

# Classification robuste sur l'espace des matrices de covariance: Application à l'imagerie polarimétrique radar

Ioana ILEA<sup>1,2</sup>, Lionel BOMBRUN<sup>1</sup>, Christian GERMAIN<sup>1</sup>, Romulus TEREDES<sup>2</sup>, Monica BORDA<sup>2</sup>

<sup>1</sup>Université de Bordeaux, Laboratoire IMS (CNRS UMR 5218), Groupe Signal et Image.

<sup>2</sup>Technical University of Cluj-Napoca.

{ioana.ilea, lionel.bombrun, christian.germain}@ims-bordeaux.fr  
{romulus.terebes, monica.borda}@com.utcluj.ro

**Résumé** – Dans cet article, nous proposons une méthode robuste de classification d'images en s'appuyant sur la théorie de la détection. Tout d'abord, nous introduisons un nouveau test d'hypothèses basé sur la distance géodésique et sur l'estimateur du point fixe de la matrice de covariance pour les distributions Gaussiennes multivariées. Ensuite, nous analysons la robustesse au bruit de ce classificateur sur des données de synthèse. Enfin, nous validons la méthode proposée sur des données Polarimétrique Radar à Synthèse d'Ouverture simulées pour la classification de parcelles forestières.

**Abstract** – This paper introduces a robust image classification algorithm, based on the detection theory. First, we introduce a new hypothesis test based on the geodesic distance and on the fixed point estimator of the covariance matrix for multivariate Gaussian distributions. Next, we analyze its noise robustness for simulated data. Finally, the proposed method is validated for the classification of Polarimetric Synthetic Aperture Radar (PolSAR) images of forest stands.

## 1 Introduction

Dans la littérature, les représentations multi-échelles ont été largement plébiscitées pour de nombreuses applications en traitement d'images telles que le filtrage [1], la segmentation [2] ou la classification [3]. En classification, le principe général réside dans la modélisation des coefficients de détails par des densités de probabilités. Dans ce contexte, de nombreux modèles univariés ont été proposés, tels que la distribution Gaussienne généralisée [3]. Cependant, ces modèles univariés ne permettent pas d'exploiter pleinement les dépendances présentes dans les images (dépendances spatiales, spectrales, ...). De ce fait, des familles de lois multivariées ont été considérées telles que les distributions elliptiques [4, 5], les copules [6, 7]. L'ensemble des estimés des paramètres de ces distributions (matrice de covariance, paramètre de forme, d'échelle, ...) représente la signature de l'image.

Une fois ces signatures extraites, une règle de décision basée sur une mesure similarité est mise en place pour réaliser la classification. Pour cela, la divergence de Kullback-Leibler (KL) ou la divergence de Jeffrey sont régulièrement choisies [8]. Plus récemment, Verdoolaege *et al.* ont proposé l'utilisation de la distance géodésique (GD) qui a montré des performances de classification supérieures vis à vis de la divergence de KL [5].

Afin de réguler la probabilité de fausse alarme du classifieur, il est possible de mettre en place un test d'hypothèses. En nous appuyant sur la théorie de la détection, nous proposons un algorithme robuste de classification d'images à partir de la sta-

tistique des mesures de similarité entre deux modèles paramétriques. Inspiré par les travaux sur la divergence de KL [9] et sur la famille des  $(h, \phi)$ -divergences [10, 11], nous présentons ici un nouveau test d'hypothèses basé sur la GD. Nous présentons également son extension au cas des estimateurs robustes [12].

Cet article est organisé comme suit : la partie 2 présente le test d'hypothèses basé sur la GD. Une application au cas particulier des distributions Gaussiennes multivariées est ensuite proposée. Afin de rendre robuste le classifieur, nous nous intéressons à la théorie des M-estimateurs et notamment l'estimateur du point fixe (FP) de la matrice de covariance. Le classifieur ainsi obtenu est ensuite validé dans la partie 3, sur des images radar polarimétriques (PolSAR) simulées, pour la différenciation de classes d'âges de forêt de pin.

