

# Classification d'images texturées basée sur K-barycentres : meilleure gestion de la diversité intra-classe

Aurélien SCHUTZ, Lionel BOMBRUN, Yannick BERTHOUMIEU

Université de Bordeaux, ENSEIRB-Matmeca, Laboratoire IMS, Groupe Signal et Image  
350 Cours de la Libération, 33405 Talence, France

{aurelien.schutz, lionel.bombrun, yannick.berthoumieu}@ims-bordeaux.fr

**Résumé** – Les images de texture naturelle présentent de fortes variabilités par classe induites par les différentes conditions d'acquisition (éclairage de la scène, différence de perspective, ...). Pour gérer l'effet négatif de cette variabilité sur les performances de classification, un nouvel algorithme de classification supervisée basée sur une forme K-barycentrique est proposé. Une étude comparative entre plusieurs algorithmes de classification supervisée sur les bases de données d'images texturées VisTex et Brodatz est réalisée et montre que la classification K-barycentres propose un bon compromis entre performances de classification et coût calculatoire.

**Abstract** – Natural texture images exhibit a high intra-class diversity due to different acquisition conditions (scene enlightenment, perspective angle, ...). To handle with the diversity, a new supervised classification algorithm based on a parametric formalism is introduced: the  $K$ -centroids-based classifier ( $K$ -CB). A comparative study between various supervised classification algorithms on the VisTex and Brodatz image databases is conducted and reveals that the proposed  $K$ -CB classifier obtains relatively good classification accuracy with a low computational complexity.

## 1 Introduction

Que ce soit pour étudier des minéraux, observer l'infiniment petit ou vérifier automatiquement la qualité de pièces manufacturées, l'analyse de texture est une étape importante. La caractérisation de texture est un problème déterminant pour de nombreux systèmes d'analyse d'images. Pour bon nombre de systèmes, il est souvent question de classification supervisée ou non. Lorsqu'il est possible de constituer un corpus d'apprentissage permettant l'émergence de méthodes supervisées, on observe en pratique qu'un corpus présente une variabilité importante et cela à l'intérieur même de chacune des classes constituées. Une telle variabilité, ou diversité - qui découle d'un contenu quasi stationnaire, de changements d'éclairage ou de point de vue (voir figure 1) - est un des freins rencontrés dans les applications de classification, d'indexation ou de synthèse d'images texturées.

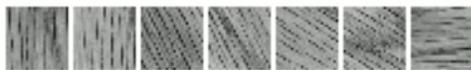


FIGURE 1 – Diversité naturellement présente dans la classe Wood de la base de données Brodatz.

La diversité intra-classe introduit une extension géométrique complexe du cluster en un ensemble de clusters nommé « macro-cluster ». Fournir des approches invariantes à la diversité intra-classe n'est réalisable qu'à travers une caractérisation précise de la géométrie de ce macro-cluster. En classification d'image, le constat précédent impose la sélection d'une approche supervisée qui prend en compte la diversité avec une large base de données d'apprentissage. Les échantillons d'une

classe sont autant d'instances de la diversité intra-classe. La classification d'échantillons peut, sans aucunes notions *a priori* sur le modèle stochastique des classes, proposer une approche basée instances comme un algorithme  $K$ -plus-proche-voisins ( $K$ -NN). Néanmoins, le principal désavantage de la méthode  $K$ -NN est le temps requis pour estimer la classe d'une image de test. Considérons  $N_a$  échantillons d'apprentissage dans un espace de dimension  $d$ , la complexité calculatoire est en  $\mathcal{O}(N_a d)$ . Si les instances sont stockées intelligemment, avec des arbres de décision par exemple, alors la recherche des voisins les plus proches croit avec la décroissance de la dimension  $d$ . Typiquement la dimension est inférieure à 20 ce qui est généralement trop bas en pratique [7].

