

Une distance de style stochastique pour la synthèse de textures multispectrales

Sélim OLLIVIER^{1,2} Yann GOUSSEAU¹ Sidonie LEFEBVRE²

¹LTCI, Télécom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France

²DOTA/AILab, ONERA, Université Paris-Saclay, 91120 Palaiseau, France

Résumé – Nous proposons dans cet article une *distance de style* multispectrale reposant sur un réseau de neurones pré-entraîné sur des images naturelles à 3 canaux. Elle consiste en une évaluation stochastique d’une *distance de style* classique sur les images formées par des triplets aléatoires de bandes spectrales. Elle permet une extension simple et efficace de l’état de l’art des méthodes de synthèse de texture aux données multispectrales, sans nécessiter l’entraînement d’un réseau dédié.

Abstract – In this paper, we propose a novel multispectral *style distance* that relies on a RGB network. It consists in the stochastic evaluation of a classical *style distance*, over images formed by random triplets of spectral bands. It constitutes a simple and efficient extension of state-of-the-art methods for multispectral texture synthesis, while avoiding the additional training of a multispectral network

1 Introduction

Ces dernières années, de nombreux efforts ont été consacrés au développement et à la conception d’instruments multispectraux, capables de capturer des images d’une scène dans plusieurs bandes spectrales. D’autre part, l’entraînement et l’évaluation des algorithmes de détection et de classification nécessitent d’avoir accès à une grande quantité de données, pour permettre de prendre en compte la variabilité des objets d’intérêt mais aussi des fonds, qui sont souvent texturés. En raison des coûts d’acquisition élevés, l’augmentation de données par la synthèse réaliste d’images à l’aide de réseaux profonds présente un réel intérêt.

Les modèles de texture ont d’abord été élaborés pour des images mono-chromatiques, en utilisant des outils comme les matrices de co-occurrence [6] ou des transformées en ondelettes, afin de prendre en compte le comportement spatial des images. Des extensions aux images couleur et multispectrales ont été proposées en utilisant des réductions de dimensions [4] et différentes stratégies pour prendre en compte l’information de couleur et les corrélations inter-canaux [8].

Depuis les travaux de Gatys et al. [2], les réseaux de neurones convolutifs (CNN) se sont imposés comme l’état de l’art en synthèse de texture. L’idée principale de ces travaux est d’utiliser des statistiques issues des cartes de caractéristiques profondes du réseau. À partir d’une image exemple, une nouvelle image peut être générée à partir d’un bruit blanc de sorte à minimiser l’écart quadratique entre les matrices de Gram des images. On définit ainsi une *distance de style*, qui permet de générer des images réalistes qui conservent l’aspect visuel des images originales tout en autorisant une répartition spatiale différente. Cependant, la procédure originale d’optimisation des pixels de l’image est coûteuse. C’est pourquoi de nombreux travaux se sont attachés à accélérer la synthèse en proposant des architectures de réseaux dédiés *feed-forward* [10][1], tout en conservant la notion centrale de *distance de style* pour l’entraînement des réseaux.

Dans la littérature, les modèles génératifs pour l’image-

rie multispectrale ont récemment évolué vers des méthodes d’apprentissage profond [3], aboutissant à des résultats de synthèse satisfaisants pour les images satellitaires. Cependant, à notre connaissance, aucune méthode de synthèse de texture s’appuyant sur les statistiques des représentations d’un réseau pré-entraîné n’a encore été proposée pour les images multispectrales. La principale difficulté vient du fait que le réseau VGG19, qui est le plus utilisé dans la littérature, est conçu pour et a été entraîné sur des images RGB à 3 canaux. Certains travaux ont tenté de contourner ce problème avec des réseaux aléatoires [11], mais les résultats demeurent de qualité inférieure. De plus, d’autres architectures, qui fonctionnent pourtant mieux en classification, ne parviennent pas à égaler VGG en synthèse de texture et en transfert de style [12]. En outre, l’entraînement d’un réseau dédié à des données multispectrales demanderait d’avoir accès à de larges quantités de données pour la combinaison de bandes d’intérêt.

