aussi bien pour la classification de l’ensemble des variables [21, 22, 31] que de l’ensemble des objets [18, 25].

Un aspect théorique nouveau de notre contribution est développé au paragraphe 4. Il correspond à une étude comparative du critère de Ward et de la famille de critères de la Vraisemblance du Lien maximal dans le contexte de l’algorithmique proposée où on impose une contrainte de contiguïté spatiale. Nous établirons en particulier un résultat important sur la possibilité du phénomène dit de *l’inversion*. Pour ce dernier, la valeur de la plus petite dissimilarité à un niveau donné de la construction de l’arbre des classifications est plus petite que celle du niveau précédent. Ce phénomène s’avère – comme d’ailleurs, cela se confirme par l’expérience – tout à fait possible avec le critère D de l’inertie. Il est – du moins pour certains paramétrages « naturels » – impossible avec le critère Δ de la vraisemblance du lien maximal.

Les données traitées et les résultats obtenus seront évoqués au paragraphe 5. Il s’agit d’images satellitaires en niveaux de gris de l’Institut Géographique National (IGN). Ainsi, seule la variable *luminance* est prise en considération. Un résultat de la segmentation est défini par une classification en parties connexes de l’ensemble des pixels de l’image. En l’état actuel, le nombre de classes est un paramètre que l’on se fixe. Dans notre approche, la classification est obtenue en coupant l’arbre à un certain niveau. Dans ce cas, la question se pose de reconnaître les segmentations les plus « naturelles » dans le contexte d’un même critère de formation des classes (D ou Δ). On peut également envisager d’améliorer une partition obtenue à un niveau donné. Ces questions seront abordées dans la conclusion qui occupe le dernier paragraphe 6.

Ce texte donne une version étendue dont [3, 4] représentent des aspects partiels et résumés. Il reprend en la structurant une partie importante de [24]. Nous donnons ici l’ensemble de la construction méthodologique en la justifiant le plus précisément possible sur les plans théorique et expérimental.

Avant de commencer notre propre développement, signalons que le sujet général de la classification ascendante hiérarchique sous contrainte de contiguïté a occupé de façon sensible la littérature en classification [10, 11, 14, 28, 29, 33, 35]. Très souvent et comme nous l’avons déjà signalé, il s’agit d’une contrainte de contiguïté géographique relativement au problème de la classification d’un ensemble d’entités spatiales décrites par un ensemble de variables

statistiques. On souhaite des classes impérativement connexes et homogènes relativement à la description statistique.

2. Aspects généraux de l’algorithmique de la Classification Ascendante Hiérarchique sous Contrainte

2.1. Principe général et représentation

Sur le plan formel la donnée initiale est un triplet (\mathcal{O}, d, c) où \mathcal{O} désigne un ensemble fini d’objets dont le cardinal sera noté n , d est un indice de distance sur \mathcal{O} et c , une relation binaire symétrique sur \mathcal{O} définissant un graphe G sur \mathcal{O} . Dans notre application, \mathcal{O} sera l’ensemble des pixels d’une image et d sera établi à partir d’une description par des variables numériques. Comme nous l’avons déjà indiqué, il s’agira en fait dans notre application de la seule variable *luminance*. Cependant, la généralisation ne pose aucun problème. La relation c est définie par la contrainte qui sera dans notre cas une contrainte de contiguïté spatiale. c détermine ainsi un graphe G de contiguïté spatiale. On remarquera que le cas de la classification libre (sans contrainte) correspond au cas où la relation c est complète. Dans ce cas la donnée initiale peut être représentée par le couple (\mathcal{O}, d) .

La Classification Ascendante Hiérarchique sous Contrainte suppose l’extension de d et de c à la comparaison de deux parties disjointes de l’ensemble \mathcal{O} . Désignons respectivement par δ et C ces deux extensions. Nous avons indiqué ci-dessus que δ sera dans notre contexte décliné d’une part, à partir du critère de Ward et d’autre part, à partir d’un élément générique de la famille de critères de la Vraisemblance du Lien Maximal. L’extension C est définie comme suit :

DÉFINITION 2.1. — Soit $\{X, Y\}$ une paire de parties disjointes de \mathcal{O} , on a $XC Y$ si et seulement si, il existe $(x, y) \in X \times Y$ tel que xcy .

Clairement δ et C définissent respectivement, un graphe valué et un graphe discret sur l’ensemble $\mathcal{P}(\mathcal{O})$ des parties de \mathcal{O} .

L’algorithme de base de la Classification Ascendante Hiérarchique sous Contrainte est simple dans son principe. À l’initialisation, on établit la matrice de dissimilarité δ entre classes singletons, nommément :

$$\{\delta(\{x\}, \{y\}) \mid \{x, y\} \in P_2(\mathcal{O})\} \quad (1)$$

où $P_2(\mathcal{O})$ représente l’ensemble des parties à deux éléments de \mathcal{O} .

On établit d’autre part, le graphe G associé à la contrainte représentée par la relation binaire c . Dans ces conditions la progression de l’algorithme consiste en la fusion à chaque pas de la paire de classes $\{X, Y\}$ réalisant la valeur minimale de δ et pour lesquelles on a $XC Y$.

Commençons par simplifier et pour fixer les idées par considérer qu’à chaque étape de l’algorithme, la valeur minimale de δ est atteinte pour une seule

paire de classes. On obtient un arbre binaire de classifications dont nous proposons ici une expression formelle au moyen d'un ensemble étiqueté de fourches binaires qui s'organisent récursivement. Notons cet ensemble :

$$\{F(m) \mid 1 \leq m \leq 2n - 1\} \quad (2)$$

où $F(m)$ désigne la fourche d'étiquette m . Cependant, $\{F(m) \mid 1 \leq m \leq n\}$ représente l'ensemble des feuilles de l'arbre dont chacune indique un singleton comprenant un élément de l'ensemble O à organiser par la *CAH*. $\{F(m) \mid n + 1 \leq m \leq 2n - 1\}$ représente l'ensemble des fourches binaires de l'arbre, pour lesquelles on a une définition récursive :

$$F(m) = \{m, F(p(m)), F(q(m)), \nu(m)\} \quad (3)$$

m est l'étiquette du noeud de jonction entre la fourche ou feuille $F(p(m))$, représentant la « fille gauche » de $F(m)$ et $F(q(m))$ représentant la « fille droite » de $F(m)$. $\nu(m)$ représente l'indice de dissimilarité δ entre les deux classes soustendues par $p(m)$ et $q(m)$. Sans ambiguïté nous noterons :

$$\nu(m) = \delta(F(p(m)), F(q(m))) \quad (4)$$

où $\nu(m)$ définit la pondération de la fourche $F(m)$.

On a bien entendu :

$$\max(p(m), q(m)) < m \quad (5)$$

2.2. Agrégation multiple

Supposons qu'à une étape donnée t de l'algorithme de *CAH*, on obtienne la partition :

$$\pi^t = \{O_j^t \mid 1 \leq j \leq k_t\} \quad (6)$$

où O_j^t désigne la j -ème classe de la partition qui en comprend k_t classes.

Comme mentionné ci-dessus, un tri commence par déterminer la valeur minimale de la dissimilarité entre classes concernées; soit :

$$\delta_{min}^t = \min\{\delta^t(j, j') \mid 1 \leq j < j' \leq k_t\} \quad (7)$$

où $\delta^t(j, j')$ indique $\delta(O_j^t, O_{j'}^t)$, $1 \leq j < j' \leq k_t$

Comme nous l'avons exprimé dans l'introduction, une telle valeur minimale peut être atteinte plus d'une fois. En d'autres termes plusieurs paires de classes réalisent « en même temps » la plus petite valeur de la dissimilarité δ entre classes. Nous choisissons de fusionner toutes ces paires de classes avant de consulter à nouveau la matrice des dissimilarités entre classes formées. Deux versions algorithmiques peuvent être considérées. Dans la première, on procède – comme dans le cas classique – de façon binaire en agrégeant l'une après l'autre la suite des paires de classes réalisant δ_{min}^t . Dans la mesure où une même classe peut apparaître dans plus d'une seule paire, on ferme transitivement; de sorte que les classes obtenues à l'étape $t+1$ peuvent résulter de la fusion de plus de deux classes de l'étape t . C'est une fois le processus

d'agrégation achevé, qu'on opère la réactualisation des indices de dissimilarité δ ; mutuellement entre les nouvelles classes ainsi qu'entre les nouvelles classes et celles, restées invariantes entre l'étape t et l'étape $t + 1$. Nous appelons « binaire » une telle version.

Le résultat du deuxième algorithme est identique au précédent. Cependant, ce qui change c'est qu'au lieu de l'agrégation binaire pas à pas, on procède en une seule fois à l'agrégation multiple : toutes les paires de classes de π^t , réalisant δ_{min}^t sont agrégées en une seule fois. Ainsi, en considérant l'ensemble des sommets d'un graphe représentant les classes de π^t et pour lequel une arête rejoint deux sommets si la dissimilarité entre les deux classes concernées est égale à δ_{min}^t , on obtient les composantes connexes d'un tel graphe. Nous appelons « multiple » cette dernière version de l'algorithme. Elle est analysée dans [19]. Précisons que si à une étape donnée du déroulement de l'algorithme k est le nombre de classes plus proches voisines réciproques, la première version nécessite $k-1$ fusions binaires ; alors que la seconde n'en nécessite qu'une seule. Dans le graphique de la figure 4 on peut comparer les pentes d'accélération de la version binaire et de celle multiple du temps d'exécution de *CAHCVR* lorsque le nombre de sommets traités augmente. Cependant, d'autres options sont également considérées dans la littérature [5], ces dernières pouvant être plus coûteuses en nombre de comparaisons.

