

Interpolation d'attributs ondelettes pour l'indexation de bases d'images satellitaires à différentes résolutions

Bin Luo¹, Jean-François Aujol², Yann Gousseau¹, Saïd Ladjal¹,

¹GET/Télécom Paris, CNRS UMR 5141 CNES-DLR-ENST Competence Centre

²CMLA, ENS Cachan, CNRS, PRES UniverSud

{bin-luo, gousseau, ladjal}@tsi.enst.fr

Jean-Francois.Aujol@cmla.ens-cachan.fr

Résumé – Nous proposons une nouvelle méthode pour comparer des attributs issus de décompositions en ondelettes entre images de différentes résolutions. L'approche repose sur un modèle d'acquisition des images satellitaires permettant de calculer les modifications de ces attributs en fonction des changements de résolution.

Abstract – In this work, we propose a new method to compare wavelet features of images with different resolutions. By taking explicitly into account an acquisition model of satellite images, we compute the behaviour of wavelet features with respect to resolution changes.

1 Motivations

Ces dernières années, des bases d'images satellitaires très volumineuses ont été créées, en particulier par le CNES (bases dont la taille augmente chaque jour de plusieurs centaines giga bits). Dans le but de pouvoir effectuer des recherches sur ces bases, il est crucial de savoir indexer de manière efficace les images qui les composent. Une particularité de ces bases est qu'elles sont constituées d'images provenant de différents satellites et possédant donc des résolutions différentes et connues. En conséquence, et dans l'optique d'extraire les spécificités des scènes physiques capturées par les images, il est souhaitable de développer des attributs invariants par changement de résolution. Il s'agit ici de la spécificité de l'approche d'indexation proposée dans cette communication. La littérature est en effet riche en méthodes pour construire des attributs invariants par changement d'échelles [3, 11, 9], mais ces méthodes ne sont pas directement applicables à notre problème [6].

Plus précisément, nous nous intéressons à l'indexation d'images par attributs issus de décomposition en ondelettes. Sous l'hypothèse d'un changement de résolution relativement simple (convolution par un noyau gaussien puis échantillonnage), nous calculons l'influence de la résolution des images sur les attributs de texture considérés et en déduisons une méthode pour comparer les attributs de deux images de résolutions différentes. Nous validons l'approche sur des bases d'images satellitaires. Des résultats préliminaires de cette étude ont été présentés dans [5].

2 Attributs de texture

De nombreux attributs ont été proposés pour l'indexation des images satellitaires. Nous nous intéressons ici à

des images mono-spectrales. Nous avons choisi d'utiliser des attributs de texture construits à partir de caractéristiques ondelettes : de nombreux travaux ont en effet montré leur efficacité en indexation ou classification d'images [1, 10, 2]. Mallat [8] a montré expérimentalement que l'histogramme $h(u)$ d'une image texturée peut être modélisé efficacement par une distribution gaussienne généralisée (*DGG*) :

$$h(u) = Ke^{-(|u|/\alpha)^\beta} \quad (1)$$

De plus, il a été montré dans [1, 2] que les paramètres α et β des DGG sont de bons attributs pour la classification de textures. Afin de ne pas être limité à des changements de résolution dyadiques, nous utilisons la transformée en ondelettes continue [7] plutôt que la (plus classique) transformée discrète. Motivés par l'invariance par changement de résolution, comme il apparaîtra par la suite, nous choisissons d'estimer les paramètres α et β des DGG à l'aide des moments d'ordre un et deux de $|u|$, notés respectivement $m_1 = \int |u|h(u)du$ et $m_2 = \int u^2h(u)du$. Dans la suite de cet article, on utilise en fait les moments m_1 et m_2 comme attributs, car les paramètres α et β peuvent être calculées directement à partir des moments.