Dans le domaine de l'analyse de texture, le problème de la diversité a été peu traité. Varma et Zisserman [12] proposent un algorithme  $K$ -moyennes pour partitionner le macro-cluster en un nombre fini de sous-clusters. Cette partition finie permet de réduire la complexité par rapport à celle du  $K$ -NN. Néanmoins le partitionnement est effectué sur des textons avec comme métrique la norme quadratique. Les données décrivent un espace de très grande dimension dans lequel la métrique euclidienne n'est pas adaptée, cela conduit à un algorithme aux performances limitées et implique de travailler avec un important nombre de sous-clusters. Dans le contexte de la modélisation de textures, certains auteurs ont proposé d'utiliser conjointement des approches espace-échelle et des modélisations stochastiques caractérisant le contenu textural [3, 4, 5, 6]. Le processus de modélisation consiste à estimer la densité stochastique paramétrique issue des histogrammes de chaque sous-bande (le nombre de paramètres est fini), introduisant directe-

ment une réduction de la dimension pour la classification. Dans [6], Do et Vetterli considèrent la distribution Gaussienne généralisée (GGD) et montrent également qu'une mesure de similarité stochastique telle que la divergence de Kullback-Leibler – explicitée avec les paramètres stochastiques – peut être utilisée. De même, Choy et Tong proposent dans [3] d'estimer le barycentre de plusieurs vecteurs paramétriques pour chaque sous-bande pour une classe donnée. Même si l'algorithme exploite les propriétés géométriques de la variété, le principal défaut de cette dernière proposition reste que la classification avec un seul barycentre n'est pas suffisamment robuste pour gérer la diversité intra-classe.

Dans ce papier, nous présentons deux de nos contributions. Premièrement, une extension de l'algorithme de Choy en prenant en compte la direction des dérivées partielles afin d'accélérer la convergence de l'estimation d'un barycentre (Karcher mean). Deuxièmement, nous proposons d'utiliser l'algorithme  $K$ -moyennes dans l'algorithme de classification supervisée basée  $K$ -barycentres ( $K$ -CB). Cette classification supervisée (1) innove par rapport aux travaux de Varma et Zisserman qui ne considéraient pas le point de vue paramétrique et (2) améliore l'algorithme de Choy et Tong en utilisant un plus grand nombre de barycentres en accord avec la diversité naturelle présente dans les images texturées. Enfin, des applications numériques du  $K$ -CB montrent un gain en performances par rapport à l'algorithme de Choy et Tong et une complexité calculatoire plus basse que l'algorithme 1-NN.

## 2 Représentation barycentrique

Soit  $I$  une image texturée. Soient  $N_o$  et  $N_s$  le nombre d'orientations et d'échelles de la décomposition multi-échelles. Une texture  $I$  peut ainsi être décomposée en  $N_o \times N_s$  sous-bandes où chaque sous-bande est caractérisée par le vecteur de paramètres  $\lambda_{s,o} = \{\alpha, \beta\}$  associé à la densité de probabilité estimée. La collection  $T$  de ces vecteurs de paramètres représente l'image texturée  $I$ .

$$T = \{\lambda_{s,o} | s = 1, \dots, N_s, o = 1, \dots, N_o\}. \quad (1)$$

Les composants  $\lambda_{s,o}$  du vecteur  $T_I$  forment une variété paramétrique Riemannienne notée  $\mathcal{M}$ .

### 2.1 Calcul d'un barycentre

Soient  $(T_{c,n})_{n=1}^{N_a}$  une collection de  $N_a$  images d'apprentissage d'une même classe de texture  $c$ . Dans [3], Choy et Tong proposent un algorithme itératif permettant de calculer l'image barycentrique  $\bar{T}_c$ . Soit  $l_c(T)$  la fonction de coût définie par :

$$l_c(T) = \frac{1}{N_a} \sum_{n=1}^{N_a} m(T \| T_{c,n}), \quad (2)$$

le barycentre est alors la solution du problème d'optimisation suivant :

$$\bar{T}_c = \arg \min_{T \in \mathcal{M}} l_c(T). \quad (3)$$

La mesure de dissimilarité  $m$  est définie comme une somme pondérée de mesures de dissimilarités SIM définies entre les

distributions de chaque sous-bande :

$$m(T_c^n \| T_{c'}^{n'}) = \sum_{s=1}^{N_s} \sum_{o=1}^{N_o} \text{SIM}(p(x; \lambda_{c,s,o}^n) \| p(x'; \lambda_{c',s,o}^{n'})); \quad (4)$$

où  $p(x; \lambda_{c,n,s,o})$  représente la distribution stochastique qui modélise les coefficients de la sous-bande  $x$  à l'échelle  $s$  et à l'orientation  $o$ .