## 2 Mise en place d'un test d'hypothèses pour la classification robuste

### 2.1 Présentation du test d'hypothèses

Soient  $X_1 = (\mathbf{x}_1^1, \dots, \mathbf{x}_m^1)$  et  $X_2 = (\mathbf{x}_1^2, \dots, \mathbf{x}_n^2)$  deux ensembles de  $m$  et  $n$  vecteurs aléatoires  $\mathbf{x}$ , indépendants et identiquement distribués (iid) selon les modèles paramétriques  $p(\mathbf{x}|\theta_1)$  et  $p(\mathbf{x}|\theta_2)$ . Soient  $\hat{\theta}_1$  et  $\hat{\theta}_2$  les estimateurs au sens du maximum de vraisemblance (ML) des paramètres  $\theta_1$  et  $\theta_2$  calculés pour ces ensembles.

L'objectif d'un algorithme de classification est de déterminer

si les ensembles  $X_1$  et  $X_2$  sont obtenus à partir du même modèle paramétrique. Pour cela, nous pouvons considérer le test d'hypothèses suivant :

$$\begin{cases} H_0 : \theta_1 = \theta_2, \\ H_1 : \theta_1 \neq \theta_2. \end{cases} \quad (1)$$

Sous l'hypothèse nulle  $H_0$ , Salicru *et al.* ont démontré dans [10, 11] la propriété suivante :

$$S(\hat{\theta}_1, \hat{\theta}_2) = \frac{2mnv}{m+n} D(\hat{\theta}_1, \hat{\theta}_2) \xrightarrow[n, m \rightarrow \infty]{} \chi_M^2, \quad (2)$$

où  $S(\hat{\theta}_1, \hat{\theta}_2)$  est la statistique du test,  $M$  le nombre de degrés de liberté (égal à la dimension de l'espace paramétrique  $\Theta$ ).  $v$  est une constante qui dépend de la mesure de similarité  $D$  (par exemple  $v = 1$  pour la divergence de Kullback-Leibler).

Ce test d'hypothèse a été d'abord introduit pour la divergence de KL [9] et ensuite généralisé pour la famille des  $(h, \phi)$  divergences [10, 11]. Dans cet article, nous proposons d'étendre ce test d'hypothèses au cas où la mesure de similarité est la distance géodésique (GD). Comme sous l'hypothèse  $H_0$  la divergence de KL est équivalente à la moitié du carré de la GD, il en résulte que la statistique  $S_{GD}(\hat{\theta}_1, \hat{\theta}_2) = \frac{mn}{m+n} GD^2(\hat{\theta}_1, \hat{\theta}_2)$  converge asymptotiquement vers la loi du  $\chi^2$  avec  $M$  degrés de liberté pour des tailles de population  $m$  et  $n$  suffisamment grandes. Il est intéressant de noter que sous l'hypothèse  $H_0$ , la distribution de la statistique  $S_{GD}$  ne dépend ni de  $\theta_1$  ni de  $\theta_2$  permettant ainsi de régler la probabilité de fausse alarme plus aisément.

Dans la suite, nous considérons une application de ce test d'hypothèses au cas particulier des distributions Gaussiennes multivariées (MGD) de moyenne nulle. Dans ce cas, la GD est définie comme :

$$GD(\hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2) = \left[ \frac{1}{2} \sum_i (\ln \lambda_i)^2 \right]^{\frac{1}{2}}, \quad (3)$$

où  $\hat{\mathbf{M}}_1$  et  $\hat{\mathbf{M}}_2$  sont les estimateurs empiriques (SCM) des deux matrices de covariance  $\mathbf{M}_1$  et  $\mathbf{M}_2$  et  $\lambda_i, i = 1 \dots d$  sont les valeurs propres de  $\hat{\mathbf{M}}_2^{-1} \hat{\mathbf{M}}_1$ . Par conséquent, l'hypothèse nulle  $\mathbf{M}_1 = \mathbf{M}_2$  peut être rejetée à un seuil  $\alpha$  si

$$Pr(\chi_M^2 > S_{GD}(\hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2)) \leq \alpha,$$

sachant que le nombre de degrés de liberté  $M$  est égal à  $d(d+1)/2$ , où  $d$  est la dimension de la matrice de covariance.