3 Modélisation de la procédure d'acquisition

Dans ce paragraphe, nous introduisons un modèle simplifié d'acquisition des images. Notons $I : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction continue qui représente la scène réelle et $I_r : \mathbb{Z}^2 \rightarrow \mathbb{R}$ sa version discrète à la résolution r . Négligeant un certain nombre de phénomènes tels que le bruit, les vibrations du satellite, le flou de filé ou la réponse du capteur, nous supposons que I_r est obtenue à partir de I par une convolution

gaussienne suivie d'un échantillonnage. Soit

$$I_r = \Pi_r(I * k_{rp}) \quad (2)$$

où Π_r est le peigne de Dirac de période r , k_{rp} le noyau gaussien d'écart-type rp , et p un paramètre caractéristique du système d'acquisition (qui doit être suffisamment élevé pour des images non aliasées). La Section 5 montre que ce modèle, bien que très simple, est expérimentalement satisfaisant sur une base d'images fournies par le CNES.

4 Interpolation

Soit $w_{r,t}(\mathbf{x})$ le coefficient ondelette à l'échelle t de l'image I_r . Comme nous l'avons vu, nous nous intéressons aux moments d'ordre un et deux des marginales de coefficients d'ondelettes, soit $m_1(r,t) = \frac{1}{|\Omega|} \int_{\Omega} |w_{r,t}(\mathbf{x})| d\mathbf{x}$ et $m_2(r,t) = \frac{1}{|\Omega|} \int_{\Omega} (w_{r,t}(\mathbf{x}))^2 d\mathbf{x}$, où Ω est le domaine de l'image. Notons finalement $\Theta(r,t) = \{m_1(r,t), m_2(r,t)\}$ les attributs ondelettes à l'échelle t obtenus à partir de I_r .

Le but est de comparer ces attributs lorsqu'il sont obtenus à partir de deux images de résolutions différentes, c'est à dire de comparer $(\Theta(r_1, t_1))_{t_1 \in \mathbb{R}_+}$ et $(\Theta(r_2, t_2))_{t_2 \in \mathbb{R}_+}$. Si on ne tient pas compte des effets du système optique, et qu'on suppose que I_{r_1} et I_{r_2} ne diffèrent que d'un facteur de zoom, les attributs aux résolutions r_1 et r_2 sont identiques (après normalisation) lorsque

$$t_2 = \frac{r_1 t_1}{r_2} \quad (3)$$

Nous avons cependant vérifié expérimentalement, sur une base d'images multi-résolutions fournie par le CNES, qu'une telle approche n'est pas satisfaisante. C'est pourquoi nous faisons l'hypothèse que I_r est obtenue par convolution avec un noyau gaussien et échantillonnage.

De façon à pouvoir tenir compte de l'étape de convolution, nous considérons des ondelettes particulières construites à partir de dérivées de gaussiennes. Dans ce cas, le coefficient d'ondelette à l'échelle t dans la direction q (où $q = 0, 1, 2, 3$ correspond respectivement aux directions horizontale, verticale et les deux diagonales) pour l'image I_r à la résolution r est :

$$w_{r,t,q}(\mathbf{x}) = I_r * \Delta_q \tilde{k}_t \quad (4)$$

où Δ_q est l'opérateur de dérivation dans la direction q , \tilde{k}_t est le noyau gaussien discret d'écart-type t . Si on suppose $\tilde{k}_t \approx k_{rt}$, on a

$$\begin{aligned} w_{r,t,q}(\mathbf{x}) &= \Delta_q \tilde{k}_t * \Pi_r(I * k_{rp}) \\ &\approx r \Pi_r(\partial_q k_{rt} * k_{rp} * I) \\ &\approx r \Pi_r(k_t \sqrt{r^2 + p^2} * \partial_q I) \end{aligned}$$

Rappelons que

$$\begin{aligned} m_1(r,t) &= \sum |w_{r,t}|/N \\ m_2(r,t) &= \sum (w_{r,t})^2/N \end{aligned}$$

où N est le nombre de pixel. Si $t\sqrt{r^2 + p^2}$ est constant, $m_1(r,t)/r$ et $m_2(r,t)/r^2$ sont donc également constants.