Nous proposons donc une nouvelle *distance de style* multispectrale, qui permet de conserver un réseau entraîné sur des images à 3 canaux comme VGG19. Celle-ci repose sur l’évaluation d’une *distance de style* classique sur des images à 3 canaux formées par des triplets de bandes spectrales tirés aléatoirement au cours de la procédure d’optimisation. Cette méthode peut être utilisée pour une vaste gamme de textures, mais nous concentrons ici nos expériences sur les nuages. Notre approche conduit à des résultats de synthèse satisfaisants visuellement et quantitativement, qui surpassent une méthode naïve consistant à utiliser une réduction de dimension. Une première version de ce travail ainsi qu’une partie des résultats expérimentaux de ce papier ont été présentés dans [5].

2 Méthode

2.1 Synthèse par optimisation

Comme mentionné plus haut, les cartes de caractéristiques profondes extraites de CNN sont cruciales dans les méthodes

de synthèse de texture. Ces représentations sont habituellement issues d'un réseau VGG19 entraîné sur la base de données ImageNet. Chaque couche l du réseau est constituée de filtres de convolution qui fournissent des activations F_l . Les matrices de Gram $G_l = F_l F_l^T$ de ces activations constituent ainsi un modèle de texture stationnaire et multi-échelle. La *distance de style* entre deux images I et $\hat{I}$, est ensuite définie comme la somme des écarts quadratiques entre les matrices de Gram G et $\hat{G}$, pondérée par des poids w_l :

$$\mathcal{L}_{style}(I, \hat{I}) = \sum_l w_l \|G^l - \hat{G}^l\|_2^2 \quad (1)$$

Cette distance constitue l'objectif de minimisation dans les travaux de [2]. A partir d'une image de bruit blanc, une descente de gradient est utilisée pour synthétiser une image ayant des matrices de Gram correspondantes à une image exemple.

2.2 Distance de style stochastique

La difficulté de l'extension de l'approche rappelée dans le paragraphe précédent aux images multispectrales réside dans la contrainte des statistiques d'une image de dimension N à l'aide d'une procédure d'optimisation qui s'appuie sur un réseau RGB. Pour ce faire, nous proposons de contraindre les statistiques de toutes les images à 3 canaux formées par des triplets de bandes spectrales issues des N bandes.

On considère une image exemple I et une image synthétique $\hat{I}$ à N bandes spectrales. À chaque itération du processus d'optimisation de [2], nous proposons de tirer un triplet aléatoire J d'entiers de $\{1, \dots, N\}$ et d'évaluer la *distance de style* sur les images formées par les bandes spectrales indexées par J , à savoir I_J et $\hat{I}_J$. Ce faisant, on s'assure de balayer toutes les bandes spectrales et les corrélations inter-bandes au cours de l'optimisation.

Par ailleurs, l'objectif global de l'optimisation est l'espérance de la *distance de style* entre les images formées par des triplets tirés uniformément. Comme le nombre de triplets possibles est fini, cela revient à moyenner cette quantité sur tous les triplets possibles :

$$\mathcal{L}_{style}^{MS}(I, \hat{I}) = \mathbb{E}_{J \sim \mathcal{U}(\mathcal{J}_N)} [\mathcal{L}_{style}(I_J, \hat{I}_J)] \quad (2)$$

$$= \frac{1}{N_J} \sum_{J \in \mathcal{J}_N} \mathcal{L}_{style}(I_J, \hat{I}_J) \quad (3)$$

où $\mathcal{J}_N$ est l'ensemble des entiers de $\{1, \dots, N\}$, de cardinal N_J , et $\mathcal{U}(\mathcal{J}_N)$ la loi uniforme sur $\mathcal{J}_N$.

