

Dématriçage basé sur un modèle du motif Sparse3

Matthieu MULLER¹ Daniele PICONE¹ Mauro DALLA MURA^{1,2} Magnus ORN ULFARSSON³

¹Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

²Institut Universitaire de France (IUF), France

³University of Iceland, Faculty of Electrical and Computer Engineering, 101 Reykjavik, Iceland

Résumé – Dans les caméras usuelles, les pixels sont placés sous une couche de filtres spectraux suivant un motif spatial pré-défini. Le dématricage consiste à restaurer une image multispectrale à partir de l’acquisition du capteur, retrouvant les informations spectrales perdues par les filtres. Cependant cette opération est souvent spécifique aux motifs RVB communs comme les motifs de Bayer ou de Quad-Bayer, laissant de côté d’autres motifs moins connus mais commercialement utilisés. Ce travail présente un cadre général pour le dématricage de tels motifs en utilisant un modèle prédéfini. Des résultats préliminaires sont aussi présentés, appliqués au motif Sparse3.

Abstract – In conventional color cameras, the pixels are overlaid with spectral filters with a predefined spatial pattern. Demosaicking consists in reconstructing a RGB image from the raw acquisition, finding the spectral information discarded by the filters. However, this operation is often tailored for the most common RGB patterns such as Bayer and Quad-Bayer, leaving behind less common spectral filter arrays that are used in commercial cameras nonetheless. In this work, we present a generalized framework for demosaicking such patterns, with a model-based method. Some preliminary experimental results are also presented, which apply to the restoration of acquisitions of Sparse3 camera.

1 Introduction

Dans bon nombre de domaines différents comme l’imagerie médicale, la télédétection et d’autres, une acquisition spectrale (qui peut être simplement RVB par exemple) est réalisée en deux étapes : l’acquisition de l’image brute et le dématricage, comme présenté dans la figure 1. La première étape consiste à mesurer l’intensité lumineuse sur la matrice des capteurs (les pixels), ce qui donne l’image brute, qui est une image scalaire. Cependant chaque pixel ne va capter qu’une couleur grâce à la Matrice de Filtres Colorés (MFC) ou Color Filter Array (CFA) en anglais. Cette matrice est un motif de filtres colorés qui est répété sur les capteurs, de sorte que chaque pixel est couvert par un filtre.

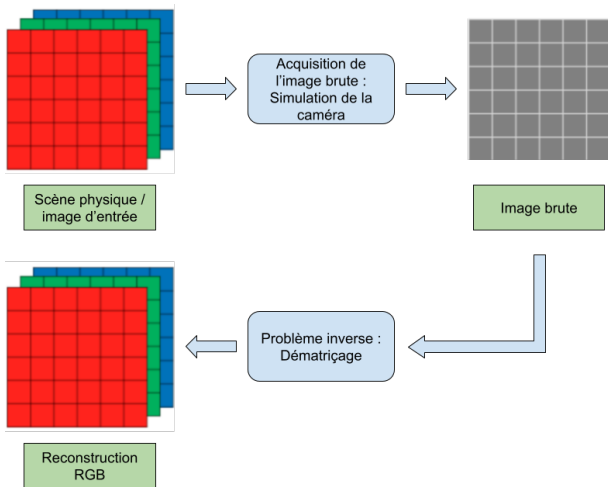


FIGURE 1 : Le processus d’une prise de vue classique, comprenant l’acquisition et la restauration.

Habituellement la MFC est prédéterminée. Bien que les motifs les plus connus soient spécifiques aux couleurs RVB (motif de Bayer [8], [16]), d’autres existent suivant les applications. Certains ont des filtres dans les rayonnements non-visibles (proche infrarouge par exemple), ou ont des filtres autres que les traditionnels rouge, vert et bleu. Nous nous intéressons particulièrement au motif Sparse3, qui est utilisé dans le projet AMICal Sat [2], pour étudier les aurores boréales depuis l’espace. Ce motif a la particularité de contenir une majorité de pixels sans filtres (le pixel capte donc toute la lumière entrante, qui sera uniquement filtrée par la réponse spectrale du détecteur), ce qui permet de conserver un bon rapport signal sur bruit pour les capteurs dans des zones peu éclairées. Les deux motifs, Bayer et Sparse3, sont présentés dans la figure 2.