On en déduit le schéma suivant d'interpolation des attributs ondelettes pour les images aux résolutions r_1 et r_2 :

- Calculer les coefficients d'ondelettes pour I_{r_1} aux échelles t_i , $i = 1, 2, 3, \dots, N$;
- Estimer les paramètres Θ_{r_1, t_i} à partir des coefficients d'ondelettes aux échelles t_i pour la résolution r_1 ;
- Pour la résolution r_2 , calculer les échelles t'_i correspondant aux t_i à l'aide de la fonction :

$$t'_i = \sqrt{\frac{r_1^2}{r_2^2}(t_i^2 + p^2) - p^2} \quad (5)$$

- Définir $\tilde{\Theta}_{r_2, t'_i} = \Theta_{r_1, t_i}$ aux échelles t'_i .

Expérimentalement, nous avons constaté que la valeur $p = 1.3$ permettait d'approcher au mieux la fonction de transfert de modulation (FTM) de nos images. Dans le cas où la FTM est connue, ce paramètre peut être fixé automatiquement. Remarquons que si $p = 0$, l'équation (5) devient $t'_i = \frac{r_1}{r_2} t_i$, c'est à dire qu'elle se réduit exactement à l'équation (3) : pour $p = 0$, on ne prend pas en compte la convolution précédant l'échantillonnage.

5 Validation numérique

Sur la Figure 1 sont affichées des scènes sur lesquelles on a les séries d'images à différentes résolutions (de $0.707m$ à $2.5m$). Chaque série d'images provient d'une même scène. Sur ces images sont calculées les valeurs $m_1(r,t)/r$ et $m_2(r,t)/r^2$. Pour les images aux résolution $0.707m$, on extrait les moments à l'échelle 5. Pour les autres résolutions, l'échelle à laquelle sont calculés les moments est déterminée en utilisant la relation (3) (courbes bleues sur la Figure 1) ou la relation (5) (courbes rouges sur la Figure 1). On peut constater que l'utilisation de (5) permet d'obtenir des attributs (ici les moments d'ordre 1 et 2) quasiment invariants par changements de résolution, contrairement aux résultats obtenus en utilisant (3).

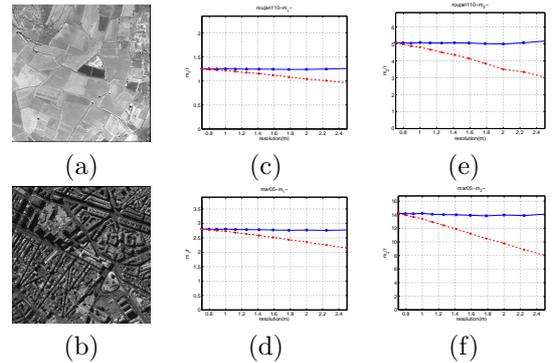


FIG. 1 – (a)-(b) Deux images (©CNES); (c)-(d) graphes de $m_1(r,t)/r$ en fonction de r , et (e)-(f) graphes de $m_2(r,t)/r^2$ en fonction de r . Les courbes bleues montrent les valeurs obtenues en utilisant la relation (5), et les courbes rouges montrent les valeurs obtenues en utilisant la relation (3).

La Figure 2 montre les résultats d'interpolation obtenus en utilisant la méthode présentée dans la Section 4. Les résolutions sont respectivement $r_1 = 1.0m$ et $r_2 = 3.175m$. Nous avons également comparé les résultats obtenus quand $p = 1.3$ et $p = 0.0$. On peut constater que lorsque l'on prend en compte la convolution pour le changement de résolution, l'interpolation des moments est beaucoup plus précise.