Dans ce papier, nous proposons d'utiliser l'algorithme du gradient naturel [1, 10] pour résoudre (3). Pour améliorer la vitesse de convergence, la matrice d'information de Fisher  $G(T)$  est utilisée avec le gradient de la fonction coût  $\nabla l_c(T)$ . Si  $\bar{T}_c$  est solution de (3), alors la suite  $(\bar{T}_{c,i})_{i=1}^{\infty}$  définie par :

$$\bar{T}_{c,i+1} = \text{Proj}_{\mathcal{M}}(\bar{T}_{c,i} - G^{-1}(\bar{T}_{c,i}) \nabla l_c(\bar{T}_{c,i})), \quad (5)$$

converge vers le barycentre  $\bar{T}_c$ . L'opérateur  $\text{Proj}_{\mathcal{M}}$  de projection sur la variété  $\mathcal{M}$  assure que le barycentre  $\bar{T}_c$  appartienne à  $\mathcal{M}$ . D'un point de vue pratique, sur la base de données Vis-Tex, l'algorithme de descente de gradient projeté [3] converge en 170 itérations là où l'algorithme de gradient naturel projeté converge en seulement 9 itérations vers la même solution.

Dans la suite, le calcul du barycentre est appliqué avec la distribution Gaussienne généralisée (GGD) puisque ce modèle stochastique a été validé pour modéliser la distribution des coefficients de sous-bande [6, 9]. Notez également que la généralisation à tout modèle stochastique associé à des expressions explicites de la mesure de similarité  $m$  et de la matrice de Fisher est possible.

### 2.2 Application à la GGD

La densité de probabilité d'une GGD univariée s'écrit :

$$p(x; \lambda) = \frac{1}{\alpha \Gamma(1/\beta + 1)} \exp \left\{ - \left( \frac{|x|}{\alpha} \right)^\beta \right\}, \quad (6)$$

avec  $\Gamma(z) = \int_{\mathbb{R}^+} t^{z-1} e^{-t} dt$  la fonction Gamma,  $\alpha$  et  $\beta$  représentent respectivement les paramètres d'échelle et de forme. Dans la suite, l'espace des paramètres pour une sous-bande est représentée par  $\lambda = \{\alpha, \beta\}$ .

La divergence de Jeffrey (*i.e.* la somme des divergences de Kullback-Leibler orientées) est considérée comme mesure de dissimilarité SIM entre deux GGDs, s'exprime [6] :

$$\begin{aligned} \text{JD}(p(x; \lambda) \| p(x; \lambda')) &= A + A' - \frac{1}{\beta} - \frac{1}{\beta'} \\ A &= \frac{\Gamma((\beta' + 1)/\beta)}{\alpha^{-\beta'} (\alpha')^{\beta'} \Gamma(1/\beta)}, \quad A' = \frac{\Gamma((\beta + 1)/\beta')}{(\alpha')^{-\beta} \alpha^\beta \Gamma(1/\beta')}. \end{aligned} \quad (7)$$

La combinaison de (4) et (7), permet l'obtention la mesure de dissimilarité  $m$  entre deux échantillons. Pour dériver la fonction de coût et pour obtenir le gradient  $\nabla l_c(T)$ , les dérivées partielles de la divergence de Jeffrey sont nécessaires :

$$\begin{aligned} \frac{\partial \text{JD}}{\partial \alpha}(\lambda, \lambda') &= \frac{\beta'}{\alpha} A + \frac{\beta}{\alpha} A', \\ \frac{\partial \text{JD}}{\partial \beta}(\lambda, \lambda') &= - \left( A' \left[ \ln \left\{ \frac{\alpha'}{\alpha} \right\} + \frac{1}{\beta'} \Psi \left( \frac{\beta + 1}{\beta'} \right) \right] - \right. \\ &\quad \left. A \left[ \Psi \left( \frac{1}{\beta} \right) - (\beta' + 1) \Psi \left( \frac{\beta' + 1}{\beta} \right) \right] \right) + \frac{1}{\beta^2}, \end{aligned} \quad (8)$$

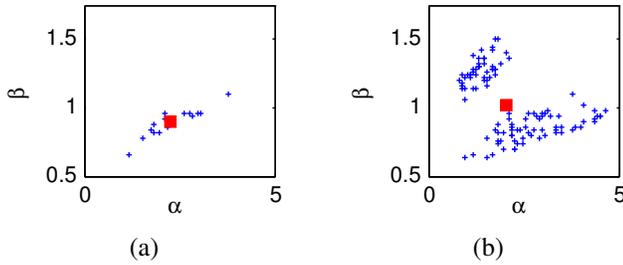


FIGURE 2 – Nuage de points dans le plan  $\alpha/\beta$  : (a) pour des textures à faible diversité intra-classe et (b) pour des textures présentant différentes orientations.

avec  $\Psi(z) = \Gamma'(z)/\Gamma(z)$  la fonction digamma.