2.3 Procédure de synthèse

Nous proposons de mettre en œuvre la fonction de perte introduite au paragraphe précédent dans le cadre de la procédure de synthèse proposée par [2]. A la différence de la procédure originale, nous choisissons ici d'utiliser les matrices de covariance et un bruit gaussien paramétré par les statistiques de couleurs (μ, Σ) de l'image exemple, ce qui conduit à de meilleurs résultats.

Il est également possible d'évaluer et de moyenner cette distance de style sur des batchs de triplets aléatoires au cours de

l'optimisation, de façon similaire aux algorithmes de descente de gradient stochastique.

L'algorithme 1 décrit la procédure de synthèse pour n pas de descente de gradient et B triplets tirés à chaque itération.

Algorithme 1 : Synthèse de texture par optimisation stochastique d'une distance de style

```

 $\hat{I} \sim \mathcal{N}(\mu, \Sigma)$ 
pour  $i \in \{1, \dots, n\}$  faire
     $\mathcal{L} \leftarrow 0$ 
    pour  $j \in \{1, \dots, B\}$  faire
         $J \sim \mathcal{U}(\mathcal{J}_N)$ 
         $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{L}_{style}(I_J, \hat{I}_J)$ 
    fin
    Faire un pas de descente  $\hat{I}$  avec le gradient  $\nabla \mathcal{L}$ 
fin
retourner  $\hat{I}$ 

```

3 Résultats

3.1 Données

Nous travaillons avec un jeu de données constitué d'images Sentinel-2 de fonds nuageux. La mission d'observation de la Terre Sentinel-2 acquiert des images dans 13 bandes spectrales dans le visible et l'infrarouge, dont 3 bandes de correction atmosphérique. Nous constituons un premier jeu de données S2-11B en excluant bande de correction B10 en raison d'un pourcentage d'absorption particulièrement élevé ainsi que la bande B8A compte tenu de son chevauchement avec la bande B8. Notre jeu de données est ainsi constitué de 130 images de fonds nuageux, de résolution 512x512 sur 11 bandes spectrales. Nous considérons par ailleurs les images constituées des 9 bandes d'observation au sein d'un second jeu de données S2-9B.

3.2 Méthodologie

Nous réalisons la synthèse de texture pour les images à 11 bandes en utilisant la distance de style introduite plus haut, à l'aide de l'algorithme 1. Nous utilisons les matrices de covariance des activations d'un réseau VGG19 avec les poids normalisés de [2] pour le modèle de texture. La descente de gradient est réalisée à l'aide de l'algorithme L-BFGS pour 500 itérations. A chaque itération, 10 triplets aléatoires sont tirés uniformément.

3.3 Métriques

Nous souhaitons synthétiser des textures à partir d'une image exemple. Il est attendu que les images générées respectent les statistiques, les propriétés et l'aspect visuel de l'image exemple, tout en présentant une répartition spatiale différente, contrairement à la reconstruction pixel à pixel plus traditionnelle. Nous détaillons ci-dessous les différentes métriques utilisées pour mesurer la qualité de la synthèse, en notant I l'image exemple et $\hat{I}$ l'image synthétique.