Afin de proposer un classifieur robuste, il convient de généraliser ce test d'hypothèses au cas des estimateurs robustes des matrices de covariance, ce qui est l'objet de la prochaine sous-partie.

## 2.2 Extension aux estimateurs robustes

La classe de M-estimateurs a été introduite dans le contexte de l'estimation robuste, afin d'identifier la présence des données aberrantes dans les jeux de données, ou d'erreurs dans le modèle. Pour des observations de moyenne nulle, le M-estimateur

de la matrice de covariance est défini [12, 13] comme la solution de :

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^N u(\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T, \quad (4)$$

où  $u(\cdot)$  est une fonction positive qui pondère chacune des observations  $\mathbf{x}_i$ . Cette fonction de pondération  $u(\cdot)$  tend vers zéro afin de garantir que les valeurs aberrantes ont une contribution faible à l'estimation de la matrice de covariance par rapport aux autres observations. L'estimateur empirique (SCM) de la matrice de covariance attribue quant à lui une pondération uniforme (*i.e.*  $u(\cdot) = 1$ ) à chaque observation, confirmant ainsi son caractère non robuste.

Les M-estimateurs ont été très largement étudiés dans la littérature. Ils peuvent être vus comme une généralisation de l'estimateur de la matrice de covariance par maximum de vraisemblance pour des distributions elliptiques. Si  $u(t) = 1/t$  on obtient un cas particulier de cet estimateur : l'estimateur du point fixe (FP), donné par [14, 15] :

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^T}{\mathbf{x}_i^T \hat{\mathbf{M}}^{-1} \mathbf{x}_i}. \quad (5)$$

Si  $\hat{\mathbf{M}}$  est une solution de (5), alors  $c\hat{\mathbf{M}}$  est aussi une solution de cette équation, quelque soit le paramètre d'échelle  $c \neq 0$ . Afin de garantir l'unicité, la matrice de covariance est normalisée. En pratique, une contrainte sur la trace de l'estimée est considérée. Cet estimateur FP peut être vu comme l'estimateur ML de la matrice de covariance normalisée pour un modèle de mélange d'échelle de Gaussiennes<sup>1</sup>, où les multiplieurs  $\tau_i$  sont des paramètres déterministes et inconnus [14].

Les propriétés de cet estimateur FP ont été analysées dans [15]. En particulier, cet estimateur FP est la solution unique, non biaisée et consistante de (5). En outre, l'estimateur FP suit une distribution de Wishart à  $(d/(d+1))N$  degrés de liberté.

## 2.3 Robustesse au bruit

Dans cette partie, nous comparons les performances des estimateurs SCM et FP en terme de robustesse au bruit pour la statistique  $S_{GD}$  introduite dans (2). Soient  $X_1$  et  $X_2$  deux ensembles de  $128^2$  réalisations iid d'une loi MGD de moyenne nulle et de matrice de covariance  $\mathbf{M} = (M_{i,j})$ , avec  $M_{i,j} = \rho^{|i-j|}$ . L'ensemble  $X_2$  est ensuite bruité par un bruit blanc Gaussien additif de matrice de covariance  $\sigma^2 \mathbf{I}_d$ , où  $\sigma^2$  est la variance du bruit et  $\mathbf{I}_d$  est la matrice identité. Pour cette expérience, le risque  $\alpha$  est fixé à 0.05 et plusieurs matrices de covariance  $\mathbf{M}$  sont testées ( $\rho = 0.25, 0.5$  et  $0.75$ ). La Figure 1 présente le taux de classification en fonction de la variance du bruit. Les courbes en pointillés et en trait plein correspondent respectivement aux estimateurs SCM et FP. Ici,  $10^3$  tirages aléatoires sont considérés pour estimer les performances de classification moyennes. Comme observé sur la Figure 1,

1. Un modèle de mélange d'échelle de Gaussiennes admet la représentation stochastique  $\mathbf{x} = \sqrt{\tau} \mathbf{z}$ , où le multiplieur  $\tau$  est une variable aléatoire ( $\tau \in \mathbb{R}^+$ ) et  $\mathbf{z}$  est un vecteur Gaussien, de moyenne nulle et de matrice de covariance  $\mathbf{M}$ .