Il y a lieu de préciser que pour tenir compte du graphe G de contiguïté spatiale, on effectuera le tri sur l'ensemble des paires de classes en présence, pour lesquelles la relation C , ici de contiguïté spatiale définie ci-dessus est satisfaite. Quelle que soit la version de l'algorithme, un tri commence par déterminer la valeur minimale de la dissimilarité entre classes concernées ; à cette fin, on effectue un parcours de l'ensemble de toutes les paires de classes ou bien, dans la mesure où une contrainte de contiguïté est imposée pour les classes à fusionner, le sous ensemble des paires de classes liées par une telle contrainte.

2.2.1. Représentation en termes de fourches binaires

Une agrégation multiple peut bien être représentée au moyen d'une suite de fourches binaires et cette circonstance permet de préserver la forme binaire de l'arbre des classifications. Considérons en effet le cas le plus simple d'une agrégation multiple faisant intervenir trois noeuds. Supposons qu'à un niveau donné t de la construction ascendante hiérarchique, la valeur minimale δ_{min}^t est atteinte deux fois au moins relativement à trois noeuds notés p , q et r de la manière suivante :

$$\delta(F(p), F(q)) = \delta(F(p), F(r)) = \delta_{min}^t \quad (8)$$

Supposons sans restreindre la généralité que $p < q < r$. Dans ces conditions, on pose :

$$\begin{aligned} F(t) &= \{t, F(p), F(q), \delta_{min}^t\} \\ F(t+1) &= \{t+1, F(t), F(r), \delta_{min}^t\} \end{aligned} \quad (9)$$

dont la représentation graphique est donnée dans la figure 1 :

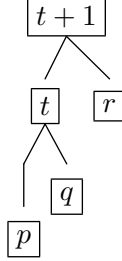


FIG 1. — Séquence de fourches.

où $\nu(t) = \nu(t + 1) = \delta_{min}^t$.

Plus généralement, si l'agrégation multiple porte sur un ensemble de k noeuds ou feuilles, elle sera représentée par un peigne dont les noeuds seront étiquetés de $t + 1$ à $t + k - 1$ et où :

$$\nu(t) = \nu(t + 1) = \dots = \nu(t + k - 1) = \delta_{min}^t \quad (10)$$

Ainsi, pour l'arbre indicé construit, un intervalle croissant et connexe de fourches également pondérées représente une agrégation multiple. Les noeuds concernés soustendent les sommets de fourches de la descendance du sommet du peigne jusqu'à la rencontre de fourches de valuation strictement plus petite que δ_{min}^t .

Maintenant, relativement à l'équation (3) et dans le cas où

$$\nu(m) \neq \max(\nu(p(m)), \nu(q(m))) \quad (11)$$

il y a – ce qui souhaitable – monotonie dans la définition de $F(m)$, si

$$\max(\nu(p(m)), \nu(q(m))) \leq \nu(m) \quad (12)$$

il y a inversion si

$$\max(\nu(p(m)), \nu(q(m))) > \nu(m) \quad (13)$$

2.2.2. Formules de réactualisation

Ces formules ont été introduites par Lance et Williams [13]. Elles permettent de façon récursive de réactualiser la matrice des similarités entre classes après une fusion de deux classes. Nous allons rappeler ces formules dans le cas des deux critères que nous aurons à comparer : le critère D de la variation de l'inertie expliquée et le critère générique Δ_ϵ de la vraisemblance du lien maximal. En effet, ces formules interviendront de façon cruciale au paragraphe 4. Le critère Δ_ϵ dépend d'un paramètre numérique ϵ compris entre 0 et 1, que nous présenterons au paragraphe 3.

Désignons par X , Y et Z trois parties mutuellement disjointes de \mathcal{O} . Il s'agira dans le contexte de la CAH de trois classes en présence à un niveau donné de l'arbre des classifications. Relativement à un indice de dissimilarité δ entre

parties disjointes de \mathcal{O} , définissant le critère de fusion entre classes, une formule de réactualisation se présente sous la forme :

$$\delta(X \cup Y, Z) = f(\delta(X, Y), \delta(X, Z), \delta(Y, Z)) \quad (14)$$

où f est une fonction pouvant faire intervenir les cardinaux de X , Y et Z , que nous noterons, respectivement, ξ , η et ζ . Ainsi pour $\delta = D$, on a :

$$D(X \cup Y, Z) = \frac{1}{\xi + \eta + \zeta} \{(\xi + \zeta)D(X, Z) + (\eta + \zeta)D(Y, Z) - \zeta D(X, Y)\} \quad (15)$$

et pour $\delta = \Delta_\epsilon$, on a :

$$\Delta_\epsilon(X \cup Y, Z) = \min\left\{\left(1 + \frac{\eta}{\xi}\right)^\epsilon \Delta_\epsilon(X, Z), \left(1 + \frac{\xi}{\eta}\right)^\epsilon \Delta_\epsilon(Y, Z)\right\} \quad (16)$$

Une généralisation de ces équations en cas d'agrégations multiples est obtenue dans [19]. Pour exprimer sa forme, considérons relativement à une partition $\pi = \{\mathcal{O}_j | 1 \leq j \leq k\}$, deux fusions multiples de classes que nous pouvons noter $\mathcal{O}_{j_1} \cup \mathcal{O}_{j_2} \cup \dots \cup \mathcal{O}_{j_r}$ et $\mathcal{O}_{h_1} \cup \mathcal{O}_{h_2} \cup \dots \cup \mathcal{O}_{h_s}$ et où $\{j_1, j_2, \dots, j_r\}$ et $\{h_1, h_2, \dots, h_s\}$ sont deux sous ensembles disjoints de $\{1, 2, \dots, j, \dots, k\}$. Relativement à un indice de dissimilarité δ entre classes, on a :

$$\begin{aligned} & \delta(j_1 \cup j_2 \cup \dots \cup j_r, h_1 \cup h_2 \cup \dots \cup h_s) \\ &= f(\{\delta(j_l, j_{l'}) | 1 \leq l \leq r, 1 \leq l' \leq r\}, \\ & \quad \{\delta(h_m, h_{m'}) | 1 \leq m \leq s, 1 \leq m' \leq s\}, \\ & \quad \{\delta(j_l, h_m) | 1 \leq l \leq r, 1 \leq m \leq s\}) \end{aligned} \quad (17)$$

où $j_1 \cup j_2 \cup \dots \cup j_r$ (resp. $h_1 \cup h_2 \cup \dots \cup h_s$) représente, avec les notations qui précèdent, l'union des classes $\mathcal{O}_{j_1}^t \cup \mathcal{O}_{j_2}^t \cup \dots \cup \mathcal{O}_{j_r}^t$ (resp. $\mathcal{O}_{h_1}^t \cup \mathcal{O}_{h_2}^t \cup \dots \cup \mathcal{O}_{h_s}^t$) et où f est une fonction numérique faisant intervenir les tailles ou poids des classes \mathcal{O}_{j_l} et \mathcal{O}_{h_m} , $1 \leq l \leq r, 1 \leq m \leq s$.

2.3. Voisins réciproques et voisinages réductibles

Le principe de l'agrégation multiple que nous avons exprimé dans le cadre de l'algorithme de base de la CAH, peut opérer dans les algorithmes dits « accélérés » de classification de « gros » ensembles, apparus dans les années 80 [6, 8, 9, 12]. Deux méthodes importantes et non sans lien sont apparues. Celle dite des « voisinages réductibles » et celle dite des « voisins réciproques » qui est maintenant la plus établie. Il s'agit en effet d'une stratégie simple et efficace d'agrégation où à chaque étape on fusionne les paires de classes qui sont respectivement, plus proches voisines réciproques l'une de l'autre.

À l'initialisation, on commence avec la partition discrète :

$$\pi^0 = \{\{x\} | x \in \mathcal{O}\} \quad (18)$$

Si

$$\pi^s = \{\mathcal{O}_j^s | 1 \leq j \leq k_s\} \quad (19)$$

est la partition obtenue à une étape donnée, $\{O_{j_1}^s, O_{j_2}^s\}$ est une paire de voisins réciproques si et seulement si

$$\begin{aligned}\delta(O_{j_1}^s, O_{j_2}^s) &= \min\{\delta(O_{j_1}^s, O_h^s) \mid 1 \leq h \leq k_s, h \neq j_1\} \\ &= \min\{\delta(O_h^s, O_{j_2}^s) \mid 1 \leq h \leq k_s, h \neq j_2\}\end{aligned}\quad (20)$$

La progression consiste alors à réunir «en même temps» (ou en parallèle) toutes les paires de classes voisines réciproques l'une de l'autre.

L'arrêt de l'algorithme se produit lorsqu'il ne subsiste qu'une seule classe \mathcal{O} .