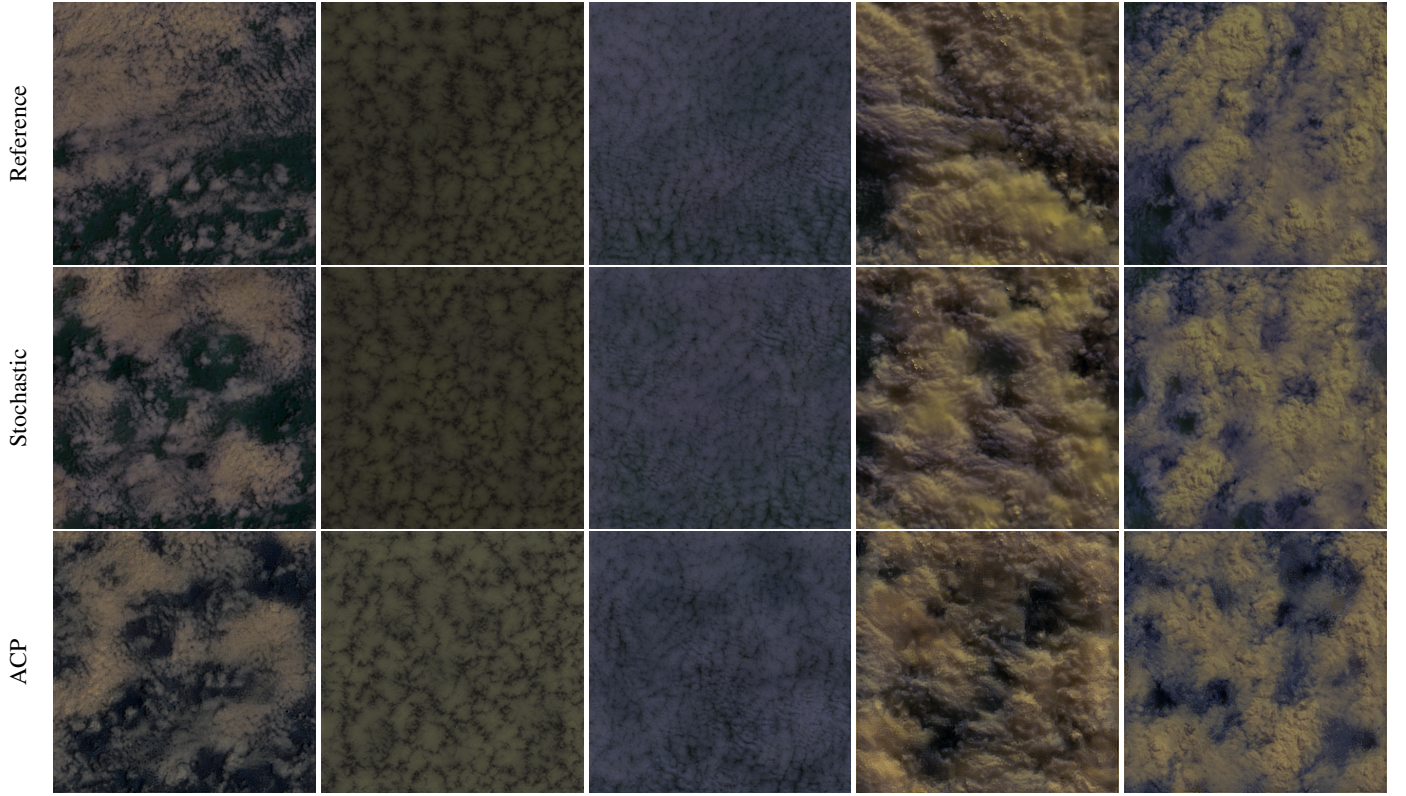


FIGURE 1 : Exemples de synthèse de texture multispectrale. Les images présentées sont le résultat d'un moyennage par groupe de bandes spectrales (canal 1 : bandes 1,2,3,4; canal 2 : bandes 5,6,7,8; canal 3 : bandes 9,11,12). La colorimétrie de certaines images a également été ajustée pour une meilleure visualisation.

3.3.1 Style

Nous utilisons la *distance de style* $\mathcal{L}_{style}^{MS}$ calculée exactement en utilisant l'ensemble des 990 triplets possibles.

3.3.2 Distribution multispectrale

De même que la couleur est une caractéristique essentielle de la description des textures RGB, la combinaison des intensités dans les différentes bandes est essentielle en imagerie multispectrale. Une image synthétique satisfaisante doit respecter la distribution des pixels multispectraux de l'image originale. La distance de Wasserstein est un moyen efficace de comparer les distributions multidimensionnelles. Comme son calcul est très coûteux en dimension supérieure à deux, nous utilisons une distance de Wasserstein par tranche SW_2 [7] :

$$\mathcal{L}_{hist}(I, \hat{I}) = SW_2(I, \hat{I}) \quad (4)$$

On utilise 1000 directions aléatoires afin de calculer cette distance de Wasserstein par tranche.