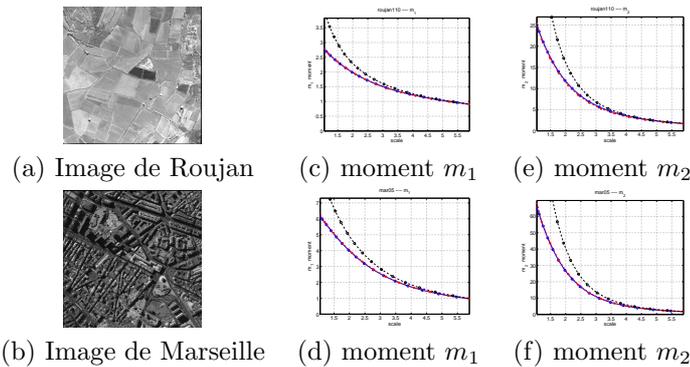


FIG. 2 – Résultats d'interpolation de m_1 et m_2 , de la résolution $1m$ à $3.175m$. Les lignes en tirets représentent les résultats d'interpolation obtenus en utilisant la méthode présentée dans la Section 4 avec $p = 0.0$ (i.e. en utilisant la relation (3)). Dans chaque figure, il a aussi deux lignes solides qui sont parfaitement superposées. L'une représente le résultat d'extrapolation obtenu quand $p = 1.3$. L'autre représente les moment m_1 et m_2 calculés directement sur l'image de résolution $3.175m$ (les valeurs de référence).

6 Expériences

Base de données : A partir d'une base d'images fournies par le CNES, on a construit une base d'images contenant 366 scènes de différents types (urbaines, rurales, forestières, maritimes). Pour chaque scène, on dispose de 5 images à 5 résolutions différentes : $0.5m$, $1m$, $2m$, $3.175m$ et $4m$. On étiquette les images manuellement selon 4 classes : villes (199 images), champs (134 images), forêts (23 images) et mer (10 images). Quelques exemples de ces images sont montrés sur la Figure 3.

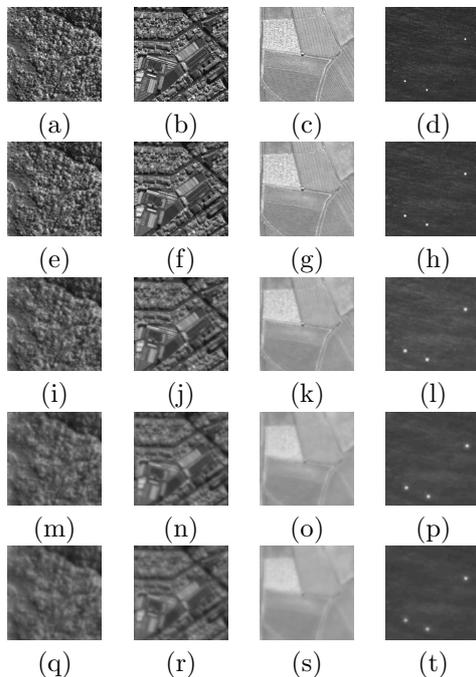


FIG. 3 – Exemples de la base de données des images multi-résolutions : (a)-(d) Images à la résolution $0.5m$; (e)-(h) Images à la résolution $1.0m$; (i)-(l) Images à la résolution $2.0m$; (m)-(p) Images à la résolution $3.175m$; (q)-(t) Images à la résolution $4.0m$;

Classification La première expérience sur cette base de données est la classification. Les images à la résolution $4m$ sont utilisées comme ensemble d'apprentissage pour entraîner l'algorithme de classification. L'objectif est de

trouver l'étiquette des images aux résolutions autres que $4m$. Les primitives (attributs) (moments m_1 et m_2 des coefficients d'ondelettes) sont tout d'abord extraites pour toutes les images dans les 4 directions (horizontale, verticale et les deux diagonales), à 21 échelles ($t = \{2^{i/6} | i = 0, 1, \dots, 20\}$). La dimension des primitives est donc $2 \times 4 \times 21 = 168$. Les primitives extraites des images aux résolutions autres que $4m$ sont ensuite interpolées à $4m$ en utilisant la méthode présentée dans la Section 4. Pour des raisons de comparaison, le paramètre p est respectivement fixé à $p = 0.0$ (correspondant au schéma sans prise en compte du modèle d'acquisition) et $p = 1.3$ (correspondant au schéma avec prise en compte du modèle d'acquisition). On effectue ensuite la classification des images aux résolutions autres que $4m$ à l'aide des primitives interpolées.