On démontre que, pourvu que δ ait une propriété dite de réductibilité ou de contractance mise en évidence par M. Bruynooghe [7], la hiérarchie de parties obtenue par le dernier algorithme se retrouve dans celle associée aux feuilles ou noeuds de l'algorithme de base de la *CAH*. Rappelons cette propriété. Elle s'exprime relativement à une partition π de l'ensemble \mathcal{O} . Soit X un élément de π et δ_0 une valeur fixée de δ . Relativement à (X, δ_0) , introduisons la boule

$$\mathcal{B}(X, \delta_0) = \{Z \mid Z \in \pi \text{ et } \delta(X, Z) \leq \delta_0\} \quad (21)$$

DÉFINITION 2.2. — *L'indice δ est contractant (on dit encore réductible) si pour toute paire de classes $\{X, Y\}$ de π et pour toute valeur δ_0 de δ , on a :*

$$\mathcal{B}(X \cup Y, \delta_0) \subset \mathcal{B}(X, \delta_0) \cup \mathcal{B}(Y, \delta_0) \quad (22)$$

En d'autres termes, pour un rayon δ_0 donné, la boule autour de la réunion est incluse dans la réunion des boules. Dans ces conditions, un sommet Z – représentant une classe de π – qui appartient au voisinage de $X \cup Y$, appartient au voisinage de X ou (non exclusif) à celui de Y .

Précisons – comme on peut l'établir à partir des formules de réactualisation – que chacun des deux indices D (moment d'inertie) et Δ_ϵ (vraisemblance du lien) sont bien contractants.

La fusion simultanée et à chaque étape de toutes les paires de classes voisines réciproques permet bien, comme nous venons de l'indiquer, de retrouver la hiérarchie de parties induite par une *CAH*. Cependant, elle ne permet pas de retrouver les niveaux indicés de l'arbre des classifications obtenu par l'algorithme de base de la *CAH*. Chacun de ces niveaux détermine une partition de l'ensemble organisé. La nécessité d'obtenir la suite des niveaux de l'arbre indicé des classifications nous conduira – notamment dans *CAHCVR* – à stocker à chaque étape l'ensemble des paires de classes plus proches voisines réciproques et à les fusionner dans l'ordre croissant des valeurs des dissimilarités entre ces paires de classes. Chacune de ces dernières valeurs donne lieu à un niveau de l'arbre des classifications et donc, à un numéro de noeud.

L'idée maîtresse concrétisée dans [8] consiste à opérer dans l'espace des classes, les fusions conformément à l'algorithme des voisins réciproques; mais, à l'intérieur des boules (dont chaque sommet représente une classe) définies dans la méthode des «voisinages réductibles». On peut ici indiquer que si

pour D , une classe admet une représentation par son centre de gravité et son poids, il n'en est pas de même pour Δ_ϵ .

Nous en parlons, parce que, dans notre algorithme *CAHCVR* nous développons une idée à laquelle nous avons été conduits de façon indépendante [1, 2, 23] et qui a un caractère beaucoup plus spécifique. La notion de voisinage est en effet fournie d'une façon exogène par la notion de contiguïté spatiale. En d'autres termes, on agrège par voisins réciproques à l'intérieur de zones géographiques connexes de l'image. Ainsi, l'agrégation se fait à l'intersection de deux graphes ; celui fourni par la contiguïté spatiale et celui fourni par la relation de voisinage réciproque au sens de la dissimilarité δ . Au cours de la progression de l'algorithme, après chaque fusion, il y a une propagation des deux graphes.

3. La famille paramétrée de critères ; le cas de la contrainte de contiguïté

3.1. Le critère de la vraisemblance du lien maximal

Le principe du critère de la vraisemblance du lien maximal est essentiellement différent de celui de la variation de l'inertie expliquée. Il peut d'ailleurs tout aussi bien être considéré pour la classification des variables que pour celle, qui nous concerne ici, des objets [16, 18, 20, 21, 22, 25, 30, 31].

L'indice de comparaison de classes d'objets est construit à partir d'un indice de comparaison des objets deux à deux. Chacun de ces deux indices est probabiliste et se réfère à une échelle $[0, 1]$. On se situe par rapport à une hypothèse d'absence de liaison ou d'indépendance entre les variables observées de description. Conceptuellement, on associe au tableau \mathcal{T} de données observées, un tableau aléatoire \mathcal{T}^* , où à la suite des variables observées, on associe une suite de variables aléatoires indépendantes ayant respectivement les mêmes distributions ou les mêmes caractéristiques distributionnelles que celles observées. Le modèle aléatoire peut ainsi être permutational ou posé *a priori*. Dans le cadre de ce modèle si $P(i^*, i'^*)$ est l'indice aléatoire associé à l'indice probabiliste $P(i, i')$ entre les deux objets i et i' , $P(i^*, i'^*)$ est une variable aléatoire uniforme sur l'intervalle $[0, 1]$.

Toutefois $P(i, i')$ s'obtient selon un processus bien identifié [18, 25]. Relativement à une variable v^j et à une paire d'objets $\{o_i, o_{i'}\}$, $1 \leq j \leq p$, $1 \leq i < i' \leq n$, on introduit la contribution brute $s^j(i, i')$ de la variable v^j à la comparaison des deux objets o_i et $o_{i'}$. Cette dernière est posée par l'expert et se déduit en général facilement de la sémantique des données et du problème posé [25].

Considérons le cas classique où la classification n'est pas contrainte par un graphe de contiguïté. Tout se passe comme si on avait un graphe G_c complet des valeurs de l'indice brut s^j sur l'ensemble de tous les couples d'objets représentés par $I \times I$, où $I = \{1, 2, \dots, i, \dots, n\}$ indexe l'ensemble O des objets. Dans ces conditions, à partir de $s^j(i, i')$, on définit la contribution normalisée

de la j -ème variable à la comparaison des deux objets o_i et $o_{i'}$, sous la forme :

$$S^j(i, i') = \frac{s^j(i, i') - \text{moy}(s^j)}{\sqrt{\text{var}(s^j)}} \quad (23)$$

où $\text{moy}(s^j)$ et $\text{var}(s^j)$ sont la moyenne et la variance de s^j sur $I \times I$.

L'intégration des contributions des différentes variables se fait de façon additive. Nous définissons :

$$S(i, i') = \sum_{j=1}^p S^j(i, i') \quad (24)$$

Cet indice global est normalisé statistiquement relativement à l'ensemble $P_2(I)$ des paires d'objets distincts pour obtenir :

$$Q_s(i, i') = \frac{S(i, i') - \text{moy}_e(S)}{\sqrt{\text{var}_e(S)}} \quad (25)$$

où $\text{moy}_e(S)$ et $\text{var}_e(S)$ sont la moyenne et la variance empiriques de $S(i, i')$ sur $P_2(I)$.

L'indice probabiliste s'obtient alors au moyen de l'expression :

$$P(i, i') = \Phi[Q_s(i, i')] \quad (26)$$

où Φ désigne la fonction de répartition de la loi normale centrée et réduite. L'équation (26) se justifie parce que sous l'hypothèse d'absence de liaison telle qu'esquissée ci-dessus, $Q_s(i^*, i'^*)$ se comporte, pour un nombre suffisant de descripteurs, comme une variable aléatoire normale, centrée et réduite. L'importance de la valeur de $P(i, i')$ mesure le degré d'exceptionnalité statistique de la grandeur de $S(i, i')$, dans le contexte de l'hypothèse d'absence de liaison.

On associera à l'indice probabiliste $P(i, i')$, l'indice de « dissimilarité informationnelle » $\delta(i, i')$ donné par :

$$\delta(i, i') = -\log_2(P(i, i')) \quad (27)$$

qui représente la quantité d'information sous-jacente à l'évènement dont la probabilité est $P(i, i')$. Cet indice de dissimilarité varie de façon monotone entre zéro et l'infini lorsque l'indice probabiliste varie entre 1 et 0.

Le premier problème qui se pose dans l'adaptation d'un critère de type VL au cas où il y a lieu de se limiter à un graphe de contiguïté spatiale G , consiste à bâtir l'indice probabiliste P [cf. (26)]. Comme ci-dessus, relativement à un couple (i, i') de G et à une variable v^j , on introduit la contribution brute $s^j(i, i')$ de la variable v^j à la comparaison des deux objets o_i et $o_{i'}$. On établit ainsi la distribution de s^j sur uniquement G :

$$\{s^j(i, i') \mid (i, i') \in G\} \quad (28)$$

L'expression de la contribution normalisée $S^j(i, i')$ de v^j à la comparaison de o_i et de $o_{i'}$ est la même que ci-dessus [cf. (23)]; mais à condition que la moyenne et la variance de s^j soient calculées sur la base de la distribution (28) ci-dessus.

On passe alors au moyen de la formule (24) à l'indice $S(i, i')$ qui représente la somme des contributions normalisées $S^j(i, i')$. Considérons alors la distribution :

$$\{S(i, i') \mid (i, i') \in G\} \quad (29)$$

et désignons par $moy_e^0(S)$ et $var_e^0(S)$ la moyenne et la variance empiriques d'une telle distribution. L'indice globalement normalisé s'écrit alors [voir 25] :

$$Q_s^0(i, i') = \frac{S(i, i') - moy_e^0(S)}{\sqrt{var_e^0(S)}} \quad (30)$$

Le passage à une échelle de probabilité s'effectue comme dans (26) en considérant l'application de la fonction de répartition de la loi normale centrée et réduite :

$$P(i, i') = \Phi[Q_s^0(i, i')], \quad (31)$$

$(i, i') \in G$.