3.3.3 Pente spectrale

Les fonds nuageux présentent un comportement fractal [9] qui peut être décrit par la décroissance du module du spectre de Fourier azimuthé. Pour une image I , on donne :

$$\mathcal{F}_{rad}(I)(r) = \frac{1}{N_r} \int_0^{2\pi} \|\mathcal{F}(I)(r \cos \theta, r \sin \theta)\| d\theta \quad (5)$$

avec $\mathcal{F}$ l'opérateur de la Transformée de Fourier.

La différence entre les spectres des deux images en niveaux de gris $\bar{I}$ et $\hat{I}$ est évaluée comme suit :

$$\mathcal{L}_{sp}(I, \hat{I}) = \left\| \log(\mathcal{F}_{rad}(\bar{I})) - \log(\mathcal{F}_{rad}(\hat{I})) \right\|_2 \quad (6)$$

3.3.4 Gradients

La dernière caractéristique clé dans le cadre de notre application aux fonds nuageux est la distribution des gradients [9]. À partir des gradients horizontaux $\nabla_x I$ et verticaux $\nabla_y I$, on obtient la magnitude des gradients sous la forme d'une image multispectrale $|\nabla I|$. Le respect des gradients est ensuite évalué à l'aide de la distance de Wasserstein par tranche qui permet de mesurer l'écart entre les distributions :

$$\mathcal{L}_{grad}(I, \hat{I}) = SW_2(|\nabla I|, |\nabla \hat{I}|) \quad (7)$$

3.4 Résultats visuels

Des exemples de synthèse sont présentés en Figure 1. Les images synthétiques sont visuellement satisfaisantes. Des résultats de synthèse issus d'une méthode plus naïve consistant à utiliser les 3 premières composantes d'une analyse en composantes principales (ACP) de chaque image pour se ramener à 3 canaux sont également présentés à titre de comparaison [5]. On observe un meilleur respect des distributions multispectrales dans les images synthétisées à l'aide de la distance de style stochastique. En outre, elle permet également d'éviter des artefacts de hautes fréquences qui sont présents dans les images synthétisées à l'aide d'une ACP.

3.5 Résultats quantitatifs

Des résultats quantitatifs sont donnés dans le Tableau 1. Nous comparons là encore notre méthode à l'utilisation d'une ACP. Notre méthode permet une amélioration de toutes les métriques évaluées ici.

Données	Méthode	$\mathcal{L}_{style}^{MS} \downarrow$	$\mathcal{L}_{sp} \downarrow$	$\mathcal{L}_{grad}(\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist}(\cdot 10^{-3}) \downarrow$
S2-9B	Stochastique	0.09	1.72	12.46	9.19
	ACP	404.54	2.01	26.24	9.48
S2-11B	Stochastique	0.17	1.96	13.52	12.63
	ACP	115.38	2.33	26.52	15.02

TABLE 1 : Évaluation de la synthèse de texture par optimisation stochastique.

On compare également les résultats obtenus sur les 3 bandes spectrales RGB en réalisant l'optimisation sur toutes les bandes spectrales à l'aide de la distance de style stochastique introduite à ceux qui auraient été obtenus en réalisant l'optimisation uniquement sur ces 3 canaux (Tableau 2). La distance de style stochastique permet des performances proches ou supérieures à la référence selon les métriques considérées.