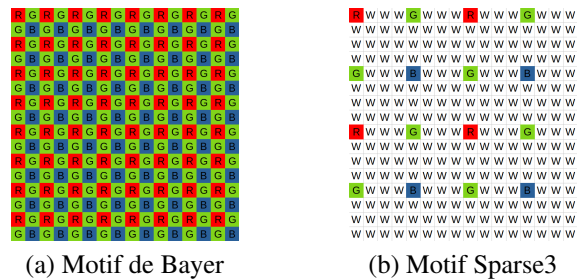


FIGURE 2 : Le motif classique de Bayer et le motif Sparse3 utilisé dans ce travail.

La deuxième étape, le dématricage, cherche à restaurer l’image en retrouvant les couleurs perdues à cause des filtres. Cette opération numérique peut être menée de différentes façons : par interpolations entre les pixels [8], par des méthodes basées sur des modèles physiques pré-définis, faisant le lien entre la scène physique en entrée de la caméra et l’acquisition

brute. Plus récemment des méthodes basées sur l'apprentissage profond sont apparues [11, 6].

Néanmoins la plupart de ces méthodes ont des inconvénients : les méthodes par interpolations ne marchent que pour certaines MFCs et sont des solutions ad-hoc. Les méthodes basées sur un modèle demandent une connaissance précise à priori du système réel. Cette connaissance est en-capsulée dans le modèle direct, représentant le processus d'acquisition de la caméra, comme celui présenté dans [9]. Cependant il n'est pas possible de coller parfaitement à la réalité. Finalement les méthodes par apprentissage ont besoin d'énormes quantités de données pour leur entraînement, qui ne sont pas aisément disponibles pour de petits systèmes. De plus ces entraînements sont spécifiques à une MFC particulière. Par exemple l'article [11] propose un réseau de neurones efficace pour le dématricage de la MFC Bayer mais ses performances ne sont pas garanties pour d'autres MFC telles que la MFC Sparse3. Il en va de même pour les travaux [16], [8] et [13] qui ne traitent que du cas du motif de Bayer.

C'est pour ces raisons que ce travail propose un modèle d'acquisition ainsi qu'une méthode générale pour le dématricage, capable de résoudre le cas du motif Sparse3 qui est utilisé dans le satellite AMICal Sat. Cette méthode est basée sur le modèle proposé et minimise une fonction objectif convexe et régularisée.

2 Modèle et méthode proposés

Cette section présente le modèle direct simulant le processus d'acquisition, puis la méthode basée sur ce modèle pour restaurer une image colorée. L'image de référence, représentant la scène physique, est encodée par $\mathbf{X} \in \mathbb{R}^{M \times N \times (C+1)}$ où C est le nombre de canaux spectraux ($C = 3$ pour une image RVB), un canal est rajouté pour l'information panchromatique. M et N représentent les dimensions spatiales. L'acquisition (image brute) est encodée dans $\mathbf{Y} \in \mathbb{R}^{M \times N}$. Il s'agit de l'image scalaire obtenue en sortie du modèle direct. Enfin nous notons par $\hat{\mathbf{X}} \in \mathbb{R}^{M \times N \times C}$ qui représente la restauration RVB à partir de \mathbf{Y} .

2.1 Le modèle direct

Le modèle direct représente la formation d'image. Il est inspiré par l'article [9], où les auteurs présentent un tel modèle linéaire dans un cadre différent. Une MFC est représentée par un masque $\mathbf{H} \in \mathbb{R}^{M \times N \times (C+1)}$, qui encode le motif voulu. Comme l'image entrante dans le système est représentée par \mathbf{X} , la sortie du modèle (donc l'application de la MFC) est l'image brute \mathbf{Y} qui suit la relation :

$$\mathbf{Y} = \sum_{k=1}^{C+1} \mathbf{X}_k \odot \mathbf{H}_k + \mathbf{N} \quad (1)$$