Considérons à présent le cas particulier de notre application où $I = \{1, 2, \dots, i, \dots, n\}$ est défini par l'ensemble des pixels d'une image qu'on munit d'un graphe de contiguïté G qui peut être G_8 ou G_4 (cf. Introduction). Sur un tel ensemble est donné la valeur de la variable *luminance* dont l'échelle numérique est définie par l'intervalle $[0, 255]$:

$$\{x_i \mid 1 \leq i \leq n\} \quad (32)$$

On peut définir la contribution brute $s(i, i')$ sous la forme :

$$s(i, i') = 255 - |x_i - x_{i'}| \quad (33)$$

pour (i, i') de G . Nous laissons au lecteur le soin de spécifier dans ce cas les formules précédentes. Compte tenu du fait qu'une seule variable est en jeu, les expressions de S et de Q_s^0 deviennent identiques. Ce n'est bien sûr plus le cas quand on a deux variables ou plus. Le passage à une échelle de probabilité via la fonction de répartition de la loi normale correspond à une solution simple et efficace. En effet, tout se passe comme si on se situait par rapport à un modèle aléatoire d'absence de liaison où $S(i^*, i'^*)$, $((i^*, i'^*) : \text{couple aléatoire de pixels indépendants mais contigus})$ suit une loi normale centrée et réduite. Ne s'agissant que d'une seule variable, le choix d'une telle loi peut paraître *a priori*. Cependant, toute autre loi d'une large famille de lois aurait eu ce même caractère et conduirait de toute façon à des résultats pratiques comparables. L'important est d'évaluer au moyen d'une échelle de probabilité posée dans une hypothèse statistique d'indépendance, l'écart centré réduit que constitue $S(i, i')$.

Il y a lieu à présent d'exprimer dans le contexte d'un graphe de contiguïté la famille de critères de la vraisemblance du lien maximal. Mais d'abord, commençons par nous abstraire d'une telle contrainte de contiguïté en supposant que les deux classes C et D sont potentiellement reliées par $|C| \times |D|$ arêtes. Dans ces conditions, si

$$\{P(c, d) | (c, d) \in C \times D\} \quad (34)$$

est la famille d'indices probabilistes, valuant ces arêtes, un critère de dissimilarité informationnelle se référant à la vraisemblance du lien maximal se met sous la forme :

$$\Delta_\epsilon(C, D) = (|C| \times |D|)^\epsilon (-\log_2(p(C, D))) \quad (35)$$

où

$$p(C, D) = \max\{P(c, d) | (c, d) \in C \times D\} \quad (36)$$

et où ϵ est un paramètre réel positif compris entre 0 et 1 [15, 20, 30]. ϵ est un paramètre à ajuster compte tenu de la nature des données traitées. Toutefois, deux valeurs de ϵ peuvent être distinguées; d'une part, $\epsilon = 1$ et d'autre part, $\epsilon = 0.5$. Dans ce dernier cas, la moyenne géométrique entre $|C|$ et $|D|$, apparaît comme facteur multiplicatif de pondération pour définir l'indice définitif à partir d'un indice qui se réfère au lien simple [cf. (35)]. Ce dernier est obtenu à partir de (36). On pourra comparer avec intérêt de façon formelle et très générale les expressions (35) et (42).

Signalons que pour notre problème de segmentation la valeur de ϵ la plus appropriée se situe, selon les cas, entre 0.35 et 0.5; et, s'il s'agit d'enlever toute ambiguïté et de ne retenir qu'une seule valeur, on prendra, aussi pour des raisons formelles, $\epsilon = 0.5$ (cf. ci-dessus).

Dans cette étude, un autre paramètre que nous appellerons π s'est avéré fondamental. Il s'agit de la valeur de l'indice brut $p(C, D)$ à partir de laquelle on autorise la possibilité d'une fusion de classes. Dans les expériences menées, une valeur de 0.45 de π s'est avérée très adéquate. Pour aboutir à une telle stratégie, on posera pour tout (i, i') appartenant au graphe G de contiguïté spatiale :

$$P(i, i') = \begin{cases} P(i, i') & \text{si } P(i, i') > \pi \\ \eta & \text{si } P(i, i') \leq \pi \end{cases} \quad (37)$$

où η est une valeur suffisamment petite (par exemple $\eta = 10^{-6}$). De la sorte, on imposera de ne plus discriminer pour la classification les valeurs de l'indice probabiliste inférieurs à π .

Dans l'élaboration de l'indice $P(i, i')$ entre unités ponctuelles, nous avons bien tenu compte du graphe G de contiguïté spatiale. Cependant tel n'est pas le cas pour l'indice $\Delta_\epsilon(C, D)$ [cf. (35)] relativement à la comparaison de deux parties connexes C et D de G . Le dernier indice est conçu dans le cas libre où C et D se trouvent reliés par $|C| \times |D|$ arêtes. Or dans le cas où il y a lieu de tenir compte de la contrainte de contiguïté, l'ensemble des arêtes de jonction prend la forme $(C \times D) \cap G$. En conséquence, aux expressions (35) et (36),

nous substituons les expressions :

$$\Delta_\epsilon^0(C, D) = a(C, D)^\epsilon (-\log_2(p_0(C, D))) \quad (38)$$

avec

$$p_0(C, D) = \max\{P(c, d) | (c, d) \in (C \times D) \cap G\} \quad (39)$$

où $a(C, D)$ représente le nombre d'arêtes de jonction entre les deux classes contiguës C et D .

C'est ce dernier indice qui a donné les résultats les plus probants dans l'application. Néanmoins, face au phénomène de l'inversion en cas de contrainte de contiguïté (cf. paragraphe 4) nous analyserons chacun des deux indices $\Delta_\epsilon(C, D)$ et $\Delta_\epsilon^0(C, D)$, correspondants à ce que nous pouvons appeler la forme complète et la forme contiguë.

Il y a ici lieu et pour la suite de souligner que l'expression de l'indice δ entre classes singletons, prend la forme (27), dont la construction pour notre application démarre avec l'indice brut (33).

3.2. Rappel du critère de Ward (variation de l'inertie expliquée)

Rappelons ce critère très classique dit de Ward [37] qui est très bien implanté dans les logiciels. Considérons pour cela, associé à la description de n objets par p variables numériques, un nuage de points :

$$\mathcal{N}(I) = \{(M_i, \mu_i)\} \quad (40)$$

dans l'espace R^p muni d'une métrique euclidienne. μ_i est le poids affecté au sommet M_i représentant le i -ème objet, $1 \leq i \leq n$. Soit maintenant $\{C, D\}$ une paire de classes définie par une paire de parties disjointes de l'ensemble des objets. Dans un tel critère la représentation d'une classe X s'effectue au moyen d'un seul sommet pesant ; ce qui se présente sous la forme d'un couple $(g(X), \mu(X))$ où $g(X)$ est le centre de gravité du sous-nuage de points représentant X et où $\mu(X)$ est le poids de ce sous-nuage (c'est-à-dire ; la somme des poids ponctuels de tous les sommets qui y appartiennent).

Ce critère de dissimilarité entre C et D est défini par le moment d'inertie du nuage suivant de deux points pesants :

$$\{(g(C), \mu(C)), (g(D), \mu(D))\} \quad (41)$$

On montre que ce dernier critère peut se mettre sous la forme :

$$\frac{\mu(C) \times \mu(D)}{\mu(C) + \mu(D)} \times d^2(g(C), g(D)) \quad (42)$$

où d est la distance associée à la métrique dont se trouve muni l'espace R^p de représentation du nuage de points. Le plus souvent et en tout cas dans notre application, μ se trouve défini par la cardinalité.

Considérons maintenant une construction binaire de l'arbre des classifications où entre les niveaux l et $l + 1$, les deux classes C et D fusionnent. Dans

ces conditions, l'expression (42) représente la perte d'inertie entre l'inertie expliquée par la partition de niveau l et celle de niveau $l+1$. Ainsi, en associant à chaque noeud interne de l'arbre (racine comprise) la perte d'inertie résultant de sa formation, on obtient une décomposition additive de l'inertie totale du nuage.

3.3. Cas de la contrainte de contiguïté

Nous venons de voir que le critère d'agrégation peut être celui de Ward ou être défini par un élément générique de la famille de critères de la Vraisemblance du Lien maximal. En cas de contrainte de contiguïté, le calcul du critère choisi ne se trouve effectué que pour les paires de classes participant au graphe de contiguïté. Précisément, donnons ici une définition plus formelle que celle exprimée dans l'introduction des deux notions de contiguïté entre pixels qui ont été utilisées. À cette fin, on désignera par $X(i)$ et $Y(i)$ les coordonnées dans le plan de l'image d'un pixel i , $1 \leq i \leq n$.

DÉFINITION 3.1. — *Deux pixels i et i' sont contigus au sens de G_8 , si et seulement s'il n'y a pas un pixel intermédiaire; ce qui se traduit par :*

$$\begin{aligned} X(i') &= X(i) + \alpha \\ Y(i') &= Y(i) + \beta \end{aligned} \tag{43}$$

où α et β sont susceptibles de prendre l'une des trois valeurs -1 , 0 ou 1 ; mais où $(\alpha, \beta) \neq (0, 0)$.

Ainsi, un sommet i a 8 sommets contigus s'il est à l'intérieur de l'image. Il n'en a plus que 5 s'il est au bord de l'image sans être au coin et il en a exactement 3 s'il est au coin.