Données	Méthode	$\mathcal{L}_{style}^{RGB} \downarrow$	$\mathcal{L}_{sp}^{mean} \downarrow$	$\mathcal{L}_{grad}(\cdot 10^{-3}) \downarrow$	$\mathcal{L}_{hist}(\cdot 10^{-3}) \downarrow$
S2-9B	RGB	0.019	2.03	28.53	14.09
	Stochastique	0.059	1.58	9.32	9.75
	ACP	565.62	1.94	20.81	9.07
S2-11B	RGB	0.019	2.03	28.53	14.09
	Stochastique	0.099	1.68	9.33	12.41
	ACP	85.15	2.09	20.55	13.02

TABLE 2 : Évaluation de la synthèse de texture par optimisation stochastique sur les bandes RGB. La ligne RGB correspond aux résultats obtenus en optimisant uniquement sur les bandes 3 RGB bandes. Toutes les métriques sont calculées sur ces 3 bandes.

4 Conclusion

Nous avons proposé une nouvelle distance de style permettant d'étendre les méthodologies usuelles de synthèse de texture à l'imagerie multispectrale et qui permet de conserver un réseau VGG19 pré-entraîné sur des images naturelles à 3 canaux. Notre approche permet d'obtenir des performances très satisfaisantes lors des expériences réalisées sur des images satellitaires de fonds nuageux.

Cette méthode ne requiert pas de réentraînement additionnel du réseau extracteur d'attributs et pourra donc aisément être utilisée pour l'entraînement de modèles génératifs *feed-forward*, par exemple.

Nous pensons par ailleurs que ce travail offre des perspectives intéressantes quant à l'utilisation de métriques perceptuelles devenues standards, comme LPIPS ou FID, pour l'imagerie multispectrale.

Références

[1] Pierrick CHATILLON, Yann GOUSSEAU et Sidonie LEFEBVRE : A geometrically aware auto-encoder for multi-

texture synthesis. *In Scale Space and Variational Methods in Computer Vision*, pages 263–275, 2023.

- [2] Leon GATYS, Alexander S ECKER et Matthias BETHGE : Texture synthesis using convolutional neural networks. *In Advances in Neural Information Processing Systems*, volume 28, 2015.
- [3] Liqin LIU, Bowen CHEN, Hao CHEN, Zhengxia ZOU et Zhenwei SHI : Diverse hyperspectral remote sensing image synthesis with diffusion models. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16.
- [4] G. MERCIER et M. LENNON : On the characterization of hyperspectral texture. *In IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 2584–2586, 2002.
- [5] Sélim OLLIVIER, Yann GOUSSEAU et Sidonie LEFEBVRE : Multispectral style distances and application to texture synthesis using rgb convolutional neural networks. *In 2024 14th Workshop on Hyperspectral Imaging and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, 2024.
- [6] Alice POREBSKI, Mohamed ALIMOUSSA et Nicolas VANDENBROUCKE : Comparison of color imaging vs. hyperspectral imaging for texture classification. *Pattern Recognition Letters*, 161:115–121, 2022.
- [7] Julien RABIN, Julie DELON et Yann GOUSSEAU : Regularization of transportation maps for color and contrast transfer. *In ICIP 2010*, page 1933, 2010.
- [8] Subhadip SARKAR et Glenn HEALEY : Hyperspectral texture synthesis using histogram and power spectral density matching. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2261–2270, 2010.
- [9] F. SZCZAP, Y. GOUR, T. FAUCHEZ, C. CORNET, T. FAURE, O. JOURDAN, G. PENIDE et P. DUBUISSON : A flexible three-dimensional stratocumulus, cumulus and cirrus cloud generator (3dcloud) based on drastically simplified atmospheric equations and the fourier transform framework. *Geoscientific Model Development*, 7(4):1779–1801, 2014.
- [10] Dmitry ULYANOV, Vadim LEBEDEV, Andrea VEDALDI et Victor LEMPITSKY : Texture networks : feed-forward synthesis of textures and stylized images. *In Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML'16) - Volume 48*, pages 1349–1357, 2016.
- [11] Ivan USTYUZHANINOV, Wieland BRENDDEL, Leon GATYS et Matthias BETHGE : What does it take to generate natural textures? *In International Conference on Learning Representations*, 2017.
- [12] Pei WANG, Yijun LI et Nuno VASCONCELOS : Rethinking and improving the robustness of image style transfer. *In The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.