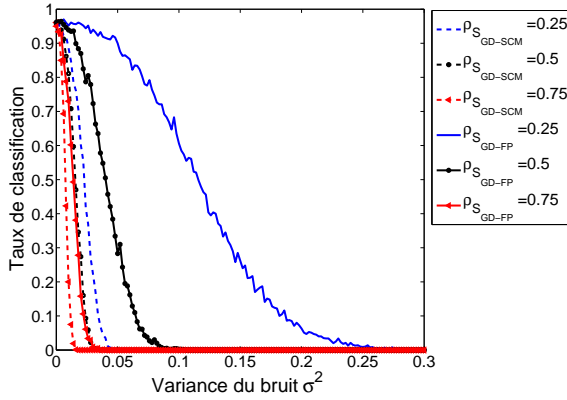


FIGURE 1 – Évolution des performances de classification en fonction de la variance du bruit  $\sigma^2$  pour les estimateurs SCM et FP.

l'estimateur FP présente de meilleures performances que l'estimateur SCM, confirmant son caractère robuste. Ce gain en performances est d'autant plus élevé que les matrices de covariances sont peu structurées ( $\rho$  faible).

### 3 Application à l'imagerie polarimétrique

Dans ce papier, nous proposons une application du classifieur proposé dans la partie 2 au cas l'imagerie polarimétrique radar (PolSAR) pour la différenciation de classes d'âges de forêts de pins. Pour cela, une base de données a été créée en utilisant le simulateur PolSARproSim [16]. Celle-ci se compose de 350 images PolSAR en bande L à 1.3 GHz, représentant des parcelles forestières de pins, qui sont groupées en 4 classes en fonction de l'âge de la parcelle : moins de 10 ans (Figure 2(a)), entre 10 et 20 ans (Figure 2(b)), entre 20 et 30 ans (Figure 2(c)), et plus de 30 ans (Figure 2(d)). La densité de la parcelle forestière est choisie pour respecter les caractéristiques typiques des forêts cultivées de pins maritimes en Aquitaine. La hauteur moyenne des arbres est calculée en utilisant le modèle théorique de Maugé [17] :  $H_{mean} = H_{max}(1 - 0.96^a)$ , où  $H_{max} = 30$  mètres est la hauteur maximale et  $a$  est l'âge du peuplement forestier. Les images sont simulées pour un capteur situé à 3500 mètres d'altitude avec un angle d'incidence égal à  $45^\circ$  et plusieurs résolutions spatiales sont considérées (0.5, 1, 2, 3 et 5 mètres).

D'un point de vue thématique, l'enjeu principal de cette étude est de déterminer la meilleure configuration d'acquisition pour la classification de forêt de pins. En d'autres termes, est-il préférable d'avoir une image SAR à très haute résolution spatiale avec un seul canal polarimétrique ou alors une image basse résolution contenant 2 ou 3 canaux polarimétriques ? A haute résolution spatiale, comme observé sur la Figure 2, il est possible de considérer l'existence d'une relation entre les variables de structure forestière (âge, diamètre de la couronne, hauteur des arbres, densité du peuplement, etc.) et l'information texturale

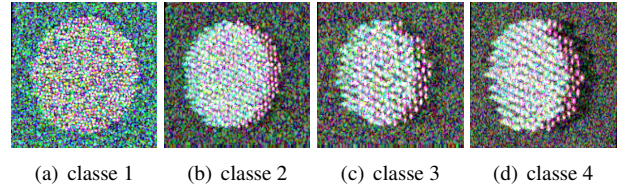


FIGURE 2 – Exemple d'images des forêts de pin en bande L, simulées avec le logiciel PolSARproSIM à une résolution d'un mètre.

présente dans les images. A basse résolution, l'information texturale n'est plus visible dans les images mais la diversité polarimétrique peut-être exploitée pour la classification.