DÉFINITION 3.2. — *Deux pixels i et i' sont contigus au sens de G_4 , si et seulement si la distance de Hamming ou «city block» entre i et i' est de 1; soit :*

$$[X(i) - X(i')] + [Y(i) - Y(i')] = 1 \tag{44}$$

Ainsi, pour G_4 un sommet i a 4 sommets contigus s'il est à l'intérieur de l'image. Il n'en a plus que 3 s'il est au bord de l'image mais sans être au coin et il en a exactement 2 s'il se trouve à l'un des quatre coins.

4. Le problème de l'inversion en cas de CAH sous contrainte de contiguïté

Considérons la construction ascendante hiérarchique d'un arbre de classification au moyen d'un indice δ de dissimilarité entre classes (cf. paragraphe 2).

DÉFINITION 4.1. — *Le critère δ est dit sans inversion ou monotone si et seulement si la valeur minimale de l'indice δ entre classes en présence au niveau $k+1$ est supérieure à celle, au niveau k .*

Ce caractère monotone est clairement acquis pour chacun des deux critères D (variation de l'inertie expliquée) et Δ_ϵ (associé à la vraisemblance du lien maximal) dès lors que la formation ascendante hiérarchique de l'arbre des classifications est libre. Il suffit de l'établir dans le cadre d'une fusion binaire. Pour ce faire, on se sert des formules de réactualisation déjà exprimées ci-dessus (cf. (15), (16)). Ces formules vont également jouer un rôle crucial dans le cas de la *CAH* sous contrainte de contiguïté.

Reprenons ci-dessous un diagramme de situation déjà considéré dans [1] (voir Fig. 2) et soient A , B et C un ensemble de trois classes. α , β et γ désignent leurs poids respectifs. D'autre part, δ désigne l'indice de dissimilarité entre classes. On suppose que (A, B) et (A, C) appartiennent au graphe G de contiguïté spatiale; mais que (B, C) en est exclu. Imaginons sans restreindre la généralité que $\delta(A, B) < \delta(A, C)$. Dans ces conditions, une situation caractéristique pouvant occasionner une inversion dans le cadre d'une agrégation binaire est fournie par :

$$\delta(B, C) < \delta(A, B) < \delta(A, C) \quad (45)$$

Dans une telle agrégation, c'est la classe $A \cup B$ qui est formée en premier, suivie de celle $A \cup B \cup C$. Le phénomène d'inversion est défini par :

$$\delta(A \cup B, C) < \delta(A, B) \quad (46)$$

Il peut être dû à la grande proximité relative entre B et C . Nous schématisons la situation au moyen du graphique suivant où les distances perçues géométriquement reflètent les dissimilarités entre classes :

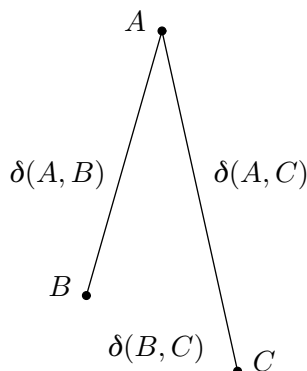


FIG 2. — Inversion possible.

Nous allons dans ces conditions étudier la possibilité d'occurrence d'une inversion pour chacun des critères D , Δ_ϵ et Δ_ϵ^0 .

4.1. Cas du critère D de l'inertie

Compte tenu de la formule de réactualisation, une inversion signifie :

$$\frac{\beta + \gamma}{\alpha + \beta + \gamma} D(B, C) + \frac{\alpha + \gamma}{\alpha + \beta + \gamma} D(A, C) - \frac{\gamma}{\alpha + \beta + \gamma} D(A, B) < D(A, B) \quad (47)$$

Cette inégalité se réduit à :

$$D(B, C) < D(A, B) - \frac{\alpha + \gamma}{\beta + \gamma} (D(A, C) - D(A, B)) \quad (48)$$

laquelle peut également s'écrire sous la forme :

$$D(A, B) - D(B, C) > \frac{\alpha + \gamma}{\beta + \gamma} (D(A, C) - D(A, B)) \quad (49)$$

Pour se rendre compte de la « facilité » avec laquelle une inversion peut se produire avec le critère D , considérons le cas $\alpha \leq \beta$. On a alors :

PROPRIÉTÉ 4.1. — *Pour $\alpha \leq \beta$, il suffit que l'écart entre les deux dissimilarités $D(A, B)$ et $D(B, C)$ soit plus grand que celui entre $D(A, C)$ et $D(A, B)$ pour que l'inversion ait lieu [voir (49)].*

4.2. Cas de la forme complète Δ_ϵ du critère de la vraisemblance du lien maximal

Reprenons avec nos actuelles notations la formule de réactualisation (16) :

$$\Delta_\epsilon(A \cup B, C) = \min \left\{ \left(1 + \frac{\beta}{\alpha}\right)^\epsilon \Delta_\epsilon(A, C), \left(1 + \frac{\alpha}{\beta}\right)^\epsilon \Delta_\epsilon(B, C) \right\} \quad (50)$$

où on suppose :

$$\Delta_\epsilon(B, C) < \Delta_\epsilon(A, B) < \Delta_\epsilon(A, C) \quad (51)$$

Dans le cas où le minimum est atteint pour

$$\left(1 + \frac{\beta}{\alpha}\right)^\epsilon \Delta_\epsilon(A, C) \quad (52)$$

l'inversion ne peut clairement pas avoir lieu en raison de l'inégalité $\Delta_\epsilon(A, C) > \Delta_\epsilon(A, B)$; puisqu'alors *a fortiori*

$$\left(1 + \frac{\beta}{\alpha}\right)^\epsilon \Delta_\epsilon(A, C) > \Delta_\epsilon(A, B) \quad (53)$$

Prenons le second cas où le minimum est atteint pour

$$\left(1 + \frac{\alpha}{\beta}\right)^\epsilon \Delta_\epsilon(B, C) \quad (54)$$

Il y a inversion si :

$$\left(1 + \frac{\alpha}{\beta}\right)^\epsilon \Delta_\epsilon(B, C) < \Delta_\epsilon(A, B) \quad (55)$$

En notant

$$d_{\min}(X, Y) = \min\{-\log_2(P(x, y)) \mid (x, y) \in X \times Y\}, \quad (56)$$

une telle inversion est théoriquement possible. Elle se produit si et seulement si d'après (35) :

$$\frac{d_{\min}(B, C)}{d_{\min}(A, B)} < \left(\frac{\gamma}{\alpha} + \frac{\gamma}{\beta}\right)^{-\epsilon} \quad (57)$$

ce qui représente une inégalité assez sévère ; d'autant plus, que la classe C est de poids γ important relativement aux poids α et β (voir le second membre de la dernière équation (57)). On peut s'en rendre compte à partir du cas particulier où $\alpha = \beta$.

PROPRIÉTÉ 4.2. — Si $d_{\min}(B, C)$ excède la part $1/\left(\frac{\gamma}{\alpha} + \frac{\gamma}{\beta}\right)^\epsilon$ de $d_{\min}(A, B)$, l'inversion est impossible.

Nous allons maintenant établir que « toutes choses égales par ailleurs », une inversion associée au critère de type Δ_ϵ a tendance à entraîner une inversion de type D (moment d'inertie). À cet égard, il y a lieu de préciser l'expression « toutes choses égales par ailleurs ».

La construction d'un arbre des classifications dépend exclusivement de :

- la définition de l'indice δ de dissimilarité entre classes singletons (dont chaque classe contient exactement un objet) ;
- la formule de réactualisation des dissimilarités δ entre classes après la fusion d'une paire de classes.

Nous avons donné en (15) et (16) ci-dessus les formules de réactualisation respectivement associées au critère D de la variation de l'inertie expliquée et à celui Δ_ϵ de la vraisemblance du lien maximal. Pour ce dernier critère, l'expression (27) donne l'indice de dissimilarité entre classes singletons. Le point de départ de cette formule est celle (33), où x_i désigne la *luminance* du pixel i . D'autre part, pour le critère D , l'indice de dissimilarité entre classes singletons $\{i\}$ et $\{i'\}$ prend la forme :

$$\delta(\{i\}, \{i'\}) = \frac{1}{2}(x_i - x_{i'})^2 \quad (58)$$

L'aptitude au phénomène des inversions dépend de la formule de réactualisation. De sorte que nous allons nous abstraire des expressions particulières de l'indice de dissimilarité entre classes singletons dans le cas de chacun des deux critères D et Δ_ϵ . Plus précisément, nous allons considérer – comme dans la figure 2 – un système de trois classes A , B et C , pour lesquelles les dissimilarités mutuelles sont $\delta(A, B)$, $\delta(A, C)$ et $\delta(B, C)$ et faire agir chacune

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

des deux formules de réactualisations respectivement associées aux critères D et Δ_ϵ , en ne retenant que leur expression formelle. Ainsi, dans chacun des deux cas δ désignera l'indice de dissimilarité entre classes. Dans le cas de la formule associée au critère D , il y a inversion si et seulement si :

$$\delta(B, C) < \frac{(\alpha + \gamma) + (\beta + \gamma)}{\beta + \gamma} \delta(A, B) - \frac{\alpha + \gamma}{\beta + \gamma} \delta(A, C) \quad (59)$$

Pour le critère Δ_ϵ , il y a inversion si et seulement si :

$$\delta(B, C) < \left(\frac{\alpha + \beta}{\beta} \right)^{-\epsilon} \delta(A, B) = \left(1 + \frac{\alpha}{\beta} \right)^{-\epsilon} \delta(A, B) \quad (60)$$

Maintenant, conformément à (60) il ne peut y avoir inversion si $\delta(B, C) \geq (1 + \frac{\alpha}{\beta})^{-\epsilon} \delta(A, B)$. A fortiori, compte tenu de (51) où on suppose $\delta(A, C) > \delta(A, B)$ (puisque A et B sont reliés en premier) il ne peut y avoir inversion si $\delta(B, C) \geq (1 + \frac{\alpha}{\beta})^{-\epsilon} \delta(A, C)$. Dans le cas où $\epsilon = 1$ et $\alpha = \beta$, cette dernière inégalité est équivalente à

$$d_{min}(B, C) \geq \frac{1}{2} d_{min}(A, C)$$

Dans ces conditions, on obtient le résultat suivant :

COROLLAIRE 4.1. — Pour $\epsilon = 1$ et $\alpha = \beta$, si $d_{min}(B, C) \geq \frac{1}{2} d_{min}(A, C)$ l'inversion est impossible.