Pour l’algorithme de gradient naturel, la matrice d’information de Fisher de taille  $2 \times 2$  doit être explicitée :

$$g_{\alpha\alpha}(\lambda) = \frac{\beta}{\alpha^2}, \quad g_{\alpha\beta}(\lambda) = \frac{1}{\alpha\beta} \left( \Psi\left(\frac{1}{\beta}\right) - \beta + 1 \right),$$

$$g_{\beta\beta}(\lambda) = \frac{\beta + 1}{\beta^4} \Psi'\left(\frac{1}{\beta}\right) + \frac{1}{\beta^2} + \frac{1}{\beta^3} \left[ \Psi^2\left(\frac{1}{\beta}\right) + 2(\beta + 1)\Psi\left(\frac{1}{\beta}\right) \right]. \quad (9)$$

Ainsi, en injectant (8) et (9) dans (5), le barycentre des sous-bandes distribuées selon une GGD peut être estimée par un processus itératif.

### 2.3 Représentation mono-barycentrique : capacités et limites

Dans cette section, plusieurs expériences sont réalisées sur des images texturées afin d’évaluer le potentiel et les limites d’une représentation mono-barycentrique. Fig. 2(a) et (b) montrent une représentation des images d’apprentissage (les croix bleues) dans l’espace de paramètres  $\alpha/\beta$  pour une sous-bande. Le carré rouge montre l’emplacement du barycentre estimé. Pour une classe de texture possédant une faible variabilité intra-classe (cf Fig. 2(a)), le nuage de points est très compact. Le barycentre représente correctement le cluster. Cependant, lorsque la diversité intra-classe est importante, une représentation mono-barycentrique n’est pas suffisante pour expliquer le cluster, comme montrée dans la figure 2.(b) où les images proviennent d’une même classe mais avec différentes orientations dominantes (cf Fig. 1). La diversité naturelle présente dans une classe de texture peut être due à plusieurs raisons telles que des différences d’éclairage de la scène, des différences d’échelle, de perspective, ... Cette diversité modifie la forme du cluster en l’allongeant ou en l’éclatant en plusieurs nuages. Pour gérer au mieux cette diversité intra-classe, une approche multi-barycentres en réadaptant l’algorithme des  $K$ -moyennes aux modèles stochastiques est proposée dans la section suivante.

## 3 La classification supervisée K-CB

### 3.1 Représentation multi-barycentrique

Pour chaque classe  $c = 1, \dots, N_c$ ,  $K$  barycentres  $(\bar{T}_{c,k})_{k=1}^K$  sont calculés à l’aide d’une version réadaptée du classifieur

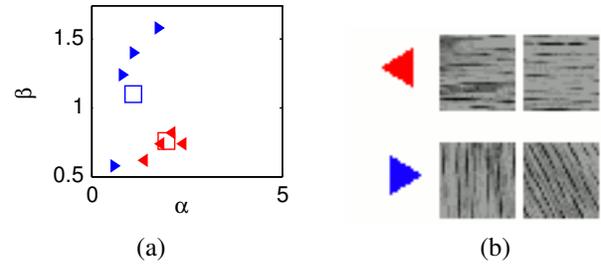


FIGURE 3 – (a) Nuage de points dans le plan  $\alpha/\beta$  pour des textures présentant une forte variabilité intra-classe et (b) exemple d’images pour chacun des clusters.

des  $K$ -moyennes (notée  $K$ -CB) pour prendre en compte la nature stochastique des modèles considérés. Soit  $w_{c,k,n} \in \{0, 1\}$  une série de poids telle que  $w_{c,k,n} = 1$  si et seulement si l’image  $T_{c,n}$  est dans le cluster numéro  $k$ , alors les  $K$  barycentres  $(\bar{T}_{c,k})_{k=1}^K$  sont estimés par minimisation de la fonctionnelle :

$$\sum_{k=1}^K \sum_{n=1}^{N_a} \frac{w_{c,k,n}}{\sum_{i=1}^{N_a} w_{c,k,i}} m(\bar{T}_{c,k} \| T_{c,n}) \quad (10)$$

Soit  $T_t$  une image de texture à classifier. Cette image est labellisée avec la classe  $\hat{c}$ , classe du barycentre le plus similaire, *i.e.*

$$\hat{c} = \arg \min_c \{ m(T_t \| \bar{T}_{c,k}) | \forall k = 1, \dots, K \}. \quad (11)$$

Notez que l’algorithme de Choy et Tong décrit à la section 2 est un cas particulier du  $K$ -CB lorsque  $K = 1$ . De plus, lorsque  $K$  est égal au nombre d’images d’apprentissage  $N_a$ , l’algorithme  $K$ -CB revient à un classifieur au sens du plus proche voisin (1-NN).