où \odot est le produit de Hadamard entre deux matrices, \mathbf{X}_k et \mathbf{H}_k sont des matrices représentant le k -ième canal de couleur des tenseurs \mathbf{X} et \mathbf{H} respectivement. Le tenseur \mathbf{H} qui encode le masque a des 1 aux positions où l'information est gardée, le reste des valeurs sont mises à 0, écartant les données à ces positions. L'image brute \mathbf{Y} est alors composée de valeurs réelles, où pour chaque pixel seule l'information du canal coloré choisi par le filtre est gardée. Finalement $\mathbf{N} \in \mathbb{R}^{M \times N}$

est un tenseur de bruit gaussien. Ce dernier permet de rendre compte du bruit physique présent sur le capteur de l'appareil. De plus il représente dans une moindre mesure l'écart entre le modèle direct et la réalité de la caméra qu'il doit représenter.

Dans le cas d'un pixel sans filtre comme dans le cas du motif étudié, il faut que l'entrée du modèle \mathbf{X} possède un canal panchromatique, qui va être utilisé pour les pixels à large bande. Par exemple le cas d'une image RVB donne $\mathbf{X} \in \mathbb{R}^{M \times N \times (3+1)}$, avec les trois bandes RVB classiques plus la bande panchromatique. Si cette bande n'est pas connue elle est approchée par une combinaison linéaire des autres bandes.

La partie déterministe de l'opération faite par une MFC, décrite suivant le modèle de eq. (1), est une opération linéaire notée $A : \mathbb{R}^{M \times N \times (C+1)} \rightarrow \mathbb{R}^{M \times N}$, il devient alors possible d'exprimer l'opération par :

$$\mathbf{Y} = A(\mathbf{X}) + \mathbf{N}. \quad (2)$$

L'opération adjointe du modèle direct sera très utile pour le dématricage présenté dans la section 2.2. Comme l'opération étudiée ici est purement réelle l'opération adjointe est simplement l'adjoint de A . Cependant pour des raisons d'optimisation il est préférable d'utiliser l'expression suivante :

$$\mathbf{Y}_k^{adj} = \mathbf{H}_k \odot \mathbf{Y}, \quad \forall k \in \{1, \dots, (C+1)\} \quad (3)$$

où $\mathbf{Y}^{adj} \in \mathbb{R}^{M \times N \times (C+1)}$. Cette opération correspond à projeter l'image brute \mathbf{Y} dans l'espace $\mathbb{R}^{M \times N \times (C+1)}$, en plaçant l'information de \mathbf{Y} dans le canal d'indice correspondant.

2.2 Dématricage par modèle direct

Le problème inverse du dématricage est un problème mal posé. En effet il y a plus d'inconnues (les valeurs de $\hat{\mathbf{X}}$) que de variables connues (les valeurs de \mathbf{Y}). L'objectif est donc de minimiser une fonction objectif régularisée :

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} (\|\mathbf{A}(\mathbf{X}) - \mathbf{Y}\|_F^2 + \lambda \|L(\mathbf{X})\|_{221}) \quad (4)$$

où $\|\cdot\|_F$ est la norme de Frobenius, $\lambda \|L(\cdot)\|_{221}$ est le terme de régularisation : L est un opérateur matriciel de projection dans un domaine où la régularisation est imposée (ici la variation totale, représentée par des différences finies), λ le poids associé au terme de régularisation et $\|\cdot\|_{221}$ représente la norme ℓ_2 sur les canaux colorés et les gradients, suivie de la norme ℓ_1 sur le domaine spatial. Ce terme est inspiré des travaux [9] et [15], où de bons résultats expérimentaux sont présentés. L'image restaurée $\hat{\mathbf{X}}$ a donc quatre canaux, mais seuls les trois correspondant aux couleurs RVB sont conservés.