Maintenant, nous allons nous rendre compte que «le plus souvent» la deuxième inégalité (60) entraîne la première (59). Pour mieux le voir, supposons, sans trop restreindre la généralité, que $\alpha = \beta$. Dans ce cas, la deuxième inégalité se réduit à :

$$\delta(B, C) < 2^{-\epsilon} \delta(A, B) \quad (61)$$

et la première à :

$$\delta(B, C) < 2\delta(A, B) - \delta(A, C) \quad (62)$$

Dans ces conditions, si

$$2^{-\epsilon} \delta(A, B) \leq 2\delta(A, B) - \delta(A, C)$$

c'est-à-dire si

$$\delta(A, C) \leq (2 - 2^{-\epsilon}) \delta(A, B) \quad (63)$$

l'inégalité (61) implique celle (62), pourvu que $\delta(A, C)$ ne soit pas «trop grand» par rapport à $\delta(A, B)$ (à titre indicatif, pour $\epsilon = 1$, le coefficient multiplicatif est de 1.5 et pour $\epsilon = 0.5$, il est de 1.3). Dans ces conditions, on peut énoncer :

THÉORÈME 4.1. — Pour $\alpha = \beta$ et $\delta(A, C) < (2 - 2^{-\epsilon}) \delta(A, B)$, une inversion associée à une formule de réactualisation de type Δ_ϵ implique une inversion associée à une formule de réactualisation de type D .

En fait et pour terminer, signalons qu'intuitivement, on peut s'attendre que, compte tenu de ce que la valeur de $\delta(B, C)$ soit la plus petite parmi $\delta(A, B)$, $\delta(A, C)$ et $\delta(B, C)$ (cf. (45)), $\delta(A, C)$ ne peut pas être « trop grand » par rapport à $\delta(A, B)$ (voir Figure 2).

4.3. Cas de la forme contiguë Δ_ϵ^0 du critère de la vraisemblance du lien maximal

C'est cette forme du critère [cf. (38)] qui a permis d'obtenir les résultats les plus performants et cela aussi bien du point de vue de la qualité que de l'efficacité.

Dans une telle situation $\Delta_\epsilon^0(B, C)$ ne se trouve pas défini puisque $a(B, C) = 0$ [cf. (35)]. On a

$$\Delta_\epsilon^0(A, B) < \Delta_\epsilon^0(A, C) \quad (64)$$

D'autre part, les arêtes reliant – pour le graphe concerné – C à $A \cup B$ sont ceux reliant C à A , puisque B et C ne sont pas contigus. Dans ces conditions et par nature :

$$\Delta_\epsilon^0(C, A \cup B) = \Delta_\epsilon^0(C, A) \quad (65)$$

Il en résulte le :

THÉOREME 4.2. — Δ_ϵ^0 ne présente pas d'inversion.

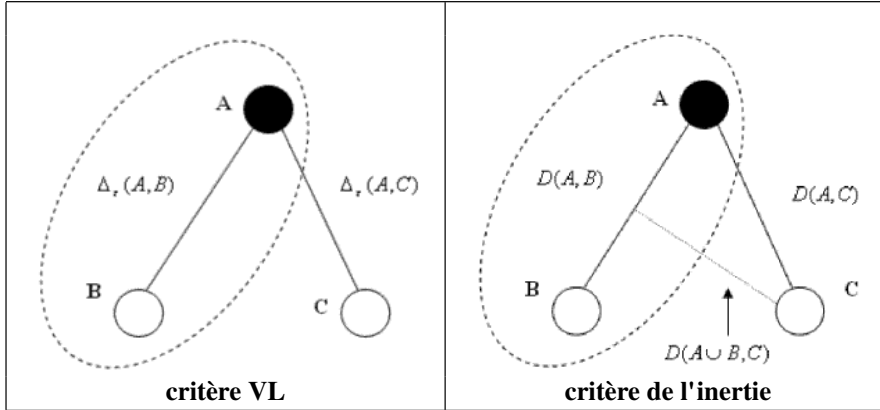


FIG 3. — Inversion.

Les propriétés que nous venons d'établir ont essentiellement un intérêt mathématique et algorithmique. Comme on verra ci-dessous, les classes qu'on obtient dépendent surtout de la nature du critère utilisé.

5. Données traitées et résultats obtenus

La *CAHCVR* a été appliquée au problème de la segmentation d'image. Le but d'une telle segmentation est l'identification des limites entre les différents objets qui constituent l'image. La notion de contiguïté spatiale est ici naturelle puisque les pixels sont situés géographiquement sur un plan. La classification sous cette contrainte de contiguïté permettra d'obtenir des classes connexes. Plus précisément, relativement à la classification hiérarchique, la partition déterminant une segmentation sera obtenue en choisissant un niveau « pertinent » de l'arbre des classifications.

5.1. Comparaison des résultats de *CAHCVR* paramétré

Deux types de paramétrage sont considérés pour *CAHCVR*. Le premier est de nature algorithmique et le second, statistique. Le premier comprend d'une part, le choix de la contrainte de contiguïté et d'autre part, la version binaire ou multiple de l'agrégation (cf. paragraphe 2). Le paramétrage statistique comprend le choix du critère δ d'association entre classes. Il peut s'agir de l'indice D de Ward ou de la famille paramétrée du critère $\Delta(\epsilon, \pi)$ de la vraisemblance du lien maximal (cf. paragraphe 3).

Deux images satellites ¹ ont été traitées. La première est appelée sat3, elle comprend $n = 758 \times 419$ pixels (cf. Fig. 5). La seconde, appelée sat2 (cf. Fig. 13) comprend $n = 459 \times 686$ pixels. Ici, un seul attribut est considéré pour la segmentation. Il s'agit de la *luminance* en niveaux de gris, codée sur l'intervalle $[0, 255]$.

Les meilleurs résultats ont été obtenus pour $\epsilon = 0.5$ et la contrainte de contiguïté correspondante au graphe G_4 et notée *Cont2*. Ces deux derniers paramètres étant fixés, différentes valeurs de π ont été essayées : $\pi = 0.4$ (Fig. 8, Fig. 11 et Fig. 15), $\pi = 0.45$ (Fig. 7) et $\pi = 0.5$ (Fig. 9).

Cherchons maintenant à déceler quelques différences visuelles. Dans l'image sat3 (Fig. 6), pour une segmentation en 32 classes, les zones 1 et 3 sont mieux définies par $\Delta(\epsilon, \pi)$ (Fig. 9) que par D (Fig. 10). À l'inverse, les zones 2 et 4 sont plus clairement identifiées par D . L'accroissement du nombre de classes (700 ici) montre une dispersion plus accentuée avec D (Fig. 12) qu'avec $\Delta(\epsilon, \pi)$ (Fig. 11).

Pour ce qui est de l'image sat2 (Fig. 14), les zones 3, 4 et 5 sont bien identifiées avec D (Fig. 16) ; alors que les zones 3 et 4 se retrouvent dans la même classe avec Δ_ϵ (Fig. 15). Cependant, on a une meilleure définition de la zone 1 avec $\Delta(\epsilon, \pi)$ (Fig. 15).

Une comparaison précise en termes de la qualité des résultats obtenus par D et Δ_ϵ n'est pas aisée. En effet, nous n'avons pas pour les images testées (qui sont « naturelles ») une claire connaissance *a priori* sur les segments qu'il y a lieu de reconnaître. Une telle connaissance serait donnée par un étiquetage systématique de tous les pixels par rapport aux différents segments. C'est

1. Nous sommes redevables à l'IGN qui nous a transmis ces données image.

ce que nous avons tenté en simulant des images structurées telles que celle de la figure 17, où la *luminance* d'un pixel donné est la réalisation d'une variable aléatoire entière dont le support est $[0, 255]$ et dont la loi dépend du segment auquel il appartient. Tous les segments définis dans cette image artificielle ont été parfaitement identifiés, aussi bien avec le critère D qu'avec celui $\Delta(\epsilon, \pi)$. Bien que très encourageantes, ces premières expériences doivent être poursuivies pour inférer intuitivement la différence de type de résultat qu'on obtient selon le critère utilisé.