Dans ce contexte, nous procédons à une méthode de validation croisée en séparant la base de données en deux : un ensemble d'apprentissage et un ensemble de test. La signature de chaque image dans les deux ensembles est calculée et utilisée pour classifier chacun des peuplements forestiers. Nous comparons ensuite différentes approches basées sur des descripteurs non paramétriques comme les matrices de co-occurrences (GLCM) [18] et des descripteurs paramétriques en utilisant l'approche présentée dans la partie 2.

L'image contenant le coefficient de rétrodiffusion en dB est quantifiée sur 32 niveaux de gris afin de calculer la GLCM et d'en extraire les descripteurs texturaux d'Haralick [18]. Les paramètres des GLCM (distance, orientation, choix des descripteurs) sont optimisés afin d'obtenir les meilleures performances de classification. Dans cette étude, les critères d'homogénéité, d'entropie, de corrélation et de moyenne sont calculés et moyennés dans le quatre directions afin de réduire l'influence de l'orientation des rangs du peuplements.

Pour les descripteurs paramétriques, nous considérons la distribution Gaussienne multivariée (MGD) avec l'estimateur FP de la matrice de covariance introduit dans la section 2.2. Plusieurs types de dépendances peuvent être considérés. Si un seul canal polarimétrique est à disposition, il est possible de modéliser les dépendances spatiales à travers une décomposition en ondelette (WT + S). Dans ce cas, pour chaque sous-bande, un ensemble de vecteurs de taille 9 contenant les coefficients en ondelette dans un voisinage spatial  $3 \times 3$  sont modélisés par une MGD. Lorsque plusieurs canaux polarimétriques sont disponibles, différentes dépendances peuvent être modélisée :

- la dépendance polarimétrique (MGD Polar) en modélisant directement les vecteurs composés des trois canaux polarimétriques.
- la dépendance polarimétrique sur les coefficients en ondelettes (MGD Polar + WT). Dans ce cas, une décomposition en ondelettes est effectuée pour chacun des canaux polarimétriques et le modèle MGD est utilisé pour modéliser ces dépendances inter-canaux.
- la dépendance polarimétrique et spatiale sur les coefficients en ondelettes (MGD Polar + WT + S). Un ensemble de vecteurs de taille 27 contenant les coefficients en ondelette dans un voisinage spatial  $3 \times 3$  pour chacun des canaux polarimétriques est modélisé par une MGD.

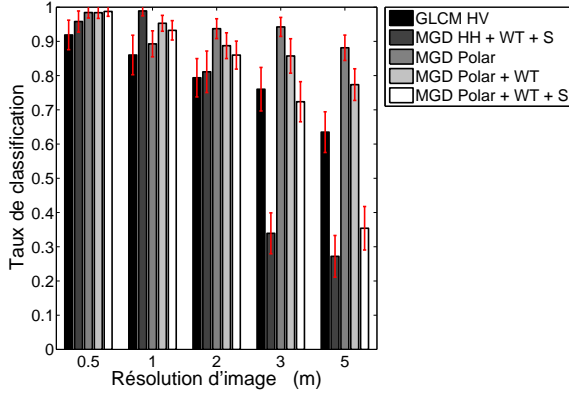


FIGURE 3 – Influence de la résolution sur les performances de classification.

Pour les modèles faisant intervenir une décomposition en ondelettes (WT), la mesure de similarité entre deux images  $I_1$  et  $I_2$  est obtenue par :

$$GD^2(I_1||I_2) = \sum_{s=1}^{N_{sc}} \sum_{o=1}^{N_{or}} GD^2(\hat{M}_{1,s,o}||\hat{M}_{2,s,o}), \quad (6)$$

où  $\hat{M}_{1,s,o}$  représente l'estimateur FP de la matrice de covariance  $M$  à l'échelle  $s$  et l'orientation  $o$  de l'image  $I_1$ .