La méthode itérative de Chambolle-Pock [5] est un algorithme classique pour résoudre des problèmes convexes de minimisation représentés par une somme de deux fonctions convexes. Le problème inverse de dématricage peut donc être résolu par cet algorithme, et peut être ré-écrit sous la forme :

$$\min_{\mathbf{X}} f(\mathbf{X}) + g(L(\mathbf{X})) \quad (5)$$

où f et g correspondent à $\|\mathbf{A}(\cdot) - \mathbf{Y}\|_F^2$ et à $\lambda \|\cdot\|_{221}$ introduits par eq. (4). En s'appuyant sur les travaux [10] et [5] l'algorithme de Chambolle-Pock consiste à itérer Q fois les deux opérations suivantes :

$$\begin{cases} \mathbf{X}^{q+1} &= \mathbf{prox}_{\tau, f}(\mathbf{X}^q - \tau L^T(\mathbf{Z}^q)) \\ \mathbf{Z}^{q+1} &= \mathbf{prox}_{\sigma, g^*}(\mathbf{Z}^q + \sigma L(2\mathbf{X}^{q+1} - \mathbf{X}^q)) \end{cases} \quad (6)$$

où $q \in \{1, \dots, Q\}$ est le compteur d'itérations, g^* est la fonction conjuguée de g , τ et σ sont les paramètres des opérateurs proximaux de f et de g^* respectivement. Les expressions algébriques des opérateurs \mathbf{prox} de l'algorithme nécessitent de pouvoir calculer efficacement l'opération adjointe de la MFC, ce qui est permis par eq. (3)

3 Résultats expérimentaux

Cette section présente nos premiers résultats, sur le cas spécifique du motif Sparse3. Nous comparons notre méthode à trois autres approches. La première est une méthode basée sur l'apprentissage profond, nommée PIPNet [11]. Cette méthode, basée sur l'architecture UNet donne de bons résultats pour le dématricage du motif de Bayer. La seconde méthode à laquelle nous comparons nos résultats est celle fournie par e2v, le constructeur du capteur Onyx, une caméra qui utilise le motif Sparse3. Cette méthode est présentée dans la notice d'utilisateur [4]. Il s'agit d'un problème de pansharpening [14], l'idée est de fusionner la luminance de l'image brute (avec une grande résolution spatiale mais en nuances de gris) avec une image colorée extraite de l'image brute (de faible résolution spatiale mais colorée). Enfin la troisième méthode est un algorithme dit Plug-and-Play [7, 12]. Il s'agit d'une modification de l'algorithme de Chambolle-Pock utilisé dans notre méthode, l'opérateur $\mathbf{prox}_{\sigma, g^*}$ est remplacé par le débruiteur BM3D [3].

Pour choisir les hyper-paramètres de notre méthode (λ , τ et σ) nous utilisons le module Python Optuna ([1]), qui permet une recherche de ces paramètres plus fine que des tests empiriques. Grâce à cette automatisation nous avons utilisé les paramètres suivants : $\lambda = 0.001$, $\sigma = 0.04$ et $\tau = \frac{0.9}{\sigma \|\mathbf{L}\|^2}$. L'algorithme a été lancé avec ces paramètres avec $Q = 400$ itérations. Augmenter le nombre d'itérations au delà n'apporte pas de gain significatif à la restauration. Enfin nous présentons nos résultats avec un bruit gaussien d'écart type 0.05 qui simule le bruit thermique retrouvé sur les capteurs réels. Pour augmenter le réalisme de nos expériences la table 1 a été obtenue sur tout le jeu de données CAVE-Multispectral [17]. Ce jeu de données est composé de 32 images de dimensions $512 \times 512 \times 31$. Ces images de 31 canaux spectraux ont été transformées en des images à quatre canaux (RVB plus panchromatique) à l'aide des filtres de réponses spectrales réelles.

PSNR (en dB)	Sans bruit	Avec bruit (std=0.05)
e2V [4]	28.59 ± 3.18	24.22 ± 1.60
PIPNet [11]	-18.11 ± 1.57	-18.06 ± 1.59
Plug-and-Play	30.48 ± 3.43	26.94 ± 1.71
Proposition	33.75 ± 4.19	30.03 ± 2.83

TABLE 1 : La moyenne et l'écart type du PSNR en décibels pour les quatre méthodes étudiées (meilleur en gras).