Fig. 3(a) montre une représentation des images d’apprentissage (triangle) dans le plan  $\alpha/\beta$  pour une sous-bande. Les images d’apprentissage sont issues de la même classe de texture mais elles ne partagent pas la même orientation. L’algorithme  $K$ -CB est appliqué avec  $K = 2$ . Les carrés bleu et rouge correspondent aux deux barycentres estimés. Dans cette expérience, l’algorithme  $K$ -CB est appliqué sur une variété  $\mathcal{M}$  de dimension 24<sup>1</sup>. Une seule projection est affichée dans la Fig. 3(a). Comme observé dans Fig. 3(b), l’algorithme proposé permet de bien discriminer les clusters selon leur orientation principale. Les deux barycentres sont alors représentatifs de la diversité dans les classes de texture.

### 3.2 Résultats et discussion

Pour évaluer les performances des algorithmes de classification supervisée, la base de données est séparée entre une base d’apprentissage et une base disjointe de test. Deux bases de données sont utilisées : VisTex [8] avec 40 classes et  $N_e = 16$  images par classe ( $128 \times 128$  pixels), et Brodatz [2] avec 13 classes et  $N_e = 112$  images par classe ( $128 \times 129$  pixels). La base VisTex contient des images texturées présentant des

1. Pour une décomposition en pyramides orientées [11] avec deux échelles et 6 orientations de décomposition et deux paramètres ( $\alpha$  et  $\beta$  par sous-bande)

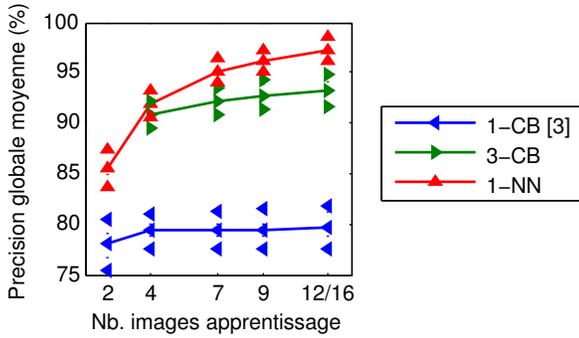


FIGURE 4 – Évolution de la précision globale moyenne en fonction du nombre d’images d’apprentissage sur la base de données VisTex.

différences dans les conditions d’éclairage, tandis que la base Brodatz présente une plus forte variabilité intra-classe causée par des différences dans l’acquisition de l’image comme des rotations. Dans la suite, 100 réalisations Monte-Carlo sont utilisées afin d’évaluer les performances des différents classifieurs (précision globale et barre d’erreurs).

Dans cette première expérience, la décomposition en ondelettes stationnaires (avec 2 échelles de décomposition) et le filtre de Daubechies db4 est utilisé. La figure 4 montre l’évolution de la précision globale moyenne en fonction du nombre d’images d’apprentissage sur la base de données VisTex [8] : 1-NN en rouge (triangle vers le haut), 1-CB [3] en bleu (triangle vers la gauche) et 3-CB en vert (triangle vers la droite). Un gain de 10 points est observé lorsque 3 barycentres sont utilisés à la place d’un seul. De ce fait, la version 3-CB permet une meilleure caractérisation de la diversité intra-classe. Notons également que les performances de l’algorithme 3-CB restent proches de celles du 1-NN. Cependant, la complexité calculatoire est significativement inférieure avec le  $K$ -CB<sup>2</sup>, puisqu’il y a seulement  $K$  calculs de dissimilarité entre l’image requête et les barycentres alors que  $N_a$  calculs sont nécessaires pour le 1-NN. Par conséquent, le  $K$ -CB est un bon compromis entre précision de la classification et complexité calculatoire.