La Figure 3 montre les performances de bonne classification en fonction de la résolution spatiale pour les différentes méthodes présentées ci-dessus. Pour les deux méthodes n'exploitant qu'un canal polarimétrique, seule la polarisation présentant les meilleurs résultats est affichée (canal HV pour les GLCM et canal HH pour MGD+WT+S). Aussi, on peut conclure que, pour cette application, il est préférable d'avoir une image radar à très haute résolution spatiale ayant une seule polarisation qu'une image contenant les 3 canaux polarimétrique à basse résolution. Un gain d'environ 8.7% est observé ( $95.8 \pm 3.1\%$  pour MGD HH + WT + S à 0.5 mètres versus  $88.1 \pm 3.7\%$  pour MGD Polar à 5 mètres).

## 4 Conclusion

Dans cet article, un algorithme robuste de classification sur l'espace des matrices de covariance est proposé. Celui-ci repose sur la mise en place d'un test d'hypothèses basé sur la distance géodésique et sur un estimateur robuste de la matrice de covariance : l'estimateur du point fixe. Une expérimentation sur données simulées a montré un gain significatif par rapport à l'estimateur empirique de la matrice de covariance sur des données bruitées. Enfin, une application à la classification d'images polarimétriques a été proposée validant ainsi le potentiel cette approche.

## 5 Remerciements

Les travaux présentés dans ce papier ont été financés par le ministère des Affaires étrangères et du Développement international, ainsi que par l'Agence exécutive pour l'enseignement supérieur, la recherche, le développement et l'innovation dans le cadre du projet numéro 32619VL (France) et PNII Capacitati 779/27.06.2014 (Roumanie).

## Références

- [1] D.L. Donoho, "Denosing by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [2] J.-F. Aujol, G. Aubert, and L. Blanc-Feraud, "Wavelet-based level set evolution for classification of textured images," *IEEE Trans. Image Proces.*, vol. 12, no. 12, pp. 1634–1641, 2003.
- [3] M.N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Proces.*, vol. 11, pp. 146–158, 2002.
- [4] L. Bombrun, Y. Berthoumiou, N.-E. Lasmar, and G. Verdoolaege, "Multivariate texture retrieval using the geodesic distance between elliptically distributed random variables," in *Proc. IEEE ICIP*, 2011, pp. 3637–3640.
- [5] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *J. Math. Imag. Vision*, vol. 43, no. 3, pp. 180–193, 2012.
- [6] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Proces.*, vol. 19, no. 1, pp. 241–253, 2010.
- [7] Y. Stitou, N.-E. Lasmar, and Y. Berthoumiou, "Copulas based multivariate gamma modeling for texture classification," in *Proc. IEEE ICASSP*, 2009, pp. 1045–1048.
- [8] S. Kullback and R.A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [9] M. Kupperman, *Further applications of information theory to multivariate analysis and statistical inference*, Ph.D. thesis, George Washington University, 1957.
- [10] M. Salicru, D. Morales, M.L. Menendez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *Journal of Multivariate Analysis*, vol. 51, no. 2, pp. 372–391, 1994.
- [11] A.D.C. Nascimento, R.J. Cintra, and A.C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, pp. 373–385, 2010.
- [12] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.
- [13] D. E. Tyler, "A distribution-free  $m$ -estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.
- [14] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [15] F. Pascal, P. Forster, J. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans. Signal Proces.*, vol. 56, no. 6, pp. 2206–2217, June 2008.
- [16] M.L. Williams, "PolSARproSim : A coherent, Polarimetric SAR Simulation of forests for PolSARPro. Design document and algorithm specification," 2006.
- [17] J.P. Mauge, *Le pin maritime premier résineux de France*, Centre de productivité et d'action forestière d'Aquitaine, Institut pour le développement forestier, Paris, 1987.
- [18] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.