L'algorithme de classification utilisé est celui du plus proche voisin. La mesure utilisée est la distance euclidienne. Pour une image à résolution autre que $4m$, l'algorithme de classification lui donne l'étiquette de l'image de l'ensemble d'apprentissage (à la résolution de $4m$) qui est son plus proche voisin dans l'espace des primitives.

On compare également les résultats de classification avec les primitives de Haralick [4]. Les primitives de Haralick sont calculées à partir des matrices de co-occurrence de l'image. La matrice de co-occurrence est construite à partir de la distribution jointe des niveaux de gris entre deux pixels dans une certaine direction et avec une certaine distance d . Cette distance entre deux pixels peut aussi être considérée comme un paramètre d'échelle. Pour des images aux différentes résolutions, il est possible de calculer des matrices de co-occurrence à différentes échelles. Par exemple, dans notre expérience, on utilise $d = 3, 4, 6, 12, 24$ respectivement pour les images aux résolutions $4m, 3.175m, 2m, 1m$ et $0.5m$ afin d'assurer que $d \times r$ reste constant. Les primitives de Haralick sont ici 13 valeurs statistiques calculées sur les matrices de co-occurrence [4]. La dimension totale des primitives de Haralick est donc $(4+2) \times 13 = 78$.

La Figure 4 montre les erreurs de classification en utilisant l'interpolation des primitives ondelettes avec $p = 1.3$ et $p = 0.0$ et les erreurs obtenues en utilisant les primitives de Haralick.

Résolution	$p = 1.3$	$p = 0.0$	Haralick
0.5m	0.00%	21.86%	31.9%
1m	0.00%	19.95%	22.4%
2m	0.00%	10.66%	13.6%
3.175m	0.00%	1.91%	10.1%
4m	0.00%	0.00%	0.00%

FIG. 4 – Erreurs de classification

On peut constater qu'avec l'interpolation sans prise en compte du modèle d'acquisition (i.e. $p = 0$), l'erreur de classification augmente d'autant plus que la résolution de l'image s'éloigne de $4m$. Par contre lorsque l'on prend en compte le modèle d'acquisition ($p = 1.3$), il n'y a pas d'erreur. Ceci est dû à la précision de l'interpolation (voir la Section 5). Les faibles erreurs lors de l'interpolation des attributs ondelettes ne sont pas de nature à perturber la

classification. Même si l'algorithme de classification n'a pas trouvé exactement la même scène, il parvient à trouver une scène suffisamment similaire pour attribuer une étiquette correcte.

De plus, la performance de classification par les primitives de Haralick est mauvaise par rapport aux résultats obtenus par des primitives ondelette (même sans prise en compte du modèle d'acquisition). Nous avons vérifié numériquement que cette propriété restait vraie en diminuant la taille des primitives ondelettes (en ne prenant que 3 échelles au lieu de 21). Le point important ici est de noter que la simple prise en compte du facteur de zoom d'une résolution à l'autre sur les primitives de Haralick n'est pas suffisante pour atteindre de bonnes performances de classification. Or dans ce cas, ces primitives étant non linéaires, il est difficile de prendre en compte l'étape de convolution de l'acquisition des images.

Mise en correspondance Il s'agit ici une expérience plus exigeante que la précédente. Etant donnée une image à résolution autre que $4m$, on ne cherche plus simplement à retrouver la classe de l'image (ville, mer, forêt, champs, ...), mais la scène exacte à résolution $4m$ dont provient l'image considérée. Autrement dit, on peut considérer que chaque scène a une étiquette unique. On utilise le même algorithme de classification (plus proche voisin) et les mêmes primitives que précédemment. Pour des raisons de comparaison, on utilise toujours deux valeurs de p ($p = 0$ et $p = 1.3$ qui correspond à la prise en compte du noyau de convolution).