TABLE 1 – Taux de reconnaissance moyen pour différents classifieurs supervisés sur les bases VisTex et Brodatz.

$N_a$	1-NN $K$	1-CB [3] $N_e/2$	$K$ -CB $N_e/2$	1-NN $N_e/2$
VisTex	79 % $\pm$ 2	73 % $\pm$ 2	89 % $\pm$ 2	94 % $\pm$ 1
Brodatz	70 % $\pm$ 3	70 % $\pm$ 1	97 % $\pm$ 2	99 % $\pm$ 1

Le tableau 1 montre la précision globale moyenne pour différents algorithmes de classification supervisée (1-NN, 1-CB et  $K$ -CB) sur les bases de données VisTex et Brodatz (respectivement  $K$  égale 3 et 10 pour chacune des bases de données). Deux classifieurs 1-NN sont considérés, l’un avec seulement  $K$

2. Le calcul des barycentres est effectué pendant la phase d’apprentissage pour les  $K$ -CB.

images d’apprentissage par classe (pour avoir la même complexité calculatoire que le  $K$ -CB) et l’autre avec  $N_a$  images d’apprentissage par classe. Dans cette expérience, la décomposition en pyramides orientées avec deux échelles et 8 orientations de décomposition est utilisée. Comme nous pouvons le voir dans le tableau 1, pour une même complexité calculatoire, la classification  $K$ -CB proposée montrent un gain de 10 et 27 points comparé à une classification 1-NN avec  $N_a = K$ . De ce fait, une sélection adaptée des barycentres avec le  $K$ -moyennes permet d’accroître la précision globale. Notez également que la performance du  $K$ -CB reste très proche de celle du 1-NN avec  $N_a$  images d’apprentissage. De plus, le classifieur  $K$ -CB proposé basé sur un point de vue paramétrique fait bien mieux qu’une approche basée textures telle que l’algorithme proposé par Varma et Zisserman dans [12] qui affiche une précision globale moyenne d’environ 82% sur la base de données VisTex avec 20 barycentres.

## 4 Conclusion

Dans cet article, une classification supervisée  $K$ -barycentres a été introduite pour gérer la diversité naturelle intra-classe des images texturées. Plusieurs résultats de classification sur les bases de données de texture ont montré que le classifieur  $K$ -CB proposé permet l’obtention d’une bonne précision de classification avec une complexité calculatoire relativement basse. Les prochains travaux porteront sur le développement d’un algorithme de classification supervisée avec un nombre de barycentres par classe de texture automatique et adapté (qui dépendra de la diversité intra-classe).

## Références

- [1] S. Amari and S. Douglas. Why natural gradient? In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, volume 2, pages 1213–1216, Mai 1998.
- [2] P. Brodatz. *Textures : A Photographic Album for Artists and Designers*. Dover, New York, 1966.
- [3] S.-K. Choy and C.-S. Tong. Supervised texture classification using characteristic generalized Gaussian density. *Journal of Mathematical Imaging and Vision*, 29 :35–47, Août 31 2007.
- [4] O. Cula and K. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1) :33–60, 2004.
- [5] O. Cula, K. Dana, F. Murphy, and B. Rao. Bidirectional imaging and modeling of skin texture. *Biomedical Engineering, IEEE Transactions on*, 51(12) :2148–2159, 2004.
- [6] M. Do and M. Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *Image Processing, IEEE Transactions on*, 11(2) :146–158, Fév. 2002.
- [7] A. Kibriya and E. Frank. An empirical comparison of exact nearest neighbour algorithms. *Knowledge Discovery in Databases : PKDD 2007*, pages 140–151, 2007.
- [8] MIT Vision and Modeling Group. Vision Texture. Disponible : <http://vis-mod.media.mit.edu/pub/VisTex>.
- [9] M. Novey, T. Adali, and A. Roy. A complex generalized Gaussian distribution – characterization, generation, and estimation. *Signal Processing, IEEE Transactions on*, 58(3) :1427–1433, 2010.
- [10] A. Schutz, Y. Berthoumieu, F. Turcu, C. Nafornita, and A. Isar. Barycentric distribution estimation for texture clustering based on information-geometry tools. In *Electronics and Telecommunications (ISETC), 10th International Symposium on*, Nov. 2012.
- [11] E. Simoncelli, W. T. Freeman, E. Aldelson, and D. Heeger. Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2) :587–607, Mar. 1992.
- [12] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62 :61–81, 2005.