La figure 3 présente l'image brute acquise avec le motif Sparse3 via le modèle direct, avec et sans bruit gaussien. On



FIGURE 3 : Résultats du modèle direct. L'image d'entrée (gauche), l'image brute (centre) et l'image brute avec bruit gaussien (droite).

remarque le motif, notamment dans la partie rouge de l'image (il est recommandé d'agrandir la figure).

La figure 4 présente les résultats des trois méthodes comparées pour trois images différentes (du jeu de données CAVE). Bien que le réseau de neurones PIPNet soit efficace pour restaurer des images avec un motif de Bayer, il échoue à restaurer une image avec le motif Sparse3. Au contraire les trois autres méthodes proposent des restaurations plausibles, surtout la méthode proposée qui est moins sujette aux artefacts.

La méthode d'e2v a l'avantage d'être très rapide à implémenter mais elle comporte deux inconvénients principaux. Premièrement la méthode varie suivant les paramètres de la caméra. elle ne marche que pour le motif Sparse3. Deuxièmement cette approche est dépendante des interpolations utilisées, bien que de meilleures interpolations amènent de meilleurs résultats, le coût de l'algorithme augmente avec. Les méthodes basées sur les réseaux de neurones nécessitent entraînement lourd pour chaque cas spécifique. L'algorithme de Plug-and-Play fournit des résultats concluants mais systématiquement en dessous de la méthode proposée. Cela montre que le choix du débruiteur (ici BM3D) est crucial dans la méthode de Plug-and-Play et que tous les débruiteurs ne peuvent pas convenir suivant les applications.

La méthode proposée, basée sur un modèle direct et un algorithme d'optimisation convexe bien connu, permet d'obtenir des résultats satisfaisants dans des situations peu communes comme celle du motif Sparse3. Un changement dans le processus d'acquisition de l'image brute n'a des conséquences que dans l'expression de l'application A , le reste de la méthode reste inchangée. Cela permet une bonne adaptation à une grande variété de problèmes pour un coût minimal.

4 Conclusion

Pour conclure cet article présente un modèle direct capable de simuler l'action d'une caméra avec le motif Sparse3 ainsi qu'une méthode basée sur la variation totale pour résoudre le problème de dématricage de ce motif. Cette méthode devrait être aisément applicable à d'autres motifs car seul le modèle direct encode l'information du problème en lui-même.

Références

- [1] T. AKIBA, S. SANO, T. YANASE, T. OHTA et M. KOYAMA : Optuna. *In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019.