La Figure 6 montre le pourcentage des images pour lesquelles on n'a pas trouvé les mêmes scènes en utilisant 21 échelles ($t = \{2^{i/6} | i = 0, 1, \dots, 20\}$) et 4 directions d'ondelettes. On peut constater que l'erreur augmente considérablement par rapport aux résultats de la classification présentés dans la Figure 4, surtout pour $p = 0.0$. Ceci est dû au fait qu'il y a beaucoup de scènes très similaires qui appartiennent à une même classe (ville, mer, forêt, champs), ce qui implique qu'une petite erreur d'interpolation peut conduire à une mauvaise mise en correspondance. Pour $p = 1.3$, l'erreur de mise en correspondance est beaucoup moins grande, ce qui confirme la précision de l'interpolation lorsque le modèle d'acquisition est pris en compte.

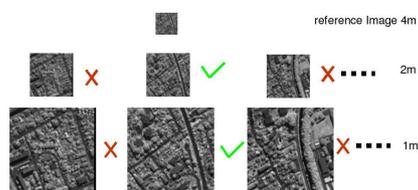


FIG. 5 – Exemple de mise en correspondance

résolution	0.5	1.0	2.0	3.175
p=1.3	4.1	0.27	0	1.64
p=0	81.97	79.51	65.57	34.97

FIG. 6 – Pourcentage d'erreur de mise en correspondance avec des primitives d'ondelette extraites dans 4 directions et sur 21 échelles.

7 Conclusion

Cet article présente une nouvelle approche pour indexer des images satellitaires indépendamment de leur résolution. En tenant compte explicitement d'un processus simplifié d'acquisition des images nous calculons l'effet du changement de résolution sur des attributs issus de décompositions en ondelettes de type dérivées de gaussiennes, ce qui permet de proposer une procédure de comparaison des attributs entre images de résolutions différentes. Cette procédure est appliquée à la classification d'images et à la mise en correspondance. Des expériences sont en cours pour montrer la robustesse de la méthode par rapport au nombre d'attributs utilisés. Nous étudions par ailleurs l'influence de la FTM de l'instrument optique sur la modélisation utilisée. Nous envisageons également l'application de l'approche à d'autres modalités d'imagerie, en particulier l'imagerie microscopique.

Références

- [1] Aujol, J.F., G. Aubert et L. Blanc-Féraud: *Wavelet-based level set evolution for classification of textured images*. IEEE Transactions on Image Processing, 12(12) :1634–1641, 2003.
- [2] Do, Minh N. et Martin Vetterli: *Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance*. IEEE Transactions on Image Processing, 11(2) :146–158, Février 2002.
- [3] Fan, Fernand S. Cohen Zhigang et M. A. Patel: *Classification of Rotated and Scaled Textured Images Using Gaussian Markov Random Field Models*. IEEE PAMI, 13(2) :192–202, Février 1991.
- [4] Haralick, R.M., K. Shanmugam et I. Dinstein: *Textural Features for Image Classification*. IEEE Transactions on Systems Man and Cybernetics, SMC-3(6), 1973.
- [5] Luo, B., J.F. Aujol, Y. Gousseau et S. Ladjal: *Interpolation of wavelet features for satellite images with different resolutions*. Dans *IGARSS 2006*, Denver USA, 2006.
- [6] Luo, B., J.F. Aujol, Y. Gousseau, S. Ladjal et H. Maître: *Characteristic scale in satellite images*. Dans *ICASSP 2006*, Toulouse France, 2006.
- [7] Mallat, S.G.: *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [8] Mallat, Stéphane G.: *A Theory for Multiresolution Signal Decomposition : The Wavelet Representation*. IEEE PAMI, 11(7) :674–693, Juillet 1989.
- [9] Pun, C. M. et M. C. Lee: *Log-Polar Wavelet Energy Signatures for Rotation and Scale Invariant Texture Classification*. IEEE PAMI, 25(5) :590–603, 2003.
- [10] Wouwer, Gert Van de: *Wavelets for Multiscale Texture Analysis*. Thèse de doctorat, Universiteit Antwerpen, 1998.
- [11] Zhang, Jianguo et Tieniu Tan: *Affine Invariant Texture Signatures*. Dans *ICIP 2001*, Thessaloniki Grèce, 2001.