5.2. Comportement de *CAHCVR* en termes de complexité

Le graphique suivant montre le comportement de *CAHCVR* en termes de complexité. C'est le critère Δ_ϵ qui est utilisé ici; cependant, le comportement du critère de Ward est parfaitement similaire. Les résultats théoriques mentionnés ci-dessus (dans l'introduction et le paragraphe 2) quant au caractère linéaire de la complexité [2, 24], sont parfaitement vérifiés. L'axe des abscisses indique le nombre de pixels et celui des ordonnées, la moyenne en secondes du temps de calcul. D'autre part, cette expérimentation a pu montrer que l'accélération est d'autant plus grande que *Cont2* (issu du graphe G_4) est considéré, préférentiellement à *Cont1* (issu du graphe G_8) et qu'on utilise l'agrégation multiple plutôt que celle binaire. Ainsi, l'agrégation multiple et la contrainte de contiguïté *Cont2* permettent le meilleur temps de traitement. Relativement à l'agrégation binaire, celle multiple permet approximativement 50% de gain en temps calcul. D'autre part, relativement à *Cont1* (G_8), *Cont2* (G_4) permet un gain de temps de l'ordre de 70%. Surtout, et cela était imprévisible, *Cont2* produit des résultats de qualité notablement meilleure que ceux obtenus avec *Cont1*. Clairement, cette qualité dépend d'un ajustement des valeurs de ϵ et de π .

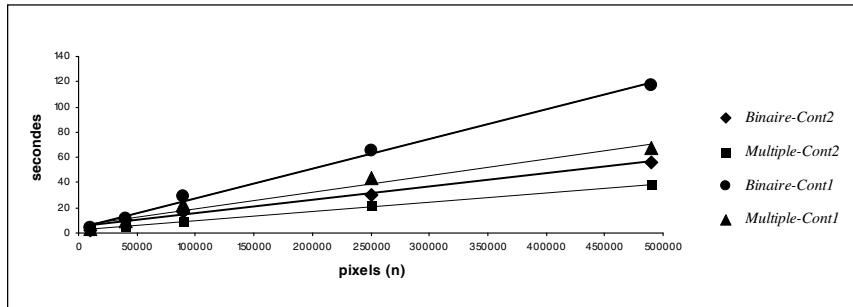


FIG 4. — Temps de traitement en secondes.

6. Conclusion et Perspectives

L'objectif de cette étude a consisté à asseoir la méthodologie *CAHCVR* (Classification Ascendante Hiérarchique sous Contrainte et par agrégation de

Voisins Réciproques) et cela par rapport à deux types de critères. Il s'agit d'une part du critère de Ward défini par un moment d'inertie et d'autre part, du critère paramétré de la vraisemblance du lien maximal.

L'application en segmentation d'image présente un intérêt spécifique. Elle a très fortement motivé l'étude et a permis la validation de l'approche et des résultats. Il s'agissait de la description de pixels par une seule variable, la *luminance*. L'extension au cas de la description de blocs de pixels par plusieurs variables doit être testée expérimentalement. L'approche est en effet très générale et intervient dès lors qu'une contrainte de contiguïté exogène à la description statistique est posée.

Un intérêt majeur de l'algorithmique est son caractère linéaire, en moyenne et dans un certain sens. C'est pour cette raison que nous avons voulu la situer dans le contexte général des algorithmes «accélérés» de la *CAH* (cf. paragraphe 2). D'autre part, l'étude expérimentale, établie dans le cas le plus complexe du graphe de contiguïté G_8 , a parfaitement conforté nos résultats théoriques [2, 24]. Compte tenu de la consistance de leur élaboration, nous nous gardons la possibilité dans le futur de proposer un article développant ces aspects théoriques.

Un des problèmes fondamentaux consiste en la détermination de niveaux de l'arbre des classifications qu'il s'agit de retenir pour déterminer des partitions «significatives», chacune ayant un nombre «naturel» de classes. Ce problème a longtemps occupé notre recherche. Nous avons mis au point une technique fondée sur l'élaboration combinatoire et statistique d'un critère d'adéquation entre une partition et une similarité [16, 17]. Une telle technique est bien implantée dans *CHAVLH* [26, 34], ainsi d'ailleurs que dans le programme *LLAhclust* (Likelihood Linkage Analysis hierarchical clustering) qui en est issu et qui est implanté dans l'environnement logiciel dit **R** (lien : <http://cran.r-project.org/src/contrib/Descriptions/LLAhclust.html>). La version classique de ce critère est quadratique par rapport au nombre d'observations. Cependant, une dernière version, valide dans le cas de données numériques ou nominales, est linéaire par rapport au nombre d'observations [27]. Il serait intéressant de pouvoir l'implanter au niveau de *CAHCVR*. Relativement à une partition récoltée à un niveau donné de l'arbre des classifications, ce critère permettrait, moyennant un remaniement algorithmique linéaire (e.g. deux pas, comprenant un centrage et une allocation contrainte par $G^{(0)}$ d'un K-means), de l'améliorer encore.

Maintenant, il existe bien des méthodes relativement au problème fondamental de la segmentation d'images ; par exemple [32, 36]. Il importe dans ces conditions d'y situer l'intérêt de notre approche. Deux aspects peuvent ressortir de notre contribution ; le premier est défini par la construction méthodologique relativement au domaine de la classification automatique et le second correspond à l'efficacité de la méthode obtenue en segmentation d'image. Ce dernier aspect sera développé à la faveur de nouvelles expérimentations.

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

Images originale et segmentées (critère (VL ou D), nombre de classes



FIG 5. — Image sat3.



FIG 6. — Zones de l'image.

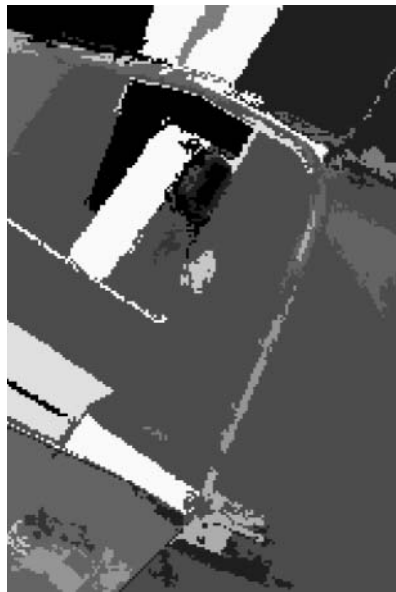


FIG 7. — (VL, 32).

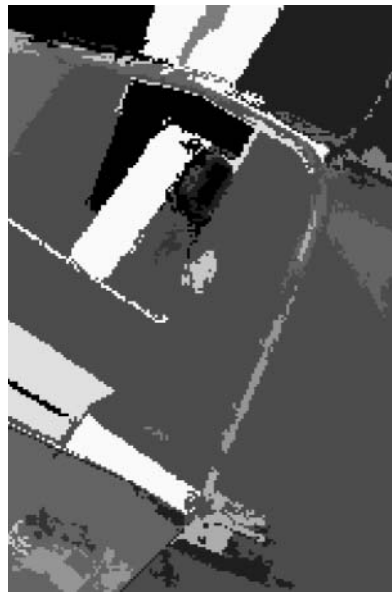


FIG 8. — (VL, 40).

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

Images originale et segmentées (critère (VL ou D), nombre de classes (suite))

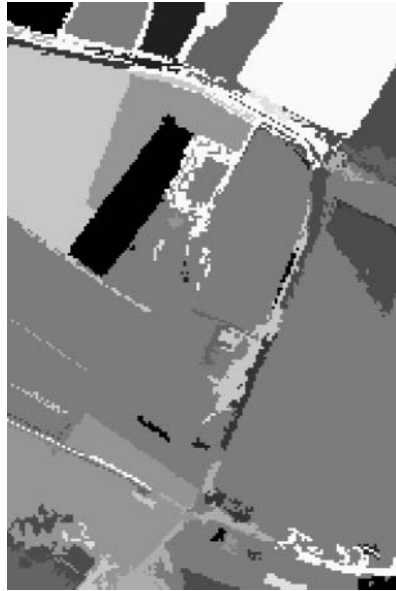


FIG 9. — (VL, 32).



FIG 10. — (D, 32).

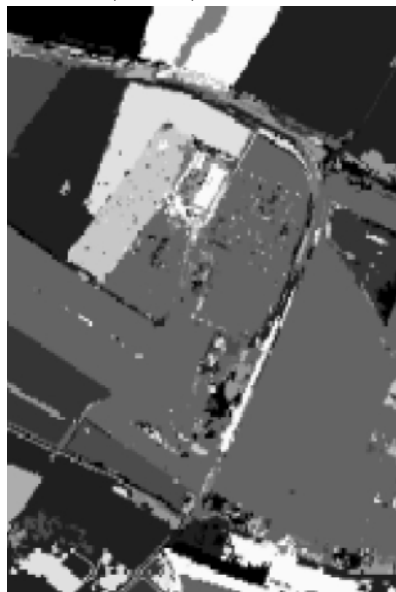


FIG 11. — (VL, 700).



FIG 12. — (D, 700).

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

Images originale et segmentées (critère (VL ou D), nombre de classes (suite))



FIG 13. — Image sat2.

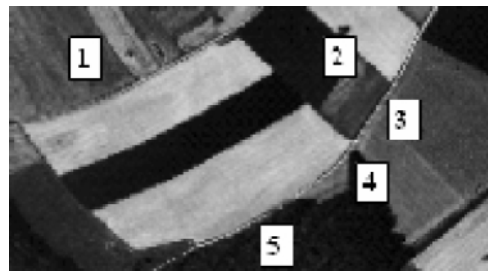


FIG 14. — Zones de l'image.