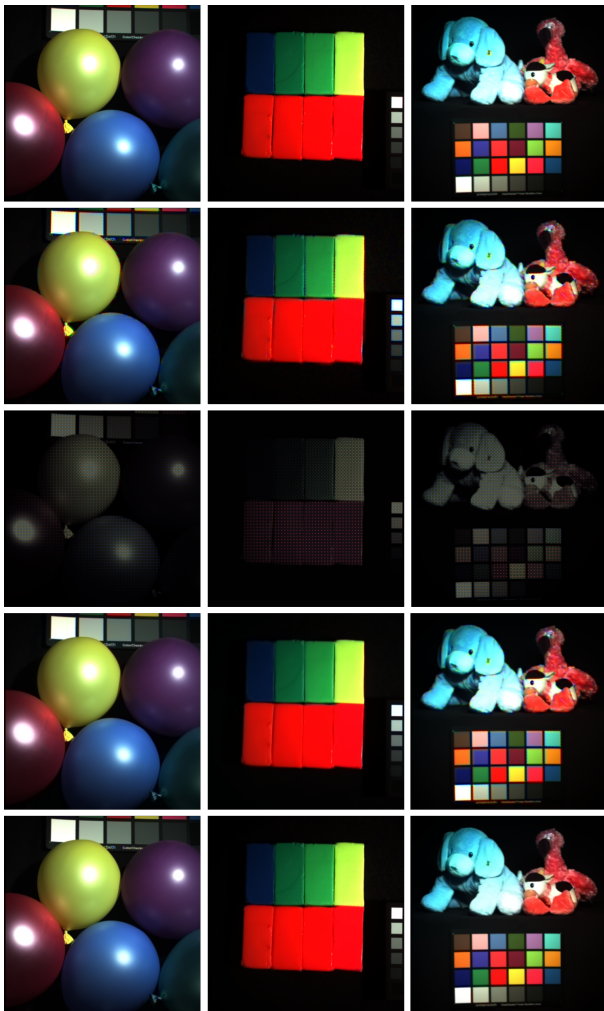


FIGURE 4 : Exemples de reconstruction. L'image d'entrée (1ère ligne), la restauration d'e2v (2ème ligne), la restauration de PIPNet (3ème ligne), la restauration Plug-and-Play (4ème ligne) et restauration proposée (5ème ligne).

- [2] M. BARTHELEMY, E. ROBERT, V. KALEGAEV, V. GRENNERAT, T. SEQUIES, G. BOURDAROT, E. LE COARER, J.J. CORREIA et P. RABOU : Amical sat : A sparse rgb imager on board a 2u cubesat to study the aurora. *IEEE Journal on Miniaturization for Air and Space Systems*, 3(2):36–46, 2022.
- [3] A. DANIELYAN, V. KATKOVNIK et K. EGIAZARIAN : Bm3d frames and variational image deblurring. *IEEE Transactions on Image Processing*, 21(4):1715–1728, 2012.
- [4] Teledyne E2V : Application note : How to interpolate the sparse 3 onyx sensor, 2018.
- [5] E. ESSER, X. ZHANG et T. F. CHAN : A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.
- [6] K. FENG, Y. ZHAO, J. C.-W. CHAN, S. G. KONG, X. ZHANG et B. WANG : Mosaic convolution-attention network for demosaicing multispectral filter array images. *IEEE Transactions on Computational Imaging*, 7:864–878, 2021.
- [7] U. S. KAMILOV, C. A. BOUMAN, G. T. BUZZARD et B. WOHLBERG : Plug-and-play methods for integrating physical and learned models in computational imaging : Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97, 2023.
- [8] J. S. J. LI et S. RANDHAWA : Colour filter array demosaicking using cubic spline interpolation. In *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, volume 1, pages I–865–I–868, 2007.
- [9] D. PICONE, M. DALLA MURA et L. CONDAT : Joint demosaicing and fusion of multiresolution coded acquisitions : A unified image formation and reconstruction method. *IEEE Transactions on Computational Imaging*, 9:335–349, 2023.
- [10] T. POCK et A. CHAMBOLLE : Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *International Conference on Computer Vision*. IEEE, 2011.
- [11] S M A SHARIF, R. ALI NAQVI et M. BISWAS : Beyond joint demosaicking and denoising : An image processing pipeline for a pixel-bin image sensor. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 233–242, 2021.
- [12] Y. SUN, B. WOHLBERG et U. S. KAMILOV : An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3):395–408, 2019.
- [13] Y. UEKI, T. YAMAGUCHI et M. IKEHARA : Adaptive guided upsampling for color image demosaicking. In *European Signal Processing Conference (EUSIPCO)*, pages 2240–2244, 2018.
- [14] G. VIVONE, M. DALLA MURA, A. GARZELLI, R. RESTAINO, G. SCARPA, M. O. ULFARSSON, L. ALPARONE et J. CHANUSSOT : A new benchmark based on recent advances in multispectral pansharpening : Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):53–81, 2021.
- [15] Y.-W. WEN, Michael K. NG et Y.-M. HUANG : Efficient total variation minimization methods for color image restoration. *IEEE Transactions on Image Processing*, 17(11):2081–2088, 2008.
- [16] T. YAMAGUCHI et M. IKEHARA : Image demosaicking via chrominance images with parallel convolutional neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1702–1706, 2019.
- [17] F. YASUMA, T. MITSUNAGA, D. ISO et S.K. NAYAR : Generalized Assorted Pixel Camera : Post-Capture Control of Resolution, Dynamic Range and Spectrum. Rapport technique, Nov 2008.