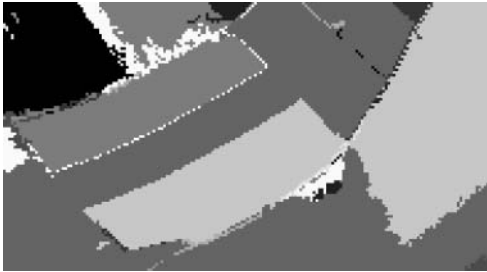


FIG 15. — (VL, 38).

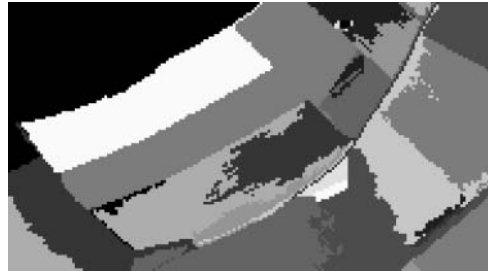


FIG 16. — (D, 32).

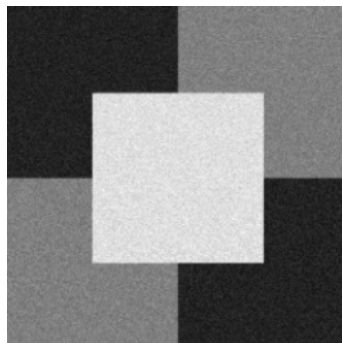


FIG 17.

Références

- [1] K. BACHAR. *Contributions en analyse factorielle et en classification ascendante hiérarchique sous contrainte de contiguïté. Applications à la segmentation d'images*. PhD thesis, Université de Rennes 1, Décembre 1994.
- [2] K. BACHAR and I.C. LERMAN. Statistical conditions for a linear complexity for an algorithm of hierarchical classification under constraint of contiguity. In Bock H.-H. Rizzi A., Vichi M., editor, *Advances in Data Science and Classification/IFCS'98*, pages 131-136. Springer-Verlag, 1998.
- [3] K. BACHAR and I.C. LERMAN. Étude d'un comportement paramétré de CAHCVR sur des données réelles en imagerie numérique. In Melfi G. Dodge Y., editor, *Méthodes et Perspectives en Classification*, pages 63-66. Presses Académique Neuchâtel, 2003.
- [4] K. BACHAR and I.C. LERMAN. Fixing parameters in the constrained hierarchical classification method : Application to digital image segmentation. In Banks D. et al., editor, *Classification Clustering and Data Mining Applications*, pages 85-94. Springer, 2004.
- [5] J.P. BARTHELEMY and A. GUENOCHÉ. *Trees and Proximity Representations*. J. Wiley, 1991.
- [6] J.P. BENZECRI. Construction d'une classification ascendante hiérarchique par la recherche en chaîne des voisins réciproques. *Les Cahiers de l'Analyse des Données*, (VII) : 209-218, 1982.
- [7] M. BRUYNNOOGHE. Classification ascendante hiérarchique des grands ensembles de données; un algorithme rapide fondé sur la construction des voisinages réductibles. *Les Cahiers de l'Analyse des Données*, (III) : 7-33, 1978.
- [8] M. BRUYNNOOGHE. *Nouveaux algorithmes en classification automatique applicables aux très grands ensembles de données, rencontrés en traitement d'images et en reconnaissance des formes*. PhD thesis, Thèse d'État, Université de Paris 6, janvier 1989.
- [9] C. DE RAHM. La classification hiérarchique ascendante selon la méthode des voisins réciproques. *Les Cahiers de l'Analyse des Données*, (V, n°2) : 135-144, 1980.
- [10] A.D. GORDON. A survey of constrained classification. *Computational Statistics and Data Analysis*, (1) : 17-29, 1996.
- [11] P. HANSEN, B. JAUMARD, C. MEYER, B. SIMEONE, and V. DORING. Maximum split clustering under connectivity constraints. *Journal of Classification*, (20) : 143-180, 2003.
- [12] J. JUAN. Programme de classification hiérarchique par l'algorithme de recherche en chaîne des voisins réciproques. *Les Cahiers de l'Analyse des Données*, (VII) : 219-225, 1982.
- [13] G.N. LANCE and W.T. WILLIAMS. A general theory of classification sorting strategies : clustering systems. *Computer Journal*, (10) : 271-277, 1967.
- [14] L. LEBART. Programme d'agrégation avec contraintes. *Les Cahiers de l'Analyse des Données*, (III(3)) : 275-287, 1978.
- [15] I.C. LERMAN. Sur l'analyse des données préalable à une classification automatique. proposition d'une nouvelle mesure de similarité. *Mathématiques et Sciences Humaines*, (32) : 5-15, 1970.
- [16] I.C. LERMAN. *Classification et analyse ordinale des données*. Dunod, 1981.

- [17] I.C. LERMAN. Sur la signification des classes issues d'une classification automatique. In J. Felsenstein, editor, *Numerical Taxonomy*, pages 179-198. Springer-Verlag, 1983.
- [18] I.C. LERMAN. Construction d'un indice de similarité entre objets décrits par des variables d'un type quelconque. application au problème du consensus en classification. *Revue de Statistique Appliquée*, (XXXV(2)) : 39-60, 1987.
- [19] I.C. LERMAN. Formules de réactualisation en cas d'agrégations multiples. *RAIRO, série R.O.*, (vol 23, n°2) : 151-163, 1989.
- [20] I.C. LERMAN. Foundations of the likelihood linkage analysis(IIa) classification method. *Applied Stochastic Models and Data Analysis*, (vol 7,1) : 6376, march 1991.
- [21] I.C. LERMAN. Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles I. *Revue Mathématique Informatique et Sciences Humaines*, (118) : 35-52, 1992.
- [22] I.C. LERMAN. Conception et analyse de la forme limite d'une famille de coefficients statistiques d'association entre variables relationnelles II. *Revue Mathématique Informatique et Sciences Humaines*, (119) : 75-100, 1992.
- [23] I.C. LERMAN and K. BACHAR. Agrégations multiples et contraintes de contiguïté dans la classification ascendante hiérarchique utilisant les voisins réciproques et le critère de la vraisemblance des liens. In *8-èmes Rencontres de la Société Francophone de Classification*, pages 232-237. Université Pointe-à-Pitre, 2001.
- [24] I.C. LERMAN and K. BACHAR. Construction et justification d'une méthode de classification ascendante hiérarchique accélérée fondée sur le critère de la vraisemblance du lien maximal en cas de données de contiguïté. Application en imagerie numérique. Rapport de Recherche 1616, IRISA, avril 2004.
- [25] I.C. LERMAN and Ph. PETER. Indice probabiliste de vraisemblance du lien entre objets quelconques ; analyse comparative entre deux approches. *Revue de Statistique Appliquée*, (LI(1)) : 5-35, 2003.
- [26] I.C. LERMAN, Ph. PETER, and H. LEREDDE. Principes et calculs de la méthode implantée dans le programme CHAVL 1-ère partie. *La Revue de Modulad*, (12) : 33-70, Décembre 1993.
- [27] I.C. LERMAN, J.F. PINTO DA COSTA, and H. SILVA. Validation of very large data sets clustering by means of a nonparametric criterion. In A. Sokolowski K. Jajuga and H.-H. Bock, editors, *Classification, Clustering and Data Analysis, Recent Advances and Applications*, pages 147-157. Springer-Verlag, 2002.
- [28] P. MONESTIEZ. Méthode de classification automatique sous contraintes spatiales. *Statistique et Analyse des Données*, (3) : 75-84, 1977.
- [29] F. MURTAGH. A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal*, (28) : 82-88, 1985.
- [30] F.C. NICOLAÛ and H. BACELAR-NICOLAÛ. Some trends in the classification of variables. In *Data Science, Classification and Related Methods, IFCS'96*, pages 89-98. Springer-Verlag, 1996.
- [31] M. OUALI-ALLAH. *Analyse en préordonnance des données qualitatives. Application aux données numériques et symboliques*. PhD thesis, Université de Rennes 1, Décembre 1991.
- [32] E.J. PAUWELS and G. FREDERIX. Finding salient regions in images. *Computer Vision and Image Understanding*, (75) : 73-85, 1999.
- [33] C. PERRUCHET. Constrained agglomerative hierarchical classification. *Pattern Recognition*, (16(2)) : 213-217, 1983.

CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

- [34] P. PETER, H. LEREDDE, and I.C. LERMAN. Notice du programme CHAVLH (Classification Hiérarchique par Analyse de la Vraisemblance des Liens en cas de variables Hétérogènes). Dépôt APP (Agence pour la Protection des Programmes) IDDN.FR.001.240016.000.S.P.2006.000.20700, Université de Rennes 1, Décembre 2005.
- [35] A. PRODHOMME. *Indices d'explication des classes obtenues par une méthode de classification hiérarchique respectant la contrainte de contiguïté spatiale. Application à la viticulture Girondine et à la construction de logements dans les Bouches-du Rhône*. PhD thesis, Université de Rennes 1, Décembre 1980.
- [36] M. TABB and N. AHUJIA. Multiscale image segmentation by integrated edge and region detection. *IEEE Trans. Image Processing*, (6) : 642-655, 1997.
- [37] J.H. WARD. Hierarchical grouping to optimise an objective function. *Journal of the American Statistical Association*, (58) : 238-244, 1963.